Stat
The ISI's Journal for the Rapid Dissemination of Statistics Research
(wileyonlinelibrary.com) DOI: 10.100X/sta.0000

Bivariate smoothing of mortality surfaces with cohort and period ridges

Alexander Dokumentov[a*], Rob J Hyndman[b], Leonie Tickle[c]

Received 6 June 2018; Accepted 6 July 2018

Mortality rates typically vary smoothly over age and time. Exceptions occur due to events such as wars and epidemics which create ridges in the age-period surface of mortality rates in a particular year or for cohorts born in a particular year. We propose a new practical method for modelling the age-period surface of mortality rates. Our approach uses L_1 regularization with bivariate smoothing, and allows for age-varying period and cohort ridges in the otherwise smooth surface. Cross validation on data from many countries and from simulations demonstrates that our approach is superior to existing approaches in estimating the "true" age-period mortality surface. It also provides greater insight into the underlying mortality dynamics, informing mortality modelling, analysis and forecasting. Although designed for the modelling of mortality rates, our method can also be applied to any bivariate data with occasional ridges, and extends the statistical literature on quantile smoothing.

Keywords: Bivariate data; nonparametric smoothing; graduation; cohort effects; period effects

1. Introduction

Mortality rates are used to compute life tables, life expectancies, insurance premiums and reserves, and other items of interest to demographers and actuaries. However, observed mortality rates are noisy, and it is useful to smooth or "graduate" them in order to obtain estimates of the underlying rates. Events such as wars and epidemics can affect mortality rates as either period effects (different mortality rates in a particular calendar year) or cohort effects (different mortality rates among those born in a particular year). These appear as ridges in the otherwise smooth mortality surface, and any effective smoothing method applied to mortality data needs to allow for such features.

Figure 1 shows log mortality rates $m_{x,t} = \log(M_{x,t})$ for females in France from 1950 to 1970. The observed mortality rate at age x in year t is $M_{x,t} = D_{x,t}/E_{x,t}$, where $D_{x,t}$ is the number of deaths aged x in year t, and $E_{x,t}$ is the total

[a]13 Beacon St., Parkdale VIC 3195, Australia
[b]Department of Econometrics & Business Statistics, Monash University, Clayton VIC 3800, Australia
[c]Department of Actuarial Studies and Business Analytics, Macquarie University, North Ryde NSW 2109, Australia
*Email: alexander.dokumentov@gmail.com

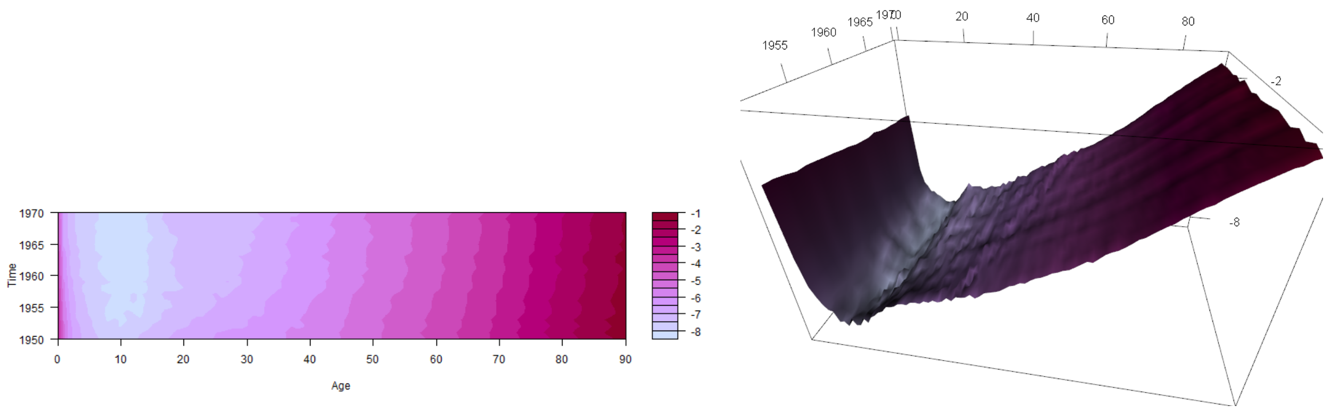

Figure 1. Natural logarithms of French female mortality rates (years 1950–1970).

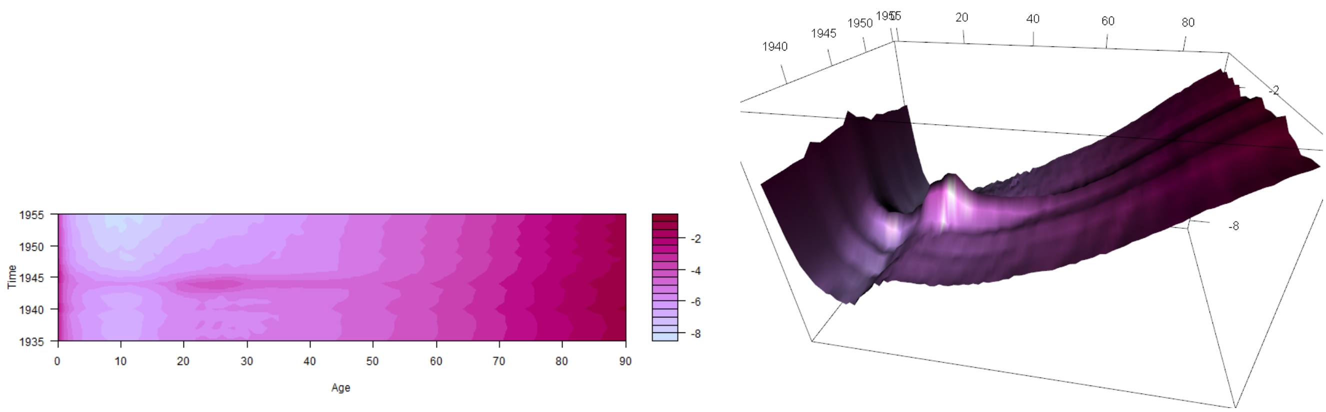

Figure 2. Natural logarithms of French female mortality rates (years 1935–1955).

number of years lived at age x during year t, which can be approximated by the mid-year population. The data was sourced from the Human Mortality Database (2008).

