Article
Graphical Diagnostic for Mortality Data Modeling

M.L. Gamiz *,**†**, **M.D. Martinez-Miranda †** and **R. Raya-Miranda †**

   Department of Statistics and O.R., University of Granada, 18010 Granada, Spain;
mmiranda@ugr.es (M.D.M.-M.); rraya@ugr.es (R.R.-M.)

*   Correspondence: mgamiz@ugr.es

†   These authors contributed equally to this work.

Abstract: The main contribution of this paper is to develop a graphical tool to evaluate the suitability of a candidate parametric model to fit a data set. The practical motivation is to check the adequacy of the so called SAINT model proposed in Jarner and Kryger (2011). This model has been implemented in practice by an important European pension fund concerned with setting annuity reserves for all current or former employees of Denmark. So, one particular relevant question is whether this mortality model is actually fitting old-age. Our graphical test can be described as follows. First we transform the data by means of the parametric model which is being evaluated. Should the model be correct, the transformed data will be in accordance with an Exponential distribution with rate 1. Then we construct a family of local linear hazard estimates based on the data on the transformed scale. Finally we use the statistical tool SiZer to assess the goodness-of-fit of the Exponential distribution to the data on the transformed scale. If the model is correct the SiZer map should not reveal any significant feature in the family of kernel estimates. Our method allow us to establish a diagnostic regarding the validity of the SAINT model when describing mortality patterns in four different countries.

Keywords: SAINT model; SiZer; local linear fitting; mortality data

---

0. Introduction

The main contribution of this paper is to develop a graphical tool to evaluate the suitability of a candidate parametric model to fit a data set. The practical motivation is to check the adequacy of the so called SAINT model proposed in Jarner and Kryger (2011). This model has been implemented in practice by an important European pension fund concerned with setting annuity reserves for all current or former employees of Denmark. So, one particular relevant question is whether this mortality model is actually fitting old-age.

The SAINT model has been evaluated previously in Spreeuw, Nielsen and Jarner (2013), where a visual inspection technique is introduced in order to detect the credibility of the parametric assumptions of the proposed model. The technique consists of first transforming the original mortality data according to the proposed parametric hazard model, and then non parametrically estimate the density based on transformed data. The density of the transformed data should be close to the uniform distribution when the model assumptions are correct. To estimate the density on the transformed axis the local linear density estimator based on filtered data and proposed in Nielsen, Tanggaard and Jones (2009) is used.

In this paper we focus on hazards estimation instead of the densities. In Section 1 we describe the parametric frailty hazard model (SAINT model) to be tested and describe the estimation procedure of the parameters involved. In Section 2 we present the local linear fitting of the hazard and its first derivative in the context of aggregated survival data. Section 3 is devoted to the description of the exploratory tool SiZer. In Section 4 we carry out a study with real mortality data where we evaluate the goodness-of-fit of the SAINT frailty model in different scenarios determined by mortality data concerning four different countries and two calendar years. Conclusions are given in Section 5.

1. Mixed hazard model for mortality data

1.1. Model specifications

The parametric model for mortality data suggested in Jarner and Kryger (2011), SAINT model is a multiplicative frailty model where the mortality rate of a individual is decomposed in two factors: one is a standard intensity function and the other is a non-negative random variable, the frailty. In Spreeuw *et al.* (2013) this model is evaluated. The authors discuss the adequacy of several candidates for the frailty variable, a Gamma model, an Inverse Gaussian model, and degenerate (i.e. no frailty) to properly adjust hazard rates based on mortality data. In their study the authors introduce a visual test to assess the validity of the mixed model to explain mortality rates based on data concerning female population of ages ranging from 40 onwards and coming from four different countries, i.e., United States, United Kingdom, Denmark and Iceland, during the year 2006. In their conclusions, the authors argue that the Gamma frailty model provides a more accurate fit to the data in all real cases considered, specially at the oldest old-ages for which the Gamma frailty model better accommodate to the presence of an old-age mortality plateau. The parametric mixed mortality model considered generalizes the classical Gompertz survival model by including more parameters and by including a multiplicative frailty component.

Specifically, the frailty model presented in Jarner and Kryger (2011) and Spreeuw *et al.*(2013) introduces the individual effect for life's mortality multiplicatively. So, it is assumed that the individual effect for a particular subject is described by a random variable X, in the sense that the conditional mortality hazard at age t, given $X = x$, is expressed as

$$\alpha(t, x) = x\alpha_0(t), \tag{1}$$

where $\alpha_0(t)$ represents the baseline mortality hazard at age t, that is, the hazard corresponding to an individual with frailty level $x = 1$. Also, it is assumed that, given a cohort of n individuals, $X_1, X_2, \ldots, X_n$ are independent and identically distributed with mean 1.

We adopt here the specifications given in Jarner and Kryger (2011) for national and international mortality modeling, later analyzed in Spreeuw *et al.* (2013), which consist of defining the baseline mortality hazard as

$$\alpha_0(t) = \exp\left(a_0 + a_1 t + a_2 t^2\right). \tag{2}$$

Also, following the recommendation of Spreeuw *et al.* (2003) we consider for the frailty X a gamma distribution with mean 1 and variance σ^2.

From Equation (1), the relationship between the corresponding survival functions can be written as $S_x(t) = e^{-\Lambda_x(t)} = e^{-x\Lambda_0(t)}$, where $\Lambda_x(t) = \int_0^t \alpha_x(s)ds$ is the cumulative hazard mortality for an individual for which $X = x$ and $\Lambda_0(t) = \int_0^t \alpha_0(s)ds$ is the baseline cumulative hazard. Then the cohort survival function is given by

$$S(t) = \int S_x(t)f(x)dx = \int e^{-x\Lambda_0(t)}f(x)dx,$$

or, equivalently,

$$S(t) = L_X\left(\Lambda_0(t)\right),$$

where, $L_X(s) = E\left[e^{-sX}\right]$ denotes the Laplace transform of the distribution of the frailty random variable, X. From here we can express the frailty model as

$$\alpha(t) = -\left(\frac{d\ln L_X(s)}{ds}\bigg|_{\Lambda_0(t)}\right)\alpha_0(t) \tag{3}$$

Considering (2) and that X has a gamma distribution as stated above, $L_X(s) = E\left[e^{-sX}\right] = \left(1 + \sigma^2 s\right)^{-1/\sigma^2}$, we have that the SAINT model specifies the cohort mortality hazard as

$$\alpha(t) = \frac{\alpha_0(t)}{1 + \sigma^2 \Lambda_0(t)} = \frac{\exp\left(a_0 + a_1 t + a_2 t^2\right)}{1 + \sigma^2 \left(\int_0^t \exp\left(a_0 + a_1 s + a_2 s^2\right) ds\right)} \tag{4}$$

