Gámiz Pérez, M. L., Martinez-Miranda, M. D. & Nielsen, J. P. (2013). Smoothing survival densities in practice. Computational Statistics and Data Analysis, 58(1), pp. 368-382. doi: 10.1016/j.csda.2012.09.011

Smoothing survival densities in practice

M. Luz Gámiz Pérez *

María Dolores Martínez Miranda †

Jens Perch Nielsen ‡

September 6, 2012

Abstract

Many nonparametric smoothing procedures consider independent identically distributed stochastic variables. While many important smoothing applications accommodate to this type of data, there are also many nonparametric smoothing applications, where the data is more complicated. This paper considers survival data or filtered data defined as following Aalen's multiplicative hazard model and aggregated versions of this model. Aalen's model based on counting process theory allows multiple left truncations and multiple right censoring to be present in the data. This type of filtering is omnipresent in biostatistical and demographical applications, where people can join a study, leave the study and perhaps join the study again. This paper provides a data application to aggregated national mortality data, where immigrations to and from the country correspond to respectively left truncation and right censoring. The estimation methodology is based on a recent class of local linear density estimators to which we develop a new stable bandwidth-selector, the do-validated estimator. Our aggregated mortality data study illustrates that our new practical density estimators provide us with an important extra element in our visual toolbox for understanding survival data.

* *University of Granada, Spain*

† *Cass Business School, City University, London, U.K.*

‡ *Cass Business School, City University, London, U.K.*

Keywords and Phrases: Aalen's multiplicative model, bandwidth, crossvalidation, do-validation, filtered data, local linear estimation

1 Introduction

Smooth density estimation has been widely studied for identically and independent distributed data. One reason being the importance of providing density curves and functionals of density curves of this type of data. Another and perhaps as important reason is that the analysis of this type of data is the simplest possible and that the insights from studying this relatively simple case can be diffused into other areas of applications in mathematical statistics. In this paper we take advantage of recent detailed theoretical contributions of this simple type of data, see Mammen, Martínez-Miranda, Nielsen and Sperlich (2011). That paper describes in theory and practice why their new Do-validation bandwidth selector behave so well. In this paper we generalize their Do-validation estimator to filtered data allowing for left truncation and right censoring and we also show how this estimator should be calculated if only aggregated data is available. Our approach is a real generalization of Mammen et al. (2011) in the sense that had their been no filtering and no aggregation and had we had only classical independent and identically distributed data, then one element in our class of density estimators and its bandwidth selection simply reduces to the approach of Mammen et al.(2011). Our motivating application in this paper is old-age survival based on classical actuarial aggregated filtered data. Most practical mortality studies seem to be based on this type of data. We argue that while nonparametric hazard estimators become unstable when old-age is big then density estimators do no seem to have this problem. They keep stable - also for old age. In our particular study at hand, we consider a real life example visualizing the fit of a mortality model recently implemented in one of Europe's biggest pensions funds holding annuities for all current or former employees in Denmark. Clearly such a pension fund is not only concerned about the overall fit of their mortality model. Old age mortality is important to control for a pension fund paying lifelong guaranteed benefits to an entire population. The methodology provided by this

paper shows that the parametric mortality model used by the pension fund fits very well to the observed data. The fit is a little bit on the safe-side (seen from the pension company) in the sense that people live a little too long according to the model. The point of view of this paper is that such a conclusion could not have been produced from standard hazard methodology and we therefore recommend classical densities to be part of the toolbox also for statisticians working with survival data. Another point of this paper is that it really has to be filtered survival data one works with. Many good academic papers have for example worked with right censored data. While this of course is related to our data, one could not apply such methods to our aggregated data application with many left truncations, and one would not be able to apply such methods to our simulated data either. When discussing what density estimator to start from, one is to some extent in the same situation as in the studies of density estimation with independent identically distributed data. One could ask whether one should use standard kernel estimation as in Rosenblatt (1956) or whether one should proceed to more sophisticated principles as the local linear estimator of Jones (1993) or perhaps even the asymmetric kernels of Scaillet (2004). We think the recent class of local linear density estimators for filtered data by Nielsen, Tanggaard and Jones (2009) suffices for our purposes. It generalizes the well accepted local linear estimator of Jones (1993) to our filtered data case. For right censored data Kulasekera and Padgett (2006) and Bouezmarni, El Ghouch and Mesfioui (2011) develop positive density estimators generalizing the approach of Scaillet (2004). However, even in the area of independent, identically distributed data, there is still a debate on the best possible way to do asymmetric kernels and whether these estimators based on asymmetric kernels do actually outperform more classical kernel methods. For a very good new paper in this direction see Hirukawa and Sakudo (2012). In this paper we consider the well accepted and well understood approach of local linear estimation. The purpose of this paper is not to define new estimators for our type of data, but to take one good stable such estimator and make it practical by providing the final element of choosing the level of smoothing. So far only the crossvalidation approach has been defined for our type of estimator for our type of data. Cheng (1997) suggests a plug-in type of estimator for the local linear

estimator of Jones (1993). Plug-in estimation is, however, nontrivial to generalize to our setting. Also, the recent study by Mammen, Martínez-Miranda, Nielsen and Sperlich (2011) indicates that it might not be necessary after all to develop plug-in type estimators for our density estimator. The more practical Do-validation method of choosing a bandwidth seems to work even better than the best plug-in estimators. We therefore choose to generalize the much simpler approach of Do-validation to our situation and we show that the performance of Do-validation is excellent also for our type of data. The Do-validation approach of Mammen *et al.* (2011) is based on recent developments of Hart and Yi (1998), Hart and Lee (2005), Martínez-Miranda, Nielsen and Sperlich (2009) and Savshuk, Hart and Sheather (2010a, 2010b) that all consider indirect type of bandwidth selections. The idea being to start with a rough problem, where crossvalidation tends to perform well and then readjust to the original case. Do-validation seems to be the simplest possible version of this idea. It first considers an estimator only using the left side of the kernel and then an estimator using only the right side of the kernel. The final bandwidth is the readjusted version of these two one-sided bandwidths. For recent contributions to local linear hazard estimation working with right censored data, see Spierdijk (2008), Bagkavos and Patil (2008) and Bagkavos (2011).

The paper is organized as follows. In Section 2 we set out our theoretical setting. In Section 3, we derive crossvalidation and Do-validation bandwidth for our local linear density estimators and Section 4 contains a finite sample study showing that Do-validation indeed is a strong bandwidth selector also for survival data. In Section 5 we consider aggregated discrete data in general. In sections 6 and 7 we go through the concrete aggregated mortality application. Section 8 is the conclusion.

2 Model

We consider the same model of Nielsen *et al.* (2009). We observe n individuals $i = 1, 2, \ldots, n$. Let N_i count observed failures for the ith individual in the time interval $[0, \tau]$: 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 $\mathcal{F}_t$, $t \in [0, \tau]$, i.e. one that obeys *less conditions habituelles*, see Andersen, Borgan, Gill and Keiding (1993, p.60). We model the random intensity as

$$\lambda_i(t) = \alpha(t) Y_i(t) \tag{1}$$

with no restriction on the functional form of $\alpha(\cdot)$. Also 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.

2.1 The class of local linear estimators

The local density estimators of Nielsen *et al.* (2009) involve a pilot estimator of the survival function. We use the well known Kaplan-Meier estimator. One could easily replace this Kaplan-Meir pilot estimator with any other pilot estimator of the survival function if this seems appropriate for the problem at hand.

The Aalen estimator of the conditional integrated hazard function $\Lambda(t) = \int_0^t \alpha(s) ds$ can be expressed as

$$\widehat{\Lambda}(t) = \sum_{i=1}^{n} \int_0^t \left\{ Y^{(n)}(s) \right\}^{-1} dN_i(s),$$

where $Y^{(n)}(s) = \sum_{i=1}^{n} Y_i(s)$ is the risk set also called here the exposure process, which means, the number of individuals under observation by time s. The corresponding product integral estimator for the survival function,

$$\widehat{S}(t) = \prod_{s \leq t} \left\{ 1 - d\widehat{\Lambda}(s) \right\}$$

is the Kaplan-Meier product limit estimator, see Fleming and Harrington (1991). Tsai, Jewell and Wang (1987) adapt this estimator by modifying the expression $Y^{(n)}(s)$ in order to account for truncated in addition to censored data.