The following features of the log mortality surface are evident:

- In the age dimension, the log mortality decreases rapidly at the early ages and reaches a minimum at about age 10. There is an "accident hump" around age 20 in some years (Heligman & Pollard, 1980), after which log mortality increases almost linearly to the oldest ages.
- At almost every age, mortality decreases over time during the period, more steeply (on the log scale) for younger ages than for older ages.
- There are diagonal ridges (along $t = x + k$ for a specific k) due to cohort effects (these are somewhat difficult to discern in Figure 1 but are analyzed in depth in later sections). Studies have indicated a relationship between mortality and year of birth in numerous countries including France and the UK (Willets, 2004; Janssen & Kunst, 2005; Andreev & Vaupel, 2005; O'Connell & Dunstan, 2009), particularly for males (Richards, 2008). Cohort differences in smoking prevalence as well as pre-natal and early life factors are thought to play a role (Janssen & Kunst, 2005; Richards et al., 2006).

- There are horizontal ridges (for a fixed calendar year t) due to period effects. Such patterns are usually due to extreme environmental events such as wars and pandemics, which affect all people (with different magnitude) during a particular year. Figure 2, covering the period 1935 to 1955, illustrates marked period effects due to World War II. Major period effects in other periods are due to the Spanish influenza epidemic of 1918, which resulted in worldwide deaths of the order of 50 million (Johnson & Mueller, 2002), and to World War I. Less extreme period effects are also evident in other years.

Ideally, any model of the mortality surface should distinguish and model all components: the smooth underlying bivariate age and time mortality surface; additional smooth functions of age for some cohorts; additional smooth functions of age in some calendar years; and random errors. Mortality analysis or forecasting may then proceed with a more comprehensive and nuanced understanding of the relevant patterns and their underlying causes.

Existing smoothing methods address this problem to some extent, but not completely. A number of methods that have been developed are not tailored to the mortality surface. Early work in functional data analysis and nonparametric smoothing tended to involve general methods that could be applied to a wide range of problems (see, for example Silverman & Ramsay, 2005; Ferraty & Vieu, 2006; Horváth & Kokoszka, 2012). More recently, the trend in functional data research has been to develop or adapt methods that take account of the specific features of the problem at hand — such as jump discontinuities in neural images (Zhu et al., 2014) and amplitude and phase variation in nuclear magnetic resonance spectral data (Marron et al., 2015) — or that are applied to new types of data including trees (Shen et al., 2014) and graphs (Zhu et al., 2016). Ideas from functional data analysis are also being applied in other areas of mathematics and statistics including delay differential equations (Brunel et al., 2014) and automated variable selection in functional linear models (Gertheiss et al., 2013). The methods that will be developed in this paper contribute to this broader literature and address particular issues that arise in the modelling of the mortality surface.

Some smoothing methods have been proposed that are tailored to mortality rates but which smooth only in the age direction. Examples include the nonparametric functional data analysis approach of Hyndman & Ullah (2007) which finds a smooth function $f_t(x) = \mathrm{E}[m_{x,t}]$ separately for each t, the Whittaker-Henderson method and generalizations (Whittaker, 1922; Henderson, 1924; Schuette, 1978), and the many approaches that fit parametric functions to mortality rates across age in a particular year (Forfar et al., 1988), including Heligman & Pollard (1980), Carriere (1992) and Hannerz (2001). However, in the absence of wars and other extreme events, it is reasonable to expect that mortality rates will progress smoothly over time as well as age, and can therefore be modelled using a smooth bivariate function $f(x,t) = \mathrm{E}[m_{x,t}]$. By allowing the assumption of smoothness in both dimensions, better performance should be possible due to the additional information included in the estimation.

Currie et al. (2004) use two-dimensional P-splines for smoothing and forecasting mortality rates, and Camarda (2012) implement this approach along with a one-dimensional version in the R package MortalitySmooth. By using a global smoother across the whole age range, these methods can fail to deal with the steep drop in mortality between birth and age 10 (Camarda, 2012). Camarda et al. (2012) use special bases for P-splines to better model the steep decline in mortality at the youngest ages for the univariate case, and Camarda et al. (2010) for both the univariate and bivariate cases. Other bivariate smoothing approaches include regularized singular value decomposition (Huang et al., 2009) and repeated functional observations (Chen & Müller, 2012). However, none of these methods handle period and cohort effects.

One issue with many bivariate smoothing approaches is their sensitivity to outliers due to the use of an L_2 norm rather than an L_1 norm in the estimation process. The L_2 norm penalizes large changes in the smoothed function much more heavily than smaller and more gradual changes and therefore is more likely to over-smooth abrupt features in the data compared to an L_1 norm. The L_1 norm has been found to improve robustness in the univariate case when the data contains outliers (Schuette, 1978; Portnoy, 1997).

Finally, most bivariate mortality smoothing methods fail to preserve cohort and period effects, instead removing them from the estimated mortality surface. Kirkby & Currie (2010) extend the approach of Currie et al. (2004) to model period (but not cohort) effects: the method is based on a Poisson model of deaths using a GLM for estimation, with the period effects estimated in a multi-step procedure. Barbi & Camarda (2011) extend Currie et al. (2004) to incorporate age-independent period and cohort effects. However, they acknowledge that use of age-independent effects is simplistic, and they call for the development of more sophisticated models that incorporate age-varying cohort and period effects.

In Section 2 we propose a new practical bivariate mortality smoothing method that addresses the shortcomings of existing approaches. Our approach incorporates age-varying cohort and period effects, and uses an L_1 norm to ensure preservation of features in the data and robustness to outliers. It can be viewed as an extension of quantile smoothing splines (see, for example, Koenker et al., 1994; Portnoy, 1997; He et al., 1998) with partial differential regularization (Sangalli, 2014), and may be applied to any bivariate data with ridges. We call our proposed method RESPECT, an acronym for REgularized Smoothing with PEriod and CohorT effects. It also emphasises our attempt to respect what is in the data, rather than to smooth it away.

