# Proportionality adjustment in credibility weighted hazard estimation

by

Jens Perch Nielsen and Bjørn Lunding Sandqvist

Codan

**Abstract**

This paper extends the credibility weighted intensity estimation ideas of Hardy and Panjer (1998) and Nielsen and Sandqvist (2000) to the proportional intensity case, where the level of each intensity is estimated from data and used to construct the credibility weighted estimator in areas of data sparsity.

## Introduction

Inspired by the credibility approach to hazard estimation of Hardy and Panjer (1998), Nielsen and Sandqvist (2000) considered hazards of different groups assum-

**Key Words and Phrases:** Counting process theory; Kernel hazard estimation; Credibility; Bühlmann-Straub model; Proportional hazard models.

ing the hazard of each group to fluctuate across a common baseline hazard. This fluctuation is modelled by a heterogeneity parameter capturing the particular properties of each group allowing for a surprisingly simple credibility estimation procedure. The new estimation procedure extends the applicability of nonparametric smoothing techniques to a number of data sets from the insurance industry or other places. The estimation of credit risk could be one important example from finance, where a number of possible transitions from one credit risk group to another is 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 for one group is used to estimate the hazards of other groups. The effect of credibility in areas of high information will contribute to improve estimation in areas of low information.

Nielsen and Sandqvist (2000) showed through a practical application and a simulation study, that this credibility method considerable improves estimation in areas of data sparsity. 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). This is relevant in situations, where data sparsity makes it impossible to construct reliable nonparametric estimators in certain areas. Consider for example a situation, where mortality is estimated for woman and men and data sparsity is present for women in certain age intervals. We do, however, have sufficient data to conclude that the level of mortality of women is lower than the level of mortality of men. We suggest to weight the individual estimator towards the proportional intensity estimator in this situation. However, 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 other words 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. This automatic method therefore has the traditional advantage of the credibility method. We always have an estimator to use. Even if we start a new line of insurance business with very little information on risk. Over time more information accumulates and the crude proportionality assumption is gradually replaced by an estimation technique relying exclusively on this new line of business.

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

$$\gamma_i(t) = D_i \ \beta_i(t),$$

where

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

and $\alpha(t)$ is a smooth baseline intensity and $E\{\theta_i(t)\} = 1$. In average we have the proportional intensity model.

We show that the proportionality factors, $D_i'$s, can be estimated through a simple integral of some nonparametric estimator. We also argue that this leads to square-root-$n$ consistent estimators. This parametric rate of convergence of the proportionality factors is superior to the rate of convergence present in smoothing problems. Therefore, the asymptotics of the smooth estimators of the group hazards are estimated as well with our estimated proportionality factors as they would have been, if the proportionality factors have actually been known beforehand. This is not surprising and follow standard semiparametric analysis, see Bickel et al. (1993).

In this paper we use nonparametric smoothing to estimate different hazards, e.g. the $\alpha$. Although nonparametric smoothing techniques have been developed quite a bit 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 lends 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.

## 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 $N_{ij}$ count observed failures for the $j$'th individual in the $i$'th group in the time interval $[0, T]$. We assume that $N_{ij}$ is taking 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, p60).

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, $\theta_i(t) \geq 0$ 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. $D_1, \ldots, D_k$ are parameters playing the same type of role as parameters in traditional problems.

We assume that the stochastic processes $\theta_1, \ldots, \theta_k$ defined on the time interval $[0, T]$ are independent identically distributed and that in the conditional distribution given $\theta_1, \ldots, \theta_k$, the counting processes $(N_{11}, Y_{11}), .., (N_{kn_k}, Y_{kn_k})$ are independent for the $m$ individuals and within the $i$'th group then the counting processes $(N_{i1}, Y_{i1}), .., (N_{in_i}, Y_{in_i})$ are identically distributed.

We assume that

$$E\theta_i(t) = 1 \tag{1}$$

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

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

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

for some weighting function $w \geq 0$, 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 $\delta_i's$. The kernel $K$ is a probability density function symmetric about zero and $K_b(\cdot) = b^{-1}K(\cdot/b)$ for any bandwidth $b > 0$.

