International Statistical Review (2006), **74,** 2, 215–233, Printed in Wales by Cambrian Printers

A Comparison of Nonparametric Methods in the Graduation of Mortality: Application to Data from the Valencia Region (Spain)

Ana Debón[1], Francisco Montes[2] and Ramón Sala[2]

[1]*Universidad Politécnica de Valencia, E-46022, Valencia, Spain. E-mail: andeau@eio.upv.es*
[2]*Universitat de València, Spain*

Summary

The nonparametric graduation of mortality data aims to estimate death rates by carrying out a smoothing of the crude rates obtained directly from original data. The main difference with regard to parametric models is that the assumption of an age-dependent function is unnecessary, which is advantageous when the information behind the model is unknown, as one cause of error is often the choice of an inappropriate model. This paper reviews the various alternatives and presents their application to mortality data from the Valencia Region, Spain. The comparison leads us to the conclusion that the best model is a smoothing by means of Generalised Additive Models (GAM) with splines. The most interesting part of this paper is the development of a plan that can be applied to mortality data for a wide range of age groups in any geographical area, allowing the most appropriate table to be chosen for the data in hand.

Key words: GAM; Kernel smoothing; Life tables; LOESS; Splines.

1 Introduction

Nonparametric graduation aims to obtain new values from the original data, in which any influence that does not come from the predictor variable is eliminated and, contrary to parametric graduation, a function that expresses the relationship between death rates and age is not provided.

Although this paper only applies graduation to the probability of death, q_x, the results can be extended to the force of mortality at age x, μ_x or m_x central mortality rates at age x.

Parametric and nonparametric methods can complement each other in certain cases, in the sense that the result of nonparametric graduation can more adequately describe the type of equation to be used. Moreover, nonparametric methods can be used to carry out a diagnosis of parametric models or simply to examine data.

Univariate smoothing techniques are compared in an extensive simulation in Breiman & Peters (1992).

A review of smoothing life tables can be found in Wang *et al.* (1998) and Wang (2005). They provide rigorous asymptotic results and show that the transformation approach is supported by both asymptotic and simulation results.

The representation of mortality data by means of nonparametric models attracted the attention of actuaries, demographers and statisticians throughout the past century. Nielsen (2003) reviews papers on smoothing and prediction, most of these are papers with theoretical considerations; however, applications to actuarial science, biostatistics and finance are also discussed. Another review of these methods, though only applicable to medicine, can be found in Zhang (2004).

This paper intends to review the various nonparametric methods, all of which are based on smoothing techniques. The forerunner of these methods was moving averages (Hoem & Linnemann, 1988; Benjamin & Pollard, 1992), a method that is not dealt with in this paper due to the fact that this concept has since been clearly improved upon by other alternatives. In section 2, the main methods are described: kernel smoothing, splines, locally-weighted regression (LOESS) and GAM. In section 3, the various nonparametric methods are applied to mortality data from the Valencia Region (Spain), specifically to the three-year period from 1999–2001. The last section is dedicated to a comparison of how well each method fits and the subsequent conclusions that can be drawn.

2 Nonparametric Methods

We consider a set of mortality rates in the form of life tables. We wish to produce smoother estimates, $\hat{q}_x$, of the true but unknown mortality probabilities q_x from the set of crude mortality rates, $\dot{q}_x$, for each age x. The crude rate at age x_i is typically based on the corresponding number of deaths recorded, d_i, relative to initial exposed to risk, E_i.

Nonparametric methods obtain an estimate, $\hat{q}_x$, at a particular age, x, by means of the weighted averages of crude mortality rates of neighbouring ages, $\{x_1, x_2, \ldots, x_r\}$. The differences among nonparametric estimates mainly lie in the number of points in the average r, and the differences in weights. Habitually, the size of the neighbour is represented by the bandwidth, b, the maximum distance from age x to any age of the neighbour, x_i, so $|x_i - x| \leq b$.

The power of modern computers makes all these techniques considerably more accessible and easier to apply (see the software `S-plus` and `R`, for example). In the following subsections, each one of these techniques is discussed in more detail.

2.1 Kernel Smoothing

This technique is a generalisation of the moving averages method used in temporal series to soften them. It is a flexible technique, initially developed to estimate density functions. The first application of this method to graduation is found in Copas & Haberman (1983). To achieve its objective, this method uses a weighted average with weights based on a kernel function, K, according to the equation

$$\omega_{x_i} = \frac{K\left(\frac{x - x_i}{b}\right)}{\sum_{i=1}^{r} K\left(\frac{x - x_i}{b}\right)}, \tag{1}$$

where parameter b is the bandwidth. Different K function choices result in different types of kernel estimations: Normal, Gaussian, Triangular, Parzen and Epanechnikov's are the best known and, particulary, with the latter providing the best results (Epanechnikov, 1969).

Gavin *et al.* (1993) describes and compares two kernel estimators of q_x that are attributed to Copas & Haberman (1983),

$$\hat{q}_x^{CH} = \frac{\sum_{i=1}^{r} d_i K\left(\frac{x - x_i}{b}\right)}{\sum_{i=1}^{r} E_i K\left(\frac{x - x_i}{b}\right)}, \tag{2}$$

and Nayadara–Watson (Nayadara, 1964; Watson, 1964),

$$\hat{q}_x^{NW} = \frac{\sum_{i=1}^{r} \left(\frac{d_i}{E_i}\right) K\left(\frac{x - x_i}{b}\right)}{\sum_{i=1}^{r} K\left(\frac{x - x_i}{b}\right)}, \tag{3}$$

respectively, in which no transformation whatsoever is considered. Both kernel methods are discussed in more detail, together with the criteria for bandwidth choice, in Gavin *et al.* (1993, 1994, 1995) and Verrall (1996).

One problem that arises with this technique is bias in the estimation of the upper and lower bands. Various solutions to avoid this problem are discussed in Gavin *et al.* (1995) and Verrall (1996). The former defines and discusses a general adapting kernel estimation with a bandwidth that varies according to age and studies the convenience of transforming the data. The solution proposed by Verrall (1996) is a data transformation to which a straight line may be fitted. The literature relating to the use of this technique leads us to conclude that the choice of kernel function K does not have as much influence as the value of the smoothing parameter b (Gasser & Müller, 1979; Gasser *et al.*, 1991; Brockman *et al.*, 1993; Montenegro, 2001).

Felipe *et al.* (2001) propose a nonparametric smoothing method to visualize the evolution of mortality rates in Denmark and Spain. The smoothing method is based on a two-dimensional kernel. This methodology has also been applied by Guillen *et al.* (2006) for studying the evolution of mortality rates in different countries and Fledelius *et al.* (2004) in the study of Swedish old-age mortality.

2.2 *Splines Smoothing*

The method proposed by Whittaker (1923) and Henderson (1924), looks for the estimator of q_x, $\hat{q}_x$, by minimizing

$$W = \sum_{j=1}^{n} w_j(\dot{q}_j - \hat{q}_j)^2 + \lambda \sum_{j=1}^{n-z} (\Delta^z \hat{q}_j)^2, \tag{4}$$

a criteria that combines fitting and smoothness and where w_j are the weights, q_j denotes q_{x_j}, λ is the smoothing parameter, Δ is the difference operator defined by

$$\Delta q_j = q_j - q_{(j-1)}$$

and Δ^z is the difference operator applied z times. This graduation is expressed as a generalised dynamic linear model by Verrall (1993a, l994).

The method of Whittaker–Henderson is a precedent of the splines smoothing method (Wang, 2005). Really, when we consider $\hat{q}_x = f(x)$, a continuous function, $w_j = 1$ and $z = 2$, this criterion becomes the penalized sum of squares defined by

$$\sum_{j=1}^{n} (\dot{q}_j - f(x_j))^2 + \lambda \int_{x_1}^{x_n} (f''(t))^2 dt, \tag{5}$$

where the first term measures proximity by means of the squares of the differences between the original and adjusted values. This term is penalized by the second one, which increases with fluctuations of the $f(x)$. It can be shown that (5) has a unique minimizer, which is a natural cubic spline with knots at the values of x_j, $j = 1, \ldots, n$ (Hastie *et al.*, 2001). The parameter λ determines the relative importance of the terms and has the same role as the bandwidth in kernel estimation.

