View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Universidade do Minho: RepositoriUM

Nonparametric Location-Scale Models for Censored Successive Survival Times

Ingrid Van Keilegom[*] Jacobo de Uña-Álvarez[†]
Luis Meira-Machado[‡]

September 9, 2010

Abstract

Let (T_1, T_2) be gap times corresponding to two consecutive events, which are observed subject to (univariate) random right-censoring. The censoring variable corresponding to the second gap time T_2 will in general depend on this gap time. Suppose the vector (T_1, T_2) satisfies the nonparametric location-scale regression model $T_2 = m(T_1) + \sigma(T_1)\varepsilon$, where the functions m and σ are 'smooth', and ε is independent of T_1. The aim of this paper is twofold. First, we propose a nonparametric estimator of the distribution of the error variable under this model. This problem differs from others considered in the recent related literature in that the censoring acts not only on the response but also on the covariate, having no obvious solution. On the basis of the idea of transfer of tail information (Van Keilegom and Akritas, 1999), we then use the proposed estimator of the error distribution to introduce nonparametric estimators for important targets such as: (a) the conditional distribution of T_2 given T_1; (b) the bivariate distribution of the gap times; and (c) the so-called transition probabilities. The asymptotic properties of these estimators are obtained. We also illustrate through simulations, that the new estimators based on the location-scale model may behave much better than existing ones.

Key words: Bivariate distribution; Conditional distribution; Error distribution; Progressive three-state model; Recurrent events; Transfer of tail information; Transition probabilities.

[*]Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, E-mail: ingrid.vankeilegom@uclouvain.be

[†]Departamento de Estadística e Investigación Operativa, Facultad de CC. Económicas y Empresariales, University of Vigo, Campus Universitario Lagoas-Marcosende, 36310 Vigo, Spain, E-mail: jacobo@uvigo.es

[‡]Department of Mathematics and Applications, University of Minho, Guimarães, Portugal, E-mail: lmachado@math.uminho.pt

1 Introduction

Consider two random variables T_1 and T_2, that represent the duration times between successive events. This is often encountered in practice. Consider e.g. an AIDS study, where T_1 represents the time between HIV infection and AIDS, and T_2 represents the time between development of AIDS and death of the patient. This model, also called a progressive three state-model in the context of multi-state models, is also encountered in studies where T_1 is the time between diagnosis of a disease and transplantation, and T_2 is the survival time after transplantation, or in studies where T_1 is the time until recurrence of a disease, and T_2 is the time from recurrence to death. Other examples include studies where one reports the time patients stay in a certain stage of a disease, and studies where recurrent events are observed (like infections, asthma attacks, ...).

The fact that the variables T_1 and T_2 are recorded successively, rather than simultaneously, is important when the variables are subject to censoring. We consider here random right censoring. Let C be the (univariate) censoring variable, supposed to be independent of (T_1, T_2). When T_1 and T_2 would be recorded simultaneously, we would observe $T_1 \wedge C$, $T_2 \wedge C$ and the corresponding censoring indicators. However, in the present context of successive events, we only observe the second gap time if the first failure time is uncensored. More precisely, the observable variables are given by $(\widetilde{T}_1, \widetilde{T}_2, \Delta_1, \Delta_2)$, where $\widetilde{T}_1 = T_1 \wedge C$, $\Delta_1 = I(T_1 \leq C)$, $\widetilde{T}_2 = T_2 \wedge C_2$ and $\Delta_2 = I(T_2 \leq C_2)$, where $C_2 = (C - T_1)I(T_1 \leq C)$ is the censoring variable for the second gap time. Note that $\Delta_2 = 1$ implies $\Delta_1 = 1$. Hence, $\Delta_2 = \Delta_1\Delta_2 = I(Y \leq C)$ is the censoring indicator pertaining to the total time $Y = T_1 + T_2$. Let $(\widetilde{T}_{1i}, \widetilde{T}_{2i}, \Delta_{1i}, \Delta_{2i})$, $1 \leq i \leq n$, be i.i.d. data with the same distribution as $(\widetilde{T}_1, \widetilde{T}_2, \Delta_1, \Delta_2)$.

Due to the independence assumption between C and (T_1, T_2), the marginal distribution of the first gap time T_1 can be consistently estimated by the Kaplan-Meier estimator based on the $(\widetilde{T}_{1i}, \Delta_{1i})$'s. Similarly, the distribution of the total time may be consistently estimated by the Kaplan-Meier estimator based on the $(\widetilde{T}_{1i} + \widetilde{T}_{2i}, \Delta_{2i})$'s, since $T_1 + T_2$ is independent of C. However, since T_2 and C_2 will in general be dependent, the estimation of the marginal distribution of the second gap time is not such a simple issue. For the same reason, it is not clear in principle how the conditional distribution $F(y \mid x) = P(T_2 \leq y \mid T_1 = x)$ can be efficiently estimated. This situation has been discussed in several papers, see for example Wang and Wells (1998), Lin *et al.* (1999), Schaubel and Cai (2004) or Van Keilegom (2004).

In this paper we will propose and study estimators of the conditional distribution

$F(y \mid x) = P(T_2 \leq y \mid T_1 = x)$ of T_2 given $T_1 = x$, the bivariate distribution $F(x, y) = P(T_1 \leq x, T_2 \leq y)$ of T_1 and T_2, and the so-called transition probabilities (see (2.2), (2.3) and (2.4) below for their definitions). We will estimate these quantities assuming that the relationship between the gap times T_1 and T_2 is given by the following model :

$$T_2 = m(T_1) + \sigma(T_1)\varepsilon, \tag{1.1}$$

where the error ε is independent of T_1, and where $\int_0^1 F_\varepsilon^{-1}(s)J(s)ds = 0$ and $\int_0^1 F_\varepsilon^{-1}(s)^2 J(s)ds = 1$ (F_ε being the distribution of the error ε). Here, J is a score function satisfying $\int_0^1 J(s)ds = 1$. This implies that m and σ are L-functionals, given by

$$m(x) = \int_0^1 F^{-1}(s \mid x)J(s)ds, \tag{1.2}$$

$$\sigma^2(x) = \int_0^1 F^{-1}(s \mid x)^2 J(s)ds - m^2(x), \tag{1.3}$$

where $F^{-1}(s \mid x) = \inf\{t : F(t \mid x) \geq s\}$ is the quantile function of T_2 given $T_1 = x$. Note that when $J(s) = I(0 \leq s \leq 1)$, then $m(x) = E(Y|x)$ and $\sigma^2(x) = \text{Var}(Y|x)$. Expressions (1.2) and (1.3) are motivated by the fact that under right censoring, it is in general impossible to consistently estimate the conditional mean and variance in a completely nonparametric way, whereas the above choices of m and σ can be estimated consistently, provided the score function J is chosen appropriately. Model (1.1) has been extensively studied in Van Keilegom and Akritas (1999) when T_1 is completely observed and T_2 is subject to random right censoring.

Note that one often transforms the variable T_2 (which is usually positive) in such a way that the transformed variable ranges from $-\infty$ to $+\infty$ (use e.g. a logarithmic transformation). We will continue denoting the response by T_2, but should keep in mind that this might represent a transformation of T_2.

To explain the motivation for model (1.1), consider for simplicity the case where T_1 and T_2 are positively correlated. Then, the higher the value of T_1, the higher the probability that T_1 is censored (since T_1 and C are independent), and also the higher the probability that T_2 is censored (note that whenever T_1 is censored, T_2 is also censored). This means that for large values of T_1, the conditional distribution $F(y|x)$ of T_2 given $T_1 = x$ will be hard to estimate in a completely nonparametric way, especially in the right tail, whereas this will not be the case for small values of T_1. Under model (1.1) these conditional distributions are given by

$$F(y \mid x) = F_e\left(\frac{y - m(x)}{\sigma(x)}\right), \tag{1.4}$$

where $F_e(y) = P(\varepsilon \leq y)$, so an estimator for $F(y \mid x)$ can be obtained by plugging in proper estimators of the error distribution and of the location and scale functions. Since the error distribution can be estimated globally (i.e. by using all data points, not only those in a neighborhood of x), we will be able to estimate well the conditional distribution even for large values of T_1.

Before concluding this section, note that the validity of the location-scale model (1.1) needs to be verified for a particular data set. This problem has been considered for the case where all data are completely observed in Einmahl and Van Keilegom (2008a, 2008b) and Neumeyer (2009). The extension of the tests proposed in these papers to the current situation of censored data, can be considered but will not at all be straightforward.

The paper is organized as follows. In the next section, we introduce some notations and give the precise definitions of the estimators of $F_e(y)$, $F(y|x)$, $F(x,y)$ and of the transition probabilities. Section 3 deals with the asymptotic properties of these estimators, while their finite sample performance is investigated in Section 4. Finally, the proofs of the asymptotic results are collected in the Appendix.

2 The estimators

We will assume throughout that the support of T_1 lies in $(0, \infty)$, whereas T_2 is defined on (a subset of) $(-\infty, +\infty)$. This is because T_2 is allowed to represent a transformation of the second gap time (see Section 1 for more details).

