Hiabu, Munir; Martínez-Miranda, María Dolores; Perch Nielsen, Jens; Spreeuw, Jaap;
Tanggaard, Carsten; Villegas, Andrés M.
Global Polynomial Kernel Hazard Estimation

Global Polynomial Kernel Hazard Estimation

Ajuste polinomial global para la estimación kernel de la función de riesgo

Munir Hiabu[1,a], María Dolores Martínez-Miranda[1,2,b],
Jens Perch Nielsen[1,c], Jaap Spreeuw[1,d], Carsten Tanggaard[3,e],
Andrés M. Villegas[1,f]

[1]Faculty of Actuarial Science and Insurance, Cass Business School, City University London, United Kingdom

[2]Department of Statistics and O.R., Faculty of Sciences, University of Granada, Spain

[3]CREATES, Aarhus University, Denmark

Abstract

This paper introduces a new bias reducing method for kernel hazard estimation. The method is called global polynomial adjustment (GPA). It is a global correction which is applicable to any kernel hazard estimator. The estimator works well from a theoretical point of view as it asymptotically reduces bias with unchanged variance. A simulation study investigates the finite-sample properties of GPA. The method is tested on local constant and local linear estimators. From the simulation experiment we conclude that the global estimator improves the goodness-of-fit. An especially encouraging result is that the bias-correction works well for small samples, where traditional bias reduction methods have a tendency to fail.

Key words: Kernel Estimation, Hazard Function, Local Linear Estimation, Boundary Kernels, Polynomial Correction.

Resumen

En este artículo se introduce un nuevo método de corección del sesgo para la estimación núcleo de la función de riesgo. El método, denominado ajuste polinomial global (APG), consiste en una corrección global que es aplicable a

[a]Ph.D. Student. E-mail: Munir.Hiabu.1@cass.city.ac.uk

[b]Associate Professor. E-mail: mmiranda@ugr.es

[c]Professor. E-mail: Jens.Nielsen.1@city.ac.uk

[d]Senior Lecturer. E-mail: J.Spreeuw@city.ac.uk

[e]Professor. E-mail: ctanggaard@creates.au.dk

[f]Ph.D. Student. E-mail: Andres.Villegas.1@cass.city.ac.uk

cualquier tipo de estimador núcleo de la función de riesgo. Se comprueba que APG posee buenas propiedades asintóticas y que consigue reducir el sesgo sin incrementar la varianza. Se realizan estudios de simulación para evaluar las propiedades del APG en muestras finitas. Dichos estudios muestran un buen comportamiento en la práctica del APG. Esto es especialmente alentador dado que para muestras finitas los métodos tradicionales de reducción del sesgo tienden a tener un comportamiento bastante pobre.

Palabras clave: estimación kernel, funciones de riesgo, estimación local lineal, kernels de frontera, corrección polinomial.

1. Introduction

In this paper we introduce a polynomial correction to a kernel hazard estimator based on standard survival data formulated via counting processes. The global correction attempts to obtain the best of both the two dominating non-parametric estimation worlds: kernel estimation and spline estimation. One clear advantage of kernel estimators is their analytical tractability whereas many practitioners seem to prefer polynomial based methods since they often give a good performance.

In this paper we start with an arbitrary kernel hazard pilot estimator, $\widetilde{\alpha}$, and correct it multiplicatively by a global polynomial. The parameters of the polynomial are chosen in such a way that the empirical moments of the counting process fit the resulting estimator up to some chosen order, r, of the adjusting polynomial. The resulting estimator is called global polynomial adjustment (GPA).

Such a polynomial adjustment is known from density estimation, see Efron & Tibshirani (1996). They do however use an exponential transformation of the data. This procedure requires more computations, but it does have the advantage of preserving positivity.