The normalization assumptions (1) and (2) identify the model. They imply that $\int_0^T \alpha(t)w(t)dt = 1$, which determines the level of $\alpha$. They also imply that every realization of the process $\theta_i$ will be just as much above as below 1 (measured with respect to the probability measure with density $\alpha(\cdot)w(\cdot)$).

Define

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

which will be used in the estimation procedure.

The group exposure

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

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

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

## 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.

These four intensities are clearly much too volatile to be useful predictors of the disability intensities involved. Nielsen and Sandqvist (2000) alleviated this problem through a credibility weighting towards a common baseline hazard, see Figure 2.

We see that the credibility weighted estimators of Figure 2 are much less volatile than the individual nonparametric estimators of Figure 1. They can therefore be expected to be much 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, 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

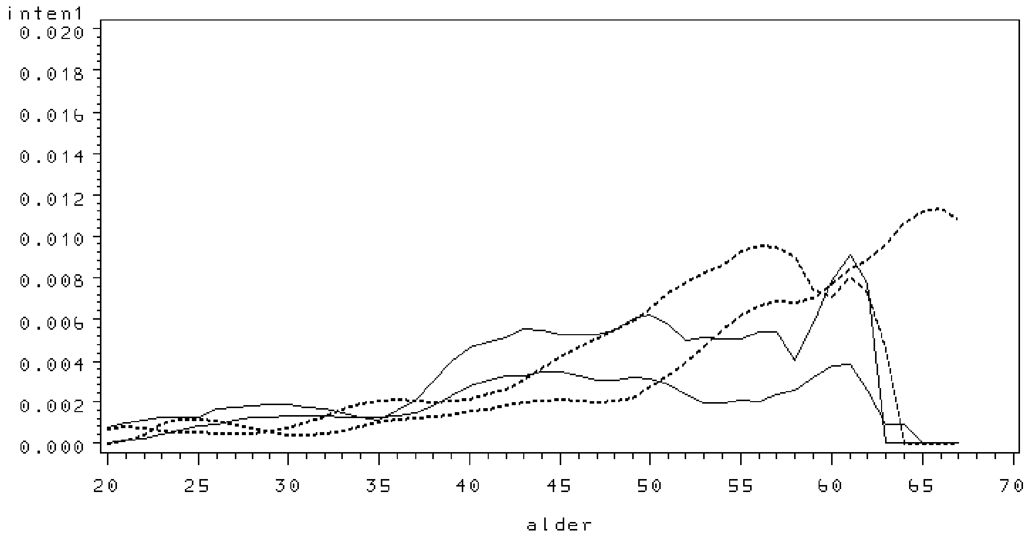


Figure 1: The four estimated individual hazards (male: dotted, female: solid)

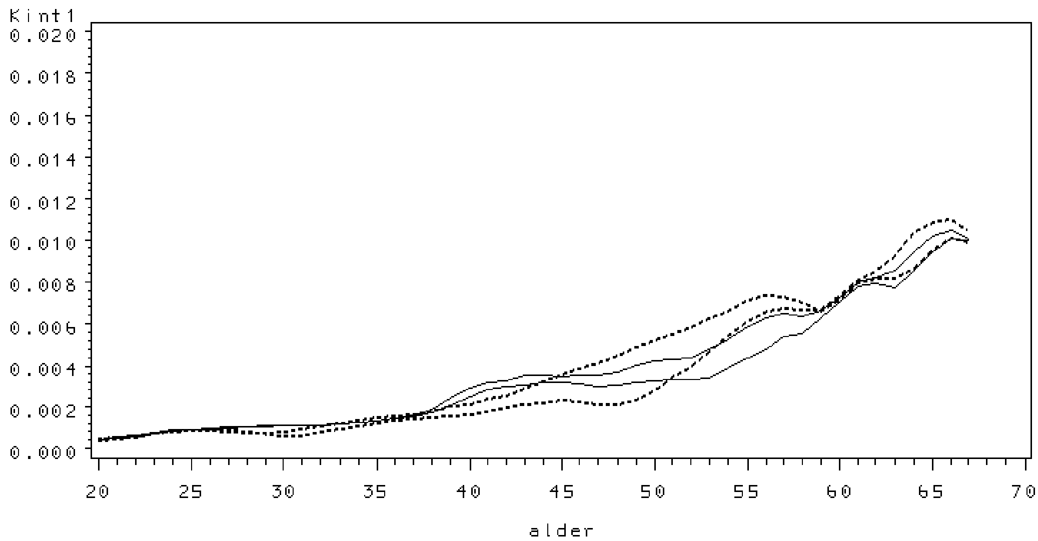


Figure 2: The four estimated credibility hazards (male: dotted, female: solid) from Nielsen and Sandqvist (2000)

