CREDIBILITY WEIGHTED HAZARD ESTIMATION

By

JENS PERCH NIELSEN AND BJØRN LUNDING SANDQVIST

Codan

ABSTRACT

Credibility weighting is helpful in many insurance applications where sparse data crave information from other sources of data. In this paper we aim at estimating a hazard curve using the nonparametric kernel method, where a credibility weighting principle is used locally, so that areas of sparse data for one subgroup can be alleviated by available information from other subgroups. The credibility estimator is found through a Hilbert space projection formulation of Bühlmann-Straub's credibility approach.

KEYWORDS AND PHRASES

Counting process theory, Kernel hazard estimation, Credibility, Bühlmann-Straub model.

1. INTRODUCTION

Credibility weighting of hazard estimators is a natural approach in situations where data is available on a number of different subgroups, but where there is a problem of sparse data within some or perhaps all of the subgroups. Hardy & Panjer (1998) considered a Bühlmann-Straub type of credibility estimator of the hazard loss ratio for a company relative to a standard mortality table, allowing them to predict the expected number of excess mortalities. In this paper we take a slightly different approach. We aim at estimating the entire hazard curve using the nonparametric kernel method. Credibility weighting is used locally for the different time areas, so that sparse data for one subgroup in certain areas can be alleviated by available information from other subgroups in those areas. We also use the Bühlmann-Straub credibility approach, but we concentrate on the particular interpretation of the Bühlmann-Straub method as a projection in a relevant Hilbert space onto a linear subspace. The result is a locally weighted credibility type estimator where the usual trade-off between within-subgroup variability and between-subgroups variability decides the weighting between a

local and a global hazard estimator. We use a slightly different version of the kernel type estimator of Ramlau-Hansen (1983), namely the local constant kernel estimator of Hjort (1992). The idea of using Bayesian estimation techniques to include prior information is well known in nonparametric survival analyses, see among others Hjort (1990). However, the natural and fairly simple approach of credibility weighting of smooth hazard curves seems to be new. The use of kernel smoothing in credibility estimation is not unknown, for example: Young (1997, 1998) uses kernel smoothing of densities to estimate the shape of the prior in a credibility model.

In §2 we outline the counting process formulation for the hazard model. In §3 we define the credibility estimator and in §4 we estimate the parameters and get the empirical credibility estimator. In §5 we motivate the estimator using the traditional set-up of the Bühlmann-Straub model. In §6 we apply the method to different subgroups of disability and in §7 a simulation study illustrates the effectiveness of the method. The proofs are deferred to the appendix.

2. A COUNTING PROCESS FORMULATION OF THE MODEL

We observe m individuals for k different subgroups and use the indexes $i = 1, ..., k$ and $j = 1, ..., n_i$ for respectively the k subgroups and the n_i individuals in the i'th subgroup. Let N_{ij} count observed failures for the j'th individual in the i'th subgroup in the time interval $[0, T']$. We assume that N_{ij} is a one-dimensional counting process with respect to an increasing, right continuous, complete filtration $\mathcal{F}_t$, $t \in [0, T']$, i.e. one that obeys *les conditions habituelles*, see Andersen *et al.* (1992, p. 60).

We model the random intensity as

$$\lambda_{ij}(t) = \gamma_i(t) Y_{ij}(t) = \theta_i(t)\alpha(t) Y_{ij}(t)$$

where $\gamma_i(t) = \theta_i(t)\alpha(t)$. Here, Y_{ij} is a predictable process taking values in $\{0,1\}$, indicating (by the value 1) when the j'th individual in the i'th subgroup is under risk, $\theta_i(t)$ is the continuous risk parameter for the i'th subgroup and $\alpha(\circ)$ is a deterministic unknown baseline hazard with no restriction on the functional form apart from smoothness assumptions. We assume that the stochastic processes $\theta_1, ..., \theta_k$ defined on the time interval $[0, T']$ are independent identically distributed. We furthermore assume that conditional on $\theta_1, ..., \theta_k$ we have the following: the counting processes $(N_{11}, Y_{11}), ..., (N_{kn_k}, Y_{kn_k})$ are independent for the m individuals and within the i'th subgroup the counting processes $(N_{i1}, Y_{i1}), ..., (N_{in_i}, Y_{in_i})$ are identically distributed. We also assume that $E\{\theta_i(t)\} = 1$, $Var\{\theta_i(t)\} = \sigma_t^2$ and $E\{Y_{ij}(t)\} = \delta_i(t)$ for $t \in [0, T']$.

