

THE INSTITUTE  
OF STATISTICS

THE CONSOLIDATED UNIVERSITY  
OF NORTH CAROLINA



BOOTSTRAP SELECTION OF THE SMOOTHING  
PARAMETER IN NONPARAMETRIC HAZARD RATE ESTIMATION

by

W. González-Manteiga, R. Cao and J.S. Marron

February 1994

Micro Series #2315

DEPARTMENT OF STATISTICS  
Chapel Hill, North Carolina

# Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation\*

W. González-Manteiga , R. Cao and J.S. Marron

February 14, 1994

## Abstract

An asymptotic representation of the mean weighted integrated squared error for the kernel based estimator of the hazard rate in the presence of right censored samples is obtained for different bootstrap resampling methods. As a consequence, a new bandwidth selector based on the bootstrap is introduced. Very satisfactory simulations results are obtained in comparison to the cross-validation selector for different models, using WARPed, i.e. binned versions of the estimators.

**AMS Subject Classification:** 62G07

**KEY WORDS:** Asymptotic bootstrap mean integrated squared error; bandwidth; binning; censored data; bandwidth; nonparametric hazard rate estimation; warping.

## 1 Introduction.

In survival analysis and reliability, a curve which provides useful information about a variable  $X^0$  is the “hazard rate”

$$r_0(x) = \lim_{dx \rightarrow 0} P(X^0 \leq x + dx | X^0 > x) \frac{1}{dx} = \frac{f^0(x)}{1 - F^0(x)}.$$

---

\*Partially supported by the Grants PTR91-0058, PB91-0794 and XUGA20701B92 for the first two authors and by NSF Grant DMS-9203135 for the third author

The curve  $r_0$  represents the instantaneous failure time at every  $x$ , when  $X^0$  is assumed to denote the lifetime of a system. See Watson and Leadbetter (1964a) for discussion of the many useful properties of  $r_0$ . It can be represented in terms of the density  $f^0$  and the cumulative distribution  $F^0$  of  $X^0$  as  $r_0(x) = \frac{f^0(x)}{1-F^0(x)}$ . When no parametric assumptions are made about the underlying distribution, kernel estimators provide a useful tool for estimating  $r_0$ . There are several such available, as defined and discussed in section 2.1.

It is frequently not possible to observe a complete random sample of size  $n$  from the  $X^0$  population. An effective model for such situations is random right censorship, as discussed by Koziol and Green (1976). Mathematical formulation of this estimation is given in section 2.2.

An important problem when estimating the hazard rate function by kernel methods is the choice of the smoothing parameter (i.e. bandwidth) denoted here as  $h$ . Following ideas used in density estimation, Tanner and Wong (1984) proposed a maximum likelihood cross-validation method for uncensored hazard estimation. We do not consider that method here, because least squares cross-validation has been shown to be more appropriate. See Cao, Cuevas and González-Manteiga (1993) for convincing simulation evidence, and see Hall (1987) for theoretical insights.

Least squares cross-validation was adapted to censored density estimation by Marron and Padgett (1987). It was extended to uncensored hazard estimation by Patil (1990) and by Sarda and Vieu (1991) and a version in the censored hazard case was proposed in Patil (1990,1993). In our context the cross-validated bandwidth is the minimizer  $\hat{h}$  of the score function 2 defined in section 2.3. These authors have established asymptotic optimality, in the sense

$$\frac{ISE(\hat{h})}{\inf_{h \in H_n} ISE(h)} \rightarrow 1, a.s.$$

where  $ISE$  is the integrated squared error  $ISE(h) = \int (\hat{r}_{0,h}^1 - r_0)^2 w$ , for a suitable weight function  $w$ . This weight function plays an important role. In particular, compact support, contained in  $(0, T)$ , is required for  $w$ , where  $T$  is a value such that  $1 - F^0(T) > 0$ . The range of values for the smoothing parameter  $h$ ,  $H_n$ , is set following regularity conditions as stated in Patil (1990) or in Sarda and Vieu (1991). There exist other studies concerning local choices of the bandwidth. This has been done by Müller and Wang (1990), for instance.

While the above asymptotic optimality seems encouraging, it has been

observed in a variety of related contexts that the rate of convergence for cross-validation methods is unacceptably slow, see Härdle, Hall and Marron (1988) and references therein. This has led to a search for improved methods of bandwidth selection. Among the best of these is the “smoothed bootstrap” approach, see Marron (1992) for an introduction and more references in the context of density estimation. The main goal of this paper is to develop such effective methods of bootstrap bandwidth selection in the case of hazard and censored hazard function estimation. In Section 3 we introduce several bootstrap mechanisms to select the parameter  $h$  for the estimator  $\hat{r}_{0,h}$ . Some asymptotic results are obtained which validate the resampling method proposed.

As fast computation is often an important need in both curve estimation and bandwidth selection, we discuss some fast algorithms in Section 4. These are based on WARP (i.e. binning) ideas as discussed in Härdle (1991). See Fan and Marron (1994) for additional insights, for compelling evidence as to how much faster this method is, and for alternative fast methods of calculation. We present fast methods for computing both the estimator given in 1 and also the various bandwidth selectors discussed here.

In Section 5 we give some results of a Monte Carlo comparison of our smoothed bootstrap bandwidth selection methods, with least squares cross-validation. As anticipated, the smoothed bootstrap is much more stable and effective.

The Appendix contains the proofs.

## 2 Technical Background

In this section, we give formal definitions of kernel estimators of the hazard rate  $r_0$ , and also formulate the censored data setting. Readers familiar with these topics may want to scan this section quickly with the idea of only becoming familiar with our notation.

### 2.1 Kernel hazard estimators.

Kernel estimators of the hazard function  $r_0$ , from a random sample  $X_1^0, \dots, X_n^0$ , have been developed by Grenander (1956), Murthy (1965), Rice and Rosenblatt (1976), Földes, Retjö, and Winter (1981) and Sethuraman and Singpurwalla (1981). While the bandwidth selection ideas we develop apply equally

well to all of these proposals, we focus here only on

$$\hat{r}_{0,h}(x) = \int K_h(x-u) \frac{dF_n^0(u)}{1-F_n^0(u)} \simeq \sum_{i=1}^n \frac{1}{h} K\left(\frac{x-X_{(i)}^0}{h}\right) \frac{1}{n-i+1},$$

where  $\{X_{(i)}^0\}_{i=1}^n$  is the ordered sample. This estimator was introduced by Watson and Leadbetter (1964a, 1964b). This one has been chosen because of its simplicity, because it gives a smooth estimate, and is the “most natural” in the sense of Patil, Wells and Marron (1994).

## 2.2 Censored data kernel estimators.

Under random right censorship, the data from the population of interest,  $X_1^0, \dots, X_n^0$  are not completely observable (e.g. time until death by a certain cause can be rendered unobservable when the subject dies first from a different cause). For each  $i = 1, \dots, n$ , under the random right censoring model there is a “censoring variable”  $Z_i$  (also unobserved, e.g. the time of death by some different cause), which obscures  $X_i^0$  when it is smaller. The observed data are a collection of pairs:  $\{(X_i, \Delta_i)\}_{i=1}^n$ , where  $X_i = \min\{X_i^0, Z_i\}$ ,  $\Delta_i = 1_{\{X_i = X_i^0\}}$ ,  $i = 1, 2, \dots, n$ . The variable  $\Delta_i$  codifies the presence or absence of censoring by

$$\Delta_i = \begin{cases} 0 & \text{if } X_i = Z_i \\ 1 & \text{if } X_i = X_i^0 \end{cases}.$$

We will assume that  $X^0$  and  $Z$  are independent.

As above,  $f^0$  and  $F^0$  denote the density and distributions for  $X^0$ . Let  $h^0$  and  $H^0$  denote the analogs for  $Z$ . The distribution function of the observed variable  $X$ , with density  $f$ , can be written as  $F = 1 - (1 - F^0)(1 - H^0)$ , because of the independence.