For the last decades there has been a lot of research aimed at finding new ways of reducing the bias of basic kernel smoothers. See Jones & Signorini (1999) for an overview and a comparative simulation study in the case of kernel density. Recently there is a huge development in the case of filtered data, see e.g. Nielsen, Tanggaard & Jones (2009), Gámiz Pérez, Martínez Miranda & Nielsen (2013) and Spreeuw, Nielsen & Jarner (2013). The case of kernel hazard estimation is studied in Nielsen & Tanggaard (2001). A multivariate study is given in Gámiz Pérez, Janys, Martínez Miranda & Nielsen (2013). However, none of these papers use the GPA technique introduced in this paper but rather stop with local adjustments. This is where our paper starts. We show that easy applicable global adjustments, here the GPA, can improve these estimates significantly.

In this paper, we consider as pilot estimators the estimators considered in Nielsen & Tanggaard (2001) who introduced the concepts of local constant and local linear kernel hazard estimators and rephrased the well known estimator of Ramlau-Hansen (1983) in this framework. Nielsen & Tanggaard (2001) also reformulated the traditional multiplicative bias correction method of Jones, Linton & Nielsen (1995) and Nielsen (1998) as the result of minimising a loss function

and introduced a local additive bias correction method based on the same minimisation.

The GPA method could be considered as a semiparametric approach to non-parametric estimation along the lines of Copas (1995), Eguchi & Copas (1998), Hjort & Glad (1995), Hjort & Jones (1996) and Loader (1996). However, as the asymptotic theory points out, then our type of semiparametric estimation is indeed quite different from the more common semiparametric estimation techniques. It is seen from the asymptotic theory that our estimator very often has a better asymptotic bias than if a GPA correction had not been applied. This holds even in situations where the true underlying hazard is far from the used parametric model: a polynomial in our case. This is different from these other semiparametric models, where the estimation method only works if the used parametric model is close to the true underlying hazard, see for example Hjort & Jones (1996).

The theory and the simulation experiment in Nielsen & Tanggaard (2001) demonstrated that local bias reduction methods are effective. Nevertheless, it turns out that the global polynomial correction of this paper is so good, that it is only of minor importance to use local bias adjustment in addition to the global adjustment. The step from local constant estimation to local linear estimation is, however, still important even when our global adjustment is used as a final step. In particular, we note that the global correction method gives a substantial improvement for small data sets where traditional bias reducing methods typically do not work. A feature of the traditional bias reducing methods is that they seem to work for data sets where there is no need for further precision, namely large samples, whereas they do not work well for small data sets. This is not the case for the global polynomial correction method as it works very well for both small and large data sets.

The simulation study shows that the usual conclusion holds for global polynomial adjustment, namely that it is better not to use too much bias correction when the data sets are small. It is preferable to use the more sophisticated bias correcting methods for large data sets. However, inspection of the simulation results suggests that applied researchers can do uniformly well by simply taking the local linear estimator as pilot estimator and make a global correction along the lines of this paper.

This paper proceeds as follows. In section 2 we introduce the general principle of globally correcting a pilot kernel hazard estimator. In section 3 we give the asymptotic properties of the GPA method and in section 4 we give an extensive simulation evaluating the performance of 5 different pilot estimators.

2. Global Hazard Estimators in a Counting Process Setting

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, Borgan, Gill & Keiding (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 ith individual is under risk. We assume that $(N_1, Y_1), \ldots, (N_n, Y_n)$ are *iid* for the n individuals.

We consider any preliminary estimator, $\widetilde{\alpha}$, of α, and we define the r'th order global polynomial correction based on adjusting the empirical counting process moments to this estimator.

Let r be a positive integer and let $\widehat{\mathbf{M}} = (\widehat{M}_0, \ldots, \widehat{M}_{r-1})'$, where

$$\widehat{M}_j = \sum_{i=1}^{n} \int_0^T s^j dN_i(s), j = 0, \ldots, r-1$$

is the j'th empirical counting process moment. We construct the global polynomial estimator, $\widehat{\alpha}^r$, by adjusting the original estimator by a multiplicative polynomial correction such that

$$\widehat{\alpha}^r(s) = \psi^r(s) \widetilde{\alpha}(s), \quad s \in [0, T],$$

where

$$\psi^r(s) = \sum_{k=0}^{r-1} \widehat{c}_k s^k.$$

The parameter vector $\widehat{c} = (\widehat{c}_0, \ldots, \widehat{c}_{r-1})'$ is chosen such that the first r counting process moments are correct, i.e.