1.2. Maximum likelihood estimation of the parametric model

Consider a data set consisting of mortality statistics of n lives. More specifically, the information we register is related, on the one hand to the exposure process Y_i, which takes value 1 when individual i is alive and under observation, and, on the other hand, N_i, the counting process taking the value 1 when the individual i has died while under observation and 0, otherwise. We consider both Y_i and N_i depending on the age t, in such a way that we assume that N_i is a one-dimensional counting process with respect to an increasing right continuous complete filtration $\mathcal{F}_t$, $t > 0$. Also we assume that the corresponding intensity function of N_i satisfies the multiplicative model of Aalen (see for example Andersen *et al.*, 1993)

$$\lambda_i(t) = \alpha(t) Y_i(t),$$

where, as in Spreeuw *et al.* (2013), we consider that $\{\alpha_\theta, \theta \in \Theta\}$ is a parametric family of hazard functions to which the true hazard function $\alpha = \alpha_\theta$ belongs. In other words, the exact mortality is determined except for the value of the parameter vector θ which is estimated by maximum-likelihood methods, that is, $\widehat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta)$, with

$$\mathcal{L}(\theta) = \sum_{i=1}^n \int \log\left(\alpha_\theta(t)\right) dN_i(t) - \sum_{i=1}^n \int \alpha_\theta(t) Y_i(t) dt$$

1.3. Transformation of data

In this section we use the parametric fit explained in the previous section to transform the data in such a way that the underlying hazard would become constant in case the parametric fit indeed would have been the correct underlying model α. The parametric fit is in other words used as a kind of prior knowledge to simplify the non-parametric estimation problem. If the prior knowledge is of high quality and the parametric model has good approximation properties, then the resulting non-parametric problem is simplified in the second step compared to a fully non parametric treatment of the original dataset (not transformed). Next a local linear hazard estimator is applied to the transformed data. If the parametric specification α_θ were true then after the time transformation $x = \Lambda_\theta(t)$ with $\Lambda_\theta(t) = \int_0^t \alpha_\theta(s) ds$, the functional form of the hazard on the transformed scale would be simply equal to the standard Exponential, i.e. the hazard would be equal to 1 on the transformed scale.

For a given $\theta \in \Theta$ we put $\widetilde{N}_i^\theta = N_i \circ \Lambda_\theta^{-1}$ and $\widetilde{Y}_i^\theta = Y_i \circ \Lambda_\theta^{-1}$. This transformed process follows Aalen's multiplicative hazard model with transformed stochastic hazard

$$\lambda_i^\theta(x) = g_\theta(x) \times Y_i\left(\Lambda_\theta^{-1}(x)\right)$$

where $g_\theta(x) = \alpha\left(\Lambda_\theta^{-1}(x)\right) / \alpha_\theta\left(\Lambda_\theta^{-1}(x)\right)$, α is the true hazard and α_θ is the assumed parametric value. We now carry out our non-parametric local linear smoothing technique on the transformed processes $\widetilde{N}_i^\theta$ and $\widetilde{Y}_i^\theta$ and obtain an estimate $\widehat{g}_\theta$ of g_θ on the transformed time axis. As in Spreeuw et al. (2013) we use a parametric mixed hazard model based on a gamma frailty mortality model and we estimate the parameters by a standard maximum likelihood procedure. Then we transform the survival data as just described. If the parametric model is correct, then the estimated hazard $\widehat{\alpha}_\theta$ should be very close to the true curve, and in consequence the transformed time scale data would

produce a sample from an Exponential distribution with rate 1. In other words the estimated curve $\widehat{g}_\theta$ should be very close to the line $y = 1$.

2. Local linear estimation with aggregated data

In this section we present a local linear estimation method for aggregated survival data which is based on a least squares minimization problem. We follow the general formulation proposed in Nielsen and Tanggaard (2001) and we obtain an estimator for the hazard function as well as for its first derivative. Also we derive an expression for the variance of the estimator of the derivative of the hazard function needed to develop the SiZer tool. The details are given in the following. We observe n individuals $i = 1, \ldots, n$. Let N_i count observed failures for the ith individual in the time interval $[0, T]$. N_i can take values 0 or 1. We assume that N_i is a one-dimensional counting process with respect of an increasing, right continuous, complete filtration $\mathcal{F}_t$, $t \in [0, T]$, i.e. it obeys *les conditions habituelles*. We assume Aalen's multiplicative model where the random intensity is written as

$$\lambda_i(t) = \alpha(t)Y_i(t), \tag{5}$$

with no restriction on the functional form of the hazard function $\alpha(\cdot)$. Y_i is a predictable process taking values in $\{0, 1\}$, indicating (by the value 1) when the ith individual is at risk. We assume that $(N_1, Y_1), \ldots, (N_n, Y_n)$ are i.i.d. for the n individuals. Note that this formulation contains the important particular case of a longitudinal study with left truncation and right censoring. In that case we observe triplets (L_i, Z_i, δ_i) $(i = 1, .., n)$ where L_i is the time an individual enters the study, Z_i is the time he/she leaves the study and δ_i is binary and equal to 1 if death is the reason for leaving the study (censoring indicator). Then, the process Y_i above would be $Y_i(t) = I[L_i \leq t < Z_i]$ and $N_i(t) = I[Z_i \leq t]\delta_i$, where $I(\cdot)$ is the indicator function. Hereafter we will work in the general model. Expressions for particular sampling schemes such as left truncation and right censoring can be derived just by substituting the particular expressions of the processes.

The data that we will use are divided into discrete numbers of occurrences and exposures. We suppose that the following aggregated values of occurrences and exposures are available

$$O_r = \sum_{i=1}^{n} \int_{x_{r-1}}^{x_r} dN_i(x), \quad \text{and} \quad E_r = \sum_{i=1}^{n} \int_{x_{r-1}}^{x_r} Y_i(x)dx,$$

for $r = 1, \ldots, m$. Here, $x_1, \ldots, x_m$ are some time points. For simplicity in the formulation of the problem we consider this points equidistant however we can easily generalize for the non equidistant case. With $\Delta_r = x_r - x_{r-1}$ we can define $Y_r = E_r / \Delta_r$. This is the average number of individuals which are at risk in the interval $[x_{r-1}, x_r)$ for $r = 1, \ldots, m$ and with $x_0 = 0$.