Let K be a probability density function and define $K_b(\cdot) = b^{-1} K(b^{-1} \cdot)$ for any bandwidth $b > 0$. The local linear density estimator defined in Nielsen *et al.* (2009) is

$$\widehat{f}_{b,K}(t) = \sum_{i=1}^{n} \int \overline{K}_{t,b}(t - s) W(s) Y_i(s) \widehat{S}(s) dN_i(s), \tag{2}$$

where

$$\overline{K}_{t,b}(t-s) = \frac{a_2(t) - a_1(t)(t-s)}{a_0(t)a_2(t) - \{a_1(t)\}^2} K_b(t-s)$$

and

$$a_j(t) = \int K_b(t-s)(t-s)^j W(s) Y^{(n)}(s) ds, \quad j = 0, 1, 2. \tag{3}$$

Notice that

$$\int \overline{K}_{t,b}(t-s) W(s) Y^{(n)}(s) ds = 1, \int \overline{K}_{t,b}(t-s)(t-s) W(s) Y^{(n)}(s) ds = 0,$$

$$\int \overline{K}_{t,b}(t-s)(t-s)^2 W(s) Y^{(n)}(s) ds > 0,$$

so that $\overline{K}_{t,b}$ can be interpreted as a second order kernel with respect to the measure μ, where $d\mu(s) = W(s)Y^{(n)}(s)ds$ and $W(s)$ is an arbitrary weight function that might depend on the data. For any given $W(s)$ the pointwise asymptotic behavior is the same. Two weight functions are particularly interesting. The first is simply the unit weight function $W(s) = 1$ and the second is the so called Ramlau-Hansen weighting $W(s) = \{n/Y^{(n)}(s)\} I(Y^{(n)}(s) > 0)$. The Ramlau-Hansen weigting is the weighting that generalizes classical kernel density estimation in the simple case of independent and identically distributed observations. We follow Nielsen *et al.* (2009) and use the unit weighting that have some stability advantages over the Ramlau-Hansen weighting. The final expression of our density estimator is therefore

$$\widehat{f}_{b,K}(t) = \sum_{i=1}^{n} \int \overline{K}_{t,b}(t-s) Y_i(s) \widehat{S}(s) dN_i(s). \tag{4}$$

Our Do-validation method uses directly the pointwise asymptotic theory of the local linear estimator. We therefore quickly go through this theory in the next section.

2.2 Asymptotic theory

We assume that the following general assumptions hold:

- There exist a function $\gamma \in C_1([0,\tau])$ positive in t which is the limit of the exposure function, that is

$$\sup_{s \in [0,\tau]} \left| Y^{(n)}(s)/n - \gamma(s) \right| \xrightarrow{P} 0;$$

- The bandwidth b satisfies that $b \to 0$ and $nb \to \infty$ as $n \to \infty$;
- $f \in C_4([0, \tau])$.

We write the error term $\widehat{f}_{b,K}(t) - f(t)$ as a variable part V_t converging in distribution plus a stable part B_t converging in probability. Here V_t is not exactly the variance and B_t is not exactly the bias, but V_t and B_t are analytically tractable quantities that are asymptotically equivalent to the variance and the bias. Assuming that the kernel K is symmetric, the asymptotic theory of the local linear filtered density estimator is

$$(nb)^{1/2}\left(\widehat{f}_{b,K}(t) - f(t) - b^2 B_t\right) \xrightarrow{\mathcal{D}} N(0, V_t) \tag{5}$$

where $B_t = \frac{1}{2}\kappa_2 f''(t)$ with

$$\kappa_2 = \int K(v)v^2 dv, \tag{6}$$

and $V_t = g_1 f(t) S(t) \{\gamma(t)\}^{-1}$ with

$$g_1 = \int K^2(v)dv. \tag{7}$$

For a one-sided kernel $K^*(v)$ the above kernel constants simply change to $\overline{\kappa}_2 = \int \overline{K}^*(v)v^2 dv$ and $\overline{g}_1 = \int [\overline{K}^*(v)]^2 dv$, involving the equivalent kernel

$$\overline{K}^*(v) = \frac{\kappa_2 - \kappa_1 v}{\kappa_2 - \kappa_1^2} K(v),$$

where $\kappa_1 = \int K(v)v dv$. This is key to understanding our readjustment constant in our Do-validation method.

The martingale nature of the problem transfers weak convergence into $\mathcal{L}_2$–convergence (see Andersen *et al.* 1993), when we add the assumption

$$\sup_{s \in [0,\tau]} \left| E\left\{Y^{(n)}(s)\right\}^{-1} - \{\gamma(s)\}^{-1} \right| \longrightarrow 0.$$

Under this new assumption we then get that

$$E\left(\widehat{f}_{b,K}(t) - f(t)\right)^2 = b^4 B_t^2 + (nb)^{-1} V_t + o\left(b^4 + (nb)^{-1}\right).$$

An application of the Fubini's Theorem therefore leads to the following asymptotic expansion of the Mean Integrated Square Errors (MISEs)

$$E\int_0^\tau \left(\widehat{f}_{b,K}(t) - f(t)\right)^2 dt = b^4 \int_0^\tau (B_t)^2 dt + (nb)^{-1} \int_0^\tau V_t dt + o\left(b^4 + (nb)^{-1}\right).$$

These asymptotic expansions of MISEs are important when developing the Do-validation procedure in the next section.

The MISE derivations lead to the following optimal deterministic bandwidth selector for the local linear density with symmetric kernel K:

$$b_{MISE} = \left\{ \frac{g_1}{\kappa_2^2} \frac{\int_0^\tau f(t)S(t)\{\gamma(t)\}^{-1}dt}{\int_0^\tau \left\{\frac{1}{4}f''(t)^2\right\} dt} \right\}^{1/5} n^{-1/5}. \tag{8}$$

Plug-in approaches for estimating b_{MISE} could be defined using smooth pilot estimators of the density f. However, as pointed out in the introduction, the extra practical disadvantages of considering such pilot estimators have made us decide that it is not worth the trouble. In particular because Do-validation seems to work better for finite samples than the more complicated plug-in estimator anyway, see Mammen *et al.* (2011).

3 Bandwidth selection by crossvalidation and do-validation

Let $\widehat{f}_{b,K}$ be a density estimator depending on a bandwidth b and a kernel K. The crossvalidation procedure below reduces to classical least-squares crossvalidated kernel density estimation when the Ramlau-Hansen weighting is being used. We therefore see the approach below as a natural generalization of classical crossvalidation to our more general filtered data case. Ideally, we would like to choose the smoothing parameter as the minimizer of

$$Q_0(b) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \widehat{f}_{b,K}(s) - f(s) \right\}^2 Y_i(s)ds \tag{9}$$

which is equivalent to minimizing

$$n^{-1} \left\{ \sum_{i=1}^n \int_0^\tau \left[\widehat{f}_{b,K}(s) \right]^2 Y_i(s)ds - 2\sum_{i=1}^n \int_0^\tau \widehat{f}_{b,K}(s) f(s) Y_i(s)ds \right\}.$$

Only the second of these terms depends on the unknown density and therefore only the second term needs to be estimated from data. Our estimator of $Q_0(b)$ is

$$\widehat{Q}_0(b) = n^{-1} \left\{ \sum_{i=1}^n \int_0^\tau \left[\widehat{f}_{b,K}(s) \right]^2 Y_i(s)ds - 2\sum_{i=1}^n \int_0^\tau \widehat{f}_{b,K}^{[i]}(s) \widehat{S}(s) dN_i(s) \right\}, \tag{10}$$

where $\widehat{f}^{[i]}_{b,K}(s)$ is the estimator arising when the data set is changed by setting the stochastic process N_i equal to 0 for all $s \in [0, \tau]$ and $\widehat{S}(s)$ is a suitable estimator (KM estimator in case of censoring and TJW estimator in case of censoring and truncation) of the survival function. The cross validation bandwidth $\widehat{b}_{CV}$ is defined as the minimizer of $\widehat{Q}_0(b)$. See also Nielsen *et al.* (2009) for more details and motivation of this crossvalidation procedure.

The Do-validation method is based on a combination of onesided crossvalidations that were introduced by Martínez-Miranda *et al.* (2009) for classical kernel density estimation

In our filtered data case, we first define the onesided crossvalidation score. Considering the leftsided kernel $K_L(u) = 2K(u)\mathbf{1}_{(-\infty,0)}$ from a symmetric kernel K. The left-onesided crossvalidation criterion is analogous to the crossvalidation criterion in (10) but involving a (leftsided) local linear density estimator $\widehat{f}_{b,K_L}$, namely the same estimator as in (4) but replacing K by K_L. Specifically we use the notation

$$\mathrm{OSCV}_L(b) = n^{-1}\left\{\sum_{i=1}^{n}\int_0^{\tau}\left[\widehat{f}_{b,K_L}(s)\right]^2 Y_i(s)ds - 2\sum_{i=1}^{n}\int_0^{\tau}\widehat{f}^{[i]}_{b,K_L}(s)\widehat{S}(s)dN_i(s)\right\}. \tag{11}$$

In exactly the same way we define the right-OSCV criterion, OSCV_R, except that $\widehat{f}_{b,K_L}$ in (11) is replaced by $\widehat{f}_{b,K_R}$, using the rightsided kernel, $K_R(u) = 2K(u)\mathbf{1}_{(0,\infty)}$.

