Bias in nearest-neighbor hazard estimation

Rafael Weißbach[†*] & Holger Dette[‡]

[†]Institut für Wirtschafts- und Sozialstatistik, Universität Dortmund, Dortmund, Germany

[‡]Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany

June 23, 2008

*address for correspondence: Rafael Weißbach, Institut für Wirtschafts- und Sozialstatistik, Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany, email: Rafael.Weissbach@tu-dortmund.de, Fon: +49/231/7555419, Fax: +49/231/7555284.

JEL classifications. C13, C14, C41, G32, AMS subject classification 62M02

Keywords. hazard rate, kernel estimation, bandwidth selection, variable bandwidth, nearest neighbor bandwidth, rule of thumb.

Abstract

In nonparametric curve estimation, the smoothing parameter is critical for performance. In order to estimate the hazard rate, we compare nearest neighbor selectors that minimize the quadratic, the Kullback-Leibler, and the uniform loss. These measures result in a rule of thumb, a cross-validation, and a plug-in selector. A Monte Carlo simulation within the three-parameter exponentiated Weibull distribution indicates that a counter-factual normal distribution, as an input to the selector, does provide a good rule of thumb. If bias is the main concern, minimizing the uniform loss yields the best results, but at the cost of very high variability. Cross-validation has a similar bias to the rule of thumb, but also with high variability.

Keywords: hazard rate, kernel smoothing, bandwidth selection, nearest neighbor bandwidth, rule of thumb, plug-in, cross-validation, credit risk.

1 Introduction

In the finance literature, the hazard rate or intensity of default is part of fundamental pricing formulae [Bielecki and Rutkowski (2002)]. Furthermore, it is an important parameter for rating matrices. The latter play a crucial role for regulatory capital [Basel Committee on Banking Supervision (2004)] and economic capital [Gupton et al. (1997)]. In medicine, the hazard rate is used in cancer research with death or tumor relapse as endpoints [Günther et al. (2005); Siu et al. (1998); Weißbach et al. (2008a)].

Although the constant hazard rate has its merits, a time-homogeneous failure process is occasionally rejected by goodness-of-fit tests [see Weißbach and Dette (2007); Weißbach et al. (2008b); Kiefer and Larson (2007)]. Kernel hazard rate estimation has attracted some attention over the last few decades and is usually based on kernel smoothing [Marron and de Uña Álvarez (2004); Marron (1996); Müller and Wang (1994); Patil (1993); Schäfer (1985)]. Estimation for an entire rating transition matrix is very similar [Andersen et al. (1993); Aalen and Johansen (1978)].

As in density estimation, the bandwidth selection is crucial for the performance of an estimate. The bandwidth must not be too large, so as to avoid over-smoothing, i.e. substantial bias, and must not be too small either, so as to avoid detecting the underlying structure. The concept of nearest neighbors is currently attracting interdisciplinary interest [(see Abarbanel, 1996; Sugihara and May, 1990; Ralescu, 1995; Wagner, 1975, e.g.)]. This concept helps balance the problem of fixed bandwidth along the time axis, by using more observations where density is high and fewer

where it is low. The idea is to widen the window that is used for estimation if the density is low, and to narrow it if the density is high. The advantage, for failure time analysis is, that it can also be adapted to censored data [Dette and Gefeller (1995)].

Comparisons of bandwidth selectors are often asymptotical and in the context of density estimation [Park and Marron (1990); Jones et al. (1996)]. The purpose of the present paper is to study the finite-sample bias of kernel hazard rate estimates under automated bandwidth selection, by means of a simulation study. We concentrate our investigations on bias, because this is of primary concern in banking. The Banking Committee on Banking Supervision (2004, p. 86) , for example, requests almost totally unbiased estimation for credit risk parameters.

The first specific goal for bandwidth selection is to minimize the quadratic loss. To this end, Weißbach et al. (2008a) translate Silverman's rule of thumb for a fixed bandwidth to the nearest neighbor bandwidth. Hall (1978) attempts to minimize the Kullback-Leibler loss for a fixed bandwidth by means of a leave-one-out cross-validation. We use an analogous implementation for the nearest neighbor bandwidth [Gefeller et al. (1996)]. Weißbach (2006) suggests minimizing the uniform absolute loss, and we use plug-in to implement an optimal number of nearest neighbors.

Motivated by empirical results about intensity shapes in credit risk modeling [Lando and Skødeberg (2002)], we base our Monte Carlo simulation on a three-parameter extension of the Weibull distribution. The main finding confirms what is already known on fixed bandwidth selection in kernel density estimation. Assuming a normal distribution in order to define an optimal bandwidth ideal from a practical point of view. Cross-validation has a similar bias, but higher variability and required a greater computational effort. Plug-in, however, results in a small maximal bias, but with very substantial variability.

2 Model

The typical examples of (right-)censoring in applications are studies with "time till event" as the variable in question T and a censoring mechanism preventing the observation of the "event". For such studies, the hazard rate $\alpha(\cdot)$ has proven easier for the purpose of interpretation than the density, because of its notion as instantaneous failure rate

$$\begin{aligned}\alpha(t) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\left(T \in [t, t+\Delta t] \mid T \geq t\right) \\ &= \frac{f(t)}{S(t)} = \frac{f(t)}{1-F(t)},\end{aligned}$$

where $f(\cdot)$ denotes the pertinent density and $F(\cdot)$ and $S(\cdot)$ the cumulative distribution function (CDF) and survival function. One can smooth the empirical process so as to obtain an estimate of the density, or, similarly, it is possible to smooth the Nelson-Aalen estimate of the cumulative hazard rate $A(t) = \int_0^t \alpha(s)ds$

$$A_n(t) = \sum_{i:X_{(i)} \leq t} \frac{\delta_{(i)}}{n - i + 1} \tag{1}$$

to obtain an estimate of the hazard rate. The observations $X_i = \max\{T_i, C_i\}$ are either failure times (T_i) or times of censoring (C_i). The $\delta_i = I_{\{X_i = T_i\}}$ indicate the censoring for $i = 1, \ldots, n$ independent observations. The order of the $\delta_{(i)}$'s refers to the pertinent ordered $X_{(i)}$'s.

