Boundary and Bias Correction in Kernel Hazard Estimation

JENS PERCH NIELSEN
Codan

CARSTEN TANGGAARD
The Aarhus School of Business

ABSTRACT. A new class of local linear hazard estimators based on weighted least square kernel estimation is considered. The class includes the kernel hazard estimator of Ramlau-Hansen (1983), which has the same boundary correction property as the local linear regression estimator (see Fan & Gijbels, 1996). It is shown that all the local linear estimators in the class have the same pointwise asymptotic properties. We derive the multiplicative bias correction of the local linear estimator. In addition we propose a new bias correction technique based on bootstrap estimation of additive bias. This latter method has excellent theoretical properties. Based on an extensive simulation study where we compare the performance of competing estimators, we also recommend the use of the additive bias correction in applied work.

Key words: boundary kernels, counting process theory, hazard functions, kernel estimation, local linear estimation

1. Introduction

Local linear estimation has recently become popular in statistical inference (see Fan & Gijbels, 1996 or Wand & Jones, 1995 for excellent reviews). While local linear estimation was introduced to one-dimensional density estimation in Jones (1993), it has until now not been considered in the area of one-dimensional hazard estimation. In this paper we derive a new class of local linear one-dimensional hazard estimators based on least squares kernel estimation with different weighting schemes. We show that the local linear estimator can be based on different weightings and that the estimator of Ramlau-Hansen is related to one particular weighting. If Ramlau-Hansen’s estimator is boundary corrected with a particular set of the Gasser & Müller (1979) type of boundary kernels, then it can be interpreted as a local linear hazard estimator with one particular type of weighting which we call the Ramlau-Hansen weighting. Starting with a different point of view Andersen et al. (1993, p. 251) arrived at exactly these boundary kernels while implementing Ramlau-Hansen’s kernel hazard estimator. We tend to prefer a different weighting principle which we call the natural weighting, even though the pointwise asymptotic properties are identical for the entire class of local linear estimators. We argue that the Ramlau-Hansen weighting is more sensitive to volatile exposure patterns than the natural weighting principle. While asymptotic theory does not catch this point, a simulation experiment shows the importance of what we call the exposure robustness of the natural weighting principle. As the Ramlau-Hansen weighting principle does not have any particular advantages to the natural weighting principle, we advise practitioners to use the latter.

We also consider two bias correction methods, namely multiplicative and additive bias corrections. The former was introduced for non-parametric regression by Linton & Nielsen (1994) and transferred to density estimation by Jones et al. (1995) and to kernel hazard estimation by Nielsen (1998b). We show that multiplicative bias correction can be interpreted as the result of a minimization of the same least squares principle that we use for our local linear
estimators. The second bias reduction principle is also based on a minimization and is an
additive bias reduction which subtracts a data dependent fraction of the estimated bias from the
original estimator to obtain the bias reduced estimator. While Jones & Signorini (1997) advocate
that the multiplicative bias reduction principle seems to be the best among the known bias
reduction principles, we conclude from a simulation study that the additive bias correction
principle is often better than the multiplicative bias correction. This is especially pronounced
with large data sets. The theoretical reason is that the additive bias correction method has a
superior bias of order $O_P(b^6)$ compared to the order $O_P(b^4)$ for the multiplicative bias correction
estimator, where $b$ is the smoothing parameter. Both bias correction principles have the iteration
property that they, at least theoretically, reduce bias on any estimator. This iteration property is
also known from the transformation approach to bias reduction which was analysed in detail by
Hössjer & Ruppert (1995). While Hössjer & Ruppert (1995) considered the $m$th time iterated
estimator, we only apply our bias reduction principle twice. The simulation study shows that for
both multiplicative bias correction and additive bias correction, it requires a very large data-set
before more than one single iteration becomes useful. Furthermore, the simulation study shows
that the one time additive bias reduction method performs better than the two times iterated
multiplicative bias reduction method. This leads us to the conclusion that the bias reduction
techniques should be applied only once.

The outline of the paper is as follows. In section 2 we outline the counting process framework.
In section 3 we consider the general split of a kernel hazard estimator into a variable part and a
stable part paralleling the variance and the bias in kernel density estimation. In section 4 we
consider the class of local constant estimators. The local constant estimator that Hjort (1992)
derived from the point of view of local parametric estimation, is identical to our local constant
estimator based on what we call the natural weighting principle. In section 5 we introduce the
new class of local linear least squares hazard estimators and we derive the pointwise asymptotic
theory of these estimators. In section 6 we introduce the multiplicative bias correction method
as a general method that can be applied to any pilot estimator. We also show that the
multiplicative correction principle can be carried out as a minimization of a least squares
criterion, namely the same local least squares criterion that we used for our local linear
estimator. In section 7 we introduce the new bias reduction principle based on additive bias
reduction. In section 8 we go through an extensive simulation study comparing our five
estimators.

2. Notation and a counting process formulation of the model
We observe $n$ individuals $i = 1, \ldots, n$. Let $N_i$ count observed failures for the $i$th individual in
the time interval $[0, T]$. We assume that $N_i$ 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. (1993, p. 60). We model the random intensity as

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

with no restriction on the functional form of $\alpha(\bullet)$. Here, $Y_i$ is a predictable process taking values
in $\{0, 1\}$, indicating (by the value 1) when the $i$th individual is under risk. We assume that
$(N_1, Y_1), \ldots, (N_n, Y_n)$ are i.i.d. for the $n$ individuals. All martingales and predictable processes
are defined with respect to the filtration $\mathcal{F}_t$. With these definitions, $\lambda_i$ is predictable and the
processes $M_i(y) = N_i(y) - A_i(y), i = 1, \ldots, n$, with $A_i(y) = \int_0^y \lambda_i(s) \, ds$, are squared integrable
local martingales. We use $\rightarrow$ to denote convergence in probability, and $\Rightarrow$ to signify weak
convergence.

3. Splitting an estimator of $\alpha$ into a variable part and a stable part

All estimation errors of the kernel hazard estimators of this paper can be presented in the same way, namely as a sum of a variable part and a stable part. Properly normalized, the variable part converges to a normal distribution and properly normalized, the stable part converges in probability to a constant. If $\tilde{\alpha}$ is some estimator of $\alpha$ the split of the estimation error can be written as

$$\tilde{\alpha}(t) - \alpha(t) = V_t + B_t,$$

where the variable term can be linearized as

$$V_t = \sum_{i=1}^{n} \int_{0}^{T} h_i'(s) \, dM_i(s),$$

for some set of functions $(h_i')$. The variable part, $V_t$, can typically be analysed by martingale techniques. However, for many bias corrected estimators the functions $(h_i')$ are not predictable with respect to the filtration $\mathcal{F}_t$; this predictability issue can, however, be overcome (see section 7).

If an estimator, $\tilde{\alpha}$, of $\alpha$ can be written on the above form, we will call it a linearized estimator. In section 6 and section 7 we define bias reduction techniques that will work on any linearized estimator.

In this paper we will discuss five different estimators of $\alpha$ based on least squares local constant or local linear estimation using some weight function, $W$. We use the notation $\tilde{\alpha}_{1,W}(t)$, $\ldots$, $\tilde{\alpha}_{5,W}(t)$ for these estimators and $V_{1,W}(t)$, $\ldots$, $V_{5,W}(t)$ and $B_{1,W}(t)$, $\ldots$, $B_{5,W}(t)$ for the corresponding variable and stable parts. In the cases considered in this paper we will have that $(nb)^{1/2} V_{i,W}(t)$ converges to a normal distribution, where $n$ is the sample size and $b$ is the smoothing parameter. Notationally, we use $\sigma_{i,W}^2(t)$ to denote the variance of this normal distribution.

4. Local constant estimators

Let $K$ be a probability density function with support $[-1, 1]$ which is symmetric about zero and let $K_s(\bullet) = b^{-1} K(\bullet/b)$ for any $b$. Define also $Y^{(n)}(s) = \sum_{i=1}^{n} Y_i(s)$ and $J^{(n)}(s) = I(Y^{(n)}(s) > 0)$.

