PROPORTIONAL HAZARD ESTIMATION ADJUSTED BY CONTINUOUS CREDIBILITY

BY

JENS PERCH NIELSEN AND BJØRN LUNDING SANDQVIST

ABSTRACT

This paper extends the continuous credibility weighting introduced to hazard estimation in Hardy and Panjer (1998) and Nielsen and Sandqvist (2000) to the more general case, where the common basis is a proportional hazard model.

KEYWORDS

Counting process theory; Kernel hazard estimation; Continuous credibility; Bühlmann-Straub model; Proportional hazard models.

1. INTRODUCTION

Inspired by the credibility approach to hazard estimation of Hardy and Panjer (1998), Nielsen and Sandqvist (2000) considered hazards of different groups assuming the hazard of each group to fluctuate across a common baseline hazard. They modelled this fluctuation by a heterogeneity parameter capturing the particular properties of each group allowing for a surprisingly simple credibility estimation procedure. The new estimation procedure can extend nonparametric smoothing techniques to a number of data sets from the insurance industry or other places including, for example, credit risk, where transitions from one credit risk group to another are modelled. Nielsen and Sandqvist (2000) considered hazards of the i'th group given by

$$\theta_i(t)\alpha(t),$$

where $\alpha(t)$ is a smooth baseline intensity and $\theta_i(t)$ is a stochastic risk process with $E\{\theta_i(t)\} = 1$. This model was estimated by a combination of nonparametric estimation (in the t-dimension) and credibility (between groups: e.g. the i-dimension) where information from one group is guiding estimation of hazards of other groups.

In this paper we consider the situation, where the proportional hazard model takes over in case of data sparsity instead of a common baseline hazard

as in Nielsen and Sandqvist (2000). In groups with a lot data, the proportionality assumption of the hazards might be too restrictive compared to the individual information we actually do have in such a group. In groups where there is enough data to model the hazard by nonparametric smoothing techniques, one should not force some proportional hazard assumption upon the data. For groups with too little data to model the entire hazard, there can still be areas with high levels of information where the individual group hazard can indeed be modelled by standard smoothing techniques. The method of the current paper captures all these situations. The credibility weighting automatically adjust the fragile nonparametric estimator in areas of too little information and leave it alone in areas were there is sufficient information. Our method is automatic with the traditional advantage of the credibility approach: We always have an estimator to use.

Mathematically our model introduces a group-specific proportionality factor D_i resulting in the following model for the hazard in group i:

$$\gamma_i(t) = D_i \theta_i(t) \alpha(t),$$

where $\alpha(t)$ is a smooth baseline intensity and $E\{\theta_i(t)\} = 1$.

We show that the proportionality factors, D_i's, can be estimated through a simple integral of some nonparametric estimator resulting in square-root-n consistent estimators. This is not surprising and follow standard semiparametric analysis, see Bickel et al. (1993). If our unobserved heterogeneity parameter $\theta_i(t)$ did not depend on time, then our model would be a traditional frailty model. Frailty models of survival data were extensively treated in Hougaard(2000). In chapter 11.6, Hougaard points out that "it would be desirable to have a model where the frailty varies smoothly with time" and he goes on and cite Woodbury and Manton(1977) and Aalen(1994) for having considered a complicated parametric diffusion process as the underlying structure of the time dependent frailty. Our credibility approach to a closely related problem gives a direct, intuitive and nonparametric estimator of such a "smoothly varying frailty parameter" as Hougaard was looking for.

In this paper we use nonparametric smoothing to estimate different hazards, e.g. the α. Although nonparametric smoothing techniques have developed a lot since the important paper of Ramlau-Hansen (1983), we do not try to use the most recent or most sophisticated smoothing techniques. We choose to consider the simple local constant hazard estimator introduced in Hjort (1992) and reconsidered in Nielsen and Tanggaard (2001). However, modern smoothing techniques such as the bias correction methods of Nielsen (1998) and Nielsen and Tanggaard (2001) readily lend themselves to our credibility approach.

In §2 we outline the counting process formulation for the hazard model. The example which will be considered throughout the paper in introduced in §3. §4 estimate the proportionality factors and the baseline intensity. We define the proportionality adjusted continuous credibility estimator in §5 and in §6 we estimate the variance. In §7 we get the empirical credibility estimator and we conclude in §8.

2. A COUNTING PROCESS FORMULATION OF THE MODEL

We observe m individuals from k different groups and use the indexes $i = 1, .., k$ and $j = 1, .., n_i$ for respectively the k groups and the n_i individuals in the i'th group. Later, in the asymptotic considerations we need $n = min(n_{1,...,}n_k)$. Let the counting process N_{ij} count observed failures for the j'th individual in the i'th group in the time interval $[0,T]$. We assume that N_{ij} takes values in $\{0,1\}$ and 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

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

where

$$\gamma_i(t) = D_i\beta_i(t) \text{ and } \beta_i(t) = \theta_i(t)\alpha(t).$$

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 group is under risk. The positive stochastic process θ_i is the continuous risk parameter for the i'th group and $\alpha(\cdot)$ is a differentiable and deterministic unknown baseline hazard with no restriction on the functional form apart from the smoothness assumption. $D_1, ..., D_k$ are parameters playing the same type of role as parameters in traditional problems.

We assume that the stochastic processes $\theta_1, ..., \theta_k$ are independent identically distributed. Given these unobservable stochastic processes, the counting processes $(N_{11}, Y_{11}), ..., (N_{kn_k}, Y_{kn_k})$ are independent for the m individuals and $(N_{i1}, Y_{i1}), ..., (N_{in_i}, Y_{in_i})$ are identically distributed for each i.

We assume that

$$E\theta_i(t) = 1 \qquad (1)$$

and that $\theta_i(t)$ and $\{Y_{ij}(t)\}_{j \in \{1,...,n_i\}}$ are independent for i in $\{1, ..., k\}$.

We also assume that the baseline hazard α and the stochastic processes $\theta_1, ..., \theta_k$ are normalized such that

$$\int_0^T \theta_i(t)\alpha(t)w(t)dt = 1 \text{ (for all sample paths and all } i) \qquad (2)$$

for some positive weighting function w, where $\int_0^T w(t)dt$ is a finite integral. Furthermore we assume that $Var\,\theta_i(t) = \sigma_t^2$ and $E\{Y_{ij}(t)\} = \delta_i(t)$ for $t \in [0,T]$ for positive and continuous δ_i's. The kernel K is a probability density function symmetric about zero and $K_b(\cdot) = b^{-1}K(\cdot/b)$ for any positive bandwidth.

The normalization assumptions (1) and (2) identify the model and they imply the integral equation $\int_0^T \alpha(t)w(t)dt = 1$.