We have chosen cubic splines for two reasons, because they minimise (5) and use the smallest number of parameters possible (the principle of parsimony). This supposes that the nodes must coincide with the ages, x_j, $j = 1, \ldots, n$, the smooth, spline function available in `S-plus` and in `R` works on this assumption. A cubic spline with knots $x_1, x_2, \ldots, x_n$ is a cubic polynomial whose expression is

$$s(x) = s_j(x) = \sum_{l=0}^{3} a_l^{(j)}(x_{j+1} - x)^l, \quad j = 1, \ldots n - 1$$

in each interval $[x_j, x_{j+1}]$ $j = 1, \ldots n - 1$. This minimization problem is reformulated by Chan *et al.* (1986) and they show how this graduation can be formulated as a linear programming problem in some cases and a quadratic programming problem in others.

Detailed information on smoothing splines can be found in Hastie *et al.* (2001). In a recent article, Kaishev *et al.* (2004) present a new algorithm that allows us to choose the number and partition of the knots.

The use of this spline graduation begins with McCutcheon (1980, 1981) and in McCutcheon (1987), it is applied to English Life Tables ELT No. 14. Later, Forfar *et al.* (1988) provide a brief introduction to spline graduation. As far as Spain is concerned, Betzuen (1997) uses this model to graduate data for a set of active people.

In this section, we describe a variant of the cubic spline method that does not require the optimal choice of knots, smoothing splines. Haberman (1997) gives a brief summary of the method as applied to English Life Tables ELT No. 15.

2.3 Locally-weighted Regression (LOESS)

The locally-weighted regression smoother was introduced by Cleveland (1979) and is discussed at length in Fan & Gijbels (1996). This method locally (in neighbourhoods) adjusts polynomials of low degree, whose estimates, $\hat{q}_x$, are obtained by the regression of $\dot{q}_x$ over age x by means of weighted least squares in each neighbourhood $N(x)$. ω_{x_i} weightings are assigned to each point of $N(x)$, using the tricubic weighting function

$$\omega_{x_i} = T\left(\frac{|x - x_i|}{\Delta(x)}\right)$$

where

$$T(u) = \begin{cases} (1 - u^3)^3, & \text{for } 0 \le u \le 1 \\ 0, & \text{in other case,} \end{cases}$$

and $\Delta(x) = \max_{x_i \in N(x)} |x - x_i|$.

The amount of smoothness is determined by the span parameter, the proportion of the number of elements in the neighbourhood in comparison with the total number of points. The relative importance of each neighbour in the estimate depends on its distance at point x, thereby obtaining a better fit to valleys and peaks.

In Verrall (1996), this method, which as far as we know has not been used for the graduation of mortality data, is considered to be promising. Finally, it is worth pointing out that the LOESS method is related to the generalised linear models proposed by Verrall (1993a) and to the graduation by Whittaker, all of which are based on a straight line whose gradient varies smoothly with age (Verrall, 1993b).

2.4 Generalised Additive Models (GAM)

Generalised Linear models (GLM) are an extension of linear models for distributions of the non-normal response variable and non-linear transformations. A regression model constitutes a specification for the average m of the variable in terms of a small number of unknown parameters

$\beta_0, \beta_1, \ldots, \beta_p$. In the particular case of linear models, we wish to find a linear function such that $E(Y|X) = m = \beta_0 + \sum_{i=1}^{p} \beta_i x_i$, and for that a constant variance of Y is supposed, $\text{var}(Y) = \sigma^2$.

In a different way from those, a generalised linear model (GLM) provides a method for estimating a function of the average of the response variable as a linear combination of the set of predictive variables, which is

$$l(E(Y|X)) = l(m) = \eta(x) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i.$$

The function of the response average, $l(m)$, is called function *link*, and is considered to be the same as a linear function of the predictors, $\eta(x)$, which is called *linear predictor*.

The distribution of each component y_i of Y has a density that can be expanded according to

$$f(y_i; \theta_i, \varphi) = \exp\left[A_i\{y_i\theta_i - \gamma(\theta_i)\}/\varphi + \tau(y_i, \varphi/A_i)\right] \tag{6}$$

where φ is a parameter of scale (possibly known), A_i is known, and the parameter θ_i depends on the linear predictor.

In our case, D_i number of deaths at age x_i follows a Binomial distribution, $D_i \sim Bi(E_i, q_i)$. If we take $y_i = d_i/E_i$ as responses, the density can be expressed in the form

$$\log f(y_i) = E_i\left[y_i \log\left(\frac{q_i}{1-q_i}\right) + \log(1-q_i)\right] + \log\binom{E_i}{E_i y_i},$$

which, when compared with (6), it can be deduced that $A_i = E_i$, $\varphi = 1$, θ_i will be the logit transformation of q_i, $\log(q_i/(1-q_i))$ and $\gamma(\theta_i) = -\log(1-q_i) = \log(1+e^{\theta_i})$. A detailed description of GLM can be found in McCullagh & Nelder (1989).

Generalised additive models (GAM) are a natural extension of GLM in the sense that they adjust nonparametric functions to study the relationship between predictive variables and the answer. In effect, the predictor η no longer has to be a linear function of predictive variables, but is

$$\eta = \alpha + \sum_{j=1}^{p} f_j(x_j) + \epsilon \tag{7}$$

where f_j is a smooth function (splines, locally-weighted regression, ...). In graduation, the variable is the age, x, and the answer is the logit-transformation of q_x. These models are a semi-parametric approach, an improvement with respect to the nonparametric techniques because they allow us to use, in a similar way to generalised linear models, the exact distribution that corresponds to the number of deaths, the Binomial distribution under the hypothesis of independence among the deaths. The importance these models have in graduation is that they allow for an improvement in the formulation of some of the nonparametric methods already described in actuarial theory. For example, cubic spline graduation uses Normal approximate distribution. However, this is not necessary with the GAM approach as these models are able to use an exact distribution. Comprehensive research on GAM can be found in Hastie & Tibshirani (1990) and Hastie *et al.* (2001). Verrall (1996) describes how graduation theory can be incorporated and enlarged within a GAM scheme. As in the case of the LOESS method, GAM have not been used in the graduation of mortality data.

2.5 *Bandwidth or Span Smoothing Parameter Choice*

The optimum choice of the bandwidth parameter or the span parameter is an issue that has been widely commented upon in the literature on data smoothing. A common statistical technique in actuarial science is to firstly choose the model that best fits the data and later verify its smoothness by standard actuarial graduation tests: chi-squared, run and autocorrelation tests. In modern statistics, both steps are combined by using a method to choose bandwidth that balances variance and bias.

Gavin *et al.* (1994) use cross-validation, which is later compared with that attributed to Bloomfield & Haberman (1987). In both cases, insurance tables from 1967–70 (CMI data) are studied and similar choices are obtained, although cross-validation is more intuitive and theoretically firmer. The cross-validation criterium in a more general context has been examined in Stone (1974) and Silverman (1984).

According to Verrall (1996), the cross-validation method in the context of graduation consists of choosing the bandwidth value b, which minimizes

$$CV(b) = \frac{1}{n}\sum_{i=1}^{n}(\dot{q}_i - \hat{q}_i^{(-i)})^2,$$

where $\hat{q}_i^{(-i)}$ is the estimate obtained on using all crude values except $\dot{q}_{x_i}$, which, in order to simplify, is written as $\dot{q}_i$.

Finally, degrees of freedom for a smoother is a concept that is equivalent to the number of parameters in parametric graduation. They are related to the bandwidth through the linear smoother matrix S_λ ($df = tr(S_\lambda)$), where $S = s_{ij}$ is a $n \times n$ matrix whose element s_{ij} is the weight of observation j in the estimation of q_i.

3 Application of Nonparametric Methods to Mortality Data from the Valencia Region

The methods described in the previous section are applied to mortality data from the Valencia Region, a Spanish region on the coast of the Mediterranean Sea. The data consist of aggregate population and death values corresponding to the three-year period 1999–2001. These two data sets were published by the Spanish National Institute of Statistics (INE) and are classified by age (ranging from 0 to 100 or older) and sex. They both refer to the Valencia Region as the place of residence, which means the two sets of data correspond to each other coherently.