3. THE CREDIBILITY ESTIMATOR

Our approach is first to define an estimator $\tilde{\gamma}_i(t)$ of $\gamma_i(t) = \theta_i(t)\alpha(t)$ based on the data from the i'th subgroup. We then follow classical credibility theory and define a Hilbert space projection of the stochastic variable $\gamma_i(t)$ down at the

linear space spanned by a constant and our observed candidate $\tilde{\gamma}_i(t)$ from the i'th subgroup. In Section 5 it is pointed out that the analouge to classical credibility theory goes even further than the Hilbert projection interpretation of this section, since an approach based on the jumps of each individual, parallelling the original approach of Bühlmann & Straub (1970) that took its starting point at the individual claims in each subgroup, leads to exactly the same theory as considered in this section.

Let K be a probability density function symmetric about zero and let $K_b(\circ) = b^{-1}K(\circ/b)$ for any $b > 0$. Consider the nonparametric local constant estimator in the i'th subgroup, see Hjort (1992)

$$\tilde{\gamma}_i(t) = \frac{\sum_{j=1}^{n_i} \int K_b(t-s)dN_{ij}(s)}{\tilde{Y}_{i.}(t)}$$

where $\tilde{Y}_{i.}(t) = \sum_{j=1}^{n_i} \tilde{Y}_{ij}(t)$ and $\tilde{Y}_{ij}(t) = \int K_b(t-s)Y_{ij}(s)ds$ is the smoothed exposure.

For a fixed $t \in [0, T']$ we consider the Hilbert space projection of $\gamma_i(t)$ onto the linear space $\{a_t + b_t\tilde{\gamma}_i(t)|a_t, b_t \in R\}$. Based on this Hilbert space projection we obtain the following expression for the optimal linear credibility estimator minimizing $E\{\gamma_i(t) - a_t - b_t\tilde{\gamma}_i(t)\}^2$, namely

$$(1 - z_{i,t})E\{\tilde{\gamma}_i(t)\} + z_{i,t}\tilde{\gamma}_i(t),$$

where

$$z_{i,t} = \frac{COV\{\gamma_i(t), \tilde{\gamma}_i(t)\}}{VAR\{\tilde{\gamma}_i(t)\}}.$$

In the appendix we derive the following equations

$$E\{\tilde{\gamma}_i(t)\} = \{1 + o_P(1)\}\alpha(t),$$

$$COV\{\gamma_i(t), \tilde{\gamma}_i(t)\} = \{1 + o_P(1)\}\sigma_t^2\alpha^2(t)$$

and

$$VAR\{\tilde{\gamma}_i(t)\} = \{1 + o_P(1)\}\left[C_2\alpha(t)\{\tilde{Y}_{i.}(t)\}^{-1} + \sigma_t^2\alpha^2(t)\right]$$

where $C_2 = \int K^2(u)du$ and where $o_P(1)$ is the little-o function related to the limit $b \to 0$ and $bn_i \to \infty$ (for all i).

Therefore, the optimal linear credibility estimator is approximately equal to

$$\hat{\gamma}_i(t) = (1 - z_{i,t})\alpha(t) + z_{i,t}\tilde{\gamma}_i(t)$$

where

$$z_{i,t} = \{1 + o_P(1)\}\frac{\sigma_t^2\alpha^2(t)\tilde{Y}_{i.}(t)}{C_2\alpha(t) + \sigma_t^2\alpha^2(t)\tilde{Y}_{i.}(t)}.$$

Notice that we have one less parameter than the traditional estimator in Bühlmann & Straub (1970). This is due to the one-to-one relationship between variance and mean for hazard estimators. To get a preliminary estimator of the underlying credibility parameter $\theta_i(t)$ we divide the above credibility equality by $\alpha(t)$. We get

$$\widehat{\theta}_i(t) = (1 - z_{i,t}) + z_{i,t}\tilde{\gamma}_i(t)/\alpha(t).$$

4. THE EMPIRICAL CREDIBILITY ESTIMATOR

A possible estimator of the baseline hazard α is

$$\tilde{\alpha}(t) = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} \int K_b(t-s)dN_{ij}(s)}{\sum_{i=1}^{k} \tilde{Y}_{i.}(t)} = \sum_{i=1}^{k} w_i(t)\tilde{\gamma}_i(t),$$

where

$$w_i(t) = \tilde{Y}_{i.}(t)\left\{\sum_{q=1}^{k} \tilde{Y}_{q.}(t)\right\}^{-1}.$$

Thus $\sum_{i=1}^{k} w_i(t) = 1$. Other possibilities of weighting exist.