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 $\alpha$ of the model defined in Section 2. We establish a simple estimator of the proportionality constants based on integrating out a relevant nonparametric kernel estimator. If this kernel estimator is undersmoothed, 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 $\alpha$ 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 $\alpha$

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 $\gamma_i$. This is

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

where $\overline{\gamma}_i$ is the local constant kernel hazard estimator, see Hjort (1992) or Nielsen and Tanggaard (2001),

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

with bandwidth $h$.

Under standard conditions, $\widehat{D}_1, \ldots, \widehat{D}_k$ 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 $f_i > 0$ for $i \in \{1, \ldots, k\}$, and where the bandwidth $h$ is undersmoothed, see Appendix A. Since smoothing estimators converge slower than square-root-$n$ consistently this leaves the estimation error of the proportionality constants to be of lower order of magnitude.

To estimate $\alpha$ we use a different bandwidth $b > h$ and a modification of the local constant kernel hazard estimator adjusted for $\widehat{D}_1, \ldots, \widehat{D}_k$ :

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

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

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

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 $\overline{\alpha}$ is an estimator of $\alpha$.

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

$$\widehat{\alpha}(t) = \frac{\widetilde{\alpha}(t)}{\int_0^T \widetilde{\alpha}(s)w(s)ds}.$$

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

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

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 $\overline{\eta}_i(t)$ we note that the nominator estimates $D_i\theta_i(t)\alpha(t)Y_{i.}(t)$ while the denominator estimates $\alpha(t)Y_{i.}(t)$.

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 $\eta_i$ to

$$\widetilde{\eta}_i(t) = \widehat{D_i}\overline{\eta}_i(t) / \int_0^T \overline{\eta}_i(s)\widehat{\alpha}(s)w(s)ds,$$

such that

$$\widehat{D_i} = \int_0^T \widetilde{\eta}_i(s)\widehat{\alpha}(s)w(s)ds.$$

From the point of view of theoretical analyzes, we can think of $\widehat{\alpha}$ as an accurate estimator of $\alpha$, 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 $\widehat{\alpha}$ by $\alpha$, 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

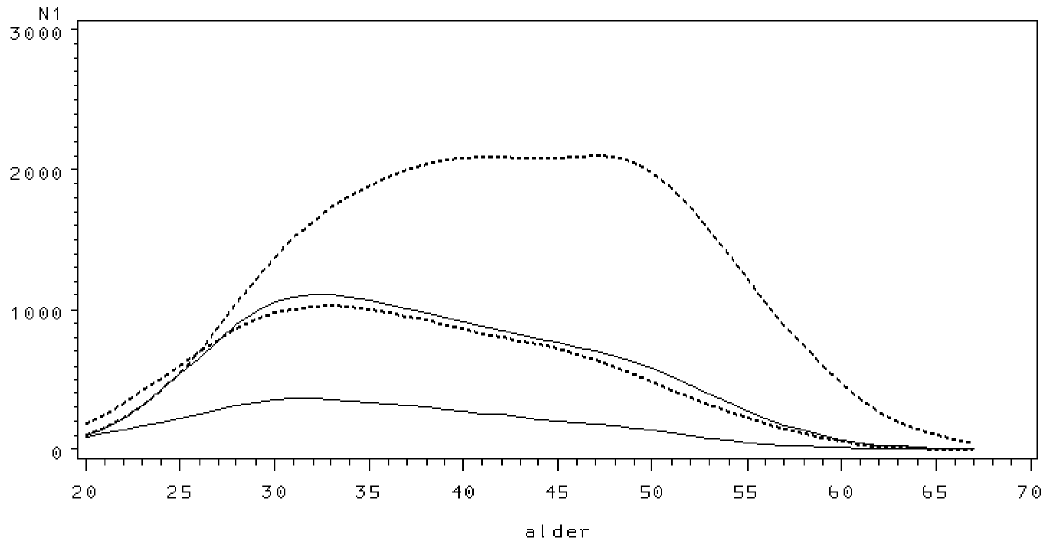


Figure 3: The smoothed exposure in the four groups (male: dotted, female: solid).