Since C and (T_1, T_2) are independent, and since the error ε is independent of the 'covariate' T_1, we have that ε and (T_1, C) are independent. As a result, the error distribution based on those errors for which T_1 is uncensored (i.e. $T_1 \leq C$) coincides with the true error distribution. Moreover,

$$F_e(y) = P(\varepsilon \leq y) = P(\varepsilon \leq y \mid \Delta_1 = 1) = P(\varepsilon \leq y \mid \Delta_1 = 1, T_1 \leq \tau_1),$$

for any constant τ_1, since ε and T_1 are independent. This relation will be used to estimate $F_e(y)$. To do so, we first need to estimate $m(\cdot)$ and $\sigma(\cdot)$, for which we follow the approach used in Van Keilegom and Akritas (1999), but with the Beran (1981)-estimator replaced by a proper conditional distribution which copes with the censoring on the covariate. More precisely, we define

$$\widehat{m}(x) = \int_0^1 \widetilde{F}^{-1}(s \mid x) J(s) ds, \quad \widehat{\sigma}^2(x) = \int_0^1 \widetilde{F}^{-1}(s \mid x)^2 J(s) ds - \widehat{m}^2(x),$$

where

$$\widetilde{F}(y \mid x) = 1 - \prod_{T_{2i} \le y, \Delta_{2i}=1} \left[1 - \frac{B_{ni}(x; a_n)}{\sum_{j=1}^{n} B_{nj}(x; a_n) I(T_{2j} \ge T_{2i})} \right]$$

for $y \le T_{2(n)}$ (the largest order statistic of the T_{2i}'s), and $\widetilde{F}(y \mid x) = 1$ for $y > T_{2(n)}$. Here

$$B_{ni}(x; a_n) = \frac{\Delta_{1i} K\left((x - T_{1i})/a_n\right)}{\sum_{j=1}^{n} \Delta_{1j} K\left((x - T_{1j})/a_n\right)},$$

a_n stands for a sequence of bandwidths, K is a probability density function (kernel), and by convention $0/0 = 0$. Note that $B_{ni}(x; a_n) = 0$ whenever $\Delta_{1i} = 0$. This is justified by the fact that, under independence of C and (T_1, T_2), we have that $F(y \mid x) = P(T_2 \le y \mid T_1 = x, \Delta_1 = 1)$, and so one can restrict to the uncensored 'covariates'. See also Van Keilegom (2004).

Now, define

$$\widehat{E}_i = \frac{\widetilde{T}_{2i} - \widehat{m}(\widetilde{T}_{1i})}{\widehat{\sigma}(\widetilde{T}_{1i})}$$

$(i = 1, \ldots, n)$ and let $\widehat{F}_e$ be the Kaplan-Meier estimator of F_e (which we suppose to be continuous) based on the $(\widehat{E}_i, \Delta_{2i})$'s for which $\Delta_{1i} = 1$ and $\tilde{T}_{1i} \le \tau_1$, i.e.

$$\widehat{F}_e(y) = 1 - \prod_{\widehat{E}_{(i)} \le y, \Delta_{2(i)}=1} \left(1 - \frac{1}{N_{u\tau} - i + 1} \right) \qquad (2.1)$$

for $y \le \widehat{E}_{(N_{u\tau})}$, and $\widehat{F}_e(y) = 1$ for $y > \widehat{E}_{(N_{u\tau})}$. Here, $\widehat{E}_{(i)}$ $(i = 1, \ldots, N_{u\tau})$ is the i-th order statistic of the $\widehat{E}_j$, $j = 1, \ldots, n$, for which $\Delta_{1j} = 1$ and $\tilde{T}_{1j} \le \tau_1$, $\Delta_{2(i)}$ is the corresponding censoring indicator, and $N_{u\tau} = \sum_{i=1}^{n} I(\Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1)$. The restriction to $\widetilde{T}_{1i}$'s that are smaller than τ_1 is purely of technical nature (it has to do with the uniform consistency of $\widehat{m}(x)$ and $\widehat{\sigma}(x)$, which is only established on a compact interval), and can be dropped when $\widetilde{T}_1$ has bounded support. We will see below that τ_1 can be chosen arbitrarily close to the right endpoint of the support of the variable $\widetilde{T}_1$, and is therefore of no importance in practical calculations.

When $P(\Delta_1 = 1) = 1$ and $\widetilde{T}_1$ has bounded support, asymptotic properties of $\widehat{F}_e$ were investigated in Van Keilegom and Akritas (1999). By following similar arguments, properties for the current, more general case, will be derived. However, we will have to cope with the risk of censoring on the 'covariate' T_1. As is clear from (2.1), the censored $\widetilde{T}_{1i}$'s are not used to estimate $F_e(y)$. However, they will play a role when computing the empirical version of the bivariate distribution and the transition probabilities.

The estimator $\widehat{F}_e(y)$ defined in (2.1), together with relation (1.4), is the key for the construction of an estimator of $F(y \mid x)$:

$$\widehat{F}(y \mid x) = \widehat{F}_e\left(\frac{y - \widehat{m}(x)}{\widehat{\sigma}(x)}\right).$$

This estimator will be more efficient than $\widetilde{F}(y \mid x)$, since it allows for the transfer of tail information from lightly censored areas to heavily censored ones. See Van Keilegom *et al.* (2001) for illustrative simulation results in this regard, in the case where the covariate is uncensored.

Next, we estimate the bivariate distribution $F(x, y)$. First, note that

$$F(x,y) = \int_0^x F\,(y \mid u)\, F_1(du) = \int_0^x F_e\left(\frac{y - m(u)}{\sigma(u)}\right) F_1(du),$$

where $F_1(x) = P(T_1 \leq x)$ is the marginal distribution of T_1. We estimate $F_1(x)$ by the Kaplan-Meier estimator based on the $(\widetilde{T}_{1i}, \Delta_{1i})$'s :

$$\widehat{F}_1(x) = 1 - \prod_{\widetilde{T}_{1(i)} \leq x, \Delta_{1(i)}=1} \left(1 - \frac{1}{n - i + 1}\right)$$

for $x \leq \widetilde{T}_{1(n)}$, and $\widehat{F}_1(x) = 1$ for $x > \widetilde{T}_{1(n)}$. Here, $\widehat{T}_{1(i)}$ $(i = 1, \dots, n)$ is the i-th order statistic of the $\widehat{T}_{1j}$, $j = 1, \dots, n$ and $\Delta_{1(i)}$ is the corresponding censoring indicator, and define

$$\widehat{F}(x,y) = \int_0^x \widehat{F}\,(y \mid u)\, \widehat{F}_1(du) = \int_0^x \widehat{F}_e\left(\frac{y - \widehat{m}(u)}{\widehat{\sigma}(u)}\right) \widehat{F}_1(du).$$

Note that the estimation of the bivariate distribution from censored gap times has been considered in several papers, including Wang and Wells (1998) and Lin *et al.* (1999). Recently, de Uña-Álvarez and Meira-Machado (2008) introduced a simple method which (unlike previous proposals) leads to a proper distribution function, avoiding the negative weighting of data points. In Section 4 we carry out some simulations which will indicate that the estimator $\widehat{F}(x, y)$ based on the location-scale model is more efficient.

Consider now the problem of estimating a transition probability $p_{ij}(s, t)$. In general, $p_{ij}(s, t)$ stands for the probability of being in state j at time t, conditionally on being in state i at time s. The transition probability is defined for a stochastic process that at any point in time may occupy one state among a discrete set of states. This type of process is modelled through the so-called multi-state models (see for example Hougaard, 2000), and then the survival prognosis is performed via the estimation of these transition probabilities or related curves (such as the transition intensities). Recurrent events data (or gap times) may be seen as arising from a three-state model, in which state 1 represents 'no event', and states 2 and 3 represent the occurrence of the first and the second events, respectively. Possible transitions are from state 1 to state 2, and from state 2 to state 3. In this model, there are essentially three transition probabilities to estimate (with $s < t$) (note that $p_{23}(s,t) = 1 - p_{22}(s,t)$, $p_{13}(s,t) = 1 - p_{11}(s,t) - p_{12}(s,t)$ and $p_{33}(s,t) = 1$) :

$$p_{11}(s,t) = P(T_1 > t \mid T_1 > s) = \frac{1 - F_1(t)}{1 - F_1(s)}, \tag{2.2}$$

$$\begin{aligned} p_{12}(s,t) &= P\left(T_1 \le t, T_1 + T_2 > t \mid T_1 > s\right) \\ &= \frac{1}{1 - F_1(s)} \int_s^t \left[1 - F(t - u \mid u)\right] F_1(du), \end{aligned} \tag{2.3}$$

$$\begin{aligned} p_{22}(s,t) &= P(T_1 \le t, T_1 + T_2 > t \mid T_1 \le s, T_1 + T_2 > s) \\ &= \int_0^s \left[1 - F(t - u \mid u)\right] F_1(du) \Big/ \int_0^s \left[1 - F(s - u \mid u)\right] F_1(du)\ . \end{aligned} \tag{2.4}$$

Replace F_1 by $\widehat{F}_1$ and $F(y \mid x)$ by $\widehat{F}(y \mid x)$ to get the following estimators:

$$\widehat{p}_{11}(s,t) = \frac{1 - \widehat{F}_1(t)}{1 - \widehat{F}_1(s)},$$

$$\widehat{p}_{12}(s,t) = \frac{1}{1 - \widehat{F}_1(s)} \int_s^t \left[1 - \widehat{F}(t - u \mid u)\right] \widehat{F}_1(du),$$

$$\widehat{p}_{22}(s,t) = \int_0^s \left[1 - \widehat{F}(t - u \mid u)\right] \widehat{F}_1(du) \Big/ \int_0^s \left[1 - \widehat{F}(s - u \mid u)\right] \widehat{F}_1(du)\ .$$

In Section 3 the asymptotic properties of a slightly modified version of these estimators are derived. Interestingly, these properties do not rely on the Markov assumption, typically used in multi-state models, nor on any other simplifying assumption. Since $\widehat{p}_{11}(s,t)$ is a simple function of ordinary Kaplan-Meier estimators, it will not be considered anymore in this manuscript. Simulations reported in Section 4 compare $\widehat{p}_{22}(s,t)$ to the non-Markovian estimator introduced in Meira-Machado *et al.* (2006), which does not make use of the information contained in the location-scale model. These simulations suggest that the transfer of tail information may improve dramatically the estimation of the transition probabilities.