Define also

$$\eta_i(t) = D_i\theta_i(t).$$

Finally, the group exposure,

$$\tilde{Y}_{i.}(t,b) = \sum_{j=1}^{n_i} \tilde{Y}_{ij}(t,b),$$

is defined as the aggregate of the individual smoothed exposure processes:

$$\tilde{Y}_{ij}(t,b) = \int_0^T K_b(t-s)Y_{ij}(s)\,ds.$$

3. An example. An application to disability insurance

We introduce a real data example to illustrate the results. The analysis of the data example will be updated along with our development of the necessary theory. The data example considered is taken from Nielsen and Sandqvist (2000) and reconsidered in the light of the theoretical developments of this paper. In that paper, we considered four groups divided after sex and standard/ substandard tables and estimated the disability intensities in the four groups with and without the credibility approach of that paper. The standard non-parametric kernel hazard estimator of each group separately, see Hjort (1992), is shown in Figure 1.

For high ages – above 50 years – these four intensities are too volatile to be useful predictors. Nielsen and Sandqvist (2000) alleviated this problem through a credibility weighting towards a common baseline hazard, see Figure 2.

For high ages the credibility weighted estimators of Figure 2 are less volatile than the individual nonparametric estimators of Figure 1. They can therefore be expected to be better predictors of the underlying disability intensities. However, we are left with a feeling that we can still do better. While the individual estimators of Figure 1 clearly are too volatile at high ages, some information of the credibility weighted estimators of Figure 2 seems to have been lost by downweighting the individual hazards towards the same baseline hazard. While the individual hazards of Figure 1 clearly are too volatile, they do seem to contain enough information to allow the levels of the disability intensities to be estimated. This observation was in fact the motivation of this paper, where we reformulate the basic model and the credibility approach, such that the stabilizing effect of credibility is maintained, while the available knowledge of intensity levels are used to our advantage. We now continue developing the approach based on credibility and the proportional hazard model. We return to the data example in Section 4.2, Section 5, Section 6 and Section 7.

4. Estimating the proportionality constants and the underlying hazard

In this section we construct estimators of the proportionality constants $D_1,\ldots,$ D_k, and the underlying hazard α of the model defined in Section 2. We establish
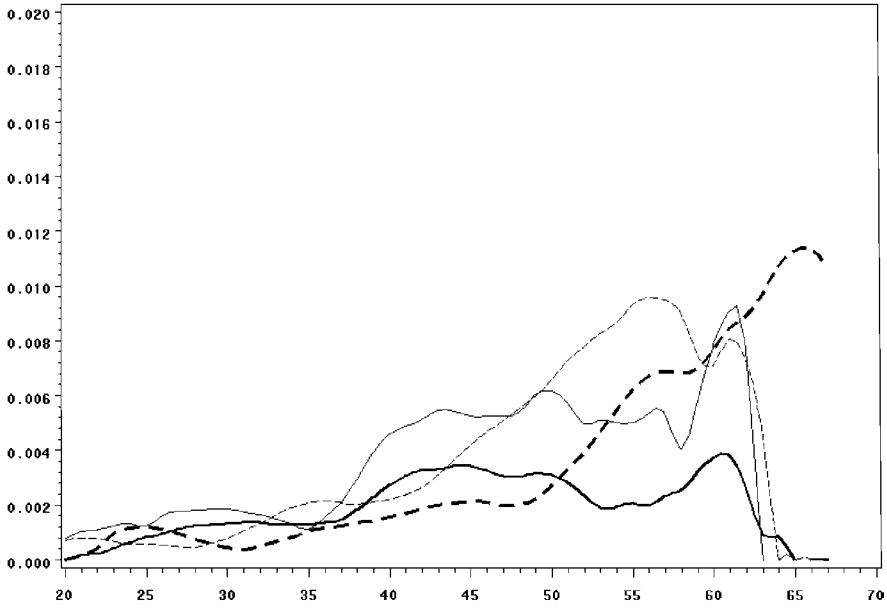
FIGURE 1. The four estimated individual hazards using standard non-parametric estimation methods.
(standard male lives: thick dotted line, standard female lives: thick solid line,
substandard male lives: thin dotted line, substandard female lives: thin solid line)
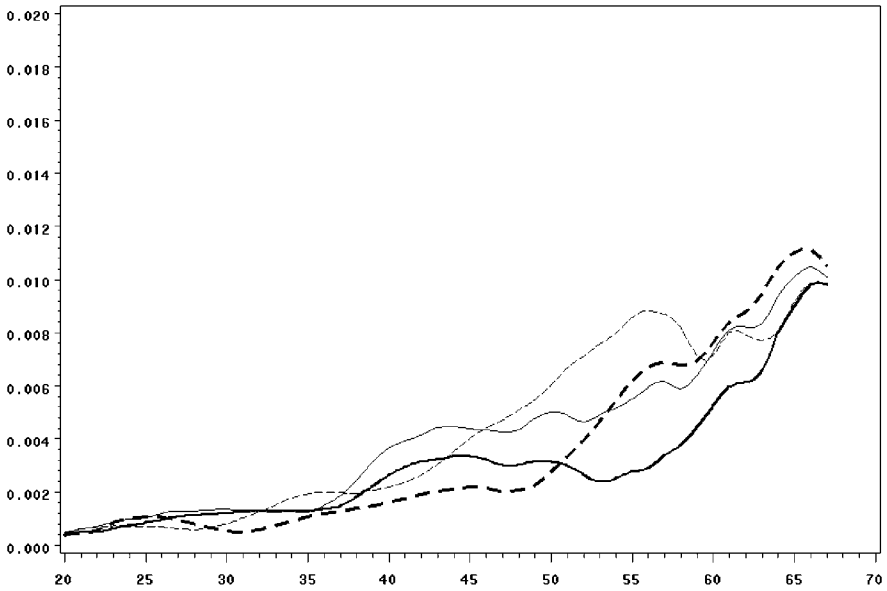

FIGURE 2. The four estimated credibility hazards from Nielsen and Sandqvist (2000). With correction note.
(standard male lives: thick dotted line, standard female lives: thick solid line,
substandard male lives: thin dotted line, substandard female lives: thin solid line)

a simple estimator of the proportionality constants based on integrating out a relevant nonparametric kernel estimator. If this kernel estimator is under-smoothed, then the resulting estimator of the proportionality constants are square-root-n consistent just like parameters tend to be in ordinary parametric estimation problems. We estimate the underlying hazard α by a standard kernel smoother adjusting appropriately for the observed differences in level. Since kernel smoothers have an inferior rate of convergence compared to the parametric rate, the uncertainty of estimating the proportionality constants are irrelevant for the asymptotic performance when the estimated underlying hazard or the resulting credibility weighted kernel hazard are considered.