In Section 3, we demonstrate the RESPECT method by applying it to the data shown in Figures 1 and 2. The performance of the RESPECT method against its major competitors is evaluated in Section 4 using a cross validation procedure. We provide some discussion and conclusions in Section 5.

2. Regularized smoothing with period and cohort effects

2.1. *Thin plate splines*

A two-dimensional thin plate regression spline is defined as the function $f(x, t)$ which minimizes

$$J(\{y_i\}_{i=1}^n, f) = \sum_{i=1}^{n} (y_i - f(x_i, t_i))^2 + \lambda \int \left[\left(\frac{\partial^2 f}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial t}\right)^2 + \left(\frac{\partial^2 f}{\partial t^2}\right)^2 \right] \mathrm{d}x\, \mathrm{d}t \qquad (1)$$

for some smoothing parameter $\lambda > 0$, knots $\{(x_i, t_i)\}_{i=1}^n$ and values $\{y_i\}_{i=1}^n$ (see, for example, Wood, 2006). The first term measures lack of fit between the observed and smoothed log mortality rates, the second term measures lack of smoothness via second derivatives across age, time and their interaction, and λ is the weight given to smoothness relative to fit in the estimation. The L_2 norm (i.e. squared measures) is used for both the fit and smoothness elements.

Thin plate regression splines work well in many cases. However, direct application of thin plate splines (or adaptive thin plate splines where flexibility varies) to the log mortality rates does not lead to good results, especially for early ages. This is the same problem noted by Camarda (2012). We speculate that the reasons for the problems are twofold:

(a) The log mortality rate surface is steep at early ages, and twisted along the time dimension due to a more rapid decrease in mortality for younger ages compared to older ages; and
(b) thin plate splines penalize big errors much more heavily than small errors.

This leads to the situation where abrupt jumps in the mortality rates generate errors in the proximity of the jumps, causing unsatisfactory performance of thin plate splines over the abrupt surface.

2.2. Bivariate quantile smoothing

If the knots form a fine regular grid, then the integral in (1) can be approximated by a sum and so $J(\{y_i\}_{i=1}^n, f)$ can be approximated as

$$J\left(\{y_i\}_{i=1}^n, f\right) \approx \sum_{i=1}^n (y_i - f(x_i,t_i))^2 + \frac{\lambda}{n}\sum_{i=1}^n \left[\left(\frac{\partial^2 f}{\partial x^2}(x_i,t_i)\right)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}(x_i,t_i)\right)^2 + \left(\frac{\partial^2 f}{\partial y^2}(x_i,t_i)\right)^2\right].$$

Also if the knots form a fine regular grid, then the second partial derivatives at knots can be approximated as linear combinations of function values at nearby knots. Denoting $\{y_i\}_{i=1}^n$ as vector y and $\{f(x_i, y_i)\}_{i=1}^n$ as vector z, then $J(\{y_i\}_{i=1}^n, f)$ can be approximated as

$$J(y,z) \approx \|y-z\|_{L_2}^2 + \frac{\lambda}{n}\left(\|D_{xx}z\|_{L_2}^2 + 2\|D_{xt}z\|_{L_2}^2 + \|D_{tt}z\|_{L_2}^2\right)$$

where D_{xx}, D_{xt} and D_{tt} are linear operators (matrices) which calculate approximations of vectors $\left\{\frac{\partial^2 f}{\partial x^2}(x_i,t_i)\right\}_{i=1}^n$, $\left\{\frac{\partial^2 f}{\partial x \partial t}(x_i,t_i)\right\}_{i=1}^n$ and $\left\{\frac{\partial^2 f}{\partial t^2}(x_i,t_i)\right\}_{i=1}^n$.

Using the above expression, we can approximate a thin plate spline computed at its knots as

$$S(y) = \arg\min_z \left(\|y-z\|_{L_2}^2 + \frac{\lambda}{n}\left(\|D_{xx}z\|_{L_2}^2 + 2\|D_{xt}z\|_{L_2}^2 + \|D_{tt}z\|_{L_2}^2\right)\right). \quad (2)$$

Here, y contains the two-dimensional data packed as a vector. The order of packing affects only the representation of matrices D_{xx}, D_{xt} and D_{tt}.

The above model uses the L_2 norm for both the fit and smoothness components, which is optimal in the case of normality (of errors in the case of fit, and second derivatives in the case of smoothness), because it gives a minimizing function that is equivalent to a maximum likelihood approach. In the case of fit, normality may be justified as deaths follow a Poisson distribution (Brillinger, 1986) which can be approximated by a normal distribution when there are many deaths, and hence errors can also be assumed close to normal. However, it is not robust to outliers. In the case of smoothness, a normal distribution of second derivatives is not always supported by data. For example, Figure 3 shows that the second derivatives of log mortality rates can deviate considerably from normality in some time periods. The distributions have longer tails (or outliers) and therefore assumptions of normality can lead to suboptimal smoothing.

Hence there is justification for replacing the L_2 norm with the L_1 norm in (2) giving L_1 regularized median smoothing. In addition, we use three different λ coefficients before every derivative to separately adjust the influence of each derivative on the smoothing. Thus we define smoothing as $Q(y) = \arg\min_z(K(y,z))$ where

$$K(y,z) = \|y-z\|_{L_1} + \lambda_{xx}\|D_{xx}z\|_{L_1} + \lambda_{xt}\|D_{xt}z\|_{L_1} + \lambda_{tt}\|D_{tt}z\|_{L_1},$$

y, D_{xx}, D_{xt} and D_{tt} are as described above, and the L_1 norm of vector v is defined as $\|v\|_{L_1} = \sum_i |v_i|$.

Minimization of $K(y,z)$ can be reduced to a quantile regression problem (described for example in Wood, 2006), which then can be solved with existing software (Koenker, 2015). First, matrices I, $\lambda_{xx}D_{xx}$, $\lambda_{xt}D_{xt}$, and $\lambda_{tt}D_{tt}$ are stacked on top of each other to give $R = [I,\ \lambda_{xx}D'_{xx},\ \lambda_{xt}D'_{xt},\ \lambda_{tt}D'_{tt}]'$. Then vector y is extended by zeros until its length is equal to the number of rows in R: $y_{ext} = [y',\ 0']'$. Finally, $K(y,z) = \|y-z\|_{L_1} + \lambda_{xx}\|D_{xx}z\|_{L_1} + \lambda_{xt}\|D_{xt}z\|_{L_1} + \lambda_{tt}\|D_{tt}z\|_{L_1}$ is replaced with the equivalent expression $K(y,z) = \|y_{ext} - Rz\|_{L_1}$. Then finding $Q(y) = \arg\min_z(K(y,z))$ is a quantile regression problem.