The nearest neighbor bandwidth definition in the presents of censored data is [Dette and Gefeller (1995)]

$$R_n^{NN}(t) := \inf \left\{ r > 0 \mid |S_n\left(t - \frac{r}{2}\right) - S_n\left(t + \frac{r}{2}\right)| \geq \frac{k}{n} \right\}.$$

Here

$$S_n(t) = \prod_{\{i:X_{(i)} \leq t\}} \left(\frac{n - i}{n - i + 1} \right)^{\delta_{(i)}}$$

denotes the Kaplan-Meier product limit estimate of the survival function [Kaplan and Meier (1958)].

Combining the estimate (1) with the nearest-neighbor bandwidth, we propose the following estimate for the hazard rate

$$\begin{aligned} \alpha_n(t) &= \int_{\mathbb{R}_0^+} \frac{1}{R_n(s)} K\left(\frac{t - s}{R_n(s)} \right) dA_n(s) \\ &= \sum_{i=1}^{n} \frac{\delta_{(i)}}{(n - i + 1) R_n^{NN}(X_{(i)})} K\left(\frac{X_{(i)} - t}{R_n^{NN}(X_{(i)})} \right). \end{aligned} \tag{2}$$

We use the quadratic kernel with support $z \in [-1/2, 1/2]$

$$K(z) = I_{[-\frac{1}{2}, \frac{1}{2}]} \frac{240}{23} \left(\frac{1}{4} - z^2 \right)^2.$$

Fortunately, the impact of the kernel on the performance of the estimation is known to be small [(see Wand and Jones, 1995, p. 31)].

3 Selecting the number of nearest neighbors

Three types of bandwidth selectors are referred to frequently in the literature. First of all, one may assume a (counterfactual) distribution. Minimizing, say, the mean integrated squared error with respect to the bandwidth results in an analytical expression for the optimal bandwidth [Parzen (1962); Silverman (1986)]. Without a specific distributional assumption, the second alternative is to numerically minimize an error by cross-validation, usually, leave-one-out estimation is used in this context [Scott and Terrell (1987); Chow et al. (1983); Hall (1978); Marron (1987); Patil (1993)]. Another idea without distributional assumptions is plug-in. The optimal bandwidth depends on the distribution to be estimated, therefore one plugs an arbitrarily estimate into the analytic expression of the optimal bandwidth [Hall et al. (1991)]. One of each type will be used in our simulation study.

For the fixed bandwidth b in kernel density estimation, Silverman's rule of thumb assumes normality of the data and minimizes asymptotically the mean integrated squared error [Silverman (1986)]. The solution is explicitly given by

$$h^{RoT} = \left(\frac{8\pi^{\frac{1}{2}} \int K^2(z)dz}{3(\int zK^2(z)dz)^2 n} \right)^{1/5} \hat{\sigma}, \tag{3}$$

where σ is the standard deviation of the observations and to be estimated. Weißbach et al. (2008a) give a modification of that rule of thumb for the hazard rate under random censoring and for the nearest neighbor bandwidth. The adoption to the nearest neighbor bandwidth is achieved by identifying the fixed bandwidth to imply a linear approximation of the CDF, upon which can be improved by stochastic approximation, with the empirical process. The number of nearest neighbors is given by

$$k^{RoT} = \left[n \cdot |\hat{\beta}| \cdot h^{RoT} \right],$$

with $\hat{\beta}$ as the regression slope through the points $(X_i, S_n(X_i))$ of the Kaplan-Meier survival estimate. For the characteristics $\int K^2$ and the second moment of the bi-square kernel in (3), see Wand and Jones (1995, p. 176) . In order to estimate the standard deviation, we used the unbiased variance estimate restricted to the uncensored observations, knowing that this is not statistically, but rather computationally efficient. The Gaussian brackets $[\cdot]$ ensure k to be an integer.

Bandwidth selection for the nearest neighbor bandwidth in hazard rate estimation can be implemented by cross-validation, as in Gefeller et al. (1996). Maximizing a leave-one-out likelihood results in an optimal bandwidth. Hall (1978) shows that this

asymptotically minimizes the expected Kullback-Leibler loss for the corresponding density. Looking at this in more detail, the likelihood is decomposed into the hazard rate and the survival function, both estimated by cross-validation.

$$k^{CV} = argmax_{k\in\{1,\ldots,n\}} \prod_{i=1}^{n} \alpha_n^{-i}(X_i)^{\delta_i} S_n^{-i}(X_i).$$

Here, $h_n^{-i}(X_i)$ estimates the hazard rate at time X_i by (2) based on the entire sample except for X_i. The same applies to the Kaplan-Meier estimate $S_n^{-i}(\cdot)$ of the survival function. Enumeration over the possible numbers of nearest neighbors yields the optimum.

The uniform error for data-driven bandwidth functional estimators is analyzed by decomposing into two additive components. The first is the stochastic error, the second, the deterministic bias [Einmahl and Mason (2005)]. More specifically, one can use an asymptotic bound on the uniform error to establish the consistency of the nearest neighbor hazard rate estimate [Weißbach (2006)]. In order to balance deterministic and stochastic error, we minimize the bound with respect to the number of nearest neighbors and yield an optimal bandwidth selector with respect to the uniform loss:

$$k^{PI} = \left(\frac{D_1 + D_2}{2D_3}\right)^{\frac{2}{3}} (\log n)^{\frac{1}{3}} n^{\frac{2}{3}}$$

The constants are given as $D_1 = 18(1-G(B))^{-\frac{1}{2}} M\tilde{M}^2\tilde{m}^{-2}(\sup(K)\tilde{m}^{-1}+L_K\tilde{M}\tilde{m}^{-1})$, $D_2 = 9(1 - F^{obs}(B)))^{-\frac{1}{2}}\tilde{M}M^{\frac{1}{2}}V(K)\tilde{m}^{-\frac{1}{2}}$, and $D_3 = 2\tilde{M}^3 M L_K L_{\tilde{\psi}}\tilde{m}^{-5} + 2\sup(K)L_{\tilde{\psi}}\tilde{M}^2 M\tilde{m}^{-4} + L_\psi\tilde{m}^{-1}$. They mainly depend on the distribution of the failure time T_i. In detail, $\tilde{M}$ and $\tilde{m}$ denote the maximum and the minimum of the density. M is the maximum of the hazard rate. The Lipschitz constants of the density and the hazard rate are $L_{\tilde{\psi}}$ and L_ψ. All quantities are to be seen restricted to the support $[A, B]$ to which the uniform consistency must be restricted. Clearly, the censoring leaves its traces, so that the CDF for the censoring C_i, $G(\cdot)$, and of the actual observations X_i, $F^{obs}(\cdot)$ enter the constants. All those quantities are estimated by plug-in here. As plug-in estimates, we use the selectors k^{RoT} and k^{CV} from the previous paragraphs.