4.1. Estimation of D_i and α

It follows from the normalization assumption (2) that the proportionality constant can be written as

$$D_i = \int_0^T \gamma_i(t) w(t) dt$$

and estimated simply by integrating out an estimator of γ_i. This is

$$\hat{D}_i = \int_0^T \bar{\gamma}_i(t,h) w(t) dt,$$

where $\bar{\gamma}_i$ is the local constant kernel hazard estimator

$$\bar{\gamma}_i(t,h) = \frac{\sum_{j=1}^{n_i} \int_0^T K_h(t-s) \, dN_{ij}(s)}{\tilde{Y}_{i.}(t,h)},$$

with bandwidth h, see Hjort (1992) or Nielsen and Tanggaard (2001).

Under standard conditions, $\hat{D}_1, \ldots, \hat{D}_k$ are square-root-n consistent estimators of the parameters $D_1, \ldots, D_k$, where $n = min(n_1, \ldots, n_k)$ and $n_i/n \to f_i$, where f_i is positive and the bandwidth h is undersmoothed, see Appendix A.

To estimate α we use bandwidth b – bigger than h – and a modification of the local constant kernel hazard estimator adjusted for $\hat{D}_1, \ldots, \hat{D}_k$:

$$\tilde{\alpha}(t) = \frac{\sum_{i=1}^k \left(\{\hat{D}_i\}^{-1} \sum_{j=1}^{n_i} \int_0^T K_b(t-s) \, dN_{ij}(s) \right)}{\sum_{i=1}^k \tilde{Y}_{i.}(t,b)}.$$

The square-root-n consistency of the estimated parameters implies that the asymptotic properties of $\tilde{\alpha}(t)$ equal the asymptotic properties of

$$\bar{\alpha}(t) = \frac{\sum_{i=1}^k \left(\{D_i\}^{-1} \sum_{j=1}^{n_i} \int_0^T K_b(t-s) \, dN_{ij}(s) \right)}{\sum_{i=1}^k \tilde{Y}_{i.}(t,b)}$$

that has the exact same asymptotic behavior as the estimator of the underlying hazard had in Nielsen and Sandqvist (2000). In Appendix B we justify that $\bar{\alpha}$ is an estimator of α.

We know that $\int_0^T \alpha(t)w(t)dt = 1$, we therefore adjust $\tilde{\alpha}(t)$ to

$$\hat{\alpha}(t) = \frac{\tilde{\alpha}(t)}{\int_0^T \tilde{\alpha}(s)\,w(s)\,ds}.$$

We now estimate $\eta_i(t) = D_i\theta_i(t)$ by locally correcting the counting processes N_{ij}, with stochastic intensities

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

for $\hat{\alpha}(t)Y_{ij}(t)$ resulting in the estimator

$$\bar{\eta}_i(t) = \frac{\sum_{j=1}^{n_i}\int_0^T K_b(t-s)\,dN_{ij}(s)}{\sum_{j=1}^{n_i}\int_0^T K_b(t-s)\,\hat{\alpha}(s)\,Y_{ij}(s)\,ds},$$

see Appendix C.

We know such multiplicative corrections from the literature of bias corrections, see Jones, Linton and Nielsen (1995) and Nielsen (1998) for respectively the density and the hazard case. Nielsen and Tanggaard (2001) showed that this multiplicative correction principle can be understood as a minimization criteria using the least squares criterion introduced in Nielsen (1998). For a motivation of the estimator $\bar{\eta}_i(t)$ we note that the nominator approximates $D_i\theta_i(t)\alpha(t)Y_{i.}(t)$ while the denominator approximates $\alpha(t)Y_{i.}(t)$.

Due to (2) we know that

$$D_i = \int_0^T D_i\theta_i(s)\alpha(s)w(s)ds = \int_0^T \eta_i(s)\alpha(s)w(s)ds.$$

We therefore adjust the estimator of η_i to

$$\tilde{\eta}_i(t) = \hat{D}_i\bar{\eta}_i(t) / \int_0^T \bar{\eta}_i(s)\hat{\alpha}(s)w(s)ds.$$

From the point of view of theoretical analyzes, we can think of $\hat{\alpha}$ as an accurate estimator of α, since it is based on information from all groups. This is a well known trick from credibility theory, where the global mean often is considered known, since it is based on all groups and therefore is estimated quite precisely. In Appendix D, we therefore replace $\hat{\alpha}$ by α, while considering asymptotic properties of our estimators.

4.2. The example continued

We now return to our example of disability hazards. We use the bandwidth $h = 3$ years (undersmoothing) while estimating the D_i's and elsewhere we use the bandwidth $b = 5$ years. The kernel function used is

$$K(t) = \cos\left(\frac{\pi}{2}t\right)I\{t \in (-1,1)\}.$$

We choose our weight function to be

$$w(t) = \sum_{i=1}^{4}Y_{i.}(t)\,I(t \in [20,67])\,/\left(\frac{1}{67-20}\int_{20}^{67}\sum_{i=1}^{4}Y_{i.}(s)\,ds\right)$$

which for the age interval [20,67] is the total exposure per age normalized to an average level of 1. There are other possibilities for the choice of weight function. One could for example let w be a constant on some interval of preference and zero outside this interval. However, in the current application, we have chosen to put more weight in areas of high exposure.

In Figure 3 we illustrate the smoothed exposure $\tilde{Y}_{i.}(t,b)$ for the different groups.

Our estimation procedure leads to the following estimators of the proportionality constants

$\hat{D}_1$	$\hat{D}_2$	$\hat{D}_3$	$\hat{D}_4$
0,101	0,172	0,095	0,160

Based on these estimators we have the following preliminary conclusions: Normally insured women (group 1) have a risk of disability which is about 6% higher than normally insured men (group 3). Substandard insured men (group 4) have a risk of disability which is about 68% higher than normally insured men while substandard insured women (group 2) have a risk of disability which is about 70% higher than normally insured women. However, the above percentages are based on the proportionality assumption. We will see that fluctuations around this proportionality assumption seem to be present in the data.

Based on $\hat{D}_1,...,\hat{D}_4$, we construct non-parametric estimators, $\hat{\alpha}$ of α, and $\tilde{\eta}_i(t)$ of $\eta_i(t) = D_i\theta_i(t)$. The above estimators lead to the individual estimators

$$\tilde{\gamma}_i(t) = \tilde{\eta}_i(t)\hat{\alpha}(t)$$

which are the starting point of our credibility estimator presented below. The credibility estimator is a locally weighted average of the individual estimators $\tilde{\gamma}_i(t)$ and the proportional hazard estimators $\hat{D}_i\hat{\alpha}(t)$ presented in Figure 4.