The leftsided crossvalidation bandwidth comes from rescaling the minimizer of the left-OSCV score in (11), denoted by $\widehat{b}_L$. Specifically using the notation $\widehat{b}_{L,\mathrm{OSCV}}$, it is defined by

$$\widehat{b}_{L,\mathrm{OSCV}} = \mathrm{C}\widehat{b}_L, \tag{12}$$

where the rescaling constant is defined as the ratio of the MISE-optimal bandwidth for the local linear estimator $\widehat{f}_{b,K}$ (involving the symmetric kernel K) and its leftsided version $\widehat{f}_{b,L}$ (that involves the leftsided kernel K_L).

Using the asymptotic expressions for the MISE-optimal bandwidths in (8) the rescaling constant becomes

$$\mathrm{C} = \left(\frac{\overline{g}_1}{g_1}\frac{\kappa_2^2}{\overline{\kappa}_2^2}\right)^{-1/5}, \tag{13}$$

for the local linear estimator $\widehat{f}_{b,K}$. The constants $\overline{\kappa}_2$ and $\overline{g}_1$ are those defined in Section 2 but replacing K by its leftsided version K_L.

Finally the Do-validation bandwidth $\widehat{b}_{DO}$ is given by

$$\widehat{b}_{DO} = \frac{1}{2}(\widehat{b}_{L,OSCV} + \widehat{b}_{R,OSCV}),$$

with $\widehat{b}_{R,OSCV}$ analogous to (12) but involving the kernel K_R. As Mammen *et al.* (2011) pointed left-onesided crossvalidation and right-onesided crossvalidation are asymptotically equivalent, but not identical because of differences in the boundary. In fact their simulations reveal that Do-validation delivers a good stable compromise in finite samples.

4 Simulation studies

In this section we describe a simulation study close to that provided by Nielsen *et al.* (2009). We assume a similar setting, but with a filtered data scheme involving truncation and censoring and we investigate the performance of the bandwidth selectors. We compare four bandwidth estimates (standard crossvalidation, left and right onesided crossvalidation and Do-validation) for the local linear density estimator $\widehat{f}_{b,K}$.

4.1 Simulations with complete data

In our experiments we simulate seven different survival densities labeled f^k ($k = 1, \ldots, 7$). The three first densities are gamma distributions and the later consist of mixtures. The density f^1 is the gamma with parameters $\lambda = 1$, $r = 1$, where $r/\lambda = 1$ is the mean and $r/\lambda^2 = 1$ is the variance. The density f^2 has mean 1.5 and variance 1, while f^3 has mean 3 and variance 1. Introduce also the gamma density g with mean 6 and variance 1. Then, the mixtures are constructed from f^2, f^3 and g by using weight vectors, w, given by:

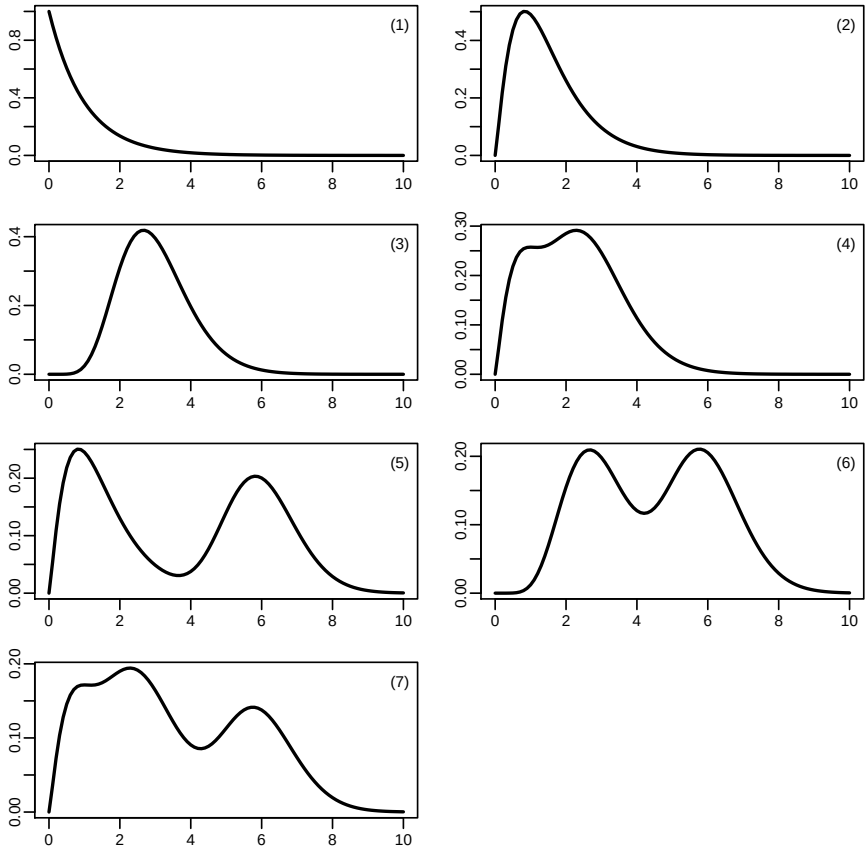

Figure 1: Simulated survival densities.

$$f^4 : w = (1/2, 1/2, 0), \qquad f^5 : w = (1/2, 0, 1/2),$$
$$f^6 : w = (0, 1/2, 1/2), \qquad f^7 : w = (1/3, 1/3, 1/3).$$

This set of densities is shown in Figure 1. For each model, an i.i.d. sample of size n was simulated randomly providing observations $X_1, \ldots, X_n$. Specifically we have considered sizes $n = 50, 200$ and 1000 and $R = 250$ repetitions of each model and sample size. Under the complete data situation the corresponding failure process at each estimation point is given by $N(t) = \sum_{i=1}^{n} I[X_i \leq t]$, and the risk process is simply $Y^{(n)}(t) = n - N(t^-)$. The local linear density estimator is calculated using the kernel

$$K(x) = \frac{3003}{2048}(1 - x^2)^6 I(-1 < x < 1). \tag{14}$$

We consider two infeasible strategies for bandwidth selection. The first of these methods is the *best possible bandwidth* that finds, for each simulated sample, the best possible bandwidth, in the sense of having smallest error in estimating the true density. The following is our measure of estimation error

$$Q_{0,l}(\widehat{f}^{(l)}_{b,K}, f) = n^{-1} \int_0^{\tau} \left[\widehat{f}^{(l)}_{b,K}(s) - f(s) \right]^2 Y^{(n)}(s) ds \tag{15}$$

for the estimator $\widehat{f}^{(l)}_{b,K}$ of the target density f, with bandwidth b and kernel K, computed from the lth simulated sample ($l = 1, \ldots, R$). We also tried an *average best bandwidth* strategy (labeled as ABB hereafter), which amounts to finding the bandwidth, which minimizes the averaged criterion

$$\overline{Q}_0(\widehat{f}_{b,K}, f) = \frac{1}{R} \sum_{l=1}^{R} Q_0(\widehat{f}^{(l)}_{b,K}, f). \tag{16}$$

We use these two infeasible bandwidth selectors as a benchmark to give us some guidance to how well our feasible bandwidth selectors are doing. The feasible bandwidth to be compared are the left and right-onesided crossvalidation $\widehat{b}_{L,OSCV}$, $\widehat{b}_{R,OSCV}$, the Do-validation method $\widehat{b}_{DO}$ and the standard crossvalidation bandwidth $\widehat{b}_{CV}$. In Section 4.4 we provide an explicit algorithm to carry out these simulations experiments with the software R.

Table 1 shows the performance averaged measure of the bandwidth estimates defined in (15) for each model and sample size. Also to asses how much the Do-validation bandwidth outperforms on the standard crossvalidation bandwidth we have calculated the relative error above Q_0 with respect to the infeasible best possible bandwidth (labeled as BB hereafter). Such measure is defined by

$$\texttt{Rerr} = \frac{m_{DO} - m_{BB}}{m_{CV} - m_{BB}} \tag{17}$$

where $m_\bullet$ is the average of the error measure Q_0 across the R samples and considering a specific criterion for the bandwidth selection. Note that `Rerr` indicates the reduction of the error Q_0 using Do-validation instead of standard crossvalidation. The resulting values for each model and sample size have been reported in Table 2.