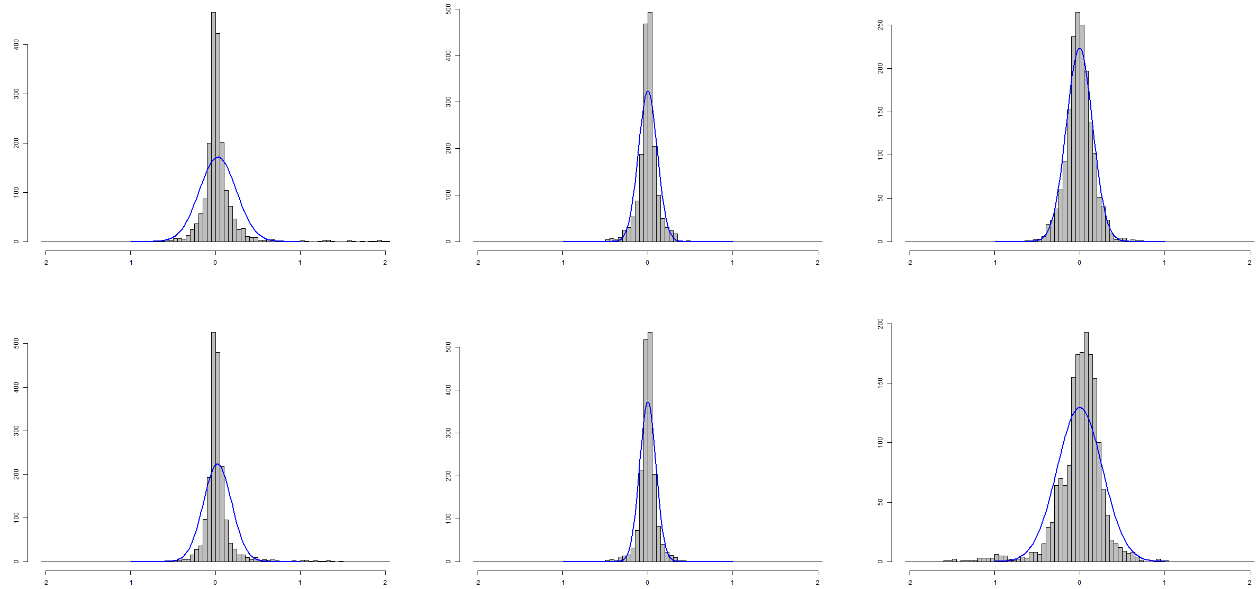

Figure 3. Histograms of second differences $\left(\frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial x \partial t}, \frac{\partial^2}{\partial t^2}\right)$ of natural logarithms of French female mortality rates with estimated normal densities. Top row: years 1950–1970. Bottom row: years 1935–1955.

Missing values in vector y can be handled by simply removing the corresponding indexes when computing $\|y - z\|_{L_1} = \sum_i |y_i - z_i|$ (the first term of $K(y, z)$). The other terms of $K(y, z)$ stay intact in this case.

We estimate the parameters, λ_{xx}, λ_{xt} and λ_{tt}, by minimizing the mean absolute error based on five-fold cross validation. We use Nelder-Mead optimization with the `optim` function in R (R Core Team, 2018).

For each fold of the cross-validation, the training data has about 20% of missing values with the same (but shifted) pattern (Figure 4). The test data are omitted in a regular pattern to ensure the distance between them is as large as possible to ensure independent errors.

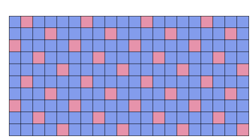

Figure 4. Missed data pattern for pseudo leave-one-out cross validation.

2.3. Cohort and period effects

To identify the period and cohort effects, we compute the residuals from the L_1 regularized median smoothing algorithm described in the previous section.

To identify the cohort effects, we split the matrix of the residuals into a set of vectors (of different length) representing diagonals, and carry out the following tests over each diagonal.

1. Perform two-sided t-tests of the residuals over all diagonals to find diagonals with residual mean values significantly different from zero.
2. Perform one-sided sample correlation tests for residual diagonals to find diagonals with positively correlated errors (every diagonal is tested for serial correlation with lag one).

To identify the period effects, we perform the same procedure over every column (representing the same year) of the residuals:

3. Perform two-sided t-tests of the residuals over all years to find years with residual mean values significantly different from zero.
4. Perform one-sided sample correlation tests to find years with positively correlated errors (residuals for every particular year are tested for serial correlation with lag one along age dimension).

Because we run multiple tests, there is a high probability that some of them will give false positive results. However, these are done only for the purpose of reducing computational complexity of the minimization problem. Some or all of these cohort and period effects will be dropped in the subsequent minimization.

Our RESPECT method involves summing four components: smooth mortality rates, cohort effects being non zero only along the diagonals identified in tests 1 and 2 above, period effects being non zero only along years identified in tests 3 and 4 above, and the noise. These four components are estimated using the following model:

$$Q(y) = \underset{z_{sm}, z_{coh}, z_{per}}{\arg\min} \left(K(y, z_{sm}, z_{coh}, z_{per})\right), \tag{3}$$

where

$$\begin{aligned} K(y, z_{sm}, z_{coh}, z_{per}) = \|y - (z_{sm} + z_{coh} + z_{per})\|_{L_1} + \lambda_{xx}\|D_{xx}z_{sm}\|_{L_1} + \lambda_{xt}\|D_{xt}z_{sm}\|_{L_1} + \lambda_{tt}\|D_{tt}z_{sm}\|_{L_1} \\ + \lambda_{coh}\|D_{coh}z_{coh}\|_{L_1} + \theta_{coh}\|z_{coh}\|_{L_1} + \lambda_{per}\|D_{per}z_{per}\|_{L_1} + \theta_{per}\|z_{per}\|_{L_1}; \end{aligned} \tag{4}$$