Note that, compared to Figure 1, the picture in Figure 4 is radically changed under the proportional assumption. The proportional hazards for all groups – except for the normally insured men – are higher for older ages than the group specific hazards in Figure 1.
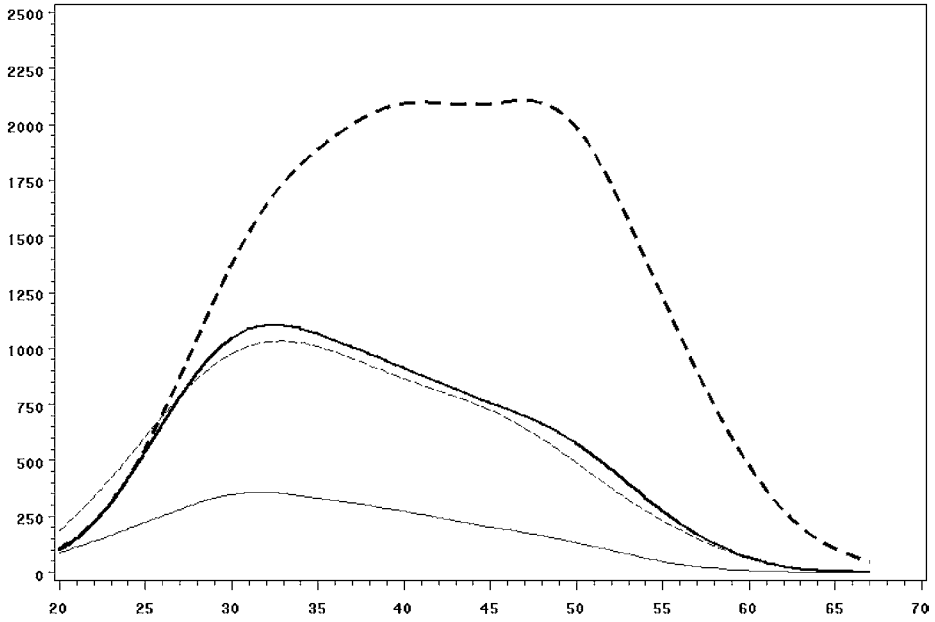
FIGURE 3. The smoothed exposure in the four groups (standard male lives: thick dotted line, standard female lives: thick solid line, substandard male lives: thin dotted line, substandard female lives: thin solid line).
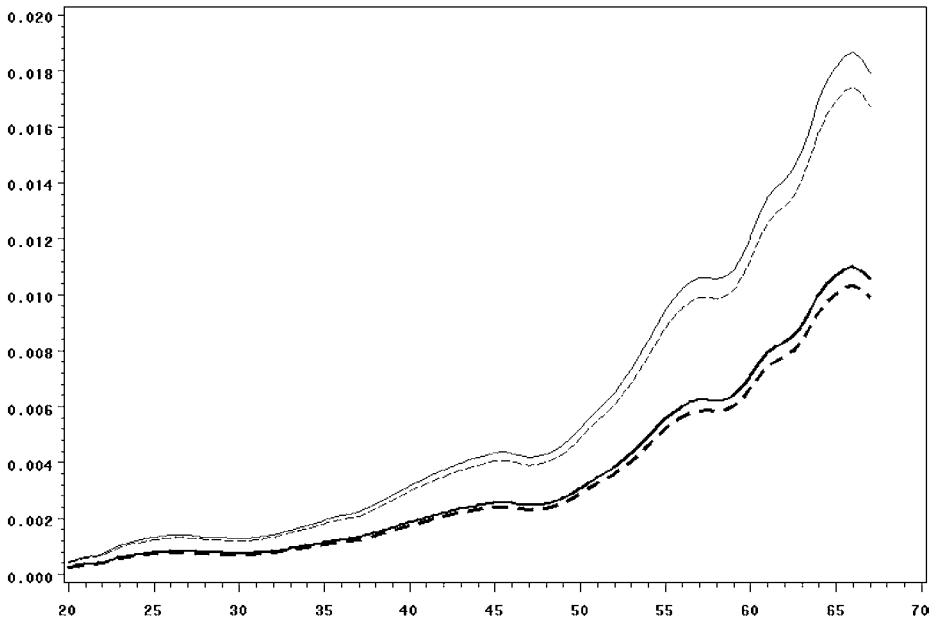

FIGURE 4. The proportional hazards $\hat{D}_i\hat{\alpha}(t)$ (standard male lives: thick dotted line, standard female lives: thick solid line, substandard male lives: thin dotted line, substandard female lives: thin solid line).

5. The proportionality adjusted credibility estimator
of $\theta_1, \ldots, \theta_k$.

To indicate how much the individual estimator $\tilde{\gamma}_i(t)$ differs from the proportional estimator $\hat{D}_i\hat{\alpha}(t)$ we look at the behavior of $\{\hat{D}_i\}^{-1}\tilde{\eta}_i(t)$ which is a raw non-credibility estimator of $\theta_i(t)$. This is shown in Figure 5.

We see that the graphs fluctuate considerable, in particular in areas of low exposure. Clearly a stabilizing estimation procedure based on credibility weights seems appropriate after a first glance at Figure 5.

We start by adjusting our problem such that it becomes comparable to classical credibility theory and the methodology of Nielsen and Sandqvist (2000). Consider the Hilbert space projection of the stochastic variable

$$\beta_i(t) = \theta_i(t)\alpha(t)$$

down at the linear space spanned by a constant and our estimated candidate

$$\tilde{\beta}_i(t) = \{\hat{D}_i\}^{-1}\tilde{\eta}_i(t)\hat{\alpha}(t).$$

This parallels the original approach of Bühlmann and Straub (1970) that also has a Hilbert space interpretation, see Norberg(2004).

Consider now the Hilbert space projection of $\beta_i(t)$ onto the linear space

$$\{a_t + b_t\tilde{\beta}_i(t) \,|\, a_t, b_t \in R\}$$

for a fixed t. Based on this Hilbert space projection we obtain the following expression for the optimal linear credibility estimator minimizing

$$E\{\beta_i(t) - a_t - b_t\tilde{\beta}_i(t)\}^2,$$

namely

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

where

$$z_{i,t}^* = \frac{COV\{\beta_i(t), \tilde{\beta}_i(t)\}}{VAR\{\tilde{\beta}_i(t)\}}.$$