As censuses are carried out during the first year of each decade, this means that only the data corresponding to 2001 are actually real, as the rest are census estimates published in INE reports (INE 1997 and INE 2001).

The first step is to calculate the crude estimations of q_x from these data. From among the different existing proposals for carrying out such estimations, we have used that of Navarro *et al.* (1995) expressed by

$$\dot{q}_x = \frac{d_{x(t-1)} + d_{xt}}{1/2P_{x(t-1)} + P_{xt} + 1/2P_{x(t+1)} + 1/2(d_{x(t-1)} + d_{xt})}, \qquad (8)$$

where P_{xt} is the population of people whose ages are between x and $x+1$ years old on 1-st January of the year t, and d_{xt} is the number of individuals' deaths whose ages are between x and $x+1$ during the year t.

This choice is made as we do not have the deaths classified according to the year of birth, but according to age and sex, thus the expression (8) allows us to avoid this difficulty because it supposes a uniform death distribution throughout the year. The denominator of the expression is an estimation of those initial exposed to risk.

The graphic representation of the logarithms of the crude estimations led us to take a range of ages between 0 and 96, which seems to us compatible with the use of the maximum possible and with the demand for relatively stable behavior. Beyond this age the logarithms decreased, showing behavior that was difficult to explain.

In the period and range under study, there were nearly 3.96 million men at risk and 4.11 million women. Of these, 77% were between 21 and 96 years of age. In the same period, nearly 39.3 thousand men died and 51.4 women, with the great majority, approximately 99%, doing so between the ages of 21 and 96 (see Table 1).

Table 1

Age, x_i, number of initial exposed to risk, E_i, and number of deaths, d_i, observed in the period 1999–2001 in the Valencia Region (Spain).

	MEN		WOMEN			MEN		WOMEN	
x_i	E_i	d_i	E_i	d_i	Age	E_i	d_i	E_i	d_i
0	39199.70	180.00	36955.70	182.00					
1	38315.50	21.00	36211.50	17.00	49	48797.50	212.00	50280.50	137.00
2	38139.50	5.00	35822.50	11.00	50	48223.00	234.00	49786.00	132.00
3	38096.00	7.00	35797.00	10.00	51	47697.50	254.00	49363.50	151.00
4	38345.00	6.00	36114.50	6.00	52	47301.50	298.00	49073.50	160.00
5	39066.00	8.00	36755.50	7.00	53	46565.50	265.00	48400.50	180.00
6	39971.50	9.00	37689.00	1.00	54	45335.00	318.00	47270.00	195.00
7	40772.50	10.00	38576.00	11.00	55	43837.00	345.00	45854.00	189.00
8	41449.00	8.00	39300.50	7.00	56	42854.00	349.00	44903.00	205.00
9	42046.50	2.00	39956.00	8.00	57	42080.50	376.00	44140.00	199.00
10	42696.50	8.00	40522.50	9.00	58	40376.50	359.00	42484.00	227.00
11	43574.50	5.00	41207.00	3.00	59	38943.50	457.00	41196.00	268.00
12	44781.00	5.00	42271.00	4.00	60	38785.00	440.00	41292.50	248.00
13	46223.00	10.00	43662.50	9.00	61	38629.50	476.00	41424.50	303.00
14	47927.00	11.00	45348.50	8.00	62	38341.50	487.00	41368.00	347.00
15	50014.50	20.00	47414.00	19.00	63	39109.00	591.00	42455.00	398.00
16	52505.50	34.00	49799.50	19.00	64	39885.50	656.00	43643.50	416.00
17	55265.50	50.00	52413.50	26.00	65	39758.50	706.00	43911.00	488.00
18	58202.00	66.00	55256.00	24.00	66	39325.00	744.00	43958.00	507.00
19	61265.00	49.00	58248.50	29.00	67	38834.50	868.00	43998.50	614.00
20	64133.00	66.00	61040.00	30.00	68	37958.00	939.00	43610.00	665.00
21	66470.00	60.00	63360.50	22.00	69	36676.50	922.00	42844.00	748.00
22	68160.50	58.00	65063.50	36.00	70	35290.50	994.00	42032.50	821.00
23	68999.50	59.00	66045.50	31.00	71	33869.00	1043.00	41144.50	881.00
24	69101.50	70.00	66303.00	31.00	72	32269.50	1146.00	40023.00	977.00
25	68737.50	56.00	66035.00	33.00	73	30646.50	1245.00	38763.00	1136.00
26	68178.00	71.00	65598.00	29.00	74	29109.00	1276.00	37516.50	1281.00
27	67544.00	73.00	65109.00	33.00	75	27468.00	1273.00	36204.50	1383.00
28	66988.50	69.00	64620.50	40.00	76	25565.00	1327.00	34564.00	1542.00
29	66568.50	81.00	64415.50	40.00	77	23332.00	1484.00	32491.50	1579.00
30	66311.00	71.00	64476.50	37.00	78	20987.50	1488.00	30287.50	1715.00
31	66111.00	99.00	64526.00	41.00	79	18440.50	1290.00	27841.50	1704.00
32	65994.00	77.00	64630.00	52.00	80	15972.00	1199.00	25308.50	1796.00
33	65963.00	93.00	64775.00	65.00	81	13771.50	1148.00	22870.00	1856.00
34	65564.00	111.00	64497.50	41.00	82	12148.50	1169.00	20833.00	2095.00
35	64618.50	109.00	63797.00	75.00	83	10820.00	1040.00	18958.50	2096.00
36	63513.00	123.00	63059.50	72.00	84	9785.50	1094.00	17317.00	2256.00
37	62503.50	102.00	62347.00	68.00	85	8764.50	1074.00	15721.50	2325.00
38	61467.00	117.00	61572.50	69.00	86	7716.50	1021.00	14113.00	2332.00
39	60471.00	128.00	60830.50	83.00	87	6670.50	973.00	12400.50	2306.00
40	59447.50	136.00	59983.00	89.00	88	5589.50	845.00	10596.00	2141.00
41	58063.00	128.00	58719.00	85.00	89	4631.00	733.00	8933.00	2051.00
42	56272.50	164.00	57007.00	90.00	90	3737.00	593.00	7320.00	1809.00
43	54271.00	162.00	55168.50	104.00	91	2995.50	550.00	5883.00	1654.00
44	52509.00	165.00	53575.00	102.00	92	2301.00	462.00	4558.50	1396.00
45	51236.50	165.00	52366.50	122.00	93	1702.50	334.00	3416.50	1097.00
46	50214.50	173.00	51369.50	118.00	94	1269.50	234.00	2535.00	936.00
47	49358.50	202.00	50587.50	116.00	95	852.50	179.00	1740.00	667.00
48	48993.50	188.00	50376.00	133.00	96	576.50	130.00	1148.50	530.00

The performance of the nonparametric methods will be evaluated with the goodness-of-fit statistics: log-likelihood, deviance and χ^2. These three statistics measure the distance between observed values $\dot{q_x}$ and adjusted values $\hat{q_x}$. The likelihood is

$$\begin{aligned} L(q) &= \prod_{i=1}^{n} \binom{E_i}{d_i} q_i^{d_i} (1 - q_i)^{E_i - d_i} \\ &\propto \prod_{i=1}^{n} q_i^{d_i} (1 - q_i)^{E_i - d_i} \end{aligned}$$

and the log-likelihood (without constants) is

$$\log L(q) = \sum_{i=1}^{n} \left(d_i \log(q_i) + (E_i - d_i) \log(1 - q_i) \right).$$

From the log-likelihood function, we can calculate the deviance,

$$D(\hat{q}) = 2 \log L(\dot{q}) - 2 \log L(\hat{q}).$$

The discrepancy between observed and expected deaths is measured with the corresponding χ^2,

$$\chi^2 = \sum_{i=1^n} \frac{(d_i - E_i \hat{q}_i)^2}{E_i \hat{q}_i (1 - \hat{q}_i)}.$$

We want to find the model with maximum log-likelihood and minimum deviance and χ^2. When comparing two models, the difference between deviances can be related approximately to the chi-square distribution having a number of degrees of freedom equal to the difference between the number of parameters in each model. We then choose the model with a significant improvement. To explore the changes in the number of terms in a model, we consider the Mallows statistic C_p defined by

$$C_p = RSS + 2\sigma^2 p,$$

where RSS is the residual squared sum and p is the number of parameters. This statistic penalizes the complexity of models, because it increases with the number of parameters.