$$\begin{aligned}
\widehat{M}_j &= \sum_{i=1}^{n} \int_0^T s^j \widehat{\alpha}^r(s) Y_i(s) ds \\
&= \sum_{k=0}^{r-1} \widehat{c}_k \sum_{i=1}^{n} \int_0^T s^{j+k} \widetilde{\alpha}(s) Y_i(s) ds \\
&= \sum_{k=0}^{r-1} \widehat{c}_k \widehat{H}_{k,j},
\end{aligned}$$

where

$$\widehat{\mathbf{H}} = (\widehat{H}_{k,j})$$

is the $r \times r$ matrix with

$$\widehat{H}_{k,j} = \sum_{i=1}^{n} \int_0^T s^{j+k} \widetilde{\alpha}(s) Y_i(s) ds$$

for $k, j \in \{0, \ldots, r-1\}$. The final global polynomial corrected estimator is therefore

$$\widehat{\alpha}^r(s) = \left\{ \sum_{k=0}^{r-1} \widehat{c}_k s^k \right\} \widetilde{\alpha}(s),$$

where the estimator $\widehat{\mathbf{c}} = (\widehat{c}_0, \ldots, \widehat{c}_{r-1})'$ is given via

$$\widehat{c} = \widehat{\mathbf{H}}^{-1}\widehat{\mathbf{M}}$$

This estimator clearly depends on an invertible $\widehat{\mathbf{H}}$.

Note. One could have chosen to multiply an exponential function of a polynomial instead of a polynomial itself. This would have the attraction of preserving positivity, but it would also have the disadvantage of requiring more complicated numerical computations.

3. Asymptotic Theory of the Global Polynomial Estimators

The general approach in non-parametric kernel estimation is to divide the estimation error into a stable part and a variable part. Under the standard conditions in that smoothing theory (cf. e.g. Nielsen & Tanggaard, 2001 for the local linear estimator) it is easy to verify that the asymptotic properties of the variable part, and thus the asymptotic variance, will remain unchanged from the global adjustment. The bias will, however, change. While the order of magnitude of the bias is unchanged, the leading term of the bias will be substantially reduced.

Let's assume that before the global polynomial adjustment, the bias of the estimator has the form

$$Bias(t) = g(t)b^x + o_P(b^x),$$

where b is a chosen bandwidth and $g(t)$ is a function of the hazard argument, t. For the estimators we will consider in the next chapters, we will have $x = 2$ for the local constant and the local linear estimators, $x = 4$ for the multiplicative bias corrected estimator, and $x = 6$ for the two times multiplicatively bias corrected estimator and the additively bias corrected estimator. After a global polynomial adjustment it is a straightforward exercise to show that under the conditions of the theorems in Nielsen & Tanggaard (2001) the bias will be adjusted to

$$Bias(t) = \overline{g}(t)b^x + o_P(b^x),$$

where x is unchanged and g is replaced by

$$\overline{g}(s) = g(s) \sum_{k=0}^{r-1} \overline{c}_k s^k.$$

The coefficients, $\overline{c}_0, \ldots, \overline{c}_{r-1}$, are chosen so that

$$\int_0^T s^k \overline{g}(s)\alpha(s)\gamma(s)ds = 0$$

for $k \in \{0, \ldots, r-1\}$ where $\gamma(s)$ is the deterministic limit of $n^{-1}Y^{(n)}(s)$. Hence $\overline{g}(s)$ is a polynomial down-weight of $g(s)$ towards 0. The larger r is, the heavier the

down-weight will typically be. Therefore, the global polynomial method reduces the leading term of the bias without changing the variance.

Accordingly, the global polynomial kernel hazard estimation approach has good theoretical properties and appears to be a promising method for applied research. In practice one should expect that r should be rather small when the observed data set is small, whereas it takes a relatively large data set to allow for the increased complexity in the estimation step introduced by large r's in the global polynomial regression. This conjecture is supported by the study of the small sample properties in the next section.