From Appendix D we get the following three results

$$E\{\tilde{\beta}_i(t)\} = \{1 + o(1)\}\alpha(t),$$

$$COV\{\beta_i(t), \tilde{\beta}_i(t)\} = \{1 + o(1)\}\sigma_t^2\alpha^2(t)$$

and

$$VAR\{\tilde{\beta}_i(t)\} = \{1 + o(1)\}\left[D_i^{-1}C_2\alpha(t)\{b\tilde{Y}_{i.}(t,b)\}^{-1} + \sigma_t^2\alpha^2(t)\right],$$
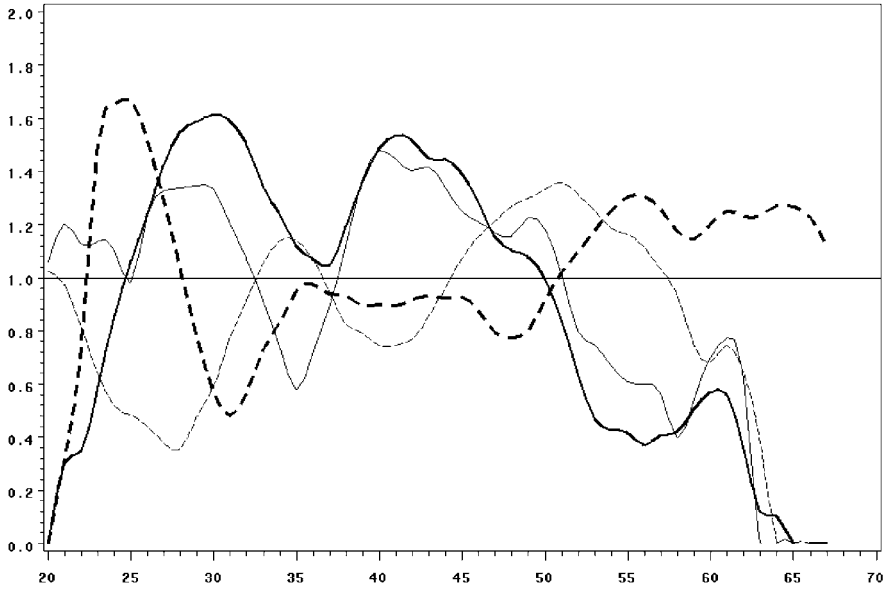
FIGURE 5. The raw non-credibility estimator of $\theta_i(t)$: $\{\hat{D}_i\}^{-1}\tilde{\eta}_i(t)$ (standard male lives: thick dotted line, standard female lives: thick solid line, substandard male lives: thin dotted line, substandard female lives: thin solid line).

where $C_2 = \int K^2(u)\,du$ and where $o(1)$ 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 of $\beta_i(t)$ is approximately equal to

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

where

$$z_{i,t} = \frac{D_i \sigma_t^2 \alpha(t) b \tilde{Y}_{i.}(t, b)}{C_2 + D_i \sigma_t^2 \alpha(t) b \tilde{Y}_{i.}(t, b)}. \tag{3}$$

We get a preliminary estimator of the underlying credibility parameter $\theta_i(t)$ by dividing the above credibility estimator by $\alpha(t)$:

$$\bar{\theta}_i(t) = (1 - z_{i,t}) + z_{i,t}\tilde{\beta}_i(t)\{\alpha(t)\}^{-1}.$$

If $D_i = 1$ for all i, then the credibility weighting $z_{i,t}$ is equal to credibility weighting Nielsen and Sandqvist (2000) arrived at, see also the correction note to that paper. Note that $z_{i,t}$ is increasing as a function of $D_i\alpha(t)b\tilde{Y}_{i.}(t,b)$. This latter quantity is asymptotically proportional to the expected number of disability cases in the time interval $[t - b, t + b]$. It is reasonable that the weight $z_{i,t}$ for the individual estimator will increase when this quantity increases. It is also reasonable

that $z_{i,t}$ is increasing in σ_t^2. All required parameters have been estimated at this point except the variance σ_t^2 that we estimate in the next section.

6. Estimating the variance

While estimating the credibility weights, we need an estimator of the variance. A natural choice is

$$\tilde{\sigma}_t^2 = (k-1)^{-1} \sum_{i=1}^{k} \left[\tilde{\eta}_i(t) \{\hat{D}_i\}^{-1} - 1 \right]^2$$

where we have employed that $E\theta_i(t) = 1$ and that $\tilde{\eta}_i(t)$ is an estimator of $\eta_i(t) = D_i \theta_i(t)$ and therefore $\tilde{\eta}_i(t)\{\hat{D}_i\}^{-1}$ (illustrated in Figure 5) is a raw estimator of $\theta_i(t)$.

We expect the estimated variance in the model presented here to be considerable smaller than the estimated variance in Nielsen and Sandqvist (2000) will be for the same data set. This is due to the fact that a lot of the variation of the method considered in Nielsen and Sandqvist (2000) will be caught by the parameters $D_1, \ldots, D_k$.

We set $\sigma_t^2 = \sigma^2$, since σ_t^2 can not be estimated with sufficient accuracy when we only have four different groups in our study. Without this simplifying assumption the credibility estimators would depend on the wiggly estimator of σ_t^2. A natural choice of estimator in the simplifying case is the weighted average:

$$\tilde{\sigma}^2 = \left(\int_0^T \tilde{Y}_{..}(t,b)\,dt \right)^{-1} \int_0^T \tilde{\sigma}_t^2\, \tilde{Y}_{..}(t,b)\,dt$$

with $\tilde{Y}_{..}(t,b) = \sum_{i=1}^{k} \tilde{Y}_{i.}(t,b)$.

Let us consider the variance estimator in our example. In Figure 6 we see both the time varying variance parameter $\tilde{\sigma}_t^2$ and the constant variance estimator $\tilde{\sigma}^2 = 0{,}144$.

We can also see from Figure 6 that the estimator of the time varying variance fluctuates a lot, in particular in areas with low exposure levels. In the following we continue to use the notation $\tilde{\sigma}_t^2$ for the variance estimator.

7. The final estimator

Now we have all the components to obtain the final estimators of the credibility weights $z_{i,t}$:

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

Then we get

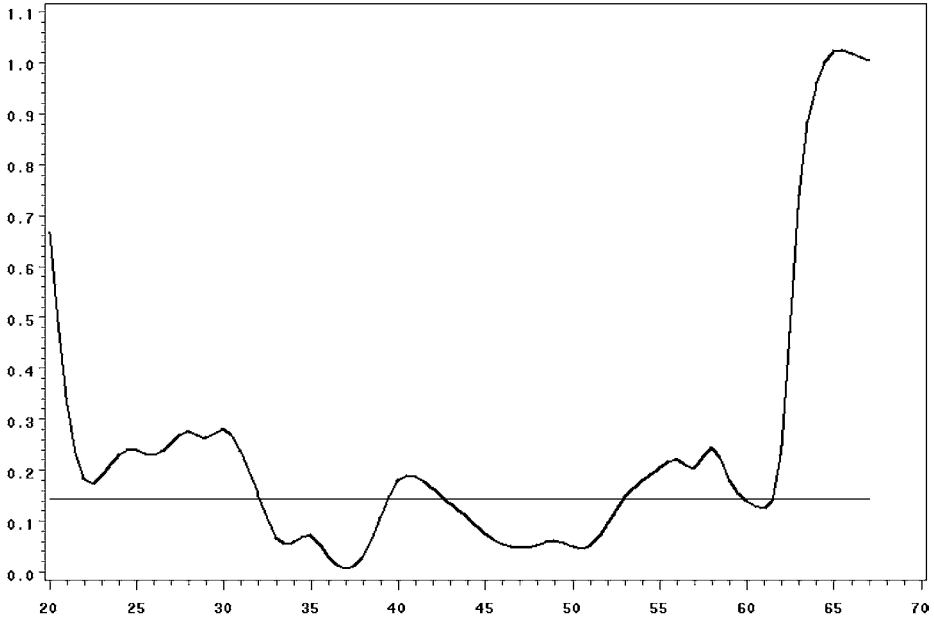$$\tilde{\theta}_i(t) = (1 - \tilde{z}_{i,t}) + \tilde{z}_{i,t} \tilde{\beta}_i(t)\{\hat{\alpha}(t)\}^{-1}$$

