

CORE DISCUSSION PAPER 2006/118

Density and Hazard Rate Estimation for Censored and  $\alpha$ -mixing  
Data Using Gamma Kernels

Taoufik Bouezmarni\*      Jeroen V.K. Rombouts†

December 12, 2006

**Abstract**

In this paper we consider the nonparametric estimation for a density and hazard rate function for right censored  $\alpha$ -mixing survival time data using kernel smoothing techniques. Since survival times are positive with potentially a high concentration at zero, one has to take into account the bias problems when the functions are estimated in the boundary region. In this paper, gamma kernel estimators of the density and the hazard rate function are proposed. The estimators use adaptive weights depending on the point in which we estimate the function, and they are robust to the boundary bias problem. For both estimators, the mean squared error properties, including the rate of convergence, the almost sure consistency and the asymptotic normality are investigated. The results of a simulation demonstrate the excellent performance of the proposed estimators.

*Key words and phrases.* Gamma kernel, Kaplan Meier, density and hazard function, mean integrated squared error, consistency, asymptotic normality.

We would like to thank Song Chen, Alain Guay for useful comments.

---

\*HEC Montréal, CREF, Institute of Statistics (Université catholique de Louvain).

†Institute of Applied Economics at HEC Montréal, CIRANO, CIRPEE, CORE (Université catholique de Louvain), CREF, 3000 Cote Sainte Catherine, H3T 2A7 Montréal (QC), Canada.

# 1 Introduction

Survival times from clinical trials are often dependent and censored due to the nature of the experiment. To ensure consistent estimation, this has to be taken into account when the marginal density of the survival data is fitted nonparametrically. An additional problem arises when there is a high concentration of durations near zero. See for example Jones (1993) for some solutions to this boundary problem in the context of uncensored and independent data. This paper makes use of the gamma kernel to estimate nonparametrically the marginal density and the hazard function of positive dependent survival data that suffer from right censoring. Both estimators are robust with respect to boundary problems.

We start by introducing some notations and by reviewing the relevant literature. Let  $T_1, \dots, T_n$  (survival times) and  $C_1, \dots, C_n$  (censoring times) be two nonnegative random sequences with distribution functions  $F$  and  $G$ , respectively. We assume that the censoring times  $C_i$  are i.i.d. and independent of the survival times  $T_i$ . We consider right censoring, that is instead of observing  $T_i$ , we observe the pair  $(X_i, \delta_i)$ , where  $X_i = \min(T_i, C_i)$  and  $\delta_i = I(T_i \leq C_i)$  where  $I(\cdot)$  is the indicator function. Furthermore, we suppose that the survival times  $T_i$  are  $\alpha$ -mixing. Recall the definition of  $\alpha$ -mixing. Let  $\mathcal{F}_i^k(T)$  be the  $\sigma$ -field of events generated by  $\{T_j, i \leq j \leq k\}$ .

**Definition 1.** Let  $\{T_i, i \geq 1\}$  be a sequence of random variable. Given a positive integer  $n$ , set

$$\alpha(n) = \sup_k |P(A \cap B) - P(A)P(B)|, A \in \mathcal{F}_1^k(T) \quad \text{and} \quad B \in \mathcal{F}_{k+n}(T).$$

The sequence is  $\alpha$ -mixing if the mixing coefficient  $\alpha(n) \rightarrow 0$  as  $n \rightarrow \infty$ .

The  $\alpha$ -mixing condition, also called strong mixing, is the weakest among mixing conditions known in the literature. Many stochastic process satisfy the  $\alpha$ -mixing condition, see for example Doukhan (1994).

In the case of censoring, it is well known that the empirical distribution is not a consistent estimator for the distribution function  $F$ . Therefore, Kaplan and Meier (1958) proposed a consistent estimator, hereafter denoted the KM estimator, for the distribution function  $F$  which is defined as follows:

$$\Gamma_n(x) = \begin{cases} 1 - \prod_{i:1 \leq i \leq n, X_{(i)} \leq x} \left( \frac{n-i}{n-i+1} \right)^{\delta_{(i)}} & \text{if } x < X_{(n)} \\ 1 & \text{Otherwise.} \end{cases}$$

$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  are the order statistics of  $X_1, X_2, \dots, X_n$  and  $\delta_{(i)}$  is the concomitant of  $X_{(i)}$ . For independent survival times Breslow and Crowley (1974) state the weak convergence of the KM estimator. Lo and Singh (1985) expressed the KM estimator as an i.i.d. mean process with a remainder of negligible order. Wang (1987) established the uniform weak convergence of KM estimator Stute and Wang (1993) proved the convergence in mean and almost surely of  $\int \varphi d\Gamma_n$  for any function  $\varphi$  that is  $F$ -integrable.