- y, D_{xx}, D_{xt} and D_{tt} are as described above;
- z_{sm}, z_{coh} and z_{per} are estimated components representing respectively smooth mortality surface, cohort effects restricted to some diagonals and period effects restricted to some years;
- D_{coh} is a linear differentiation operator representing a discrete version of the second directional derivative in the diagonal cohort direction;
- $D_{per} = D_{tt}$ is a linear differentiation operator representing a discrete version of the second derivative along the years axis;
- $\lambda_{xx}, \lambda_{xt}$, and λ_{tt} are parameters controlling the smoothness of the mortality surface;
- λ_{coh} and λ_{per} are parameters controlling the smoothness of the cohort effects and the period effects respectively;
- θ_{coh} and θ_{per} are parameters controlling shrinking (respectively) the cohort effects and the period effects towards zero.

It may appear that components z_{sm}, z_{coh} and z_{per} duplicate each other. However, this is not the case because the best cross-validation performance is achieved when λ_{coh} and λ_{per} are much greater than values of parameters λ_{xx}, λ_{xt} and λ_{tt}; then z_{coh} and z_{per} tend to reflect long 1d features and z_{sm} tends to reflect 2d features (which are smaller in diameter than the length of the 1d features). Thus, z_{coh} and z_{per} reflect cohort and period effects and z_{sm} reflects the smooth surface. We speculate that the separation of features is the reason for the improvement of cross-validation performance. Knowing such behavior, we restrict λ_{coh} and λ_{per} values to be much higher than values of parameters λ_{xx}, λ_{xt} and λ_{tt} to speed up the process of parameter estimation (technically it is done by setting ranges within which the parameters can vary).

We optimize five parameters, λ, and two parameters, θ, using the approach described in Section 2.2. First we run the L_1 regularized median smoothing method to find starting values for the λ_{xx}, λ_{xt} and λ_{tt} parameters; these are then optimized along with the other parameters in fitting model (3).

To minimize $K(y, z_{sm}, z_{coh}, z_{per})$, we solve the quantile regression problem in which:

- vectors z_{sm}, z_{coh} and z_{per} are stacked on top of each other as a single vector, $z_{ext} = [z'_{sm},\ z'_{coh},\ z'_{per}]'$;
- matrices I, $\lambda_{xx}D_{xx}$, $\lambda_{xt}D_{xt}$, $\lambda_{tt}D_{tt}$, $\lambda_{coh}D_{coh}$, $\lambda_{per}D_{per}$, $\theta_{coh}I$, and $\theta_{per}I$ are combined in one matrix,

$$R = \begin{bmatrix} I & I & I \\ \lambda_{xx}D_{xx} & 0 & 0 \\ \lambda_{xt}D_{xt} & 0 & 0 \\ \lambda_{tt}D_{tt} & 0 & 0 \\ 0 & \lambda_{coh}D_{coh} & 0 \\ 0 & \theta_{coh}I & 0 \\ 0 & 0 & \lambda_{per}D_{per} \\ 0 & 0 & \theta_{per}I \end{bmatrix};$$

- vector y is extended by zeros to have its length equal to the number of rows in R: $y_{ext} = [y',\ 0']'$;
- and

$$K(y, z_{sm}, z_{coh}, z_{per}) = \|y - (z_{sm} + z_{coh} + z_{per})\|_{L_1} + \lambda_{xx}\|D_{xx}z_{sm}\|_{L_1} + \lambda_{xt}\|D_{xt}z_{sm}\|_{L_1} + \lambda_{tt}\|D_{tt}z_{sm}\|_{L_1} + \lambda_{coh}\|D_{coh}z_{coh}\|_{L_1} + \theta_{coh}\|z_{coh}\|_{L_1} + \lambda_{per}\|D_{per}z_{per}\|_{L_1} + \theta_{per}\|z_{per}\|_{L_1}$$

  is replaced with the equivalent expression $K(y, z_{ext}) = \|y_{ext} - Rz_{ext}\|_{L_1}$.

Where necessary, we can use weights to control for heteroscedasticity. Thus, equation (4) becomes

$$K(y, z_{sm}, z_{coh}, z_{per}) = \|w(y - (z_{sm} + z_{coh} + z_{per}))\|_{L_1} + \lambda_{xx}\|D_{xx}z_{sm}\|_{L_1} + \lambda_{xt}\|D_{xt}z_{sm}\|_{L_1} + \lambda_{tt}\|D_{tt}z_{sm}\|_{L_1} + \lambda_{coh}\|D_{coh}z_{coh}\|_{L_1} + \theta_{coh}\|z_{coh}\|_{L_1} + \lambda_{per}\|D_{per}z_{per}\|_{L_1} + \theta_{per}\|z_{per}\|_{L_1}, \quad (5)$$

where weights w_i are taken as the inverse of the estimated standard deviations of y_i. However, we show in Section 4 that this has little effect on the results.

3. Application to French female mortality

The RESPECT method is applied to French female mortality rates, and the resulting smooth surface (minus the cohort and period effects) is shown in Figure 5, the cohort effects are seen in Figure 6, and the residuals are plotted in Figure 7. No period effects were identified in this data set.