While estimating the variance σ_t^2, then we employ that $E\{\theta_i(t)\} = 1$ and that $\tilde{\gamma}_i(t)$ is an estimator of $\gamma_i(t) = \theta_i(t)\alpha(t)$ and therefore $\tilde{\gamma}_i(t)\{\tilde{\alpha}(t)\}^{-1}$ is an estimator of $\theta_i(t)$. One nice smooth estimator of σ_t^2 is therefore

$$\tilde{\sigma}_t^2 = (k-1)^{-1} \sum_{i=1}^{k} \left[\tilde{\gamma}_i(t)\{\tilde{\alpha}(t)\}^{-1} - 1\right]^2.$$

By substitution of σ_t^2 with $\tilde{\sigma}_t^2$ and $\alpha(t)$ with $\tilde{\alpha}(t)$ in the formula for $z_{i,t}$ we can estimate $z_{i,t}$ by

$$\tilde{z}_{i,t} = \frac{\tilde{\sigma}_t^2\tilde{\alpha}^2(t)\tilde{Y}_{i.}(t)}{C_2\tilde{\alpha}(t) + \tilde{\sigma}_t^2\tilde{\alpha}^2(t)\tilde{Y}_{i.}(t)}.$$

The empirical credibility estimator for the ith subgroup is therefore

$$\widehat{\tilde{\gamma}}_i(t) = (1 - \tilde{z}_{i,t})\tilde{\alpha}(t) + \tilde{z}_{i,t}\tilde{\gamma}_i(t),$$

corresponding the following credibility estimator for the underlying θ in the ith subgroup:

$$\widehat{\tilde{\theta}}_i(t) = (1 - \tilde{z}_{i,t}) + \tilde{z}_{i,t}\tilde{\gamma}_i(t)/\tilde{\alpha}(t).$$

5. A MOTIVATION OF THE MODEL VIA THE TRADITIONAL BÜHLMANN-STRAUB SET-UP

Our credibility estimator does not only have the projection interpretation in common with methodology, originally defined in Bühlmann & Straub (1970). We can also motivate that the traditional set-up of Bühlmann-Straub models apply to our case and indeed will lead to the same approach as we use above. We give a short verification of this fact.

Define

$$\tilde{\gamma}_{ij}(t) = \frac{\int K_b(t - s)dN_{ij}(s)}{\tilde{Y}_{ij}(t)}.$$

which is the nonparametric local constant estimator for the j'th individual in the i'th subgroup.

Calculations analog to the appendix with $\tilde{\gamma}_i$ and $\tilde{Y}_{i.}$ replaced with $\tilde{\gamma}_{ij}$ and $\tilde{Y}_{ij}$ give

$$E\{\tilde{\gamma}_{ij}(t)|\theta_i(t)\} = \{1 + o_P(1)\}\theta_i(t)\alpha(t)$$

and

$$VAR\{\tilde{\gamma}_{ij}(t)|\theta_i(t)\} = \{1 + o_P(1)\}C_2\theta_i(t)\alpha(t)\{\tilde{Y}_{ij}(t)\}^{-1}.$$

For fixed t this is a special case of the Bühlmann-Straub model. The use of this approach will lead to the same results as in Section 3 above.