When the survival times are dependent, the behavior of the KM estimator has been studied by many authors. In fact, the consistency, the asymptotic normality and the limiting variance of the KM estimator are developed by Ying and Wei (1994) for  $\phi$ -mixing survival time, Cai and Roussas (1998) when the survival time exhibit some particular dependence mode, called positive or negative association, Leonenko and Sakhno (2001) for long-range dependent survival and censoring times and, Cai (2001) for  $\alpha$ -mixing survival and censoring times. Cai (1998a) represents the KM estimator as the mean of random variables with a remainder of order  $O(n^{-1/2}(\log n)^\lambda)(\lambda > 0)$ , and states the asymptotic normality in  $\alpha$ -mixing context. In this paper, this approximation is used to derive some properties for the gamma kernel density and the hazard rate function estimators.

This paper concentrates on the estimation of the density and the hazard function. Based on the smooth kernel technique, the standard kernel estimator for the density function  $f$  is defined as follows:

$$\hat{f}_a(x) = \frac{1}{a} \int K\left(\frac{x-t}{a}\right) d\Gamma_n(t) = \frac{1}{a} \sum_{i=1}^n K\left(\frac{x-X_i}{a}\right) \omega_i \quad (1)$$

and the hazard rate estimator is:

$$\hat{h}_a(x) = \frac{1}{a} \sum_{i=1}^n K\left(\frac{x-X_{(i)}}{a}\right) q_i. \quad (2)$$

where  $a$  is the bandwidth parameter depending on the sample size,  $K$  is a symmetric density function, the weight  $\omega_1 = \Gamma_n(X_{(1)})$  and  $\omega_i = \Gamma_n(X_{(i)}) - \Gamma_n(X_{(i-1)})$  for  $i = 2, \dots, n$  and  $q_i = \frac{\delta_{(i)}}{n-i+1}$ . For simplicity we omit the support of the integral which is  $[0, \tau]$ , where  $\tau = \sup\{t \geq 0, H(t) < 1\}$

and  $H$  is the distribution function of  $X_i$  which is given by:

$$H(t) = 1 - (1 - F(t))(1 - G(t)), \quad \text{for } t \geq 0.$$

The consistency of these estimators is well documented for independent and dependent survival data. When the survival data are i.i.d., Tanner and Wong (1983) derive the mean and the variance and establish the asymptotic normality of the standard kernel estimator for the hazard rate function. Mielniczuk (1986) studies the consistency of the kernel estimator for the density function. Lo, Mack, and Wang (1989) derive the mean integrated squared error, consistency and asymptotic normality of the standard kernel estimator for the density and the hazard rate functions. When the survival times are  $\alpha$ -mixing, Cai (1998b) derives the consistency including its rate and the asymptotic normality of the standard kernel estimator of the density and the hazard rate function. Liebscher (2002) improves the rate of convergence in Cai (1998b) for the standard kernel estimators.

The above results ignore the boundary bias problem. Several techniques are proposed to remove this problem in the uncensored case. For example Schuster (1985) develops the reflection method. Müller (1991), Lejeune and Sarda (1992), Jones (1993) and Jones and Foster (1996) use boundary kernels in the boundary region and the standard kernel away from the boundary. Marron and Ruppert (1994) propose to transform the data before using the standard kernel, Chen (2000) and Scaillet (2004) consider asymmetric kernels. These methods can be adapted to the censored data case, which is the subject of this paper.

The rest of the paper is organized as follows. In Section 2, we present the gamma kernel estimator for the density and hazard rate functions. The mean integrated square error, the asymptotic normality and the strong consistency are proved for the two estimators in Section 3. Section 4 provides Monte Carlo results concerning the finite sample properties of the estimators for various distributions and parameter values. Section 5 concludes. The proofs are given in the appendix.