For the kernel, the supremum, $\sup(K)$, and the Lipschitz constant, L_K, and the total variation, $V(K)$, gear the constants. The biquadratic kernel, we use here, implies $\sup(K)$ to be 0.652173913 and L_K to be 2.008174849 and $V(K)$ to be 0.326086957.

4 Simulation design

The preceding sections proposes three numbers for the nearest neighbors k^{RoT}, k^{CV}, and k^{PI} for the hazard rate estimate (2). Plugging the modified rule of thumb k^{RoT} into k^{PI}, denoted by $k^{PI}(\leftarrow RoT)$, and the cross-validation k^{CV}, denoted by $k^{PI}(\leftarrow CV)$, yields four bandwidth selection methods for the purpose of comparison.

For a simulation, it is necessary to select a parametric family for the comparison of true and estimated hazard rates. The benchmark is a constant hazard rate. For instance, In finance, the homogeneous Markov process is popular for modeling rating migrations. In particular, its generalization for several rating states enables convenient calculation of rating migration matrices containing the one-, two-, and three-year transition probabilities by means of the generator method [see Bluhm et al. (2002)]. The constant hazard rate, that is, the exponential distribution, is a special case of the Weibull distribution. It is advisable to have more parameters, in order to avoid reducing the smoothness of the family. Few parameters will automatically lead to a dominance of methods that originate from the case of bandwidth selection for a gaussian density, as is the case for bandwidth h^{RoT} (3). Lando and Skødeberg (2002) present empirical evidence that external rating intensities tend to decrease. In order to model this effect, we use the three-parameter (exponentiated) Weibull family [Mudholkar et al. (1995)]. Decreasing hazards are incorporated, and additionally, bath-tub shapes. The latter is useful, because decreasing hazards are not plausible as a long-term effect in credit risk. Migration rates would be decreasing, an unexperienced effect. In medicine, the bath-tub shape is the general death hazard after birth.

The exponentiated Weibull distribution is defined in terms of the survival function

$$S(t) = 1 - \left(1 - \exp\left(-\left(\frac{t}{\gamma}\right)^{\kappa}\right)\right)^{\theta},$$

with $0 < x < \infty$, $\kappa > 0$, $\theta > 0$ and $\gamma > 0$, so that the family contains the original Weibull distribution for $\theta = 1$. In terms of the hazard rate, the three-parameter extension becomes

$$\alpha(t) = \frac{\kappa\theta\left(1 - \exp\left(-\left(\frac{t}{\gamma}\right)^{\kappa}\right)\right)^{\theta-1} \exp\left(-\left(\frac{t}{\gamma}\right)^{\kappa}\right)\left(\frac{t}{\gamma}\right)^{\kappa-1}}{\gamma\left(1 - \left(1 - \exp\left(-\left(\frac{t}{\gamma}\right)^{\kappa}\right)\right)^{\theta}\right)}.$$

Four shapes of the hazard rate modeled in this family are identifiable by parameter space segments, increasing, decreasing, unimodal and convex (bath-tub) shaped. The limiting lines are $\kappa = 1$ and $\kappa\theta = 1$. The simulation was conducted for representatives of the four shapes, but the convex hazard rate is the most challenging of the

four (see Figure 1). It increases steeply towards the left boundary, the time-origin. Additionally, it increases steeply towards the left, where fading data constitutes the main problem. For the sake of brevity, we restrict the discussion mainly to the estimation of that type, with the parameters $(5, 0.1, 100)$. The findings are similar for the three other (increasing, unimodal and decreasing) shapes. There is one aspect in which the similarity is not true for the decreasing shape, this will be analyzed.

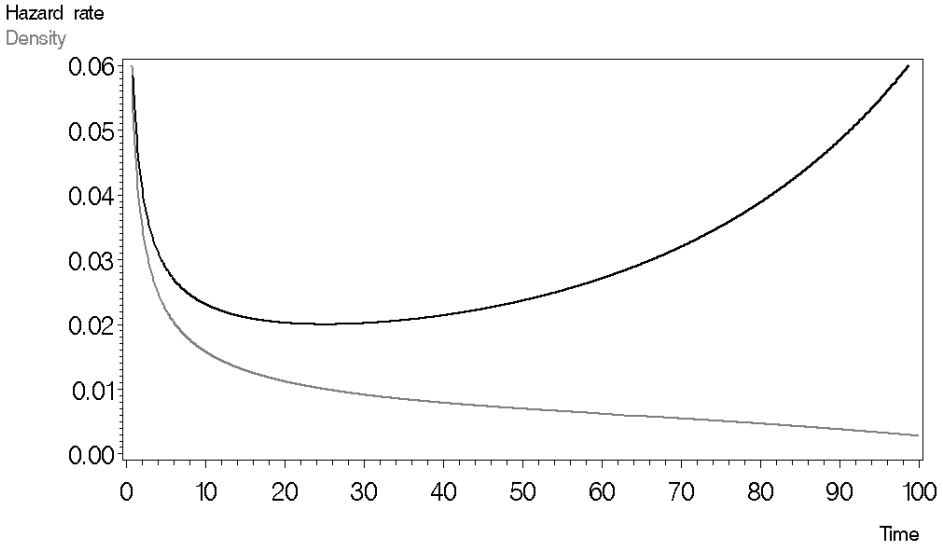

Figure 1: Convex hazard rate (black) and its density (gray) of an exponentiated Weibull distribution with parameters $\kappa = 5$, $\theta = 0.1$, $\gamma = 100$

Preliminary simulations have shown that it is advisable to restrict estimation to the inner 80% area, $[F^{-1}(0.1), F^{-1}(0.9)]$, for our bath-tub hazard it becomes $[1; 84.5]$. We consider only one degree of censoring, namely 40%. To our experience there is no typical value, but this is an average value for data sets we have analyzed. In fact lays in between the degrees of the two sets we analyze later on. The degree of censoring is achieved by choosing survival times T_i and censoring times C_i, both due to the exponentiated Weibull family with similar parameters. Starting from an expected 50% censoring for similar distributions, we use the monotony of the expected degree of censoring with respect to the parameter θ, in order to simulate the desired 40%. The resulting parameter set is $(5, 0.15, 100)$. (The parameter sets for the decreasing hazard rates are $(0.5, 9.5, 5)$ for the failure time and $(0.5, 14.3, 5)$ for the censoring time. The 80% support is $[11.8; 101.6]$.) Note that the censoring was simulated with respect to the entire support, but was found to hold for the 80% support with high accuracy in the study.