3 Main results

In this section we state the asymptotic expansion and the weak convergence of the estimators of $F_e(y)$, $F(y|x)$, $F(x,y)$, $p_{12}(s,t)$ and $p_{22}(s,t)$. The asymptotic expansion is useful, since it decomposes the estimator in a sum of i.i.d. terms and a remainder term of smaller order. Based on this decomposition, the weak convergence of the estimator can then be established. This weak convergence result immediately leads to the construction of uniform confidence bands and test statistics for hypotheses concerning the functions of interest.

The notations used in the asymptotic results below are given in Appendix A, whereas the assumptions under which these results are valid are stated in Appendix B.

Theorem 3.1 *Assume (A1)-(A4).*

(i) Then,

$$\widehat{F}_e(y) - F_e(y) = (np_{1\tau})^{-1} \sum_{i=1}^{n} I(\Delta_{1i} = 1, \widetilde{T}_{1i} \le \tau_1)\varphi(\tilde{T}_{1i}, \tilde{T}_{2i}, \Delta_{2i}, y) + R_n(y),$$

where $\sup\{|R_n(y)|; -\infty < y \le \tau_e\} = o_P(n^{-1/2})$.

(ii) The process $n^{1/2}(\widehat{F}_e(y) - F_e(y))$ $(-\infty < y \le \tau_e)$ *converges weakly to a zero-mean Gaussian process* $Z(y)$ *with covariance function*

$$Cov(Z(y), Z(y')) = p_{1\tau}^{-2} E\Big\{I(\Delta_1 = 1, T_1 \le \tau_1)\varphi(\tilde{T}_1, \tilde{T}_2, \Delta_2, y)\varphi(\tilde{T}_1, \tilde{T}_2, \Delta_2, y')\Big\}.$$

Note that, for the usual Kaplan-Meier estimator, the function φ in the above i.i.d representation needs to be replaced by the function ξ_e defined in Appendix A (see Lo and Singh (1986)). The extra term in the representation above is caused by the fact that our estimator of $F_e(y)$ is based on the estimated quantities $(\widetilde{T}_{2i} - \widehat{m}(\widetilde{T}_{1i}))/\widehat{\sigma}(\widetilde{T}_{1i})$, instead of the unobserved $(\widetilde{T}_{2i} - m(\widetilde{T}_{1i}))/\sigma(\widetilde{T}_{1i})$ $(i = 1, \dots, n)$.

Theorem 3.2 *Assume (A1)-(A4).*

(i) Then,

$$\widehat{F}(y|x) - F(y|x) = (na_n p_1)^{-1} h_{1|u}(x)^{-1} \sum_{i=1}^{n} K\Big(\frac{x - \tilde{T}_{1i}}{a_n}\Big) I(\Delta_{1i} = 1) h_{x,y}(\tilde{T}_{2i}, \Delta_{2i}) + R_n(y|x),$$

where $\sup\{|R_n(y|x)|; -\infty < (y - m(x))/\sigma(x) \le \tau_e$ *and* $x \le \tau_1\} = o_P((na_n)^{-1/2})$.

(ii) The process $(na_n)^{1/2}(\widehat{F}(y|x) - F(y|x))$ *(*$x \le \tau_1$ *fixed,* $(y - m(x))/\sigma(x) \le \tau_e$*) converges weakly to a zero-mean Gaussian process* $Z(y|x)$ *with covariance function*

$$Cov(Z(y|x), Z(y'|x)) = p_1^{-2} h_{1|u}(x)^{-1} \|K\|_2^2\, E[I(\Delta_1 = 1) h_{x,y}(\tilde{T}_2, \Delta_2) h_{x,y'}(\tilde{T}_2, \Delta_2) | \tilde{T}_1 = x],$$

where $\|K\|_2^2 = \int K^2(u)\, du$.

Next, we give the asymptotic analysis for the bivariate distribution function estimator. In the spirit of Akritas (1994) and Van Keilegom and Akritas (1999), the results will be established for the following slightly modified version of $\widehat{F}(x,y)$:

$$\widehat{F}_\tau(x,y) = \int_0^x \widehat{F}(y \wedge T_t|t)\, d\widehat{F}_1(t),$$

where $T_t \leq \tau_e\sigma(t) + m(t)$ of all t. Actually, this $\widehat{F}_\tau(x,y)$ is an estimator for $F_\tau(x,y) = \int_0^x F(y \wedge T_t|t)\, dF_1(t)$. Note, however, that $F_\tau(x,y)$ can become arbitrarily close to $F(x,y)$ if $\inf\{y; H_e(y)=1\} = \inf\{y; F_e(y)=1\}$ and T_t, respectively τ_e, is chosen sufficiently close to $\tau_e\sigma(t)+m(t)$, respectively $\inf\{y; H_e(y)=1\}$, for all t.

Theorem 3.3 *Assume (A1)-(A4).*

(i) Then,

$$\widehat{F}_\tau(x,y) - F_\tau(x,y) = n^{-1}\sum_{i=1}^n g_{x,y}(\tilde{T}_{1i}, \Delta_{1i}, \tilde{T}_{2i}, \Delta_{2i}) + R_n(x,y),$$

where $\sup\{|R_n(x,y)|; y \in I\!R \text{ and } x \leq \tau_1\} = o_P(n^{-1/2})$.

(ii) The process $n^{1/2}(\widehat{F}_\tau(x,y) - F_\tau(x,y))$ $(x \leq \tau_1, y \in I\!R)$, converges weakly to a zero-mean Gaussian process $Z(x,y)$ with covariance function

$$Cov(Z(x,y), Z(x',y')) = E\{g_{x,y}(\tilde{T}_1, \Delta_1, \tilde{T}_2, \Delta_2) g_{x',y'}(\tilde{T}_1, \Delta_1, \tilde{T}_2, \Delta_2)\}.$$

As for the bivariate distribution, the results pertaining to the transition probabilities imply preliminary modification of the estimators. Introduce

$$\widehat{p}_{12,\tau}(s,t) = \frac{1}{1-\widehat{F}_1(s)} \int_s^t \left[1 - \widehat{F}((t-r) \wedge T_r \mid r)\right] \widehat{F}_1(dr),$$

$$\widehat{p}_{22,\tau}(s,t) = \int_0^s \left[1 - \widehat{F}((t-r) \wedge T_r \mid r)\right] \widehat{F}_1(dr) \Big/ \int_0^s \left[1 - \widehat{F}((s-r) \wedge T_r \mid r)\right] \widehat{F}_1(dr),$$

and define also

$$p_{12,\tau}(s,t) = \frac{1}{1-F_1(s)} \int_s^t [1 - F((t-r) \wedge T_r \mid r)]\, F_1(dr)$$

$$p_{22,\tau}(s,t) = \int_0^s [1 - F((t-r) \wedge T_r \mid r)]\, F_1(dr) \Big/ \int_0^s [1 - F((s-r) \wedge T_r \mid r)]\, F_1(dr).$$

Theorem 3.4 *Assume (A1)-(A4).*

(i) Then,

$$\begin{aligned}
&\widehat{p}_{12,\tau}(s,t) - p_{12,\tau}(s,t) \\
&= n^{-1}\{1-F_1(s)\}^{-1} \sum_{i=1}^{n} \Big[l_{s,t}(\widetilde{T}_{1i},\Delta_{1i},\widetilde{T}_{2i},\Delta_{2i}) + p_{12,\tau}(s,t)\xi_1(\widetilde{T}_{1i},\Delta_{1i},s) \Big] + \widetilde{R}_n(s,t),
\end{aligned}$$

where $\sup\{|\widetilde{R}_n(s,t)| : 0 \le s < t \le \tau_1\} = o_P(n^{-1/2})$.

(ii) The process $n^{1/2}(\widehat{p}_{12,\tau}(s,t) - p_{12,\tau}(s,t))$ $(0 \le s < t \le \tau_1)$ *converges weakly to a zero-mean Gaussian process* $\widetilde{Z}(s,t)$ *with covariance function*

$$\begin{aligned}
&Cov(\widetilde{Z}(s,t), \widetilde{Z}(s',t')) \\
&= \{1-F_1(s)\}^{-1}\{1-F_1(s')\}^{-1} E\Big\{ \Big[l_{s,t}(\widetilde{T}_1,\Delta_1,\widetilde{T}_2,\Delta_2) + p_{12,\tau}(s,t)\xi_1(\widetilde{T}_1,\Delta_1,s) \Big] \\
&\quad \times \Big[l_{s',t'}(\widetilde{T}_1,\Delta_1,\widetilde{T}_2,\Delta_2) + p_{12,\tau}(s',t')\xi_1(\widetilde{T}_1,\Delta_1,s') \Big] \Big\}.
\end{aligned}$$

Theorem 3.5 *Assume (A1)-(A4).*

(i) Then,

$$\begin{aligned}
&\widehat{p}_{22,\tau}(s,t) - p_{22,\tau}(s,t) \\
&= \Big\{ \int_0^s [1 - F((s-r)\wedge T_r \mid r)]\, F_1(dr) \Big\}^{-1} \\
&\quad \times n^{-1} \sum_{i=1}^{n} \Big[l^*_{s,t}(\widetilde{T}_{1i},\Delta_{1i},\widetilde{T}_{2i},\Delta_{2i}) - p_{22,\tau}(s,t) l^*_{s,s}(\widetilde{T}_{1i},\Delta_{1i},\widetilde{T}_{2i},\Delta_{2i}) \Big] + R^*_n(s,t),
\end{aligned}$$

where $\sup\{|R^*_n(s,t)| : 0 \le s < t \le \tau_1\} = o_P(n^{-1/2})$.