The key to extending the estimator in section 2.1 to the censored case is direct replacement of the empirical distribution function  $F_n^0$  by the “censored version” introduced by Kaplan and Meier (1958):

$$1 - F_n^0(x) = \begin{cases} 1 & \text{if } x < X_{(1)} \\ \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1}\right)^{\Delta_j} & \text{if } X_{(i-1)} \leq x < X_{(i)} \\ 0 & \text{if } x \geq X_{(n)} \end{cases},$$

which is usefully rewritten as

$$\hat{r}_{0,h}(x) = n^{-1} \sum_{i=1}^n \frac{K_h(x - X_i)\Delta_i}{1 - F_n(X_i)} \quad (1)$$

where  $F_n(x) = n^{-1} \sum_{i=1}^n 1_{[X_i \leq x]}$ , the empirical distribution function constructed using the observed sample  $\{X_1, X_2, \dots, X_n\}$ .

This and similar adaptations of other kernel hazard estimators have been studied by Blum and Susarla (1980), Yandell (1981,1983), by Tanner and Wong (1984), by Cheng (1987), by Lo, Mack and Wang (1989), and by Patil (1990). It is seen there that these estimators have the same asymptotic variance:

$$\frac{r_0(x)}{1 - F(x)} \int K^2 \frac{1}{nh}$$

but different asymptotic bias. In fact, the asymptotic bias of the estimator in section 2.1 is given by:

$$E(\hat{r}_{0,h}(x)) = \int \frac{1}{h} K\left(\frac{x-u}{h}\right)(1 - F(u)^n) d\Lambda_0(u),$$

where

$$\Lambda_0(x) = \int_0^x \frac{f_1^0(t)p}{1 - F(t)} dt,$$

$f_1^0$  is the conditional density of  $X|\Delta = 1$  and  $p = P(\Delta = 1)$  is the proportion of censoring.

### 2.3 Least Squares Cross Validation

See Rudemo (1982) and Bowman (1984) for motivation of this method of selecting the bandwidth for kernel density estimation. Their idea was adapted to the case of censored hazard estimation by Patil (1990), see page 36. The Least Squares Cross Validated bandwidth is the minimizer of

$$CV(h) = \int \hat{r}_{0,h}(x)^2 w(x) dx - 2n^{-1} \sum_{i=1}^n \frac{\hat{r}_{0,h}^{-i}(X_i)}{(1 - F_n(X_i))} w(X_i^0) \Delta_i \quad (2)$$

where  $\hat{r}_{0,h}^{-i}$  is the "leave one out" estimator given by

$$\hat{r}_{0,h}(x) = (n-1)^{-1} \sum_{j=1, j \neq i}^n \frac{K_h(x - X_j)\Delta_j}{1 - F_n(X_j)}.$$

### 3 Bootstrap bandwidth selectors

To select the bandwidth  $h$  in 1, we will focus on the estimation of the  $h$  which minimizes the mean integrated squared error:

$$MISE_w(h) = E \left[ \int (\hat{r}_{0,h} - r_0)^2 w \right] \quad (3)$$

where the weight function  $w$  satisfies the assumptions of Section 1.

An asymptotic representation of  $MISE_w$  is given in the following result, proved by Tanner and Wong (1983),

**Theorem 3.1** *Under the assumptions given in Section 1, when  $n \rightarrow \infty$ ,  $h \rightarrow 0$  and  $nh \rightarrow \infty$ :*

$$MISE_w(h) = AMISE_w(h) + o(AMISE_w(h))$$

where

$$AMISE_w(h) = \int \left( p(K_h * \frac{f_1^0}{1-F} - \frac{f_1^0}{1-F}) \right)^2 w + \frac{1}{nh} \int K^2 p \int \frac{f_1^0 w}{(1-F)^2}$$

\* denotes convolution.

The expression given for  $AMISE_w$  comes from the integrated squared bias and the asymptotic integrated variance. See Patil (1990) for a variety of similar results.

The bootstrap method is a resampling device to approximate the distribution of statistics of interest. In this case the statistic of interest is  $\hat{r}_{0,h}$ , and the aspect of its distribution that we will approximate is  $MISE_w$ . We define the *bootstrap selector of the bandwidth*,  $h^*$ , to be the minimizer of the estimated  $MISE_w$ . As noted in Marron (1992), “smoothed bootstraps” are needed to properly approximate the important “bias” part of  $MISE_w$ .

Bootstrapping is more complicated than usual in the censored case, as there are several possible approaches. Here we discuss three, (although they are really only two, since two turn out to be the same) which mostly differ in when one should do the “smooth” part:

#### 3.1 SC resampling plan

This *smoothed-censored* (SC) resampling plan is a smooth version of Efron’s resampling, originally introduced by Efron (1981). Here we introduce “smooth

bootstrap" ideas as early as possible. Since  $F^0$  and  $H^0$  are absolutely continuous, they can be estimated by the smoothed Kaplan-Meier estimators:  $F_n^0 * K_{g_1}$  and  $H_n^0 * K_{g_2}$ . Here,  $F_n^0$  and  $H_n^0$  denote the Kaplan-Meier estimators of  $F^0$  and  $H^0$  and  $g_1$  and  $g_2$  are two initial smoothing parameters (or pilot bandwidths). Now, the method proceeds as in Efron (1981), but replacing the Kaplan-Meier estimators by these smoothed versions of them.

1. First draw bootstrap resamples  $\{X_1^{0*}, \dots, X_n^{0*}\}$  from  $F_n^0 * K_{g_1}$ .
2. Generate independent bootstrap resamples  $\{Z_1^*, \dots, Z_n^*\}$  from  $H_n^0 * K_{g_2}$ .
3. Finally construct  $\{(X_i^{0*}, \Delta_i^*)\}_{i=1}^n$ , where  $\Delta_i^* = 1_{\{X_i^{0*} = Z_i^*\}}$  and  $X_i^* = \min\{X_i^{0*}, Z_i^*\}$ ,  $i = 1, 2, \dots, n$ .

Similar arguments to those used in the proof of Theorem 3.1 lead to:

$$AMISE_w^* = \int \left[ \hat{p}_{g_1, g_2} \left( K_h * \frac{\hat{f}_{1, g_1, g_2}^0}{1 - \hat{F}_{g_1, g_2}} - \frac{\hat{f}_{1, g_1, g_2}^0}{1 - \hat{F}_{g_1, g_2}} \right) \right]^2 w + \frac{1}{nh} \int K^2 \hat{p}_{g_1, g_2} \int \frac{\hat{f}_{1, g_1, g_2}^0 \cdot w}{(1 - \hat{F}_{g_1, g_2})^2}, \quad (4)$$

where

$$1 - \hat{F}_{g_1, g_2} = (1 - F_n^0 * K_{g_1})(1 - H_n^0 * K_{g_2}), \quad \hat{p}_{g_1, g_2} = P_*\{X^{0*} \leq Z^*\}$$

and

$$\hat{f}_{1, g_1, g_2}^0 = \frac{\hat{f}_{g_1}^0 (1 - \hat{F}_{g_1, g_2})}{\hat{p}_{g_1, g_2} (1 - F_n^0 * K_{g_1})}.$$

The distributions  $F_n^0 * K_{g_1}$  and  $H_n^0 * K_{g_2}$ , in the first two steps of the algorithm, denote those ones having densities  $\hat{f}_{g_1}^0(x) = \int K_{g_1}(x - u) dF_n^0(u)$  and  $\hat{h}_{g_2}^0(x) = \int K_{g_2}(x - u) dH_n^0(u)$  respectively.

The expression 4, shows how the bootstrap mechanism mimics the theoretical asymptotic integrated weighted squared error in 3.1. As indicated above, smooth resampling (which is the only difference with respect to Efron's (1981) proposal) is essential for the bias part of 4. A possible bootstrap bandwidth selector could be defined as the minimizer of 4. An alternative is to minimize  $MISE_w^*$  itself. However this is not easy since no explicit expression is available for  $MISE_w^*$  in this context, which has been very important in avoiding the time-consuming Monte Carlo approximation



(see, for instance, Marron (1992) or Cao, Cuevas and González-Manteiga (1993), where formulas which avoid simulation have been developed in the context of density estimation).

### 3.2 CS resampling plans

With this notation we refer to resampling plans where the *censoring part is done first and then the smoothing is performed*. We will consider two possibilities for these plans.