We now define a local constant kernel hazard estimator based on the least squares minimization principle introduced in Nielsen (1998a). It is closely related to a similar minimization principle introduced for density estimation in Jones (1993). Let

$$\tilde{\alpha}(t) = \arg \min_{a} \lim_{w \to 0} \sum_{i=1}^{n} \int_{0}^{T} \left( \frac{1}{w} \int_{s-w}^{s} dN_i(u) - a \right)^2 K_b(t - s) W(s) Y_i(s) \, ds$$

$$= \arg \min_{a} \sum_{i=1}^{n} \int_{0}^{T} \left( \Delta N_i(s) - a \right)^2 K_b(t - s) W(s) Y_i(s) \, ds,$$

where we have adopted the notation $\int \Delta N_i(s) W(s) \, ds = \int W(s) \, dN_i(s)$. The criterion function itself is not very tractable, but the differentiated criterion function is. This is clearly seen from the following first order condition in $a$

$$\sum_{i=1}^{n} \int_{0}^{T} K_b(t - s) W(s) Y_i(s) \, dN_i(s) = a \sum_{i=1}^{n} \int_{0}^{T} K_b(t - s) W(s) Y_i(s) \, ds$$

with solution
We will call $\hat{\alpha}_{1,W}(t)$ the local constant estimator of $\alpha(t)$ based on the weighting function $W$. Ramlau-Hansen’s estimator, with the traditional cut-and-normalized kernel at the boundary (see Gasser & Müller, 1979; Zhang et al., 1999), corresponds to the local constant estimator with weighting function $W(s) = [nY(n)/Y(n)](s)$, whereas the simple or natural weighting function $W(s) = 1$ corresponds to an estimator that Hjort (1992) arrived at from a different minimization principle based on a local likelihood. We call the first choice of weighting the Ramlau-Hansen weighting, $W_{R}$, and the latter choice of weighting the natural weighting $W_{N}$.

Below we give the asymptotic properties of the two local constant estimators. Now define the kernel moments: $\kappa_0 = \int_{-1}^{1} v^2 K(v) \, dv$ and $\kappa_2 = \int_{-1}^{1} K^2(v) \, dv$.

### General assumptions (GA)

Define the following general assumptions: $A_{t,b}$ is defined as the *local neighbourhood* $A_{t,b} = [t - 10b, t + 10b]$, where $b$ is the bandwidth and $b \to 0$, $nb \to \infty$ as $n \to \infty$. The positive functions $\gamma, w$ are the *local limits* of respectively the exposure and the weighting function, i.e.

$$\sup_{s \in A_{t,b}} |Y^{(n)}(s/n) - \gamma(s)| \xrightarrow{P} 0$$

and

$$\sup_{s \in A_{t,b}} |W(s) - w(s)| \xrightarrow{P} 0,$$

where $w, \gamma \in C_1([0, T])$.

The first part of the following theorem, on the asymptotic theory of the local constant estimator with the Ramlau-Hansen weighting, can be found in Nielsen (1998b). The second part, on the asymptotic theory of the local constant estimators with the natural weighting, can be found in Hjort (1992).

### Theorem 1

Assume that the general assumptions GA hold and assume that $\alpha \in C_2([0, T])$. Then

$$(nb)^{1/2} V_{1,W}(t) \Rightarrow N\{0, \sigma_{1,W}^2(t)\}$$

and

$$B_{1,W}(t) = b^2 m_{W}(t) + o_P(b^2),$$

where

$$m_{W}(t) = \kappa_0 \{\alpha''(t)/2\}, \quad m_N(t) = \kappa_0 \{\alpha'(t)\gamma'(t)/\gamma(t) + \alpha''(t)/2\}$$

and

$$\sigma_{1,W}^2(t) = \kappa_2 \alpha(t) \{\gamma(t)\}^{-1}.$$
While the two competing local constant estimators have the same asymptotic variance, there is a little difference with respect to the bias. The bias associated with the local constant estimator using the natural weighting has one more term than the bias term associated with the local constant estimator using the Ramlau-Hansen weighting. This extra term depends on the limit level of the exposure and this may be a reason to prefer the Ramlau-Hansen weighting in this particular case. It is easy to see that none of these estimators are asymptotically better for every combination of $\alpha$ and $\gamma$. As noted above, the asymptotic theory only holds in the interior of the interval. Close to boundary regions the bias of these local constant estimators has an inferior rate of convergence, see Jones (1993, p. 137) for a precise account of this phenomenon in the similar case of local constant density estimation.

5. Local linear estimators

In this section we define the new class of local linear estimators. There is one local linear estimator for each choice of weights. While the asymptotic theory of the local constant least squares estimators depends on the choice of weighting, the situation is different in the local linear case. These estimators have the same pointwise behaviour independent of the choice of weighting. We consider two particular choices of weighting: the Ramlau-Hansen weighting and the natural weighting. It is pointed out in section 8 that the small sample properties of these estimators are almost identical as long as the exposure has a typical slowly varying character. However, if the exposure has a lot of abrupt volatility, perhaps caused by some rather special truncation and censoring pattern, the simulation study shows that the natural weighting outperforms the Ramlau-Hansen weighting considerably (see section 8). We therefore conclude that the natural weighting is more robust to volatile exposure patterns than the Ramlau-Hansen weighting. Since the Ramlau-Hansen weighting does not have any particular advantages over the natural weighting we recommend practitioners to change their habit of using Ramlau-Hansen’s weighting and begin to use the natural weighting.

The local linear estimator corresponding to the weighting function $W$ is defined as

$$\hat{\alpha}_{2,W}(t) = \hat{\Theta}_0,$$

where

$$
\begin{align*}
\begin{pmatrix}
\hat{\Theta}_0 \\
\hat{\Theta}_1
\end{pmatrix} &= \arg \min_{\Theta_0, \Theta_1} \sum_{i=1}^{n} \int_{0}^{T} [\Delta N_i(s) - \Theta_0 - \Theta_1(t-s)]^2 K_b(t-s) W(s) Y_i(s) \, ds.
\end{align*}
$$

Let

$$a_j(t) = \int_{0}^{T} K_b(t-s)(t-s)^j W(s) Y_i(s) \, ds$$

for $j = 0, 1, 2$ and

$$G_j(t) = \sum_{i=1}^{n} \int_{0}^{T} K_b(t-s)(t-s)^j W(s) dN_i(s)$$

for $j = 0, 1$.

Then to find $(\hat{\Theta}_0, \hat{\Theta}_1)$ amounts to solving the equations

$$G_0(t) = \Theta_0 a_0(t) + \Theta_1 a_1(t)$$
$$G_1(t) = \Theta_0 a_1(t) + \Theta_1 a_2(t).$$

This results in the local linear estimator $\hat{\Theta}_0 = \hat{\alpha}_{2,W}(t)$, where
\[ \hat{\alpha}_{2,W}(t) = \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) dN_{i}(s), \]

with

\[ K_{t,b}(t-s) = \frac{a_2(t)K_{b}(t-s) - a_1(t)K_{b}(t-s)(t-s)}{a_0(t)a_2(t) - \{a_1(t)\}^2} W(s). \]

Note that

\[ \int_{0}^{T} K_{t,b}(t-s) Y^{(n)}(s) ds = 1, \quad \int_{0}^{T} K_{t,b}(t-s)(t-s) Y^{(n)}(s) ds = 0, \]

\[ \int_{0}^{T} K_{t,b}(t-s)(t-s)^2 Y^{(n)}(s) ds > 0. \]

Therefore \( K_{t,b} \) can be interpreted as a second order kernel with respect to the measure \( \mu \), where \( d\mu(s) = Y^{(n)}(s) ds \). We use this fact in deriving the pointwise asymptotic bias below.

**Remark 1.** The local linear estimator with the Ramlau-Hansen weighting function \( R(s) = nJ_{2}(s)\{Y^{(n)}(s)\}^{-1} \) equals Ramlau-Hansen’s estimator on internal intervals. In our case with estimation interval \((0, T)\), the local linear estimator with weighting function \( R(s) \) corresponds to Ramlau-Hansen’s estimator with boundary kernels \( k_{t,b} \) as defined in appendix 1. If a fixed boundary is considered these boundary kernels comply with the definition of boundary kernels in Gasser & Müller (1979). Andersen et al. (1993, p. 250) arrived at these boundary kernels and Jones (1993, p. 138) derived them from a local linear principle in the context of density estimation. Nielsen (1998b) used exactly this estimator in an extensive simulation study comparing what this paper would call the local constant and the local linear versions of Ramlau-Hansen’s estimator with and without multiplicative bias correction. The local linear estimator with the natural weighting function \( W(s) = 1 \) results in an estimator which perhaps has a more natural form. As pointed out above, the exposure robustness of this estimator makes us recommend this weighting instead of the Ramlau-Hansen weighting. The natural weighting is also the closest analogue of our local linear estimators to the traditional approach of local linear regression; see Fan & Gijbels (1996).

We now turn to the pointwise asymptotic properties of our class of local linear estimators. The details of the proof are deferred to appendix 2. Let

\[ \alpha_{2,W}^{x}(t) = \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) dA_{i}(s). \]

The pointwise asymptotic distribution of \( \alpha_{2,W}(t) \) is analysed by splitting the error \( \hat{\alpha}_{2,W}(t) - \alpha(t) \) into the variable part \( V_{2,W}(t) = \hat{\alpha}_{2,W}(t) - \alpha_{2,W}(t) \) and the stable part \( B_{2,W}(t) = \alpha_{2,W}(t) - \alpha(t) \).