(ii) The process $n^{1/2}(\widehat{p}_{22,\tau}(s,t) - p_{22,\tau}(s,t))$ $(0 \le s < t \le \tau_1)$ *converges weakly to a zero-mean Gaussian process* $Z^*(s,t)$ *with covariance function*

$$\begin{aligned}
&Cov(Z^*(s,t), Z^*(s',t')) \\
&= \Big\{ \int_0^s [1 - F((s-r)\wedge T_r \mid r)]\, F_1(dr) \Big\}^{-1} \Big\{ \int_0^{s'} [1 - F((s'-r)\wedge T_r \mid r)]\, F_1(dr) \Big\}^{-1} \\
&\quad \times E\Big\{ \Big[l^*_{s,t}(\widetilde{T}_1,\Delta_1,\widetilde{T}_2,\Delta_2) - p_{22,\tau}(s,t) l^*_{s,s}(\widetilde{T}_1,\Delta_1,\widetilde{T}_2,\Delta_2) \Big] \\
&\qquad \times \Big[l^*_{s',t'}(\widetilde{T}_1,\Delta_1,\widetilde{T}_2,\Delta_2) - p_{22,\tau}(s',t') l^*_{s',s'}(\widetilde{T}_1,\Delta_1,\widetilde{T}_2,\Delta_2) \Big] \Big\}.
\end{aligned}$$

4 Simulation study

In this section we carry out some simulations to demonstrate the behavior of the proposed estimators for finite sample sizes. For comparison purposes, we include in the simulations some existing estimators for the conditional and the bivariate distribution functions, as well as estimators for the transition probabilities, which do not use the information given by the location-scale model. In this way we can get an idea of the improvement associated with the transfer of tail information. We concentrate on the estimation in the case of heavy censoring, because the methods proposed in this paper are mainly relevant when the uncensored information is scarce. An exception to this will be the simulations for the transition probabilities, for which the effect of an increasing censoring level will be fully illustrated.

The vector of gap times (T_1, T_2) is simulated as follows. The first gap time T_1 is simulated according to an $Exp(1)$ distribution. Given $T_1 = x$, the second gap time T_2 is drawn from an exponential distribution with rate parameter $1/(1 + 5x + 2x^2)$. The censoring time C is independently generated following also an exponential distribution, with rate chosen in order to get a pre-determined censoring proportion on the total time $Y = T_1 + T_2$: 25% (3.7% and 22.1% for the first and second gap times, respectively), 50% (11.4% and 43.6%) and 75% (28.3% and 65.1%). We consider sample sizes $n = 100$ and $n = 150$. The number of replications is 500. For the kernel weights in the Beran-type estimator we use the biquadratic kernel $K(u) = (15/16)(1 - u^2)^2 I(|u| \le 1)$, while the bandwidth sequence a_n is chosen by minimizing the asymptotic mean-squared error (AMSE):

$$\text{AMSE}(a_n) = \text{AsVar}(\tilde{F}(y|x)) + \big(\text{AsBias}(\tilde{F}(y|x))\big)^2 = (na_n)^{-1} s^2(y|x) + a_n^4 b^2(y|x),$$

where the formulas for $s(y|x)$ and $b(y|x)$ are given in Van Keilegom et al. (2001). The only difference with that paper is that here the variable T_1 is subject to censoring, and therefore we restrict attention to the subsample where $\Delta_{1i} = 1$. Hence the optimal choice for the bandwidth sequence is given by

$$a_n = a_n(y|x) = \Big(\frac{s^2(y|x)}{4b^2(y|x)}\Big)^{1/5} n^{-1/5}.$$

Further details about the choice of the optimal bandwidth and its implementation to real data sets can be found in Van Keilegom et al. (2001).

For the score function $J(\cdot)$ we follow Remark B.1 with $\delta = 0$, i.e. we take $J(s) = I(a \le s \le b)/(b - a)$, where $a = 0$ and $b = \min_i \widetilde{F}(+\infty|X_i)$. Note that since $\widetilde{F}^{-1}(s|x)$ is only defined for s less than $\widetilde{F}(+\infty|x)$, this choice of b ensures that $\widehat{m}(x)$ and $\widehat{\sigma}(x)$ are always defined.

To show that the location-scale model $T_2 = m(T_1) + \sigma(T_1)\varepsilon$, ε independent of T_1, holds in the simulated scenario, it suffices (as noted by Van Keilegom *et al.*, 2001) to check the model when $m(x)$ is the conditional mean function and $\sigma(x)$ is the conditional standard deviation. This follows easily by showing that $P(\varepsilon \leq y \mid T_1 = x) = \{1 - \exp[-(y+1)]\} I(y \geq -1)$, which is independent of x.

In Table 1 we report the mean, the standard deviation, and the mean squared error (MSE) of the estimator $\widehat{F}(y \mid x)$ for $x = 0.5$, for percentiles 80, 90 and 99% of the conditional distribution, and the heavily censored case (75% of censoring). We focus on the right tail of the distribution, because the censoring effects are strong at those points. We also give the figures pertaining to the Beran-type estimator $\widetilde{F}(y \mid x)$ of the conditional distribution, adapted to censoring of the covariate (Van Keilegom, 2004). Note that the latter estimator does not rely on the location-scale model. From Table 1 we see that the new method outperforms the Beran-type estimator. This is particularly clear at the far right tail (percentile 99%), where the Beran estimator shows a large bias and a relative efficiency (measured as the ratio of the MSE's) of about 50% or below. We have also performed simulations in the case of light (25%) and medium (50%) censoring (results not shown). In these cases, the use of transfer of tail information in the estimation of $F(y \mid x)$ does not improve the efficiency so clearly. We note, however, that if the location-scale model does not hold, then the proposed estimator may not perform as well as the Beran-type estimator.

[Insert Table 1 here]

Table 2 displays the results achieved by the estimator of the bivariate distribution $\widehat{F}(x, y)$. As (x, y)-pairs we take $(0.500, 2.157)$, $(1.000, 5.400)$ and $(4.000, 12.870)$, giving respectively $F(x, y)$-values of 0.25, 0.50 and 0.80. In this case, we consider the simple estimator $\widetilde{F}(x, y)$ of the bivariate distribution introduced in de Uña-Álvarez and Meira-Machado (2008) as a natural competitor. This latter estimator is constructed by attaching to each pair of observed gap times $(\widetilde{T}_{1i}, \widetilde{T}_{2i})$ the ordinary Kaplan-Meier weight of $\widetilde{T}_{1i} + \widetilde{T}_{2i}$ when estimating the distribution of the total time $Y = T_1 + T_2$. As was the case for the conditional distribution, the simulation results indicate that the location-scale model leads to more efficient estimation of the bivariate distribution. Results obtained for other censoring levels confirm this finding (results not included).

[Insert Table 2 here]

The performance of the empirical transition probability $\widehat{p}_{22}(s,t)$ in the simulations is summarized in Tables 3 and 4. Three different values of $p_{22}(s,t)$ are considered, corresponding to $s = 0.5$ and $t = 3.709$ ($p_{22}(s,t) = 0.25$), $t = 2.022$ ($p_{22}(s,t) = 0.50$), and $t = 1.112$ ($p_{22}(s,t) = 0.75$). The results refer to three censoring levels (25%, 50% and 75%). In Table 3, we provide the mean and the standard deviation (along the 500 trials) for $\widehat{p}_{22}(s,t)$ and also for two different estimators, $\widetilde{p}_{22}(s,t)$ and $\widehat{p}_{22}^{AJ}(s,t)$. The notation $\widehat{p}_{22}^{AJ}(s,t)$ stands for the Aalen-Johansen estimator of $p_{22}(s,t)$, see Aalen and Johansen (1978), which is a consistent estimator when the gap times satisfy the Markov assumption. Our simulated scenario is non-Markov, a fact that can be easily checked by using the exponential distribution of T_2 given $T_1 = x$. The estimator $\widetilde{p}_{22}$ is that proposed in Meira-Machado *et al.* (2006), and is defined as a ratio of two multivariate Kaplan-Meier integrals with respect to the marginal distribution of the total time $Y = T_1 + T_2$. This is justified because

$$p_{22}(s,t) = \frac{E\left[I(T_1 \leq s, Y > t)\right]}{E\left[I(T_1 \leq s, Y > s)\right]}.$$

The censoring on T_1 is overcome because, in the context of our model for gap times, T_1 is uncensored whenever this is the case for Y. See Meira-Machado *et al.* (2006) for more details. The estimator $\widetilde{p}_{22}$ was proposed as an alternative to the Aalen-Johansen estimator $\widehat{p}_{22}^{AJ}$ in non-Markov situations. Of course, unlike $\widehat{p}_{22}(s,t)$, $\widetilde{p}_{22}(s,t)$ and $\widehat{p}_{22}^{AJ}(s,t)$ do not make use of the location-scale model.