**The CS1 plan.** It is a *smooth version of the naive resampling*:

1. For  $i = 1, \dots, n$  generate  $\varepsilon_i$  from the density  $K_{g_1}$  if  $\Delta_i = 1$  and with density  $K_{g_2}$  if  $\Delta_i = 0$ .
2. Draw a random sample  $\{(X_i^*, \Delta_i^*)\}_{i=1}^n$  choosing every observation uniformly from the set  $\{(X_i + \varepsilon_i, \Delta_i)\}_{i=1}^n$ .

An equivalent formulation is to draw independently  $I_i$ , uniformly from  $\{1, 2, \dots, n\}$ ,  $R_i$  from the density  $K$ , and then define  $X_i^* = X_{I_i} + \Delta_i^* g_1 R_i + (1 - \Delta_i^*) g_2 R_i$ .

**The CS2 plan.** The idea behind this method is to *modify Efron's resampling* in order to incorporate the smoothing effect.

1. First draw bootstrap resamples  $\{\bar{X}_i^{0*}\}_{i=1}^n$ , from  $F_n^0$ .
2. Generate independent bootstrap resamples  $\{\bar{Z}_i^*\}_{i=1}^n$ , from  $H_n^0$ .
3. For  $i = 1, 2, \dots, n$  let  $\Delta_i^* = 1_{\{\bar{X}_i^{0*} < \bar{Z}_i^*\}}$ .
4. Draw  $X_i^*$  from  $F_n^0 * K_{g_1}$  if  $\Delta_i^* = 1$  and from  $H_n^0 * K_{g_2}$  if  $\Delta_i^* = 0$ .

An equivalent formulation is to draw  $I_{1,i}$  and  $I_{2,i}$  independently from the set  $\{1, 2, \dots, n\}$  according to the Kaplan-Meier distributions  $F_n^0$  and  $H_n^0$  respectively. Define  $\Delta_i^* = 1_{\{X_{I_{1,i}} < X_{I_{2,i}}\}}$  and

$$X_i^* = \begin{cases} X_{I_{1,i}} + g_1 R_i & \text{if } \Delta_i^* = 1 \\ X_{I_{2,i}} + g_2 R_i & \text{if } \Delta_i^* = 0 \end{cases}$$

where  $R_i$ ,  $i = 1, 2, \dots, n$  is a random sample generated from the common density  $K$ .

The following result connects the last two resampling plans:

**Theorem 3.2** *Under the assumptions in 3.1, the resampling plans CS1 and CS2 are equivalent. Under either of these resamplings we have the following asymptotic representation of the bootstrap MISE:*

$$AMISE_w^* = \int \left[ \hat{p} \left( K_h * \frac{\hat{f}_{1,g_1}^0}{1 - \hat{F}_n} - \frac{\hat{f}_{1,g_1}^0}{1 - \hat{F}_n} \right) \right]^2 w + \frac{1}{nh} \int K^2 \hat{p} \int \frac{\hat{f}_{1,g_1}^0 w}{(1 - \hat{F}_n)^2}, \quad (5)$$

where  $\hat{f}_{1,g_1}^0$  is the kernel density estimator of  $f_1^0$  constructed with the uncensored observations,  $\hat{p} = n^{-1} \sum \Delta_i$  and  $\hat{F}_n = \hat{p}(K_{g_1} * F_{1,n}^0) + (1 - \hat{p})(K_{g_2} * H_{1,n}^0)$ . The functions  $F_{1,n}^0$  and  $H_{1,n}^0$  are the empirical distribution constructed with the uncensored and censored observations respectively.

Observe that these two equivalent resamplings also mimic the theoretical  $AMISE_w$ , as the SE method did. Now, the estimators are simpler (compare for instance  $\hat{p}$  versus  $\hat{p}_{g_1, g_2}$ ). The equivalence in 3.2 ensures that the method can be implemented in such a way that no Kaplan-Meier weights need to be computed in order to generate the resamples (in other terms the presentation given in CS1).

Once more, expression 5 gives no explicit formula for  $MISE_w^*$ , but for  $AMISE_w^*$ . As above, we decided to use the minimizer,  $h^*$ , of  $AMISE_w^*$  instead of carrying out a Monte Carlo approximation of  $MISE_w^*$ . This is the bootstrap bandwidth selector which is used in the simulation study of Section 5.

## 4 WARPing algorithms

This section is devoted to the design of fast algorithms to construct the estimator, and also fast algorithms for bandwidth selection, including both traditional least squares cross-validation, and our smoothed bootstrap methods.

As discussed in Härdle (1991) and Fan and Marron (1994), the key to the speed of the WARPing idea is to “bin” the data to an *equally spaced* grid, so that the number of kernel evaluations can be drastically reduced. For bins, take the sets

$$\{B_z\}_{z \in \mathbb{Z}} = \left\{ \left[ \frac{zh}{M}, \frac{(z+1)h}{M} \right) \right\}_{z \in \mathbb{Z}},$$

where every interval has length  $\delta = \frac{h}{M}$  and  $M \in \mathbb{Z}^+$ . The censored data is then summarized by the  $n_z = \sum_{i=1}^n 1_{B_z}(X_i)\Delta_i$ ,  $z \in \mathbb{Z}$ , the number of uncensored observation in the bin  $B_z$ . The large number of kernel differences  $K_h(x - X_i)$  can be approximated by the much smaller set of differences  $w_M(k) = K(\frac{k}{M})$ , for  $k = 1 - M, \dots, M - 1$ , when  $K$  is supported on  $[-1, 1]$ . Fan and Marron (1994) show that this results in computational speed savings of factors up to 100. They also discuss variations such as refined types of binning.

#### 4.1 WARPing the kernel estimator

A WARPing approximation of  $\hat{r}_{0,h}$ , at each bin indexed by  $z$  as in the previous section, is given by

$$\hat{r}_{0,M}(z) = (n\delta M)^{-1} \sum_{k=1-M}^{M-1} \frac{w_M(k)n_{z+k}}{1 - F_M(z+k)} \quad (6)$$

where

$$1 - F_M(z) = n^{-1} \sum_{k>z} \tilde{n}_k$$

for  $\tilde{n}_k = \sum_{i=1}^n 1_{B_k}(X_i)$ , the number of all (not just uncensored) observation in the bin  $B_k$ . Note that  $F_M$  is the cumulative histogram of absolute frequencies in the bins  $B_k$ .

For a fixed  $h$  and letting  $M \rightarrow \infty$ , we have  $\hat{r}_{0,M}(x) \rightarrow \hat{r}_{0,h}(x)$ . I.e. the WARPing approximation error is small when the rounding error,  $\delta$ , is small. This point is illustrated in Figure 1.

[put figure 1 about here]

FIGURE 1: *Simulated censored hazard estimation example.*

Figure 1 shows the true hazard rate function (solid thick line) for a Weibull(2,1) model, the kernel estimator (dotted thin line) –computed directly upon a sample of 100 uncensored observations of such a model– and two WARPing approximations of this curve:  $\delta = 1$  (dashed thin line) and  $\delta = 0.01$  (dashed thick line). The role of the rounding error,  $\delta$ , becomes clear: the approximation to the kernel estimator gets better as  $\delta$  decreases to zero.