In this situation, we define an empirical (naive) estimator of α_r as

$$\widehat{\alpha}_r = \frac{O_r}{E_r}.$$

From theorem A.1 (i) and (ii) in Wang, Muller and Capra (1998) it can be deduced (see also Bagkavos, 2011) that

$$E(\widehat{\alpha}_r) \approx \alpha(x_r),$$

and

$$Var(\widehat{\alpha}_r) \approx \frac{\alpha(x_r)}{S(x_r)} \tag{6}$$

Moreover, the pairs $(x_r, \widehat{\alpha}_r)$, $r = 1, 2, \ldots, M$ can be considered approximately independent.

Under a local linear approach we assume that, for each x_r, the hazard function α can be locally approximated (in a neighborhood of x_r) by a linear function, this is, for all $x \in (x_r - h, x_r + h)$, $\alpha(x) = \theta_0 + \theta_1 (x - x_r)$, for an appropriate bandwidth h. Thus, at the grid point x_r $(r = 1, \ldots, M)$,

an estimation of the intercept θ_0 would provide an estimation of $\alpha\left(x_r\right)$, whereas an estimation of the slope θ_1 would provide an estimation of the derivative $\alpha'\left(x_r\right)$.

To this goal, we formulate the following least squares problem

$$\left(\begin{array}{c}\widehat{\theta}_0 \\ \widehat{\theta}_1\end{array}\right) = \underset{\left(\begin{array}{c}\theta_0 \\ \theta_1\end{array}\right)}{\arg\min} \sum_{m=0}^{M} \left(\widehat{\alpha}_m - \theta_0 - \theta_1\left(x_m - x_r\right)\right)^2 K_h\left(x_m - x_r\right),$$

where $\widehat{\alpha}_m = \frac{O_m}{E_m}$.

The solution to the equations above is easy to obtain and can be expressed as

- The hazard function, α

$$\widehat{\alpha}_{LL}\left(x\right) = \frac{S_2\left(x\right)T_0\left(x\right) - S_1\left(x\right)T_1\left(x\right)}{S_2\left(x\right)S_0\left(x\right) - \left(S_1\left(x\right)\right)^2} \tag{7}$$

- The first derivative of the hazard function, α'

$$\widehat{\alpha'}_{LL}\left(x\right) = \frac{S_0\left(x\right)T_1\left(x\right) - S_1\left(x\right)T_0\left(x\right)}{S_2\left(x\right)S_0\left(x\right) - \left(S_1\left(x\right)\right)^2} \tag{8}$$

where, as usual, we define the moments S_j, for $j = 0, 1, 2$,

$$S_j(x) = \sum_{m=0}^{M} \left(x_m - x\right)^j K_h\left(x_m - x\right),$$

and, for $j = 0, 1$, we denote

$$T_j(x) = \sum_{m=0}^{M} \left(x_m - x\right)^j K_h\left(x_m - x\right)\widehat{\alpha}_m,$$

being $\widehat{\alpha}_m = \frac{O_m}{E_m}$, for $m = 1, \ldots, M$, and $K_h\left(\cdot\right) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$, with K a symmetric kernel function.

Using matrix notation, we can write the estimators as follows

$$\widehat{\alpha}_{LL}(x) = \mathbf{e}_1^t\left(\mathbf{X}^t\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{W}\widehat{\alpha},$$

and

$$\widehat{\alpha'}_{LL}(x) = -\mathbf{e}_2^t\left(\mathbf{X}^t\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{W}\widehat{\alpha},$$

where

$$\mathbf{X} = \left(\begin{array}{cc}1 & x_1 - x \\ \vdots & \vdots \\ 1 & x_M - x\end{array}\right), \quad \mathbf{W} = diag\left(K\left(\frac{x_1 - x}{h}\right), \ldots, K\left(\frac{x_M - x}{h}\right)\right),$$

and

$$\mathbf{e}_1 = \left(\begin{array}{c}1 \\ 0\end{array}\right); \qquad \mathbf{e}_2 = \left(\begin{array}{c}0 \\ 1\end{array}\right)$$

Denoting $A_x = -\mathbf{e}_2^t\left(\mathbf{X}^t\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{W}$, then the variance of the estimator of $\alpha'(x)$ can be expressed as

$$V\left(\widehat{\alpha'}_{LL}\right)(x) = A_x Var\left(\widehat{\alpha}\right) A_x^t, \tag{9}$$

where, using independence, we have that $\mathbf{V} = Var\left(\widehat{\alpha}\right) = diag\left(V_1, \ldots, V_M\right)$, with $V_r = Var\left(\widehat{\alpha}_r\right)$, for $r = 1, 2, \ldots, M$.

Expanding the expression in (9) we get

$$V\left(\widehat{\alpha'}_{LL}\right)(x) = \mathbf{e}_2^t \left(\mathbf{X}^t\mathbf{W}\mathbf{X}\right)^{-1} \mathbf{R} \left(\mathbf{X}^t\mathbf{W}\mathbf{X}\right)^{-1} \mathbf{e}_2,$$

where $\mathbf{R} = \mathbf{X}^t\mathbf{W}\mathbf{V}\mathbf{W}\mathbf{X}$.

On the one hand, it can be checked that

$$\left(\mathbf{X}^t\mathbf{W}\mathbf{X}\right)^{-1} = \frac{1}{S_2(x)S_0(x) - (S_1(x))^2} \begin{pmatrix} S_2(x) & -S_1(x) \\ -S_1(x) & S_0(x) \end{pmatrix},$$

and, on the other hand, we have that $\mathbf{R}$ is a symmetric matrix with entries

$$\mathbf{R} = \begin{pmatrix} R_0(x) & R_1(x) \\ R_1(x) & R_2(x) \end{pmatrix}$$

$$R_j(x) = \sum_{m=1}^{M} (x_m - x)^j \left(K\left(\frac{x_m - x}{h}\right)\right)^2 V_m, \quad j = 0, 1, 2.$$

After easy algebra we obtain that

$$V\left(\widehat{\alpha'}_{LL}\right)(x) = \frac{S_1^2(x)R_0(x) - 2S_1(x)S_0(x)R_1(x) + S_0^2(x)R_2(x)}{\left(S_2(x)S_0(x) - (S_1(x))^2\right)^2}$$