[Insert Tables 3 and 4 here]

From Table 3 we see that the Aalen-Johansen estimator is systematically biased. This is because the non-Markov nature of the simulated gap times. When comparing $\widehat{p}_{22}(s,t)$ and $\widetilde{p}_{22}(s,t)$, we see that the estimator based on the location-scale model behaves more efficiently. This is more clearly seen in Table 4, where the MSEs of both estimators are compared. Indeed, Table 4 suggests that using the information contained in the location-scale model is more relevant in the heavily censored case, and for the estimation of large values of $p_{22}(s,t)$. Interestingly, the relative efficiency of $\widetilde{p}_{22}(s,t)$ is as poor as about 30% in some cases.

Appendix A: Notations

In this Appendix we collect all the notations that are needed for the asymptotic results stated in Section 3. The notations are introduced in the order in which they are required in Section 3.

Let $p_1 = P(\Delta_1 = 1)$, $p_{1\tau} = P(\Delta_1 = 1, T_1 \le \tau_1)$, with $\tau_1 < \inf\{t; H_1(t) = 1\}$, where $H_1(t) = P(\tilde{T}_1 \le t)$. Furthermore, $E = (\tilde{T}_2 - m(T_1))/\sigma(T_1)$, $H_e(y) = P(E \le y|\Delta_1 = 1)$, $\tau_e < \inf\{y; H_e(y) = 1\}$, $H_e^u(y) = P(E \le y, \Delta_2 = 1|\Delta_1 = 1)$, $H_e(y|x) = P(E \le y|\Delta_1 = 1, T_1 = x)$, $H_e^u(y|x) = P(E \le y, \Delta_2 = 1|\Delta_1 = 1, T_1 = x)$, $H(y|x) = P(\tilde{T}_2 \le y|\Delta_1 = 1, T_1 = x)$, and $H^u(y|x) = P(\tilde{T}_2 \le y, \Delta_2 = 1|\Delta_1 = 1, T_1 = x)$. Also, let $h_e(y|x)$ and $h_e^u(y|x)$ denote the probability (sub)density functions attached to $H_e(y|x)$ and $H_e^u(y|x)$, respectively. Moreover, introduce the functions

$$\xi_e(t,\delta,y) = (1 - F_e(y))\left\{-\int_{-\infty}^{y\wedge t} \frac{dH_e^u(s)}{(1-H_e(s))^2} + \frac{I(t \le y, \delta = 1)}{1 - H_e(t)}\right\},$$

$$\xi(t,\delta,y|x) = (1 - F(y|x))\left\{-\int_{-\infty}^{y\wedge t} \frac{dH^u(s|x)}{(1-H(s|x))^2} + \frac{I(t \le y, \delta = 1)}{1 - H(t|x)}\right\},$$

$$\eta(t,\delta|x) = \int_{-\infty}^{+\infty} \xi(t,\delta,v|x) J(F(v|x))\, dv\, \sigma^{-1}(x),$$

$$\zeta(t,\delta|x) = \int_{-\infty}^{+\infty} \xi(t,\delta,v|x) J(F(v|x)) \frac{v - m(x)}{\sigma(x)}\, dv\, \sigma^{-1}(x),$$

$$\gamma_1(y|x) = \int_{-\infty}^{y} \frac{h_e(s|x)}{(1-H_e(s))^2}\, dH_e^u(s) + \int_{-\infty}^{y} \frac{d\, h_e^u(s|x)}{1 - H_e(s)},$$

$$\gamma_2(y|x) = \int_{-\infty}^{y} \frac{s h_e(s|x)}{(1-H_e(s))^2}\, dH_e^u(s) + \int_{-\infty}^{y} \frac{d\,(s h_e^u(s|x))}{1 - H_e(s)},$$

and let with $S_e = 1 - F_e$,

$$\begin{aligned}\varphi(t_1,t_2,\delta_2,y) \;&=\; \xi_e\left(\frac{t_2 - m(t_1)}{\sigma(t_1)}, \delta_2, y\right) - S_e(y)\eta(t_2,\delta_2|t_1)\gamma_1(t_2|t_1)\\ &\quad - S_e(y)\zeta(t_2,\delta_2|t_1)\gamma_2(t_2|t_1).\end{aligned}$$

Introduce $H_{1|u}(t) = P(\tilde{T}_1 \le t|\Delta_1 = 1)$, and let $f_e(y)$ and $h_{1|u}(t)$ denote the probability density functions attached to $F_e(y)$ and $H_{1|u}(t)$, respectively. Furthermore, let $H_1^u(t) = P(\tilde{T}_1 \le t, \Delta_1 = 1)$, and define

$$h_{x,y}(t,\delta) = f_e\Big(\frac{y - m(x)}{\sigma(x)}\Big)\Big[\eta(t,\delta|x) + \frac{y - m(x)}{\sigma(x)}\zeta(t,\delta|x)\Big].$$

In addition, let

$$g_{x,y}(t_1,\delta_1,t_2,\delta_2) = \sum_{j=1}^{3} g^j_{x,y}(t_1,\delta_1,t_2,\delta_2),$$

where

$$g^1_{x,y}(t_1,\delta_1,t_2,\delta_2) = p_{1\tau}^{-1} I(\delta_1=1, t_1\le\tau_1) E\Big\{\varphi\Big(t_1,t_2,\delta_2,\frac{y\wedge T_{T_1} - m(T_1)}{\sigma(T_1)}\Big) I(T_1\le x)\Big\}$$

$$g^2_{x,y}(t_1,\delta_1,t_2,\delta_2) = p_1^{-1} I(\delta_1=1) h_{t_1,y\wedge T_{t_1}}(t_2,\delta_2) I(t_1\le x)$$

$$g^3_{x,y}(t_1,\delta_1,t_2,\delta_2) = \int_0^x F_e\Big(\frac{y\wedge T_t - m(t)}{\sigma(t)}\Big)\, d\xi_1(t_1,\delta_1,t),$$

and

$$\xi_1(t,\delta,x) = (1-F_1(x))\Big\{-\int_0^{x\wedge t}\frac{dH_1^u(s)}{(1-H_1(s))^2} + \frac{I(t\le x,\delta=1)}{1-H_1(t)}\Big\}.$$

Moreover, we need to introduce

$$l_{s,t}(t_1,\delta_1,t_2,\delta_2) = \sum_{i=1}^{3}\big\{l^i_{t,t}(t_1,\delta_1,t_2,\delta_2) - l^i_{s,t}(t_1,\delta_1,t_2,\delta_2)\big\}$$

and

$$l^*_{s,t}(t_1,\delta_1,t_2,\delta_2) = \sum_{i=1}^{3} l^i_{s,t}(t_1,\delta_1,t_2,\delta_2),$$

where

$$l^1_{s,t}(t_1,\delta_1,t_2,\delta_2) = -p_{1\tau}^{-1} I(\delta_1=1,t_1\le\tau_1) E\left\{\varphi\left(t_1,t_2,\delta_2,\frac{(t-T_1)\wedge T_{T_1} - m(T_1)}{\sigma(T_1)}\right) I(T_1\le s)\right\},$$

$$l^2_{s,t}(t_1,\delta_1,t_2,\delta_2) = -g^2_{s,t-t_1}(t_1,\delta_1,t_2,\delta_2),$$

$$l^3_{s,t}(t_1,\delta_1,t_2,\delta_2) = \int_0^s\left[1 - F_e\left(\frac{(t-r)\wedge T_r - m(r)}{\sigma(r)}\right)\right] d\xi_1(t_1,\delta_1,r).$$

B Appendix B: Proofs

In this section, we give the proofs of the results stated in Section 3. The proofs of Theorems 3.1(*ii*), 3.2(*ii*), 3.3(*ii*), 3.4(*ii*) and 3.5(*ii*) are similar to the proofs of Corollaries 3.2, 3.4 and 3.6 in Van Keilegom and Akritas (1999) and are therefore omitted.

Let $\tilde{T}_x$ be any value less than the upper bound of the support of $H(\cdot|x)$ such that $\inf_{x\le\tau_1}(1-H(\tilde{T}_x|x)) > 0$. The assumptions we need for the proofs of the main results are listed below.

(A1)(i) $na_n^4 \to 0$ and $na_n^{3+d} \to \infty$ for some $d > 0$.

(ii) K has compact support, $\int uK(u)\,du = 0$ and K is twice continuously differentiable.

(A2)(i) There exist $0 \leq s_0 \leq s_1 \leq 1$ such that $s_1 \leq \inf_{x \leq \tau_1} F(\tilde{T}_x|x)$, $s_0 \leq \inf\{s \in [0,1]; J(s) \neq 0\}$, $s_1 \geq \sup\{s \in [0,1]; J(s) \neq 0\}$ and $\inf_{x \leq \tau_1} \inf_{s_0 \leq s \leq s_1} f(F^{-1}(s|x)|x) > 0$.

(ii) J is twice continuously differentiable on the interior of its support, $\int_0^1 J(s)\,ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$.

(iii) T_x is twice continuously differentiable in x.

(A3) $H_1(x)$ and $H_{1|u}(x)$ are thrice continuously differentiable, $\inf_{x \leq \tau_1} h_{1|u}(x) > 0$, and $\inf_{x \leq \tau_1} \sigma(x) > 0$.

(A4) $H(y|x)$ and $H^u(y|x)$ are thrice continuously differentiable with respect to x and y, $\sup_{x,y} |y^3 H'(y|x)| < \infty$, and similarly for all other derivatives of $H(y|x)$ and $H^u(y|x)$ with respect to x and y up to order 3.