Sample sizes n of 50 and 100 observations mimic the situation depicted in Weißbach

and Dette (2007). A moderate number of 300 observations resembles the situation in Weißbach et al. (2008b). In our experience, 500 simulation runs are sufficient for point estimation purposes. For the 300-observation situation, 250 replications reduce the computation time. The remaining variability inherent in the mean of the point estimates enabled us to detect high variability in an investigated method.

As random number generator, we use the generator of uniformly distributed random numbers on $[0, 1]$ as implemented in SAS/IML and map to our distribution family via the inverse CDF

$$F^{-1}(u) = \gamma[-\log(1 - u^{\frac{1}{\theta}})]^{\frac{1}{\kappa}}, \quad 0 < u < 1.$$

All samples are generated once for all methods, in order to avoid a bias from multiple random number generation, between the different smoothing methods.

5 Results

We now analyze the bias obtained by the different methods for the selection of the smoothing parameter. To that end, we plot the average over all 500 (or 250) simulation loops. We see that, along with the bias assessment, some aspects of estimation variability can be deduced from those graphs. The graphical analysis is amended by a statistical analysis. The averages and standard deviations of the losses, the selected numbers of nearest neighbors, the integrated squared biases and variances are discussed.

Before comparing the nearest-neighbor selectors, we study (i) the effect of censoring and (ii) the effect of the sample size on the bias.

The effect of censoring can be seen, even for the decreasing hazard in Figure 2, for k^{RoT} and the cross-validatory selector k^{CV}. For 40% censoring, large observations are more likely to be censored than small ones. It is easier to estimate the left boundary, as compared to the right tail, especially for the small sample size of 50 observations.

The effect of sample size is depicted for the convex hazard in Figure 4 for the number of nearest neighbors selected according to the four selectors (each in a row). With increasing sample size (left to right) the bias decreases. However, the benefit differs across the support interval. On the left boundary, the interval's part that benefits is much smaller than on the right tail. The reason is mainly the larger density near the origin. As deduced from Figure 2, censoring compounds this effect by further reducing data towards the right tail. The fit between the quartiles of the

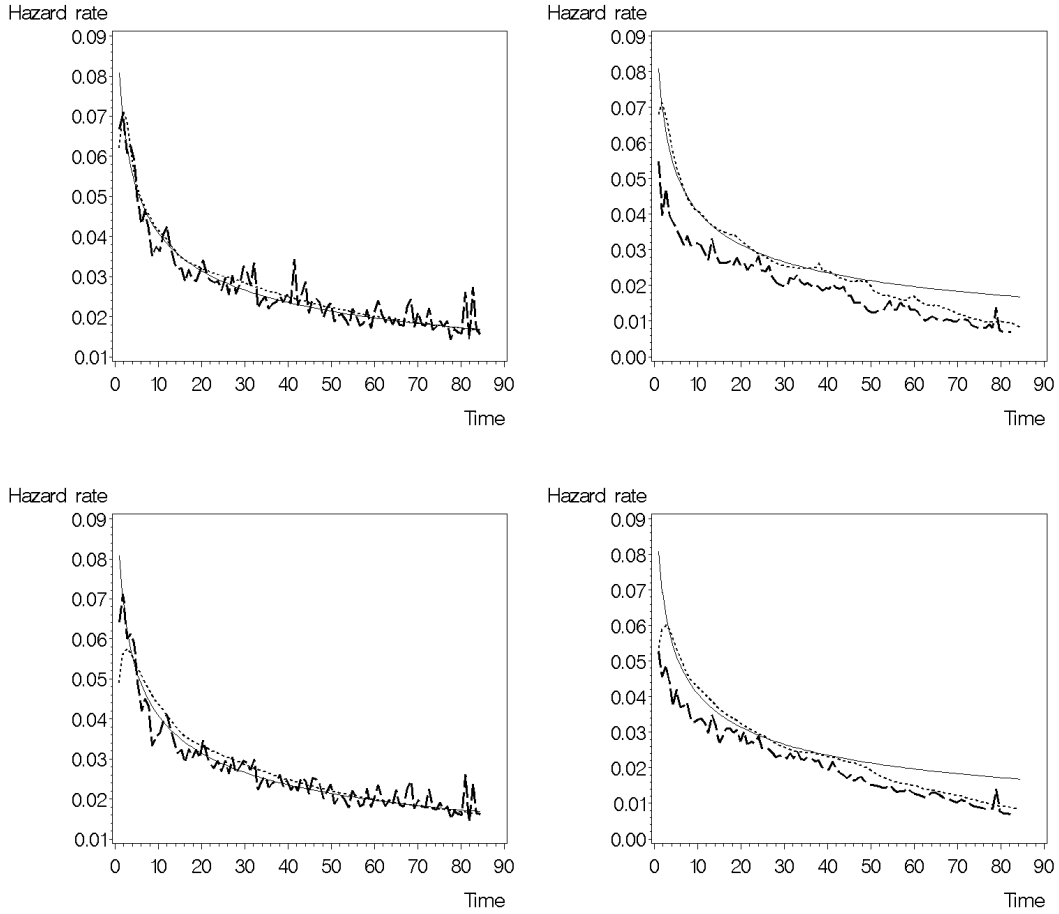

Figure 2: Decreasing hazard rate (solid line) and average of 500 estimations with k^{RoT} (top/dotted) and $k^{PI}(\leftarrow RoT)$ (top/dashed), and k^{CV} (bottom/dotted) and $k^{PI}(\leftarrow CV)$ (bottom/dashed) nearest neighbors from 50 uncensored observations (left) and 50 observations with 40% censoring (right)

distribution, however, is only slightly improved. The reason is that the hazard rate is near to constant here.

In fact, the rule of thumb and cross-validation behave very similarly with respect to the bias. The plug-in estimate behaves differently, but before assessing the effect of uniform absolute error minimization k^{PI}, it is useful to ensure that the plug-in choice does not exert a dominant influence. As plug-in for estimation the convex hazard, Figure 4 shows the rule of thumb k^{RoT} (second row) and the modified likelihood maximization k^{CV} (forth row). The bias is similar, that is not influenced by the choice of plug-in.