These measures of goodness-of-fit along with the graphical representations of estimated values have been used in order to choose the optimal model of each nonparametric method.

3.1 Kernel Graduation

A graduation of death probabilities, q_x, is carried out using different bandwidth choices. The goodness-of-fit tests mentioned above lead us to conclude that the Nayadara–Watson estimator fits better than that of Copas–Haberman for both sexes.

Figure 1 shows the graphic corresponding to the fit for men and women, respectively. The graphic of residuals against estimated values confirms the existence of heteroscedasticity (Figures 2 and 3), which leads us to propose the transformation of the crude probabilities. The transformations proposed are: $\log(q_x)$, $\text{logit}(q_x)$ and $\text{clog}(q_x) = \log(-\log(1 - q_x))$.

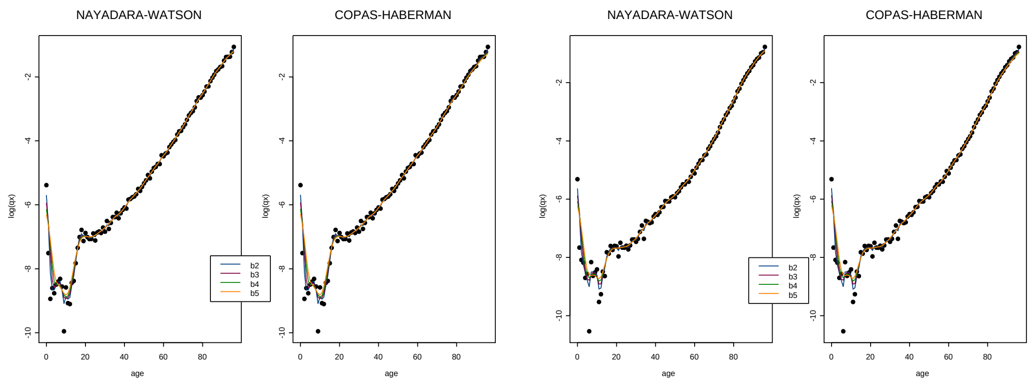

Figure 1. *Comparison of kernel estimations for men (left) and women (right).*

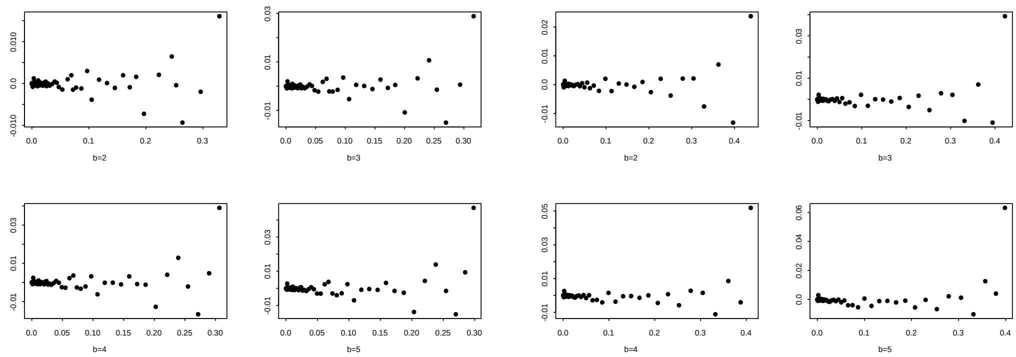

Figure 2. *Residuals versus Nayadara–Watson estimations for men (left) and women (right).*

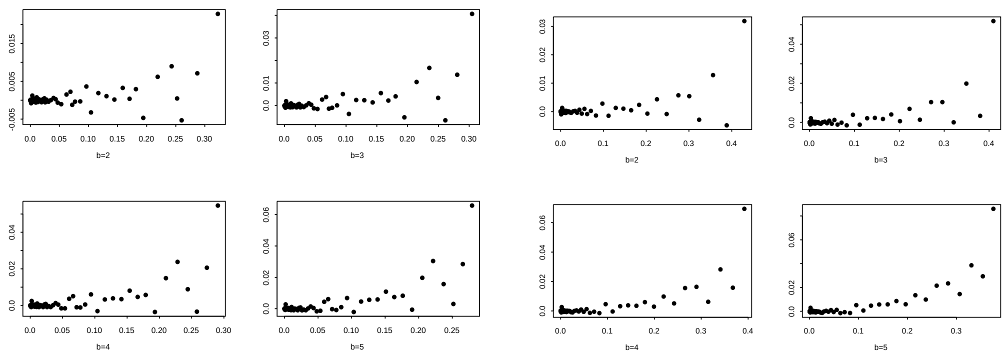

Figure 3. *Residuals versus Copas–Haberman estimations for men (left) and women (right).*

The goodness-of-fit statistics in Table 2 show that the best results are obtained for q_x. Among all the transformations, logit behaves best. It is worth remembering that we need to carry out transformation of the data in order to eliminate heteroscedasticity. Logit transformation is that which behaves best among all the transformations, and we use it to compare kernel density estimation with other methods. The success in removing heteroscedasticity is confirmed by Figure 4 showing the graph of residuals versus the estimates of transformed scores. Nevertheless, estimates for early ages still show high deviation.

Table 2

Goodness-of-fit statistics for the Nayadara–Watson estimator (kernel).

	MEN				WOMEN			
$\mathbf{q_x}$	$b=2$	$b=3$	$b=4$	$b=5$	$b=2$	$b=3$	$b=4$	$b=5$
Deviance	77.90	160.08	228.36	287.75	84.24	158.83	222.71	290.68
log-likelihood	-169391	-169432	-169466	-169496	-190080	-190118	-190149	-190183
χ^2	73.05	156.73	235.95	313.35	76.52	154.72	231.87	318.93
$\mathbf{log(q_x)}$	$b=2$	$b=3$	$b=4$	$b=5$	$b=2$	$b=3$	$b=4$	$b=5$
Deviance	100.44	240.61	362.31	457.64	113.48	238.37	334.66	417.60
log-likelihood	-169402	-169472	-169533	-169580	-190094	-190157	-190205	-190247
χ^2	118.47	343.00	610.46	879.72	136.54	340.51	548.30	762.43
$\mathbf{logit(q_x)}$	$b=2$	$b=3$	$b=4$	$b=5$	$b=2$	$b=3$	$b=4$	$b=5$
Deviance	100.25	240.22	361.73	456.81	113.11	237.38	332.58	413.44
log-likelihood	-169402	-169472	-169472	-169533	-190094	-190156	-190156	-190204
χ^2	118.18	342.26	609.22	877.97	136.05	339.15	545.60	757.45
$\mathbf{clog(q_x)}$	$b=2$	$b=3$	$b=4$	$b=5$	$b=2$	$b=3$	$b=4$	$b=5$
Deviance	100.34	240.41	361.99	457.17	113.28	237.81	333.46	415.17
log-likelihood	-169402	-169472	-169472	-169533	-190094	-190157	-190157	-190205
χ^2	118.32	342.62	609.80	878.76	136.28	339.77	546.79	759.57

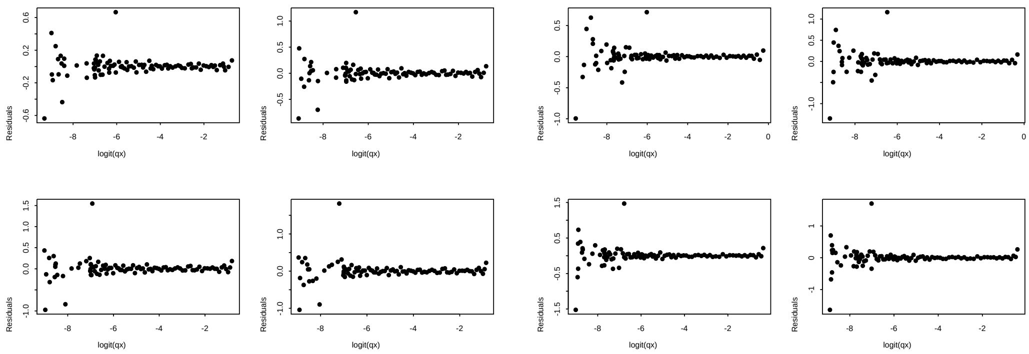

Figure 4. *Residuals versus estimation of logit*(q_x) *for men (left) and women (right).*

Bandwidth choice is carried out according to values obtained by cross-validation (Table 3), where age zero is excluded. We have eliminated age 0 because it has greater curvature and presents more difficulties when it comes to adjustment. Haberman (1997) does the same. In accordance with this criteria $b = 5$, this value is a trade-off between smoothness and goodness-of-fit.