FIGURE 6. The time dependent variance estimator $\tilde{\sigma}_t^2$ and the constant variance estimator $\tilde{\sigma}^2 = 0,144$.
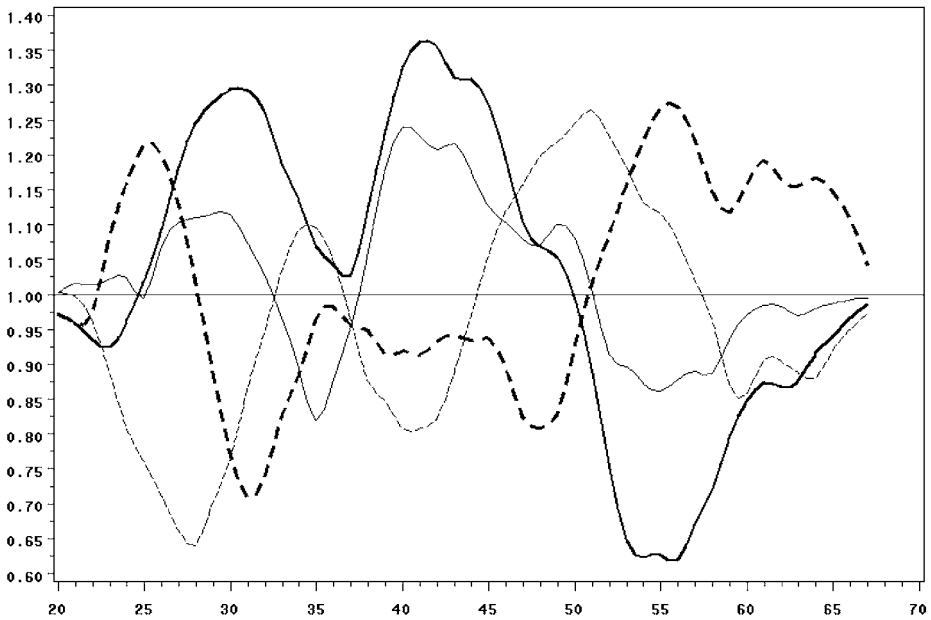

FIGURE 7. The credibility estimator $\hat{\theta}_i(t)$ (standard male lives: thick dotted line, standard female lives: thick solid line, substandard male lives: thin dotted line, substandard female lives: thin solid line).

and normalize this due to (2) to the final estimator of $\theta_i(t)$:

$$\hat{\theta}_i(t) = \tilde{\theta}_i(t) / \int_0^T \tilde{\theta}_i(s)\hat{\alpha}(s)w(s)ds.$$

The final estimator of $\theta_i(t)$ is shown in Figure 7.

It is seen that this final credibility estimator fluctuates less than the raw preliminary estimator of $\theta_i(t)$ given in Figure 5. Our conclusion is that the credibility based estimator of Figure 7 is much more plausible than the raw preliminary estimator given in Figure 5. The estimator $\hat{\theta}_i(t)$ indicates how far the credibility estimator is from the proportional hazard model. In nearly the entire age interval we see that the proportional assumption holds within ±30%. While this fluctuation is within reasonable limits, it does, however, indicate important deviations from the proportional hazard model.

The above leads to the final credibility estimator of the hazard $\gamma_i(t) = D_i\theta_i(t)\alpha(t)$:

$$\hat{\gamma}_i(t) = \hat{D}_i\hat{\theta}_i(t)\hat{\alpha}(t).$$

In areas of data sparsity $\hat{\gamma}_i(t)$ is close to the estimator one would have obtained from the proportional hazard model $D_i\alpha(t)$ and in areas with a lot of data, $\hat{\gamma}_i(t)$ is close to the individual estimator $\bar{\gamma}_i(t,b)$ of the i'th group.

The final credibility hazards in our example are shown in Figure 8.

We can see that the proportional assumption still has influence on the estimators, but there are also important deviations from this such as indicated in Figure 7. For example, the intensity for normally insured women (thick solid line) is above the intensity for normally insured men (thick dotted line) for ages below 50 years while the opposite is true for ages above 50 years. This can be compared to the earlier conclusion from the proportional hazard model, Figure 4, that the overall disability risk for normally insured women is about 6% above normally insured men. Therefore, the credibility approach of this paper, Figure 8, captures effects in data which can not be seen under the proportional assumption in Figure 4. Note also that the credibility based estimators of Nielsen and Sandqvist (2000) given in Figure 2 give a rather different impression than the estimator resulting from the proportional hazard approach of this paper presented in Figure 8, especially for substandard lives above 50 years.

One can conclude that the model assumptions play an important role. In areas with low exposure we essentially arrive at the estimator $\hat{D}_i\hat{\alpha}(t)$ in this paper and $\hat{\alpha}(t)$ in Nielsen and Sandqvist (2000). While Figure 8 and Figure 2 therefore are quite different, they are clearly both much better at ages above 50 years than the original nonparametric estimators of Figure 1. In areas with relative high exposure – e.g. below age 50 – Figure 1, Figure 2 and Figure 8 are quite close to each other.