Both graphs series - compared to the direct use of k^{RoT} (first row) and k^{CV} (third row) nearest neighbors - visualize the behavior of the asymptotic minimization of uniform loss. The maximal bias is smaller than for k^{RoT} and k^{CV}, and to the left

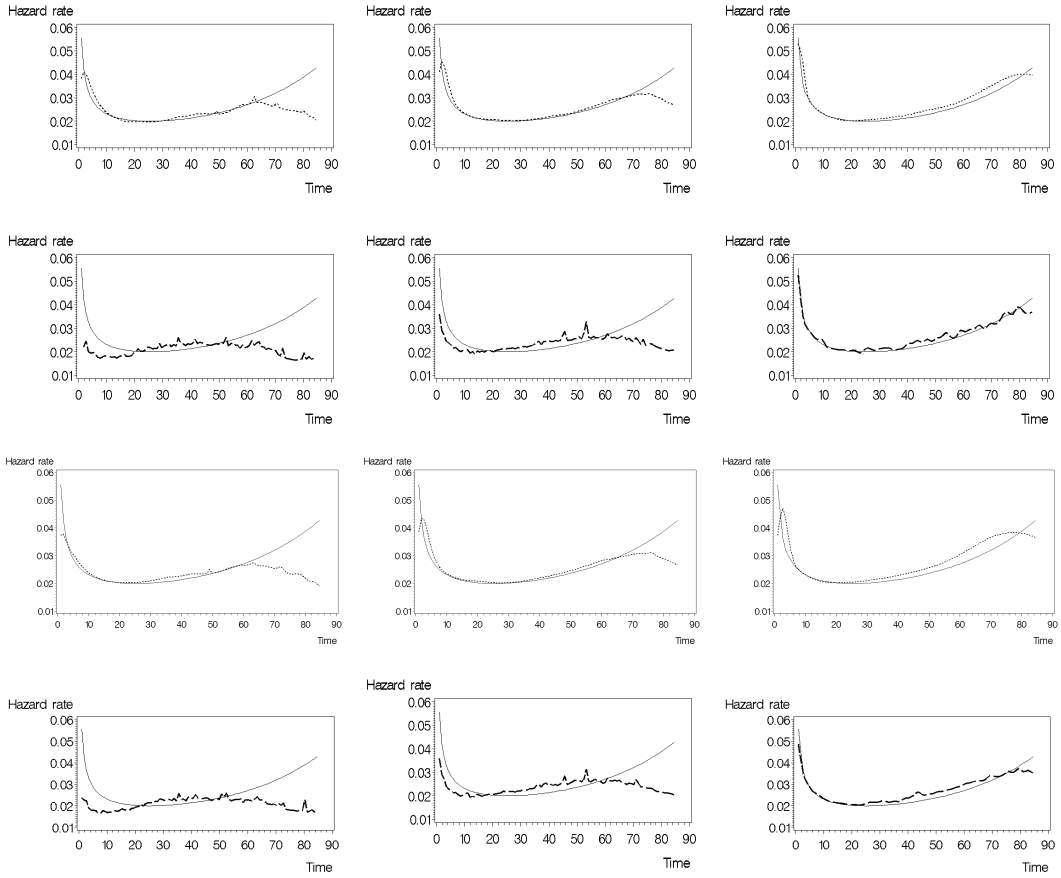

Figure 3: Convex hazard rate (solid line) and average of estimations with k^{RoT} (dotted/first row), $k^{PI}(\leftarrow RoT)$ (dashed/second row), k^{CV} (dotted/third row), and $k^{PI}(\leftarrow CV)$ (dashed/forth row) nearest neighbors and 40% censoring (dashed) for 500 estimations from 50 observations (left), 500 estimations from 100 observations (middle) and 250 estimations from 300 observations (right)

boundary for $n = 300$ even non-existent. As a consequence, the small bias leads to large variance. The wriggly appearance of the hazard rate already indicates a very substantial estimation variability, including very small numbers of nearest neighbors. This can also be seen in the numeric results. The exception, the smoothness on the left from 0 to 20, is caused, once again, by the fading density causing more than 50% – mostly uncensored – observations in that region.

An aggregate assessment can be made on the basis of mean integrated squared error (MISE), the uniform absolute error (UAE), the mean integrated Kullback-Leibler error (MIKLE), the number of nearest neighbors (NN), the integrated squared bias (IBIAS2), and the integrated variance (IVARIANCE). The averages and standard deviation over the simulations are displayed in Table 1 for the decreasing hazard and Table 2 for the convex hazard.

Table 1: *Loss averages, bandwidth averages, average integrated squared biases, and average integrated variances for decreasing hazard under 40% censoring.*

	k^{RoT}	k^{CV}	$k^{PI}(\leftarrow RoT)$	$k^{PI}(\leftarrow CV)$
Criterion	Average(±Std)	Average(±Std)	Average(±Std)	Average(±Std)
		$n = 50$		
MISE	0.042(±0.07)	0.027(±0.04)	0.165(±0.87)	0.140(±0.87)
UAE	0.069(±0.06)	0.055(±0.03)	0.185(±0.31)	0.153(±0.30)
MIKLE	7.2(±2.2)	7.1(±2.1)	5.6(±2.8)	6.0(±2.7)
NN	13(±2)	21(±7)	26(±64)	34(±65)
IBIAS2	0.041(±0.07)	0.025(±0.04)	0.159(±0.88)	0.136(±0.88)
IVARIANCE	0.0027(±0.002)	0.0030(±0.003)	0.0089(±0.009)	0.0064(±0.007)
		$n = 100$		
MISE	0.020(±0.03)	0.015(±0.02)	0.173(±0.88)	0.107(±0.28)
UAE	0.046(±0.03)	0.043(±0.02)	0.192(±0.33)	0.161(±0.22)
MIKLE	7.8(±2.0)	7.8(±1.9)	6.2(±2.8)	6.5(±2.6)
NN	22(±2)	34(±9)	30(±76)	43(±114)
IBIAS2	0.020(±0.03)	0.014(±0.02)	0.169(±0.88)	0.104(±0.28)
IVARIANCE	0.0007(±0.001)	0.0011(±0.001)	0.0058(±0.007)	0.0042(±0.005)
		$n = 300$		
MISE	0.007(±0.01)	0.006(±0.01)	1.012(±13.85)	0.103(±0.20)
UAE	0.027(±0.01)	0.027(±0.01)	0.186(±0.73)	0.129(±0.15)
MIKLE	8.1(±1.4)	8.1(±1.4)	6.5(±5.1)	6.5(±2.1)
NN	52(±3)	78(±8)	29(±80)	37(±106)
IBIAS2	0.007(±0.01)	0.006(±0.01)	1.006(±13.8)	0.100(±0.21)
IVARIANCE	0.0002(±0.000)	0.0008(±0.001)	0.0102(±0.095)	0.0044(±0.006)