4. A Simulation Study

In this section we conduct a Monte Carlo simulation study for several hazard estimators and the global polynomial adjustment of these estimators. We follow Nielsen & Tanggaard (2001) and consider five different kernel hazard estimators: $\widehat{\alpha}_1$ (local constant), $\widehat{\alpha}_2$ (local linear), $\widehat{\alpha}_3$ (multiplicatively bias corrected local linear), $\widehat{\alpha}_4$ (double multiplicatively bias corrected local linear), $\widehat{\alpha}_5$ (additively bias corrected local linear). More precisely, these estimators are defined as follows.

For a kernel K and a bandwidth b let

$$a_l(t) = \sum_{i=1}^{n} \int_0^T K_b(t-s)(t-s)^j W(s) Y_i(s) ds, \quad l = 0, 1, 2,$$

where $K_b(u) = b^{-1} K(u/b)$. The local constant estimator with the weighting $W(s)$ is

$$\widehat{\alpha}_1(t) = \sum_{i=1}^{n} \int_0^T \widetilde{K}_{t,b}(t-s) dN_i(s),$$

where

$$\widetilde{K}_{t,b}(t-s) = \frac{K_b(t-s)}{a_0(t)} W(s).$$

The local linear estimator with weighting $W(s)$ equals

$$\widehat{\alpha}_2(t) = \sum_{i=1}^{n} \int_0^T \overline{K}_{t,b}(t-s) dN_i(s),$$

with

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

The multiplicatively bias corrected estimator of the local linear estimator is defined as

$$\widehat{\alpha}_3(t) = \widehat{\alpha}_2(t) \widehat{g}_M(t),$$

where the local linear estimator of the multiplicative error, $g_M(t) = \alpha(t)/\widehat{\alpha}_2(t)$ and given by

$$\widehat{g}_M(t) = \sum_{i=1}^{n} \int_0^T \overline{K}_{t,b}(t-s) \{\widehat{\alpha}_2(s)\}^{-1} dN_i(s),$$

is constructed with the weighting function

$$\widetilde{W}(s) = \{\widehat{\alpha}_2(s)\}^2 \, W(s).$$

We also consider the two times multiplicatively bias corrected estimator, $\widehat{\alpha}_4$, which is constructed by bias correcting $\widehat{\alpha}_3$.

The additively bias corrected estimator is based on a preliminary bias estimator, $\widehat{B}_t$, which is constructed by bootstrapping the bias (see Nielsen & Tanggaard 2001). The result is

$$\widehat{B}_t = \int_0^T \overline{K}_{t,b}(t-s) \left\{\widehat{\alpha}_2(t) - \widehat{\alpha}_2(s)\right\} ds.$$

Now

$$\widehat{g}_A(t) = \sum_{i=1}^n \int_0^T \overline{K}_{t,b}(t-s) d\widetilde{N}_i(s),$$

and

$$d\widetilde{N}_i(s) = \widehat{B}_s^{-1} dN_i(s) - \widehat{B}_s^{-1} \widehat{\alpha}_1(s) Y_i(s) ds,$$

where $\overline{K}_{t,b}$ is constructed with the weighting function

$$\overline{W}(s) = \widehat{B}_s^2.$$

We call

$$\widehat{\alpha}_5(t) = \widehat{\alpha}_2(t) + \widehat{g}_A(t) \widehat{B}_t$$

the local linear additively bias corrected estimator.

In the simulation studies we use the weighting $W(s) = 1$ and we consider global corrections of order, $r = 0, \ldots, 8$ (note that $r = 0$ amounts to no correction). All estimators are based on the kernel function

$$K(x) = (1 - x^2)^6,$$

where we have left out the normalization constant. As in Nielsen (1998) and Nielsen & Tanggaard (2001) we use as the true hazard one of the four functions:

$$\begin{aligned}
\gamma_1(t) &= B(t, 2, 2), \\
\gamma_2(t) &= B(t, 4, 4), \\
\gamma_3(t) &= 0.6 * \left[B(t, 0.5, 0.5) + B(t, 7, 7)\right], \\
\gamma_4(t) &= 0.6 * \left[B(t, 0.5, 0.5) + B(t, 4, 2) + B(t, 2, 4)\right],
\end{aligned}$$

where $B(t, \alpha, \beta)$ is for $t \in (0,1)$ the value of the Beta-density with parameters, α, β (see (Nielsen 1998) for a graph of the four functions).