8. CONCLUSION

In this paper we have developed a continuous credibility adjustment to the proportional hazard model. This method is applicable when the proportional
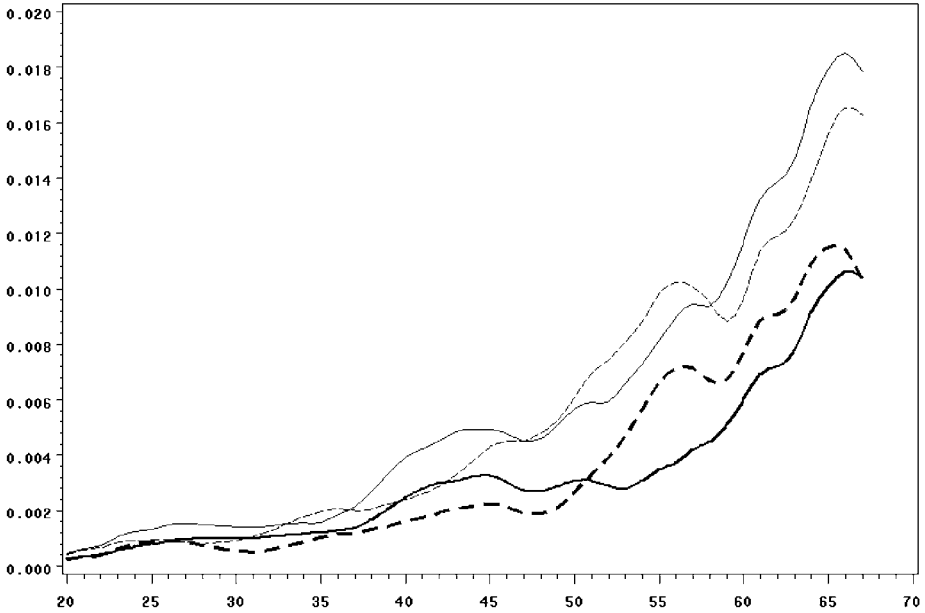
FIGURE 8. The final credibility hazard $\hat{\gamma}_i(t)$ in the proportional model $D_i\theta_i(t)\alpha(t)$, where $\hat{D}_1 = 0{,}101$, $\hat{D}_2 = 0{,}172$, $\hat{D}_3 = 0{,}095$ and $\hat{D}_4 = 0{,}160$. (standard male lives: thick dotted line, standard female lives: thick solid line, substandard male lives: thin dotted line, substandard female lives: thin solid line)

hazard model seems more appropriate as a starting point than the simple model in Nielsen and Sandqvist (2000) assuming that all groups are identical. However, sometimes the proportional hazard model is not better as starting point than the simple model assuming identical groups. A test could clarify which of the two starting points that are closest to the data. However, such tests have a tendency to put most weight in areas with most data. In the end, it is really up to the individual actuary to determine which starting point to use. Both methods adjust automatically to the true underlying structure when the amount of data increases.

APPENDIX A

In this appendix, we shortly comment on the fact that the estimated proportionality factors are square-root-n consistent. It is a well known fact from

semiparametric analyses that undersmoothing of the nonparametric compo-
nent is often necessary to obtain square-root-n consistency and efficiency of
the parametric component. For a review of semiparametric analyses, see Bickel,
Klaassen, Ritov and Wellner (1993) and for a semiparametric hazard model
within the counting process framework of this paper, see Nielsen, Linton and
Bickel (1998).

Remember that the proportionality constant

$$D_i = \int_0^T \gamma_i(t)w(t)dt$$

is estimated by integrating out the smooth estimator

$$\bar{\gamma}_i(t,h) = \frac{\sum_{j=1}^{n_i} \int_0^T K_h(t-s)\,dN_{ij}(s)}{\sum_{i=1}^k \tilde{Y}_{i\cdot}(t,h)},$$

such that

$$\hat{D}_i = \int_0^T \bar{\gamma}_i(t,h)w(t)dt.$$

It follows from standard kernel smoothing technique, see Nielsen and Tang-
gaard (2001), that the estimation error $\bar{\gamma}_i(t,h) - \gamma_i(t)$ can be written as a sum of
a variable part $V_i(t)$ and a stable part $B_i(t)$ resulting in the following expression
of the estimation error of the proportionality constant:

$$\hat{D}_i - D_i = \bar{V}_i + \bar{B}_i,$$

where

$$\bar{V}_i = \int_0^T V_i(t)w(t)dt$$

and

$$\bar{B}_i = \int_0^T B_i(t)w(t)dt.$$

The point is that while the variable part $V_i(t)$ has rate of convergence of the
order of magnitude $O_P\{(nh)^{-1}\}$ then $\bar{V}_i$ integrates out the smoothing effect
resulting in a square-root-n consistent quantity. The stable part $B_i(t)$ is of the
order of magnitude $O_P(h^2)$. Integrating does not change the order of magnitude
and $\bar{B}_i = O_P(h^2)$ as well. Therefore, if an undersmoothed bandwidth $h = o_P(n^{-1/4})$
is chosen, then $\bar{B}_i = o_P(n^{-1/2})$ and

$$\hat{D}_i - D_i = \bar{V}_i + \bar{B}_i,$$

is square-root-n consistent, since $\bar{V}_i$ is square-root-n consistent. The exact same
arguments are present in Nielsen, Linton and Bickel (1998) and is omnipresent

in the semiparametric literature when considering the parametric part of the semiparametric problem. We omit further details for the sake of clarity.

Appendix B

We pointed out in Section 4.1 that $\bar{\alpha}$ and $\tilde{\alpha}$ are equivalent from an asymptotic point of view. Here we verify that $\bar{\alpha}$ estimates α. Note that the difference of

$$\bar{\alpha}(t) = \frac{\sum_{i=1}^{k}\left(\{D_i\}^{-1}\sum_{j=1}^{n_i}\int_0^T K_b(t-s)\,dN_{ij}(s)\right)}{\sum_{i=1}^{k}\tilde{Y}_{i\cdot}(t,b)}$$

and

$$\bar{\alpha}^*(t) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\int_0^T K_b(t-s)\,\theta_i(s)\,\alpha(s)\,Y_{ij}(s)\,ds}{\sum_{i=1}^{k}\tilde{Y}_{i\cdot}(t,b)}$$

is equal to

$$\bar{\alpha}(t) - \bar{\alpha}^*(t) = \frac{\sum_{i=1}^{k}\left(\{D_i\}^{-1}\sum_{j=1}^{n_i}\int_0^T K_b(t-s)\,dM_{ij}(s)\right)}{\sum_{i=1}^{k}\tilde{Y}_{i\cdot}(t,b)}$$

that can be analyzed by standard martingale techniques, see Ramlau-Hansen (1983) or Nielsen and Tanggaard (2001), as $M_{ij}(s) = N_{ij}(s) - D_i\theta_i(s)\alpha(s)Y_{ij}(s)$ is a martingale. This will lead to that

$$(nb)^{-1/2}\{\bar{\alpha}(t) - \bar{\alpha}^*(t)\}$$

converges to a normal distribution.

Note furthermore that $\bar{\alpha}^*(t)$ is close to

$$\alpha(t)\frac{\sum_{i=1}^{k}\tilde{Y}_{i\cdot}(t,b)\,\theta_i(t)}{\sum_{i=1}^{k}\tilde{Y}_{i\cdot}(t,b)}$$

for small b. The term

$$\frac{\sum_{i=1}^{k}\tilde{Y}_{i\cdot}(t,b)\,\theta_i(t)}{\sum_{i=1}^{k}\tilde{Y}_{i\cdot}(t,b)}$$

converges to the value one for k going to infinity, since $E\{\theta_i(t)\} = 1$. Therefore $\bar{\alpha}^*(t)$ estimates $\alpha(t)$.