		complete data						filtered data					
Density	n	BB	CV	OSCV$_L$	OSCV$_R$	DO	ABB	BB	CV	OSCV$_L$	OSCV$_R$	DO	ABB
1	50	10.88	18.11	16.74	23.08	18.66	15.06	12.67	23.22	21.50	28.41	23.52	19.39
	200	3.18	6.64	4.46	6.68	4.97	3.67	3.51	6.66	5.79	7.82	6.31	4.26
	1000	0.85	1.46	1.11	1.37	1.14	1.01	0.93	1.65	1.27	1.58	1.33	1.06
2	50	8.56	15.71	10.40	10.91	10.51	9.23	9.33	16.22	12.90	12.97	12.80	11.58
	200	2.82	4.41	3.69	3.50	3.55	3.08	2.86	4.73	3.74	3.89	3.77	3.26
	1000	0.88	1.25	1.04	1.04	1.03	0.99	0.90	1.36	1.18	1.12	1.13	1.01
3	50	5.13	9.77	6.51	6.30	6.35	5.48	5.25	9.78	7.01	6.78	6.84	5.37
	200	1.91	3.09	2.40	2.36	2.37	2.08	1.78	3.28	2.25	2.23	2.22	1.97
	1000	0.58	0.83	0.67	0.67	0.66	0.62	0.59	0.92	0.72	0.71	0.71	0.65
4	50	5.67	9.58	7.33	7.50	7.37	6.21	6.25	9.82	9.87	9.85	9.71	7.87
	200	1.99	3.23	2.52	2.58	2.52	2.30	1.90	3.08	2.90	3.08	2.96	1.98
	1000	0.36	0.81	0.77	0.76	0.75	0.67	0.70	1.07	0.96	1.00	0.97	0.85
5	50	7.20	10.29	8.25	8.33	8.21	7.61	6.47	9.71	8.28	8.71	8.35	6.67
	200	2.51	3.46	3.18	3.06	3.09	2.71	2.85	4.03	3.45	3.55	3.44	3.14
	1000	0.70	0.86	0.79	0.79	0.78	0.76	1.18	1.40	1.34	1.33	1.32	1.23
6	50	5.02	7.35	5.78	5.67	5.68	5.21	4.46	6.86	6.01	5.80	5.84	4.48
	200	1.70	2.56	2.14	2.11	2.12	1.83	2.28	3.59	2.83	2.88	2.84	2.48
	1000	0.52	0.68	0.60	0.60	0.59	0.55	1.46	1.80	1.67	1.68	1.67	1.52
7	50	4.84	7.07	5.84	5.94	5.86	5.18	6.02	8.68	8.10	8.51	8.13	6.13
	200	1.68	2.46	2.04	2.00	1.98	1.86	2.92	3.88	3.43	3.63	3.51	3.14
	1000	0.55	0.66	0.66	0.63	0.64	0.58	1.92	2.24	2.14	2.20	2.17	1.99

Table 1: Average of errors $Q_{0,l}(\widehat{f}^{(l)}_{b,K}, f)$ over the replicas $l = 1, \ldots, 250$. The numbers have been multiplied by 10^3.

Density	n	complete data	filtered data
1	50	1.076	1.028
	200	0.518	0.888
	1000	0.483	0.553
2	50	0.273	0.504
	200	0.459	0.487
	1000	0.396	0.512
3	50	0.262	0.350
	200	0.391	0.297
	1000	0.325	0.363
4	50	0.435	0.969
	200	0.424	0.895
	1000	0.689	0.728
5	50	0.327	0.580
	200	0.608	0.499
	1000	0.523	0.658
6	50	0.284	0.574
	200	0.485	0.429
	1000	0.473	0.619
7	50	0.456	0.795
	200	0.406	0.612
	1000	0.844	0.753

Table 2: Relative errors (`Rerr`): Do-validation is compared with crossvalidation.

4.2 Simulating data involving truncation and censoring

We focus in this section on the case where the lifetime variable X cannot be completely observed due to random left-truncation and random right-censoring. Left-truncation may occur if the time origin of the lifetime precedes the time origin of the observation period. Only those individuals that fail after the start of the study are being followed, otherwise they are left-truncated. We also include right-censoring in our simulated data. Let (T, X, C) be a random vector where X is the lifetime with continuous distribution function (c.d.f.) F, T is the random left-truncation time with c.d.f. L and C is the random right-censoring time with c.d.f. G. We assume that T, X, and C are mutually independent and that we observe the sample (T, Z, δ) only if $T \leq Z$, where $Z = \min\{X, C\}$ and $\delta = I\{Z = X\}$ is the random indicator of censoring. The data thus consist of independent samples such as $\{(T_1, Z_1, \delta_1), \ldots, (T_n, Z_n, \delta_n)\}$, with $T_i \leq Z_i$, for all $i = 1, \ldots, n$. In case that $T > Z$ there is no observation. If we denote $\beta = Pr\{T > Z\}$ as the probability of being left-truncated, it is obvious that $\beta < 1$ in order to have nonempty sample sets. On the other hand, let H be the c.d.f. of Z, then we have that $H = 1 - (1 - F)(1 - G)$.

To conduct the simulation study we consider for the lifetime X the same seven true density functions specified in Section 4.1. We independently generate random censoring times from the uniform $U(0, a_C)$, where a_C is selected so that the desired percentage of censoring is achieved on average across all iterations. Let denote by $\alpha \times 100\%$ such percentage.

To implement the random left-truncation, we also consider a uniform distribution for variable T, say $U(0, a_T)$. We set the value of a_T in order to achieve approximately the level of truncation indicated by β. The presence of truncation in the simulations is forcing us to generate extra values that will not be included in the sample, since when T gives a value greater than the value observed in Z the corresponding data is discarded. To assess the effect of truncation in our procedure we think that a very high level of truncation is not necessary so we limit ourselves to small values of β. This procedure results in a sample of triplets (T_i, Z_i, δ_i) with $T_i \leq Z_i$, and where approximately a $\alpha \times 100\%$ of the observations are $\delta_i = 0$, that is censored.

We consider $n = 50, 200$ and 1000 as in the i.i.d. case above. From this simulated sample set we construct the failure process $N(t)$ and the risk process $Y^{(n)}(t)$ at each estimation point t by $N(t) = \sum_{i=1}^{n} I[Z_i \leq t]\delta_i$ and $Y^{(n)}(t) = \sum_{i=1}^{n} I[T_i \leq t \leq Z_i]$, respectively.

Using the analogous performance measures described above for non-filtered data we compare the behavior of the four bandwidth estimates with the two unfeasible strategies. The results are summarized in the right hand of the tables 1 and 2.

4.3 Discussion of the simulation results

We find that Do-validation systematically outperforms crossvalidation significantly. Only in one case, the first model with 50 observations, do crossvalidation outperform Do-validation. But even in this case the difference is small. In general Do-validation outperforms crossvalidation with relative errors varying around half of those of crossvalidation. We also see that the left and right onesided crossvalidation do not have the same performance. Do-validation therefore provide a good stable compromise between the two avoiding unstable behavior when one of them does not work well.

4.4 Implementation and some computational issues

In the just presented simulation study we have implemented the continuous version of the local linear estimator defined in (2), which runs from continuous data as formulated in the model in Section 2. It is a different approach to the previously considered by Nielsen *et al.* (2009). These authors carried out their simulations experiments on an equally-spaced grid on the interval $[0, \tau]$, and used a discrete version of the local linear estimator, which runs on aggregated data as we will describe later in Section 5.2. Under such continuous formulation we describe in this subsection the algorithm we have used to perform the simulation experiments described above. We also discuss some issues about optimization and numerical integration, which is involved in the algorithm. And finally we provide a brief runtime analysis to give an insight to the reader about the practicability of the methods under R running on Windows platforms.

Let consider for any given model, $f \in \{f^k : k = 1, \ldots, 7\}$, and sample size, n, the general case of filtered data where a_C and a_T are defined to provide the desired percentage of censoring and truncation, respectively. Let define intervals $I_b = [range(f)/n, range(f)/2]$ for bandwidth selection in each case. The simulation algorithm can be described through the following steps:

Algorithm

Step 1. Simulate a sample: $\{(T_1, Z_1, \delta_1), \ldots, (T_n, Z_n, \delta_n)\}$ holding the desired levels of censoring and truncations as was described in Section 4.2.

Step 2. Calculate the best possible bandwidth (reported by BB in Table 1) for a given sample as the minimizer of the performance measure in (15), as a function of b in the interval I_b.

Step 3. Calculate the crossvalidation bandwidth estimate, $\widehat{b}_{CV}$, as the minimizer in the interval I_b of the crossvalidation score

$$\widehat{Q}_0(b) = n^{-1} \left\{ \int_0^\tau \left[\widehat{f}_{b,K}(s) \right]^2 Y^{(n)}(s) ds - 2 \sum_{i=1}^n \widehat{f}_{b,K}^{[i]}(Z_i) \widehat{S}(Z_i) \delta_i \right\}. \tag{18}$$

Step 4. Calculate the left-onesided crossvalidation bandwidth estimate as $\widehat{b}_{L,\mathrm{OSCV}} = \mathrm{C}\widehat{b}_L$, where $\widehat{b}_L$ is the minimizer of the (left) onesided crossvalidation score (11). The minimization is performed inside the rescaled interval I_b/C, where C = 0.5874 is the constant (13) calculated for the kernel in (14).