## 4.2 WARPing cross-validation

A WARPed version of the score function 2 in 2.3 involves essentially replacing the conventional kernel estimators by their WARPed versions. However, some additional work is required to get a version which may be rapidly computed. Define the “leave one out bin counts”  $n_z^{-i} = \sum_{j=1, j \neq i}^n 1_{B_z}(X_j) \Delta_j$ ,  $z \in Z$ , and use the approximations

$$\int \hat{r}_{0,h}(x)^2 w(x) dx \simeq \delta \sum_{z \in Z} \hat{r}_{0,M}(z)^2 w_{M,h}(z)$$

and

$$\begin{aligned} & -2n^{-1} \sum_{i=1}^n \frac{\hat{r}_{0,h}^i(X_i) w(X_i) \Delta_i}{1 - F_n(X_i)} \simeq -2n^{-1} \sum_{i=1}^n \frac{\sum_{z \in Z} 1_{B_z}(X_i) \hat{r}_i}{1 - F_n(X_i)} = \\ & = -2n^{-1} \sum_{i=1}^n \frac{\sum_z 1_{B_z}(X_i) \left[ (n\delta M)^{-1} \sum_{k=1-M}^{M-1} \frac{w_M(k) n_{z+k}^{-i}}{1 - F_M^{-1}(z+k)} \right] w_{M,h}(z) \Delta_i}{1 - F_n(X_i)} \simeq \\ & \simeq -2n^{-1} \sum_{i=1}^n \frac{\sum_z 1_{B_z}(X_i) (n\delta M)^{-1} \left[ \sum_{k=1-M}^{M-1} \frac{w_M(k) n_{z+k}}{1 - F_M(z+k)} - \frac{w_M(0)}{1 - F_M(z)} \right] w_{M,h}(z) \Delta_i}{1 - F_n(X_i)} \simeq \\ & \simeq -2n^{-1} \sum_{i=1}^n \sum_{z \in Z} \frac{1_{B_z}(X_i) \frac{n}{n-1} \hat{r}_{0,M}(z) w_{M,h}(z) \Delta_i}{1 - F_M(z)} + \\ & + 2n^{-1} \sum_{i=1}^n \sum_{z \in Z} \frac{1_{B_z}(X_i) \frac{w_M(0)}{(n-1)\delta M(1 - F_M(z))} w_{M,h}(z) \Delta_i}{1 - F_M(z)} = \\ & = -\frac{2}{n-1} \sum_{z \in Z} \frac{n_z \hat{r}_{0,M}(z) w_{M,h}(z)}{1 - F_M(z)} + \frac{2w_M(0)}{n(n-1)\delta M} \sum_{z \in Z} \frac{n_z w_{M,h}(z)}{(1 - F_M(z))^2}, \end{aligned}$$

where  $w_{M,h}(z)$  denotes the value of  $w$  at the lower limit of  $B_z$ . Putting these together yields the approximated score function

$$\begin{aligned} \overline{CV}(M) &= \delta \sum_{z \in Z} \hat{r}_{0,M}(z)^2 w_{M,h}(z) \\ & - \frac{2}{n-1} \sum_{z \in Z} \frac{n_z \hat{r}_{0,M}(z) w_{M,h}(z)}{1 - F_M(z)} + \frac{2w_M(0)}{(n-1)\delta M} \sum_{z \in Z} \frac{n_z w_{M,h}(z)}{n(1 - F_M(z))^2} \end{aligned} \quad (7)$$

Note that for  $h$  fixed and  $\delta$  small enough,  $\overline{CV}(M) \simeq CV(h)$ . We suggest fixing the rounding error  $\delta$  and minimizing the function  $\overline{CV}$  in  $M$ . This leads to a very fast WARPing approximation of the bandwidth  $h_{CV}$ .

### 4.3 WARPing the smoothed bootstrap

A WARPed version of the function  $AMISE_w^*$  in 5 is similarly obtained by replacing conventional kernel estimates with their WARPed version, but again some additional work is needed to get a fast algorithm. Our fast algorithm is most clearly motivated through the observation that  $AMISE_w(h)$ , admits the equivalent representation:

$$AMISE_w(h) = \int \left( E\bar{r}_{0,h}(x) - r_0(x) \right)^2 w(x) dx + \frac{1}{nh} \int K^2 p \int \frac{f_1^0(x)w(x)}{(1-F(x))^2} dx$$

where

$$\bar{r}_{0,h}(x) = n^{-1} \sum_{i=1}^n \frac{K_h(x - X_i)\Delta_i}{1 - F(X_i)},$$

i.e. the “known denominator” version of  $\hat{r}_{0,h}$ . This version of the bias admits the WARPed approximation:

$$\begin{aligned} E \left( \bar{r}_{0,h}(x) - r_0(x) \right) &\simeq E \left[ (n\delta M)^{-1} \sum_j 1_{B_j}(x) \sum_{k=1-M}^{M-1} \frac{w_M(k)n_{j+k}}{1 - F_M(j+k)} - r_0(x) \right] = \\ &= (n\delta M)^{-1} \sum_j 1_{B_j}(x) \sum_{k=1-M}^{M-1} w_M(k) \frac{nP(X \in B_{j+k}, \Delta = 1)}{1 - F_M(j+k)} - r_0(x) = \\ &= (n\delta M)^{-1} \sum_j 1_{B_j}(x) \sum_{k=1-M}^{M-1} w_M(k) \frac{np \int_{B_{j+k}} f_1^0(u) du}{1 - F_M(j+k)} - r_0(x) \simeq \\ &\simeq (n\delta M)^{-1} \sum_j 1_{B_j}(x) \sum_{k=1-M}^{M-1} w_M(k) n\delta r_{0,M}(j+k) - r_0(x) = \\ &= M^{-1} \sum_j 1_{B_j}(x) \sum_{k=1-M}^{M-1} w_M(k) r_{0,M}(j+k) - r_0(x). \end{aligned}$$

where  $F_M(z)$  and  $r_{0,M}(z)$  denote the values of  $F$  and  $r_0$  at the lower limit of  $B_z$ .

Our fast method requires some smoothed c.d.f. estimates, so we study WARPed versions of these. First recall that the WARPed version of the kernel density estimator:  $\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K_h(x - X_i)$  is given by

$$\hat{f}(x) = \frac{1}{nh} \sum_j 1_{B_j}(x) \sum_{k=1-M}^{M-1} w_M(k) \tilde{n}_{j+k},$$

which be denoted by  $\hat{f}_M(z)$  when it is evaluated at the lower limit of  $B_z$ . Hence, a natural WARP approximation to the smoothed empirical distribution function:  $\hat{F}_n(x) = K_h * F_n(x) = \int_{-\infty}^x \hat{f}_h(t) dt$ , is:

$$\hat{F}(x) = \int_{-\infty}^x \hat{f}_M(t) dt = \frac{1}{nh} \sum_{k=1-M}^{M-1} w_M(k) \sum_j \int_{-\infty}^x 1_{B_j}(t) dt \tilde{n}_{j+k}$$

and

$$\hat{F}_M(z) = \hat{F}\left(\frac{zh}{M}\right), \quad z \in \mathcal{Z}$$

A fast approximation, which is constant in each bin is, for  $x \in B_l$ ,

$$\hat{F}(x) \simeq \frac{1}{n\delta M} \sum_{k=1-M}^{M-1} w_M(k) \delta \sum_{j=-\infty}^l \tilde{n}_{j+k}.$$

Hence redefine the WARPed version of the smoothed empirical distribution function as:

$$\hat{F}(x) = \frac{1}{nM} \sum_j 1_{B_j}(x) \sum_{k=1-M}^{M-1} w_M(k) \bar{n}_{j+k}, \quad (8)$$

where  $\bar{n}_z$  denotes the number of observations less than the right endpoint of the bin  $B_z$ .

Now for the bootstrap approximation of  $AMISE_w(h)$ , first write

$$\begin{aligned} AMISE_w^*(h) &= \int \left( E^* \bar{r}_{0,h}^*(x) - \frac{\hat{p}_{\hat{f}_{1,g}}^p(x)}{1 - \hat{F}_n(x)} \right)^2 w(x) dx \\ &\quad + \frac{1}{nh} \int K^2 \hat{p} \int \frac{\hat{f}_{1,g}^p \cdot w}{(1 - \hat{F}_n)^2}, \end{aligned}$$

where

$$\bar{r}_{0,h}^*(x) = n^{-1} \sum_{i=1}^n \frac{K_h(x - X_i^*) \Delta_i^*}{1 - \hat{F}_n(X_i^*)}.$$

Then a WARPing approximation as above yields

$$AMISE_w^*(h) \simeq \int \left( \sum_j 1_{B_j}(x) \left( M^{-1} \sum_{k=1-M}^{M-1} w_M(k) \hat{r}_{0,M_1}(j+k) - \hat{r}_{0,M_1}(j) \right) \right)^2 w(x) dx \\ + \frac{1}{nh} \int K^2 \hat{p} \int \frac{\hat{f}_{1,M_1}^0(x) \cdot w(x)}{2} dx,$$

where  $\hat{f}_{1,M_1}^0$  is the WARPing approximation to the nonparametric estimation of the true density  $f_1^0$ , of the uncensored data, with parameter  $M_1 = \frac{q_1}{\delta}$ , and where

$$\hat{F}(x) = \hat{p} \hat{F}_{1,M_1}^0(x) + (1 - \hat{p}) \hat{H}_{1,M_2}^0(x),$$

and

$$\hat{F}_{M_1,M_2}(z) = \hat{F} \left( \frac{zh}{M} \right), \quad z \in Z$$

is the WARP version of the estimator  $\hat{F}$ , in Theorem 3.2, using suitable WARPing approximations  $\hat{F}_{1,M_1}^0$  and  $\hat{H}_{1,M_2}^0$ , constructed by means of expression 8 with uncensored and censored data respectively.