The strongest effects, in order, are for cohorts born in 1920, 1916, 1919, 1915 and 1926. Caselli et al. (1987) found evidence that Italian males and females born during the war years 1914 to 1918 had higher mortality than adjacent cohorts for at least 30 years. They did not, however, observe a similar effect in France. Our findings are consistent with the lack of an overall cohort effect, because the 1915 and 1920 cohorts show higher mortality and the 1916 and 1919 cohorts lower mortality, with only a small net increase for the combined cohort. They are also consistent with the findings of Richards (2008) of lower mortality for the 1919 cohort in England and Wales. Richards (2008) traces this pattern, in part, to a surge of births in late 1919 and early 1920 that resulted in the 1919 cohort being young on average and the 1920 cohort old on average; similar surges were observed in France and other European countries

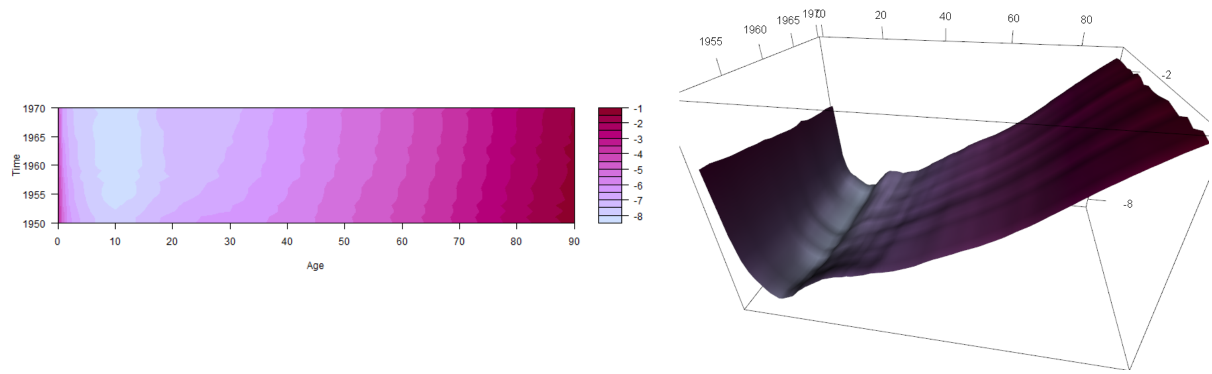

Figure 5. French female log mortality rates smoothed with the RESPECT method (with cohort effects removed).

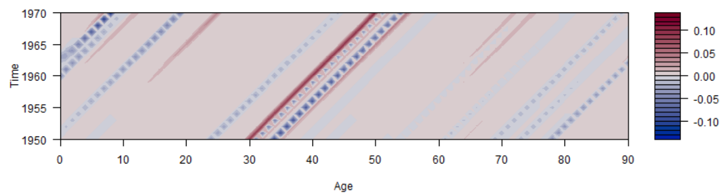

Figure 6. Cohort effects of French female log mortality rates.

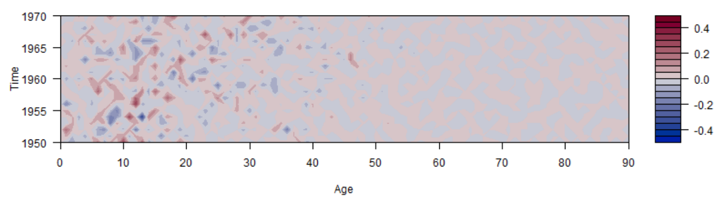

Figure 7. Residuals of French female log mortality rates after smoothing with the RESPECT method.

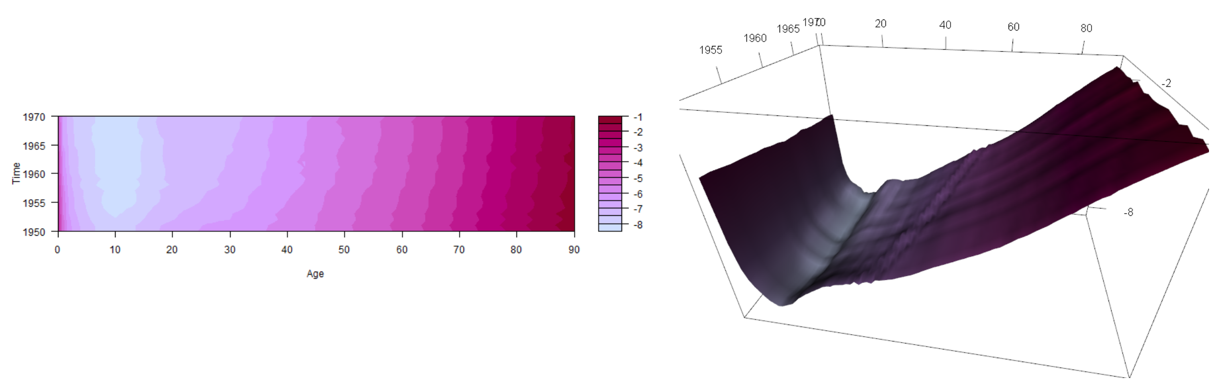

Figure 8. French female mortality rates smoothed with the RESPECT method (including cohort effects).

(Vandenbroucke, 2014). Higher perinatal mortality for the 1919 cohort due to the Spanish influenza epidemic (in a study by Harris 1919, 26% of pregnant women who survived the epidemic lost the child) potentially also resulted in lower ongoing mortality among the select, hardy survivors. Finally, data artefacts due to difficulties estimating population size because of wild swings in the pattern of births may have contributed to observed cohort patterns (Richards, 2008).

Figure 8 depicts the complete surface with the cohort and period effects added to the smooth surface. It is the "signal" which we have separated from the "noise" (represented by the residuals).

4. Comparison

We compare our proposed RESPECT method against three competing approaches.

1. HU: The Hyndman & Ullah (2007) method is implemented in the demography R package (Hyndman, 2014), and smooths mortality rates only in the age dimension. It uses weighted penalized spline regression, constrained to be monotonic for older ages, and applied independently for each year.
2. CDE: The Currie et al. (2004) method smooths mortality rates in both dimensions using two-dimensional penalized B-splines. It is implemented in the MortalitySmooth package for R (Camarda, 2012).
3. RMS: The L_1 regularized median smoothing discussed in Section 2.2.

We also compare a weighted version of our RESPECT method to allow for heteroskedasticity, using (5).

We use a 20-fold cross validation procedure for comparing the different smoothing methods, where each fold data randomly omits 5% of the original observations. Four subsets of the French female mortality data are used in our comparisons.

