International Journal of Population Studies

RESEARCH ARTICLE

Nonparametric graduation techniques as a common framework for the description of demographic patterns

Anastasia Kostaki[1], Javier M. Moguerza[2], Alberto Olivares[3] and Stelios Psarakis[1]

[1] Department of Statistics, Athens University of Economics and Business, Greece
[2] Department of Computer Science and Statistics, Rey Juan Carlos University, Spain
[3] Department of Signal Theory and Communications and Telematic Systems and Computing, Rey Juan Carlos University, Spain

Abstract: The graduation of age-specific demographic rates is a subject of special interest in many disciplines as demography, biostatistics, actuarial practice, and social planning. For estimating the unknown age-specific probabilities of the various demographic phenomena, some graduation technique must be applied to the corresponding empirical rates, under the assumption that the true probabilities follow a smooth pattern through age. The classical way for graduating demographic rates is parametric modelling. However, for graduation purposes, nonparametric techniques can also be adapted. This work provides an adaptation, and an evaluation of kernels and Support Vector Machines (SVM) in the context of graduation of demographic rates.

Keywords: graduation, mortality pattern, fertility pattern, kernels, Support Vector Machines

*Correspondence to: Anastasia Kostaki, Department of Statistics, Athens University of Economics and Business, Greece; Email: kostaki@aueb.gr

Received: August 24, 2015; **Accepted:** October 4, 2015; **Published Online:** October 17, 2015

Citation: Kostaki A, Moguerza J M, Olivares A, *et al*. (2016). Nonparametric graduation techniques as a common framework for the description of demographic patterns. *International Journal of Population Studies*, vol.2(1): 1–20. http://dx.doi.org/10.18063/IJPS.2016.01.001.

1. Introduction

The graduation of demographic rates is a subject of special interest in demographic analysis, biostatistics, actuarial practice, and social planning. The demographer needs to describe the age-specific patterns of the various demographic phenomena in a population for various purposes such as providing population projections, constructing life tables and multiple decrement tables, as well as for calculating reproduction rates. The actuary needs a mortality and fertility basis suitable for calculations in life insurance and in designing of social security systems. Social planning also requires estimations and projections of the age-specific demographic patterns for many purposes, e.g., for designing health care systems, as well as for analysing and projecting the labour force.

In order to estimate the unknown age-specific probabilities of the various demographic phenomena underlying the empirical age-specific rates which are affected by random fluctuations, the typical way is the utilization of some graduation techniques to be applied to the empirical age-specific rates, under the assumption that the true probabilities follow a smooth pattern through age. A graduation technique, focusing to eliminate random fluctuations affecting the empirical measures, can therefore serve in order to provide a clear description of the real shape of the various age-specific patterns, and consequently provide a real basis for population analysis and projections. The classical way to graduate empirical demographic rates is to fit a model that presents these rates as a parametric function of age. For the graduation of the age-specific rates of each one of the three demographic phenomena, specific parametric models have been proposed.

Several parametric models have been proposed for the graduation of the age-specific mortality rates and many authors have contributed to the problem of estimating their parameters (Heligman and Pollard, 1980; Keyfitz, 1982; Forfar, McCutcheon, and Wilkie, 1988; Kostaki, 1992; Hannerz, 1999; Karlis and Kostaki, 2000). A variety of models presenting the empirical age-specific fertility rates as a parametric function of age have also been proposed for the graduation of the age-specific fertility rates. A thorough description of these models is provided by Kostaki and Peristera (2007). Finally, for the description of nuptiality patterns alternative parametric models have been proposed (Coale and McNeil, 1972; Liang, 2000). However, for graduation purposes, a possible way to smooth demographic rates is the utilization of non-parametric smoothing techniques. Kernels have already been used for graduating mortality patterns (Copas and Haberman, 1983; Gavin, Haberman, and Verrall, 1993; Gavin, Haberman, and Verrall, 1994; Felipe, Guillen, and Nielsen, 2000). An evaluation of kernels as tools for graduating mortality patterns is provided by Peristera and Kostaki (2005).

An alternative nonparametric way for graduating age-specific demographic rates would be the utilization of Support Vector Machines (SVM). These techniques appeared in 1995 in the framework of Vapnik's Statistical Learning Theory (Vapnik, 1995; Moguerza and Muñoz, 2006) for classification and regression purposes. In particular, SVM have been used in a number of applications (Chongfuangprinya, Kim, Park *et al*., 2011; Erdogan, 2013). They have been also used successfully for smoothing noisy data such as neighbourhood curves (Muñoz and Moguerza, 2005) and nonlinear profiles (Moguerza *et al*., 2007). Therefore, they can *a priori* be considered as a promising tool for demographic graduation tasks. In addition, the use of SVMs is affordable by practitioners with a lack of advanced statistical or computational skills. The reason is that documentation at all levels is available through the Internet and new libraries and easy-to-use software are continuously being developed (see Weka[1] or the software package known as "R"[2]).

The focus of this paper is to evaluate and compare the performance of kernels and SVMs for graduation purposes of demographic rates for each one of the three basic demographic phenomena. Both kernels and SVMs have been adjusted and applied to empirical data sets of mortality, fertility, and nuptiality rates of a variety of populations and years. In particular, a cross-validation approach has been conducted for the SVM models and a plug-in technique has been used for kernel models, in order to fit their corresponding parameters. For comparison purposes parametric models are also fitted to the same empirical data sets. In the next section, a short description of existing parametric models for fitting mortality, fertility, and nuptiality data is provided. Sections 3 and 4 are devoted to a presentation of kernels and SVMs, respectively. Then, Section 5 provides the results of our calculations in order to assess the utilization of kernels and SVM techniques as tools for estimating age-specific mortality, fertility, and nuptiality patterns. Some concluding remarks and some issues for further research are given in Section 6.

2 Parametric Models

2.1 Mortality Models

A wide variety of mortality laws has been presented in the literature (Brass, 1971; Mode and Busby, 1982) since the first attempt by de Moivre in 1725. Among all of these laws, the most successful attempt to describe the age-specific mortality pattern for the total life span through a parametric

1 https://weka.wikispaces.com/LibSVMor
2 https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Classification/SVM

model and the most widely used since 1980 is the one proposed by Heligman and Pollard (1980). This model (hereafter HP) is described by the formula,

$$\frac{q_x}{p_x} = A^{(x+B)^C} + De^{-(E(\ln x - \ln F)^2} + GH^x,$$

where q_x is the probability of dying within a year, $p_x=1-q_x$, and A to H are parameters to be estimated.

2.2 Fertility Models

A variety of alternative have been proposed in the literature. In this section, a summary of parametric models for fitting the age-specific fertility curve is provided.

The Hadwiger function (Hadwiger, 1940; Gilje, 1969) takes the form:

$$f(x) = \frac{ab}{c}\left(\frac{c}{x}\right)^{\frac{3}{2}} \exp\left\{-b^2\left(\frac{c}{x} + \frac{x}{c} - 2\right)\right\},$$

where a, b, and c are parameters to be estimated and x is the age of the mother at birth.

The Gamma function (Hoem, Madsen, Nielsen *et al*., 1981) is expressed by:

$$f(x) = R\frac{1}{\Gamma(b)c^b}(x-d)^{b-1} \exp\{-((x-d/c))\}, \qquad fo \;\; xr > d$$

where R, b, c, and d are parameters that should be estimated.

The Beta function by Hoem *et al*. (1981) is given by:

$$f(x) = R\frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}(\beta-\alpha)^{-(A+B-1)}(x-\alpha)^{A-1}(\beta-x)^{B-1}, \quad for \;\; \alpha < x < \beta,$$

where R determines the level of fertility, and A, B, α, and β are calculated as:

$$B = \left\{\frac{(v-\alpha)(\beta-v)}{\tau^2} - 1\right\}\frac{\beta-v}{\beta-\alpha} \text{ and } A = B\,\frac{v-\alpha}{\beta-v},$$

v being the mean and τ^2 the variance.

The Schmertmann (2003) model for representing age-specific fertility schedules is obtained using a piecewise quadratic spline function by defining three index ages that describe the shape of the age-specific fertility:

$$f(x) = \begin{cases} R * \sum_{k=0}^{4} \theta_k (x-t_k)_+^2, & \alpha \leqslant x \leqslant \beta \\ 0, & otherwise, \end{cases}$$

with Knots $t_0 < t_1 < \ldots < t_4$ falling in the interval between ages α and β, where $t_0 = \alpha$, (the lowest age of childbearing) and $(t-t_k)_* \equiv MAX[0, x-t_k]$.

A deviation from its classical shape in terms of a bulge in fertility rates of younger women is exhibited by recent fertility patterns of some developed countries. Chandola *et al*. (1999) developed a two-component mixture model of the Hadwiger function for the description of distorted fertility patterns:

$$f(x) = am\left(\frac{b_1}{c_1}\right)\left(\frac{c_1}{x}\right)^{3/2} \exp\left\{-b_1^2\left(\frac{c_1}{x} + \frac{x}{c_1} - 2\right)\right\} + (1-m)\left(\frac{b_2}{c_2}\right)\left(\frac{c_2}{x}\right)^{3/2} \exp\left\{-b_2^2\left(\frac{c_2}{x} + \frac{x}{c_2} - 2\right)\right\},$$

Where x is the age of the mother at birth, and parameters m, α, b_1, c_1, b_2, and c_2 are to be estimated, resulting in a seven parameter model by the inclusion of an additional parameter (Ortega and Kohler, 2000).

A model with two versions capturing both the classical and the distorted fertility pattern was proposed by Kostaki and Peristera (2007). The simple version of the Peristera-Kostaki model (hereafter P-K model) takes the form:

$$f(x) = c_1 \exp\left[-\left(\frac{x-\mu}{\sigma(x)}\right)^2\right],$$

Where $f(x)$ is the age-specific fertility rate at age x, c_1, μ, and σ are parameters to be estimated, while $\sigma(x) = \sigma_{11}$ *if* $x \leqslant \mu$, and $\sigma(x) = \sigma_{12}$ *if* $x > \mu$.