The similarity of the selectors k^{RoT} and k^{CV} in the visual bias assessment extents to the aggregated characteristics. The average losses differ at most by 36% - in the case of the MISE of 0.042, compared to 0.027 - for both hazard shapes and all three losses. The comparison of integrated bias and variance suggests that the bias is slightly larger for k^{RoT} than for k^{CV}, up to 64% - 0.025 compared to 0.041. Consequently, the inverse relation holds for the variance. The higher variance of the cross-validation has already been reported for the fixed-bandwidth kernel density estimation (cf. Hall et al. (1987)). Along with the variance of the estimate, the standard deviation of the nearest-neighbor number k^{RoT} is smaller than k^{CV}, up to seven times - 1 compared to 7. It is puzzling that this coincides with a smaller average k^{RoT}, compared to the average k^{CV}, up to 50% - 52 compared to 78.

The consistency of the kernel hazard rate estimates using k^{RoT} and k^{CV} is well documented in Tables 1 and 2 for MISE and UAE. The mean integrated squared error and uniform absolute error decrease as the sample size n increases, at a rate

Table 2: *Loss averages, bandwidth averages, average integrated squared biases, and average integrated variances for convex hazard under 40% censoring.*

	k^{RoT}	k^{CV}	$k^{PI}(\leftarrow RoT)$	$k^{PI}(\leftarrow CV)$
Criterion	Average(±Std)	Average(±Std)	Average(±Std)	Average(±Std)
		$n = 50$		
MISE	0.037(±0.01)	0.029(±0.05)	143.076(±3198.10)	0.048(±0.14)
UAE	0.058(±0.08)	0.051(± 0.06)	0.915(±18.42)	0.084(±0.13)
MIKLE	7.4(±1.9)	7.3(±1.9)	7.5(±22.3)	6.4(±2.0)
NN	14(±1)	18(±7)	88(±165)	100(±154)
IBIAS2	0.034(±0.01)	0.026(±0.06)	142.803(±3185.78)	0.040(± 0.14)
IVARIANCE	0.0051(± 0.005)	0.0055(±0.005)	0.5504(±12.306)	0.0090(±0.007)
		$n = 100$		
MISE	0.020(±0.03)	0.017(±0.03)	0.049(±0.26)	0.042(±0.25)
UAE	0.040(±0.03)	0.038(±0.02)	0.080(±0.16)	0.073(±0.14)
MIKLE	7.8(±1.65)	7.8(±1.6)	7.2(±1.8)	7.2(±1.7)
NN	25(±1)	28(±5)	78(±128)	82(±126)
IBIAS2	0.018(±0.03)	0.016(±0.03)	0.044(±0.26)	0.037(±0.25)
IVARIANCE	0.0023(±0.002)	0.0025(±0.002)	0.0055(±0.005)	0.0032(±0.003)
		$n = 300$		
MISE	0.008(±0.01)	0.007(±0.01)	0.032(±0.05)	0.015(±0.02)
UAE	0.026(±0.01)	0.026(±0.01)	0.066(±0.07)	0.041(±0.03)
MIKLE	8.3(±1.2)	8.4(±1.1)	8.1(±1.3)	8.2(±1.1)
NN	61(±2)	75(±0.4)	45(±55)	67(±58)
IBIAS2	0.008(±0.01)	0.006(±0.01)	0.032(±0.05)	0.014(±0.02)
IVARIANCE	0.0007(±0.001)	0.0012(±0.001)	0.0006(±0.001)	0.0009(±0.001)

close to n^{-1}. The Kullback-Leibler error, however, is influenced only by the sample size.

Now consider the plug-in selector k^{PI}. The unimportance of the plug-in, k^{RoT} or k^{CV}, already mentioned in the graphic evaluation, can also be seen in the numerical assessment in Tables 1 and 2. Average losses for $k^{PI}(\leftarrow RoT)$ and $k^{PI}(\leftarrow CV)$ are similar for both hazard shapes, only the 300-observation simulation for the decreasing hazard and the 50-observation simulation for the convex hazard are obviously dominated by pathological outliers. However, some minor differences are revealed. In general, plugging k^{RoT} into k^{PI} leads to a larger average loss, larger integrated squared bias, and larger integrated variance. The average number of nearest neighbors is smaller throughout, compared to plugging in k^{CV}.

Compared to the rule of thumb and cross validation, the plug-in selectors are clearly inferior. Only for the convex hazard rate (with the exception of $n = 50$ and $k^{PI}(\leftarrow RoT)$), mean integrated squared error and the uniform absolute error are of a similar

magnitude as for the direct methods (Table 2). For the decreasing hazard rate, MISE and UAE are - with the outlier at $n = 300$ and $k^{PI}(\leftarrow RoT)$ - around ten times higher (Table 1). The main reason is clearly the difficulty of estimating the constants D_1, D_2, and D_3. The contained elements, such as maxima and minima, must be estimated with statistics that converge to the parameters at the known slow rates. As evidence for this, virtually no decrease in the average losses with respect to the small to medium sizes is evident. In contrast, the average Kullback-Leibler losses for $k^{PI}(\leftarrow RoT)$ and $k^{PI}(\leftarrow CV)$ are smaller than for k^{RoT} or k^{RoT}.

In conclusion, the use of a specific loss, in order to asymptotically define the optimal number of nearest neighbor, does not automatically lead to a superiority with respect to that loss, in a finite sample situation. A preference for a specific loss should be combined with the size of the sample to be analyzed.

6 Applications

After the assessment of bias in simulations, that is the averages over several samples, it is of interest to determine how the estimate performs in single sample situations. We apply now the nearest neighbor estimate (2) to two sets of study data, a financial and a medical study. The hazard rate estimates with the four bandwidth selectors, namely the numbers of nearest neighbors k^{RoT}, k^{CV}, $k^{PI}(\leftarrow RoT)$, and $k^{PI}(\leftarrow CV)$, are depicted in Figure 4.