**Theorem 2**

Assume that the general conditions GA hold and that \( \alpha \in C_{2}[0, 1] \). Then

\[ (nb)^{1/2} V_{2,W}(t) \Rightarrow N\{0, \sigma_{2,W}^{2}(t)\} \]

and

\[ B_{2,W}(t) = b^2 m_{2,W}(t) + o_P(b^2), \]

where \( m_{2,W}(t) = \kappa_{0} \{ \alpha'(t)/2 \} \) and \( \sigma_{2,W}^{2}(t) = \kappa_{2} \alpha(t) \{ \gamma(t) \}^{-1} \).
Remark 2. Hjort (1992) introduced the local likelihood kernel hazard estimation principle. The local likelihood principle was considered in more detail in the local likelihood kernel density estimation paper of Hjort & Jones (1996). It is interesting to notice that the local likelihood principle up to first order is identical to our least squares principle with the weighting

\[ W(s) = \{\alpha(s)\}^{-1}. \]

This fact implies that our least squares local linear estimator has the same asymptotic properties as the local linear version of the local likelihood kernel hazard estimator of Hjort (1992).

6. Multiplicative bias correction

The multiplicative bias correction was introduced for kernel hazard estimation in Nielsen (1998b), where the Ramlau-Hansen kernel hazard estimator was used as the starting point. In this section we show that the multiplicative correction can be constructed using the general minimization principle introduced in section 4 and section 5. We also generalize the multiplicative type of bias correction to work on any linearized estimator as we defined it in section 3.

We adopt the notation from section 3 and consider as pilot estimator a general estimator of the form

\[ \hat{\alpha}(t) = Y_t + B_t + \alpha(t), \]

with linearized variable term

\[ V_t = \sum_{i=1}^{n} \int_{0}^{T} h'_i(s) dM_i(s), \]

and where the set of functions used in the linearization is \( (h'_i) \).

The two estimators \( \hat{\alpha}_{3,W}, \hat{\alpha}_{4,W} \) will be given below as examples of this general principle (applying two different pilot estimators). The two pilot estimators are the local linear kernel hazard estimator of section 5 and the one time multiplicatively corrected estimator. This latter estimator is called the double multiplicatively corrected kernel hazard estimator.

Before turning to these examples, let us first consider the local linear minimization aiming at estimating the multiplicative error \( g_M(t) = \alpha(t)\{\hat{\alpha}(t)\}^{-1} \) of the pilot estimator, \( \hat{\alpha}(t) \).

Let

\[ \left( \hat{\Theta}_0 \right) = \arg \min \sum_{i=1}^{n} \int_{0}^{T} \left[ \Delta N_i(s) - \{\Theta_0 - \Theta_1(t - s)\} \hat{\alpha}(s)\right]^2 K_{\hat{\alpha}}(t - s) W(s) Y_i(s) ds. \]

Then it follows from the results in section 5 on local linear estimation that \( \hat{\Theta}_0 = \hat{g}_M(t) \), where

\[ \hat{g}_M(t) = \sum_{i=1}^{n} \int_{0}^{T} K_{s,b}(t - s) \{\hat{\alpha}(s)\}^{-1} dN_i(s), \]

and \( K_{s,b} \) is constructed with the weighting function

\[ \hat{W}(s) = \{\hat{\alpha}(s)\}^{2} W(s). \]

We call \( \hat{g}_M(t) \) the local linear estimator of the multiplicative error, \( g_M(t) = \alpha(t)\{\hat{\alpha}(t)\}^{-1} \). The multiplicative bias corrected estimator of the general pilot estimator \( \hat{\alpha}(t) \) is

\[ \hat{\alpha}_M(t) = \hat{\alpha}(t) \hat{g}_M(t). \]
Still using the general notation from section 3 on the variable part and the stable part, we get that the variable part and stable part of \( \hat{\alpha}_M(t) \) is called \( V_M(t) \) and \( B_M(t) \), such that

\[
\hat{\alpha}_M(t) = V_M(t) + B_M(t) + \alpha(t).
\]

The calculation in appendix 3 shows that

\[
V_M(t) = \sum_{i=1}^{n} \int_{0}^{T} f^i(s) dM_i(s),
\]

where

\[
f^i(s) = h^i(s) + \overline{K}_{t,b}(t-s)\overline{a}(t)\{\overline{a}(s)\}^{-1} - \int_{0}^{T} \overline{K}_{t,b}(t-u)h^i(s)\overline{a}(t)\{\overline{a}(u)\}^{-1} Y^{(n)}(u) du,
\]

and

\[
B_M(t) = \int_{0}^{T} \overline{K}_{t,b}(t-s)\overline{a}(t)\{\delta(t) - \delta(s)\} Y^{(n)}(s) ds,
\]

where

\[
\delta(t) = B_t\{\overline{a}(t)\}^{-1},
\]

and where \( B_t \) is the stable term of our general pilot estimator, \( \overline{a}(t) \), as defined above and in section 3. In the two examples below we consider two particular choices of pilot estimators.

**Remark 3.** It is seen that the local linear multiplicative bias correction method is computationally attractive. It more or less amounts to running the local linear estimation procedure twice.

**Remark 4.** If the preliminary estimator is carried out with the Ramlau-Hansen weighting and the second step with the weighting \( W(s) = \{\overline{a}(s)\}^{-1} R(s) \), we would arrive at the estimator considered in Nielsen (1998b).

**Example 1.** The multiplicative bias correction of local linear kernel hazard estimators. In this example we take the preliminary estimator, \( \overline{a}(t) \), to be the local linear hazard estimator, \( \hat{\alpha}_2(s) \). The third estimator, \( \hat{\alpha}_{3,W}(t) \), is the resulting multiplicatively corrected estimator, \( \hat{\alpha}_M(t) \). In this case we get that

\[
h^i_j(s) = K_{t,b}(t-s).
\]

With

\[
f_{1,3,W}(s) = K_{t,b}(t-s)\{1 + \overline{a}(t)\{\overline{a}(s)\}^{-1}\}
\]

\[
- \int_{0}^{T} \overline{K}_{t,b}(t-u)K_{u,b}(u-s)\overline{a}(t)\{\overline{a}(u)\}^{-1} Y^{(n)}(u) du.
\]

we therefore get that in this particular case,

\[
f^i_j(s) = f_{1,3,W}(s).
\]

Let

\[
\tilde{f}_{1,3,W}(s) = \Gamma_{K_b}(t-s)\{Y^{(n)}(s)\}^{-1},
\]

where

\[
\Gamma_{K_b}(u) = 2K_b - K_b * K_b(u),
\]

\( \odot \) Board of the Foundation of the Scandinavian Journal of Statistics 2001.
where $K_b$ is defined in the beginning of section 4. Note that

$$\Gamma_K(u) = 2K - K * K(u)$$

is the fourth order kernel obtained by twicing (Stuetzle & Mittal, 1979).

The asymptotic normality of $V_{3,W}(t)$ is derived using a standard technique, where the key element is to approximate

$$\sum_{i=1}^{n} \int_{0}^{T} f_{i,3,W}(s) dM_i(s)$$

with

$$\sum_{i=1}^{n} \int_{0}^{T} \hat{f}_{i,3,W}(s) dM_i(s)$$

and then use the fact that $\hat{f}_{i,3,W}$ is a predictable process. This follows exactly the same routine as the proof of th. 1 in Nielsen (1998b). For this reason the details of this variance approximation are omitted here. Theorem 3 below states the pointwise asymptotic theory for the multiplicatively bias corrected estimator

$$\hat{\alpha}_{3,W}(t) = V_{3,W}(t) + B_{3,W}(t) + \alpha(t),$$

where the bias considerations follow from the general bias calculation in the introduction that implies that

$$B_{3,W}(t) = \int_{0}^{T} \bar{K}_{t,b}(t-s) \hat{\alpha}_{2,W}(t) \{ \delta_{3,W}(t) - \delta_{3,W}(s) \} Y^{(n)}(s) \, ds,$$

where

$$\delta_{3,W}(t) = B_{2,W}(t) \{ \hat{\alpha}_{2,W}(t) \}^{-1}.$$

Using the fact that

$$\int_{0}^{T} \bar{K}_{t,b}(t-s) Y^{(n)}(s) \, ds = 1,$$

the asymptotic expansion of $B_{3,W}(t)$ is straightforward. We are therefore ready to state theorem 3 below.