The version capturing the distorted fertility pattern of the Peristera and Kostaki model (hereafter P-K mixture model) is a mixture model given by:

$$f(x) = c_1 \exp\left[-\left(\frac{x-\mu_1}{\sigma_1(x)}\right)^2\right] + c_2 \exp\left[-\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right],$$

Where $f(x)$ is the age-specific fertility rate at mother age x, while $\sigma(x) = \sigma_{11}$ if $x \leqslant \mu$ and $\sigma(x) = \sigma_{12}$ *if* $x > \mu$ and c_1, c_2, μ_1, μ_2, σ_{11}, σ_{12}, σ_{11}, σ_2 are parameters to be estimated.

2.3 Nuptiality Models

Next, we provide a brief description of different parametric models proposed in literature for the fitting of empirical first-marriage rates.

Coale and McNeil (1972) defined the probability density function (hereafter C-M) for the age distribution of first-marriages as:

$$f(x) = \frac{\beta}{\Gamma(a/\beta)} \exp\left[-a\left(x - \mu - \exp\{-\beta(x-\mu)\}\right)\right],$$

where Γ denotes the gamma function, and α, β, μ are parameters to be estimated.

The generalized log gamma model (hereafter GLG) proposed by Kaneko (1991, 2003) is expressed by:

$$f(x; C, u, b, \lambda) = C \frac{|\lambda|}{b\Gamma(\lambda^{-2})} \left(\lambda^{-2}\right)^{\lambda^{-2}} \exp\left[\lambda^{-1}\left(\frac{x-u}{b}\right) - \lambda^{-2} \exp\left\{\lambda\left(\frac{x-u}{b}\right)\right\}\right],$$

where $f(x)$ is the age-specific first marriage rate at age x, C, λ, and u are parameters to be estimated and Γ denotes the gamma function.

Since in recent years a considerable variation is observed in the pattern of first-marriage in data sets of several populations, Liang (2000) built a mixture model using the double-exponential distribution. This model, denoted as the mixture Coale-McNeil model (hereafter MC-M), is described by:

$$f(x; m, \alpha_1, \lambda_1, \mu_1, \alpha_2, \lambda_2, \mu_2) = \frac{m\lambda_1}{\Gamma\left(\frac{\alpha_1}{\lambda_1}\right)} \exp(-\alpha_1(x-\mu_1) - e^{-\lambda_1(x-\mu_1)})$$

$$+ \frac{(1-m)\lambda_2}{\Gamma\left(\frac{\alpha_2}{\lambda_2}\right)} \exp\left(-\alpha_2(x-\mu_2) - e^{-\lambda_2(x-\mu_2)}\right),$$

where m, α_1, λ_1, μ_1, α_2, λ_2, and μ_1 are parameters to be estimated.

3. Kernel Techniques

Let (x_i, y_i), $i = 1,\ldots,p$ be a set of observations of two variables X and Y whose relation is given by an unknown regression function $m(x)$:

$$y_i = m(x_i) + \varepsilon_i, \; i = 1, \cdots, p,$$

where ε_i are independent random variables with zero mean and constant variance. In order to estimate the unknown function *m* at a point *x*, an averaging of the values of the response variable is locally done. The smoothness of the resulting estimator is controlled by a bandwidth determining the width of the neighbourhood over which the averaging is performed. As a result, the estimator of the function *m* takes the form:

$$\hat{m}_h(x) = n^{-1}\sum W_h(x; X_1, X_2, \cdots, X_n)Y_i,$$

where W_h is a weight function depending on the bandwidth parameter *h* and variables $X_1, X_2, \cdots, X_n$. The shape of the weight function W_h is represented by a so-called kernel function, which includes the bandwidth *h* that adjusts the size and the form of the weights around *x*, acting as a scale parameter. Hence, kernel regression estimators correspond to local weighted averages of the response variable, with weights determined by the kernel function *K*, depending on the size of the weights on the bandwidth parameter. Usually, for regression purposes, *K* performs and has the properties of a probability density function: it is generally a positive, smooth function, decreasing monotonically as the bandwidth parameter increases in size and peaking at zero.

A detailed review of the formulae proposed in the literature for the kernel estimator $\hat{m}$ of the regression mean function *m* can be consulted in Peristera and Kostaki (2005), where it is shown that the Gasser-Müller estimator (Gasser and Müller, 1979, 1984) is an adequate estimator for the graduation of mortality data, its formula being:

$$\hat{m}_{GM}(x) = \sum_{i=1}^{n} Y_{[i]} \int_{(x_{(i)}+x_{(i-1)})/2}^{(x_{(i+1)}+x_{(i)})/2} K_h(x - x_i)dx,$$

where $x_0 = -\infty$, $x_n = \infty$, x_i denotes the i^{th} largest value of the observed covariate values and $Y_{[i]}$ the corresponding response value.

Regarding the selection of the bandwidth parameter, a description of techniques can be consulted in Hardle (1990, 1991), and Peristera and Kostaki (2005). A typical way to select the bandwidth parameter is to build a direct plug-in estimator of the optimal smoothing parameter *h*. Gasser *et al*. (1991) described how unknown quantities can be effectively estimated and explicit expressions for *h* appropriate to the Gasser-Müller estimator are provided. The selection of a global or a local bandwidth is another crucial decision. A local selection allows the use of a smaller bandwidth in areas of high density, while for areas of low density a larger bandwidth can be adopted (Brockmann *et al*., 1993; and Hermann, 1997, for discussions on the advantages of using kernel regression estimators with a local bandwidth). The underlying idea of the plug-in method is to select the optimal bandwidths by estimating the asymptotically optimal mean integrated squared error bandwidths. Hermann (1997) developed a generalization of the global iterative plug-in algorithm of Gasser *et al*. (1991) for the selection of a local bandwidth, and the advantages of the local selection over the global plug-in rule and the cross-validation method are shown.

4. Support Vector Machines

The SVM technique is part of the regularisation methods (Moguerza and Muñoz, 2006). These methods also include Splines. In fact, there is a close relation between both methodologies — SVM and Splines (Pearce and Wand, 2006). Next, we provide a brief description of the regression version of SVM and its main features. SVM can be presented from its geometrical interpretation. Basically, the method works by solving an optimization problem of the form (Tikhonov and Arsenin, 1977):

$$\min_{f \in H_K} \frac{1}{p}\sum_{i=1}^{p} L\left(f(x_i) - y_i\right) + M\|f\|_K^2,$$

where (x_i, y_i), $i = 1, K$, and *p* are a set of data with $x_i \in \Re^n$ and $y_i \in \Re$, L is a loss function, $M > 0$ is a constant that penalizes non-smoothness, H_K is a space of functions known as Reproducing Kernel Hilbert Space (RKHS) (Aronszajn, 1950; Moguerza and Muñoz, 2006), and $\|f\|_K$ is the norm of f in the RKHS. The loss function L measures the estimation error of the method and $\|f\|_K$ is a measure for non-smoothness. The smaller $\|f\|_K$ is, the smoother f becomes. This means that the function $f^* \in H_K$ obtained as the solution of this optimization problem will be the result of a compromise between accuracy and smoothness. As a consequence, this way to proceed seems to be a nice approximation for the graduation of demographic data. Moreover, the optimization problem to solve is convex and therefore, without local minima. This convex property is one of the main differences with other methods, avoiding the possible existence of local solutions.

Another key issue of SVM is its ability to map the data into a higher-dimensional space (known as "feature space"). To achieve this task, a kernel approach is used in order to operate in the feature space. A kernel K is a real-valued function $K(x,y) \in \Re$ where usually $x, y \in \Re^n$, which makes the role of a scalar product in the feature space. In this way, the explicit coordinates in this higher-dimensional space are never calculated, as only the inner products between the images of all pairs of data in the feature space are needed. Three of the most widely used kernels: the linear kernel $K(x,y) = x^T y$ which corresponds to the identity mapping; the polynomial kernel $K(x,y) = (c + x^T y)^d$, where c and d are constants, which maps the data into a finitely dimensional space; and the Gaussian kernel $K(x,y) = e^{-\frac{|x-y|^2}{\sigma}}$, where σ is a positive constant, which maps the data into an infinitely dimensional space. The Gaussian kernel, given its approximation capacity, is the most extensively used (Moguerza and Muñoz, 2006), and the one that we suggest for graduation purposes.

In practical implementations of the method, such as the one provided by the software R, the accuracy and smoothing properties are achieved by fixing a band determined by a constant $\varepsilon > 0$ around the solution $f^* \in H_K$. In order to penalize strong violations of the band, another constant $C > 0$ is used. The constant ε makes the role of the loss function and C performs the control of smoothness. As a consequence, three parameters are to be fixed when using SVM with the Gaussian kernel, namely: ε, σ, and C. In practice, a grid of parameters can be determined visually taking into account that the problem at hand is one-dimensional. Then, a so called cross-validation is performed, that is, a random search within the grid is done in order to find the best combination of the parameters.

5. Evaluation and Comparisons

5.1 Numerical Results for Mortality

In our calculations we used the empirical age-specific mortality rates of the male and female populations in Sweden, for the time periods of 1981–1985, 1984–1988, and 1991–1995, as well as those in France and Japan for the years of 1990, 1991, and 1995. The Swedish data sets were taken from Statistics Sweden while the French and Japanese ones were parts of the Berkeley mortality database (2005) available from the web.

For kernel applications, the subroutine "*lokerns*" of the library "*lokern*" for the R-package is used for the calculation of Gasser-Muller estimators with local bandwidth parameter. This is available from http://cran.r-project.org/web/packages/lokern/index.html. In order to select bandwidth for a local linear Gaussian kernel regression estimator, a direct plug-in technique (Ruppert, Sheather, and Wand *et al*., 1995) is used. The initial bandwidth parameter is derived using the KernSmooth library in R package. In particular, for this implementation we obtained an initial bandwidth h = 2.3849.