$b = 5$ years. The kernel function used is

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

We choose our weight function to be

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

which for the age interval $[20, 67]$ is the total exposure per age normalized to an average level of 1.

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

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

| $\widehat{D}_1$ | $\widehat{D}_2$ | $\widehat{D}_3$ | $\widehat{D}_4$ |
|---|---|---|---|
| 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 $\widehat{D}_1, ..., \widehat{D}_4$, we construct non-parametric estimators, $\widehat{\alpha}$ of $\alpha$, and $\widetilde{\eta}_i(t)$ of $\eta_i(t) = D_i\theta_i(t)$. The above estimators lead to the individual estimators

$$\widetilde{\gamma}_i(t) = \widetilde{\eta}_i(t)\widehat{\alpha}(t)$$

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

Note that, compared to Figure 1, the picture is radically changed under the proportional assumption. In particular for substandard lives in the age interval above 55

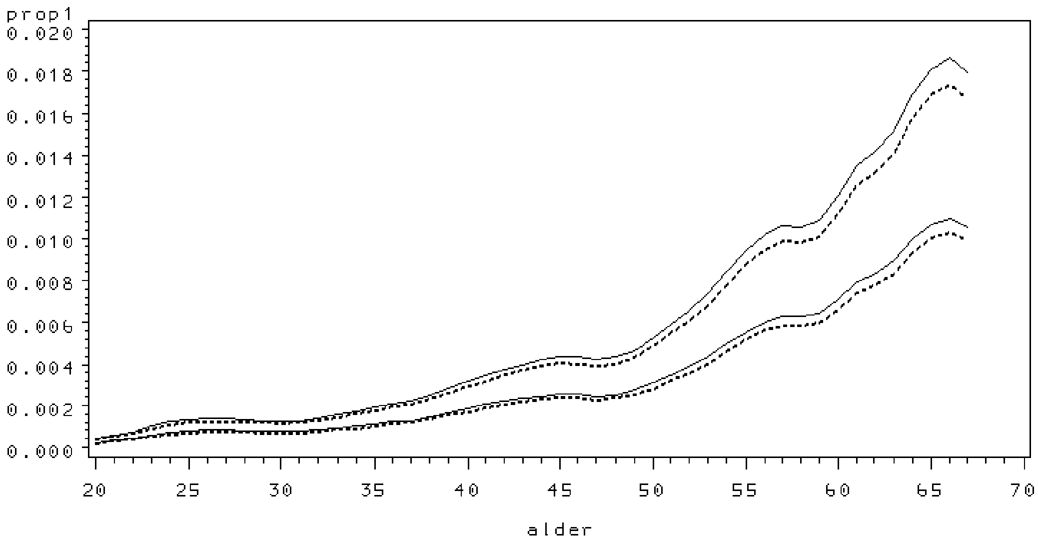


Figure 4: The proportional hazards $\widehat{D}_i\widetilde{\alpha}(t)$ (male: dotted, female: solid).

years. The proportional hazards for all groups, except for the normally insured men, are much higher in the older ages than the group specific hazards in Figure 1.

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

To indicate how much the individual estimator $\widetilde{\gamma}_i(t)$ differs from the proportional estimator $\widehat{D}_i\widehat{\alpha}(t)$ we look at the behavior of $\left\{\widehat{D}_i\right\}^{-1}\widetilde{\eta}_i(t)$ 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.