Step 5. Calculate the right-onesided crossvalidation bandwidth estimate, $\widehat{b}_{R,\mathrm{OSCV}}$, as in Step 4 but from the right onesided crossvalidation score.

Step 6. Calculate the do-validation bandwidth as $\widehat{b}_{DO} = (\widehat{b}_{L,\mathrm{OSCV}} + \widehat{b}_{R,\mathrm{OSCV}})/2$.

Step 7. Repeat steps 1–6 a large number of times and afterwards the best average bandwidth in (16) is calculated as the minimizer of the average of the performance measure along the simulated samples (as a function of b in the interval I_b).

Optimization. The above algorithm is a direct implementation of the bandwidth estimates using one dimensional optimization methods designed for continuous functions. In our experiments we have used the function `optimize` in the package `stats` in R. This function implements a method being a combination of golden section search and successive parabolic interpolation based on the Algol 60 procedure `localmin` by Brent 1973). Using this function we have calculated in the simulations the bandwidth estimates by applying the function `optimize` to the corresponding scores. In this aim these scores are considered as continuous functions of the bandwidth parameter, b, and the optimization is defined in the interval $I_b = [range(f)/n, range(f)/2]$ for each model $f \in \{f^k : k = 1, \ldots, 7\}$ and sample size n.

Numerical integration. In the algorithm the local linear estimator in (2) should be evaluated in the optimization, using the symmetric kernel and also its two onesided versions. Then it is required the calculation of the involved moments

$$a_j(t) = \int K_b(t-s)(t-s)^j Y^{(n)}(s)ds, \tag{19}$$

for $j = 0, 1, 2$. Note that these integrals can be exactly calculated for polynomial kernels as that in (14). However it cannot be done with the performance measure and also with the CV and OSCV scores. In this case we have considered numerical integration through the R function `integrate`, which implements unidimensional adaptive Gauss-Kronrod quadrature based on the Fortran functions `DQAGE`, and `DQAGIE` from QUADPACK (Piessens, deDoncker-Kapenga, Uberhuber and Kahaner 1983). We use this function considering the maximum number of intervals being 100 and the relative accuracy of $1.11e-14$. Such choice is enough to provide stable approximations of the involved integrals as we can see from Table 3. This table shows the results from the methods when we increase this number from 100 to 200. Here we only report results for the two first models and sample sizes in the case of non-filtered data. However similar conclusions were derived from the other models, filtering schemes but also bigger number of subdivisions. We can confirm from these results that a bigger number of subdivisions does not alter the conclusions from our experiments reported in Table 1.

Density	n	sbd.	BB	CV	OSCV$_L$	OSCV$_R$	DO
1	50	100	10.88	18.11	16.74	23.08	18.66
		200	10.88	18.05	16.74	22.93	18.53
		%Rdiff.	3.0e-5	3.3e-3	3.7e-4	0.85	0.43
	200	100	3.18	6.64	4.46	6.68	4.97
		200	3.19	6.64	4.38	6.70	4.83
		%Rdiff.	0	9.4e-3	1.1e-3	0.67	0.52
2	50	100	8.56	15.71	10.40	10.91	10.51
		200	8.56	15.69	10.40	10.88	10.51
		%Rdiff.	6.1e-4	5.4e-2	1.4e-4	0.31	0.17
	200	100	2.82	4.41	3.69	3.50	3.55
		200	2.82	4.41	3.67	3.48	3.53
		%Rdiff.	0	0.02	6.3e-3	0.55	0.77

Table 3: Sensitivity of the adaptive Gauss-Kronrod quadrature approximation (implemented in the function `integrate` in R) to the fixed maximum number of subdivisions. The two first rows for each model and sample size show the average of errors as in Table 1 in the complete data case, with maximum number of subdivisions (sbd.) equal to 100 and 200, respectively. The rows labeled "%Rdiff." shows the medians of the relative differences, $100 \times \frac{|Q_{0,l}^{[100]}(\widehat{f}_{\widehat{b},K}^{(l)},f) - Q_{0,l}^{[200]}(\widehat{f}_{b,K}^{(l)},f)|}{Q_{0,l}^{[100]}(\widehat{f}_{b,K}^{(l)},f)}$, along the replicas.

n	Step 2	Step 3	Steps 4-5-6	Total
50	1.35	1.68	2.57	5.91
200	3.54	3.28	6.08	15.91
1000	8.88	44.92	72.60	169.87

Table 4: Approximated run-time complexity of the algorithm in minutes. Column 5 shows the total time to perform all the calculations required for one replica in the simulations experiments. Columns 2–4 show the time in performing the main steps in the algorithm (Step 2 $\rightarrow$ best possible bandwidth, Step 3 $\rightarrow \widehat{b}_{CV}$ and Steps 4-5-6 $\rightarrow \widehat{b}_{DO}$). Each reported time is the average over one simulated sample from each of the models f^k $(k = 1, \ldots, 7)$.

Run-time complexity. Cross-validation strategies and therefore our do-validation method are expensive to compute in the simulations. To overcome the time complexity of the algorithm described above, we have used the R package Rmpi to perform the simulations experiments reported in previous sections. Such package provides an interface (wrapper) to MPI APIs and interactive R slave environment. However we provide in Table 4 an idea of the computational complexity of the algorithm running in a standard computer (Pentium (R) Dual-Core CPU-E5700 with 3.00GHz and 2.00GB-RAM with R working under Windows 7-32 bits). Specifically we have evaluated the run-time of one arbitrary simulated (complete) sample from each of the models, f^k $(k = 1, \ldots, 7)$, for increasing values of the sample size, n. We have carried out these experiments considering numerical integration through the R function `integrate` with a maximum number of intervals of 100 and optimization through `optimize`, which are the same choices used to provide the simulation results summarized in the paper. The the resulting times is reported in Table 4 for increasing values of n.

5 Discrete and aggregated data

In this section we show that our approach is simple to adjust to discrete and aggregated data.

5.1 Discrete occurrences and exposures: a common data representation of survival data

Survival data often takes the shape of occurrences and exposures defined on intervals. It is for example common practice that national bureaus of statistics publish yearly mortality occurrences and exposures for their population. The national bureaus take their original continuous stochastic processes and transfer them to the more easily publishable yearly occurrences and exposures. We define such occurrences and exposures to be approximations to the following discrete approximations and aggregations of the continuous counting processes and their exposure:

$$O_r = \sum_{i=1}^{n} \int_{X_{r-1}}^{X_r} dN_i(x),$$

and

$$E_r = \frac{1}{\Delta_r} \sum_{i=1}^{n} \int_{X_{r-1}}^{X_r} Y_i(x)dx,$$

where $X_0, \ldots, X_m$ are discrete points and $\Delta_r = X_r - X_{r-1}$ $(r = 1, \ldots, m)$.

Survival data in the shape of occurrences and exposures are also well known in many other survival data applications and many applied statisticians and actuaries are so used to working with them that they think of them as original data. Different data providers might approximate their discrete occurrences and exposures from their continuous data in different ways, but everyone tends to agree that the discretization should be sufficiently accurate to allow for both a nonparametric and a parametric survival analysis.

5.2 Local linear estimation for discrete data

In this section we provide the computational aspects of calculating our local linear estimator when the data provider has provided a discrete set of occurrences and exposures. Note that our approach below also works when the discretization is not equidistant. That will become important in our application below. While data providers most often provide occurrences and exposures in an equidistant way, modern survival techniques might transform the data and analyze the data on a different

time scale for example. In such cases the transform will no longer be equidistant and this is exactly the situation in our data study below. Based on discrete occurrences and exposures our local linear density estimator (4) can be defined as

$$\widetilde{f}_{b,K}(t) = \sum_{r=1}^{m} \overline{K}_{d,t,b}(t - X_r^*)\widehat{S}(X_r^*)O_r$$

where

$$\overline{K}_{d,t,b}(t - X_r^*) = \frac{a_{d,2}(t) - a_{d,1}(t)(t - X_r^*)}{a_{d,0}(t)a_{d,2}(t) - \{a_{d,1}(t)\}^2} K_b(t - X_r^*)$$

and

$$a_{d,j}(t) = \sum_{r=1}^{m} K_b(t - X_r^*)(t - X_r^*)^j E_r \Delta_r, \quad j = 0, 1, 2,$$

with $X_r^* = (X_{r-1} + X_r)/2, \quad r = 1, \ldots, m$. Here $\widehat{S}(X_r^*)$ is the estimator of the survival function defined by $\widehat{S}(X_r^*) = \exp\left(-\sum_{i=1}^{r} \frac{O_i}{E_i}\right)$, for $r = 1, \ldots, m$.