This results in a WARPed version of  $AMISE_w^*(h)$  as

$$\overline{AMISE_w^*}(h) = \delta \sum_{k \in Z} \left( M^{-1} \sum_{k=1-M}^{M-1} w_M(k) \hat{r}_{0,M_1}(z+k) - \hat{r}_{0,M_1}(z) \right)^2 w_{M,h}(z) \\ + \frac{1}{nh} \int K^2 \hat{p} \sum_{z \in Z} \frac{\hat{f}_{1,M_1}^0(z) w_{M,h}(z) \delta}{(1 - \hat{F}_{M_1,M_2}(z))^2}. \quad (9)$$

The above formulas are frequently computationally unstable, because the estimated c.d.f.'s can take on the value one. To avoid this problem, we used slightly modified estimates of the form

$$\tilde{F}_n = \frac{n-1}{n} F_n$$

and

$$\tilde{F}_M = \frac{n-1}{n} F_M,$$

which yield obvious modifications of the above formulas.

## 5 Simulation Study

In this section we compare least squares cross-validation and the bootstrap bandwidth selectors for moderate sample sizes. We considered three underlying distributions:

a) **Weibull model.** The distribution of interest is taken as the Weibull,  $F^0 = F_{\alpha,\lambda}$ , ( $\alpha, \lambda > 0$ ):

$$F_{\alpha,\lambda}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-(\lambda x)^\alpha} & \text{if } x \geq 0 \end{cases}$$

with density  $f^0 = f_{\alpha,\lambda}$ , satisfying  $f^0 = f_{\alpha,\lambda}(x) = \alpha\lambda(\lambda x)^{\alpha-1}e^{-(\lambda x)^\alpha}1_{[0,\infty)}(x)$  and hazard rate function  $r_0(x) = r_{\alpha,\lambda}(x) = \alpha\lambda(\lambda x)^{\alpha-1}1_{[0,\infty)}(x)$ .

We chose a “proportional censorship” model (see Koziol and Green (1976)) which means  $H^0$  is defined by  $1 - H^0 = (1 - F^0)^\eta$  for certain  $\eta > 0$ . This gives a probability of censoring of  $\frac{\eta}{1+\eta}$ . This model allows simple control of the amount of censoring in the simulation through choice of  $\eta$ . The random number generation was carried out using the routine RNUNF of the IMSL library (1991) to draw pseudo-uniform random numbers  $u$ , and then the inversion method was used to compute random deviations following the Weibull distribution:

$$y = \left[ \frac{-\ln(1-u)}{\eta} \right]^{1/\alpha} \frac{1}{\lambda}.$$

b) **Gumbel model.** The underlying distribution is  $F^0 = F_{\alpha,\lambda}$ , of Gumbel type ( $\alpha, \lambda > 0$ ):

$$F_{\alpha,\lambda}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\alpha(e^{\lambda x} - 1)} & \text{if } x \geq 0 \end{cases}$$

with density function  $f^0(x) = f_{\alpha,\lambda}(x) = \alpha\lambda e^{\lambda x - \alpha(e^{\lambda x} - 1)}1_{[0,\infty)}(x)$  and hazard rate  $r_0(x) = r_{\alpha,\lambda}(x) = \alpha\lambda e^{\lambda x}1_{[0,\infty)}(x)$ .

The proportional censorship model is also used and the random number generation, again using the inversion method, is carried out through

$$y = \ln \left[ 1 - \frac{\ln(1-u)}{\eta\alpha} \right] \frac{1}{\lambda}.$$

c) **Truncated normal model.** Here again we used a proportional censorship model, but efficient simulation is now accomplished through separately considering two cases. If the proportion of censoring is greater than or equal to  $1/2$ , then  $F^0$  is taken as the distribution of the random variable  $X^0 = Y|Y \geq 0$ , where  $Y \in N(\mu, \sigma^2)$ . I.e.

$$F^0(x) = \frac{\Phi_{\mu,\sigma}(x) - \Phi_{\mu,\sigma}(0)}{1 - \Phi_{\mu,\sigma}(0)},$$



where  $\Phi_{\mu,\sigma}$  is the distribution of  $Y$ . Observe that the density function of  $X^0$  is given by:

$$f^0(x) = \frac{\phi_{\mu,\sigma}(x)}{1 - \Phi_{\mu,\sigma}(0)} 1_{[0,\infty)}(x),$$

( $\phi_{\mu,\sigma}(x) = \Phi'_{\mu,\sigma}(x)$ ) and hazard rate function:

$$r_0(x) = \frac{\phi_{\mu,\sigma}(x)}{1 - \Phi_{\mu,\sigma}(x)},$$

which coincides with the hazard rate function of a  $N(\mu, \sigma^2)$  distribution (a simple way of seeing this is to recall that the hazard rate is the infinitesimal risk of "becoming dead" at a given "age". This does not change if we maintain the shape but the range).

The distribution  $H^0$  is defined through the relation  $1 - H^0 = (1 - F^0)^\eta$ , with  $\eta = 1, 2, \dots$ , which is nothing else but the minimum of  $\eta$  truncated normal random variables. Note that the censoring proportion can only attain the values  $p = 1/2, 2/3, 3/4, \dots$

A probability of less than 1/2 can be obtained by interchanging the roles of the distributions  $F^0$  and  $H^0$  above. The IMSL routine RNNOF was used to simulate the normal random variables.

Table 1 contains the results of 1000 trials of sample size 100 corresponding to the following models:

- Weibull models with parameters  $\lambda = 1$  and  $\alpha = 1, 2, 3$ , without censoring (denoted by  $W(1, 1)$ ,  $W(2, 1)$ , and  $W(3, 1)$ ) and also with 25% of censoring (denoted by  $CW(1, 1)$ ,  $CW(2, 1)$  and  $CW(3, 1)$ ).
- Gumbel models with parameters  $\lambda = 1$  and  $\alpha = 1, 2, 3$ , without censoring ( $G(1, 1)$ ,  $G(2, 1)$ , and  $G(3, 1)$ ) and with 25% censoring ( $CG(1, 1)$ ,  $CG(2, 1)$  and  $CG(3, 1)$ ).
- A truncated normal distributions with parameters  $\mu = 1$  and  $\sigma = 0.5$  for an uncensored situation and also with a censoring of 25% ( $N(1, 0.5)$ ,  $CN(1, 0.5)$ ).

Although the only sample size in the simulations presented here is  $n = 100$ , similar results were observed in the whole simulation study carried out also for  $n = 50$  and  $n = 200$ .