1. Data for years 1950–1970 and ages 10–60 represent a relatively smooth surface. This comparison is useful to ensure that the most "responsive" algorithms using the L_1 norm do not perform any worse than the more "stable" algorithms based on the L_2 norm. This data set is also important because it is the only comparison satisfying the requirements for Method 2 (Camarda, 2012) which is designed to work for ages starting from 10 years and where there are no outliers.
2. Data for years 1950–1970 and ages 0–60 represent a period when no outliers happened — there were no global wars or large pandemics. The younger ages from 0 to 9 have abrupt changes in mortality rates which are more challenging to smooth.
3. Data for years 1935–1955 and ages 10–60 represent a period including outliers due to World War II. It is important to mention that such an outlier should be considered as an outlier only along the time dimension and not over age. Therefore in our case, when only uncorrelated errors are considered as noise, such a one-dimensional outlier should be preserved during smoothing as signal. These data are important for testing the smoothing abilities of the algorithms in the presence of one-dimensional outliers.
4. Data for years 1935–1955 and ages 0–60 represents the most complex dataset containing the one-dimensional outliers (WWII) and also a period of abrupt mortality changes for ages 0 to 9.

To ensure an automatic and fair selection of the smoothing parameters, we simply used all methods with default arguments as set by the R package authors.

Table 1 shows Mean Square Error (MSE) and Mean Absolute Error (MAE) multiplied by 100 for the five methods. In all cases the error in a particular age and time cell is the difference between the observed log mortality rate and the modelled / smoothed log mortality rate incorporating the smooth bivariate surface as well as any cohort and period

Years	Ages	MSE HU	MSE CDE	MSE RMS	MSE RESPECT	MSE wRESPECT	MAE HU	MAE CDE	MAE RMS	MAE RESPECT	MAE wRESPECT
1950–1970	10–60	1.15	0.72	0.57	0.46	0.51	6.39	5.86	5.12	4.64	4.73
1950–1970	0–60	0.78		0.43	0.41	0.40	5.61		4.81	4.73	4.60
1935–1955	10–60	1.05		0.45	0.39	0.38	4.91		4.59	4.11	4.04
1935–1955	0–60	0.85		0.40	0.39	0.38	5.09		4.38	4.13	4.15

Table 1. Cross validation performance of different smoothing methods against French female mortality data (HU is MSE for cross validation of Method 1 mutiplied by 100, HU is MAE for cross validation of Method 1 mutiplied by 100, . . . , WDHT is MAE for cross validation of Method 6 mutiplied by 100)

Years	Ages	MSE HU	MSE CDE	MSE RMS	MSE RESPECT	MSE wRESPECT	MAE HU	MAE CDE	MAE RMS	MAE RESPECT	MAE wRESPECT
1950–1970	10–60	2.83	1.51	1.36	1.31	1.35	9.07	7.97	7.44	7.25	7.20
1950–1970	0–60	3.57		1.37	1.35	1.38	9.65		7.62	7.49	7.53
1935–1955	10–60	2.01		0.94	0.88	0.88	7.82		6.46	6.11	6.09
1935–1955	0–60	2.45		1.05	0.98	0.98	8.48		6.88	6.57	6.62

Table 2. Average cross validation results across all 24 populations.

effects. Results indicate that the RESPECT method (with and without taking heteroscedasticity into account) shows better or similar performance compared to the other methods.

We carried out a similar comparison for 12 countries and both sexes. All data were taken from Human Mortality Database (2008). The countries selected are all those for which the Human Mortality Database has data available for the period 1935–1970. Table 2 shows the average MSE and MAE across all 24 populations for each of the four period and age subsets. The results clearly indicate that the previous conclusions based on French females also apply for this more comprehensive dataset.

5. Conclusion

In this paper we have proposed a new method to smooth mortality data in two dimensions with cohort and period effects. Comparisons with other methods show that it provides superior estimates of the underlying mortality surface, in addition to providing insights into the existence of cohort and period effects that might otherwise be overlooked.

Additional improvements are possible. While only partially adaptive splines were used for the L_1 regularized median smoothing methods, a fully adaptive approach may provide further improvements and requires further investigation.

The λ coefficients used in the regularized median smoothing methods were estimated using computationally lengthy numerical methods. A simpler procedure, similar to that used by Camarda (2012), would improve their practical usefulness. We leave this to a later paper.

Another potential improvement is to model several countries simultaneously, taking account of the existence of similar features in related countries. However, this would substantially increase the complexity of the optimization routines, and it is not clear that there is sufficient information available in the cross-country correlations that is not already exploited by smoothing within countries.

The RESPECT method is publicly available in the R package "smoothAPC" (Dokumentov & Hyndman, 2017) on CRAN. Code for this article can be accessed from the on-line repository at http://bit.ly/respectsmooth.

References

Andreev, K & Vaupel, J (2005), 'Patterns of mortality improvement over age and time: Estimation, presentation and implications,' in *Population Association of America 2005 Annual Meeting*, vol. 31.

Barbi, E & Camarda, CG (2011), 'Period and cohort effects on elderly mortality: a new relational model for smoothing mortality surfaces,' *Statistica*, **71**(1), pp. 51–69.

Brillinger, DR (1986), 'The natural variability of vital rates and associated statistics,' *Biometrics*, **42**, pp. 693–734.

Brunel, NJB, Clairon, Q & d'Alché Buc, F (2014), 'Parametric estimation of ordinary differential equations with orthogonality conditions,' *Journal of the American Statistical Association*, **109**(505), pp. 173–185.

Camarda, CG (2012), 'MortalitySmooth: An R package for smoothing Poisson counts with P-splines,' *Journal of Statistical Software*, **50**(1), pp. 1–24.

Camarda, CG, Eilers, PHC & Gampe, J (2010), 'Sums of smooth exponentials,' in *Proceedings of the 25th International Workshop on Statistical Modelling*, pp. 113–118.

Camarda, CG, Eilers, PHC & Gampe, J (2012), 'Additive decomposition of vital rates from grouped data,' in *Proceedings of the 27th International Workshop on Statistical Modelling*, pp. 57–62.

Carriere, JF (1992), 'Parametric models for life tables,' *Transactions of the Society of Actuaries*, **44**, pp. 77–99.

Caselli, G, Vallin, J, Vaupel, JW & Yashin, A (1987), 'Age-specific mortality trends in France and Italy since 1900: Period and cohort effects,' *European Journal of Population / Revue européenne de Démographie*, **3**(1), pp. 33–60.

Chen, K & Müller, HG (2012), 'Modeling repeated functional observations,' *Journal of the American Statistical Association*, **107**(500), pp. 1599–1609.