The parameters in the Heligman-Pollard model are estimated using an iterative routine of the Nag library that is based upon a modification of the Gauss-Newton algorithm, described by Gill and Murray (1978). The model was fitted using weighted non-linear least squares, minimizing the following sum of squares:

$$\sum_{x} w_x \left(\hat{q}_x - q_x\right)^2 \tag{4.1}$$

With weights w_x the reciprocals of the estimated variances of the age-specific mortality rates $w_x = E_x / q_x(1-q_x)$, where E_x is the exposed-to-risk population at age x and q_x is the mortality rate at age x.

For the SVM applications, the subroutine "*svm*" of the library *e*1071 for the *R*-package is used for the derivation of the SVM model parameters. This is available from http://cran.r-project.org/. In order to select the parameters ε, σ, and C for the ε-regression procedure, the previously tioned cross-validation technique was conducted. Since the search within the grid of parameters involves randomness, for the sake of replicability, we provided the final combination of parameters used in the experiments. In particular, the values $\varepsilon = 0.02$, $\sigma = 125$ and $C = 2200$ have been chosen for this SVM implementation.

Although the graphical representation of the observed and the graduated rates is a useful way for deriving conclusions, we also used a statistical criterion in order to evaluate the performance of the alternative estimators. To check the closeness of the graduated rates to the observed ones, we used the χ^2 criterion, (4.1) that was used as minimizing criterion for fitting HP model.

The values of the criterion (4.1) for all the data sets used, and all the graduation techniques applied, are presented in Table 1. Examining these values, one can easily observe that the SVM

Table 1. Values of (4.1) at the exit of the estimation procedure for HP, SVM, and Kernels

		HP	SVM	Kernels
Sweden				
	Females			
	1981–1985	950	725	2842
	1984–1988	861	293	1817
	1991–1995	1468	882	2507
	Males			
	1981–1985	180	717	3813
	1984–1988	191	485	3125
	1991–1995	268	490	3340
Japan				
	Females			
	1990	4370	453	1767
	1991	3849	568	1859
	1995	3516	320	1601
	Males			
	1990	1140	495	2219
	1991	951	300	2047
	1995	542	394	2023
France				
	Females			
	1990	2887	594	3508
	1991	1995	639	2897
	1995	879	366	1839
	Males			
	1990	983	786	4685
	1991	687	999	4625
	1995	987	1117	2697

graduation proves adequate in terms of goodness of fit. Considering the value of 4.1 quantities, these values are lower in all female cases for SVM than for the HP model and kernels for the Swedish and the Japanese data sets. In males, SVM are better than kernels in all cases, they are a bit higher compared to the HP model for the Swedish data and in comparable levels in the French and Japanese data sets.

Figures 1–6 illustrates the results of each technique separately for some chosen cases. It is clear in these figures that the results of SVM are closer to the empirical data than those of HP formula, the later exhibiting some systematic deviations in the early adult ages. It is also clear that SVM provides smoother results than kernels. In order to do that more clearly, we compared only kernels and SVMs (Figures 2A and 4A).

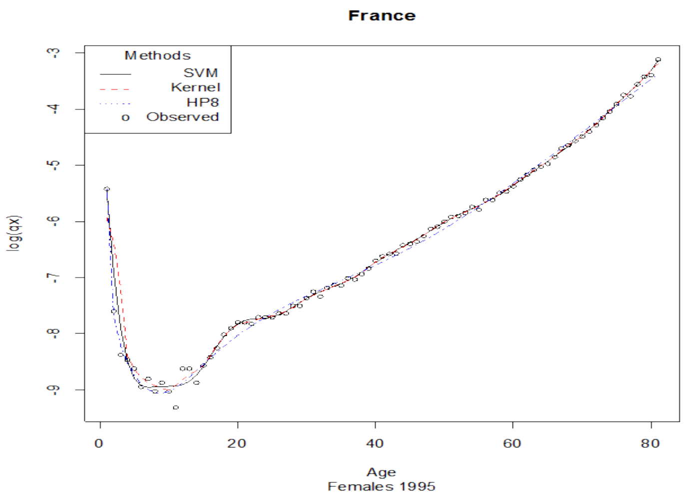

Figure 1. Empirical and graduated q_x-values, French females, 1995.

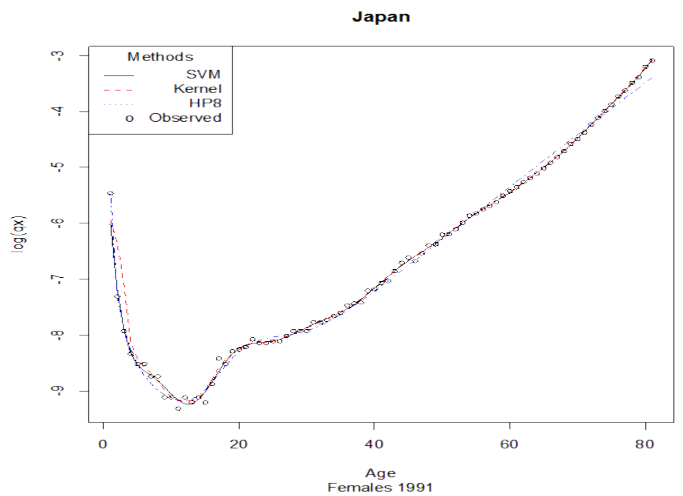

Figure 2. Empirical and graduated q_x-values, Japanese females, 1991.

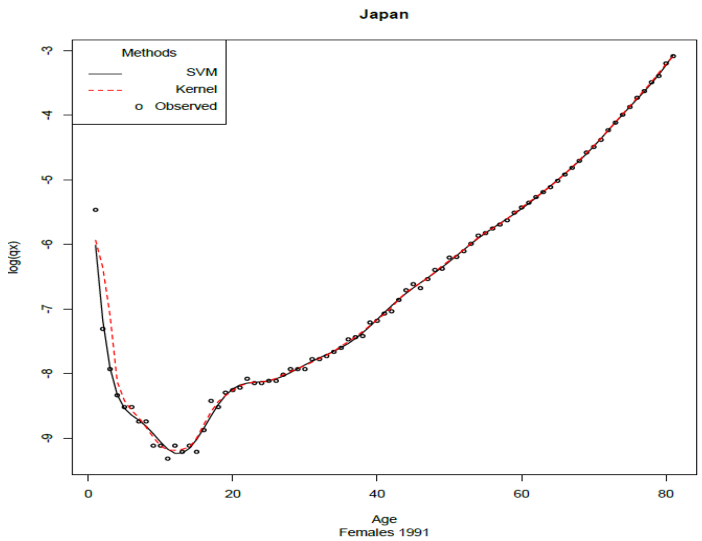

Figure 2A. Empirical and graduated q_x-values, Japanese females, 1991.

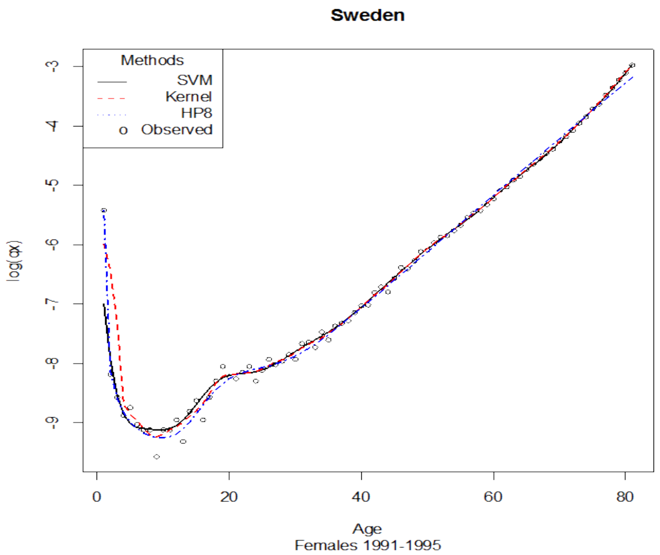

Figure 3. Empirical and graduated q_x-values, Swedish females, 1991–1995.

5.2 Numerical Results for Fertility

In order to evaluate SVM as a tool for graduating age-specific fertility patterns, we used period age-specific fertility rates for the populations of Sweden, Norway, and Denmark (1996 and 2000); Belgium (1993 and 1995); Greece and Italy (1995 and 2000); UK (1992 and 2000); Ireland (1995 and 2000); the white and black populations of the USA (2003); and for Spain (1942 and 1963). The empirical data sets were obtained from Eurostat New Cronos database (http://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal/NewCronos.aspx). Additionally, single year age-specific fertility rates for the US were derived for the 2003 Natality Data Set, obtained after a request from the US National Center of Health Statistics (http://www.cdc.gov/nchs/). Cohort data are also used for Spain for the generations born in 1943 and 1962, obtained from the Eurostat New Cronos database. It should be noted that even for cohorts not yet completed, Eurostat provides estimates of the fertility rates for older women by using the rates observed for previous generations, without waiting for the cohort to reach the end of the reproductive period. Parity-specific birth rates were computed as occurrence exposure rates based on parity in marriage.

Then, we applied SVMs and kernels and also provided the fits of the alternative parametric models to these data sets, the latter initially calculated by Kostaki and Peristera (2007). In populations with no apparent early-age hump, except of kernels and SVMs, the fits of Hadwiger, Gamma, and Beta models (Chandola, Coleman, and Hiorns, 1999; Hoem, Madsen, Nielsen *et al*., 1981), P-K model (Kostaki

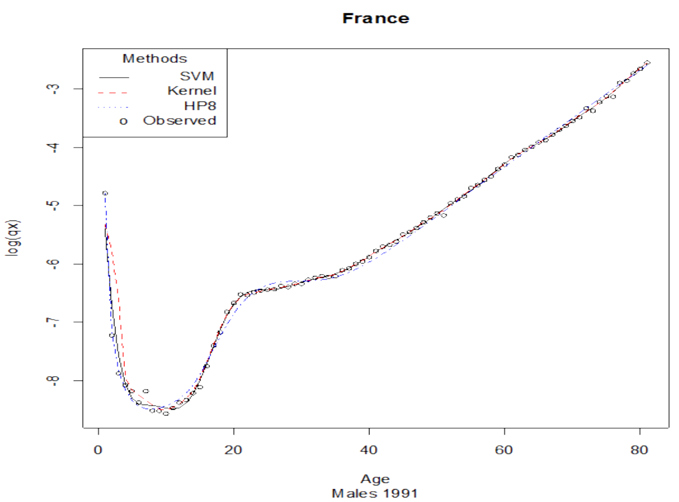

Figure 4. Empirical and graduated q_x-values, French males, 1991.