Table 3

Goodness-of-fit statistics for kernel estimation applied with logit-transformed crude probabilities for the bandwidth obtained by cross-validation.

	MEN	WOMEN
Bandwidth	5.218196	4.554322
Deviance	113.2749	96.087
log-likelihood	-169408.8	-190086.2
χ^2	117.979	96.01836
Df	20.44698	23.35616

3.2 Smoothing Spline Graduation

The choice of smoothness (degrees of freedom) to be applied to the various transformations of crude probabilities is decided by cross-validation. Cross-validation is easier to apply with splines than the previous and following methods because it is implemented in `S-plus` and `R` software .

Age zero was excluded for the above-mentioned reasons, allowing us to obtain a reasonably smooth final curve, as can be seen in Figure 5.

The same transformations that are applied to kernel methods are used to avoid the problem of heteroscedasticity. As in the previous case, results improve. Log transformation is chosen as it offers the best fit (Table 4), despite the fact that there is not much difference with respect to other methods (Figure 5). An advantage of the smoothing spline over kernel density estimation is that we can obtain forecasts for age values greater than 96 using the last piece of spline function. This extrapolation beyond the age range requires, as Haberman (1997) points out, that q_x approaches the value of 1 as age increases.

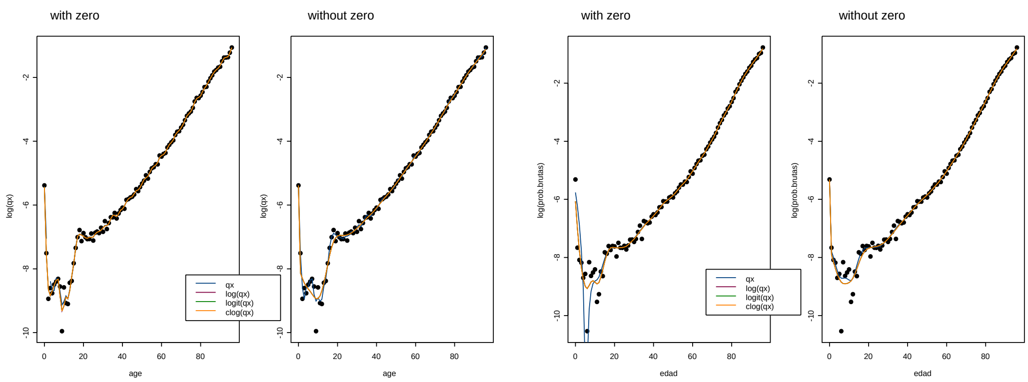

Figure 5. *Spline smoothing for men (left) and women (right).*

Table 4

Goodness-of-fit statistics for spline adjustment.

	MEN				WOMEN			
with zero	q_x	$\log(q_x)$	$\text{logit}(q_x)$	$\text{clog}(q_x)$	q_x	$\log(q_x)$	$\text{logit}(q_x)$	$\text{clog}(q_x)$
Deviance	NA	41.19	41.12	41.16	235.73	160.49	160.80	160.66
log-likelihood	NA	-169373	-169373	-169373	-190156	-190118	-190119	-190118
χ^2	NA	41.66	41.59	41.62	391.58	19	186.21	186.13
Df	57.67	52.54	52.60	52.56	22.80	24.99	24.98	24.99
without zero	q_x	$\log(q_x)$	$\text{logit}(q_x)$	$\text{clog}(q_x)$	q_x	$\log(q_x)$	$\text{logit}(q_x)$	$\text{clog}(q_x)$
Deviance	40.08	116.98	117.16	117.06	69.75	88.70	89.68	89.16
log-likelihood	-169372	-169411	-169411	-169411	-190074	-190083	-190083	-190083
χ^2	39.53	119.88	120.06	119.96	69.75	88.70	89.68	89.16
Df	56.79	17.79	17.77	17.78	22.54	15.93	15.91	15.92

3.3 Graduation with Locally-weighted Regression

This smoother has several possibilities as it allows us to manipulate the span parameter and the local polynomial degree used, which can be linear or quadratic. In the first place, the different possibilities are tested and, afterwards, cross-validation is used to choose the parameters.

Although the results in Table 5 show better behaviour for q_x than for its transformations, as was done in previous methods, the transformations have been applied to avoid the problem of heteroscedasticity. We will use logit transformation in cross-validation because it is slightly better.

In Figure 6, the comparison of results of the linear and quadratic local polynomial regressions with span chosen by cross-validation can be observed. The quadratic solution for men forms quite a rough curve in ages close to zero and for this reason it shows better goodness-of-fit (Table 6). Nevertheless, the linear regression is chosen because it is much smoother. However, for women, contrary behaviour occurs and quadratic regression is chosen.

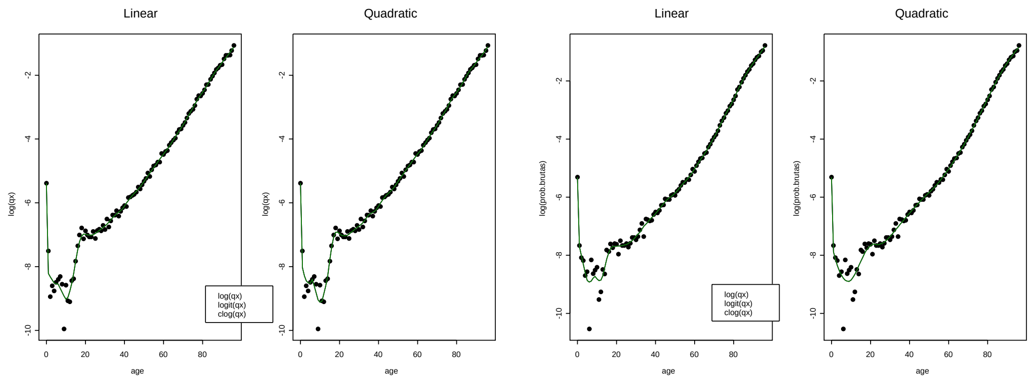

Figure 6. *Linear and Quadratic LOESS comparison with span chosen by cross-validation for men (left) and women (right).*

Table 5
Goodness-of-fit statistics for LOESS adjustment.