6. AN APPLICATION TO DISABILITY INSURANCE

We consider a portfolio of disability insurance policies. We divide the portfolio into four subgroups after sex and standard/substandard tables. The subgroups correspond to a subset of data used in Codan for rating of disability insurance. We have chosen not to reveal the details of these data sets for reasons of confidentiality. However, the estimation performance of these subsets is very similar to the estimation performance we have seen in a number of "in-house" applications – therefore we consider this application to be representative of our methodology. Figure 1 shows the four individual kernel hazard estimators for the four subgroups without using any credibility estimation. It is seen that there is big differences between the estimators and in particular a lot of fluctuation is present. The thick line in Figure 1 is the baseline hazard for the four subgroups. Note that this baseline hazard does not fluctuate too much. Figure 2 shows the credibility weighted hazards of the four different subgroups. While the four different credibility weighted hazards still are different, then it is noted that the fluctuations are much less dominant. In particular it can be noted that the hazard estimators look a lot more realistic in the age-group above 55 years where data is sparse.
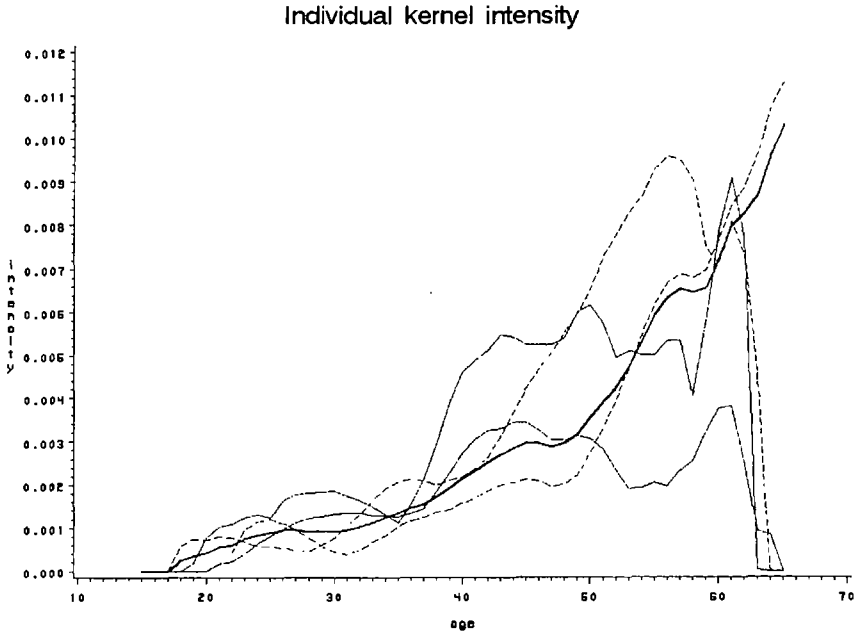
Individual kernel intensity

FIGURE 1: The four estimated individual disability hazards (male: dotted, female: solid) and the baseline (thick).

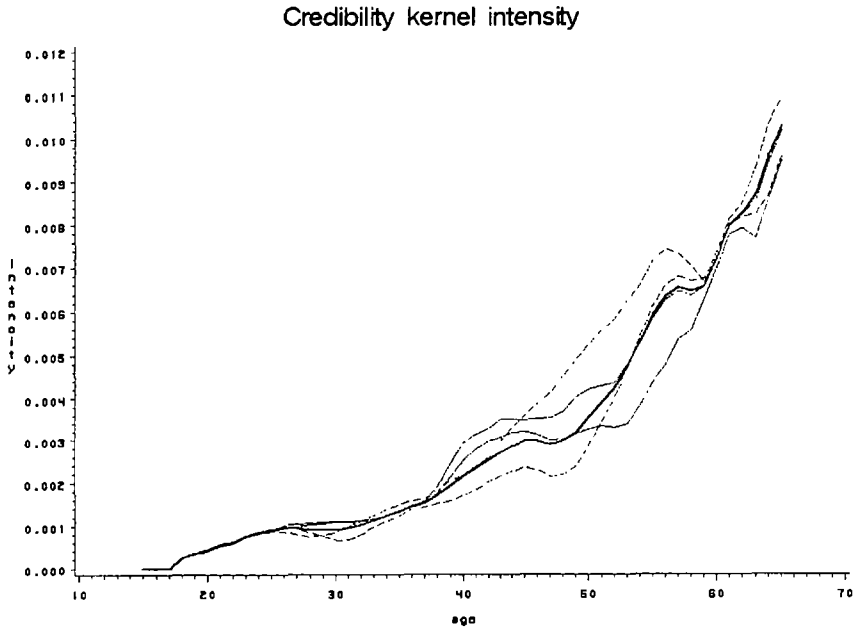## Credibility kernel intensity

FIGURE 2: The four estimated credibility disability hazards (male: dotted, female: solid) and the baseline (thick).

7. A SIMULATION STUDY

First we consider a model for simulating the stochastic processes $\theta_i(t)$. We take the approach of letting

$$\theta_i(t) = (1 - t)X_i + tY_i,$$

for $t \in [0, 1]$, where X_i and Y_i are independent uniformly distributed stochastic variables with support $[0.75, 1.25]$. The condition $E\{\theta_i(t)\} = 1$ is clearly fulfilled and $VAR\{\theta_i(t)\} = \sigma_t^2 = \frac{1}{48}\left[(1 - t)^2 + t^2\right]$. For clarity we chose to use the same bandwidth, namely $b = 0.1$ in all the simulated cases below. We use the following four baseline hazard functions:

$$\alpha_1(t) = B(t, 4, 4),$$
$$\alpha_2(t) = B(t, 2, 2),$$
$$\alpha_3(t) = 0.6 * [B(t, 0.5, 0.5) + B(t, 7, 7)],$$
$$\alpha_4(t) = 0.6 * [B(t, 0.5, 0.5) + B(t, 4, 2) + B(t, 2, 4)],$$

where $B(t, 4, 4)$ is the value of the density of a beta distribution with parameters $(4, 4)$. These hazards were used in Nielsen (1998) to compare bias corrected hazard estimators to traditional hazard estimators.