**Theorem 3**

Suppose that $nJ^{(n)} / Y^{(n)} \stackrel{P}{\rightarrow} 1/\gamma$ uniformly in a neighbourhood of a given $t \in [0, T]$ as $n \rightarrow \infty$, that $\alpha \in C_4([0, T])$ and that $nb^3 \rightarrow \infty$.

Then

$$(nb)^{1/2} V_{3,W}(t) \Rightarrow N\{0, \sigma_{3,W}^2(t)\}$$

and

$$B_{3,W}(t) = \frac{1}{4} b^4 \kappa_0^2 \alpha(t)(\alpha''/\alpha')^2(t) + o_P(b^4),$$

where $\sigma_{3,W}^2(t) = \alpha(t) \{ \gamma(t) \}^{-1} \int_{-1}^{1} \Gamma_k^2(v) \, dv$.

**Example 2.** The double multiplicative bias correction of local linear kernel hazard estimators. In this second example we simply take the multiplicative estimator, $\hat{\alpha}_{3,W}$, as our preliminary
estimator, $\hat{\alpha}(t)$. We call the corresponding multiplicatively corrected estimator, $\hat{\alpha}_M(t)$, for our fourth estimator, $\hat{\alpha}_{4,W}(t)$. In this case we get

$$h_i'(s) = f_{i,3,W}(s).$$

With

$$f_{i,4,W}(s) = f_{i,3,W}(s) + K_{i,b}(t-s)\hat{\alpha}(t)\{\hat{\alpha}(s)\}^{-1}$$

$$- \int_{0}^{T} K_{i,b}(t-u)f_{u,3,W}(s)\hat{\alpha}(t)\{\hat{\alpha}(u)\}^{-1}Y^{(n)}(u) \, du.$$  

we therefore get that in this particular case

$$f_i'(s) = f_{i,4,W}(s).$$

Now $f_{i,4,W}(s)$ can be approximated with

$$\tilde{f}_{i,4,W}(s) = \mathbb{F}_{Kb}(t-s)\{Y^{(n)}(s)\}^{-1},$$

where we get the sixth order kernel $\mathbb{F}_K$ is defined by

$$\mathbb{F}_K(u) = K + \Gamma_K - K \ast \Gamma_K$$

and therefore also

$$\mathbb{F}_{Kb}(u) = K_b + \Gamma_{Kb} - K_b \ast \Gamma_{Kb}.$$

$\Gamma_K$ and $\Gamma_{Kb}$ are defined in example 1 above. Notice that the kernel $\mathbb{F}_K$ is exactly the same kernel as Hössjer & Ruppert (1995) arrived at while using their transformation bias reducing technique twice. Theorem 4 below states the pointwise asymptotic theory for the double multiplicatively bias corrected estimator

$$\hat{\alpha}_{4,W}(t) = V_{4,W}(t) + B_{4,W}(t) + \alpha(t),$$

where the variance considerations follow as in example 1 and the bias considerations follow from the general bias calculation in the introduction that implies that

$$B_{4,W}(t) = \int_{0}^{T} K_{i,b}(t-s)\hat{\alpha}_{3,W}(t)\{\delta_{4,W}(t) - \delta_{4,W}(s)\}Y^{(n)}(s) \, ds,$$

where $\delta_{4,W}(t) = B_{3,W}(t)\{\hat{\alpha}_{3,W}(t)\}^{-1}$. We are now ready to state theorem 4.

**Theorem 4**

Suppose that $nJ^{(n)}/Y^{(n)} \stackrel{P}{\rightarrow} 1/\gamma$ uniformly in a neighbourhood of a given $t \in [0, T]$ as $n \rightarrow \infty$ that $\alpha \in C_6([0, T])$ and that $nb^3 \rightarrow \infty$.

Then

$$(nb)^{1/2}V_{4,W}(t) \Rightarrow N\{0, \sigma_{4,W}^2(t)\}$$

and

$$B_{4,W}(t) = \frac{1}{8} b^6 K_b^2 \alpha(t)(\alpha''/\alpha)^{(n)}(t) + o_P(b^6),$$

where $\sigma_{4,W}^2(t) = \alpha(t)\{\gamma(t)\}^{-1} \int_{-1}^{1} \mathbb{F}'_K(v) \, dv$.

7. Additive bias correction

In this section we introduce a new bias reducing technique based on additive bias correction. It
is decided locally how much weight is given to an estimator of the bias of a pilot estimator that subsequently is subtracted from the pilot estimator to give the final estimator. The theoretical asymptotic variance of this estimator equals a noise component which is due to the estimation of the pilot estimator of the bias used plus the asymptotic variance known from the multiplicative bias correction case. The theoretical bias of the additively corrected estimator is of lower order than in the multiplicative bias correction case, namely $O_p(b^\alpha)$. The method is as general as the multiplicative bias correction method in the sense that it works for any generalized linear estimator, $\bar{\alpha}$. This generalized pilot estimator is used exactly the same way as in the case of multiplicative bias correction. We therefore again consider some generalized pilot estimator

$$\tilde{\alpha}(t) = V_t + B_t + \alpha(t),$$

where the variable part

$$V_t = \sum_{i=1}^{n} \int_{0}^{T} h_i'(s) dM_i(s)$$

is linearized by the functions $(h_i')$. We define the general additive bias correction principle for this general pilot estimator, $\tilde{\alpha}$, using some estimator $\hat{B}_t$ of the true bias $B_t$ of $\tilde{\alpha}$. In section 7.2 we describe a general bootstrap procedure to obtain a bias estimator $\hat{\sigma}_t$ of the true bias $B_t$. In section 7.3 we give the theoretical properties of the local additive bias correction of the local linear estimator.

### 7.1. The general additive bias correction procedure

The general bias estimation procedure is based on the local linear estimation principle. Let $\hat{B}_t$ be some estimator of the bias, $B_t$, of the pilot estimator $\tilde{\alpha}$.

Let

$$\left( \hat{\Theta}_0, \hat{\Theta}_1 \right) = \arg \min_{\Theta_0, \Theta_1} \sum_{i=1}^{n} \int_{0}^{T} \left[ AN_i(s) - \tilde{\alpha}(s) - \left( \Theta_0 - \Theta_1(t-s) \right) \hat{B}_s \right]^2 K_b(t-s) Y_i(s) \, ds.$$ 

$$= \arg \min_{\Theta_0, \Theta_1} \sum_{i=1}^{n} \int_{0}^{T} \left[ \hat{B}_s^{-1} A N_i(s) - \hat{B}_s^{-1} \tilde{\alpha}(s) - \left( \Theta_0 - \Theta_1(t-s) \right) \right]^2 \hat{B}_s^2 K_b(t-s) Y_i(s) \, ds.$$ 

Now let $\hat{\Theta}_0 = \hat{g}_A(t)$, where

$$\hat{g}_A(t) = \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) d\tilde{N}_i(s)$$

and

$$d\tilde{N}_i(s) = \hat{B}_s^{-1} dN_i(s) - \hat{B}_s^{-1} \tilde{\alpha}(s) Y_i(s) \, ds,$$

and where $K_{t,b}$ is constructed with the weighting function

$$\tilde{W}(s) = \hat{B}_s^2.$$ 

The general local linear additive bias corrected estimator of the generalized linear pilot estimator $\tilde{\alpha}(t)$ is defined as

$$\hat{\alpha}_A(t) = \tilde{\alpha}(t) + \hat{g}_A(t) \hat{B}_t.$$
Remark 5. It is seen that the local linear additive bias correction method is almost as computationally attractive as the local linear multiplicative bias correction. It amounts to running the local linear estimation procedure twice with a redefinition of the occurrences based on a preliminary bias estimator in the second step (the multiplicative bias correction did not need such a preliminary bias estimator).

7.2. A bootstrap bias estimation procedure

In this section we give a general bias estimating procedure that is valid (in principle) for any estimator, $\hat{\alpha}(t)$, of the underlying hazard $\alpha(t)$. The estimation principle is based on bootstrapping the bias from the local linear estimator $\hat{\alpha}_{2,W}$. This works as follows. Let $\Psi$, be the functional of the underlying data that results in the estimator $\hat{\alpha}(t)$, i.e

$$\alpha(t) = \Psi_{1}\{(N_1, Y_1), \ldots, (N_n, Y_n)\}.$$

Furthermore, let $\overline{\alpha}(t)$ be the bootstrapped estimator of $\hat{\alpha}(t)$:

$$\overline{\alpha}(t) = \Psi_{b}\{(\hat{A}_1, Y_1), \ldots, (\hat{A}_n, Y_n)\},$$

where the $\hat{A}_i$s are the integrated local linear estimators of the observed counting processes:

$$\hat{A}_i(t) = \int_0^t \hat{\alpha}_{2,W}(s) Y_i(s) ds.$$

The final estimator of the bias is

$$\hat{B}_i = \overline{\alpha}(t) - \hat{\alpha}_{2,W}(t).$$

In the case of estimating the bias of the local linear estimator we get

$$\hat{B}_{2,W}(t) = \sum_{i=1}^{n} \int_0^{T} {K}_{i,b}(t-s) \{\hat{\alpha}_{2,W}(s) - \hat{\alpha}_{2,W}(t)\} Y_i(s) ds.$$

7.3. Pointwise asymptotic theory of the bias estimation procedure

We consider the bias corrected estimator based on the standard local linear estimator. It turns out that this additively bias corrected estimator has exactly the same variance properties as the one time multiplicatively bias corrected estimator plus some extra noise component due to the variability of the bias estimator. The additively bias corrected estimator does, however, have a better rate of convergence than the multiplicatively bias corrected estimator. Theoretically, at least, the additive bias correction therefore appears promising.