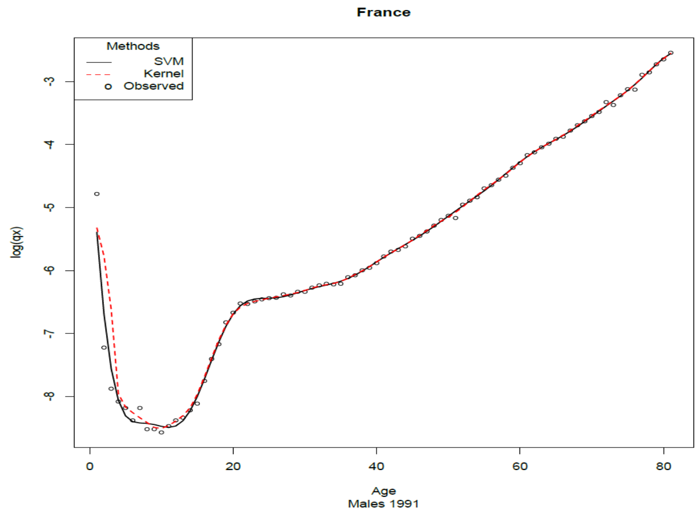

Figure 4A. Empirical and graduated q_x-values, French males, 1991.

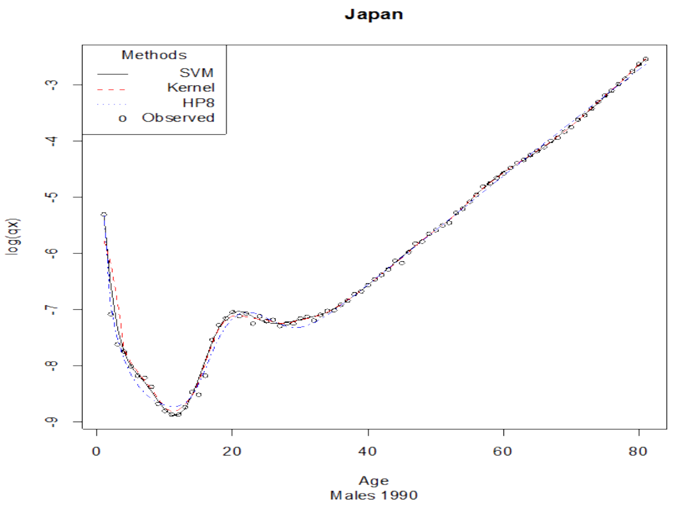

Figure 5. Empirical and graduated q_x-values, Japanese males, 1990.

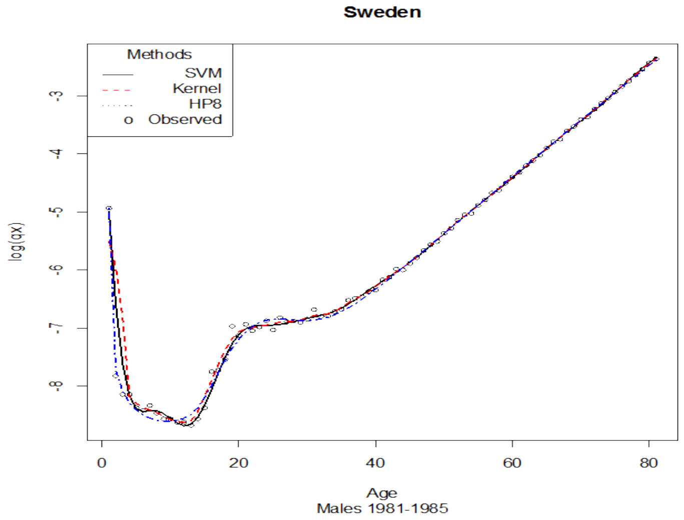

Figure 6. Empirical and graduated q_x-values, Swedish males, 1981–1985.

and Peristera, 2007), and the quadratic Spline model (Schmertmann, 2003) are provided, while in the cases of distorted fertility distributions, the Hadwiger mixture model (Chandola, Coleman, and Hiorns, 1999; 2002) and the P-K mixture model (Kostaki and Peristera, 2007) are provided.

In order to avoid heterogeneity, we also used data differentiated by order of birth from both cohort and period data sets. Finally, in the case of the USA, the fits of the alternative models are provided for the white and the black population separately. Details for fitting the alternative parametric models are given by Kostaki and Peristera (2007).

The parameters of the various models have been estimated by means of a non-weighted non-linear least-squares procedure, minimizing the following sum of squares:

$$\sum_{x}\left(\hat{f}_x - f_x\right)^2, \tag{4.2}$$

where $\hat{f}_x$ is the estimated marriage rate at age x and f_x is the corresponding empirical one. This minimizing criterion has been used as most appropriate for fertility graduation by Kostaki and Peristera (2007) and also suggested by Hoem *et al*. (1981) as providing equal good fits as the more complicated weighted one, with weights reciprocal to the estimated variances of the age-specific rates, the latter being most appropriate when fitting mortality rates.

For kernel applications, in the case of mortality data, the subroutine "*lokerns*" of the library "*lokern*" for the R-package was used for the calculation of Gasser-Muller estimators with local bandwidth parameter. In a similar way, the initial bandwidth parameter was derived using the KernSmooth library in R package. An initial bandwidth of h =1.9066 was obtained particularly for this implementation.

As in the case of mortality data, for the SVM techniques, the subroutine *svm* of the library *e*1071 for the R-package is used, and a similar two-step cross-validation technique is used to select the parameters ε, σ, and C of the ε-regression procedure. Parameters ε, σ, and C play the same role as explained in the mortality study. In particular, the values ε = 0.0001, σ = 40 and C = 1.8, have been obtained for this SVM implementation.

The values of (4.2) for all the data sets used, and all graduation techniques applied, are presented in Tables 2 and 3. The results of fitting the parametric models were first presented by Kostaki and Peristera (2007). Figures 7–12 provide illustrations for some chosen cases. In all cases, we used ages ranging from 15 to 48, so each schedule has 34 rates.

As stated in the tables and figures, the results of SVM prove superior to the corresponding ones of all the other models. SVM produced results that in the vast majority of cases are closer to the empirical rates, with a sole exception, the results for the USA data differentiated by order of birth and race, where the performance of the P-K mixture model were somewhat superior. Regarding the figures, one can easily observe that the results of SVM were closer to the empirical values especially for the ages in the tails and the peak of the fertility curve.

Table 2. Values of (4.2) multiplied by 100.000, at the exit of the estimation procedure for P-K model, Beta model, Gamma model, Hadwiger model, quadratic Spline model, kernels, and SVM

SSE*10^6	P-K Model	BetaModel	Gamma Model	Hadwiger Model	Quadratic Spline Model	Kernel	SVM
Period Data							
Sweden							
1996	115	108	132	326	174	67	72
2000	117	181	321	689	174	30	11
Norway							
1992	242	175	265	656	263	65	61
2000	233	225	640	329	287	40	10
Denmark							
1992	103	107	130	383	169	54	20

Continued table2

SSE*10^6	P-K Model	BetaModel	Gamma Model	Hadwiger Model	Quadratic Spline Model	Kernel	SVM
2000	225	363	575	1073	287	51	6
Belgium							
1993	401	396	380	540	462	68	15
1995	346	374	376	558	525	78	30
Greece							
1995	190	137	184	289	101	26	14
2000	34	114	491	617	55	14	13
Italy							
1995	20	58	139	352	49	18	11
2000	47	71	524	908	82	14	3
Cohort Data							
Spain							
1943	732	1005	1159	1547	5450	452	562
1962	295	259	1113	184	3720	69	67

Table 3. Values of (4.2), multiplied by 100.000, at the exit of the estimation procedure, for P-K mixture model, Hadwiger mixture, and SVM for the US data

SSE*10^6	P-K Mixture Model	Hadwiger Mixture Model	Kernel	SVM
Period Data Total Births				
UK				
1992	154	35	37	14
2000	99	22	40	14
Ireland				
1995	437	97	62	90
2000	78	177	65	43
Spain				
1999	29	17	30	12
2000	23	15	31	6
Cohort Data Total Births				
Spain				
1963	77	85	59	62
Period Data First Births				
UK				
2004	5	8	47	4
Ireland				
2000	73	53	61	62
Period Data Second Births				
UK				
2004	4	5	45	3
Ireland				
2000	31	31	25	28
USA 2003				
Total	150	28	63	58
White	28	156	63	51
Black	39	190	103	86

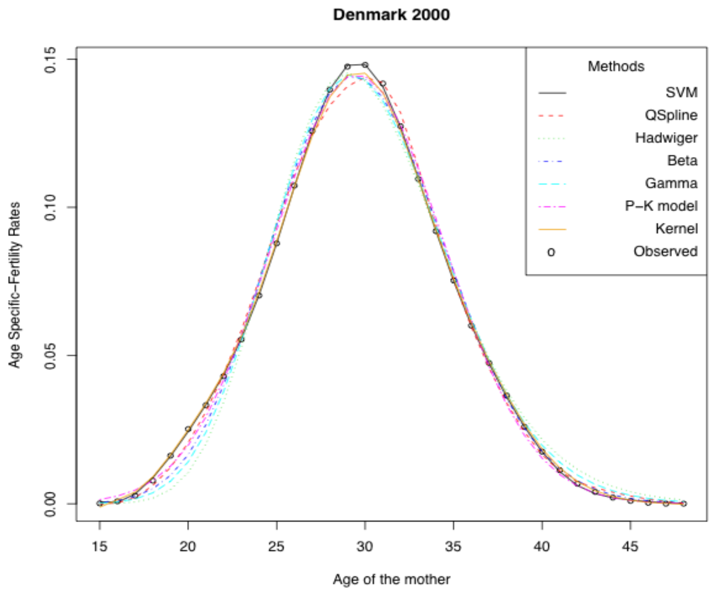

Figure 7. Observed and estimated period age-specific fertility rates for Denmark, 2000.