Remark B.1 *In practice, the function J can be chosen as follows : $J(s) = I(a \leq s \leq b)/(b-a)$, where $a = \delta > 0$ and $b = \min_i \widetilde{F}(+\infty|X_i) - \delta$. This choice ensures that conditions (A2)(i)-(ii) are a.s. satisfied, provided that the density $f(y|x)$ is strictly positive for all y and x. In practice δ can be taken very small, or even equal to zero (as is done in the simulation study). Also note that the function T_x will be twice differentiable in x (as is required in (A2)(iii)) when, e.g., T_x is defined as $T_x = T_e\sigma(x) + m(x)$ for some $T_e < \tau_e$ and $m(x)$ and $\sigma(x)$ are twice differentiable.*

For the proofs, we need to introduce the following estimators of $H_{e\tau}(y) = P(E \leq y|\Delta_1 = 1, T_1 \leq \tau_1)$ and $H^u_{e\tau}(y) = P(E \leq y, \Delta_2 = 1|\Delta_1 = 1, T_1 \leq \tau_1)$: $\widehat{H}_{e\tau}(y) = N_{u\tau}^{-1} \sum_{i=1}^n I(\widehat{E}_i \leq y, \Delta_{1i} = 1, \tilde{T}_{1i} \leq \tau_1)$ and $\widehat{H}^u_{e\tau}(y) = N_{u\tau}^{-1} \sum_{i=1}^n I(\widehat{E}_i \leq y, \Delta_{1i} = 1, \tilde{T}_{1i} \leq \tau_1, \Delta_{2i} = 1)$. We start with an auxiliary result, which gives an i.i.d. representation for the estimators $\widehat{m}(x)$ and $\widehat{\sigma}(x)$.

Proposition B.2 *Assume (A1)-(A4). Then,*

$$\widehat{m}(x) - m(x) = -h_{1|u}(x)^{-1}\sigma(x)(na_np_1)^{-1} \sum_{i=1}^n K\Big(\frac{x - \tilde{T}_{1i}}{a_n}\Big) I(\Delta_{1i} = 1)\eta(\tilde{T}_{2i}, \Delta_{2i}|x) + R_n(x),$$

and

$$\widehat{\sigma}(x) - \sigma(x) = -h_{1|u}(x)^{-1}\sigma(x)(na_np_1)^{-1} \sum_{i=1}^n K\Big(\frac{x - \tilde{T}_{1i}}{a_n}\Big) I(\Delta_{1i} = 1)\zeta(\tilde{T}_{2i}, \Delta_{2i}|x) + \tilde{R}_n(x),$$

where $\sup\{|R_n(x)|; x \le \tau_1\} = O((na_n)^{-3/4}(\log n)^{3/4})$ *a.s., and* $\sup\{|\tilde{R}_n(x)|; x \le \tau_1\} = O((na_n)^{-3/4}(\log n)^{3/4})$ *a.s.*

Proof. The result is an easy consequence of Theorems 4.8 and 4.9 in Van Keilegom and Akritas (1999), where representations for $\widehat{m}(x)$ and $\widehat{\sigma}(x)$ are obtained when T_1 is completely observed. Consider for example $\widehat{m}(x)$. Put $N_u = \sum_{i=1}^n I(\Delta_{1i} = 1)$. We have

$$\widehat{m}(x) - m(x) = -h_{1|u}(x)^{-1}\sigma(x)(N_u a_n)^{-1}\sum_{i=1}^n K\Big(\frac{x - \tilde{T}_{1i}}{a_n}\Big)I(\Delta_{1i} = 1)\eta(\tilde{T}_{2i}, \Delta_{2i}|x) + R_n(x)$$

$$= -h_{1|u}(x)^{-1}\sigma(x)(na_n p_1)^{-1}\sum_{i=1}^n K\Big(\frac{x - \tilde{T}_{1i}}{a_n}\Big)I(\Delta_{1i} = 1)\eta(\tilde{T}_{2i}, \Delta_{2i}|x) + R_n(x),$$

since

$$np_1\Big[\frac{1}{N_u} - \frac{1}{np_1}\Big] = -\frac{1}{np_1}\sum_{i=1}^n [I(\Delta_{1i} = 1) - p_1] + o_P(n^{-1/2}) = O_P(n^{-1/2}). \quad \text{(B.1)}$$

Remark B.3 *Propositions 4.5-4.7 in Van Keilegom and Akritas (1999), which deal with the uniform consistency (with rate of convergence) of* $\widehat{m}(x)$, $\widehat{\sigma}(x)$ *and their derivatives can be easily adapted to the present context. We omit the details.*

Lemma B.4 *Assume (A1)-(A4). Then,*

$$\sup_{-\infty<y<+\infty}\Big|(N_{u\tau})^{-1}\sum_{i=1}^n \Big\{I(\widehat{E}_i \le y, \Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) - I(E_i \le y, \Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1)\Big\}$$

$$-\Big\{P(\widehat{E} \le y|\mathcal{X}_n, \Delta_1 = 1, T_1 \le \tau_1) - P(E \le y|\Delta_1 = 1, T_1 \le \tau_1)\Big\}\Big| = o_P(n^{-1/2}),$$

where $P(\widehat{E} \le y|\mathcal{X}_n, \Delta_1 = 1, T_1 \le \tau_1)$ *is the distribution of* $\widehat{E} = (\tilde{T}_2 - \widehat{m}(T_1))/\widehat{\sigma}(T_1)$ *conditioning on* $\Delta_1 = 1$*, on* $T_1 \le \tau_1$ *and on* $(\tilde{T}_{1j}, \tilde{T}_{2j}, \Delta_{1j}, \Delta_{2j})$, $j = 1, \ldots, n$.

Proof. The expression between absolute values can be written as

$$(np_{1\tau})^{-1}\sum_{i=1}^n \Big\{I(\widehat{E}_i \le y, \Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) - I(E_i \le y, \Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1)\Big\}$$

$$-\Big\{P(\widehat{E} \le y|\mathcal{X}_n, \Delta_1 = 1, T_1 \le \tau_1) - P(E \le y|\Delta_1 = 1, T_1 \le \tau_1)\Big\}$$

$$+\Big[\frac{1}{N_{u\tau}} - \frac{1}{np_{1\tau}}\Big]\sum_{i=1}^n \Big\{I(\widehat{E}_i \le y, \Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) - I(E_i \le y, \Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1)\Big\}$$

$$= T_1 + T_2.$$

In a very similar way as in Lemma A.1 in Van Keilegom and Akritas (1999), it can be shown that $T_1 = o_P(n^{-1/2})$ (the only difference with that lemma is the conditioning on $\Delta_1 = 1$ and on $\tilde{T}_1 \le \tau_1$). For T_2, note that

$$
\begin{aligned}
&(np_{1\tau})^{-1} \sum_{i=1}^n \Big\{ I(\widehat{E}_i \le y, \Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) - I(E_i \le y, \Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) \Big\} \\
&= P(\widehat{E} \le y | \mathcal{X}_n, \Delta_1 = 1, T_1 \le \tau_1) - P(E \le y | \Delta_1 = 1, T_1 \le \tau_1) + o_P(n^{-1/2}) = o_P(1),
\end{aligned}
$$

uniformly in y, where the latter equality follows from the uniform consistency of $\widehat{m}$ and $\widehat{\sigma}$ (see Remark B.3). Hence, it follows from the analogue of (B.1) for $N_{u\tau}$, that $T_2 = o_P(n^{-1/2})$. □

Proposition B.5 *Assume (A1)-(A4). Then,*

$$
\begin{aligned}
&\widehat{H}_{e\tau}(y) - H_{e\tau}(y) \qquad \text{(B.2)} \\
&= -(np_{1\tau})^{-1} \sum_{i=1}^n I(\Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) \Big\{ \eta(\tilde{T}_{2i}, \Delta_{2i} | \tilde{T}_{1i}) + \zeta(\tilde{T}_{2i}, \Delta_{2i} | \tilde{T}_{1i}) y \Big\} h_e(y | \tilde{T}_{1i}) \\
&\quad + (np_{1\tau})^{-1} \sum_{i=1}^n I(\Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) \Big\{ I(E_i \le y) - H_{e\tau}(y) \Big\} + R_n(y),
\end{aligned}
$$

where $\sup\{|R_n(y)|; -\infty < y < +\infty\} = o_P(n^{-1/2})$.

Proof. Using Lemma B.4,

$$
\begin{aligned}
&\widehat{H}_{e\tau}(y) - H_{e\tau}(y) \\
&= p_{1\tau}^{-1} \int_0^{\tau_1} \left\{ H_e\left(\frac{y\widehat{\sigma}(x) + \widehat{m}(x) - m(x)}{\sigma(x)} \middle| x \right) - H_e(y|x) \right\} dH_1^u(x) \\
&\quad + N_{u\tau}^{-1} \sum_{i=1}^n I(\Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) \Big\{ I(E_i \le y) - H_{e\tau}(y) \Big\} + o_P(n^{-1/2}) \\
&= p_{1\tau}^{-1} \int_0^{\tau_1} h_e(y|x) \frac{y(\widehat{\sigma}(x) - \sigma(x)) + (\widehat{m}(x) - m(x))}{\sigma(x)} \, dH_1^u(x) \qquad \text{(B.3)} \\
&\quad + \frac{1}{2p_{1\tau}} \int_0^{\tau_1} h_e'(\xi|x) \left(\frac{y(\widehat{\sigma}(x) - \sigma(x)) + (\widehat{m}(x) - m(x))}{\sigma(x)} \right)^2 dH_1^u(x) \\
&\quad + (np_{1\tau})^{-1} \sum_{i=1}^n I(\Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) \Big\{ I(E_i \le y) - H_{e\tau}(y) \Big\} \\
&\quad + \Big[(N_{u\tau})^{-1} - (np_{1\tau})^{-1} \Big] \sum_{i=1}^n I(\Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) \Big\{ I(E_i \le y) - H_{e\tau}(y) \Big\} + o_P(n^{-1/2}),
\end{aligned}
$$

where ξ is between y and $(y\widehat{\sigma}(x) + \widehat{m}(x) - m(x))/\sigma(x)$. The first term above can be treated in a very similar way as in the proof of Proposition A.2 in Van Keilegom and Akritas (1999), and leads to the first term on the right hand side of (B.2), while the second term is $o_P(n^{-1/2})$ by the rates of the uniform consistency of $\widehat{m}$ and $\widehat{\sigma}$ (see Remark B.3). For the fourth term, use the derivation in (B.1) together with the fact that $(np_{1\tau})^{-1}\sum_{i=1}^n I(\Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1)\{I(E_i \le y) - H_{e\tau}(y)\} = o_P(1)$ uniformly in y. $\square$