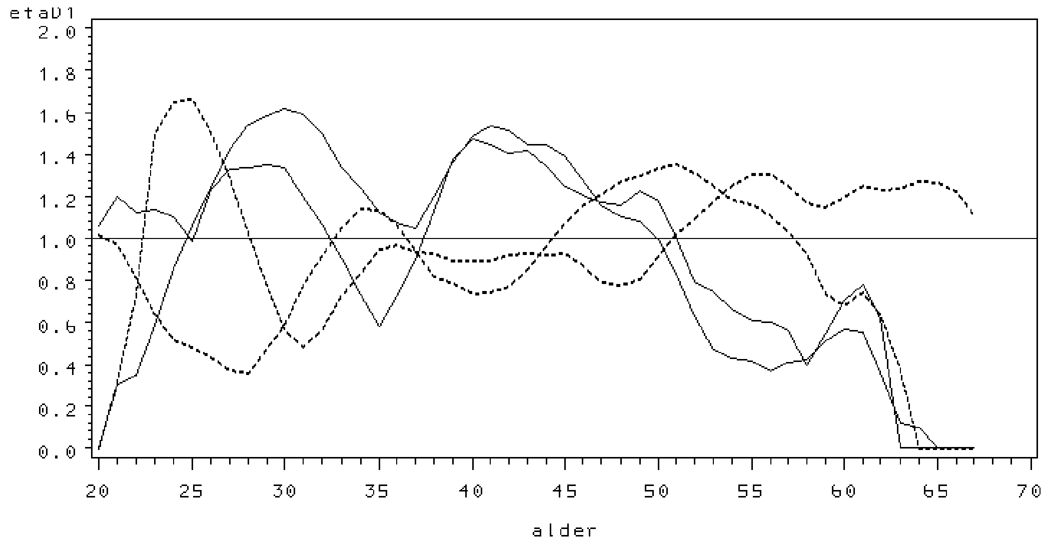


Figure 5: The raw non-credibility estimator of $\theta_i\,(t)$, $\left\{\widehat{D_i}\right\}^{-1}\widetilde{\eta}_i(t)$ (male: dotted, female: solid).

We start by adjusting our problem such that it becomes comparable to classical credibility theory and the methodology of Nielsen and Sandqvist (2000) and we define a 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 observed candidate

$$\widetilde{\beta}_i(t) = \left\{\widehat{D_i}\right\}^{-1}\widetilde{\eta}_i(t)\widehat{\alpha}(t)$$

from the $i$'th group. This parallels the original approach of Bühlmann and Straub (1970) that also has a Hilbert space interpretation.

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

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

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

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

namely

$$(1 - z_{i,t})E\left\{\widetilde{\beta}_i(t)\right\} + z_{i,t}\widetilde{\beta}_i(t),$$

where

$$z_{i,t} = \frac{COV\left\{\beta_i(t), \widetilde{\beta}_i(t)\right\}}{VAR\left\{\widetilde{\beta}_i(t)\right\}}.$$

From Appendix D we get the following three results

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

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

and

$$VAR\left\{\widetilde{\beta}_i(t)\right\} = \{1 + o(1)\}\left[D_i^{-1}C_2\alpha(t)\left\{\widetilde{Y}_{i\cdot}(t,b)\right\}^{-1} + \sigma_t^2\alpha^2(t)\right],$$

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

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

where

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

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

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

Note that if we had $D_i = 1$ for all $i$, then the credibility weighting $z_{i,t}$ is equal to the one Nielsen and Sandqvist (2000) arrived at. Note that $z_{i,t}$ is increasing as a function of $D_i \alpha(t) \widetilde{Y}_{i.}(t,b)$. This latter quantity can be interpreted as the expected number of disability cases at time $t$ and 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 $\sigma_t^2$. We have earlier estimated all the above needed components except for the variance $\sigma_t^2$ which will be estimated in the next section.

## 6. Estimating the variance

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

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

where we have employed that $E\theta_i(t) = 1$ and that $\widetilde{\eta}_i(t)$ is an estimator of $\eta_i(t) = D_i\theta_i(t)$ and therefore $\widetilde{\eta}_i(t) \left\{ \widehat{D}_i \right\}^{-1}$ (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 $D_1, \ldots, D_k$.

When $k$ is small it might be better to assume that $\sigma_t^2 = \sigma^2$, since $\widetilde{\sigma}_t^2$ is estimated with little accuracy. A natural choice of estimator in this case is the weighted average:

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

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

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

In our example we choose to use the constant variance assumption, since we only have four groups, $k = 4$. We can also see from Figure 6 that the estimator of the time varying variance fluctuate a lot, in particular in areas with low exposure levels. In the following we continue to use the notation $\widetilde{\sigma}_t^2$ for the variance estimator.