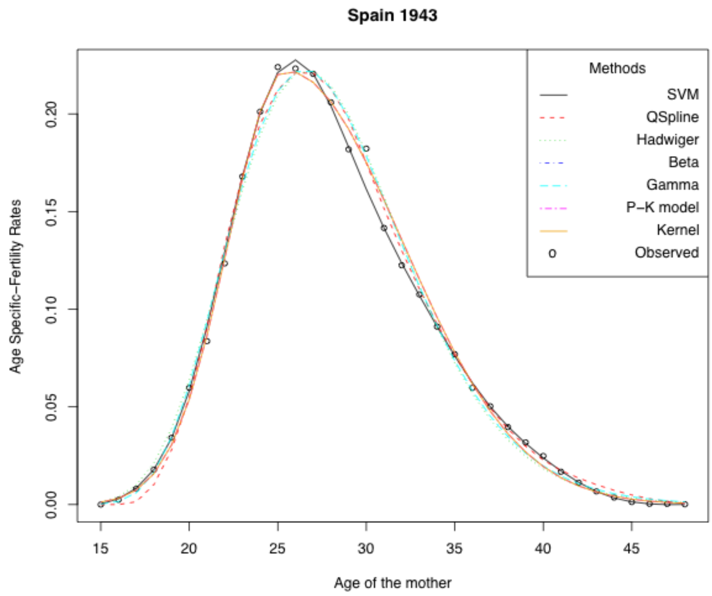

Figure 8. Observed and estimated cohort age-specific fertility rates for Spain, 1943.

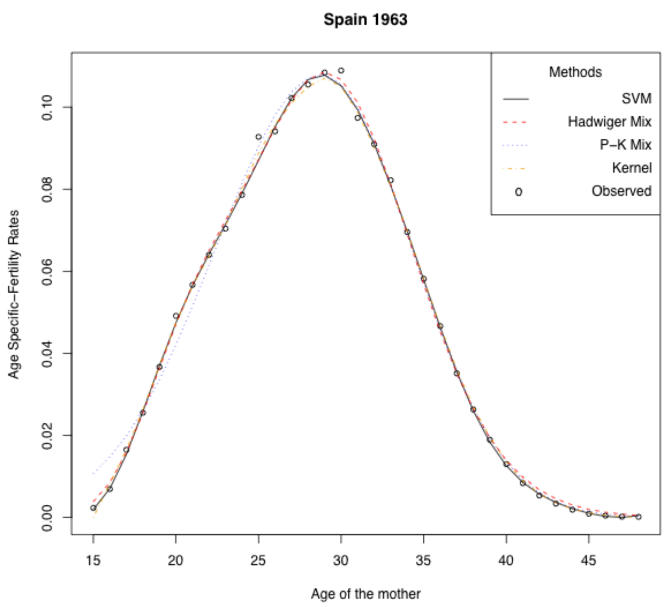

Figure 9. Observed and estimated cohort age-specific fertility rates for Spain, 1963.

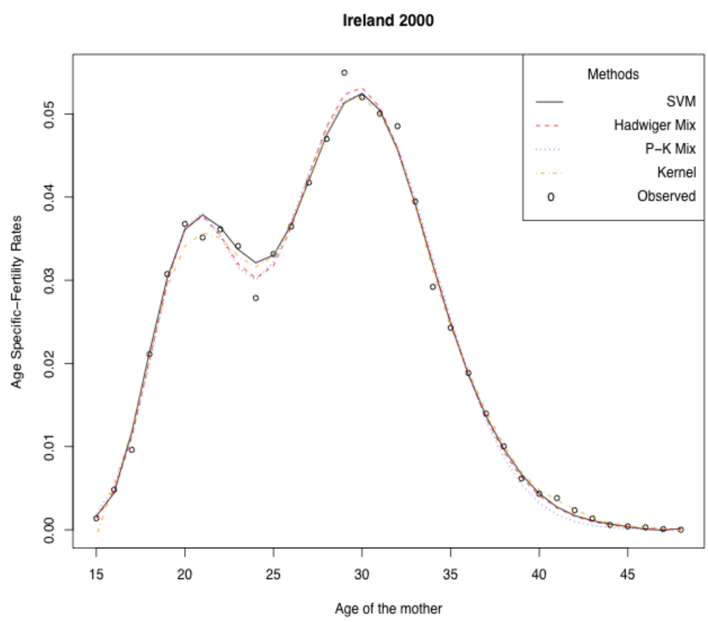

Figure 10. Observed and estimated age-specific fertility rates of Ireland, 2000. First births.

5.3 Numerical Results for Nuptiality

In order to evaluate the adequacy of SVM for nuptiality graduation purposes, we applied SVM to a variety of empirical data sets. For comparison reasons, we also fit the same data sets to the SC-M, C-M, and GLG models. In cases where these simple models fail to adequately estimate the nuptiality pattern in data sets expressing heterogeneity, we fitted the MC and the P-K mixture models.

Once again, for kernel techniques, in the case of mortality and fertility data, the subroutine "*lokerns*" of the library "*lokern*" for the R-package was used for the calculation of Gasser-Müller

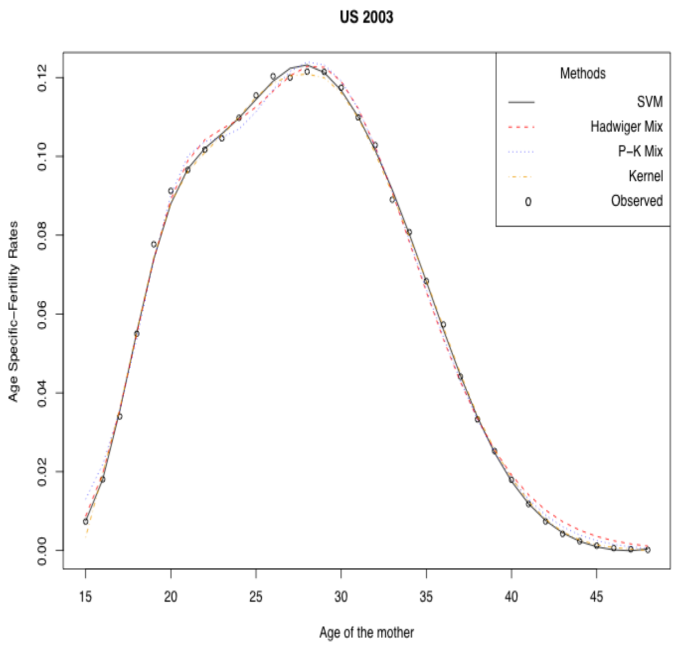

Figure 11. Observed and estimated age-specific fertility rates of US, 2003. 2003. White population.

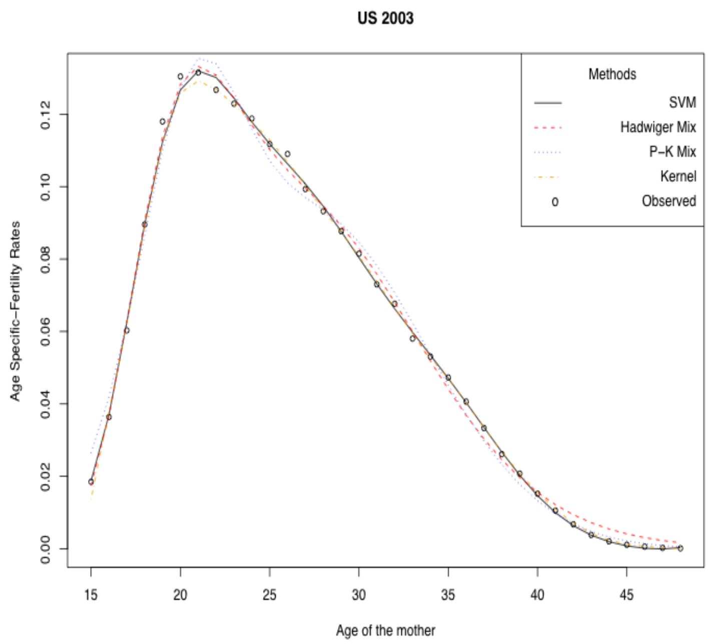

Figure 12. Observed and estimated age-specific fertility rates of US, 2003. Black population.

estimators with local bandwidth parameter. And in a similar way, the initial bandwidth parameter is derived using the KernSmooth library in R package. In particular, for this implementation, an initial bandwidth of h =1.3118 was obtained.

Also, in the case of mortality and fertility data for the SVM applications, the subroutine *svm* of the library *e*1071 for the R-package was used, and a similar two-step cross-validation technique was used to select the parameters ε, σ, and C of the ε-regression procedure. In particular, the values ε = 0.00008, σ = 0.0424, and C = 0.0662, have been obtained for this SVM implementation.

Single-year age-specific first-marriage rates for the female populations of Spain, Greece, Italy, Germany, the Netherlands, Norway, Sweden, Finland, Ireland, and the UK for the available years were used. These data were obtained from the Eurostat New Cronos database (http://www.eui.eu/Research/Library/ResearchGuides/ Economics/Statistics/DataPortal/NewCronos.aspx).

The parameters of the parametric models are estimated, as in the case of fertility for the same reasons mentioned before, by means of a least-squares procedure by minimizing the following sum of squares.

$$\sum_x (\hat{f}_x - f_x)^2, \tag{4.3}$$

where $\hat{f}_x$ is the estimated first-marriage rate at age x and f_x is the empirical one.