Remark B.6 *In a similar way as has been done in the proofs of Lemma B.4 and Proposition B.5 above, it can be shown that Proposition A.3 and Corollary A.5 in Van Keilegom and Akritas (1999) can be adapted to the present context. They are needed in the following proof.*

Proof of Theorem 3.1(i). Note that

$$\begin{aligned}
&\widehat{F}_e(y) - F_e(y) \\
&= -(1 - F_e(y))\Big\{\log(1 - \widehat{F}_e(y)) - \log(1 - F_e(y))\Big\} + o_P(n^{-1/2}) \\
&= -(1 - F_e(y))\Big\{\log(1 - \widehat{F}_e(y)) + \int_{-\infty}^{y} \frac{1}{1 - H_{e\tau}(s)} dH_{e\tau}^u(s)\Big\} + o_P(n^{-1/2}),
\end{aligned}$$

uniformly in y, where the latter equality follows from the independence of ε and $\{C - T_1 - m(T_1)\}/\sigma(T_1)$ given $\Delta_1 = 1$ and $T_1 \le \tau_1$. Next, write

$$\begin{aligned}
&\log(1 - \widehat{F}_e(y)) + \int_{-\infty}^{y} \frac{1}{1 - H_{e\tau}(s)} dH_{e\tau}^u(s) \\
&= \log(1 - \widehat{F}_e(y)) + \int_{-\infty}^{y} \frac{1}{1 - \widehat{H}_{e\tau}(s-)} d\widehat{H}_{e\tau}^u(s) \\
&\quad - \int_{-\infty}^{y} \frac{1}{1 - \widehat{H}_{e\tau}(s)} d\widehat{H}_{e\tau}^u(s) + \int_{-\infty}^{y} \frac{1}{1 - H_{e\tau}(s)} dH_{e\tau}^u(s) + O_P(n^{-1}).
\end{aligned}$$

Using a Taylor expansion, the first term above can be written as

$$-\frac{1}{2}\sum_{i=1}^{N_{u\tau}} \frac{I(\widehat{E}_{(i)} \le y, \Delta_{(2i)} = 1)}{(N_{u\tau} - i + 1)^2} \frac{1}{(1 - R_i)^2} = O(n^{-1})$$

a.s., uniformly in y, where R_i is between 0 and $(N_{u\tau} - i + 1)^{-1}$. The second term equals

$$\begin{aligned}
&-\int_{-\infty}^{y} \left[\frac{1}{1 - \widehat{H}_{e\tau}(s)} - \frac{1}{1 - H_{e\tau}(s)}\right] dH_{e\tau}^u(s) - \int_{-\infty}^{y} \frac{1}{1 - H_{e\tau}(s)} d(\widehat{H}_{e\tau}^u(s) - H_{e\tau}^u(s)) \\
&\quad - \int_{-\infty}^{y} \left[\frac{1}{1 - \widehat{H}_{e\tau}(s)} - \frac{1}{1 - H_{e\tau}(s)}\right] d(\widehat{H}_{e\tau}^u(s) - H_{e\tau}^u(s)).
\end{aligned}$$

From Remark B.6 and Corollary A.5 in Van Keilegom and Akritas (1999), it follows that the last term on the right hand side is $o_P(n^{-1/2})$. Using the consistency of $\widehat{H}_{e\tau}$ (which can be established along the same lines as in Proposition A.3 in Van Keilegom and Akritas (1999); see Remark B.6), the sum of the first and second terms can be written as

$$
\begin{aligned}
&-\int_{-\infty}^{y} \frac{\widehat{H}_{e\tau}(s) - H_{e\tau}(s)}{(1 - H_{e\tau}(s))^2} dH_{e\tau}^u(s) - \int_{-\infty}^{y} \frac{1}{1 - H_{e\tau}(s)} d(\widehat{H}_{e\tau}^u(s) - H_{e\tau}^u(s)) + o_P(n^{-1/2}) \\
&= -(1 - F_e(y))^{-1}(np_{1\tau})^{-1} \sum_{i=1}^{n} I(\Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1)\varphi(\tilde{T}_{1i}, \tilde{T}_{2i}, \Delta_{2i}, y) + o_P(n^{-1/2}),
\end{aligned}
$$

where the equality follows by Proposition B.5 and its analogue for $\widehat{H}_{e\tau}^u$. □

Proof of Theorem 3.2(i). The proof of this result is very similar to that of Theorem 3.3 in Van Keilegom and Akritas (1999). The only difference is that the data set is here restricted to the observations for which $\Delta_{1i} = 1$. Hence, it follows from that theorem that

$$
\begin{aligned}
&\widehat{F}(y|x) - F(y|x) \\
&= (N_u a_n)^{-1} h_{1|u}(x)^{-1} \sum_{i=1}^{n} K\Big(\frac{x - \tilde{T}_{1i}}{a_n}\Big) I(\Delta_{1i} = 1) h_{x,y}(\tilde{T}_{2i}, \Delta_{2i}) + o_P((na_n)^{-1/2}) \\
&= (na_n p_1)^{-1} h_{1|u}(x)^{-1} \sum_{i=1}^{n} K\Big(\frac{x - \tilde{T}_{1i}}{a_n}\Big) I(\Delta_{1i} = 1) h_{x,y}(\tilde{T}_{2i}, \Delta_{2i}) + o_P((na_n)^{-1/2}),
\end{aligned}
$$

where the last equality follows from (B.1). □

Proof of Theorem 3.3(i). Write

$$
\begin{aligned}
&\widehat{F}_\tau(x, y) - F_\tau(x, y) \\
&= \int_0^x \Big\{\widehat{F}_e\Big(\frac{y \wedge T_t - \widehat{m}(t)}{\widehat{\sigma}(t)}\Big) - F_e\Big(\frac{y \wedge T_t - m(t)}{\sigma(t)}\Big)\Big\} dF_1(t) \\
&\quad + \int_0^x \Big\{\widehat{F}_e\Big(\frac{y \wedge T_t - \widehat{m}(t)}{\widehat{\sigma}(t)}\Big) - F_e\Big(\frac{y \wedge T_t - m(t)}{\sigma(t)}\Big)\Big\} d(\widehat{F}_1(t) - F_1(t)) \\
&\quad + \int_0^x F_e\Big(\frac{y \wedge T_t - m(t)}{\sigma(t)}\Big) d(\widehat{F}_1(t) - F_1(t)) \\
&= T_1 + T_2 + T_3.
\end{aligned}
$$

We first consider T_3 :

$$\begin{aligned}
T_3 &= \{\widehat{F}_1(x) - F_1(x)\}\, F_e\Big(\frac{y \wedge T_x - m(x)}{\sigma(x)}\Big) - \int_0^x \{\widehat{F}_1(t) - F_1(t)\}\, dF_e\Big(\frac{y \wedge T_t - m(t)}{\sigma(t)}\Big) \\
&= n^{-1} \sum_{i=1}^n \xi_1(\tilde{T}_{1i}, \Delta_{1i}, x) F_e\Big(\frac{y \wedge T_x - m(x)}{\sigma(x)}\Big) \\
&\quad -n^{-1} \sum_{i=1}^n \int_0^x \xi_1(\tilde{T}_{1i}, \Delta_{1i}, t)\, dF_e\Big(\frac{y \wedge T_t - m(t)}{\sigma(t)}\Big) + o_P(n^{-1/2}) \\
&= n^{-1} \sum_{i=1}^n g^3_{x,y}(\tilde{T}_{1i}, \Delta_{1i}, \tilde{T}_{2i}, \Delta_{2i}) + o_P(n^{-1/2}),
\end{aligned}$$

uniformly in x, y, where the second equality follows from Lo and Singh (1986). Next, note that T_1 can be written as