Currie, ID, Durban, M & Eilers, PH (2004), 'Smoothing and forecasting mortality rates,' *Statistical Modelling*, **4**(4), pp. 279–298.

Dokumentov, A & Hyndman, RJ (2017), *Smoothing of Two-Dimensional Demographic Data, Optionally Taking into Account Period and Cohort Effects*, R package version 0.2. http://cran.r-project.org/package=smoothAPC.

Ferraty, F & Vieu, P (2006), *Nonparametric functional data analysis: theory and practice*, Springer.

Forfar, DO, McCutcheon, JJ & Wilkie, AD (1988), 'On graduation by mathematical formula,' *Journal of the Institute of Actuaries*, **115**(1), pp. 1–149.

Gertheiss, J, Maity, A & Staicu, AM (2013), 'Variable selection in generalized functional linear models,' *Stat*, **2**(1), pp. 86–101.

Hannerz, H (2001), 'Presentation and derivation of a five-parameter survival function intended to model mortality in modern female populations,' *Scandinavian Actuarial Journal*, **2001**(2), pp. 176–187.

Harris, J (1919), 'Influenza occurring in pregnant women: A statistical study of thirteen hundred and fifty cases,' *Journal of the American Medical Association*, **72**(14), pp. 978–980.

He, X, Ng, P & Portnoy, S (1998), 'Bivariate quantile smoothing splines,' *Journal of the Royal Statistical Society: Series B*, **60**(3), pp. 537–550.

Heligman, L & Pollard, JH (1980), 'The age pattern of mortality,' *Journal of the Institute of Actuaries (1886-1994)*, **107**(1), pp. 49–80.

Henderson, R (1924), 'A new method of graduation,' *Transactions of the Actuarial Society of America*, **25**, pp. 29–40.

Horváth, L & Kokoszka, P (2012), *Inference for functional data with applications*, Springer.

Huang, JZ, Shen, H & Buja, A (2009), 'The analysis of two-way functional data using two-way regularized singular value decompositions,' *Journal of the American Statistical Association*, **104**(488), pp. 1609–1620.

Human Mortality Database (2008), University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), data downloaded on 20 Feb 2008.

Hyndman, RJ (2014), *demography: Forecasting mortality, fertility, migration and population data*, with contributions from Heather Booth, Leonie Tickle and John Maindonald. R package version 1.16. http://cran.r-project.org/package=demography.

Hyndman, RJ & Ullah, SM (2007), 'Robust forecasting of mortality and fertility rates: A functional data approach,' *Computational Statistics & Data Analysis*, **51**(10), pp. 4942–4956.

Janssen, F & Kunst, AE (2005), 'Cohort patterns in mortality trends among the elderly in seven European countries, 1950-99.' *International Journal Of Epidemiology*, **34**(5), pp. 1149 – 1159.

Johnson, NP & Mueller, J (2002), 'Updating the accounts: global mortality of the 1918-1920 "spanish" influenza pandemic,' *Bulletin of the History of Medicine*, **76**(1), pp. 105–115.

Kirkby, J & Currie, I (2010), 'Smooth models of mortality with period shocks,' *Statistical Modelling*, **10**(2), pp. 177–196.

Koenker, R (2015), *quantreg: Quantile Regression*, R package version 5.11. http://cran.r-project.org/package=quantreg.

Koenker, R, Ng, P & Portnoy, S (1994), 'Quantile smoothing splines,' *Biometrika*, **81**(4), pp. 673–680.

Marron, J, Ramsay, JO, Sangalli, LM & Srivastava, A (2015), 'Functional data analysis of amplitude and phase variation,' MOX Report 27/2015, Department of Mathematics, Politecnico di Milano, Italy.

O'Connell, A & Dunstan, K (2009), 'Do cohort mortality trends emigrate? insights on the U.K.'s golden cohort from a comparison with a British settler country,' *British Actuarial Journal*, **15**(S1), pp. 91–121.

Portnoy, E (1997), 'Regression-quantile graduation of Australian life tables, 1946–1992,' *Insurance: Mathematics and Economics*, **21**(2), pp. 163–172.

R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Richards, SJ (2008), 'Detecting year-of-birth mortality patterns with limited data,' *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, **171**(1), pp. 279–298.

Richards, SJ, Kirkby, JG & Currie, ID (2006), 'The importance of year of birth in two-dimensional mortality data,' *British Actuarial Journal*, **12**(1), pp. 5–38.

Sangalli, LM (2014), 'Statistical and numerical techniques for spatial functional data analysis,' in Bongiorno, EG, Salinelli, E, Goia, A & Vieu, P (eds.), *Contributions in infinite-dimensional statistics and related topics*, Società Editrice Esculapio.

Schuette, DR (1978), 'A linear programming approach to graduation,' *Transactions of Society of Actuaries*, **30**, pp. 407–431.

Shen, D, Shen, H, Bhamidi, S, Muñoz Maldonado, Y, Kim, Y & Marron, J (2014), 'Functional data analysis of tree data objects,' *Journal of Computational and Graphical Statistics*, **23**(2), pp. 418–438.

Silverman, B & Ramsay, J (2005), *Functional Data Analysis*, Springer.

Vandenbroucke, G (2014), 'Fertility and wars: The case of World War I in France,' *American Economic Journal: Macroeconomics*, **6**(2), pp. 108–136.

Whittaker, ET (1922), 'On a new method of graduation,' *Proceedings of the Edinburgh Mathematical Society*, **41**, pp. 63–75.

Willets, R (2004), 'The cohort effect: insights and explanations,' *British Actuarial Journal*, **10**(04), pp. 833–877.

Wood, SN (2006), *Generalized Additive Models: an introduction with R*, CRC Press.

Zhu, H, Fan, J & Kong, L (2014), 'Spatially varying coefficient model for neuroimaging data with jump discontinuities,' *Journal of the American Statistical Association*, **109**(507), pp. 1084–1098.

Zhu, H, Strawn, N & Dunson, DB (2016), 'Bayesian graphical models for multivariate functional data,' *Journal of Machine Learning Research*, **17**, pp. 1–27.