Besides, the least-squares crossvalidation principle given in (9) can be formulated for discretized data and the estimator $\widetilde{f}_{b,K}$ as follows. Let consider the discretized version of the optimality criteria in (9) given by

$$Q_{d,0}(b) = n^{-1} \sum_{r=1}^{m} \left(\widetilde{f}_{b,K}(X_r^*) - f(X_r^*)\right)^2 E_r \Delta_r.$$

Then the crossvalidation bandwidth is defined as the minimizer of the following estimation

$$\widehat{Q}_{d,0}(b) = \sum_{r=1}^{m} \left(\widetilde{f}_{b,K}(X_r^*)\right)^2 E_r \Delta_r - 2 \sum_{r=1}^{m} \widetilde{f}_{b,K}^{[r]}(X_r^*)\widehat{S}(X_r^*)O_r,$$

where $\widetilde{f}_{b,K}^{[r]}(X_r^*)$ is the estimator arising when the dataset is changed by setting $O_r = O_r - 1$. Similarly the onesided crossvalidation bandwidths are defined from the discretized OSCV scores and therefore the Do-validation method.

6 Visual test of parametric submodel

We apply our new practical smoothed survival function to estimate the density function of a transformation of mortality data considered in Spreeuw, Nielsen and Jarner (2012). One of the authors of this paper is a research director in one of Europe's big

pension funds responsible for a recent implementation of their mortality model and hence for setting annuity reserves for all current or former employees of Denmark. One particular important question of interest of the mortality model is whether it is actually fitting old-age. While the filtered density approach proved very good at answering these questions, the bandwidths had - somewhat unsatisfying - to be chosen by eye ball. Crossvalidation did not work for this study. The nonparametric filtered density estimator was used to develop a graphical test evaluating the suitability of a candidate parametric mixed hazard model based on a gamma frailty mortality model. Spreeuw *et al.* (2012) first estimate this candidate mixed hazard model through a standard maximum likelihood approach. Then the survival data are transformed in such a way that had the estimated parametric mixed model been the true model, then the transformed data would follow a uniform distribution. The graphical test considers the density of the transformed survival data and inspect whether it looks uniform. Since the transformed data lies in the interval $[0, 1]$, the good boundary-correction properties of the local linear estimator are crucial.

6.1 The parametric mixed mortality model

The considered parametric mixed mortality model generalizes the classical Gompertz survival model partly by including more parameters and partly by including a multiplicative frailty component. We first define this mortality model for identically independent data, where the frailty interpretation is more immediate, after that we generalize the parametric mixed mortality model to our setting.

Assume a cohort of n i.i.d. individuals and assume that the individual frailty effect can be represented by a random gamma distributed random variable. Assume now that $B_1, \ldots, B_n$ are the n individuals i.i.d frailty parameters and assume $E[B_i] = 1$. Then the conditional force of mortality at age x given $B_i = b_i$ is $\mu(x, b_i) = b_i\bar{\mu}(x)$, for $i = 1, \ldots, n$. And the cohort mortality at age x is $\mu(x) = E[B|x]\bar{\mu}(x)$.

Assume that the underlying mortality - before adjusting for the frailty - is $\bar{\mu}(x) = \exp(a_0 + a_1x + a_2x^2)$. We consider the gamma specification where $E[B|x] = (1 + \sigma^2\overline{M}(x))^{-1}$, with $\overline{M}(x) = \int_0^x \overline{\mu}(s)ds$.

6.2 Visual test of parametric survival density

The dataset consists of mortality data of women from four countries, namely United States (US), United Kingdom (UK), Denmark and Iceland. These data were obtained from Human Mortality Database and concern the calendar year 2006. The data providers have constructed this data as approximations to the yearly occurrences and exposures. In the application ages from 40 to 110 were included. The parametric mixed hazard specification of Spreeuw *et al.* (2012) assumes that the underlying continuous data come from a counting process with intensity function

$$\lambda_i(x) = \mu_\theta(x) Y_i(x)$$

as in (1) with

$$\mu_\theta(x) = \frac{\exp(a_0 + a_1 x + a_2 x^2)}{1 + \sigma^2 \int_0^x \exp(a_0 + a_1 s + a_2 s^2) ds}$$

where $\theta = (a_0, a_1, a_2, \sigma^2)$ is a four-dimensional parameter. It is the same parametric hazard shape as we had in the simplified i.i.d situation in the above section. By including it in our stochastic framework we allow ourself to work with this parametric specification for our general type of data set with repeated left truncations and right censoring. Let us denote by F_θ the distribution function associated with μ_θ. First, we calculate the parametric maximum likelihood estimators of θ following Borgan (1984). Secondly, we follow Spreeuw *et al.* (2012) and transform the time axis with the function $F_{\widehat{\theta}}$. It is of course well known that such a transformation would result in uniformly distributed survival times - if $F_{\widehat{\theta}}$ represents the true underlying cdf of these survival times. We can therefore construct our practical smoothed survival density on this transformed data and visualize our estimator of the difference between the density represented by $F_{\widehat{\theta}}$ and the true underlying density. Concretely we consider the discrete time points used for our occurrences and exposures $X_1, \ldots, X_m$ and get their transformed versions $\overline{X}_r = F_{\widehat{\theta}}(X_r)$, for $r = 1, \ldots, m$ and $\overline{X}_0 = 0$. If $F_{\widehat{\theta}}$ is really a good description of the true model then the transformed data should be uniformly distributed and in such a case our density estimator based on the time transformed occurrences and exposures should be close to the unit line.

For these data Spreeuw *et al.* (2012) get an estimator of the density by choosing

Country	$\widehat{b}_{CV}$	$\widehat{b}_{DO}$	b_{SNJ}
United States	0.044	0.175	0.083
United Kingdom	0.044	0.222	0.111
Denmark	0.084	0.485	0.167
Iceland	0.592	0.497	0.333

Table 5: Estimated bandwidth for each method and the suggested value in Spreeuw *et al.* (2012).

the bandwidth according to whatever looked best for the particular data set. It leaded the authors to consider the local linear density estimator with bandwidths $b = 1/12$ for US, $b = 1/9$ for UK, $b = 1/6$ for Denmark and $b = 1/3$ for Iceland. These subjective choices are now compared with the bandwidth estimates $\widehat{b}_{CV}$ and the Do-validation method $\widehat{b}_{DO}$. The results for each country are presented in Table 5. Our above visual test could of course also be used to test other parametric models based on filtered data as for example the classical Weibull parametric model, see for example Balakrishnan and Mitra (2012)

Figure 2 shows the resulting estimates of the transformed points using local linear estimation with the corresponding $\widehat{b}_{DO}$ (solid curve) and $\widehat{b}_{CV}$ (discontinuous curve). These bandwidths are also reported in second and third columns of Table 5). As can be appreciated from the table, the Do-validation method is the one that produces the estimated values which are closer to the values that the authors recommended (shown in the last column of the table). However standard crossvalidation is completely wrong providing very small values for the bandwidths which lead to inappropriate density estimators in all cases, as it is displayed in Figure 2 (discontinuous curve). Do-validation works well providing density estimates close to the uniform. All the estimated transformed densities based on Do-validation bandwidths provide tail densities below one indicating that the parametrically estimated mortality is too heavy-tailed. Also, all estimated densities on the transformed scale - except for the small data Iceland case - are only a few percentages away from one. The maximal deviance around eight percent for the very old British is not sufficient to suggest the use of alternative parametric models. We also considered the case without any frailty. This is we simply estimated the parametric mortality model without a frailty. Based on our do-validated bandwidths we got the same conclusion

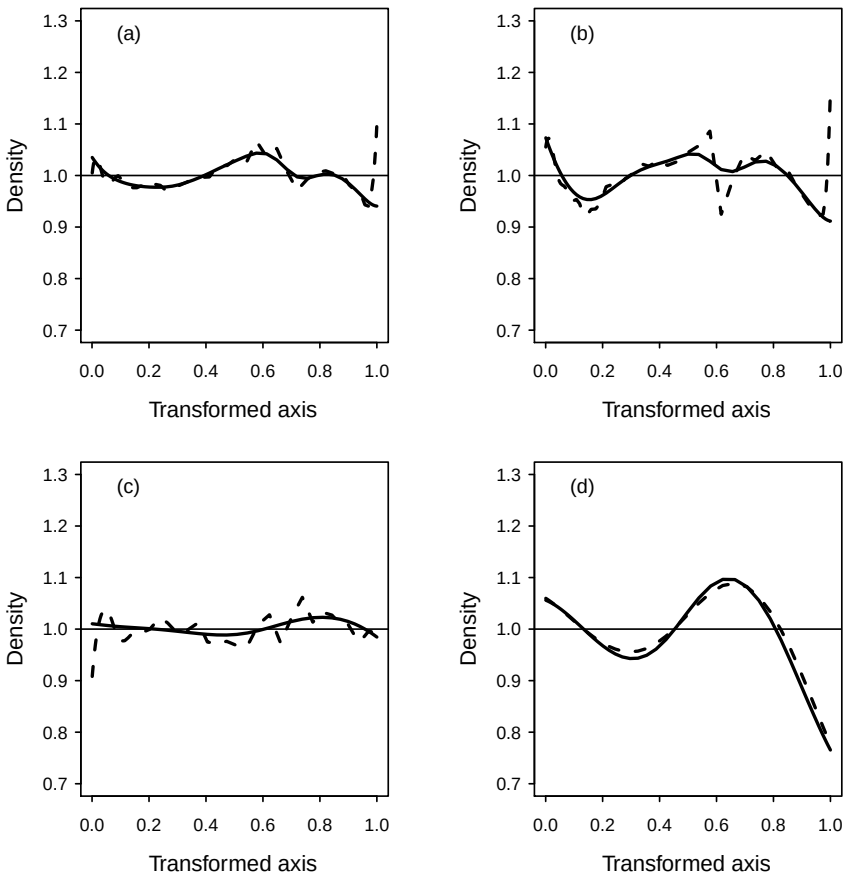

Figure 2: Mortality data application. Estimated density of the transformed data to judge the suitability of the gamma frailty specification for the SAINT model. Each panel shows local linear density estimate for each country: (a) US, (b) UK, (c) Denmark and (d) Iceland. In all the cases we have considered the Do-validation bandwidth (solid curve) and the crossvalidation bandwidth (dashed curve) which were reported in Table 5.