Example 3. The additive bias correction of local linear kernel hazard estimators. The fifth estimator, $\hat{\alpha}_{5,W}(t)$, is the additively bias corrected estimator with the local linear hazard estimator $\hat{\alpha}_{2,W}(t)$ as the pilot estimator. The derivation in appendix 4 yields

$$\hat{\alpha}_{5,W}(t) - \alpha(t) = \bar{\alpha}_{5,W}(t) + \bar{B}_{5,W}(t),$$

where
\[ V_{5,w}(t) = V_{2,w}(t) + \sum_{i=1}^{n} \int_0^T K_{t,b}(t-s)B_{2,w}(t)\left\{ \hat{B}_{2,w}(s) \right\}^{-1} dM_b(s) \]

\[ - \sum_{i=1}^{n} \int_0^T K_{t,b}(t-s)B_{2,w}(t)\left\{ \hat{B}_{2,w}(s) \right\}^{-1} V_{2,w}(s)Y_i(s) \, ds \]

and

\[ B_{5,w}(t) = - \frac{1}{2} \sum_{i=1}^{n} \int_0^T K_{t,b}(t-s)(t-s)^2 B_{2,w}(t)\left\{ \hat{B}_{2,w}(s) \right\}^{-1} \left\{ \hat{B}_{2,w}(s) - B_{2,w}(s) \right\} Y_i(s) \, ds \]

\[ + \frac{1}{2} \sum_{i=1}^{n} \int_0^T K_{t,b}(t-s)(t-s)^2 \left\{ \hat{B}_{2,w}(s) \right\}^{-1} B_{2,w}(s)Y_i(s) \, ds, \]

where \( t^* \) and \( t^{**} \) are between \( t \) and \( s \).

The analysis of \( V_{5,w}(t) \) is almost identical to the analysis of the variable term in the multiplicative bias correction case described in section 6. The analysis of \( B_{5,w}(t) \) is slightly more complicated since the variability due to estimating the bias of the pilot estimator has the same order of magnitude as \( V_{5,w}(t) \). The analysis of \( B_{5,w}(t) \) is closely tied to lemma 2 and the conclusions regarding the stable part are rather immediate. The slightly complicated thing is the volatility part stemming from bootstrapping the bias of the pilot estimator. While the second part of \( B_{5,w}(t) \) involving

\[ \hat{B}_{2,w}(t^{**}) - B_{2,w}(t^{**}) \]

only has variance of inferior order, then the first part of \( B_{5,w}(t) \) involving

\[ \hat{B}_{2,w}(t^*) - B_{2,w}(t^*) \]

does in fact add to the volatility. It follows from lemma 2 that the volatility of

\[ \hat{B}_{2,w}(t) - B_{2,w}(t) \]

is asymptotically equivalent to the volatility \( \bar{V}_{2,w}^{(i)}(t) \) that is defined in appendix 5. We call the additional variance due to the first bias term for \( V_{5,w}^B \) and note that

\[ V_{5,w}^B(t) = \frac{1}{2} \kappa_0 b^2 \bar{V}_{2,w}^{(i)}(t) + o_P\{ (nb)^{-1/2} \} \]

since

\[ \sum_{i=1}^{n} \int_0^T K_{t,b}(t-s)Y_i(s) \, ds = \{ 1 + o_P(1) \} \kappa_0 b^2. \]

This follows from the considerations in appendix 1. Since \( \bar{V}_{2,w}^{(i)} \) is of the order of magnitude \( (nb^5)^{-1/2} \), the volatility \( V_{5,w}^B(t) \) is of the order of magnitude \( (nb)^{-1/2} \), which equals the order of magnitude of the variance term \( V_{5,w}(t) \).

We therefore get the final definitions:

\[ V_{5,w}(t) = V_{5,w}(t) + V_{5,w}^B \]

and

\[ B_{5,w}(t) = B_{5,w}(t) - V_{5,w}^B. \]
With these definitions we have that
\[ \hat{\alpha}_5 (t) - \alpha(t) = \bar{V}_5 (t) + \bar{B}_5 (t) = V_5 (t) + B_5 (t). \]

In theorem 5 below the asymptotic normality of \( V_5 (t) \) is derived using the same technique as in multiplicative bias correction (see Nielsen, 1998b). While deriving the asymptotic properties of the stable terms of \( B_5 (t) \), it is also useful that
\[ \sum_{i=1}^{n} \int_0^T \bar{K}_{i,t}(t-s)Y_i(s)ds = \{1 + o_P(1)\} \kappa_0 b^2. \]

The rest of the bias consideration follows from the extensive expansion given in appendix 4 and in lemma 1 and 2.

One of the key elements in the variance consideration of Nielsen (1998b) is the solution to the predictability issue. This is because the variable part is written as a sum of non-predictable integrands with respect to counting process martingales. The solution to the predictability issue used in Nielsen (1998b) ensures that standard martingale theory can be used as if the integrands had been predictable. For more details on the predictability issue, see Nielsen et al. (1998) and Nielsen (1999).

**Theorem 5**
Assume that the general conditions GA hold and assume that the kernel \( K \) is 4 times continuously differentiable and that \( \alpha \in C_0[0, T] \).

Then
\[ (nb)^{-1/2} V_5 (t) \Rightarrow N \{0, \sigma^2_5 (t) \} \]
and
\[ B_5 (t) = \frac{1}{8} \kappa_0 b^6 \left[ -\alpha^{(iv)}(t) + \{\alpha^{(iv)}(t)\}^2 \{\alpha''(t)\}^{-1} \right] + o_P(b^6), \]

where
\[ \sigma^2_5 (t) = \alpha(t) \gamma(t)^{-1} \left[ \int_{-1}^{1} \{\Gamma_K (v) - \frac{1}{2} \kappa_0 (K - K * K)^{(iv)}(v)\}^2 dv \right], \]

where
\[ \Gamma_K (u) = 2K(u) - K * K(u) \]
also is defined in example 1.

8. A simulation study
In this section we conduct a Monte Carlo simulation study of the five estimators, \( \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \hat{\alpha}_5 \) as summarized in Table 1. The estimators use the kernel function
\[ K(x) = (1 - x^2)^6, \]
where we have left out the normalization constant.

We use as the true hazard one of the four functions:
\[ \gamma_1(t) = B(t, 2, 2), \]
\[ \gamma_2(t) = B(t, 4, 4), \]
\[ \gamma_3(t) = 0.6 \times [B(t, 0.5, 0.5) + B(t, 7, 7)], \]
\[ \gamma_4(t) = 0.6 \times [B(t, 0.5, 0.5) + B(t, 4, 2) + B(t, 2, 4)], \]

where \( B(t, \alpha, \beta) \) is for \( t \in (0, 1) \) the value of the density of the beta distribution with parameters \( \alpha, \beta \) (see Nielsen, 1998b for a graph of the four functions). These Beta functions have been chosen because of their flexible forms.

Each simulation run is constructed as follows. First, we define a discrete grid on the interval \((0, 1)\) with gridlength, \( \delta_M = 1/(M + 1) \), as \( \{ t_j : t_j = j \delta_M, j = 1, \ldots, M \} \). Then for a sample of \( n \) individuals, failures at time \( t_k \) are generated from the binomial distribution, \( \text{bi}[Y^{\text{in}}(t_k), \gamma_j(t_k) \delta_M] \). We tried several values of \( M \) of which we report only the results for \( M = 100 \). Higher values of \( M \) do not seem to alter the conclusions.

To evaluate the simulations we use the following global measure of estimation error for each true hazard function \( \gamma_1, \ldots, \gamma_4 \)

\[ \text{err}(\hat{\alpha}_h) = n^{-1} \sum_{i=1}^{n} \int_0^1 [\hat{\alpha}_h(s) - \gamma_h(s)]^2 Y_i(s) \, ds \]

for \( h \in \{ 1, 2, 3, 4, 5 \} \).