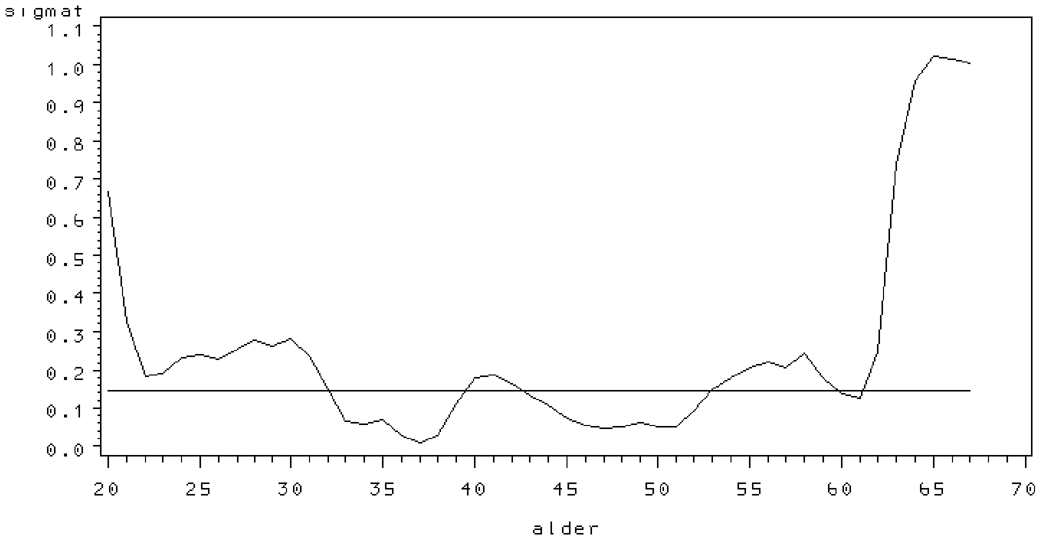


Figure 6: The time dependent variance estimator $\widetilde{\sigma}_t^2$ and the constant variance estimator $\widetilde{\sigma}^2 = 0,144$.

## 7. The final estimator

Now we have all the components to estimate the credibility weights $z_{i,t}$ and the result is

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

Then we get

$$\widetilde{\theta}_i(t) = (1 - \widetilde{z}_{i,t}) + \widetilde{z}_{i,t} \widetilde{\beta}_i(t) \left\{\widehat{\alpha}(t)\right\}^{-1}$$

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

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

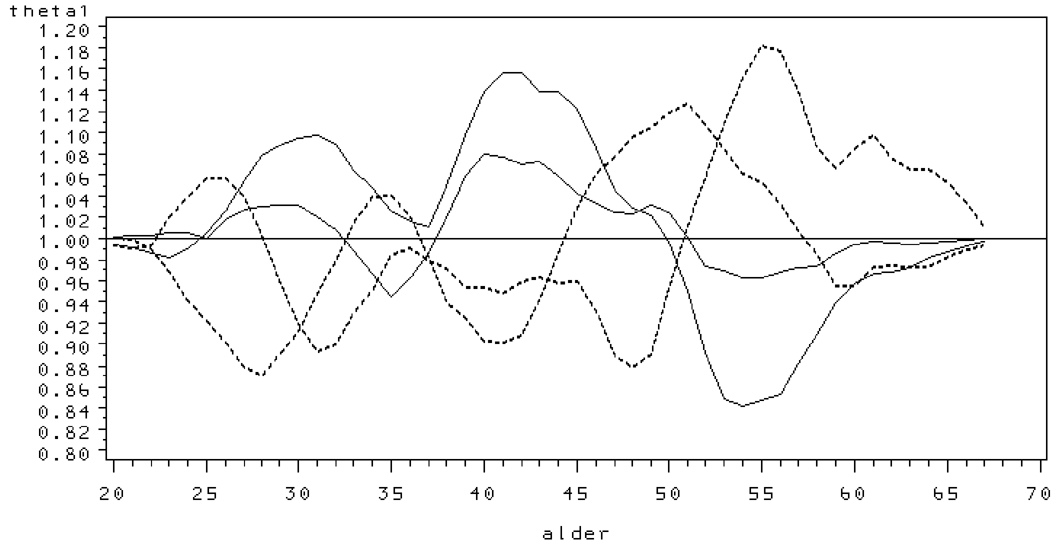


Figure 7: The credibility estimator $\widehat{\theta}_i(t)$ (male: dotted, female: solid).