APPENDIX C

The exact procedure for estimation of $\bar{\eta}_i(t)$ in Section 4.1 can be written as

$$\underset{\eta}{\arg\min} \lim_{q \to 0}$$

$$\sum_{j=1}^{n_i} \int_0^T \left(\frac{1}{q} \int_{s-q}^s dN_{ij}(u) - \eta\hat{\alpha}(s) \right)^2 K_b(t-s)\{\hat{\alpha}(s)\}^{-1} Y_{ij}(s)\, ds$$

$$= \underset{\eta}{\arg\min} \sum_{j=1}^{n_i} \int_0^T \left(\Delta N_{ij}(s) - \eta\hat{\alpha}(s) \right)^2 K_b(t-s)\{\hat{\alpha}(s)\}^{-1} Y_{ij}(s)\, ds,$$

where we have adopted the notation that $\int \Delta N_i(s) W(s)ds \equiv \int W(s) dN_i(s)$ for some function W as in Nielsen (1998) and Nielsen and Tanggaard (2001). Note that the least square criterium uses the inverse variance $K_b(t-s)\{\hat{\alpha}(s)\}^{-1} Y_{ij}(s)$ as weights. Thus the above criterium is a kind of a local likelihood principle.

The criterion function itself is not tractable, but the differentiated (after η) criterion function is. This is seen from the following first order condition

$$\sum_{j=1}^{n_i} \int_0^T K_b(t-s)\, dN_{ij}(s) = \eta \sum_{j=1}^{n_i} \int_0^T K_b(t-s)\, \hat{\alpha}(s) Y_{ij}(s)\, ds.$$

where we have used $Y_{ij}(s)dN_{ij}(s) = dN_{ij}(s)$. The solution is

$$\eta = \frac{\sum_{j=1}^{n_i} \int_0^T K_b(t-s)\, dN_{ij}(s)}{\sum_{j=1}^{n_i} \int_0^T K_b(t-s)\, \hat{\alpha}(s) Y_{ij}(s)\, ds}.$$

Thus $\bar{\eta}_i(t)$ is given by the above formula.

APPENDIX D

The asymptotic properties of

$$\bar{\beta}_i(t) = \{\hat{D}_i\}^{-1} \tilde{\eta}_i(t)\hat{\alpha}(t)$$

is crucial while deriving the formulae for the credibility weights given in Section 5. As pointed out (in a slightly different context) in Section 4, we can replace $\bar{\beta}_i(t)$ by

$$\bar{\beta}_i(t) = \{D_i\}^{-1} \tilde{\eta}_i(t)\alpha(t)$$

from the point of view of asymptotic theory.

As a consequence of standard kernel hazard estimation techniques, see Nielsen and Tanggard (2001) or the appendix of Nielsen and Sandqvist (2000) we get

$$E(\tilde{\eta}_i(t) \,\big|\, \theta_i(t)) = \{1 + o(1)\} D_i \theta_i(t)$$

and

$$Var(\tilde{\eta}_i(t) \,\big|\, \theta_i(t)) = \{1 + o(1)\} C_2 D_i \theta_i(t) \alpha^{-1}(t) \{b \tilde{Y}_{i.}(t,b)\}^{-1},$$

where $C_2 = \int K^2(u) du$. Thus

$$E(\tilde{\eta}_i(t)) = \{1 + o(1)\} D_i,$$
$$Cov(\eta_i(t), \tilde{\eta}_i(t)) = \{1 + o(1)\} D_i^2 \sigma_t^2$$

and

$$Var(\tilde{\eta}_i(t)) = \{1 + o(1)\} \big[C_2 D_i \alpha^{-1}(t) \{b \tilde{Y}_{i.}(t,b)\}^{-1} + D_i^2 \sigma_t^2 \big].$$

Due to the first lines in this section we can replace $\tilde{\beta}_i(t)$ with $\bar{\beta}_i(t) = \{D_i\}^{-1} \tilde{\eta}_i(t) \alpha(t)$ and we conclude that

$$E\{\bar{\beta}_i(t)\} = \{1 + o(1)\} \alpha(t),$$

$$COV\{\beta_i(t), \tilde{\beta}_i(t)\} = \{1 + o(1)\} \sigma_t^2 \alpha^2(t)$$

and

$$VAR\{\bar{\beta}_i(t)\} = \{1 + o(1)\} \big[D_i^{-1} C_2 \alpha(t) \{b \tilde{Y}_{i.}(t,b)\}^{-1} + \sigma_t^2 \alpha^2(t) \big].$$

REFERENCES

AALEN, O.O. (1994) Effects of frailty in survival analyses. *Statistical Methods in Medical Research*, **3**, 227-243.

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

BICKEL, P.J., KLAASSEN, C.A.J., RITOV, Y. and WELLNER, J.A. (1993) *Efficient and adaptive estimation for semiparametric models*. The John Hopkins University Press, Baltimore and London.

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 credibillity approach to mortality risk. *Astin Bulletin*, **28**, 269-283.

HASTIE, T.J. and TIBSHIRANI, R.J. (1990) *Generalized Additive Models*. Chapman and Hall.

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

HOUGAARD, P. (2000) *Analysis of multivariate survival data*. Springer-Verlag, New-York.

JONES, M.C., LINTON, O.B. and NIELSEN, J.P. (1995) A simple bias reduction method for density estimation. *Biometrika*, **82**, 327-38.

MAMMEN, E., LINTON, O.B. and NIELSEN, J.P. (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics*, **27**, 1443-1490.

NIELSEN, J.P. (1998) Multiplicative bias correction in kernel hazard estimation. *Scand. J. Statist.*, **11**, 453-466.

NIELSEN, J.P., LINTON, O.B. and BICKEL, P. (1998). On a semiparametric survival model with flexible covariate effect. *Ann. Statist.*, **26**, 215-241.

NIELSEN, J.P. and SANDQVIST, B. (2000) Credibility weighted hazard estimation. *Astin Bulletin*, **30**, 405-417.

NIELSEN, J.P. and SANDQVIST, B. (2005) Correction note to "Credibility weighted hazard estimation. Astin Bulletin, 30(2), 405-417". *Astin Bulletin* **35(1)**, 259.

NIELSEN, J.P. and TANGAARD, C. (2001) Boundary and bias correction in kernel hazard estimation. *Scand. J. Statist.*, **28**.

NORBERG, R (2004) Credibility theory. In: *Encyclopedia of Actuarial Science*, Wiley.

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

WOODBURY, M.A. and MANTON, K.G. (1977) A randon walk model of human mortality and aging. *Theor. Population Biol.*, **11**, 37-48.

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