Finally, we need to estimate the variance of $\widehat{\alpha}_r$, i.e. V_r, for $r = 1, \ldots, M$. To do it we use an appropriate estimator of the survival function and plug in the expression (6). So we propose

$$\widehat{V}_m = \frac{\widehat{\alpha}\left(x_m\right)}{\widehat{S}\left(x_m\right)} = \widehat{\alpha}_m \exp\left(\sum_{r=1}^{m} \widehat{\alpha}_r \Delta_r\right).$$

3. SiZer analysis for mortality data hazards

3.1. SiZer description

SiZer is a graphical device for the display of significant features in data with respect to a localization space and a scale space. It was proposed by Chaudhuri and Marron (1999) in the context of univariate density and regression. The authors described SiZer as a powerful exploratory tool to allow the visual assessment of significance of features in data. In real data analysis to discover new features in the data such as "bumps" or increasing/decreasing patterns might lead to important new scientific insights. In the context of goodness-of-fit and graphical residuals analysis, features and monotonic or shaped patterns in residuals lead to the rejection of postulated models. The exploratory power of the SiZer analysis has been applied recently for goodness-of-fit purposes. Some examples are Marron and Zhang (2005) for testing univariate parametric regression models and monotonicity; González-Manteiga et al. (2008) for testing additive regression models. Park and Kang (2008) and Park et al. (2014) for the comparison of regression curves; Park et al. (2004, 2009) and Rondonotti et al. (2007) for significance of trends in times series; Park et al. (2011) for testing uniformity in homogeneous Poisson processes. All these works differ from standard nonparametric methods for solving tests, which make use of difficult choices of bandwidths and provide global information that does not allow adequate diagnostics. The key idea of the SiZer analysis for goodness-of-fit purposes is to visualize the inference problem locally and considering many different scales. Since useful information can be available at different levels of smoothing, the problem is analysed considering a

wide range of smoothing levels that correspond to resolution levels. Besides possible local deviations from the postulated model can be located at specific areas in the localization space.

In this paper we are interested in estimating the hazard function, more specifically the hazard for data transformed under a parametric null hypothesis. The idea is to check the adequacy of a postulated model by assessing significance of features in the hazard function of the transformed data. If the model does fit the data then the hazard for the data transformed using the model should be constant, with no monotonic trend neither features. Features such as bumps are characterised by the significant zero crossing of first derivative of the target curve. At a bump there is a zero crossing of the derivative, curve first goes up (positive derivative) and then goes down (negative derivative). Chaudhuri and Marron (1999) described the SiZer analysis from two plots, the "family plot" that is the family of smoothers for the target function indexed by the bandwidth parameter, and a coloured map which displays the scale space inference about the first derivative of the function as follows: for each bandwidth (scale) and value in the support (localization), a confidence interval for the first derivative is calculated and the results about the sign are displayed in a two-dimensional space (map) using a colour code. Considering n_h scales and n_{grid} localizations, each pixel in the $n_{grid} \times n_h$ map is colour-coded as blue if zero is greater than the upper confidence limit (significantly increasing), red if zero is less than the lower confidence limit (significantly decreasing), purple if zero is within the confidence limits (not significantly increasing or decreasing), and gray indicating regions where the data are too sparse to make statements about significance (the effective sample size defined below is less than 5). From the map the features in data are characterized by changes in colour so a change from blue to red displays a significant bump.

Figure 1 shows the SiZer analysis for simulated data from a parametric hazard model. The details about how this SiZer has been constructed are given in the next section, now we consider this just as an example to facilitate the understanding of readers not familiarised with this graphical tool. The family plot shows the local linear hazard estimates with different choices of the bandwidth parameter. This gives an idea of the shape of the underlying hazard that the considered smoothing method is able to capture. The bigger bandwidths produce oversmoothed curves showing the general trend in the curve while small bandwidths provide a very detailed description of the hazard with many wiggles. The answer about which features are really there and whether any of these wiggles actually correspond to a significant feature is answered by looking at the colour map in Figure 1. This map shows the result of the inference about the first derivative of the hazard. In the vertical axis different smoothing levels in logarithmic scale are represented so drawing a horizontal line at a given value in the range we can see the inference results a that particular value. Following the spirit of the scale space inference real features in data will appear at some of these scales so the analyst should look at all of them to conclude about the significance. On the other hand the horizontal axis shows the values where the hazard (in this case the first derivative) is evaluated, so placing a vertical line in the map we can see the results for the inference at the corresponding point. With this in mind and the colour code described above we can see that the underlying hard is significantly increasing at the beginning (blue colour up to about 0.1), then there is a zero crossing (confidence intervals contain zero between 0.1 and 0.3) and the hazard goes down (red colour up to about 0.5). The change of colours indicates a significant bump around 0.2. Now, if we continue moving the localization up to the right then we can see that a second bump seems to occurs around the value 0.8. Note that the blue colour before 0.8 shows significant increasing interrupted by a zero crossing of the derivative (purple) and ending with significant decrease (red). The data used to construct these plots have been simulated from a distribution with hazard rate obtained as a mixture of Beta functions, that is $\alpha(t) = B(t; 6, 2) + B(t; 2, 6)$, where $B(t; a, b)$ is the density function of a Beta distribution with parameters a, and b. Note that the inference with SiZer is able to detect the two features in the true hazard, and it has been possible considering many different bandwidth choices.
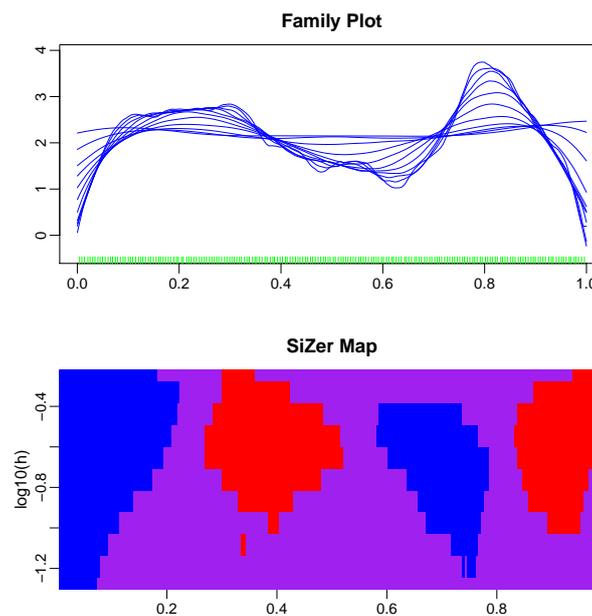
Figure 1. Illustration of SiZer with simulated data.