The simulation study is based on the best possible bandwidth measured by \( \text{err}(\hat{\alpha}_h) \). This separates the issue of estimator quality (conditional on bandwidth) from the bandwidth selection problem. Clearly our error measure are extreme values, which are unattainable in applied work. This parallels the type of simulation study carried out in Jones et al. (1995), Jones & Signorini (1997), Nielsen (1998b) and Jones et al. (1999). For some more comments on this type of simulation study and its relation with the bandwidth selection problem, see Jones et al. (1999).

Regarding bandwidth selection for one-dimensional kernel hazard estimation Nielsen (1990) gave an extensive theoretical treatment of the cross-validation principle originally suggested by Ramlau-Hansen (1981) in the unpublished part of his master thesis. Nielsen (1990) showed that cross-validation in kernel hazard estimation has exactly the same type of theoretical performance as in the i.i.d. kernel density case. It is, however, beyond the scope of this paper to pursue this issue any further. Nonetheless, we expect cross-validation to work well for the estimators studied here.

Table 2 shows the simulated estimation errors for sample size \( n = 100, n = 1,000, n = 10,000 \) and \( n = 100,000 \). Table 3 presents the estimation errors as ratios, \( 100 \times \frac{\text{err}(\hat{\alpha}_i)}{\text{err}(\hat{\alpha}_1)} \), \( i = 1, \ldots, 6 \). We did the simulations for the natural weighting as well as Ramlau-Hansen’s weighting. The results are very similar, and we only report the results for the natural weighting.

As expected, the optimal amount of bias reduction increases with \( n \). The local constant
estimator, the local linear estimator and the multiplicative bias correction methods are almost of the same quality for $n \approx 100$, where the local constant estimator is best for the complicated hazards $\gamma_3$ and $\gamma_4$, whereas bias reduction is already appropriate for the simpler hazards $\gamma_1$ and $\gamma_2$. From $n = 1000$ the multiplicative bias reduction should be used, but the local additive bias correction is competitive and outperforms the others for $n \approx 10000$ and $n \approx 100000$. The two times multiplicatively bias corrected estimator is not interesting. In the cases where it is better than the just one time multiplicatively bias corrected estimator, it is not as good as the local additively bias corrected estimator. We also tried a local additive bias correction estimator based on the estimator of bias resulting from subtracting the multiplicatively bias corrected estimator from the local linear estimator. This bias estimator turned out to do worse in all cases than the bootstrapped bias estimator presented here. Also, the two times local additively bias corrected estimator was not as good as the one time local additive bias reduced estimator for any of the hazards or sample sizes considered. For this reason we do not report the results for the alternative bias estimator and the double additive bias corrected estimator. These findings make us conclude that it is not worthwhile to do bias correction more than once when the sample is less than $n \approx 100000$.

Table 4 presents the results of a robustness study. The idea is to use a very volatile exposure pattern. The pattern assumes that the exposure has the value 1 in the intervals $0.08 - 0.09, 0.18 - 0.19, \ldots, 0.98 - 1.00$, whereas the occurrence is zero in the same intervals. In practice this type of data can be thought

<table>
<thead>
<tr>
<th>$n = 100$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>3.745</td>
<td>3.978</td>
<td>6.851</td>
<td>3.822</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>2.791</td>
<td>3.288</td>
<td>7.203</td>
<td>4.250</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>2.449</td>
<td>3.027</td>
<td>7.169</td>
<td>5.280</td>
</tr>
<tr>
<td>$\hat{\alpha}_4$</td>
<td>3.017</td>
<td>3.652</td>
<td>7.327</td>
<td>6.425</td>
</tr>
<tr>
<td>$\hat{\alpha}_5$</td>
<td>2.539</td>
<td>3.164</td>
<td>7.115</td>
<td>10.118</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 1000$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>0.800</td>
<td>0.605</td>
<td>1.922</td>
<td>1.083</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>0.432</td>
<td>0.551</td>
<td>1.359</td>
<td>0.976</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>0.377</td>
<td>0.425</td>
<td>1.198</td>
<td>0.989</td>
</tr>
<tr>
<td>$\hat{\alpha}_4$</td>
<td>0.428</td>
<td>0.481</td>
<td>1.124</td>
<td>1.073</td>
</tr>
<tr>
<td>$\hat{\alpha}_5$</td>
<td>0.318</td>
<td>0.565</td>
<td>1.225</td>
<td>1.111</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 10000$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>0.138</td>
<td>0.104</td>
<td>0.513</td>
<td>0.462</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>0.060</td>
<td>0.099</td>
<td>0.275</td>
<td>0.328</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>0.048</td>
<td>0.063</td>
<td>0.234</td>
<td>0.332</td>
</tr>
<tr>
<td>$\hat{\alpha}_4$</td>
<td>0.056</td>
<td>0.070</td>
<td>0.226</td>
<td>0.345</td>
</tr>
<tr>
<td>$\hat{\alpha}_5$</td>
<td>0.040</td>
<td>0.075</td>
<td>0.185</td>
<td>0.237</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 100000$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>0.025</td>
<td>0.017</td>
<td>0.114</td>
<td>0.169</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>0.010</td>
<td>0.017</td>
<td>0.073</td>
<td>0.076</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>0.007</td>
<td>0.009</td>
<td>0.050</td>
<td>0.070</td>
</tr>
<tr>
<td>$\hat{\alpha}_4$</td>
<td>0.008</td>
<td>0.009</td>
<td>0.052</td>
<td>0.079</td>
</tr>
<tr>
<td>$\hat{\alpha}_5$</td>
<td>0.005</td>
<td>0.008</td>
<td>0.033</td>
<td>0.037</td>
</tr>
</tbody>
</table>

The table shows averages over 100 simulation runs of $\text{err}(\hat{\alpha}_k) = n^{-1} \sum_{i=1}^{n} \left[ \hat{\alpha}_k(s) - \gamma(s) \right]_0^s Y(s) ds$. The bandwidth is chosen as the best possible bandwidth. The weighting scheme is the natural weighting.

The table shows averages over 100 simulation runs of $\text{err}(\hat{\alpha}_k) = n^{-1} \sum_{i=1}^{n} \left[ \hat{\alpha}_k(s) - \gamma(s) \right]_0^s Y(s) ds$. The bandwidth is chosen as the best possible bandwidth. The weighting scheme is the natural weighting.

Table 4 presents the results of a robustness study. The idea is to use a very volatile exposure pattern. The pattern assumes that the exposure has the value 1 in the intervals

$0.08 - 0.10, 0.18 - 0.20, \ldots, 0.98 - 1.00$, whereas the occurrence is zero in the same intervals. In practice this type of data can be thought
of as occurring when no observations are made in these intervals. The low level of exposure left in these intervals may arise from a data error. It turns out that the estimators based on Ramlau-Hansen’s weighting completely break down in this situation, whereas the natural weighting does relatively well.

Table 5 presents the decrease in estimation error due to a ten times increase in data. The numbers are presented relative to the local constant estimator. The results are in accordance with

<table>
<thead>
<tr>
<th>$n$ = 100</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>74.5</td>
<td>82.6</td>
<td>105.1</td>
<td>111.2</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>65.4</td>
<td>76.1</td>
<td>104.6</td>
<td>138.2</td>
</tr>
<tr>
<td>$\hat{\alpha}_4$</td>
<td>80.6</td>
<td>91.8</td>
<td>107.0</td>
<td>168.1</td>
</tr>
<tr>
<td>$\hat{\alpha}_5$</td>
<td>67.8</td>
<td>79.5</td>
<td>103.9</td>
<td>264.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$ = 1000</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>54.0</td>
<td>91.1</td>
<td>70.7</td>
<td>90.1</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>47.0</td>
<td>70.2</td>
<td>62.3</td>
<td>91.3</td>
</tr>
<tr>
<td>$\hat{\alpha}_4$</td>
<td>53.5</td>
<td>79.6</td>
<td>58.5</td>
<td>99.0</td>
</tr>
<tr>
<td>$\hat{\alpha}_5$</td>
<td>39.7</td>
<td>93.4</td>
<td>63.7</td>
<td>102.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$ = 10 000</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>43.5</td>
<td>95.9</td>
<td>53.6</td>
<td>71.0</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>35.0</td>
<td>60.9</td>
<td>45.7</td>
<td>72.0</td>
</tr>
<tr>
<td>$\hat{\alpha}_4$</td>
<td>40.4</td>
<td>67.1</td>
<td>44.0</td>
<td>74.7</td>
</tr>
<tr>
<td>$\hat{\alpha}_5$</td>
<td>29.0</td>
<td>72.2</td>
<td>36.0</td>
<td>51.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$ = 100 000</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>41.4</td>
<td>97.5</td>
<td>63.8</td>
<td>45.1</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>26.8</td>
<td>52.6</td>
<td>43.6</td>
<td>41.3</td>
</tr>
<tr>
<td>$\hat{\alpha}_4$</td>
<td>30.4</td>
<td>54.6</td>
<td>45.3</td>
<td>46.8</td>
</tr>
<tr>
<td>$\hat{\alpha}_5$</td>
<td>19.6</td>
<td>48.4</td>
<td>29.0</td>
<td>21.8</td>
</tr>
</tbody>
</table>

The table shows averages over 100 simulation runs of the relative estimation error, $\text{err}(\hat{\alpha}_k)/\text{err}(\hat{\alpha}_1)$. The bandwidth is chosen as the best possible bandwidth. The weighting scheme is the natural weighting.