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

It is seen that this final credibility estimator fluctuates a lot 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 $\widehat{\theta}_i(t)$ 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 $\pm 20\%$. While this fluctuation is within reasonable limits, it does, however, indicate important deviations from the proportional hazard model.

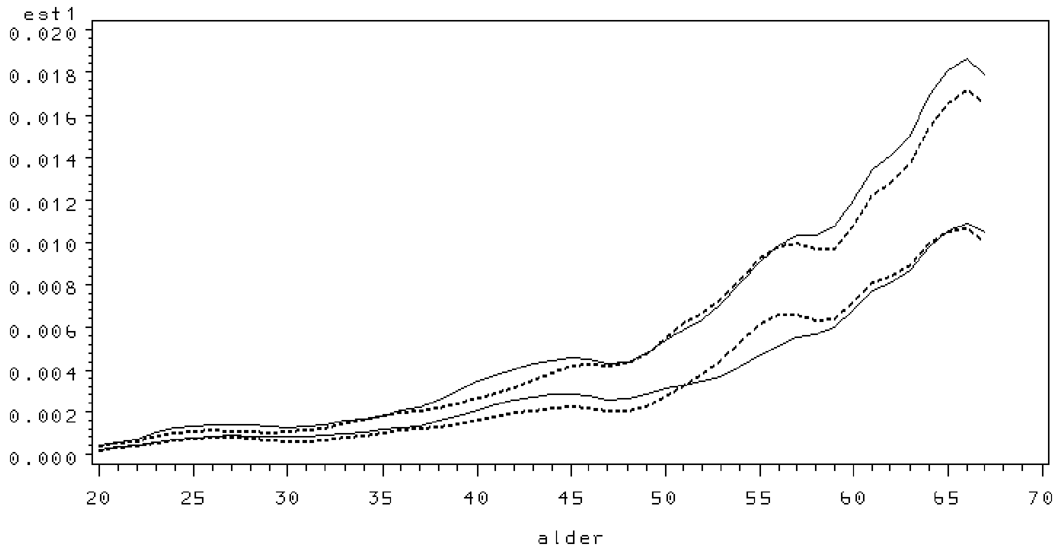


Figure 8: The final credibility hazard $\widehat{\gamma}_i(t)$ (male: dotted, female: solid) in the proportional model $D_i\theta_i(t)\alpha(t)$, where $\widehat{D_1} = 0,101$, $\widehat{D_2} = 0,172$, $\widehat{D_3} = 0,095$ and $\widehat{D_4} = 0,160$.

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

$$\widehat{\gamma}_i(t) = \widehat{D_i}\widehat{\theta}_i(t)\widehat{\alpha}(t).$$

In areas of data sparsity $\widehat{\gamma}_i(t)$ 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, $\widehat{\gamma}_i(t)$ is close to the individual estimator $\overline{\gamma}_i(t,b)$ 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 a great 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 is above the intensity for normally insured men for ages below 51 years while the opposite is true for ages above 51 years. This can be compared to the earlier conclusion that the overall disability risk for normally insured women is about 6% above normally insured men. Therefore, the credibility approach of this paper captures effects in data which can not be shown under the proportional assumption.

Note 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.

One can conclude that the model assumptions play an important role. For example, in areas with low exposure we essentially use $\widehat{D_i}\widehat{\alpha}(t)$ in this paper and $\widehat{\alpha}(t)$ in Nielsen and Sandqvist (2000). Although Figure 8 and Figure 2 have significant differences, both graphs get rid of most of the fluctuations present in Figure 1.


## Acknowledgements

We would like to thank Søren Feodor Nielsen and Philip Hougaard for valuable comments.