All the parametric models are fitted by means of a non-linear least-squares procedure and a Gauss-Newton optimization scheme. The Matlab built-in routine for non-linear parameter estimation "*lsqnonlin*" was used in order to find the unconstrained minimum of the unweighted sum of squares.

The residual sums of squares are given in Table 4. Furthermore, the empirical and graduated age-specific first-marriage rates for selected years and countries are depicted in Figures 13–22.

The observed and estimated rates of GLG, C-M, and SC-M models as well as of SVM are depicted in Figures 13–19 for the populations of Germany, Greece, Italy, the Netherlands, Norway, Spain, and the UK.

The existence of a bulge at young ages and another one at the older ones becomes obvious in recent Swedish data. The bulge during the young ages appears around age 20 and the older one around age 40. This phenomenon has also started to appear in the data sets of Finland and Ireland. In these cases, simple models fail to closely estimate the tails and the peak value of the marriage distribution. We thus fitted the mixture model MC-M to the data sets. Figures 20–22 provide illustrations of the results.

According to the values of the minimizing criterion, for the majority of the cases the C-M model provides the best fits among the parametric models. The second best fit is usually obtained by the GLG one.

As mentioned above, a variety of factors related to the socioeconomic and cultural background of male and female populations may contribute to the appearance of the heterogeneity in the first-marriage curve. However in order to be able to verify or reject all these hypotheses about heterogeneity in the first-marriage curve, further research based on empirical evidence is required.

Turning now to the SVM, we observed from the values of the residual sum of squares as well as from the graphical illustration, that their performance is superior in comparison to any other parametric approach. In the vast majority of the data sets, the values of the residual sum of squares were in significantly lower levels than those resulting by model fitting. It is probably worth mentioning that this technique works with high accuracy in both homogeneous and heterogeneous data sets, while for the later cases more complicated models are required. Taking a closer look at the figures, we observed that SVM performance is highly superior to parametric modelling, in the peaks and the tables of the marriage distributions where parametric modelling provides systematic deviations from the empirical rates.

Table 4. Values of (4.3), multiplied by 100.000, at the exit of the estimation procedure
FEMALES

SSE*10^6	Standard Coale-McNeil	GLG	Coale-McNeil	Kernel	SVM
Spain					
1995	62308	17358	14058	788	216
2002	46959	14321	12446	731	185
Greece					
2001	54891	8333	63228	331	214
2002	54891	8333	63228	289	329
Italy					
1990	38956	8654	8251	2431	2706
2000	47607	8605	6597	603	312
Germany					
1998	27263	4936	4761	786	517
2001	19583	3331	3137	279	195
Netherlands					
1996	19516	2661	2563	859	1144
2002	40108	8705	8171	372	201
Norway					
1996	29525	8001	7828	1746	3986
2002	20803	5770	5291	352	539
UK					
1996	12060	2271	2226	449	223
1999	214910	10280	10278	759	241

SSE*10^6	Standard Coale-McNeil	GLG	Coale-McNeil	Mixture Coale-McNeil	Kernel
Sweden					
1997	16568	8873	9152	8482	929
2002	32185	18168	17859	26983	327
Finland					
1993	29465	8252	8413	2251	935
1998	23595	8527	8436	6049	833
Ireland					
1993	31352	12977	12975	7389	2032
1998	62918	19153	17110	8563	886

MALES

SSE*10^6	Standard Coale-McNeil	GLG	Coale-McNeil	Kernel	SVM
Spain					
1990	52827	13188	1220	1198	496
2002	22002	9331	9261	656	250
Greece					
1998	22913	5541	5477	334	284
2002	18348	5543	5528	277	283
Italy					
1991	21338	5345	5334	829	240
2000	23853	7017	6624	360	181
Germany					
1993	13191	3127	3084	428	567
1998	21424	5043	4609	317	670
Netherlands					
1995	2350	15534	15534	603	783
2003	19898	6130	5924	195	327
Norway					
1997	13648	6041	6041	872	1744
2001	10697	4134	4138	171	228
UK					
1999	21491	3126	3088	281	288
2000	20437	2469	2460	215	223

SSE*10^6	Standard Coale-McNeil	GLG	Coale-McNeil	Mixture Coale-McNeil	Kernel	SVM
Sweden						
1997	4827	3904	3822	8482	519	1039
2001	7135	5411	5382	26983	211	273
Finland						
1995	8290	3377	4350	2251	398	453
2002	12020	7793	7791	6049	305	346
Ireland						
1996	13788	4468	4422	7389	242	317
1998	34290	12541	11680	8563	808	281

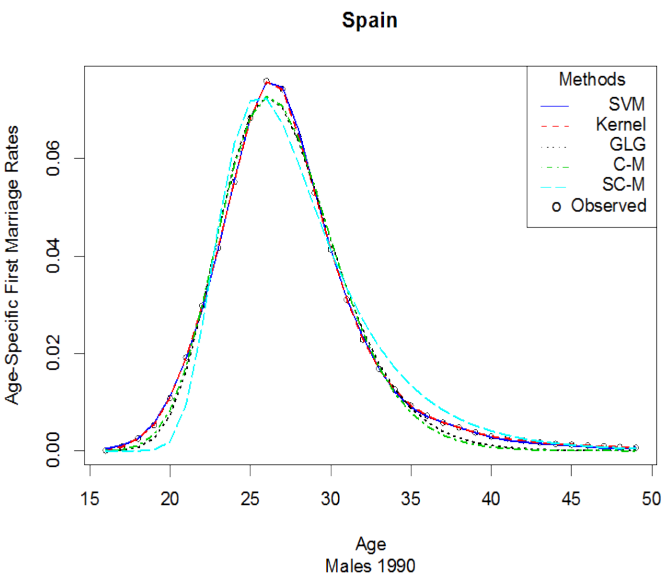

Figure 13. Observed and estimated age-specific nuptiality rates, Spain, males,

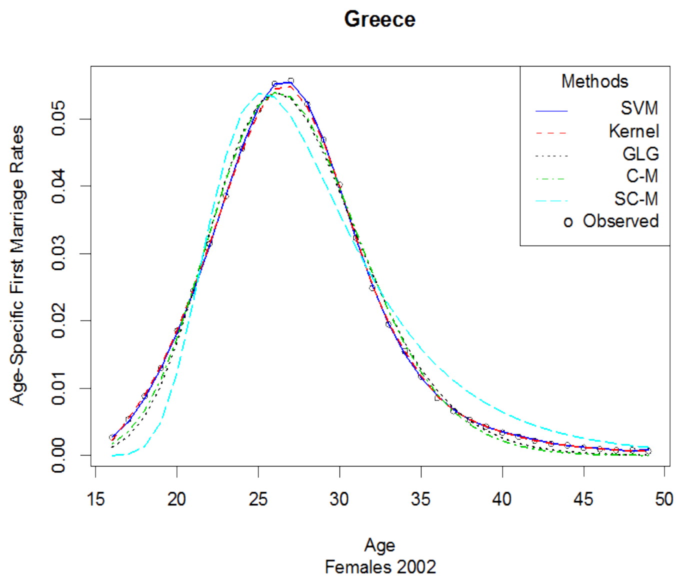

Figure 14. Observed and estimated age-specific nuptiality rates, 1990 Greece, females, 2002.

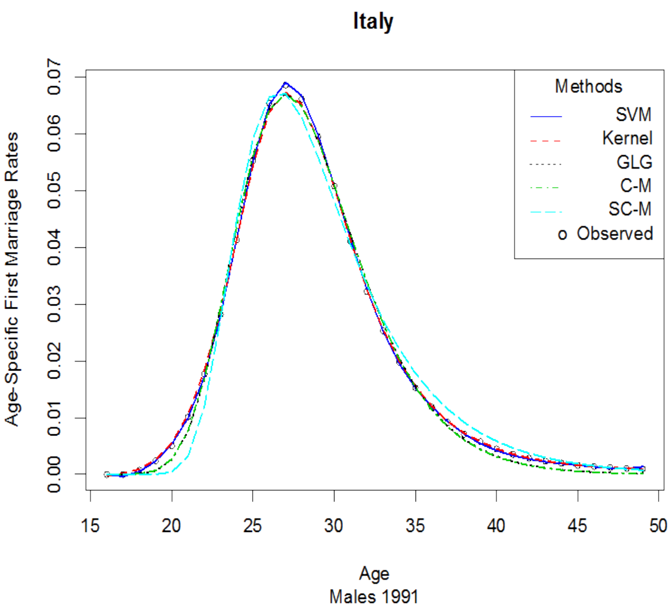

Figure 15. Observed and estimated period age-specific fertility rates, Italy, males, 1991.

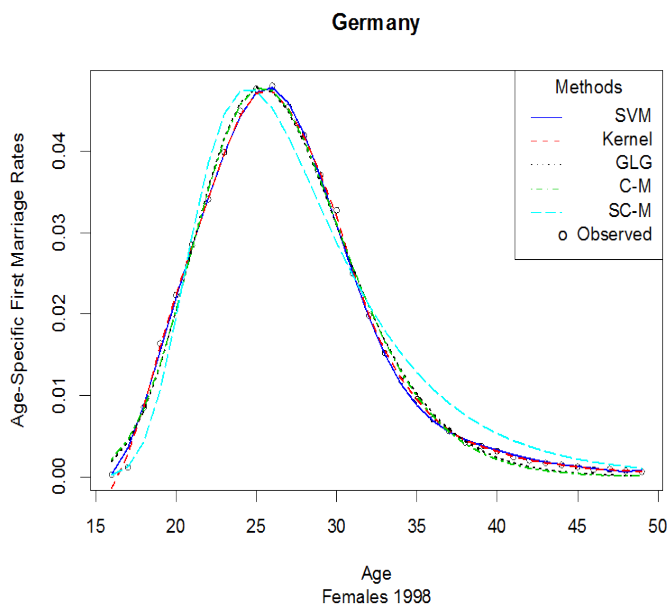

Figure 16. Observed and estimated age-specific nuptiality rates, Germany, females, 1998.