The table shows averages over 100 simulation runs of $\text{err}(\hat{\alpha}_k) = n^{-1}\sum_{i=1}^{n} \int_{[\gamma_1(s) - \gamma_2(s)]^2 Y_i(s) ds}$ with $n = 1000$. The simulations differ from Table 2 with respect to the exposure pattern. The pattern assumes that the exposure has the value 1 in the intervals

$0.08 - 0.10, 0.18 - 0.20, \ldots, 0.98 - 1.00$,

whereas the occurrence is zero in the same intervals. The left hand panel is based on the Ramlau-Hansen weighting whereas the right hand panel is based on the natural weighting. The bandwidth is best possible.
the asymptotic theory, since the estimators with the high rates of convergence have the largest relative improvements.

Acknowledgements

We thank the editors and two anonymous referees for useful comments and suggestions. The usual disclaimer applies. Research is supported by The Danish Social Science Research Council. Jens Perch Nielsen also received support for research from Bergiafonden.

References


Received October 1997, in final form December 2000

Jens Perch Nielsen, Codan, Gammel Kongevej 60, DK 1790, Copenhagen V, Denmark.

## Appendices

The appendices consist of proofs and technical details. Appendix 1 gives the asymptotic properties of the local linear kernels that is needed while deriving asymptotic properties of the local linear estimator. It is also critical for the derivation of the properties of the derivatives of the local linear estimator and of the bias estimators discussed in appendix 5 and applied in the section 7 on additive bias correction. In appendix 2 we outline the asymptotic properties of the general local linear estimator and in appendix 3 we give the technical calculations leading to the asymptotic theory of the multiplicative bias reduction method. In appendix 4 the technicalities linked to the additive bias correction method are given. These depend on the rate of convergence of the derivatives considered in appendix 5.

### Appendix 1. Asymptotic properties of the local linear kernels

Consider first the deterministic local linear kernel that results from Ramlau-Hansen’s weighting:

\[
 k_{t,b}(t-s) = \frac{c_2(t)K_b(t-s) - c_1(t)K_b(t-s)(t-s)}{c_0(t)c_2(t) - \{c_1(t)\}^2},
\]

where

\[
 c_j(t) = \int_0^T K_b(t-s)(t-s)^j \, ds
\]

for \( j = 0, 1, 2 \). This is the same kernel as Jones (1993, p. 138) arrived at in the case of local linear kernel density estimation. When \( b \to 0 \) we get that \( c_1(t) = 0 \) and \( c_0(t) = 1 \) for sufficiently large \( n \), lets say for \( n \geq N_0 \) for some natural number \( N_0 \). This in turn means that \( k_{t,b}(t-s) = K_b(t-s) \) for \( n \geq N_0 \) which eases the asymptotic theory.

Now let us turn our attention to local linear kernels constructed with the general weighting function \( W \):

\[
 K_{t,b}(t-s) = \frac{a_2(t)K_b(t-s) - a_1(t)K_b(t-s)(t-s)}{a_0(t)a_2(t) - \{a_1(t)\}^2} W(s).
\]

By a change of variables we get that
In this section we prove the asymptotic properties stated in theorem 2. First note that the bias

\[ a_j(t) = \int_0^t K_h(t-s)(t-s)^j W(s)Y^{(n)}(s) \, ds = b^j \int K(u)u^j W(t-bu)Y^{(n)}(t-bu) \, du \]

\[ = b^j \int K(u)u^j \{ W(t-bu)Y^{(n)}(t-bu) - w(t-bu)ny(t-bu) \} \, du \]

\[ + b^j \int K(u)u^j w(t-bu)ny(t-bu) \, du \]

If we assume that the general conditions GA hold, then we can employ the local uniform convergence and the differentiability of the weighting and the exposure to obtain

\[ n^{-1} a_j(t) = \{ c_j(t)w(t)y(t) + c_{j+1}(t)(w^2y)'(t) \} \{ 1 + o_P(1) \}. \]

Using this fact makes it easy to convince oneself that the kernel \( K_{1,b}(t-s) \) is asymptotically equivalent to the kernel \( k_{1,b}(t-s)Y^{(n)}(s)^{-1} \).

For the additive bias correction method we need the parallel argument for the derivatives, see appendix 5. Let \( c_j^{(r)}(t) \) and \( a_j^{(r)}(t) \), for \( j = 0, 1, 2 \), be the \( r \)th derivative of \( c_j(t) \) and \( a_j(t) \) respectively. Using the same type of argumentation as above we get

\[ a_j^{(r)}(t) = \{ c_j^{(r)}(t)w(t)y(t) + \tilde{c}_j^{(r)}(t)(w^2y)'(t) \} \{ 1 + o_P(1) \}, \]

where

\[ \tilde{c}_j^{(r)}(t) = \int_0^t \{ K_h(t-s)(t-s)^j \}^{(r)}(t-s) \, ds. \]

Since \( \tilde{c}_j^{(r)}(t) \) has one more \( (t-s) \) in the integral that \( c_j^{(r)}(t) \) we get that \( \tilde{c}_j^{(r)}(t) = o_P(b^{j+1}) \) while \( c_j^{(r)}(t) = o_P(b^j) \). This is sufficient to ensure that the kernels \( K_{1,b}(t-s) \) and \( k_{1,b}(t-s)Y^{(n)}(s)^{-1} \) also are equivalent when it comes to estimating derivatives, see also appendix 5.

### Appendix 2. Pointwise asymptotic properties of the general local linear estimator

In this section we prove the asymptotic properties stated in theorem 2. First note that the bias term can be Taylor expanded as

\[ B_{2,W}(t) = \alpha_{2,W}^*(t) - \alpha(t) = \int_0^T K_{1,b}(t-s)\{ \alpha(s) - \alpha(t) \} Y^{(n)}(s) \, ds \]

\[ = \frac{1}{2} \int_0^T K_{1,b}(t-s)(t-s)^2 \alpha''(t_s^*)Y^{(n)}(s) \, ds \]

\[ = \frac{1}{2} \int_0^T k_{1,b}(t-s)(t-s)^2 \alpha''(t_s^*) \, ds + o_P(b^2) \]

\[ = \frac{1}{2} k_0 b^2 \alpha''(t) + o_P(b^2), \]

where \( t_s^* \) is between \( [t-b, t+b] \) and \( k_{1,b}(t-s) \) is the deterministic local linear kernel corresponding to the Ramlau-Hansen weighting, see appendix 1.

Regarding the variable part we get

\[ V_{2,W}(t) = \hat{\alpha}_{2,W}(t) - \alpha_{2,W}^*(t) = \sum_{i=1}^n \int_0^T K_{1,b}(t-s) \, dM_i(s) \]

is asymptotically equivalent to \( \sum_{i=1}^n \int_0^T k_{1,b}(t-s)Y^{(n)}(s)^{-1} \, dM_i(s) \), which again is asympto-
tically equivalent to $\sum_{i=1}^{n} \int_{0}^{T} K_{i}(t-s) \{ Y^{(n)}(s) \}^{-1} dM_{i}(s)$. Hence the proof of theorem 2 follows from standard martingale theory, see Ramlau-Hansen (1983). It is seen that all local linear estimators have the same pointwise asymptotic distribution independent of the weighting function $W$. It is also seen from the bias considerations that the boundary bias is of order of magnitude $O_{P}(b^{2})$ again independent of the weighting function $W$.