3.2. SiZer specifications

We now describe the main elements to construct SiZer for inference about a hazard function in the context of aggregated mortality data described in the previous sections. The first choice to define the SiZer analysis is about the family of smoothers to be considered. Our proposal is the local linear estimator of the hazard given in (7), which works from common aggregated mortality data. For the scale space inference about the first derivative we again propose the local linear estimator given in (8), and the estimate of its variance given in (9). With this estimator and its variance we can compute confidence intervals for the first derivative using the expression

$$\left(\widehat{\alpha'}_h(t) + q_{1-p/2}\sqrt{\widehat{Var(\widehat{\alpha'}_h(t))}}, \widehat{\alpha'}_h(t) + q_{p/2}\sqrt{\widehat{Var(\widehat{\alpha'}_h(t))}} \right). \tag{10}$$

Here p is the significance level, and $q_{1-p/2}, q_{p/2}$ are appropriate quantiles. To calculate the quantiles $q_{1-p/2}, q_{p/2}$ we suggest to use the a Normal approximation. Note that kernel estimates are local averages so the pointwise asymptotic normality of these estimators can be used to construct pointwise confidence intervals, just by taking $q_{1-p/2}$ and $q_{p/2}$ equal to the $(1-p/2)$ and $p/2$ quantiles of the standard Normal distribution, respectively. However SiZer Map is a simultaneous display of statistical inferences, so there is a multiple comparison issue which needs to be addressed. We consider simultaneous inference for SiZer by defining m independent confidence interval problems, where m reflects the number of independent blocks. Thus, the simultaneous Normal quantiles in SiZer are calculated as follows

$$q_{1-p/2} = -q_{p/2} = \Phi^{-1}\left(\frac{1 + (1-p)^{1/m}}{2} \right).$$

Here Φ^{-1} is the inverse of the standard Normal distribution function and m, which depends on the bandwidth h, $m \equiv m(h)$, is the number of independents blocks of average size available from a dataset of size n. The estimation of m is based on an estimate of the effective sample size (*ESS*), which

reflects the "number of data points in each kernel window". For the kernel estimators of the intensity the *ESS* can be computed as

$$ESS(t,h) = \frac{\sum_{i=1}^{n} K_h(t - T_i)O_i}{K_h(0)}.$$

And the number of independent blocks for each h is estimated as

$$m = \frac{n}{avg_{t\epsilon \mathcal{D}_h} ESS(t,h)}$$

where $avg_{t\epsilon \mathcal{D}_h}$ denotes the average value on the set $\mathcal{D}_h = \{t : ESS(t,h) \geq 5\}$.

3.3. A motivating example to illustrate the strength of SiZer

We now show with an example the versatility of SiZer when analysing mortality data. We consider the young male population from Finland and we are interested in evaluating how mortality rate corresponding to this section of the population has evolved along the years. Specifically we estimate mortality rate for male population aged 25 years old from calendar year 1906 to the year 2012. The results are shown in Figure 2. According to the plots we can deduce that the mortality rate exhibits a long-term decreasing trend. However two significant features are highlighted in the SiZer map. The colour change blue-purple-red in the first calendar years reveals the existence of a bump around the year 1918. Also another even more prominent bump is located around the years 1943-1945 (see again the colour change blue-purple-red around these years). These two bumps indicate an unusual increase of the mortality rate for the section of population considered here (male population 25 years old) compared to the decreasing evolutionary trend of the mortality along the whole period of time considered (calendar years 1906 to 2012). Notice that this local behavior of the mortality can be explained as a consequence of the two World Wars that took place during the periods 1914–1918 and 1939–1945, respectively.
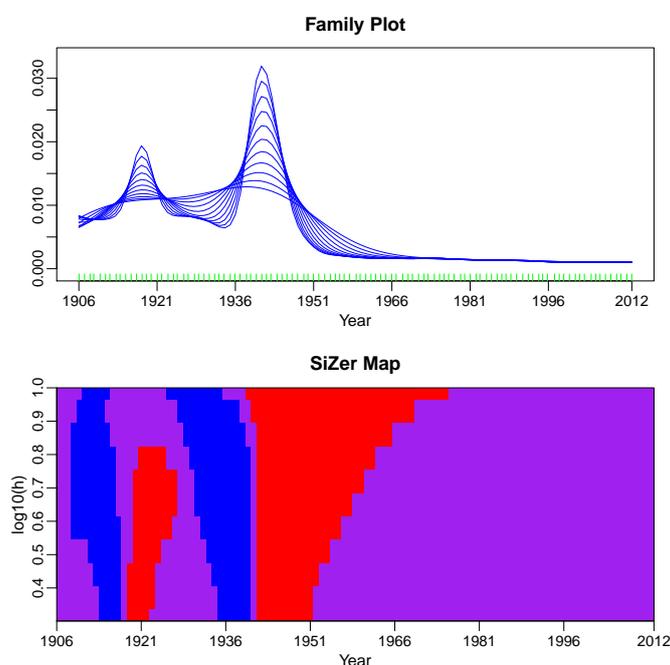

Figure 2. SiZer analysis from mortality data in Finland, for a cohort young male population during the observation period 1906–2012.

4. Graphical inspection of a candidate mortality model based on SiZer analysis

In this section we illustrate the use of SiZer analysis to check the adequacy of a candidate mortality model. We consider an application to real data based on the previous work of Spreeuw *et al.* (2013) and Gamiz *et al.* (2013). As these authors we analyse mortality data of four countries differing significantly in terms of population size, namely United States (US), United Kingdom (UK), Denmark and Iceland. For each country, the data set consists of female period mortality data, obtained from the Human Mortality Database, and concerning the calendar years 2006 and 2013, the last year from which we have complete datasets for the four countries available at Human Mortality Database. Spreeuw *et al.* (2013) used these data to check whether the SAINT model that is being used by the major Danish Pension Fund ATP for purposes of asset-liability management is a valid for representing old age mortality. Including the most recent available year 2013 we aim to check whether their conclusions from data in 2006 are still valid in 2013. In the analysis, only the ages from 40 to 110 are included and, as recommended in Spreeuw *et al.* (2013), we consider a Gamma frailty model as the candidate model to represent the data in the four countries. A visual inspection of a local linear estimator of the density based on transformed data for a single bandwidth choice they showed that the model represented fine the mortality patterns of adult at old ages. With the SiZer analysis introduced in this paper we are now able to assess the significance of these conclusions, avoiding the problem of the non-trivial bandwidth choice, see Gamiz *et al.* (2013) for a discussion of this issue.