as Spreeuw *et al.* (2012): the frailty is absolutely necessary to get satisfactory fits.

From Figure 2 we can deduce that the parametric model is a very good approximation of the data since the plots present the estimated densities very close to the baseline at 1 i.e. the uniform density. Therefore, from our analysis, we confirm the conclusions of the authors who asses that the SAINT model with gamma frailty gives a very good representation of these data. Our findings here are that Do-validation should be used to choose automatically the bandwidth.

7 The classical hazard approach

In this section we consider the same problem as above, but this time using classical techniques from the toolbox of hazard estimation. The purpose of this section is to show, that while hazard estimation techniques are indeed very useful, our density estimation methodology provides crucial information that these hazard techniques can not give.

7.1 Standard method

First we compare the empirical integrated hazard with the parametric alternative on the transformed scale. If the estimated parametric mixed model is the true model then the integrated hazards evaluated at the transformed survival times should behave according to a standard exponential distribution. So, the standard method consists in constructing the Nelson-Aalen estimator for the transformed data and compare it with the cumulative hazard function corresponding to the standard exponential distribution. This can be graphically done by comparing the graph of the Nelson-Aalen estimator with a line of slope 1.

Specifically, we now transform the data as follows. Let $X_1, \ldots, X_m$ be the discretization points of the original data and let M_θ be the cumulative hazard function associated to μ_θ, then we calculate $\widetilde{X}_r = M_{\widehat{\theta}}(X_r)$, for $r = 1, \ldots, m$, and $\widetilde{X}_0 = 0$. Here $\widehat{\theta}$ is our maximum likelihood estimator of the parametric hazard specified above. Again if $\mu_{\widehat{\theta}}$ is really a good description of the true model then the transformed

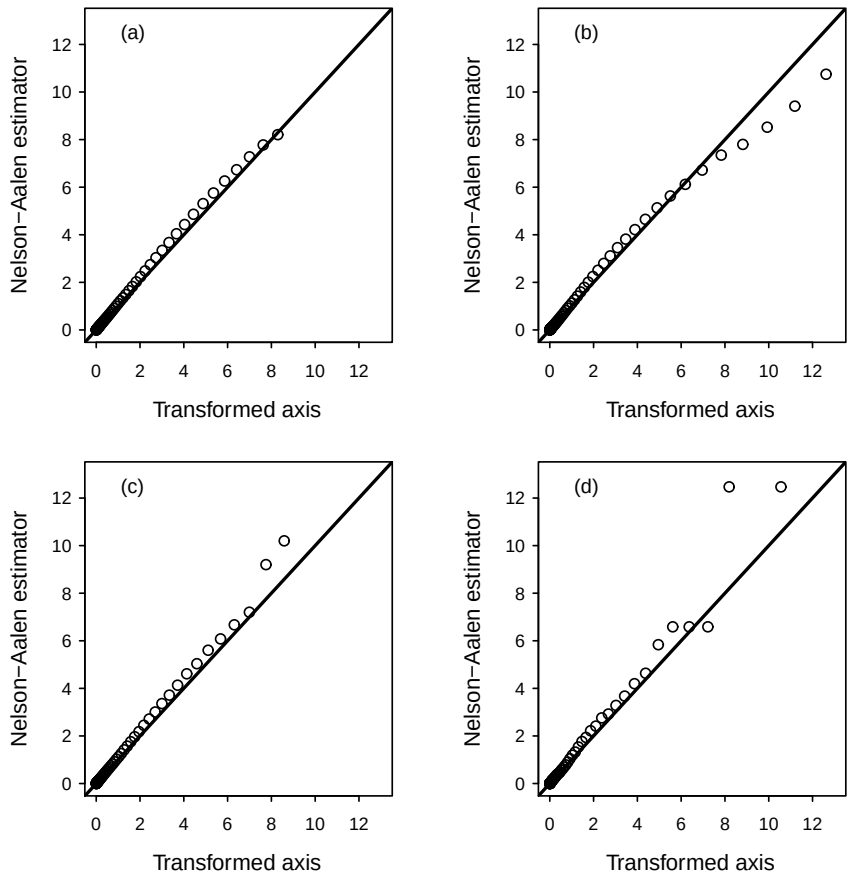

Figure 3: Nelson-Aalen estimator of the transformed data to judge the suitability of the gamma frailty specification for the SAINT model. The panels correspond to: (a) US, (b) UK, (c) Denmark and (d) Iceland.

data should be standard exponentially distributed. Let consider the Nelson-Aalen estimator given by

$$\widehat{\Lambda}(\widetilde{X}_r) = \sum_{i=1}^{r} \frac{O_i}{E_i},$$

for $r = 1, \ldots, m$. Figure 3 shows the results of the comparison for each country. The overall impression from the four graphs in Figure 3 is that three of them fit well with UK as an exception. The integrated hazard graphs have a tendency to overemphasize the importance of the early observations on the overall fit. They are less able to express right tail behavior. Also, the four graphs are not able to give us the same degree of understanding of the dynamics of the lack of fit as we get from the density graphs. Therefore, one can not with confidence express whether the assumed parametric frailty model fits well in the tail.

7.2 Visual test based on hazards

We consider transformed mortality data and provide a smooth estimate of the hazard rate on the transformed axis using the increments of the Nelson-Aalen estimator derived in the previous section. Let

$$\widehat{\alpha}_{b,K}(x) = \int_0^\infty K_b(x-s)d\widehat{\Lambda}(s) = \sum_{r=1}^{m} K_b(x-\widetilde{X}_r)\frac{O_r}{E_r},$$

for $x > 0$. This is the discrete version of the hazard estimator defined by Ramlau-Hansen (1983). If the model is suitable, the estimated hazard function should be close to the hazard function corresponding to a standard exponential random variable. Figure 4 displays the results for the four countries where we can see that the four hazards seem to be close to the exponential except for old-age. In fact the hazard seems to be very wrong for old-age and it is not clear that the true density is more heavy tailed than the estimated density. This is a clear difference between our new density based approach and classical hazard estimation: our new density approach is capable of making the conclusion that the original density is more heavy tailed than the estimated density.

8 Concluding remarks

In this paper we have provided data-driven bandwidth selectors for the local linear density estimation introduced by Nielsen *et al.* (2009) for survival data. These selectors are based on crossvalidatory scores and among them the so-called Do-validation bandwidth is shown to provide a good strategy in practice. Opposite to complicated plug-in procedures Do-validation is simpler, intuitive and does not require any pilot bandwidths. Even if a plug-in method would exist in our context, we are not sure that we would promote it after all. Mammen *et al.* (2011) argued that the known asymptotic optimality properties of plug-in estimator are not observed in practice. In practice, Do-validation seems to outperform plug-in estimation.