	MEN				WOMEN			
	Linear							
$\mathbf{q_x}$	$s=0.05$	$s=0.1$	$s=0.15$	$s=0.2$	$s=0.05$	$s=0.1$	$s=0.15$	$s=0.2$
Deviance	93.15	283.30	432.77	575.95	100.86	254.92	480.93	785.18
log-likelihood	-169399	-169494	-169568	-169640	-190089	-190166	-190279	-190431
χ^2	83.02	248.48	416.97	604.15	85.65	225.25	469.28	802.72
Df	36.8	18.2	11	8.7	36.8	18.2	11	8.7
$\mathbf{log(q_x)}$	$s=0.05$	$s=0.1$	$s=0.15$	$s=0.2$	$s=0.05$	$s=0.1$	$s=0.15$	$s=0.2$
Deviance	56.63	301.88	473.74	595.43	69.23	214.04	408.03	560.61
log-likelihood	-169380	-169503	-169589	-169650	-190073	-190145	-190242	-190318
χ^2	56.66	418.88	840.00	1273.73	70.93	266.33	701.97	1200.24
Df	36.8	18.2	11	8.7	36.8	18.2	11	8.7
$\mathbf{logit(q_x)}$	$s=0.05$	$s=0.1$	$s=0.15$	$s=0.2$	$s=0.05$	$s=0.1$	$s=0.15$	$s=0.2$
Deviance	56.64	301.78	472.79	593.60	69.42	214.55	404.47	546.05
log-likelihood	-169380	-169503	-169588	-169648	-190073	-190145	-190240	-190311
χ^2	56.67	418.33	838.15	1270.64	71.10	266.61	697.59	1184.24
Df	36.8	18.2	11	8.7	36.8	18.2	11	8.7
$\mathbf{clog(q_x)}$	$s=0.05$	$s=0.1$	$s=0.15$	$s=0.2$	$s=0.05$	$s=0.1$	$s=0.15$	$s=0.2$
Deviance	56.63	301.82	473.15	594.17	69.32	214.20	405.26	550.00
log-likelihood	-169425	-169503	-169589	-169649	-190245	-190145	-190241	-190313
χ^2	56.66	418.60	838.96	1271.84	71.01	266.38	698.78	1188.88
Df	36.8	18.2	11	8.7	36.8	18.2	11	8.7
	Quadratic							
$\mathbf{q_x}$	$s=0.08$	$s=0.1$	$s=0.15$	$s=0.2$	$s=0.08$	$s=0.1$	$s=0.15$	$s=0.2$
Deviance	77.62	NA	NA	425.51	95.41	NA	NA	343.31
log-likelihood	-169391	NA	NA	-169565	-190086	NA	NA	-190210
χ^2	69.52	NA	NA	3299.61	82.14	NA	NA	969.79
Df	47.9	35.2	20.2	15.5	47.9	35.2	20.2	15.5
$\mathbf{log(q_x)}$	$s=0.08$	$s=0.1$	$s=0.15$	$s=0.2$	$s=0.08$	$s=0.1$	$s=0.15$	$s=0.2$
Deviance	46.17	80.20	243.12	319.23	66.42	106.21	152.47	249.94
log-likelihood	-169375	-169392	-169474	-169512	-190071	-190091	-190114	-190163
χ^2	46.57	81.73	301.05	441.63	69.04	112.88	169.67	328.73
Df	47.9	35.2	20.2	15.5	47.9	35.2	20.2	15.5
$\mathbf{logit(q_x)}$	$s=0.08$	$s=0.1$	$s=0.15$	$s=0.2$	$s=0.08$	$s=0.1$	$s=0.15$	$s=0.2$
Deviance	46.18	80.21	243.05	318.97	66.69	106.28	152.79	250.06
log-likelihood	-169375	-169392	-169474	-169512	-190071	-190091	-190115	-190163
χ^2	46.58	81.72	300.71	440.99	69.30	112.90	169.91	328.60
Df	47.9	35.2	20.2	15.5	47.9	35.2	20.2	15.5
$\mathbf{clog(q_x)}$	$s=0.08$	$s=0.1$	$s=0.15$	$s=0.2$	$s=0.08$	$s=0.1$	$s=0.15$	$s=0.2$
Deviance	46.17	80.20	243.08	319.10	66.56	106.25	152.62	249.94
log-likelihood	-169419	-169392	-169474	-169512	-190242	-190091	-190114	-190163
χ^2	46.58	81.73	300.88	441.31	69.18	112.89	169.78	328.61
Df	47.9	35.2	20.2	15.5	47.9	35.2	20.2	15.5

Table 6

Bandwidth and goodness-of-fit statistics for LOESS on logit transformations of crude probabilities.

	MEN		WOMEN	
	Lineal	**Quadratic**	**Linear**	**Quadratic**
span	0.10	0.12	0.10	0.26
Deviance	108.50	87.83	73.39	100.33
log-likelihood	-169406.4	-169396	-190074.9	-190088.3
χ^2	110.79	88.36	72.87	99.40
Df	18	27.2	18	11.9

3.4 Generalised Additive Models Graduation

The graduation of logit(q_x) is now carried out using GAM. In (7) two kinds of smooth functions f_j are used: splines and locally-weighted regression.

3.4.1 GAM with splines

The result obtained with splines (see section 3.2) is taken into account when choosing the range, from 10 to 24, of the number of degrees of freedom. These graduations are carried out separately for men and women, and for ages ranging from 1 to 96 years old, age zero being excluded as in previous methods.

Table 7 compares the different models, paying attention to their increase in complexity. For this reason, the table shows the change in deviance and Mallow's C_p statistic when the degrees of freedom increase from unit to unit. A gradual improvement in goodness-of-fit can be deduced from the p-values associated with deviance as the number of degrees of freedom increase. We can see that for men the p-value of improvement in deviance is significant for all degrees of freedom and, therefore, the model with the greatest number of degrees of freedom should be accepted. For women, this significance ceases in models with 15 or 16 degrees of freedom.

Table 7

Comparison of GAM's with different degrees of freedom.

	MEN			WOMEN		
Df	**Deviance**	**p-value**	$\mathbf{C_p}$	**Deviance**	**p-value**	$\mathbf{C_p}$
10			221.00			136.96
11	6.55	0.0106	217.21	4.46	0.0351	134.39
12	7.02	0.0081	212.95	4.32	0.0377	131.96
13	7.36	0.0066	208.33	4.21	0.0406	129.65
14	7.79	0.0054	203.35	3.98	0.0453	127.53
15	7.64	0.0056	198.43	3.86	0.0497	125.56
16	7.67	0.0057	193.53	3.71	0.0554	123.77
17	7.53	0.0063	188.81	3.42	0.0642	122.23
18	7.04	0.0079	184.52	3.20	0.0741	120.92
19	6.48	0.0110	180.80	2.93	0.0868	119.87
20	6.23	0.0126	177.33	2.67	0.1022	119.09
21	5.70	0.0167	174.36	2.39	0.1204	118.56
22	5.28	0.0215	171.83	2.21	0.1404	118.27
23	4.90	0.0270	169.69	1.94	0.1639	118.22
24	4.53	0.0336	167.93	1.71	0.1881	118.37

As there is not one model that can simultaneously serve both sexes, nonparametric contrasts (Forfar *et al.*, 1988) are carried out on models with 16 or more degrees of freedom. The results shown in Table 8 would lead us to choose model $df = 21$ for men and model $df = 19$ for women.

Table 8

Nonparametric contrasts of GAM with splines.

	MEN								
df	χ^2	$z > 2$	$z > 3$	pos	p(pos)	runs	p(runs)	K-S	p(K-S)
16	149.01	6	2	46	0.3424	45	0.4407	0.062	0.993
17	140.68	5	2	46	0.3424	45	0.4407	0.062	0.993
18	132.98	5	2	45	0.2713	47	0.4770	0.062	0.993
19	125.57	5	2	43	0.1550	45	0.4472	0.062	0.993
20	118.98	5	2	44	0.2084	45	0.4444	0.062	0.993
21	112.99	3	2	45	0.2713	49	0.5119	0.062	0.993
22	107.50	3	1	45	0.2713	49	0.5119	0.062	0.993
	WOMEN								
df	χ^2	$z > 2$	$z > 3$	pos	p(pos)	runs	p(runs)	K-S	p(K-S)
16	85.56	5	0	46	0.3424	55	0.6128	0.083	0.899
17	82.25	4	0	47	0.4196	57	0.6444	0.083	0.899
18	79.19	4	0	48	0.5000	57	0.6439	0.083	0.899
19	76.42	2	0	48	0.5000	57	0.6439	0.062	0.993
20	73.92	2	0	49	0.5804	57	0.6444	0.062	0.993
21	71.71	2	0	49	0.5804	57	0.6444	0.062	0.993
22	69.69	2	0	49	0.5804	57	0.6444	0.0619	0.993

Figure 7 shows a comparison of the various models for each sex. Both are presented in log scale to enable their comparison with the rest of the graphs, even though the graduation was carried out by logit transformation.