## 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 component 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

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

such that

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

It follows from standard kernel smoothing technique, see Nielsen and Tanggaard (2001), that the estimation error $\overline{\gamma}_i(t,h) - \gamma_i(t)$ 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:

$$\widehat{D}_i - D_i = \overline{V}_i + \overline{B}_i,$$

where

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

and

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

The point is that while the variable part $V_i(t)$ has rate of convergence of the order of magnitude $O_P\left\{(nh)^{-1}\right\}$ then $\overline{V}_i$ 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 $O_P\left(h^2\right)$. Integrating does not change the order of magnitude and $\overline{B}_i = O_P\left(h^2\right)$ as well. Therefore, if an undersmoothed bandwidth $h = o_P(n^{-1/4})$ is chosen, then $\overline{B}_i = o_P\left(n^{-1/2}\right)$ and

$$\widehat{D}_i - D_i = \overline{V}_i + \overline{B}_i$$

is square-root-$n$ consistent, since $\overline{V}_i$ 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 $\overline{\alpha}$ and $\widetilde{\alpha}$ are equivalent from an asymptotic point of view. Here we verify that $\overline{\alpha}$ estimates $\alpha$. Note that the difference of

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

and

$$\overline{\alpha}^{*}(t)=\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\int_0^T K_b\left(t-s\right)\theta_i\left(s\right)\alpha(s)Y_{ij}(s)ds}{\sum_{i=1}^{k}\widetilde{Y}_{i.}(t,b)}$$

is equal to

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

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

$$(nb)^{-1/2}\left\{\overline{\alpha}(t)-\overline{\alpha}^{*}(t)\right\}$$

converges to a normal distribution.

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

$$\alpha(t)\frac{\sum_{i=1}^{k}\widetilde{Y}_{i.}(t,b)\theta_i\left(t\right)}{\sum_{i=1}^{k}\widetilde{Y}_{i.}(t,b)}$$

for small $b$.

The term

$$\frac{\sum_{i=1}^{k}\widetilde{Y}_{i.}(t,b)\theta_i\left(t\right)}{\sum_{i=1}^{k}\widetilde{Y}_{i.}(t,b)}$$

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

## Appendix C

The exact procedure for estimation of $\overline{\eta}_i(t)$ in Section 4.2 can be written as

$\arg\min_{\eta} \lim_{q\to 0}$

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

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

where we have adopted the notation that $\int \triangle N_i(s)W(s)ds \equiv \int W(s)dN_i(s)$ for some function $W$ as in Nielsen (1998) and Nielsen and Tanggaard (2001). Note that the least square criterium uses the inverse variance $K_b(t-s)\left\{\widehat{\alpha}(s)\right\}^{-1} Y_{ij}(s)$ 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 $\eta$) criterion function is. This is seen from the following first order condition

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

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

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

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

## Appendix D

The asymptotic properties of

$$\widetilde{\beta}_i(t) = \left\{\widehat{D_i}\right\}^{-1} \widetilde{\eta}_i(t)\widehat{\alpha}(t)$$

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 $\widetilde{\beta}_i(t)$ by

$$\overline{\beta}_i(t) = \{D_i\}^{-1}\,\widetilde{\eta}_i(t)\alpha(t)$$

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

$$E\left(\widetilde{\eta}_i(t)|\theta_i(t)\right) = \{1 + o(1)\}\, D_i\theta_i(t)$$

and

$$Var\left(\widetilde{\eta}_i(t)|\theta_i(t)\right) = \{1 + o(1)\}\, C_2 D_i\theta_i(t)\alpha^{-1}(t)\left\{\widetilde{Y}_{i\cdot}(t,b)\right\}^{-1},$$

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

Thus

$$E\left(\widetilde{\eta}_i(t)\right) = \{1 + o(1)\}\, D_i,$$

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

and

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

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

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

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

and

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

## References


Andersen, P.K., Borgan, Ø., Gill, R.D., and N. Keiding (1992). *Statistical models based on counting processes.* Springer-Verlag, New-York.

Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (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 R.J. Tibshirani (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.

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

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

Jens Perch Nielsen and Bjørn Lunding Sandqvist

Codan

Gammel Kongevej 60

1790 København V

Denmark

e-mail: npj@codan.dk and san@codan.dk