6. Conclusions

In this paper, we evaluated and compared SVMs and kernels for graduating age-specific demographic rates. The performance of these two nonparametric techniques has been evaluated by applying them to a set of empirical mortality, fertility, and nuptiality rates of different populations and time periods. Moreover, parametric models are fitted to these rates in order to compare their effectiveness. With regards to the values of the typical minimization criteria, the results for the two nonparametric techniques are apparently closer to the empirical values than those provided by the parametric models. This performance is probably due to the higher smoothness capacity of parametric models. A higher degree of smoothness may lead to larger distances between the graduated and the empirical values and, in many cases; it provides oversimplifications of the described patterns or systematic deviations between the empirical and the graduated values. Concerning the comparison between SVMs and kernels (the two nonparametric techniques), SVMs provided results, usually with lower values of the minimizing criteria.

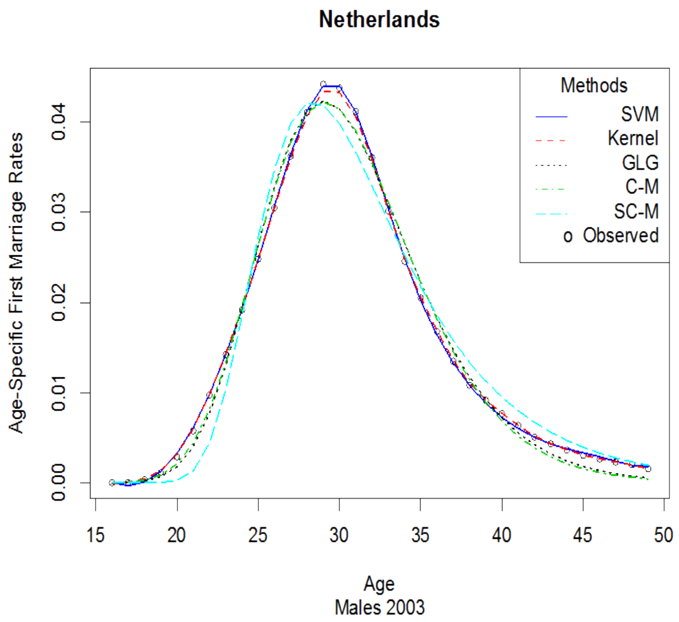

Figure 17. Observed and estimated age-specific nuptiality rates, the Netherlands, males, 2003.

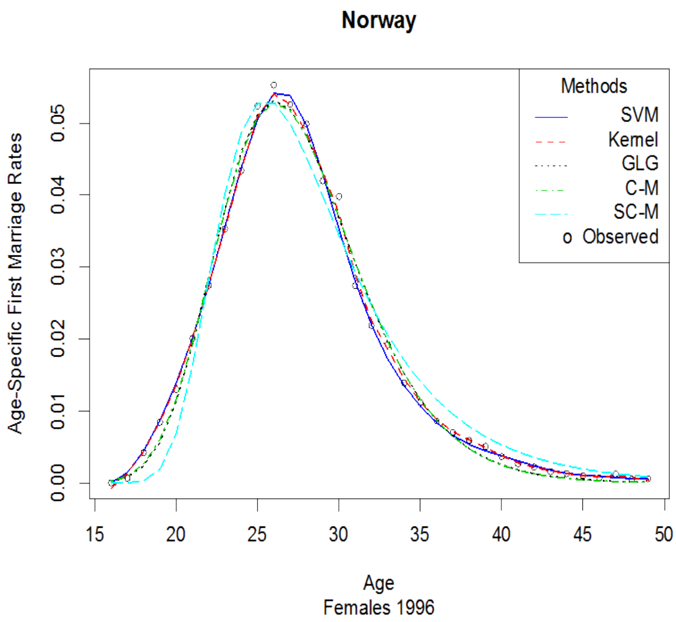

Figure 18. Observed and estimated age-specific nuptiality rates, Norway, females, 1996.

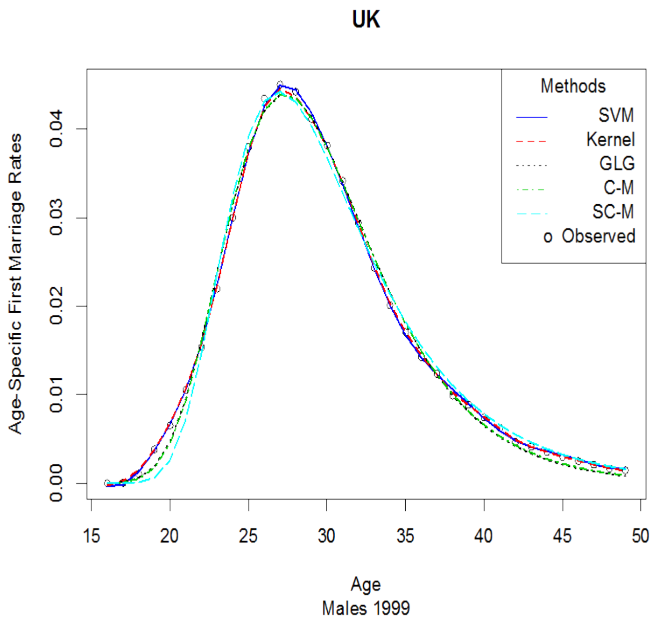

Figure 19. Observed and estimated age-specific nuptiality rates, the UK, males, 1999

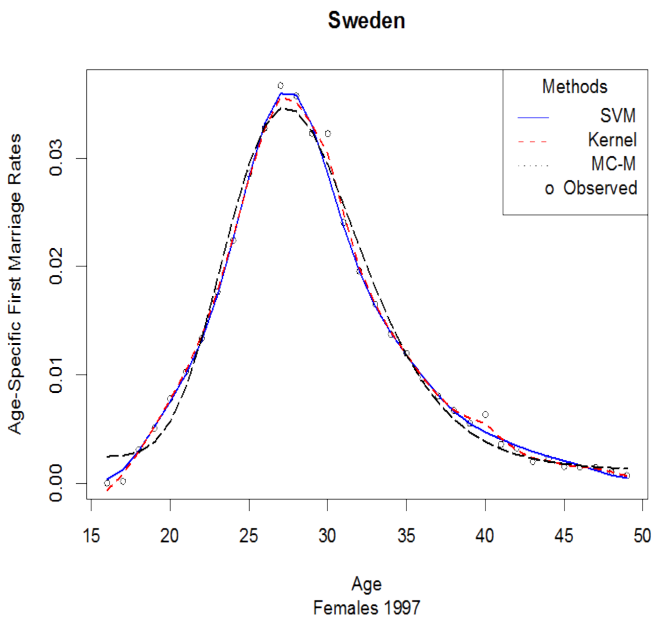

Figure 20. Observed and estimated age-specific nuptiality rates, Sweden, females, 1997.

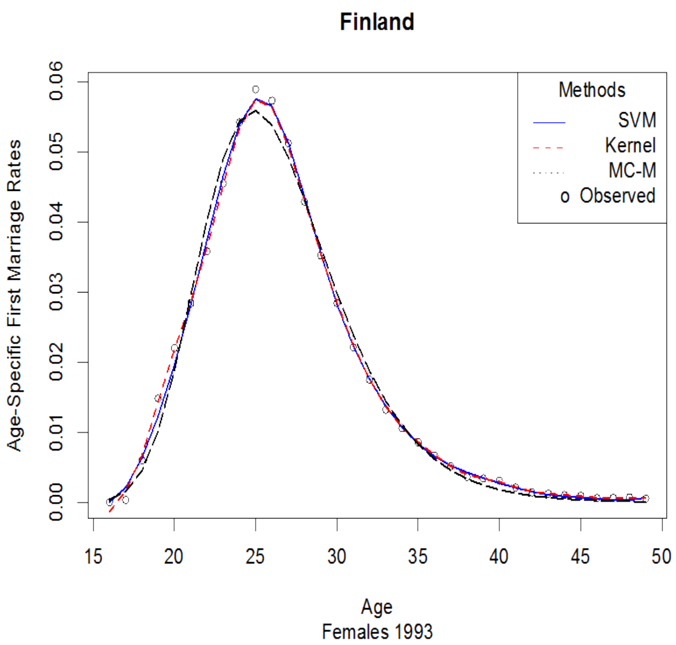

Figure 21. Observed and estimated age-specific fertility rates, Finland, females, 1993.

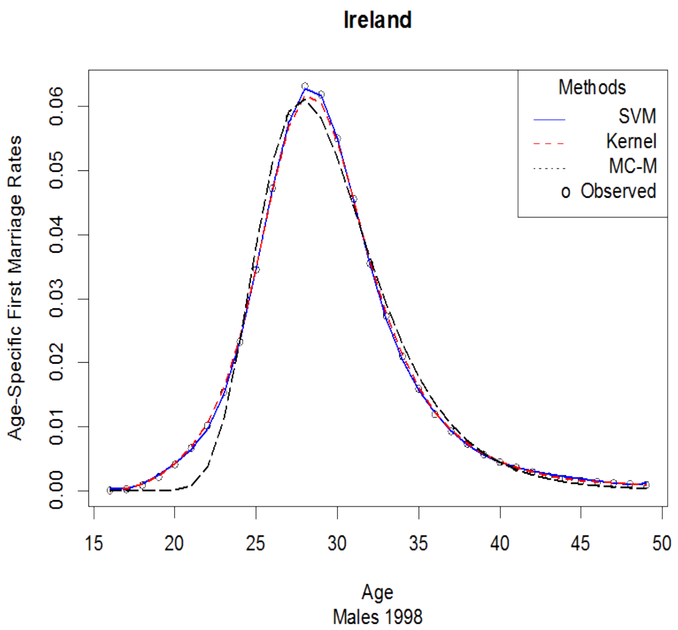

Figure 22. Observed and estimated age-specific nuptiality rates, Ireland, males, 1998.