The first step to check the adequacy of the model is to estimate the parameters by maximum likelihood. Using the methods described in Section 1.2 we have derived these estimates for the data in 2006 and 2013, the results are shown in Table 1. Then, from the estimated parametric hazard $\alpha_{\widehat{\theta}}$ and the corresponding cumulative hazard $\Lambda_{\widehat{\theta}}(t) = \int_0^t \alpha_{\widehat{\theta}}(s)ds$, the data are transformed as $x = \Lambda_{\widehat{\theta}}(t)$ (see Section 1.3. Finally the inference is performed about the underlying hazard of the transformed data. If the parametric model is correct then we should expect a hazard estimate close to the standard exponential model, i.e., constant and equal to 1. Significant features in the underlying hazard or significant monotone patterns would lead to reject the candidate model. To detect possible violations from the assumed model at each country and year we have constructed the SiZer plots with the transformed data. Figures 3 to 6 show the results for the year 2006 and the updates for the 2013 are shown in Figures 7 to 10. The total absence of red/blue colour in the maps confirm that there are not significant deviations from the Gamma frailty model for any of the countries in 2006. Also looking at the family plots we can see that the family of local linear hazards from transformed data are nearly constant around 1, as expected if the model is correct. This is the same conclusion that Spreeuw *et al.* (2013) drew in their paper. However the results from data in 2013 are slightly different. For the smaller countries Denmark and Iceland, the red-colour pattern in the maps for some of the considered bandwidths reveals a significant deviation from the constant line, which again would be expected if the parametric model is correct. A similar issue can be observed for US and UK since blue (significant increasing) and red colour (significant decreasing) arise in the maps from the 2013 data. Even these deviations are not very strong they suggest that the models are not describing the mortality rates in 2013 as well as they did in 2006.

Table 1. Maximum likelihood estimations of the parameters of the Gamma frailty hazard model for each country and calendar year. The results concerning the year 2006 have been obtained from Spreeuw *et al.* (2013).