## 2 Gamma kernel estimator

For positive data, a natural way to overcome the boundary bias problem when estimating a density nonparametrically is to consider kernels with positive support. For instance, this idea is developed for i.i.d. data by Chen (2000) which proposes a gamma kernel and Scaillet (2004) which proposes the inverse gaussian and reciprocal inverse gaussian kernel. A crucial aspect if we want to use the

gamma density as a kernel is the choice of its parameters. A simple choice is a gamma density with shape-parameter  $x/b + 1$  and scale-parameter  $b$ , for some positive  $b$  and with  $x$  the position where the density or hazard rate function is evaluated. With these parameters  $x$  is the mode, though not the mean, of the kernel. To improve the properties of the estimator, Chen (2000) proposes to take as a kernel, a gamma density with scale  $b$  and shape

$$\rho_b(x) = \begin{cases} x/b & \text{if } x \geq 2b \\ \frac{1}{4}(x/b)^2 + 1 & \text{if } x \in [0, 2b). \end{cases} \quad (3)$$

Note that outside the boundary region  $x$  is the mean of the gamma kernel and its mode is  $x - b$ . The shape  $x/b$  can not be used inside the boundary region since the gamma kernel becomes unbounded. This is the reason why the shape  $\frac{1}{4}(x/b)^2 + 1$  is used here. Other choices for the shape in the boundary region may be investigated. In this paper, we consider the gamma kernel, so that  $K$  is defined by

$$K(x, b)(t) = \frac{t^{\rho(x)-1} \exp(-t/b)}{b^{\rho(x)} \Gamma(\rho(x))}. \quad (4)$$

The shape of the gamma kernel and the amount of smoothing vary according to the position where the density is estimated. In fact, the further we move away from the boundary the more symmetric the kernel becomes. This is a crucial difference with the Gaussian or other fixed symmetric kernel estimators. The support of the gamma kernel matches the support of the probability density function to be estimated, and therefore no weight is lost when the density is estimated at the boundary region. For uncensored and independent data, Chen (2000) shows that the gamma kernel density estimator is free of boundary bias, always non-negative and achieves the optimal rate of convergence for the mean integrated squared error within the class of non-negative kernel density estimators. Furthermore, the variance reduces as the position where the smoothing is made moves away from the boundary. Bouezmarni and Scaillet (2003) state the uniform weak consistency for the gamma kernel estimator on each compact set in the positive real line when  $f$  is continuous on its support and the weak convergence in terms of mean integrated absolute error. For unbounded densities at the origin, they prove that it converges in probability to infinity at the origin where the density is unbounded. For positive uncensored and  $\alpha$ -mixing data, Bouezmarni and Rombouts (2006) derive the mean integrated squared error, almost sure convergence and asymptotic normality of the gamma kernel estimator for the density function.

For censored and dependent data, we propose the gamma kernel estimator for the density defined as:

$$\hat{f}_b(x) = \int K(x, b)(t) d\Gamma_n(t) = \sum_{i=1}^n K(x, b)(X_i)\omega_i \quad (5)$$

and the gamma kernel estimator for the hazard rate function

$$\hat{h}_b(x) = \int K(x, b)(t) d\tilde{\Delta}_n(t) = \sum_{i=1}^n K(x, b)(X_{(i)}) q_i \quad (6)$$

where  $b$  is the bandwidth parameter depending on  $n$  and the weights  $\omega_i$  and  $q_i$  defined in the introduction. Figure 1 illustrates the gamma kernel and the standard kernel estimators for the density and hazard rate functions. The first data serie(the two upper panels) considers exponential dependent survival times with uniform censoring times. The second dataset (the two lower panels) are weibull dependent survival times with weibull censoring times. The degree of censoring is 25% and the sample size is 250. We come back to these two models in Section 4 where we study the finite sample properties of the estimators. To compare, we also added the standard kernel estimator on each graph. We observe that for this standard kernel estimator there are severe problems at the boundary for both the density and the hazard function.

### 3 Convergence properties

In this section, we show the asymptotic properties of the gamma kernel estimator for both the density and hazard rate function when the data are right censored and  $\alpha$ -mixing, which satisfy the following conditions.