First we go through one particular simulation with the baseline hazard function α_1 and with 10 subgroups and 100 observations in each subgroup. The graph of the correct hazard function α_1 is shown on Figure 3 below and the ten individual hazard functions we aim at estimating, the γ's, are given in Figure 4. The individual risk credibility parameters, the θ's, are given in Figure 5.
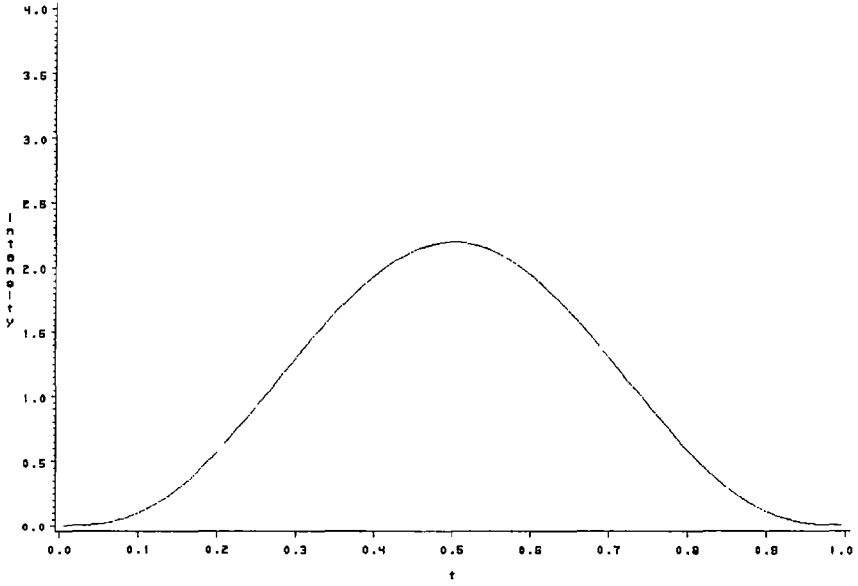
Beta baseline intensity

FIGURE 3: Beta(4, 4) baseline intensity. α_1.