Country (year)	a_0	a_1	a_2	σ
US (2006)	-8.1373	0.020140	0.0005402	0.1193
US (2013)	-7.5325	0.000114	0.0006905	0.3041
UK (2006)	-8.5070	0.01479	0.0006729	0.1632
UK (2013)	-8.1421	9.57e-10	0.0007879	0.3704
Denmark (2006)	-9.7088	0.06034	0.0003021	0.04007
Denmark (2013)	-10.6422	0.07661	0.0002253	0.02559
Iceland (2006)	-10.1136	0.05708	0.0003689	0.005556
Iceland (2013)	-10.4062	0.05928	0.0003919	0.187948
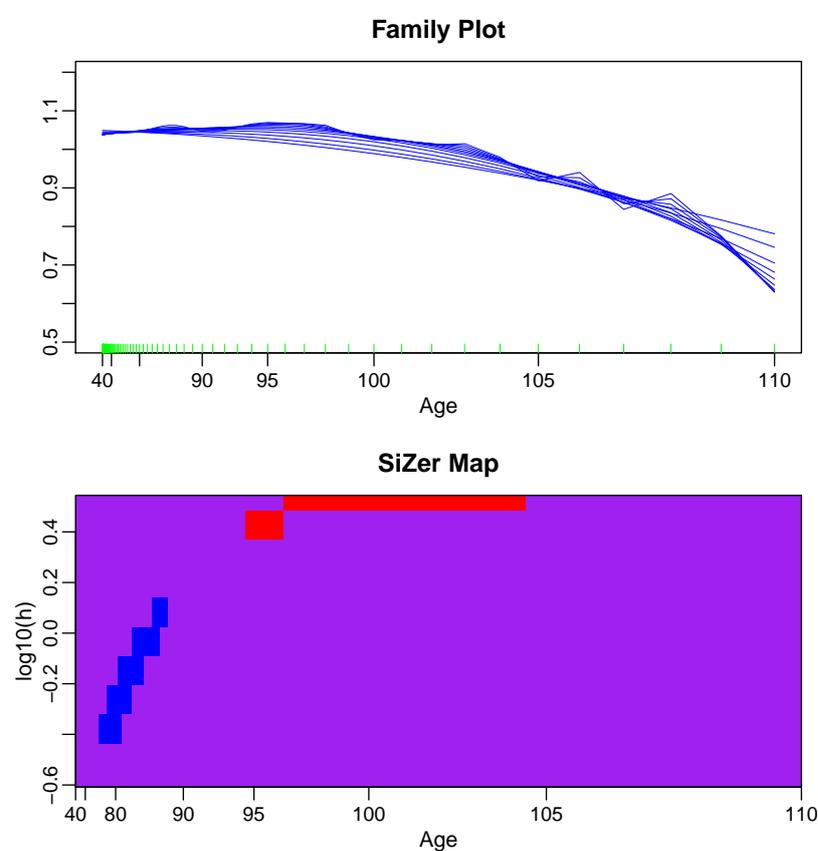
Figure 3. SiZer analysis from mortality data in US in 2006, transformed under a Gamma frailty hazard model.
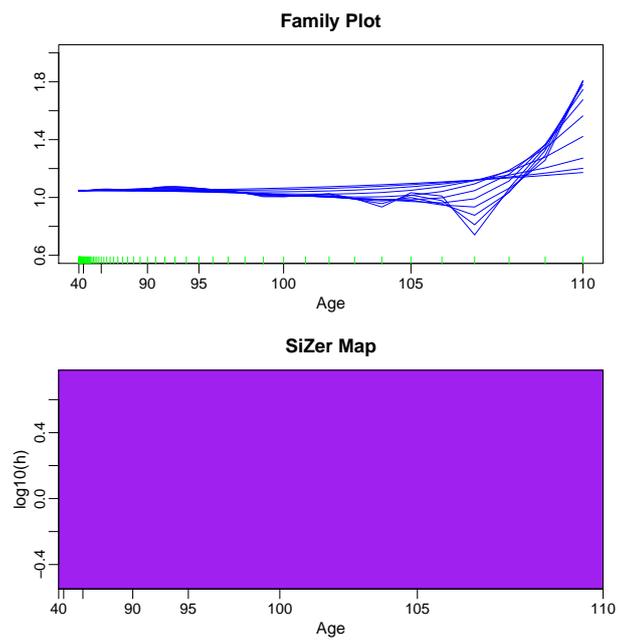
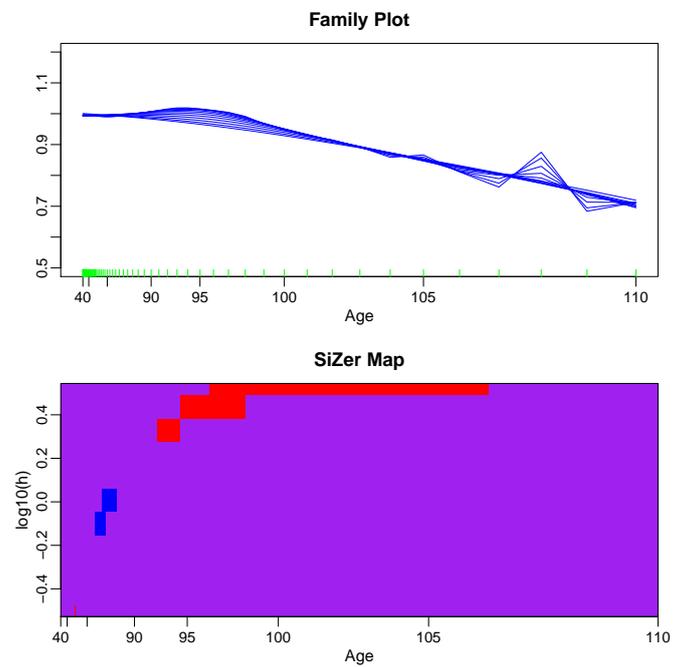**Figure 4.** SiZer analysis from mortality data in UK in 2006, transformed under a Gamma frailty hazard model.
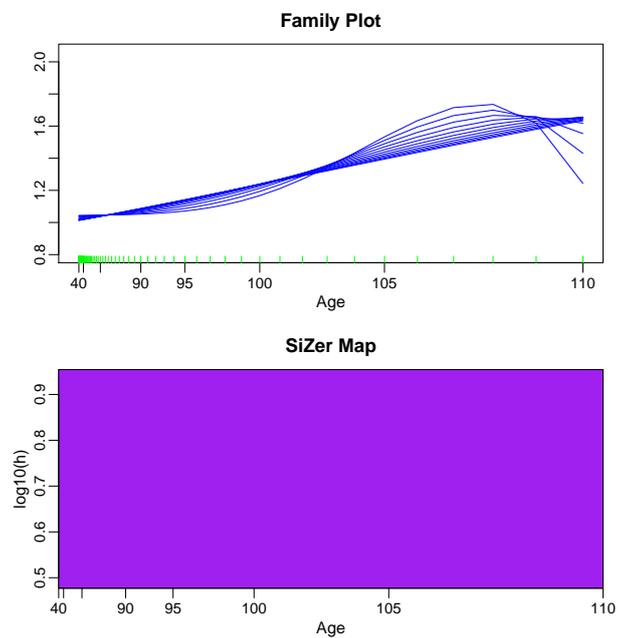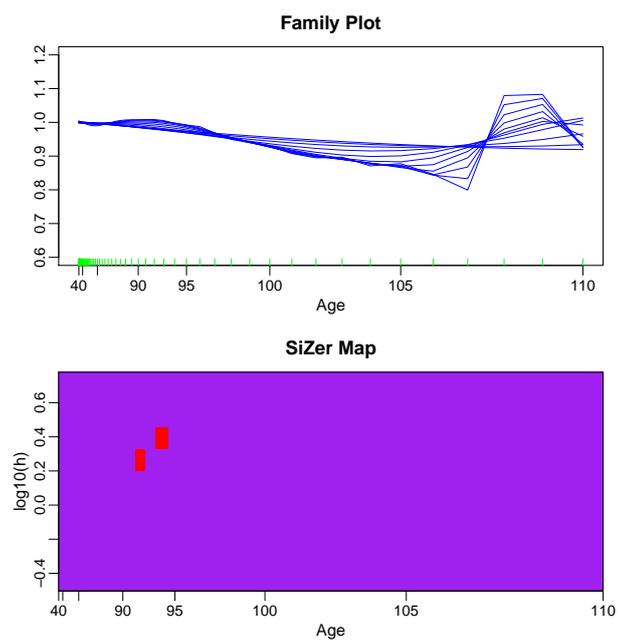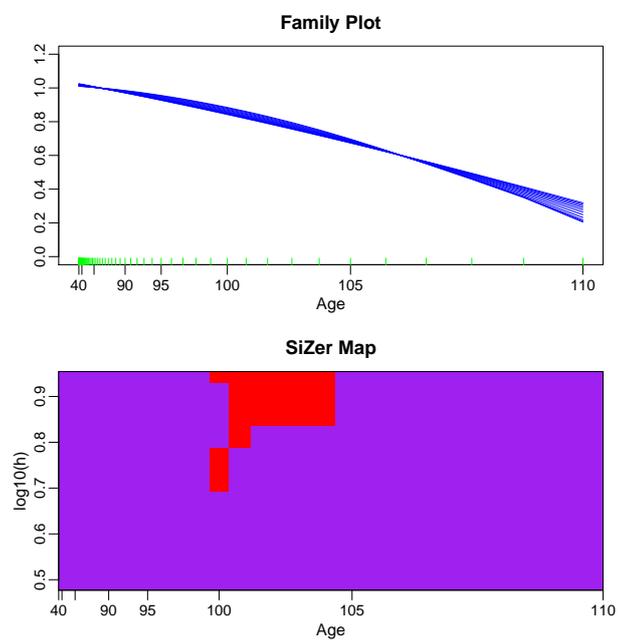