The values in Table 1 are the mean, the median and the standard deviation of the integrated squared error

$$\int (\hat{r}_{0,\hat{h}} - r_0)^2 w$$

along the 1000 samples of size 100. The bandwidth  $\hat{h}$  denotes the cross-validation bandwidth,  $h_{CV}$ , –obtained by minimizing  $\overline{CV}(M)$  in expression 7– or the bootstrap selector,  $h^*$ , –which comes from the minimization of  $\overline{AMISE}^*(M)$  in the expression 9–. Both minimizers were taken as the global minimizer over a fine grid. In both cases, the rounding error in the WARP approximation was  $\delta = 0.01$  and the weighting function was:

$$w(u) = 1_{[F_0^{-1}(0.25), F_0^{-1}(0.75)]}(u).$$

Finally, the pilot bandwidths  $g_1$  and  $g_2$  used are given by

$$g_1 = \left[ \int K''^2 d_K^{-1} \hat{a}_1^{-1} n^{-1} \right]^{-1/7},$$

$$g_2 = \left[ \int K''^2 d_K^{-1} \hat{a}_2^{-1} n^{-1} \right]^{-1/7},$$

where  $\hat{a}_1$  and  $\hat{a}_2$  are estimators of  $\int f^0 (3)^2$  and  $\int h^0 (3)^2$  and  $d_K = \int t^2 K(t) dt$ . These values are obtained by computing the exact expression of both integrals (in terms of the theoretical variance) when normality is assumed in the underlying densities and then plugging the sample variance of the uncensored or censored data, respectively. This corresponds to the asymptotically optimal bandwidth when estimating the true curvature of the density by means of the curvature of the kernel estimator (see Cao, Cuevas and González-Manteiga (1993) for details). Although the problem here is different, it is clear that it is close related to the pilot bandwidth when bootstrapping in the uncensored case (given by the expression of  $g_1$  above).

TABLE 1: *Mean, Median and standard deviation of  $h_{CV}$  and  $h^*$ .*

Table 1 shows that the bootstrap selector is much more concentrated around its mean than the cross validation bandwidth. The distribution of both are asymmetric to the right (this is more clear for  $h_{CV}$ ) and the bootstrap bandwidth tends to be smaller than the cross validation, on the average. In order to show the accuracy of both bandwidth selectors, Figures 2 and 3 present two kernel density estimators using the 1000 different values of the two selectors  $h_{CV}$  and  $h^*$  for the models  $W(1, 1)$  and  $W(3, 1)$ , as well

as some approximation of the  $h_{MISE_w}$  bandwidths, computed by minimizing the Monte Carlo approximation, based on 1000 trials, of the mean integrated squared error

$$E \left( \int (\hat{r}_{0,h} - r_0)^2 w \right),$$

in finitely many values of  $h$ , ranging from 0.1 to 1.1 with a step of 0.05. Every integrated squared error is approximated by its WARP version:

$$\delta \sum_{z \in \mathcal{Z}} \hat{r}_{0,M}^2(z) w_{M,h}(z) - 2 \sum_{z \in \mathcal{Z}} \hat{r}_{0,M}(z) r_0(z) w_{M,h}(z) + \sum_{z \in \mathcal{Z}} r_0(z) w_{M,h}(z)$$

and  $\delta = 0.01$ .

We conclude from both the numerical and graphical viewpoints that the performance of  $h^*$  is far superior to that of  $h_{CV}$ . We believe that this can be understood theoretically by studying asymptotic rates of convergence to  $h_{MISE_w}$ . Patil (1990) indicates that  $h_{CV}$  has the excruciatingly slow rate of convergence:

$$\frac{h_{CV}}{h_{MISE_w}} - 1 \simeq n^{-1/10}.$$

Based on similar rates of convergence obtained for conventional density estimation by Jones, Marron and Park (1991), we conjecture that the bootstrap bandwidth presented here has rate of convergence

$$\frac{h^*}{h_{MISE_w}} - 1 \simeq n^{-5/14},$$

and that the exponent could be improved to  $-1/2$  by the methods presented in that paper. These are some interesting open problems for future research.

## 6 Appendix: Proofs

### Proof of Theorem 3.2

First check that both resamplings are equivalent. Consider a bootstrap random observation  $(X^*, \Delta^*)$  chosen uniformly from the set  $\{(X_i, \Delta_i)\}_{i=1}^n$  and independently draw a random variable  $R$  with density  $K$ . The CS1 resampling yields the bootstrap observation  $(X^*, \Delta^*)$ , where

$$X^* = X^{t*} + [g_1 \Delta^* + g_2 (1 - \Delta^*)] R.$$

The CS2 resampling plan involves independent draws  $\overline{X}^{0*}$  from  $F_n^0$  and  $\overline{Z}^*$  from  $H_n^0$ . These are used to construct:

$$\Delta^* = 1_{\{\overline{X}^{0*} \leq \overline{Z}^*\}}$$

and

$$X'^* = \min\{X^{0*}, Z^*\}.$$

Independently generating  $R$ , with density  $K$ , yields the ME resampling,  $(X^*, \Delta^*)$ , where

$$X^* = X'^* + [g_1 \Delta^* + g_2(1 - \Delta^*)]R.$$

To compare these resampling algorithms, observe that the joint bootstrap distributions of  $R$  and  $(X'^*, \Delta^*)$  are the same under both CS1 and CS2. Hence, only the equivalence of the bootstrap distributions of  $(X'^*, \Delta^*)$  is needed to ensure that CS1 and CS2 coincide. This is a direct consequence of Efron (1981), since the absolute continuity of the distributions imply that there will be no ties, with probability one.

Many of the details needed for a rigorous derivation of the representation of  $AMISE_w$  are quite standard, so we present only an outline. The representation given by the CS1 resampling plan is most convenient. I.e. draw  $I_i$  with equiprobability from the set  $\{1, 2, \dots, n\}$ ,

$$\Delta_i^* = \Delta_{I_i},$$

$$X_i^* = X_{I_i} + \Delta_i^* g_1 R_i + (1 - \Delta_i^*) g_2 R_i$$

and  $R_i$ , with density  $K$ , independent of  $I_i$ . For every bootstrap resample  $\{(X_i^*, \Delta_i^*)\}_{i=1}^n$ , the bootstrap version of the hazard rate kernel estimator is:

$$\hat{r}_{0,h}^*(x) = n^{-1} \sum_{i=1}^n K_h(x - X_i^*) \frac{\Delta_i^*}{1 - F_n^*(X_i^*)},$$

where  $F_n^*$  is the empirical distribution function of the bootstrap resample  $(X_1^*, X_2^*, \dots, X_n^*)$ . Observe that the bootstrap observations follow the bootstrap distribution:

$$\hat{F}_n^*(x) = P_*(X^* \leq x) = \hat{p}(K_{g_1} * F_{1,n}^0)(x) + (1 - \hat{p})(K_{g_2} * H_{1,n}^0)(x).$$

So

$$E^*(\hat{r}_{0,h}^*(x)) =$$

$$\begin{aligned}
& E^* \left[ E^* \left( n^{-1} \sum_{i=1}^n K_h(x - X_i^*) \frac{\Delta_i^*}{1 - \widehat{F}_n^*(X_i^*)} \mid R_1, R_2, \dots, R_n \right) \right] \simeq \\
& \simeq E^* \left( n^{-1} \sum_{i=1}^n n^{-1} \sum_{j=1}^n K_h(x - (X_j + g_1 R_j)) \frac{\Delta_j}{1 - \widehat{F}_n(X_j + g_1 R_j)} \right) = \\
& = n^{-1} \sum_{i=1}^n n^{-1} \sum_{j=1}^n \int K_h(x - (X_j + g_1 u)) \frac{\Delta_j}{1 - \widehat{F}_n(X_j + g_1 u)} K(u) du = \\
& = n^{-1} \sum_{j=1}^n \int K_h(x - X_j - g_1 u) \frac{K(u)}{1 - \widehat{F}_n(X_j + g_1 u)} du \Delta_j \\
& = n^{-1} \sum_{j=1}^n \int K_h(v) K_{g_1}(x - v - X_j) \frac{1}{1 - \widehat{F}_n(x - v)} dv \Delta_j = \\
& = \hat{p} \int K_h(v) \frac{\hat{f}_{1, g_1}^0(x - v)}{1 - \widehat{F}_n(x - v)} dv = \hat{p} (K_h * \frac{\hat{f}_{1, g_1}^0}{1 - \widehat{F}_n})(x).
\end{aligned}$$

The asymptotic bootstrap bias follows easily from this.