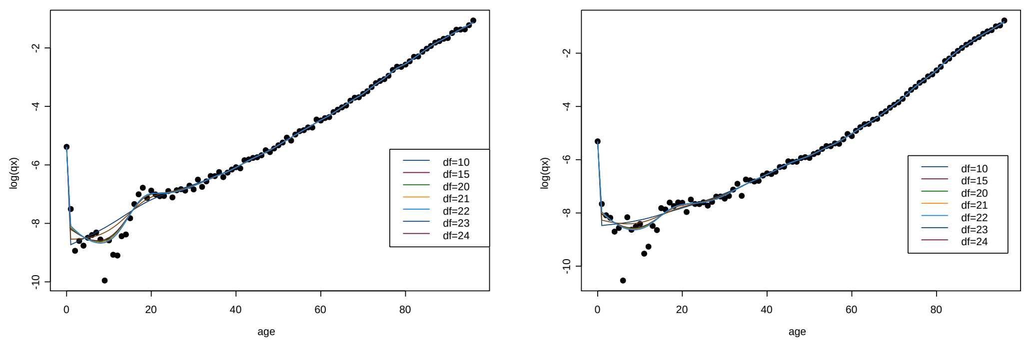

Figure 7. *Comparison of GAM's with splines for men (left) and women (right).*

3.4.2 GAM with LOESS

The locally-weighted regressions used with GAM are both linear and quadratic as in section 3.3. The comparison between the two is carried out in a similar way to that of the previous section. The p-value of deviance and C_p statistic (Table 9) does not allow us to choose the best linear model for men as the improvement is continuous. However, Figure 8 shows that the model with the lowest span can be discarded because it produces fluctuations in lower age groups. The model with the lowest span that does not produce the above-mentioned fluctuations is the model with a span of 0.1. As far as women are concerned, we would choose the models with a span of 0.1 or 0.09, in accordance with the value of the C_p statistic. The best quadratic models for both sexes are the models with a span of 0.15, according to Table 9 and Figure 8.

Table 9
Comparison of GAM's with different spans.

	MEN				WOMEN			
Linear	**Deviance**	**g.l.**	**p-value**	**C_p**	**Deviance**	**g.l.**	**p-value**	**C_p**
0.20				166.56				130.23
0.15	19.23	2.71	0.0002	152.75	14.86	2.66	0.0013	120.38
0.10	23.30	8.51	0.0041	146.46	22.89	8.48	0.0047	113.44
0.09	1.36	0.23	0.0477	145.56	0.99	0.23	0.0679	112.88
0.08	17.97	6.57	0.0092	140.72	6.83	6.59	0.4022	118.46
0.05	29.28	13.79	0.0087	138.99	12.59	13.80	0.5441	131.85
Quadratic	**Deviance**	**g.l.**	**p-value**	**C_p**	**Deviance**	**g.l.**	**p-value**	**C_p**
0.20				149.49				119.25
0.15	11.65	5.22	0.0455	148.31	13.88	5.19	0.0186	115.98
0.14	6.72	3.26	0.10	148.13	4.39	3.25	0.2550	118.25
0.13	1.16	0.25	0.0646	147.48	0.61	0.2512	0.1138	118.15
0.10	32.09	12.67	0.0019	140.81	9.40	12.69	0.7208	134.71
0.05	19.20	13.29	0.1283	148.29	9.49	13.29	0.7543	152.40

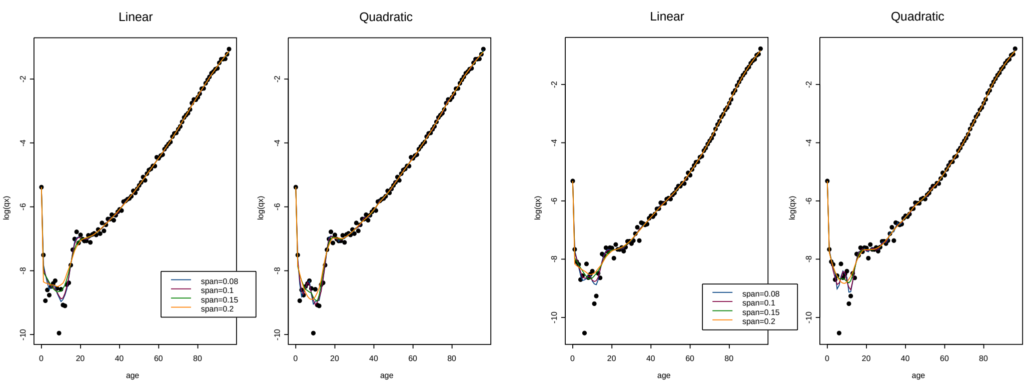

Figure 8. *Comparison of GAM with LOESS for men (left) and women (right).*

3.4.3 Choice of the best GAM

We now choose a GAM to be compared with those models chosen in the previous sections according to the results of the nonparametric contrasts. Even when results are similar, GAM with LOESS show marginally significant autocorrelation in statistical terms. In this respect, GAM with splines, despite showing a slightly worse fit, have more residual degrees of freedom. They do not show autocorrelation for women and the autocorrelation for men is almost insignificant. Consequently, the model chosen for both sexes is the GAM with splines.

4 Conclusions

We must begin this section by pointing out that, as far as we know, LOESS models and GAM have not been used to date for the graduation of mortality data.

The comparison is carried out by applying the tests proposed by Forfar *et al.* (1988). We have also obtained the values of the mean absolute percentage error (MAPE) and R^2 that Felipe & Guillén (1999) and Felipe *et al.* (2002) use in their work. Results are presented in Table 10. In summary, all models display favourable results making it difficult to choose one of them. Nevertheless, we can conclude that:

1. kernel models and splines fit worse than the other two,
2. all models fit better for women, and
3. *GAM* models provide a better fitting for both sexes and have the additional advantage of allowing us to use the real distribution of data.

Table 10

Results of nonparametric contrasts with kernel and splines.

		Kernel		Splines		LOESS		GAM	
		Men	Women	Men	Women	Men	Women	Men	Women
Relative	> 2	5	2	5	2	4	4	3	2
Desv.[a]	> 3	1	1	1	0	1	0	2	0
Signs	pos. (neg.)	52 (44)	54 (42)	50 (46)	56 (40)	49 (47)	53 (43)	45 (51)	49 (47)
test	p-value	0.7916	0.8886	0.6576	0.9401	0.5804	0.8450	0.2713	0.5804
Runs	runs	51	57	51	57	49	57	49	57
test	p-value	0.5495	0.6607	0.5495	0.6741	0.5009	0.6555	0.5119	0.6444
K-S	K-S	0.0412	0.0515	0.0412	0.0515	0.0412	0.0515	0.0619	0.0619
test[b]	p-value	1	0.9996	1	0.9996	1	0.9996	0.993	0.993
χ^2	χ^2	117.98	96.03	119.87	88.70	110.79	99.40	112.99	71.71
test[c]	g.l.	75.53	72.64	78.21	80.07	78	84.1	74.01	74.05
	p-value	0.0011	0.03085	0.0017	0.2385	0.0087	0.1219	0.0024	0.5554
$\mathbf{R^2}$		0.9346	0.9247	0.9946	1	0.9960	0.9996	0.9852	0.9910
MAPE		10.93	13.44	10.93	14.12	10.17	14.82	13.94	17.33

[a] standarized residuals
[b] Kolmogorov–Smirnov test
[c] χ^2 statistic, sum of squared standarized residuals

Finally, we should point out that the fitting of all the models presents problems for the early ages due to their irregular profile. Many authors achieve better fittings by eliminating this group of ages, which they justify by arguing that actuarial operations begin at a more advanced age. Contrary to this criterion, we have decided to include the young age groups, excluding zero, for two reasons. Firstly, it enables us to compare our results with those obtained by Navarro *et al.* (1995), who graduate mortality data for the Valencia Region for the years 1990–92 for the complete range of ages. Secondly, as far as we know, Navarro *et al.* (1995) is the only study that covers the same geographical area as ours and a comparison is vital. However, we have to point out that the inclusion of the early ages produces an increase as a result of the curvature that the data present at these ages (Gavin *et al.*, 1994). It is worth remembering that the double exponential which appears in Heligman & Pollard (1980) and related parametric models, was introduced specifically to deal with the difficulty of adjusting 0 and 1 ages.

A future line of work would be to extend the previous methods to the graduation of mortality data over time with the aim of obtaining dynamic mortality tables. The work by Felipe *et al.* (2001), Guillen *et al.* (2006) and Fledelius *et al.* (2004) go in this direction using kernel bivariante. Two-dimensional GAM have been used by Clements *et al.* (2005) to model and predict lung cancer rates.

Acknowledgements

The authors are indebted to the anonymous referees whose suggestions improved the original manuscript. This work was partially supported by a grant from MEyC (Ministerio de Educaci ón y Ciencia, Spain, project MTM-2004-06231). The research of Francisco Montes has also been partially supported by a grant from DGITT (Direcci ó General d'Investigació i Transferència Tecnològica de la Generalitat Valenciana, Project GRUPOS03/189).