Appendix 3. The expansion of the multiplicatively corrected estimator

In this section we carry out the expansion of the general multiplicative estimator

$$\hat{a}_{M}(t) = V_{M}(t) + B_{M}(t) + \alpha(t).$$

with the general linearized pilot estimator $\tilde{\alpha}$

$$\tilde{\alpha}(t) = V_{t} + B_{t} + \alpha(t)$$

as described in sections 3 and 6. Without losing generality we can assume throughout that $Y^{(n)}(s) > 0$ for $s \in (t-2b, t+2b)$ and we get

$$\hat{a}_{M}(t) = \sum_{i=1}^{n} \int_{0}^{T} \bar{K}_{i,b}(t-s) \{ \tilde{\alpha}(s) \}^{-1} dN_{i}(s)$$

$$= \sum_{i=1}^{n} \int_{0}^{T} \bar{K}_{i,b}(t-s) \{ \tilde{\alpha}(s) \}^{-1} dM_{i}(s)$$

$$+ \sum_{i=1}^{n} \int_{0}^{T} K_{i,b}(t-s) \{ \alpha(s) - \tilde{\alpha}(s) \} \{ \tilde{\alpha}(s) \}^{-1} Y_{i}(s) \, ds$$

$$+ \sum_{i=1}^{n} \int_{0}^{T} K_{i,b}(t-s) \tilde{\alpha}(t) Y_{i}(s) \, ds.$$
Appendix 4. The expansion of the additively corrected estimator

In this section we give the technical derivation for the local additive bias correction method with the pilot estimator

$$\tilde{\alpha}(t) = V_t + B_t + \alpha(t)$$

and the final estimator $\tilde{\alpha}(t)$ based on the local additive bias correction principle described in section 7 with the bias estimator $B_t$ of the pilot estimators bias term $B_t$. Following section 7, we therefore have that

$$\tilde{g}_A(t) = \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) d\tilde{N}_i(s)$$

with

$$d\tilde{N}_i(s) = \tilde{B}_s^{-1} dN_i(s) - \tilde{B}_s^{-1} \tilde{a}(s) Y_i(s) ds,$$

where the weighting function corresponding to $K_{t,b}$ is given in section 7. We get the following expansion of the additively corrected estimator $\tilde{\alpha}_A(t)$, where we use that $\int_{0}^{T} K_{t,b}(t-s)Y(s)ds = 1$ and $\int_{0}^{T} K_{t,b}(t-s)(t-s)Y(s)ds = 0$:

$$\tilde{\alpha}_A(t) - \alpha(t) = \tilde{\alpha}_A(t) + \tilde{g}_A(t)\tilde{B}_t - \alpha(t) = V_t + B_t + \tilde{B}_t \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) d\tilde{N}_i(s)$$

$$= V_t + B_t + \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) \tilde{B}_t \tilde{B}_s^{-1} dM_i(s) - \sum_{i=1}^{n} \int_{0}^{T} \tilde{B}_t \tilde{B}_s^{-1} \tilde{a}(s) Y_i(s) ds$$

$$= V_t + B_t + \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) \tilde{B}_t \tilde{B}_s^{-1} dM_i(s)$$

$$- \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) \tilde{B}_t \tilde{B}_s^{-1} \{ V_s + B_s \} Y_i(s) ds$$

$$= V_t + \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) \tilde{B}_t \tilde{B}_s^{-1} dM_i(s) - \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) \tilde{B}_t \tilde{B}_s^{-1} V_s Y_i(s) ds$$

$$+ \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t-s) \{ B_t - \tilde{B}_t \tilde{B}_s^{-1} B_s \} Y_i(s) ds$$
\[
V_t + \sum_{i=1}^{n} \int_0^T \mathcal{K}_{t,b}(t-s) \hat{B}_t \hat{B}_s^{-1} dM_i(s) - \sum_{i=1}^{n} \int_0^T \mathcal{K}_{t,b}(t-s) \hat{B}_t \hat{B}_s^{-1} V_s Y_i(s) ds
\]
\[
+ \sum_{i=1}^{n} \int_0^T \mathcal{K}_{t,b}(t-s) B_t \hat{B}_s^{-1} \{(\hat{B}_s - B_s) - (\hat{B}_t - B_t)\} Y_i(s) ds
\]
\[
+ \sum_{i=1}^{n} \int_0^T \mathcal{K}_{t,b}(t-s) \hat{B}_s^{-1} \{(B_r - B_s)(\hat{B}_t - B_t)\} Y_i(s) ds
\]
\[
= V_t + \sum_{i=1}^{n} \int_0^T \mathcal{K}_{t,b}(t-s) \hat{B}_t \hat{B}_s^{-1} dM_i(s) - \sum_{i=1}^{n} \int_0^T \mathcal{K}_{t,b}(t-s) \hat{B}_t \hat{B}_s^{-1} V_s Y_i(s) ds
\]
\[
- \frac{1}{2} \sum_{i=1}^{n} \int_0^T \mathcal{K}_{t,b}(t-s)(t-s)^2 B_t \hat{B}_s^{-1} (\hat{B}_s^{(i)} - B_t^{(i)}) Y_i(s) ds
\]
\[
+ \frac{1}{2} \sum_{i=1}^{n} \int_0^T \mathcal{K}_{t,b}(t-s)(t-s)^2 \hat{B}_s^{-1} \{B_r^{(i)}(\hat{B}_t - B_t)\} Y_i(s) ds,
\]
where \(t^*\) and \(t^{**}\) are between \(t\) and \(s\).

This expansion is used to obtain the pointwise asymptotic theory of the additively corrected estimator with the local linear kernel hazard estimator as pilot estimator.

**Appendix 5. Uniform convergence in shrinking neighbourhood of derivative and bias estimators**

In this section we establish the uniform convergence in a local neighbourhood of derivatives of the local linear estimator. This result is essential for theoretical derivatives of the local linear additive bias correction described in section 7.3. The uniform convergence in a local neighbourhood has the same rate of convergence as the original estimator. The proof of the theorem is very similar to the proof of the uniform convergence that Ramlau-Hansen gives in Ramlau-Hansen (1983, p. 123) where a partial likelihood argument makes it possible to restrict the problem to the uniform convergence of the underlying Aalen estimator that is well-known to be uniformly square-root-\(n\) consistent. The only difference in our case is that this resulting Aalen estimator is considered over a shrinking interval. This fact can be employed to obtain the slightly improved rate of convergence compared to the rate obtained in prop. 3.2.1 in Ramlau-Hansen (1983, p. 123) that we present in lemma 1. For the kernel considerations leading to the control of the differentiated local linear estimators, see appendix 1.

**Lemma 1** (Uniform convergence of derivatives in shrinking neighbourhood)

Assume that the general conditions GA hold and that \(\alpha \in C_{r_0+2}(0, T)\) and assume that the kernel \(K\) is \(r_0\) times continuously differentiable. Then

\[
\sup_{s \in A_{t,b}} |\hat{\alpha}_I^{(n)}(s) - \alpha^{(r_0)}(s)| = \frac{1}{2} k_0 \alpha^{(r_0)}(t) b^2 + \hat{\nu}_I^{(n)}(t)
\]
\[
= \frac{1}{2} k_0 \alpha^{(r_0)}(t) b^2 + O_P\{nb^{2(r_0+1)}\}.
\]

In the local additive bias reduction example in section 7.3, the pilot estimator is the local linear estimator \(\hat{\alpha}_I\) with estimated bias.
\[
\hat{B}_{2,W}(t) = \sum_{i=1}^{n} \int_{0}^{T} K_{t,b}(t - s)\{\hat{a}_{2,W}(s) - \hat{a}_{2,W}(t)\}Y_i(s)\,ds.
\]

The asymptotic properties of this bias and its derivatives are essential for the derivation of the asymptotic properties of the local additive bias correction estimator considered in section 7.3. Let \(\hat{B}_{2,W}^{(r)}(t)\) be \(\hat{B}_{2,W}(t)\) differentiated \(r\) times and

\[
B^{(r)}(t) = \frac{1}{2} k_0 \alpha^{(r+2)}(t)b^2.
\]

Let, furthermore, \(V_{2,W}^{(r_0)}(t)\) be the variable part corresponding to \(V_{2,W}^{(r_0)}(t)\), but with the kernel \(K_{t,b}(t - s)Y_i(s)\) replaced by minus that very same kernel plus that kernel folded by itself. In the limit this kernel equals \(K * K - K\).

**Lemma 2** (Uniform convergence in shrinking neighbourhood of bias estimators)

Assume that the general conditions GA hold and assume that the kernel \(K\) is \(r_0 + 2\) times continuously differentiable and that \(\alpha \in C_{r_0+4}([0, T])\). Then

\[
\sup_{s \in \Delta_{t,b}} |\hat{B}_{2,W}^{(r_0)}(s) - B^{(r_0)}(s)| = \frac{1}{4} k_0^2 \alpha^{(r_0+4)}(t)b^4 + o_p(b^4) + \frac{1}{2} k_0 b^2 \int V_{2,W}^{(r_0+2)}(t)ds.
\]