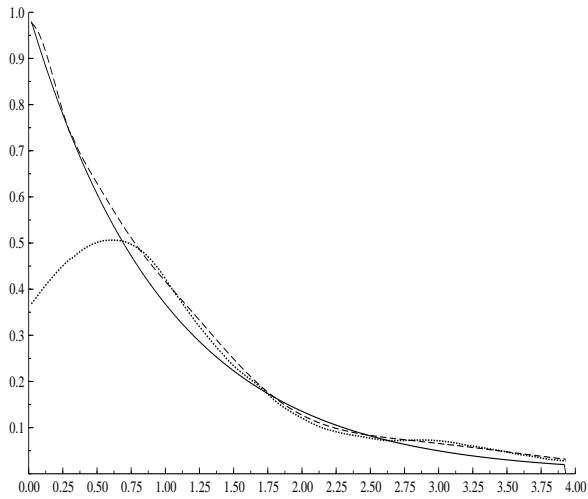
**Condition 1.**

(C1) *The survival time  $\{T_j; j \geq 1\}$  is a stationary  $\alpha$ -mixing sequence of random variables.*

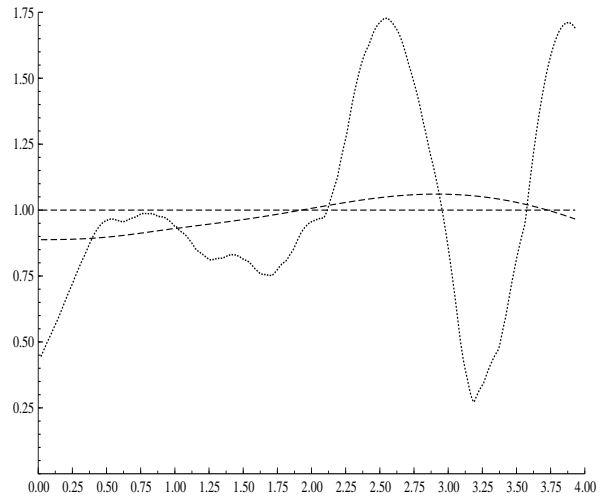
(C2) *The censoring time  $\{C_j; j \geq 1\}$  is a i.i.d. random variable and independent of  $\{T_j; j \geq 1\}$ .*

(C3)  *$\alpha(n) = O(n^{-\beta})$  for some  $\beta > 3$ .*

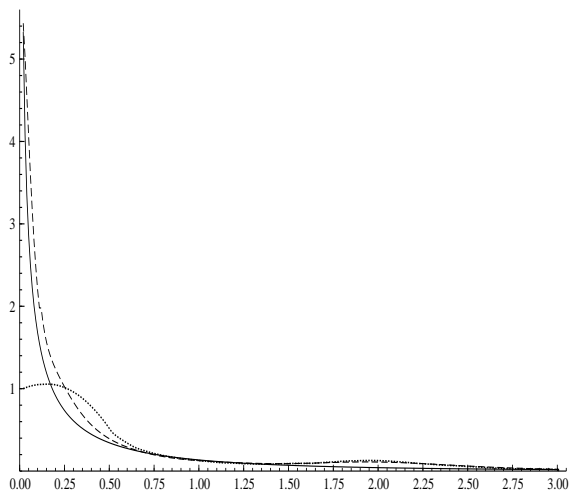
Condition C2 can be relaxed to be also a stationary  $\alpha$ -mixing sequence. To be concise, we denote by  $L$  either the density function  $f$  or the hazard rate function  $h$ .  $\hat{L}_b$  either the gamma density estimator  $\hat{f}_b$  or the gamma hazard rate estimator  $\hat{h}_b$ . The following Theorem deals with the mean integrated square error of the two estimators.