References

Benjamin, B. & Pollard, J. (1992). *The Analysis of Mortality and Other Actuarial Statistics*. Butterworth–Heinemann, London, 6th edition.

Betzuen, A. (1997). An aproach to people's in employment mortality through smoothing splines. Internal research paper, Heriot-Watt University, Edinburgh.

Bloomfield, D. & Haberman, S. (1987). Graduation: Some experiments with kernel methods. *Journal of the Institute of Actuaries*, **114**, 339–369.

Breiman, L. & Peters, S. (1992). Comparing automatic smoothers. *International Statistical Review*, **60**(3), 271–290.

Brockman, M., Gasser, T. & Herrmann, E. (1993). Locally adaptative bandwidth choice for kernel regression estimators. *Journal of American Statistical Association*, **88**, 1302–1309.

Chan, F., Chan, L., Falkenberg, J. & Yu, M. (1986). Applications of linear and quadratic programmings to some cases of the Whittaker–Henderson graduation method. *Scandinavian Actuarial Journal*, (3), 141–153.

Clements, M., Armstrong, B. & Moolgavkar, S. (2005). Lung cancer rate prediction using generalized additive models. *Biostatistics*, **6**(4), 576–589.

Cleveland, W. (1979). Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829–836.

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

Epanechnikov, V. (1969). Nonparametric estimation of a multivariate probability density. *Theory and Probability Applications*, **14**, 153–158.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman and Hall.

Felipe, A., Guillen, M. & Nielsen, J. (2001). Longevity studies based on kernel hazard estimation. *Insurance: Mathematics & Economics*, **28**(2), 191–204.

Felipe, A., Guillén, M. & Pérez-Marín, A. (2002). Recent mortality trends in the Spanish population. *British Actuarial Journal*, **8**(4), 757–786.

Felipe, M. & Guillén, M. (1999). *Evolución y Predicción de las Tablas de Mortalidad Dinámicas para la Población Española*. Cuadernos de la Fundación, Fundación Mapfre Estudios, Madrid.

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

Forfar, D., McCutcheon, J. & Wilkie, A. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries*, **115** part I(459), 1–149.

Gasser, T., Kneip, A. & Köhler, W. (1991). A flexible and fast method for automatic smoothing. *Journal of American Statistical Association*, **86**, 643–652.

Gasser, T. & Müller, H. (1979). Kernel estimation of regression functions. In *Lecture Notes in Mathematics*, **757**, pp. 23–68. Berlin: Springer-Verlag.

Gavin, J., Haberman, S. & Verrall, R. (1993). Moving weighted average graduation using kernel estimation. *Insurance:*

Mathematics & Economics, **12**(2), 113–126.
Gavin, J., Haberman, S. & Verrall, R. (1994). On the choice of bandwidth for kernel graduation. *Journal of the Institute of Actuaries*, **121**, 119–134.
Gavin, J., Haberman, S. & Verrall, R. (1995). Graduation by kernel and adaptive kernel methods with a boundary correction. *Transactions. Society of Actuaries*, **XLVII**, 173–209.
Guillen, M., Nielsen, J. & Pérez-Marín, A. (2006). Multiplicative hazard models for studying the evolution of mortality. *The Annals of Actuarial Science*, Volume **1**.
Haberman, S. (1997). English Life Tables No. 15 1990-92. Appendix B. Series DS No 14, Office for National Statistics, HMSO.
Hastie, T., Tibshirani, R. & Friedman, J. (2001). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. New York: Springer.
Hastie, T.J. & Tibshirani, R. (1990). *Generalized Aditive Models*. London: Chapman and Hall.
Heligman, L. & Pollard, J. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, **107**, 49–80.
Henderson, R. (1924). A new method of graduation. *Transactions of American Society of Actuaries*, **XXV**.
Hoem, J.M. & Linnemann, P. (1988). The tails in moving average graduation. *Scandinavian Actuarial Journal*, (4), 193–229.
Kaishev, V., Dimitrova, D., Haberman, S. & Verrall, R. (2004). Automatic, computer aided geometric desing of free-knot, regression splines. Statistical Research Paper 24, City University, London, UK.
McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models*. London: Chapman and Hall.
McCutcheon, J. (1980). Recently published U.K. mortality tables methods of construction and possible developments therefrom. *ARCH (Actuarial Research Conference)*, **61**.
McCutcheon, J. (1981). Some remarks on splines. *Transactions of the Faculty of Actuaries*, **XXXVII**(421).
McCutcheon, J. (1987). Experiments in graduating the data for the English Life Tables No. 14. *Transactions of the Faculty of Actuaries*, **XL**(135).
Montenegro, M. (2001). Kernel smoothing: Teora y Aplicaciones. PhD thesis, Departamento de Matemáticas. Universitat Jaume I, Spain.
Navarro, E., Ferrer, R., Gonzalez, C. & Nave, J. (1995). *Tablas de Mortalidad de la Comunidad Valenciana 1990–91. Censos de Població i Habitatges*, volume I. Statistics Institute of Valencia Region (IVE), Valencia.
Nayadara, E. (1964). On estimating regression. *Theory of Probability and its Aplications*, **9**, 141–142.
Nielsen, J. (2003). Smoothing and prediction with a view to actuarial science, biostatistics and finance. *Scandinavian Actuarial Journal*, (1), 51–74.
Silverman, B. (1984). Fast and efficient cross-validation method for smoothing parameter choice in spline regression. *Journal of American Statistical Association*, **79**(387), 584–589.
Spanish National Institute of Statistics (INE) (1997). *Evolución de la población de España entre los censos de 1981 y 1991*. INE, www.ine.es.
Spanish National Institute of Statistics (INE) (2001). *Evolución de la población de España entre los censos de 1991 y 2001*. INE, www.ine.es.
Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society, Series B*, **36**, 111–147.
Verrall, R. (1993a). Graduation by dynamic regression methods. *Journal of the Institute of Actuaries*, **120**, 153–170.
Verrall, R. (1993b). A state space formulation of Whittaker graduation, with extensions. *Insurance: Mathematics & Economics*, **13**, 7–14.
Verrall, R. (1994). Whittaker graduation and dynamic generalized models. *Insurance: Mathematics & Economics*, **13**, 7–14.
Verrall, R. (1996). A unified framework for graduation. Actuarial Research Paper 91, City University, London, UK.
Wang, J. (2005). *Encyclopedia of Biostatistics*, Chap. Smoothing Hazard Rates. Willey, 2nd edition.
Wang, J., Müller, H. & Capra, W. (1998). Analysis of oldest-old mortality lifetables revisited. *The Annals of Statistics*, **26**(1), 126–163.
Watson, G. (1964). Smooth regression analysis. *Sankhya A*, **26**, 359–372.
Whittaker, E. (1923). On a new method of graduation. *Proceedings of Edinburgh Mathematical Society*, **41**, 63–75.
Zhang, H. (2004). Mixed effects multivariate adaptative splines model for the analysis of longitudinal growth curve data. *Statistical Methods in Medical Research*, **13**(1), 63–82.

Résumé

La graduation non paramétrique des données sur la mortalité envisage d'estimer les differents mesures de mortalité, en effectuant un lissage des mesures brutes directement obtenues à partir des données originelles. La différence fondamentale avec les modèles paramétriques est qu'il n'est pas nécessaire de supposer une fonction dépendant de l'âge, ce qui représente un avantage lorsque l'on n'a pas d'information sur le modèle sous-jacent, puisqu'une source d'erreur en est souvent le choix inadéquat. Dans ce travail, nous en avons examiné les différentes alternatives et nous y avons montré leur application à des données sur la mortalité dans la Region de Valencia, Espagne. Nous concluons d'après cette comparaison que le meilleur modèle en est le rabotage par des modèles additifs généralisés (GAM) avec des “splines”. L'intérêt principal de notre travail est le développement d'un plan qui peut être appliqué à des données sur la mortalité pour une large plage d'âge dans n'importe quel domaine géographique, de sorte qu'il nous permet de choisir le tableau le plus adéquat à l'expérience concernée.

[*Received December 2004, accepted March 2006*]