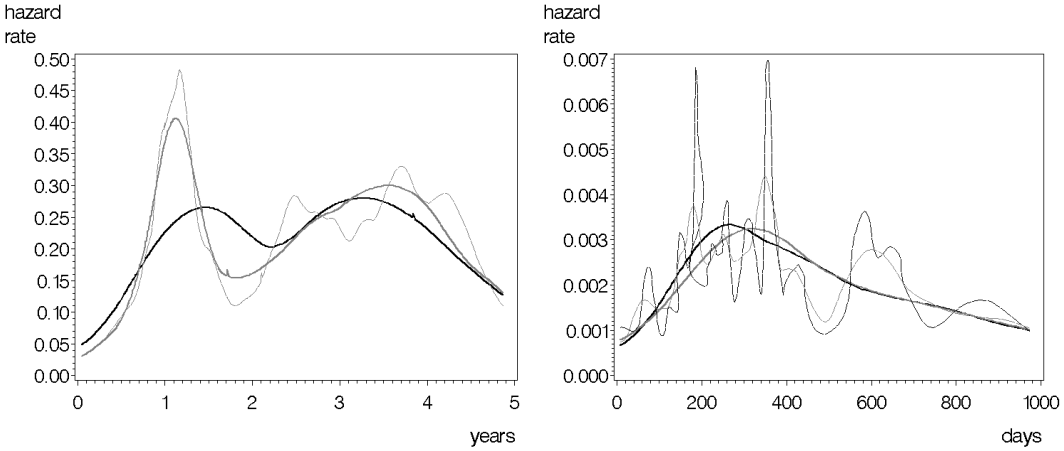

Figure 4: Hazard rate estimate with k^{RoT} (black-solid), $k^{PI}(\leftarrow RoT)$ (black-dashed), k^{CV} (grey-solid), and $k^{PI}(\leftarrow CV)$ (grey-dashed) nearest neighbors for rating study (left) (NN=123/38/78/38), bladder cancer study (right) (NN=69/9/80/15)

First, the data on credit rating migrations from Weißbach et al. (2008b); Weißbach and Dette (2007); Weißbach et al. (2009) is re-analyzed. In this study, the duration of

359 counterparts to a bank are followed, with the onset at entry into the portfolio, until they migrate to an adjacent rating class. The censoring events are either termination of contract, the end of the seven year study, or migration to any but the adjacent class. As a result, 60% of the observations are censored. The two estimates with $k^{RoT} = 123$ and $k^{CV} = 78$ are bimodal in shape, the modes are more evident for k^{CV} than for k^{RoT}. Although, in the simulations in Tables 1 and 2, the number of nearest neighbors for the rule of thumb is on average smaller, in this study, k^{RoT} is now larger than k^{CV}. An explanation is that, in most simulations, the cross-validation is more variable than the rule of thumb.

The shape supports the findings of Weißbach et al. (2008b) and Weißbach and Dette (2007), that rating migration does not follow a homogeneous Markov process, that is that migration hazard rates are not constant. However, this interpretation was made in Weißbach et al. (2008b) and Weißbach and Dette (2007) by significance testing and in Weißbach et al. (2008b) by a visual assessment of the *cumulative* hazard rates, whereas in the present paper, we are able to depict the hazard rates themselves. The plug-in number of nearest neighbors $k^{PI}(\leftarrow RoT)$ and $k^{PI}(\leftarrow CV)$ are both 38 in this example. The finding that the pilot estimate, k^{RoT} or k^{CV}, has little impact on k^{PI}, was already found in the simulations in the preceding section. However, the equality in a single sample case is remarkable. The simulations already indicate that the plug-in bandwidth is more variable than the direct methods. Accordingly, it is not surprising that the shape of the plug-in estimation differs from the two others. The difference for the k^{CV}-based estimate in the firth mode is not very pronounced. In the second mode, however, the plug-in suggests some further small modes. The plug-in also provides evidence suggesting a lack of constancy of the hazard rate. The large migration activity shortly after portfolio entrance, already found in Weißbach et al. (2008b), is supported here, especially by cross-validation and the plug-in methods.

Second, we re-analyze the bladder cancer study by Siu et al. (1998) with 114 observations and, due the lethal character of bladder cancer, with only 15% censoring. Weißbach (2006) has already explored the data in order to show that kernel estimation with k^{CV} for three groups, stratified by a categorical medical performance measure, allows almost a strict ordering of the hazard functions. Different bandwidth selectors are not compared. In general, the cancer study has a much lower scale, namely up to 0.007 compared to 0.5 for the rating study. This is mainly due to the time scale, for rating migrations, yearly and daily for cancer deaths.

Considering this in more detail, $k^{RoT} = 69$ is now smaller than $k^{CV} = 80$, in agreement with the findings in the simulation for the sample size of 100 (see again Tables 1 and 2). For the bladder cancer hazard, this implies a steeper increase and

an earlier maximum. Both k^{RoT} and k^{CV} indicate the same unimodal, left-skewed shape. As in the rating example, the plug-in bandwidths $k^{PI}(\leftarrow RoT) = 9$ and $k^{PI}(\leftarrow CV) = 15$ are smaller than the direct methods k^{RoT} and k^{CV}. Contrary to the rating example and although the pilot bandwidths are even closer to one another here, the plug-in estimates for $k^{PI}(\leftarrow RoT)$ and $k^{PI}(\leftarrow CV)$ are now different. Both estimates indicate a large number of modes. On the other hand, it is interesting to see that even maximal daily hazards of almost 0.007 can be supported, whereas the direct methods only support maxima that are half as high, and smaller bandwidths result in less biased estimates.

7 Conclusions

We based a monte carlo simulation study for the kernel hazard rate estimation on the exponentiated Weibull distribution family, thus modeling important shapes in financial applications. Motivated by the intrinsic problem of censoring in failure-time analysis, we propose the nearest neighbor bandwidth and investigate three alternatives for selecting the number of nearest neighbors. Assessing the performance by means of visual bias assessment, combined with a numerical assessment of average integrated losses, the recently developed rule of thumb for the nearest neighbor selection demonstrates superiority in many aspects. However, this can change when a particular loss is considered exclusively. Preferences about the loss criterium remain an unresolved problem, considered for example in Marron and Tsybakov (1995).