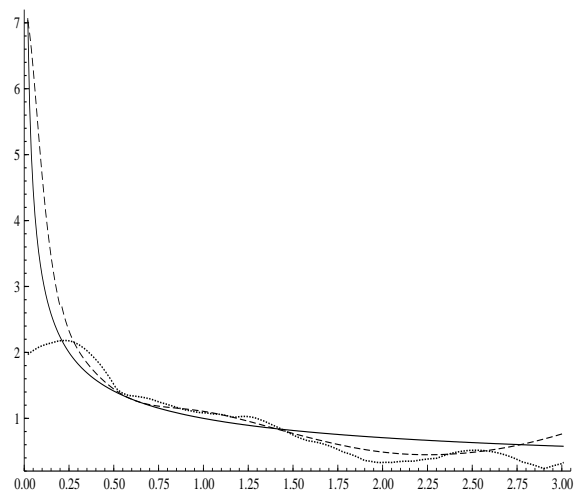
(a) Exponential density



(b) Exponential hazard rate



(c) Weibull density



(d) Weibull hazard rate

Figure 1: True density and hazard functions (solid line), gamma kernel density and hazard rate estimates (dashed line) and the standard kernel density and hazard rate estimates (dotted line).

**Theorem 1. (mean integrated square error of  $\hat{L}_b$ )**

$L$  is twice continuously differentiable. Assume that  $b = O(n^{-2/5})$ . Then, under condition 1, we have

$$E \left( \int (\hat{L}_b(x) - L(x))^2 dx \right) = b^2 B + n^{-1} b^{-1/2} V + o(b) + o(n^{-1} b^{-1/2}) \quad (7)$$

where

$$B = \frac{1}{4} \int x^2 L''^2(x) dx \quad \text{and} \quad V = \frac{1}{2\sqrt{\pi}} \int \frac{x^{-1/2} L(x)}{1 - V_1(x)} dx \quad (8)$$

and  $V_1(x) = G(x)$  for the density function and  $V_1(x) = H(x)$  for the hazard rate function. The optimal bandwidth which minimizes the asymptotic mean integrated squared error is given by

$$b^* = \left( \frac{V}{4B} \right)^{2/5} n^{-2/5}. \quad (9)$$

This leads to the optimal mean integrated squared error:

$$E \left( \int (\hat{L}_b(x) - L(x))^2 dx \right) \sim \frac{5}{4^{4/5}} V^{4/5} B^{1/5} n^{-4/5}. \quad (10)$$

■

Theorem 1 states that the gamma kernel estimators are free of boundary bias and attains the optimal rate of convergence in term of mean integrated squared error as in the uncensored case. The theorem provides the optimal theoretical choice of the bandwidth parameter. In practice, we can replace  $f$  in formula (9) by the gamma density with estimated parameters.

The asymptotic distribution of the estimators is needed for goodness of fit tests and for confidence intervals. In the next theorem we state the asymptotic normality of the gamma kernel estimators.

**Proposition 1. (asymptotic normality of  $\hat{L}_b$ )**

Under the conditions of Theorem 1, for all  $x$  such that  $f(x) > 0$ , we have for all  $x \leq \tau$

$$\left( n^{1/2} b^{1/4} \frac{\hat{L}_b(x) - E(\hat{L}_b(x))}{\sqrt{V^*(x)}} \right) \rightarrow N(0, 1) \quad (11)$$

where

$$V^*(x) = \begin{cases} \frac{1}{2\sqrt{\pi}} \frac{x^{-1/2} L(x)}{1 - V_1(x)} & \text{if } x/b \rightarrow \infty \\ \frac{\Gamma(2\kappa+1) b^{-1/2}}{2^{1+2\kappa} \Gamma^2(\kappa+1)} \frac{L(x)}{1 - V_1(x)} & \text{if } x/b \rightarrow \kappa \end{cases} \quad (12)$$

■

The results in Proposition 1 can not be used for goodness of fit or to construct the confidence intervals since the variances in the denominator depend on unknown functions  $f$  and  $h$ . However we can replace these functions by their estimates knowing that they almost surely convergence. The following proposition states the almost sure convergence of the gamma kernel estimator for the density function and the hazard rate function.