$$\begin{aligned}
T_1 &= \int_0^x \Big\{\widehat{F}_e\Big(\frac{y \wedge T_t - \widehat{m}(t)}{\widehat{\sigma}(t)}\Big) - F_e\Big(\frac{y \wedge T_t - \widehat{m}(t)}{\widehat{\sigma}(t)}\Big)\Big\} dF_1(t) \\
&\quad + \int_0^x \Big\{F_e\Big(\frac{y \wedge T_t - \widehat{m}(t)}{\widehat{\sigma}(t)}\Big) - F_e\Big(\frac{y \wedge T_t - m(t)}{\sigma(t)}\Big)\Big\} dF_1(t) \\
&= (np_{1\tau})^{-1} \sum_{i=1}^n I(\Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) \int_0^x \varphi\Big(\tilde{T}_{1i}, \tilde{T}_{2i}, \Delta_{2i}, \frac{y \wedge T_t - \widehat{m}(t)}{\widehat{\sigma}(t)}\Big) dF_1(t) \\
&\quad - \int_0^x \frac{\widehat{m}(t) - m(t)}{\sigma(t)} f_e\Big(\frac{y \wedge T_t - m(t)}{\sigma(t)}\Big)\, dF_1(t) \\
&\quad - \int_0^x \frac{\widehat{\sigma}(t) - \sigma(t)}{\sigma(t)} \frac{y \wedge T_t - m(t)}{\sigma(t)} f_e\Big(\frac{y \wedge T_t - m(t)}{\sigma(t)}\Big)\, dF_1(t) + o_P(n^{-1/2}) \\
&= (np_{1\tau})^{-1} \sum_{i=1}^n I(\Delta_{1i} = 1, \tilde{T}_{1i} \le \tau_1) E\Big[\varphi\Big(\tilde{T}_{1i}, \tilde{T}_{2i}, \Delta_{2i}, \frac{y \wedge T_X - m(X)}{\sigma(X)}\Big) I(X \le x) \Big| \tilde{T}_{1i}, \tilde{T}_{2i}, \Delta_{2i}\Big] \\
&\quad + (np_1)^{-1} \sum_{i=1}^n I(\Delta_{1i} = 1) h_{\tilde{T}_{1i}, y \wedge T_{\tilde{T}_{1i}}}(\tilde{T}_{2i}, \Delta_{2i}) I(\tilde{T}_{1i} \le x) + o_P(n^{-1/2}) \\
&= n^{-1} \sum_{i=1}^n \sum_{j=1}^2 g^j_{x,y}(\tilde{T}_{1i}, \Delta_{1i}, \tilde{T}_{2i}, \Delta_{2i}) + o_P(n^{-1/2}),
\end{aligned}$$

where the second equality follows from Theorem 3.1, and the third one follows from Lemma B.1 in Van Keilegom and Akritas (1999) and from Proposition B.2. Finally, we consider T_2. Write $\widehat{F}_1(x) - F_1(x) = [\widehat{F}_1(x) - \tilde{F}_1(x)] + [\tilde{F}_1(x) - F_1(x)]$, where

$$\tilde{F}_1(x) = \int L\Big(\frac{x - t}{h}\Big) d\widehat{F}_1(t)$$

is a smoothed Kaplan-Meier estimator, with L a given continuously differentiable cumulative distribution function with compact support and $h = h_n = a_n^2$. Hence, T_2 can be written as

$$T_2 = \int_0^x A(t,y)d(\widehat{F}_1(t) - \tilde{F}_1(t)) + \int_0^x A(t,y)(\tilde{F}_1'(t) - F_1'(t))\,dt,$$

where $A(t,y) = \widehat{F}_e\Big(\frac{y \wedge T_t - \widehat{m}(t)}{\widehat{\sigma}(t)}\Big) - F_e\Big(\frac{y \wedge T_t - m(t)}{\sigma(t)}\Big)$. The second term above is $O_P((na_n)^{-1/2}$ $(nh_n)^{-1/2}\log n) = o_P(n^{-1/2})$ uniformly in x and y, since $na_nh_n(\log n)^{-2} = na_n^3(\log n)^{-2} \to \infty$. This follows from Theorem 3.2 and from the uniform consistency of the kernel density estimator $\tilde{F}_1'(t)$ (see e.g. Diehl and Stute, 1988). For the first term above note that

$$\widehat{F}_1(x) - F_1(x) = n^{-1}\sum_{i=1}^n \xi_1(\tilde{T}_{1i}, \Delta_{1i}, x) + o_P(n^{-1/2})$$

uniformly in $x \le \tau_1$ (see Lo and Singh (1986)). Moreover, Gijbels and Veraverbeke (1989) showed that the same asymptotic representation is also valid for $\tilde{F}_1(x) - F_1(x)$, provided the bandwidth h_n satisfies $nh_n^2 = na_n^4 \to 0$. Hence, $\sup_{x\le\tau_1}|\widehat{F}_1(x) - \tilde{F}_1(x)| = o_P(n^{-1/2})$. It is now easily seen, using integration by parts, that $\int_0^x A(t,y)d(\widehat{F}_1(t) - \tilde{F}_1(t)) = o_P(n^{-1/2})$. □

Proof of Theorem 3.4(i). First note that

$$\Big[1 - \widehat{F}_1(s)\Big]\,\widehat{p}_{12,\tau}(s,t) - [1 - F_1(s)]\,p_{12,\tau}(s,t) = n^{-1}\sum_{i=1}^n l_{s,t}(\widetilde{T}_{1i}, \Delta_{1i}, \widetilde{T}_{2i}, \Delta_{2i}) + R_n(s,t),$$

where

$$\sup\{|R_n(s,t)| : 0 \le s < t \le \tau_1\} = o_P(n^{-1/2}).$$

The proof is similar to that of the bivariate distribution case. By noting that

$$\begin{aligned}&\widehat{p}_{12,\tau}(s,t) - p_{12,\tau}(s,t)\\ &= \{1 - \widehat{F}_1(s)\}^{-1}\Big[(1 - \widehat{F}_1(s))\widehat{p}_{12,\tau}(s,t) - (1 - F_1(s))p_{12,\tau}(s,t) + p_{12,\tau}(s,t)(\widehat{F}_1(s) - F_1(s))\Big],\end{aligned}$$

the result follows from Theorem 3.3, the representation of $\widehat{F}_1(s)$ as a sum of i.i.d. terms, and the fact that

$$\frac{1}{1 - \widehat{F}_1(s)} = \frac{1}{1 - F_1(s)} + O_P(n^{-1/2})$$

uniformly on $[0, \tau_1]$. □

Proof of Theorem 3.5(i). This result can be established as Theorem 3.4, after noting that

$$
\begin{aligned}
&\widehat{p}_{22,\tau}(s,t) - p_{22,\tau}(s,t) \\
&= \Big\{ \int_0^s \Big[1 - \widehat{F}((s-r)\wedge T_r \mid r)\Big] \widehat{F}_1(dr) \Big\}^{-1} \\
&\quad \times \Big[\int_0^s \big[1 - \widehat{F}((t-r)\wedge T_r \mid r)\big] \widehat{F}_1(dr) - \int_0^s \big[1 - F((t-r)\wedge T_r \mid r)\big] F_1(dr) \\
&\quad - p_{22,\tau}(s,t) \Big\{ \int_0^s \big[1 - \widehat{F}((s-r)\wedge T_r \mid r)\big] \widehat{F}_1(dr) - \int_0^s \big[1 - F((s-r)\wedge T_r \mid r)\big] F_1(dr) \Big\} \Big].
\end{aligned}
$$

□

Acknowledgements. Ingrid Van Keilegom's research was financially supported by IAP research network P6/03 of the Belgian Government (Belgian Science Policy), and by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650. Jacobo de Uña-Álvarez acknowledges financial support from the project MTM2005-01274 (FEDER support included) of the Spanish Ministerio de Educación y Ciencia and also from the projects PGIDIT06PXIC300117PN and PGIDIT07PXIB300191PR of the Xunta de Galicia. Luis F. Meira-Machado acknowledges financial support by Grant PTDC/MAT/104879/2008 (FEDER support included) of the Portuguese Ministry of Science, Technology and Higher Education and also from the project MTM2008-01603 of the Spanish Ministerio de Educación y Ciencia.

References

Aalen, O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, **5**, 141-150.

Akritas, M.G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics*, **22**, 1299–1327.

Beran, R. (1981). *Nonparametric regression with randomly censored survival data.* Technical report, Univ. California, Berkeley.

de Uña-Álvarez, J. and Meira-Machado, L.F. (2008). A simple estimator of the bivariate distribution function for censored gap times. *Statistics & Probability Letters*, **78**, 2440–2445.

Diehl, S. and Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring. *Journal of Multivariate Analysis*, **25**, 299–310.

Einmahl, J. and Van Keilegom, I. (2008a). Specification tests in nonparametric regression. *J. Econometrics*, **143**, 88-102.

Einmahl, J. and Van Keilegom, I. (2008b). Tests for independence in nonparametric regression. *Statist. Sinica*, **18**, 601-616.

Gijbels, I. and Veraverbeke, N. (1989). Smooth estimation of a distribution function and its quantiles based on a censored sample. *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics (Prague, 1988)*, 239–248, Charles Univ., Prague.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New-York.

Lin, D. Y., Sun, W. and Ying, Z. (1999). Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika*, **86**, 59-70.

Lo, S.-H. and Singh, K. (1986). The product-limit estimator and the bootstrap: some asymptotic representations. *Probability Theory and Related Fields*, **71**, 455-465.

Meira-Machado, L.F., de Uña-Álvarez, J. and Cadarso-Suárez, C. (2006). Nonparametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis*, **12**, 325-344.

Neumeyer, N. (2009). Testing independence in nonparametric regression. *J. Multivariate Anal.*, **100**, 1551-1566.

Schaubel, D. E. and Cai, J. (2004). Non-parametric estimation of gap-time survival functions for ordered multivariate failure time data. *Statistics in Medicine*, **23**, 1885-1900.

Van Keilegom, I. (2004). A note on the nonparametric estimation of the bivariate distribution under dependent censoring. *Journal of Nonparametric Statistics*, **16**, 659-670.

Van Keilegom, I. and Akritas, M. (1999). Transfer of tail information in censored regression models. *Annals of Statistics*, **27**, 1745-1784.

Van Keilegom, I., Akritas, M. and Veraverbeke, N. (2001). Estimation of the conditional distribution in regression with censored data: a comparative study. *Computational Statistics and Data Analysis*, **35**, 487-500.

Wang, W. and Wells, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika*, **85**, 561-572.