We provide a real-life analysis of mortality data based on our practical approach to filtered density smoothing and we show that our Do-validation method works

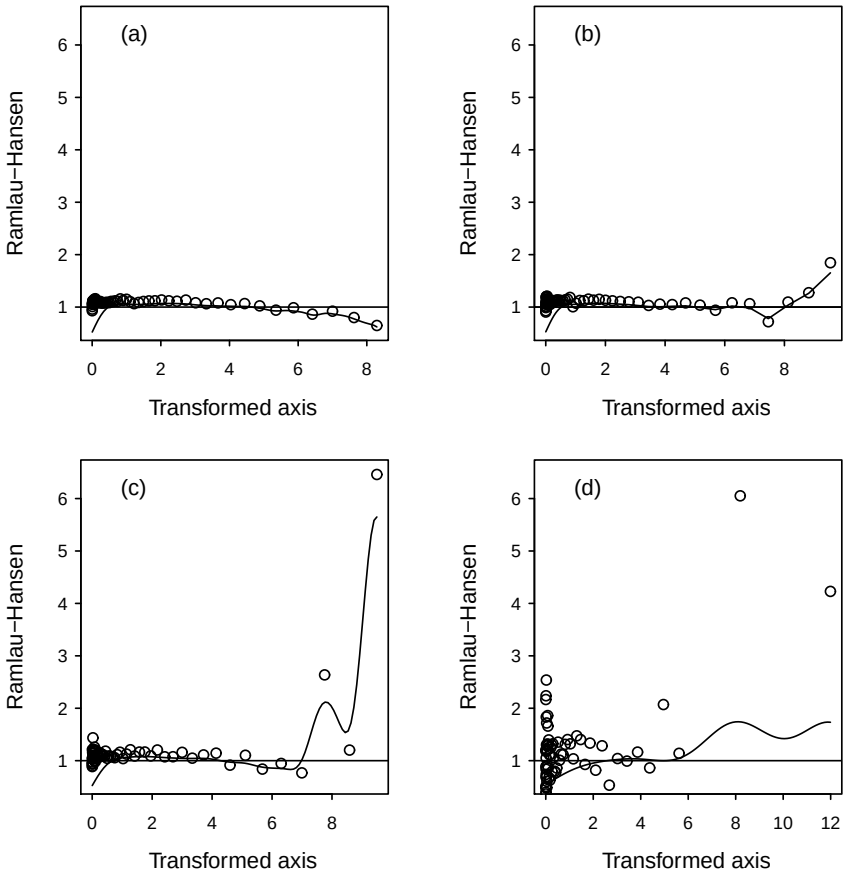

Figure 4: Mortality data application. Estimated hazard of the transformed data to judge the suitability of the gamma frailty specification for the SAINT model. The panels correspond to: (a) US, (b) UK, (c) Denmark and (d) Iceland.

well in this case, while standard crossvalidation breaks down. We also provide some motivation for having the filtered density estimator in the toolbox of the applied survival analyst. The survival density sometimes provides information that can not be extracted from hazard functions. We believe that our new filtered data bandwidth selection approach also generalizes to higher dimensional densities as in Buch-Kromann and Nielsen (2012) that is a structured filtered data version of the paper Buch-Kromann, Guillen, Linton and Nielsen (2011) and perhaps our approach even generalizes to filtered data regression, see Linton, Mammen, Nielsen and Van Keilegom (2011). Another interesting area of research is how to generalize our results to even more complicated filtering schemes, see for example the interesting general scheme of Balakrishnan and Kundub (2013)

Acknowledgments

Authors thank the Centro de Servicios de Informática y Redes de Comunicaciones (CSIRC), Universidad de Granada, for providing the computing time. The authors gratefully acknowledge the financial support of the Spanish "Ministerio de Ciencia e Innovación" by the grant MTM2008-03010 and the "Junta de Andalucía" grant TIC-06902. Second author acknowledge the support of the European Commission under the Marie Curie Intra-European Fellowship FP7-PEOPLE-2011-IEF Project number 302600. Also the authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

[1] Aalen, O.O. (1978) *Non-parametric inference for a family of counting processes.* Annals of Statistics, 6, 701–726.

[2] Andersen, P., Borgan, O., Gill, R. and Keiding, N. (1993) Statistical Models Based on Counting Processes, Springer, New York.

[3] Bagkavos, D. (2011) *Local linear hazard rate estimation and bandwidth selection.* Annals of the Institute of Statistical Mathematics, 63, 1019–1046.

[4] Bagkavos, D. and Patil, P.N. (2008) *Local polynomial fitting in failure rate estimation.* IEEE Transactions on Reliability, 57 (1),41–52.

[5] Balakrishnan, N. and Kundub, D. (2013) *Hybrid censoring: Models, inferential results and applications.* Computational Statistics and Data Analysis, 57 (1) 166–209.

[6] Balakrishnan, N. and Mitra, D. (2012) *Left truncated and right censored Weibull data and likelihood inference with an illustration.* Computational Statistics and Data Analysis, 56 (12) 4011–4025.

[7] Brent, R. (1973) Algorithms for Minimization without Derivatives. Englewood Cliffs N.J. Prentice-Hall.

[8] Borgan, O. (1984) *Maximum likelihood estimation in parametric counting process models with applications to censored failure time data.* Scandinavian Statistical Journal, 11, 1–16. Correction: 11, 275 (1984).

[9] Bouezmarni, T., El Ghouch, A. and Mesfioui, M. (2011) *Gamma kernels estimator for density and hazard rate of right censored data.* Journal of Probability and Statistics, 2011, 1–16.

[10] Buch-Kromann, T., Guillen, M., Linton, O. and Nielsen, J.P. (2011) *Multivariate density estimation using dimension reducing information and tail flattening transformations.* Insurance: Mathematics and Economics, 48, 99–110.

[11] Buch-Kromann, T. and Nielsen, J.P. (2012) *Multivariate density estimation using dimension reducing information and tail flattening transformations for truncated and censored data.* Annals of the Institute of Mathematical Statistics, 48(1), 167–192.

[12] Cheng, M.Y. (1997) *A bandwidth selector for local linear density estimators.* The Annals of Statistics, 25, 1001–1013.

[13] Fleming, T.R. and Harrington, D.P.(1991) Counting Processes and Survival Analysis, Wiley, NY.

[14] Hart, J.D. and Lee,C.L. (2005) *Robustness of one-sided crossvalidation to autocorrelation.* Journal of Multivariate Statistics, 92, 77–96.

[15] Hart, J.D. and Yi. S. (1998) *One-sided Cross-validation.* Journal of the American Statistical Association, 93, 135–146.

[16] Hirukawa, M. and Sakudo, M. (2012) *Nonparametric Multiplicative Bias Correction for Kernel-Type Density Estimation Using Positive Data.* Working paper, Faculty of Economics, Setunan University (Japan)

[17] Jones, C. (1993) *Simple boundary correction for kernel density estimation.* Statistics and Computing, 3, 135–146.

[18] Kulasekera, K.B. and Padgett, W.J. (2006) *Bayes bandwidth selection in kernel density estimation with censored data*, Journal of Nonparametric Statistics, 18 (2), 129–143

[19] Linton, O., Mammen, E., Nielsen, J.P and Van Keilegom, I. (2011) *Nonparametric regression with filtered data.* Bernoulli, 17(1), 60–87.

[20] Mammen, E., Martínez-Miranda, M.D., Nielsen, J.P. and Sperlich, S. (2011) *Do-validation for kernel density estimation*, Journal of the American Statistical Association, 106, 651–660.

[21] Martínez-Miranda, M.D., Nielsen, J.P. and Sperlich, S. (2009) *One sided cross-validation for density estimation with an application to operational risk.* In "Operational Risk Towards Basel III: Best Practices and Issues in Modelling. Management and Regulation, Ed. G.N. Gregoriou; John Wiley and Sons, Hoboken, New Jersey.

[22] Nielsen,J.P., Tanggaard, C. and Jones, M.C. (2009) *Local linear density estimation for filtered survival data.* Statistics, 43, 176–186.

[23] Piessens, R., deDoncker-Kapenga, E., Uberhuber, C. and Kahaner, D. (1983) Quadpack: A Subroutine Package for Automatic Integration. Springer Series in Computational Mathematics (Fortran routine available at `http://www.netlib.org/quadpack/`).

[24] R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna (Austria). URL: http://www.R-project.org

[25] Ramlau-Hansen, H. (1983) *Smoothing counting process intensities by means of kernel functions*. The Annals of Statistics, 11, 453–466.

[26] Rosenblatt, M. (1956). *Remarks on some nonparametric estimates of a density function*. The Annals of Mathematical Statistics, 27, 832- 837.

[27] Savchuk, O.Y., Hart, J.D. and Sheather, S. (2010a) *Indirect crossvalidation for Density Estimation*. Journal of the American Statistical Association, 105, 415–423.

[28] Savchuk, O.Y., Hart, J.D. and Sheather, S. (2010b) *An empirical study of indirect crossvalidation for Density Estimation*. IMS Lecture Notes - Festschrift for Tom Hettmansperger.

[29] Scaillet, O. (2004) *Density estimation using inverse and reciprocal inverse Gaussian kernels*. Journal of Nonparametric Statistics, 16 (1-2), 217–226.

[30] Spierdijk, L. (2008) *Nonparametric conditional hazard rate estimation: A local linear approach*, Computational Statistics and Data Analysis, 52, 2419–2434.

[31] Spreeuw, J., Nielsen, J.P. and Jarner, S.F. (2012) *A visual test of mixed hazard models*, submitted preprint.

[32] Tsai, W.Y., Jewel, N.P. and Wang, M.C. (1987) *A note on the product-limit estimator under right censoring and left truncation* Biometrika, 74, 883–886.