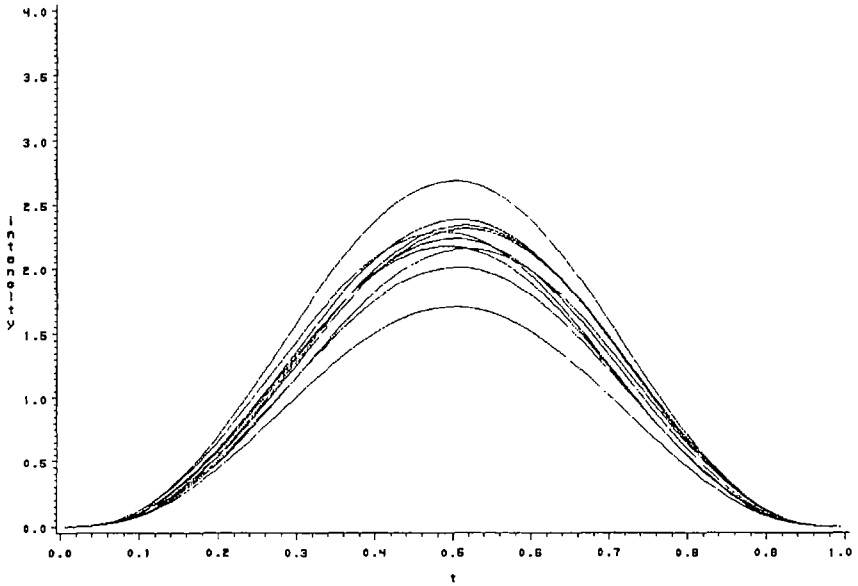## Individual intensity

FIGURE 4: The ten individual hazards.
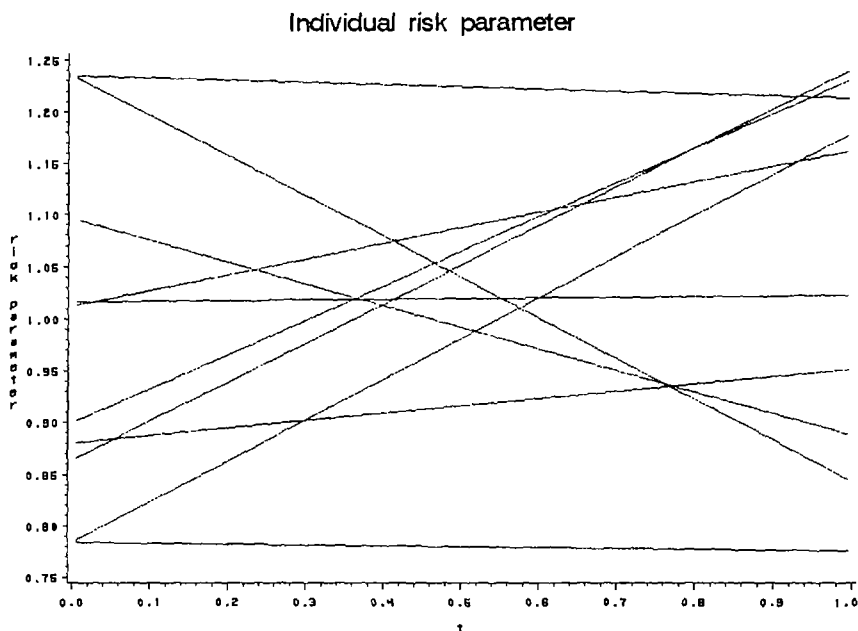
Individual risk parameter

FIGURE 5: The ten individual risk parameters.

The kernel hazard estimator of the baseline hazard is given in Figure 6 and the ten individual kernel hazard estimators are given in Figure 7 below. It is obvious that the baseline hazard estimator is a lot less volatile than the ten individual estimators. It can therefore be expected that the credibility weighted versions of the ten individual estimators will do better with respect to volatility. This is actually correct as can be seen in Figure 8, where we show the credibility weighted versions of the ten individual hazard functions.
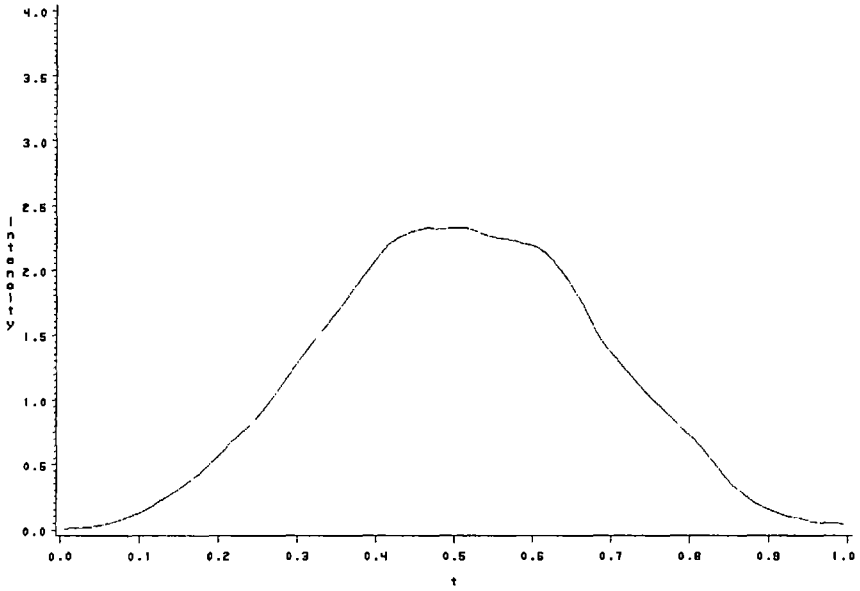
Baseline kernel estimator

FIGURE 6: Baseline kernel estimator for at study with $10 \cdot 100$ observations and $b = 0.1$.