Acknowledgement. R. Weißbach is indebted to O. Gefeller for initiating the project and to B. Bloch for editing the manuscript. Furthermore, the financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures," project B1, and Grant Ge 637/3) is gratefully acknowledged.

References

O.O. Aalen and S. Johansen. An empirical transition matrix for non-homogeneous markov chains based on censored observation. *Scandinavian Journal of Statistics*, 5:141–150, 1978.

H.D.I. Abarbanel. *Analysis of observed chaotic data.* Springer, Heidelberg, New York, 1996.

P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes.* Springer-Verlag New York, 1993.

Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards. Technical report, Bank for International Settlements, 2004.

T.R. Bielecki and M. Rutkowski. *Credit Risk: Modeling, Valuation and Hedging*. Springer, New York, 2002.

C. Bluhm, L. Overbeck, and C. Wagner. *An Introduction to Credit Risk Modeling*. Chapman & Hall, 2002.

Y.S. Chow, S. Geman, and L.D. Wu. Consistent cross-validated density estimation. *Annals of Statistics*, 11:25–38, 1983.

H. Dette and O. Gefeller. Definitions of nearest neighbour distances for censored data on the nearest neighbour kernel estimators of the hazard rate. *Journal of Nonparametric Statistics*, 4:271–282, 1995.

U. Einmahl and D.M. Mason. Uniform in bandwidth consistency of kernel-type function estimators. *Annals of Statistics*, 33:1380–1403, 2005.

O. Gefeller, R. Weißbach, and T. Bregenzer. The implementation of a data-driven selection procedure for the smoothing parameter in nonparametric hazard rate estimation using sas/iml software. In H. Friedl, A. Berghold, and G. Kauermann, editors, *Proceedings of the 13th SAS European Users Group International Conference, Stockholm*, pages 1288–1300. SAS Institute Inc. Carry, July 1996.

K. Günther, J. Leier, G. Henning, A. Dimmler, R. Weißbach, W. Hohenberger, and R. Förster. Prediction of lymph node metastasis in colorectal carcinoma by chemokine receptors CCR7 and CXCR4. *International Journal of Cancer*, 116: 726–733, 2005.

G.M. Gupton, C.C. Finger, and M. Bhatia. CreditMetrics. Technical report, J.P.Morgan, 1997.

P. Hall. Cross-validation in density estimation. *Biometrika*, 69:383–390, 1978.

P. Hall, T.C. Hu, and J.S. Marron. On the amount of noise inherent in bandwidth selection for a kernel density estimator. *Annals of Statistics*, 15:163–181, 1987.

P. Hall, S.J. Sheather, M.C. Jones, and J.S. Marron. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, 78:263–269, 1991.

M.C. Jones, J.S. Marron, and S.J. Sheather. A brief survey on bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91:401–407, 1996.

E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

N.M. Kiefer and C.E. Larson. A simulation estimator for testing the time homogeneity of credit rating transitions. *Journal of Empirical Finance*, 14:818–835, 2007.

D. Lando and T. Skødeberg. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance*, 26:423–444, 2002.

J.S. Marron. A comparison of cross-validation techniques in density estimation. *Annals of Statistics*, 15:152–163, 1987.

J.S. Marron and J. de Uña Álvarez. Sizer for length biased, censored density and hazard estimation. *Journal of Statistical Planning and Inference*, 121:149–161, 2004.

J.S. Marron and A.B. Tsybakov. Visual error for qualitative smoothing. *Journal of the American Statistical Association*, 90:499–507, 1995.

W. Gonzàlez-Manteiga R. Cao J.S. Marron. Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. *Journal of the American Statistical Association*, 91:1130–1140, 1996.

G.S. Mudholkar, D.K. Srivastava, and M. Freimer. The exponential weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*, 37:436–445, 1995.

H.-G. Müller and J.-L. Wang. Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, 50:61–76, 1994.

B.U. Park and J.S. Marron. Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85:66–72, 1990.

E. Parzen. On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

P.N. Patil. On the least square cross-validation bandwidth in hazard rate estimation. *Annals of Statistics*, 21:1792–1810, 1993.

S.S. Ralescu. The law of the iterated logarithm for the multivariate nearest neighbor density estimators. *Journal of Multivariate Analysis*, 53:159–179, 1995.

H. Schäfer. A note on data-adaptive kernel estimation of the hazard and density function in the random censorship situation. *Annals of Statistics*, 13:818–820, 1985.

D.W. Scott and G.R. Terrell. Biased and unbiased cross-validation in density estmation. *Journal of the American Statistical Association*, 82:1131–1146, 1987.

B.W. Silverman. *Density Estimation*. Chapman & Hall, London, 1986.

L.L. Siu, D. Banerjee, R.J. Khurana, X. Pan, R. Weißbach, I.F. Tannock, and M.J. Moore. The prognostic role of p53, metallothionein, p-glycoprotein, and mib-1 in muscle-invasive urothelial transitional cell carcinoma. *Clinical Cancer Research*, 4:559–565, 1998.

G. Sugihara and R.M. May. Nonlinear forecasting as a way of distinguishing chaos from measurment error in time series. *Nature*, 344:734–741, 1990.

T.J. Wagner. Nonparametric estimates of probability densities. *IEEE Transactions on Information Theory*, 21:438–440, 1975.

M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman & Hall, London, 1995.

R. Weißbach. A general kernel functional estimator with general bandwidth – strong consistency and applications. *Journal of Nonparametric Statistics*, 18:1–12, 2006.

R. Weißbach and H. Dette. Kolmogorov-Smirnov-type testing for the partial homogeneity of markov processes - with application to credit risk. *Applied Stochastic Models in Business and Industry*, 22:223–234, 2007.

R. Weißbach, F. Kramer, and C. Lawrenz. *The VAR Implementation Handbook*, chapter Aggregating and Combining Ratings, page to appear. McGraw-Hill, 2009.

R. Weißbach, A. Pfahlberg, and O. Gefeller. Double-smoothing in kernel hazard rate estimation. *Methods of Information in Medicine*, 47:167–173, 2008a.

R. Weißbach, P. Tschiersch, and C. Lawrenz. Testing time-homogeneity of rating transitions after origination of debt. *Empirical Economics*, to appear (DOI: 10.1007/s00181-008-0212-3), 2008b.