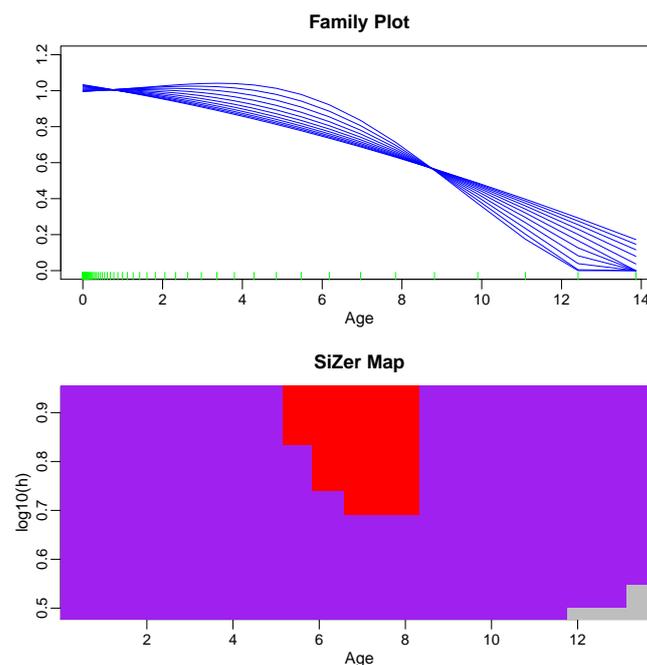**Figure 5.** SiZer analysis from mortality data in Denmark in 2006, transformed under a Gamma frailty hazard model.
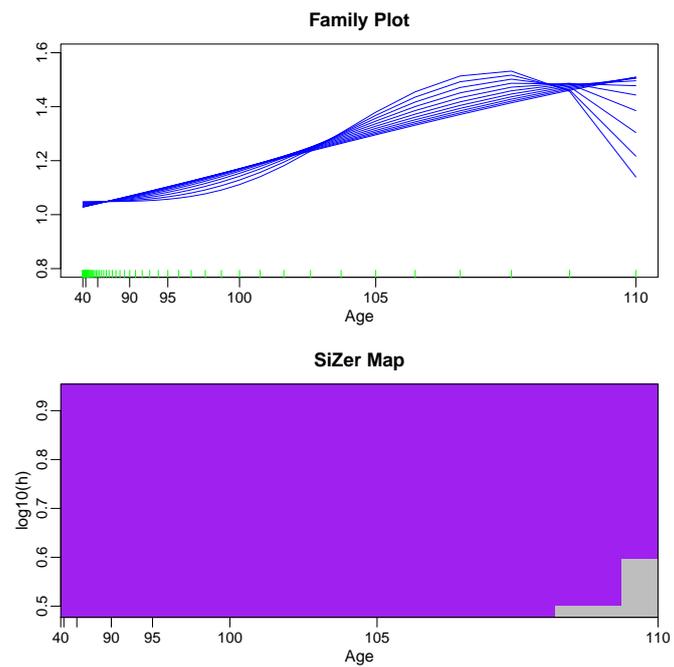
Figure 6. SiZer analysis from mortality data in Iceland in 2006, transformed under a Gamma frailty hazard model.

Figure 7. SiZer analysis from mortality data in US in 2013, transformed under a Gamma frailty hazard model.

Figure 8. SiZer analysis from mortality data in UK in 2013, transformed under a Gamma frailty hazard model.

Figure 9. SiZer analysis from mortality data in Denmark in 2013, transformed under a Gamma frailty hazard model.

Figure 10. SiZer analysis from mortality data in Iceland in 2013, transformed under a Gamma frailty hazard model.

5. Conclusions

In this paper we have developed the graphical exploratory tool SiZer map for inference of mortality rates with aggregated data. Using local linear smoothers for the hazard rate and its first derivative the developed SiZer tool is able to detect and confirm the underlying features in the hazard rate such as monotone patterns or local bumps. Besides parametric mortality models can be checked using a transformation approach and afterwards performing the SiZer analysis of the transformed data. The strength of the proposed SiZer analysis has been demonstrated on real mortality data and relevant conclusions have been drawn about a commonly used SAINT frailty model for this kind of data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Andersen, P., Borgan, O., Gill, R., Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

2. Bagkavos, D. (2011). Local linear hazard rate estimation and bandwidth selection. *Ann. Inst. Stat. Math.* **63**, 1019–1046.

3. Chaudhuri, P. and Marron, J.S. (1999). SiZer for exploration of structure in curves. *J. Amer. Statist. Assoc.*, **94**, 807–823.

4. Diggle, P. (1985). A kernel method for smoothing point process data, *Appl. Statist.*, **34**, 138–147.

5. Diggle, P. and Marron, J.S. (1988) Equivalence of Smoothing Parameter Selectors in Density and Intensity Estimation, *J. Amer. Statist. Assoc.*, **83** (403), 793–800.

6. Gámiz, M.L., Martínez-Miranda, M.D. and Nielsen, J.P. (2013) Smoothing survival densities in practice. *Comput. Statist. Data Anal.*, **58**, 368–382.

7. González-Manteiga, W., Martínez-Miranda, M.D., Raya-Miranda, R. (2008). SiZer Map for inference with additive models. Stat. Comput. **18**, 297–312.

8. Jarner, S.F. and Kryger E.M. (2011) Modelling adult mortality in small populations: The SAINT model.*ASTIN Bulletin* **41** (02) 377-418

9. Marron, J. S. and Zhang, J.T. (2005). SiZer for smoothing splines. *Computat. Statist.*, **20**, 481–502.

10. Nielsen, J. P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. *Scand. J. Statist.*, **28**, 675–698.

11. Nielsen, J. P., Tanggaard, C., and Jones, M.C. (2009). Local linear density estimation for filtered survival data. *Statistics*, **43**, 167–186.

12. Park, C. and Kang, K.-H. (2008). SiZer analysis for the comparison of regression curves. *Comput. Statist. Data Anal.*, **52**, 3954–3970.

13. Park, C., Hannig, J., and Kang, K. (2009). Improved SiZer for Time Series. *Statist. Sinica*, **19**, 1511–1530.

14. Park, C., Hannig, J., and Kang, K.-H. (2014). Nonparametric Comparison of Multiple Regression Curves in Scale-Space. *J. Comput. Graph. Statist.*, **23**, 657-677.

15. Park, C., Hernandez-Campos, F., Le, L., Marron, J. S., Park, J., Pipiras, V., Smith, F. D., Smith, R. L., Trovero, M. and Zhu, Z. (2011). Long Range Dependence Analysis of Internet Traffic. *J. Appl. Stat.*, **38**, 1407–1433.

16. Park, C., Lee, T. C. M. and Hannig, J. (2010). Multiscale exploratory analysis of regression quantiles using quantile SiZer. *J. Comput. Graph. Statist.*, **19**, 497–513.

17. Park, C., Marron, J. S. and Rondonotti, V. (2004). Dependent SiZer: goodness of fit tests for time series models. *J. Appl. Stat.*, **31**, 999–1017

18. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

19. Rondonotti, V., Marron, J. S., and Park, C. (2007). SiZer for time series: a new approach to the analysis of trends. *Electronic J. Statist.*, **1**, 268–289.

20. Spreeuw, J., Nielsen, J.P. and Jarner, S.J. (2013) A visual test of mixed hazard models. *SORT*, **37**, 149–170

21. Wang, J. L., Muller, H.G. and Capra, W. B. (1998). Analysis of oldest-old mortality: Lifetables revisited. *Ann. Statist.* **28**, 126—163.