**Proposition 2. (almost sure convergence of  $\hat{L}_b$ )**

Let  $f$  be a continuous density. Assume that the condition 1 is satisfied and  $b = O(n^{-2/5})$ . Then, for all  $x \leq \tau$  we have

$$\hat{L}_b(x) \longrightarrow L(x), \quad \text{almost surely.}$$

■

## 4 Finite sample properties

This section studies the finite sample performance of the gamma kernel estimator for the density and hazard rate function. We consider a simple underlying dynamic model for the survival times given by

$$y_t = \phi y_{t-1} + \epsilon_t, \quad t = 1, \dots, n \tag{13}$$

$\epsilon_t \sim N(0, 1)$  and  $n = 125, 250, 500, 1000$ . The survival time is given by  $T_i = F^{-1}(\Phi(y_t/\sigma))$  where  $\sigma^2 = 1/(1-\phi^2)$ . For exponential survival time, model A hereafter, the censoring times are generated from a uniform distribution  $[0, c]$  with  $c$  such that  $pc + \exp(-c) = 1$ . For Weibull survival time with scale parameter  $\gamma = 2$  and shape parameter  $\alpha = 0.25$ , model B hereafter, the censoring times are generated from weibull distribution with shape parameter  $\gamma^* = 2$  and scale parameter  $\alpha^* = \left(\frac{1-p}{2p}\right)^2$ . Note that the choice of the parameters of the censoring distribution is to ensure the degree of censoring to be equal to  $p$ . In the simulations, we fix  $\phi = 0.85$  which obviously ensures a stationary and  $\alpha$ -mixing  $T_i$  sequence and several values of  $p$  values. For a graphical representation of the two models we refer to Figure 1 in Section 2. Note that for model B, the cumulative distribution of the censoring times increases when the degree of censoring  $p$  increases. This implies, from equation (10), that the mean integrated squared error of the gamma kernel estimators increases.

To evaluate the performance of the gamma kernel estimators we compare with the local linear estimator of Jones (1993) adapted for the right censoring dependent case. The local linear estimators are defined as in (1) and (2) with the kernel  $K_l$  defined in the following way. For non-negative integer  $s$  and Epanechnikov kernel  $K$ , define

$$K_l(x, h, t) = \frac{a_2(x, h) - a_1(x, h)t}{a_0(x, h)a_2(x, h) - a_1^2(x, h)}K(t) \quad (14)$$

where

$$a_s(x, h) = \int_{-1}^{x/h} t^s K(t) dt. \quad (15)$$

Note that  $K_l(x, h, t) = K(t)$  for  $x > h$ . This type of kernel does not guarantee positive density and hazard estimates in the boundary region when the underlying functions are small in this region. Thanks to the high data concentration in the boundary region for the above two data generating processes this drawback is avoided.

We measure the performance of the estimators by analysing the mean and the standard error of the  $L_2 = \int_0^\tau (\hat{L} - L)^2$  norm, where  $L$  denotes either the density or the hazard rate function. Note that  $\tau$  is fixed for the Weibull model and changes according to the censoring degree for the exponential model, as explained in the introduction. The selected bandwidth parameters in each replication minimize the  $L_2$  error.

Table 1 considers the mean of  $L_2$  errors based on 1000 replications related to the estimation of the density function. We draw attention to the four following points. Firstly, when the sample size increases, the mean integrated square error error for the two estimators decreases. This is true for both models and all degrees of censoring. For example in model A, 15% of censoring and the gamma kernel estimator, the mean error decreases from 0.021 to 0.016 when the sample size goes from 125 to 250. For 50% of censoring, this decrease is only minor, that is from 0.033 to 0.031. In fact, for the 50% of censoring the rate at which the mean error decreases is much smaller. Secondly, Except for model A and 50% of censoring, the gamma kernel estimator outperforms strongly the local linear kernel estimator. This excellent performance does not fade away when the sample size increases. Thirdly, For model B, when the degree of censoring increases the mean integrated square error increases as expected. Note that this comparison can not be made for model A, since the support of the integral varies with the degree of censoring. Lastly, Due to the unboundedness of the density function the estimation of the density becomes more complex for model B compared

to model A, and therefore the mean error values for the former model are higher. Though, this is also due to the fact that the performance measures is not relative.

Table 2 presents the standard deviations of the  $L_2$  errors for the estimation of the density function. As expected, for both models and both estimators the variance decreases with the sample size, no matter the degree of censoring. Generally speaking for model B, the variance increases with the degree of censoring but at a lower rate for higher sample sizes. Another point to remark is that for model A, the variance of the gamma kernel estimator is smaller than the variance of the local linear estimator for  $p=0.1, 0.15$  and  $0.25$ , but not for  $p=0.5$ . However, for model B the local linear estimator is dominated in terms of variance by gamma kernel estimator in all situations.