The bootstrap variance of  $\hat{r}_{0,h}^*$ , is given by

$$\begin{aligned}
\text{Var}^*(\hat{r}_{0,h}^*(x)) & \simeq \text{Var}^* \left( n^{-1} \sum_{i=1}^n K_h(x - X_i^*) \frac{\Delta_i^*}{1 - \widehat{F}_n^*(X_i^*)} \right) = \\
& = \frac{1}{nh^2} \left[ E^* \left( K^2 \left( \frac{x - X^*}{h} \right) \frac{\Delta^*}{(1 - \widehat{F}_n^*(X_i^*))^2} \right) \right. \\
& \quad \left. - E^* \left( K \left( \frac{x - X^*}{h} \right) \frac{\Delta^*}{1 - \widehat{F}_n^*(X_i^*)} \right)^2 \right],
\end{aligned}$$

whose dominant term can be analyzed by standard methods:

$$\begin{aligned}
& \frac{1}{nh^2} E^* \left( K^2 \left( \frac{x - X^*}{h} \right) \frac{\Delta^*}{(1 - \widehat{F}_n^*(X_i^*))^2} \right) = \\
& = \frac{1}{nh^2} E^* \left( n^{-1} \sum_{j=1}^n K^2 \left( \frac{x - X_j - g_1 R}{h} \right) \frac{\Delta_j}{(1 - \widehat{F}_n(X_j + g_1 R))^2} \right) = \\
& = \frac{1}{nh^2} n^{-1} \sum_{j=1}^n \int K^2 \left( \frac{x - X_j - g_1 u}{h} \right) \frac{K(u)}{(1 - \widehat{F}_n(X_j + g_1 u))^2} du \Delta_j =
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^2 h} \sum_{j=1}^n \int \frac{K^2(z)}{(1 - \widehat{F}_n(x - hz))^2} K\left(\frac{x - hz - X_j}{g_1}\right) \frac{1}{g_1} dz \Delta_j = \\
&= \frac{1}{nh} \widehat{p} \int K^2(z) \frac{\widehat{f}_{1,g_1}^0(x - hz)}{(1 - \widehat{F}_n(x - hz))^2} dz \simeq \\
&\simeq \frac{1}{nh} \widehat{p} \int K^2 \cdot \frac{\widehat{f}_{1,g_1}^0(x)}{(1 - \widehat{F}_n(x))^2}.
\end{aligned}$$

The proof finishes using the representation of the mean integrated squared error in terms of the integrated variance and the integrated squared bias.

## 7 References

- BOWMAN, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71** 353-360.
- BLUM, J. R. AND SUSARLA, V. (1980). Maximal deviation theory of density and failure rate function estimates based on censored data. *Journal of Multivariate Analysis*, **5** 213-222.
- CAO, R., CUEVAS, A. AND GONZÁLEZ MANTEIGA, W. (1993). A comparative study of several smoothing methods in density estimation. To appear in *Computational Statistics and Data Analysis*.
- CHENG, P. E. (1987). A nearest neighbor hazard rate estimator for randomly censored data. *Communications in Statistics, Theory and Methods*, **16** 613-625.
- EFRON, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, **76** 312-321.
- FAN, J. AND MARRON, J. S. (1994) Fast implementations of nonparametric curve estimators. To appear in *Journal of Computational and Graphical Statistics*.
- FÖLDES, A., RETJÖ, L. AND WINTER, B.B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data. *Period. Math. Hung.*, **12** 15-29.
- GRENANDER, U. (1956). On the theory of mortality measurements, Part III. *Skand. Akt*, **39** 125-153.
- HÄRDLE, W. (1991). *Smoothing Techniques with implementation in S*. Springer-Verlag.

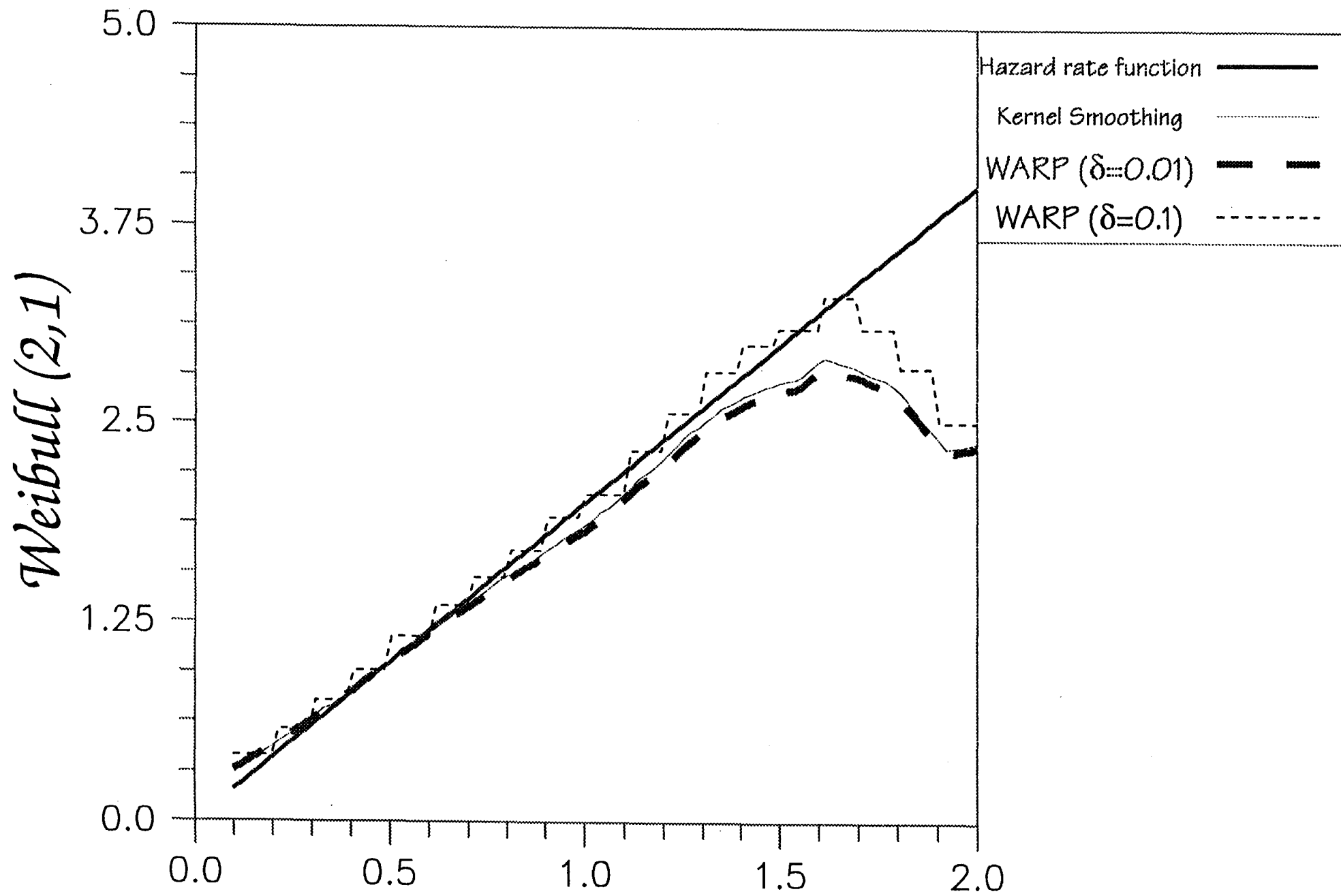
- HÄRDLE, W., HALL, P. AND MARRON, J. S. (1988) How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association*, **83** 86-101.
- HALL, P. (1987) On Kullback Leibler loss and density estimation. *Annals of Statistics*, **15** 1491-1519.
- IMSL, version 2.0 (1991), IMSL Inc.
- Jones, M. C., Marron, J. S. and Park, B. U. (1991) A simple root  $n$  bandwidth selector. *Annals of Statistics*, **19** 1919-1932.
- KAPLAN, E. L. AND MEIER, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53** 457-481.
- KOZIOL, J. AND GREEN, S. (1976). A Cramer-von Mises statistic for randomly censored data. *Biometrika*, **63** 465-474.
- LO, S. H., MACK, Y. P. AND WANG, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probability Theory and Related Fields*, **80** 461-473.
- MARRON, J. S. (1992). Bootstrap bandwidth selection. In *Exploring the limits of bootstrap*, Le Page, P. and Billard, L. Editors, 249-262.
- MARRON, J. S. AND PADGETT, J. W. (1987). Asymptotically optimal bandwidth selection for a kernel density estimator. *Annals of Statistics*, **15** 1520-1535.
- MÜLLER, H. G. AND WANG, J. L. (1990). Locally adapted hazard smoothing. *Probability Theory and Related Fields*, **85** 523-538.
- MURTHY, V. K. (1965). Estimation of jumps, reliability and hazard rate. *The Annals of Mathematical Statistics*, **36** 1032-1040.
- PARZEN, E. (1962). On estimation of a probability density and mode. *The Annals of Mathematical Statistics*, **35** 1065-1076.
- PATIL, P.N. (1990). Automatic smoothing parameter selection in hazard rate estimation. PhD Dissertation, University of North Carolina Institute of Statistics, Mimeo Series #2033.
- PATIL, P. N. (1993). Bandwidth choice for nonparametric hazard rate estimation density and mode. *Journal of Statistical Planning and Inference*, **35** 15-30.
- PATIL, P. N., WELLS, M. T. AND MARRON, J. S. (1994). Kernel based estimators of ratio functions, to appear in *Journal of Nonparametric Statistics*.
- RICE, J. AND ROSENBLATT, M. (1976). Estimation of the log survivor function and hazard function. *Sankhyā Series A*, **38** 60-78.