Each simulation run is constructed as follows. First, we define a discrete grid on the interval $(0,1)$ with grid length, $\delta_M = 1/(M+1)$, as $\{t_m : t_m = m\delta_M, m = 1, \ldots, M\}$. Then, for a sample of n individuals, failures at time t_m are generated from the binomial distribution, Binomial $\left[Y^{(n)}(t_m), \gamma_j(t_m)\delta_M\right]$. Although we tried

several values of M, we report only the results for $M = 500$. Higher values of M do not seem to alter the conclusions.

To evaluate the simulations we use the following global measure of estimation error

$$err(\widehat{\alpha}_b) = n^{-1} \sum_{i=1}^{n} \int_0^1 \left[\widehat{\alpha}_b(s) - \gamma_k(s)\right]^2 Y_i(s)ds.$$

Our simulation study is based on the best possible bandwidth and, in the case of the global polynomial fit, also on the best possible r, with $1 \leq r \leq 9$. Comparisons are therefore made in a best-case scenario, which separates choice of estimator from the bandwidth and polynomial degree selection problems.

This means that for each simulation run where GPA is applied, we choose the best two dimensional parameters, (b, r), having the lowest possible value of $err(\widehat{\alpha}_b)$. The lower bound on the estimation error is unattainable in practice. This parallels the type of simulation study carried out in Jones et al. (1995), Jones & Signorini (1999), Nielsen (1998), Jones, Signorini & Hjort (1999) and Nielsen & Tanggaard (2001).

Tables 1-5 summarise the simulation results for the five different pilot estimators, receptively, with samples of size $n = 50$, $n = 150$, $n = 500$ and $n = 1000$.

TABLE 1: The local constant estimator $\widehat{\alpha}_1$.

Model	Without GPA		With GPA			Sample size
	$100 \times err$	b	$100 \times err$	b	r	
γ_1	6.22	0.46	4.19	0.89	2.71	50
γ_2	6.34	0.34	4.00	0.58	3.07	50
γ_3	14.12	0.36	12.12	0.60	3.31	50
γ_4	8.38	0.97	8.09	0.97	0.62	50
γ_1	2.76	0.32	1.59	0.91	3.18	150
γ_2	2.78	0.28	1.62	0.58	3.40	150
γ_3	8.65	0.23	6.47	0.64	4.66	150
γ_4	4.93	0.88	4.81	0.89	0.85	150
γ_1	1.11	0.24	0.55	0.91	3.72	500
γ_2	1.00	0.22	0.50	0.58	3.90	500
γ_3	5.19	0.12	3.68	0.61	5.22	500
γ_4	3.68	0.66	3.59	0.70	2.00	500
γ_1	0.66	0.20	0.30	0.92	4.34	1000
γ_2	0.57	0.19	0.26	0.63	4.48	1000
γ_3	3.82	0.07	2.99	0.50	5.31	1000
γ_4	3.27	0.44	3.13	0.44	3.50	1000

The estimation error is measured as the average over 250 simulation runs. The minimisation is done with respect to the bandwidth, b, in the left panel and with respect to the bandwidth and the polynomial degree, (b, r), in the right panel.

The left panel shows the minimum estimation error with no global adjustments. The numbers are averages of over 250 simulation runs of $err(\widehat{\alpha}_b)$. The right panel shows the values if the GPA is applied.

We can conclude that the GPA improves the goodness-of-fit for all estimators and all sample sizes. Furthermore, the GPA works relatively better on the local

TABLE 2: The local linear estimator $\widehat{\alpha}_2$.