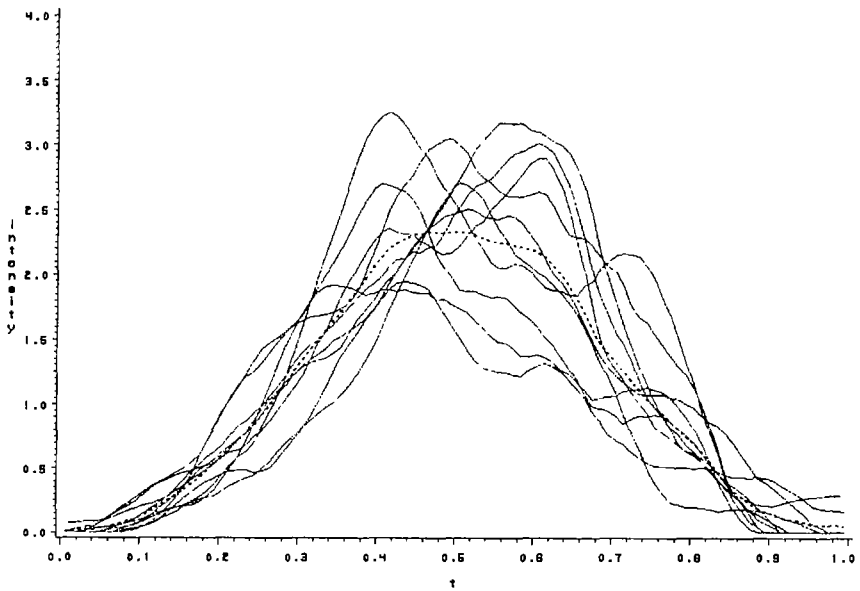## Individual kernel intensity estimator

FIGURE 7: Individual kernel estimator for the ten subgroups (solid) and the estimated baseline hazard (dotted). There are 100 observations in each subgroup and $b = 0.1$.
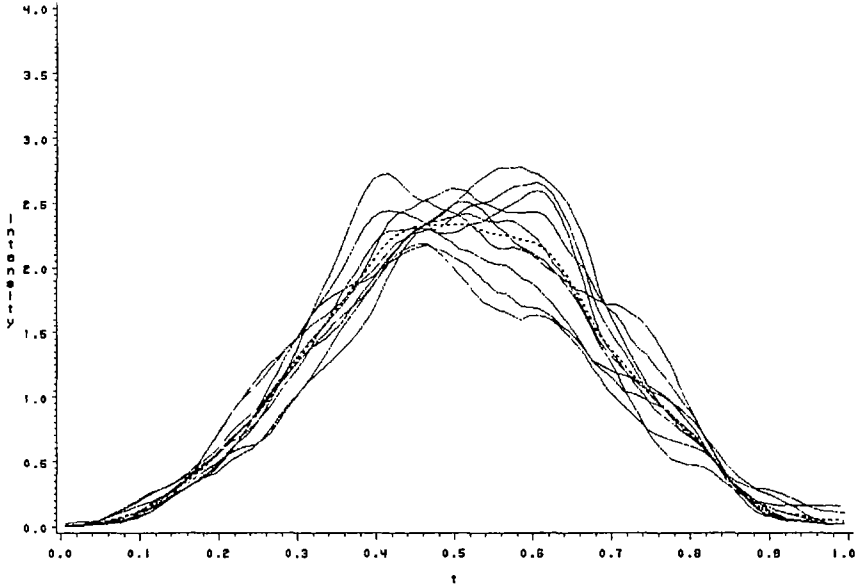
Credibility kernel intensity estimator

FIGURE 8: Credibility kernel estimator for the ten subgroups (solid) and the estimated baseline hazard (dotted). There are 100 observations in each subgroup and $b = 0.1$.

In the simulation study we use a discrete approximation of the integrated square error (ISE) performance measure

$$\sum_{i=1}^{10} \int_0^1 \left[\widehat{\bar{\gamma}}_i(s) - \theta_i(s)\alpha_q(s)\right]^2 ds, \quad q = 1, ..., 4$$

to compare the performance of the ten empirical credibility weighted estimators with the corresponding performance measure

$$\sum_{i=1}^{10} \int_0^1 \left[\tilde{\gamma}_i(s) - \theta_i(s)\alpha_q(s)\right]^2 ds, \quad q = 1, ..., 4$$

of the ten individual kernel hazard estimators. This approach is used to evaluate the four different baseline hazards with the same credibility structure defined above. We simulated 100 runs from 10 different subgroups with respectively 100 and 1000 observations in each subgroup. The averaged performance error is given in Table 1 below.

TABLE 1

THE AVERAGE OF THE INTEGRATED SQUARE ERROR (ISE) FOR THE INDIVIDUAL ESTIMATORS RESPECTIVELY THE
CREDIBILITY ESTIMATORS AND THE CORRESPONDING RATIO IN 8 DIFFERENT SIMULATION STUDIES

$n_i = 100, k = 10$	α_1	α_2	α_3	α_4	$n_i = 1000, k = 10$	α_1	α_2	α_3	α_4
Individual, $\tilde{\gamma}_i$	116	123	300	502	Individual, $\tilde{\gamma}_i$	12	13	134	133
Credibility, $\widehat{\tilde{\gamma}}_i$	49	51	197	263	Credibility, $\widehat{\tilde{\gamma}}_i$	9	9	129	119
ISE-ratio	0.42	0.42	0.66	0.52	ISE-ratio	0.74	0.73	0.96	0.90

It is interesting to notice that the improvement is consistent over the simulated datasets. The improvement is greatest for α_1 and α_2 which have a fairly simple shape compared with α_3 and α_4. The improvement is greatest for small datasets (small n's) because when the n's become large the credibility estimator and the individual estimator are close to each other as the credibility to the individual estimator grows. The results are also consistent over the individual simulation. For example: for the baseline α_1 and $n_i = 100$, the improvement of the relative error of performance of the credibility approach range from 50% and 70% corresponding to an average improvement of 58%.