- ROSENBLATT, M. (1956). Remarks on some nonparametric estimators of a density function. *The Annals of Mathematical Statistics*, **27** 832-837.
- RUDEMO, M. (1982). Empirical choices of histograms and kernel density estimates. *Scandinavian Journal of Statistics*, **9** 65-78.
- SARDA, P. AND VIEU, P. (1991). Smoothing parameter selection in hazard estimation. *Statistics and Probability Letters*, **11** 429-434.
- SETHURAMAN, J. AND SINGPURWALLA, N. (1981). Large sample estimates and uniform confidence bounds for the failure rate function based on a naive estimate. *The Annals of Statistics*, **9** 628-632.
- TANNER, M. A. AND WONG, W. H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Annals of Statistics*, **1** 989-993.
- TANNER, M. A. AND WONG, W. H. (1984). Data-based nonparametric estimation of the hazard function with applications to model diagnostics and exploratory analysis. *Journal of the American Statistical Association*, **79** 174-182.
- YANDELL, B. S. (1981). Nonparametric inference for rates and densities with censored serial data. Ph D Dissertation, University of California.
- YANDELL, B. S. (1983). Nonparametric inference for rates and densities with censored serial data. *The Annals of Statistics*, **11** 1119-1135.
- WATSON, G. S. AND LEADBETTER, M. R. (1964A). Hazard analysis I. *Biometrika*, **51** 175-184.
- WATSON, G. S. AND LEADBETTER, M. R. (1964B). Hazard analysis II. *Sankhyā Series A*, **26** 101-116.

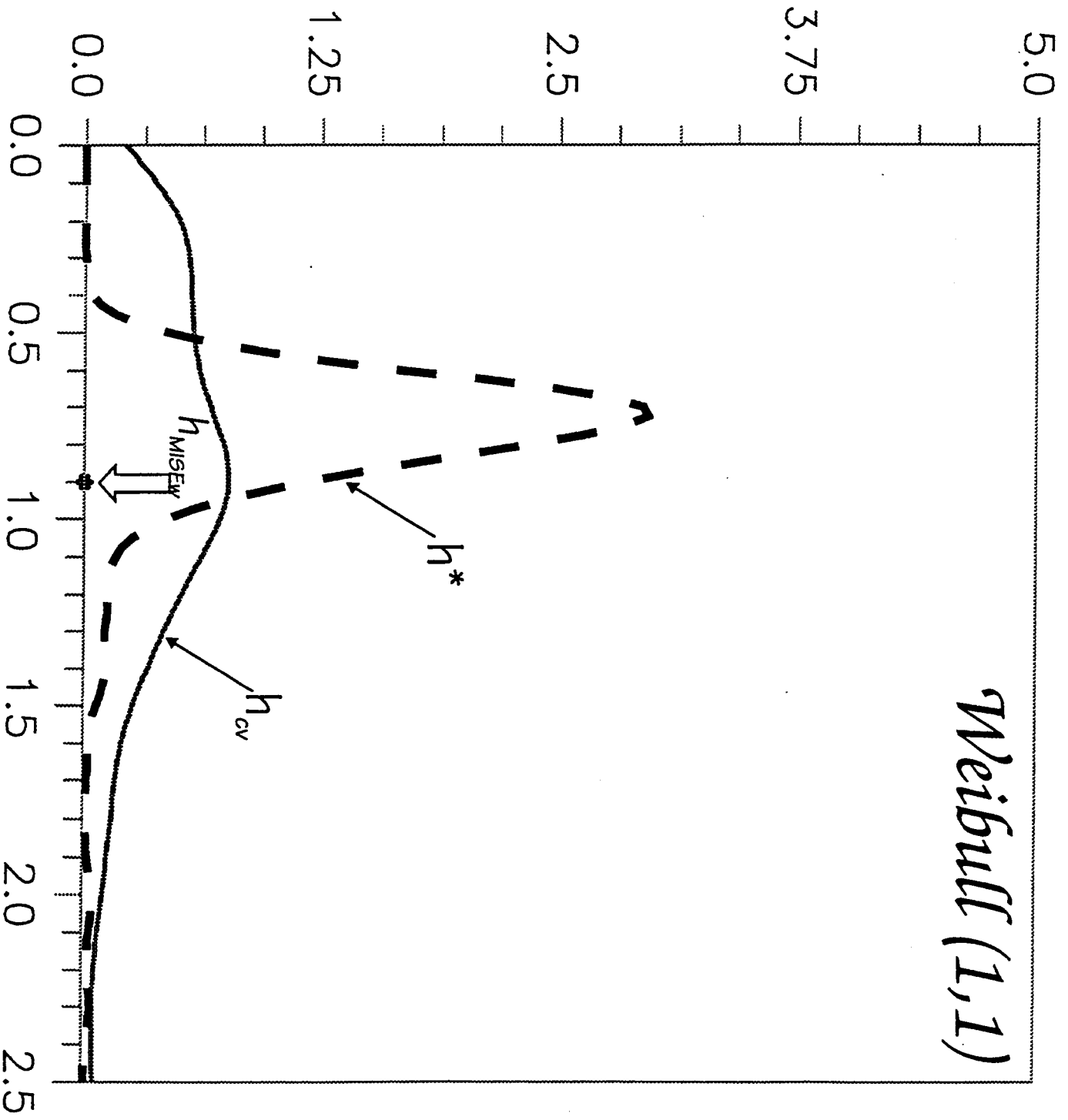


Table 1

Model	mean		median		std. dev.	
	$h_{CV}$	$h^*$	$h_{CV}$	$h^*$	$h_{CV}$	$h^*$
$W(1,1)$	0.067	0.031	0.033	0.020	0.112	0.034
$CW(1,1)$	0.118	0.083	0.095	0.068	0.104	0.062
$W(2,1)$	0.096	0.047	0.030	0.024	0.199	0.070
$CW(2,1)$	0.173	0.128	0.119	0.099	0.194	0.108
$W(3,1)$	0.145	0.083	0.091	0.054	0.197	0.096
$CW(3,1)$	0.246	0.188	0.174	0.150	0.252	0.112
$G(1,1)$	0.101	0.054	0.032	0.027	0.198	0.088
$CG(1,1)$	0.182	0.134	0.115	0.104	0.249	0.112
$G(2,1)$	0.169	0.081	0.058	0.047	0.375	0.101
$CG(2,1)$	0.303	0.223	0.214	0.181	0.365	0.172
$G(3,1)$	0.217	0.122	0.096	0.076	0.393	0.144
$CG(3,1)$	0.375	0.299	0.287	0.246	0.349	0.223
$N(1,0.5)$	0.184	0.088	0.090	0.057	0.316	0.104
$CN(1,0.5)$	0.236	0.184	0.164	0.144	0.274	0.150



# Weibull (1,1)



# Weibull (3,1)