Model	Without GPA		With GPA			Sample size
	$100 \times err$	b	$100 \times err$	b	r	
γ_1	4.64	0.59	3.60	0.86	1.79	50
γ_2	5.16	0.43	3.70	0.70	2.37	50
γ_3	15.05	0.40	12.59	0.53	3.86	50
γ_4	12.79	0.89	11.95	0.94	1.13	50
γ_1	1.75	0.49	1.20	0.78	1.87	150
γ_2	2.34	0.34	1.50	0.67	2.77	150
γ_3	7.97	0.24	6.12	0.45	4.54	150
γ_4	6.13	0.83	5.86	0.85	1.30	150
γ_1	0.64	0.38	0.39	0.65	1.97	500
γ_2	0.89	0.25	0.49	0.61	3.12	500
γ_3	4.00	0.15	3.18	0.36	4.26	500
γ_4	3.75	0.57	3.63	0.57	2.08	500
γ_1	0.36	0.33	0.21	0.56	2.08	1000
γ_2	0.52	0.22	0.25	0.60	3.32	1000
γ_3	2.78	0.10	2.41	0.24	3.89	1000
γ_4	2.93	0.28	2.81	0.26	2.60	1000

The estimation error is measured as the average over 250 simulation runs. The minimisation is done with respect to the bandwidth, b, in the left panel and with respect to the bandwidth and the polynomial degree, (b, r), in the right panel.

TABLE 3: The multiplicatively bias corrected estimator $\widehat{\alpha}_3$.

Model	Without GPA		With GPA			Sample size
	$100 \times err$	b	$100 \times err$	b	r	
γ_1	5.34	0.91	4.50	0.95	2.00	50
γ_2	5.66	0.74	4.20	0.89	2.22	50
γ_3	15.36	0.53	12.81	0.62	3.94	50
γ_4	14.30	0.93	13.56	0.95	1.38	50
γ_1	1.92	0.85	1.55	0.91	2.18	150
γ_2	2.28	0.61	1.61	0.87	2.48	150
γ_3	7.61	0.36	6.00	0.52	4.22	150
γ_4	6.67	0.89	6.47	0.89	1.61	150
γ_1	0.63	0.73	0.50	0.79	2.43	500
γ_2	0.72	0.50	0.49	0.83	2.54	500
γ_3	3.63	0.24	3.01	0.42	3.84	500
γ_4	3.85	0.63	3.74	0.58	2.52	500
γ_1	0.33	0.66	0.26	0.71	2.67	1000
γ_2	0.38	0.45	0.24	0.81	2.77	1000
γ_3	2.50	0.17	2.24	0.28	3.54	1000
γ_4	2.91	0.32	2.80	0.30	2.94	1000

The estimation error is measured as the average over 250 simulation runs. The minimisation is done with respect to the bandwidth, b, in the left panel and with respect to the bandwidth and the polynomial degree, (b, r), in the right panel.

constant and the local linear estimators, that is when no local bias correction is applied. It also seems to work better for small data sets. This is different to classical local bias corrections (See $\widehat{\alpha}_3, \widehat{\alpha}_4, \widehat{\alpha}_5$ with no GPA) which seem not to be advisable for small data sets.

TABLE 4: The two times multiplicatively bias corrected estimator $\widehat{\alpha}_4$.

Model	Without GPA		With GPA			Sample size
	$100 \times err$	b	$100 \times err$	b	r	
γ_1	6.02	0.95	4.90	0.96	2.25	50
γ_2	5.92	0.88	4.32	0.94	2.08	50
γ_3	15.65	0.59	12.79	0.66	3.95	50
γ_4	15.46	0.94	14.46	0.95	1.64	50
γ_1	2.14	0.91	1.72	0.93	2.31	150
γ_2	2.33	0.78	1.63	0.91	2.31	150
γ_3	7.44	0.43	5.91	0.57	4.16	150
γ_4	7.05	0.90	6.79	0.88	2.04	150
γ_1	0.69	0.81	0.54	0.82	2.61	500
γ_2	0.71	0.67	0.49	0.88	2.44	500
γ_3	3.48	0.30	2.91	0.46	3.82	500
γ_4	3.91	0.62	3.78	0.58	2.98	500
γ_1	0.36	0.74	0.28	0.75	2.94	1000
γ_2	0.37	0.62	0.25	0.85	2.65	1000
γ_3	2.37	0.21	2.14	0.31	3.43	1000
γ_4	2.89	0.32	2.78	0.30	3.21	1000

The estimation error is measured as the average over 250 simulation runs. The minimisation is done with respect to the bandwidth, b, in the left panel and with respect to the bandwidth and the polynomial degree, (b, r), in the right panel.