In addition, the SVM method produces results closer to the empirical rates in most cases, showing a successful performance for the graduation of empirical rates in both simple and distorted data sets. It can be observed in the figures that the results provided by SVM were closer to the empirical data than those of most alternative methods, especially for ages in the peak and the tails of nuptiality and fertility.

Nonparametric graduation techniques have the advantage of being suitable to all data sets. This is an important remark, as for data sets with distorted patterns; the use of standard parametric models is inadequate. Another advantage of the nonparametric approach is that the user has the possibility of regulating the degree of smoothness and, as a consequence, choosing a degree adapted to the goal of the graduation framework, avoiding in many cases oversimplification of age patterns.

As a future extension of the current work, we propose the use of SVM as a multivariate model for demographic forecasting.

Conflict of Interest and Funding

No conflict of interest has been reported by the authors. This research is partially financed by the Research Centre of Athens University of Economics and Business, in the framework of the project entitled "Original Scientific Publications"; and projects GROMA (MTM2015-63710-P), PPI (RTC-2015-3580-7), and UNIKO (RTC-2015-3521-7), funded by the Ministry of Economy and Competitiveness (Spain).

Acknowledgements

We would like to thank the editor and the reviewers for their constructive comments and their valuable insight, which led to significant improvements.

References

Aronszajn N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:

337–404. http://dx.doi.org/10.1090/S0002-9947-1950-0051437-7.

Berkeley mortality database. (2005). Retrieved on October 29, 2015 from http://www.demog.berkeley.edu/~bmd.

Brass W. (1971). Mortality models and their use in demography. *Transactions of the Faculty of Actuaries*, 33: 123–142. http://dx.doi.org/10.1017/S0071368600005164.

Brockmann M, Gasser T and Herrmann E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88(424): 1302–1309. http://dx.doi.org/10.1080/ 01621459.1993.10476411.

Chandola T, Coleman D A and Hiorns R W. (1999). Recent European fertility patterns: fitting curves to 'distorted' distributions. *Population Studies*, 53(3): 317–329. http://dx.doi.org/10.1080/00324720308089.

Chandola T, Coleman D A and Hiorns R W. (2002). Distinctive features of age-specific fertility profiles in the English-speaking world: common patterns in Australia, Canada, New Zealand and the United States, 1970-98. *Population Studies*, 56(2): 181–200. http://dx.doi.org/10.1080/00324720215929.

Chongfuangprinya P, Kim S B, Park S K, *et al*. (2011). Integration of support vector machines and control charts for multivariate process monitoring. *Journal of Statistical Computation and Simulation*, 81(9): 1157–1173. http://dx.doi.org/10.1080/00949651003789074.

Coale A J and McNeil D R. (1972). The distribution by age of the frequency of first marriage in a female cohort. *Journal of the American Statistical Association*, 67(340): 743–749. http://dx.doi.org/10.1080/01621459. 1972.10481287.

Copas J B and Haberman S. (1983). Non-parametric graduation using kernel methods. *Journal of Institute of Actuaries*, 110(1): 135–156. http://dx.doi.org/10.1017/S0020268100041275

Erdogan B E. (2013). Prediction of bankruptcy using support vector machines: an application to bank bankruptcy. *Journal of Statistical Computation and Simulation*, 83(8): 1543–1555. http://dx.doi.org/10.1080/00949655.2012.666550.

Felipe A, Guillen M and Nielsen J P. (2001). Longevity studies based on kernel hazard estimation. *Insurance Mathematics and Economics*, 28(2): 191–204. http://dx.doi.org/10.1016/S0167-6687(00)00076-7.

Forfar D O, McCutcheon J J and Wilkie A D. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries*, 115(1): 1–149. http://dx.doi.org/10.1017/S0020268100042633.

Gasser T, Kneip A and Köhler W. (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Society*, 86(415): 643–652. http://dx.doi.org/10.1080/01621459.1991.10475090.

Gasser T and Müller H G. (1979). Kernel estimation of regression functions. In Th Gasser and M Rosenblatt (eds), *Lecture Notes in Mathematics: Smoothing Techniques for Curve Estimation*, 757: 23–68. Berlin: Springer. http://dx.doi.org/10.1007/BFb0098489.

Gasser T and Müller H G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 11: 171–185.

Gavin J, Haberman S and Verrall R J. (1993). Moving weighted average graduation using kernel estimation. *Insurance: Mathematics and Economics*, 12(2): 113–126. http://dx.doi.org/10.1016/0167-6687(93)90821-6.

Gavin J B, Haberman S and Verrall R J. (1994). On the choice of bandwidth for kernel graduation. *Journal of the Institute of Actuaries*, 121(1): 119–134. http://dx.doi.org/10.1017/S0020268100020102.

Gilje E. (1969). Fitting curves to age-specific fertility rates: some examples. *Statistical Review of the Swedish National Central Bureau of Statistics III*, 7: 118–134.

Gill P E and Marray W. (1978). Algorithms for the solution of a non-linear least squares problem. *SIAM Journal of Numerical Analysis*, 15(5): 977–992. http://dx.doi.org/10.1137/0715063.

Hadwiger H. (1940). Eineanalytischereprodutions-funktion fur biologischeGesamtheiten. *SkandinaviskAktuarietidskrift*. 23: 101–113.

Hannerz H. (1999). *Methodology and Applications of a New Law of Mortality*. Lund: Department of Statistics, University of Lund, Sweden.

Härdle W. (1990). *Applied Non-parametric Regression*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CCOL0521382483.

Härdle W. (1991). *Smoothing Techniques with Implementation in S*. New York: Springer New York. http://dx.doi.org/10.1007/978-1-4612-4432-5.

Heligman M and Pollard J H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(1): 49–80. http://dx.doi.org/10.1017/S0020268100040257.

Hermann E. (1997). Local bandwidth choice in kernel regression estimation. *Journal of Computational and Graphical Statistics*, 6(1): 35–54. http://dx.doi.org/10.1080/10618600.1997.10474726.

Hoem J M, Madsen D, Nielsen J L, *et al*. (1981). Experiments in modelling recent Danish fertility curves. *Demography*, 18(2): 231–244. http://dx.doi.org/10.2307/2061095.

Karlis D and Kostaki A. (2000). Bootstrap techniques for mortality models. *Biometrical Journal*, 44(7): 850–866. http://dx.doi.org/10.1002/1521-4036(200210)44:7%3C850::AID-BIMJ850%3E3.0.CO;2-6.

Kaneko R. (1991). Demographic analysis of the first marriage process. *Jinko Mondai Kenkyu*, 47(3): 3–27.

Kaneko R. (2003). Elaboration of the Coale-McNeil nuptiality model as the generalised log gamma distribution: a new identity and empirical enhancements. *Demographic Research*, 9(10): 223–262. http://dx.doi.org/10.4054/DemRes.2003.9.10.

Keyfitz N. (1982). Choice of function for mortality analysis: effective forecasting depends on a minimum parameter representation. *Theoretical Population Biology*, 21(3): 329–352. http://dx.doi.org/10.1016/0040-5809(82)90022-3.

Kostaki A. (1992). Nine-parameter version of the Heligman-Pollard formula. *Mathematical Population Studies*, 3(4): 277–288. http://dx.doi.org/10.1080/08898489209525346.

Kostaki A and Peristera P. (2007). Modeling fertility in modern populations. *Demographic Research*, 16: 141–194. http://dx.doi.org/10.4054/DemRes.2007.16.6.

Liang Z. (2000). The Coale-McNeil model: theory, generalisation and application. Retrieved on October 20, 2016 from http://www.popline.org/node/172229.

Mode C J and Busby R C. (1982). An eight-parameter model of human mortality —the single decrement case. *Bulletin of Mathematical Biology*, 44(5): 647–659. http://dx.doi.org/10.1007/BF02462273.

Moguerza J M and Muñoz A. (2006). Support vector machines with applications. *Statistical Science*, 21(3): 322–336. http://dx.doi.org/10.1214/088342306000000493.

Moguerza J M, Muñoz A and Psarakis S. (2007). Monitoring nonlinear profiles using support vector machines, in L Rueda, D Mery and J Kittler (eds), *Progress in Pattern Recognition, Image Analysis and Applications*, 4789: 574–583. Berlin: Springer. http://dx.doi.org/10.1007/978-3-540-76725-1_60.

Muñoz A and Moguerza J M. (2005). Building smooth neighbourhood kernels via functional data analysis, in W Duch, J Kacprzyk, E Oja, *et al*. (eds), *Artificial Neural Networks: Formal Models and Their Applications – ICANN 2005*, 3697: 631–636. Berlin: Springer. http://dx.doi.org/10.1007/11550907_100.

Ortega Osona J A and Kohler H P. (2000). A comment on “Recent European fertility patterns: fitting curves to ‘distorted’ distributions”, by T Chandola, D A Coleman and R W Hiorns. *Population Studies*, 54(3): 347–349. http://dx.doi.org/10.1080/713779092.

Pearce N D and Wand M P. (2006). Penalized splines and reproducing kernel methods. *The American Statistician*, 60(3): 233–240. http://dx.doi.org/10.1198/000313006X124541.

Peristera P and Kostaki A. (2005). An evaluation of the performance of kernel estimators for graduating mortality data. *Journal of Population Research*, 22: 185–197. http://dx.doi.org/10.1007/BF03031828.

Ruppert D, Sheather S J and Wand M P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90 (432): 1257–1270. http://dx.doi.org/10.1080/01621459.1995.10476630.

Schmertmann C P. (2003). A system of model fertility schedules with graphically intuitive parameters. *Demographic Research*, 9: 82–110. http://dx.doi.org/10.4054/DemRes.2003.9.5.

Schölkopf B, Smola A J, Williamson RC, *et al*. (2000). New support vector algorithms. *Neural Computation*, 12(5): 1207–1245. http://dx.doi.org/10.1162/089976600300015565.

Tikhonov A N and Arsenin V Y. (1977). *Solutions of Ill-posed Problems*. New York: John Wiley and Sons. http://dx.doi.org/10.2307/2006360.

Vapnik V. (1995). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4757-2440-0.