ACKNOWLEDGMENTS

APPENDIX

First we note that given $\theta_i(t)$ the theoretical properties of $\tilde{\gamma}_i(t)$ is identical to the properties of a standard local constant kernel hazard estimator, therefore we get that

$$E\{\tilde{\gamma}_i(t)|\theta_i(t)\} = \theta_i(t)\alpha^*(t) = \{1 + o_P(1)\}\theta_i(t)\alpha(t)$$

and

$$VAR\{\tilde{\gamma}_i(t)|\theta_i(t)\} = \{1 + o_P(1)\}C_2\theta_i(t)\alpha(t)\{\tilde{Y}_{i.}(t)\}^{-1}$$

where

$$\alpha^*(t) = \int K_b(t - s)\alpha(s)ds = \{1 + o_P(1)\}\alpha(t).$$

See Hjort (1992) for more details on the local constant estimator. We can therefore make the following calculations,

$$E[\tilde{\gamma}_i(t)] = E[E\{\tilde{\gamma}_i(t)|\theta_i(t)\}] = \{1 + o_P(1)\}\alpha(t)$$

$$\begin{aligned}
COV\{\gamma_i(t), \tilde{\gamma}_i(t)\} &= E[COV\{\gamma_i(t), \tilde{\gamma}_i(t)|\theta_i(t)\}] \\
&\quad + COV[E\{\gamma_i(t)|\theta_i(t)\}, E\{\tilde{\gamma}_i(t)|\theta_i(t)\}] \\
&= COV\{\theta_i(t)\alpha(t), \theta_i(t)\alpha^*(t)\} \\
&= \sigma_t^2 \alpha(t)\alpha^*(t) \\
&= \{1 + o_P(1)\}\sigma_t^2 \alpha^2(t)
\end{aligned}$$

$$\begin{aligned}
VAR\{\tilde{\gamma}_i(t)\} &= E[VAR\{\tilde{\gamma}_i(t)|\theta_i(t)\}] + VAR[E\{\tilde{\gamma}_i(t)|\theta_i(t)\}] \\
&= E\Big[\{1 + o_P(1)\}C_2\theta_i(t)\alpha(t)\{\tilde{Y}_{i.}(t)\}^{-1}\Big] + VAR\{\theta_i(t)\alpha^*(t)\} \\
&= \{1 + o_P(1)\}C_2\alpha(t)\{\tilde{Y}_{i.}(t)\}^{-1} + \sigma_t^2\{\alpha^*(t)\}^2 \\
&= \{1 + o_P(1)\}C_2\alpha(t)\{\tilde{Y}_{i.}(t)\}^{-1} + \{1 + o_P(1)\}\sigma_t^2\alpha^2(t)
\end{aligned}$$

REFERENCES

ANDERSEN, P.K., BORGAN, Ø, GILL, R.D., and KEIDING, N. (1992) *Statistical models based on counting processes.* Springer-Verlag, New-York.

BÜHLMANN, H. and STRAUB, E. (1970) Glaubwürdigkeit für Schadensätze. *Bulletin of the Association of Swiss Actuaries* **70**, 111-133.

HARDY, M.R. and PANJER, H.H. (1998) A credibility approach to mortality risk. *ASTIN Bulletin* **28**, 269-283.

HJORT, N.L. (1990) Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18**, 1259-1294.

HJORT, N.L. (1992) Semiparametric estimation of parametric hazard rates. In KLEIN, J.P. and GOEL, P.K., eds., *Survival Analysis: State of the Art*, pp. 211-236, Kluwer, Dordrecht.

RAMLAU-HANSEN, H. (1983) Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453-466.

YOUNG, V.R. (1997) Credibility using semiparametric models. *ASTIN Bulletin* **27**, 273-285.

YOUNG, V.R. (1998) Robust Baysian credibility using semiparametric models. *ASTIN Bulletin* **28**, 187-203.

JENS PERCH NIELSEN AND BJØRN LUNDING SANDQVIST
Codan
Gammel Kongevej 60
1790 København V
Denmark
E-mail: npj@codan.dk & san@codan.dk