TABLE 5: The additively bias corrected estimator $\widehat{\alpha}_5$.

Model	Without GPA		With GPA			Sample size
	$100 \times err$	b	$100 \times err$	b	r	
γ_1	6.11	0.87	5.20	0.88	1.36	50
γ_2	6.77	0.73	4.46	0.82	2.03	50
γ_3	17.50	0.73	13.30	0.84	3.36	50
γ_4	19.48	0.90	16.96	0.90	2.03	50
γ_1	2.18	0.88	1.85	0.92	1.57	150
γ_2	2.71	0.61	1.71	0.79	2.22	150
γ_3	7.91	0.57	6.02	0.80	4.15	150
γ_4	8.06	0.85	7.24	0.84	2.16	150
γ_1	0.67	0.83	0.55	0.90	1.98	500
γ_2	0.85	0.54	0.49	0.76	2.70	500
γ_3	3.72	0.34	2.96	0.64	4.82	500
γ_4	3.91	0.68	3.50	0.63	3.57	500
γ_1	0.34	0.78	0.27	0.87	2.38	1000
γ_2	0.46	0.51	0.24	0.76	2.90	1000
γ_3	2.47	0.23	2.10	0.39	4.64	1000
γ_4	2.71	0.48	2.41	0.45	4.24	1000

The estimation error is measured as the average over 250 simulation runs. The minimisation is done with respect to the bandwidth, b, in the left panel and with respect to the bandwidth and the polynomial degree, (b, r), in the right panel.

Furthermore, we can see that the optimal r increases with the size of the data set which indicates that large data sets allow for more complexity in the estimation step.

In general, the simulation results support the asymptotic theory. We observe that when GPA is applied, a greater bandwidth is chosen than without GPA.

The asymptotic theory shows that for a fixed bandwidth, GPA reduces the bias. Therefore, a greater bandwidth (which adds bias and reduces variance) re-balances bias and variance so that the global estimation error is minimised.

5. Conclusion

In this paper, we have introduced the so-called global polynomial adjustment (GPA) for the bias reduction of kernel hazard estimators. The theoretical properties of the GPA as well as a simulation study, using several different pilot kernel hazard estimators, suggest that the GPA is very promising. Therefore, it is worth exploring the extension of the global estimation principle of this paper to the local likelihood estimation principle (see, e.g. Otneim, Karlsen & Tjøstheim 2013) and to the cases where variable bandwidth and variable kernel are considered (see Nielsen 2003, Koul & Song 2013). Also, other parametric shapes other than global polynomials could have been used, for example the Gumbel distribution of (Salinas, Pérez, González & Vaquera 2012) or multivariate structures as in Martínez-Flórez, Moreno-Arenas & Vergara-Cardozo (2013) or Lemonte, Martínez-Florez & Moreno-Arenas (2014). In the latter case, one could for example adjust the marker dependent hazard estimators of Nielsen & Linton (1995) or Nielsen (1998) according to relevant global moments. The practical choice of polynomial correction r is of course important. We are currently investigating how to adapt cross-validation and Do-validation to global polynomial kernel hazard estimation. This would involve generalising the cross-validation and do-validation procedures given in Gámiz Pérez, Martínez Miranda & Nielsen (2013) to also include picking the order of the adjusting polynomial. This again would secure that a classical variance/bias trade-off ensures that r is not being picked too big (overfitting) or too small (not taking advantage of the global correction trick).

Acknowledgements

References

Andersen, P., Borgan, O., Gill, R. & Keiding, N. (1993), *Statistical Models Based on Counting Processes*, Springer, New York.

Copas, J. B. (1995), 'Local likelihood based on kernel censoring', *Journal of the Royal Statistical Society, Series B* **57**(1), 221–235.

Efron, B. & Tibshirani, R. (1996), 'Using specially designed exponential families for density estimation', *Annals of Statistics* **24**(6), 2431–2461.

Eguchi, S. & Copas, J. (1998), 'A class of local likelihood methods and near-parametric asymptotics', *Journal of the Royal Statistical Society, Series B* **60**(4), 709–724.

Gámiz Pérez, M. L., Janys, L., Martínez Miranda, M. D. & Nielsen, J. P. (2013), 'Bandwidth selection in marker dependent kernel hazard estimation', *Computational Statistics & Data Analysis* **68**, 155–169.

Gámiz Pérez, M. L., Martínez Miranda, M. D. & Nielsen, J. P. (2013), 'Smoothing survival densities in practice', *Computational Statistics & Data Analysis* **58**, 368–382.

Hjort, N. L. & Glad, I. K. (1995), 'Nonparametric density estimation with a parametric start', *Annals of Statistics* **23**(3), 882–904.

Hjort, N. L. & Jones, M. C. (1996), 'Locally parametric nonparametric density estimation', *Annals of Statistics* **24**(4), 1619–1647.

Jones, M. C., Linton, O. B. & Nielsen, J. P. (1995), 'A simple bias reduction method for density estimation', *Biometrika* **82**(2), 327–338.

Jones, M. C. & Signorini, D. F. (1999), 'A comparison of higher-order bias kernel density estimators', *Journal of the American Statistical Association* **439**, 1063–1073.

Jones, M. C., Signorini, D. F. & Hjort, N. L. (1999), 'On multiplicative bias correction in kernel density estimation', *Sankhya* **61**(1), 422–430.

Koul, H. L. & Song, W. (2013), 'Large sample results for varying kernel regression estimates', *Journal of Nonparametric Statistics* **25**(4), 829–853.

Lemonte, A., Martínez-Florez, G. & Moreno-Arenas, G. (2014), 'Multivariate Birnbaum-Saunders distribution: Properties and associated inference', *Journal of Statistical Computation and Simulation* **85**(2), 374–392.

Loader, C. R. (1996), 'Local likelihood density estimation', *Annals of Statistics* **24**(4), 1602–1618.

Martínez-Flórez, G., Moreno-Arenas, G. & Vergara-Cardozo, S. (2013), 'Properties and inference for proportional hazard models', *Revista Colombiana de Estadística* **36**(1), 95–114.

Nielsen, J. P. (1998), 'Multiplicative bias correction in kernel hazard estimation', *Scandinavian Journal of Statistics* **25**(3), 541–553.

Nielsen, J. P. (2003), 'Variable bandwidth kernel hazard estimators', *Journal of Nonparametric Statistics* **15**(3), 355–376.

Nielsen, J. P. & Linton, O. (1995), 'Kernel estimation in a nonparametric marker dependent hazard model', *Annals of Statistics* **23**(5), 1735–1748.

Nielsen, J. P. & Tanggaard, C. (2001), 'Boundary and bias correction in kernel hazard estimation', *Scandinavian Journal of Statistics* **28**(4), 675–698.

Nielsen, J. P., Tanggaard, C. & Jones, M. C. (2009), 'Local linear density estimation for filtered survival data', *Statistics* **43**(2), 167–186.

Otneim, H., Karlsen, H. A. & Tjøstheim, D. (2013), 'Bias and bandwidth for local likelihood density estimation', *Statistics and Probability Letters* **83**(4), 1382–1387.

Ramlau-Hansen, H. (1983), 'Smoothing counting process intensities by means of kernel functions', *Annals of Statistics* **11**(2), 453–466.

Salinas, V., Pérez, P., González, E. & Vaquera, H. (2012), 'Goodness of fit tests for the gumbel distribution with type II right censored data', *Revista Colombiana de Estadística* **35**(3), 407–424.

Spreeuw, J., Nielsen, J. P. & Jarner, S. F. (2013), 'A nonparametric visual test of mixed hazard models', *SORT* **37**(2), 153–174.