Table 3 considers the mean of  $L_2$  errors based on 1000 replications related to the estimation of the hazard function. The results are similar to those of the density function except for two points. Firstly, the gamma kernel estimator is better in terms of mean integrated squared error than the local linear even for model A with 50% degree of censoring. Secondly, we observe in a few cases a small decrease of mean integrated squared error when the degree of censoring increases. However given their variance, this decrease is insignificant.

Table 4 reports the the standard deviations of the  $L_2$  errors for the estimation of the hazard function. For model B, the results are similar as for the density function. For model A, neither of the two estimators dominates.

## 5 Conclusion

This paper makes use of the gamma kernel to estimate nonparametrically the marginal density and the hazard function of positive dependent survival data that are subject to right censoring. Both estimators are robust with respect to boundary problems. We derive the mean integrated square error, the asymptotic normality and the strong consistency for the two estimators. The Monte Carlo results concerning the finite sample properties of the estimators for various distributions and parameter values show the excellent performance of the gamma kernel estimators.

Table 1: Mean of  $L_2$  error for the density function estimators.

		$n = 125$		$n = 250$		$n = 500$		$n = 1000$	
Model	% cens	G	LL	G	LL	G	LL	G	LL
A	10	0.0232	0.0367	0.0200	0.0337	0.0195	0.0354	0.0189	0.0338
	15	0.0206	0.0276	0.0161	0.0240	0.0135	0.0206	0.0129	0.0201
	25	0.0207	0.0235	0.0159	0.0207	0.0139	0.0196	0.0126	0.0182
	50	0.0325	0.0207	0.0312	0.0207	0.0311	0.0205	0.0313	0.0191
B	10	0.6244	0.9267	0.4864	0.7701	0.4287	0.7015	0.3729	0.6088
	15	0.6325	0.9653	0.5267	0.8575	0.4972	0.8216	0.4564	0.7418
	25	0.7362	1.2079	0.6624	1.1211	0.6505	1.0607	0.6397	0.9623
	50	1.4847	2.4885	1.4734	2.4249	1.4375	2.4192	1.4345	2.4108

G: gamma estimator. LL: local linear estimator. Model A: exponential survival times with uniform censoring times. Model B: Weibull survival times with Weibull censoring times. The results are based on 1000 replications.

Table 2: Standard deviation ( $\times 10^{-2}$ ) of  $L_2$  error for the density function estimators.

		$n = 125$		$n = 250$		$n = 500$		$n = 1000$	
Model	% cens	G	LL	G	LL	G	LL	G	LL
A	10	0.8632	1.2417	0.6567	1.0479	0.4928	0.8287	0.40343	0.5769
	15	1.0671	1.3481	0.7302	1.2045	0.5717	1.0251	0.49302	0.9489
	25	1.2461	1.2163	0.8024	1.0710	0.5804	1.0084	0.48691	0.9281
	50	0.9808	0.2543	0.8592	0.2273	0.7848	0.2018	0.77904	0.1038
B	10	37.040	64.085	31.545	55.920	27.511	46.876	21.890	36.691
	15	36.852	63.788	32.338	55.812	29.253	49.038	23.316	39.119
	25	41.464	68.596	34.847	56.553	31.519	49.160	26.343	43.567
	50	47.238	80.309	32.229	61.358	27.544	50.776	23.381	43.883

G: gamma estimator. LL: local linear estimator. Model A: exponential survival times with uniform censoring times. Model B: Weibull survival times with Weibull censoring times. The results are based on 1000 replications.

Table 3: Mean of  $L_2$  error for the hazard rate function estimators.

		$n = 125$		$n = 250$		$n = 500$		$n = 1000$	
Model	% cens	G	LL	G	LL	G	LL	G	LL
A	10	0.2259	0.3312	0.1427	0.2571	0.0924	0.2056	0.0547	0.1635
	15	0.2746	0.3662	0.1558	0.2509	0.0984	0.1786	0.0550	0.1233
	25	0.1533	0.2071	0.0716	0.1236	0.0400	0.0934	0.0192	0.0731
	50	0.0242	0.0413	0.0088	0.0243	0.0040	0.0184	0.0015	0.0149
B	10	2.1053	3.5894	1.4103	2.5668	0.8283	1.6377	0.5044	1.0716
	15	2.0669	3.5694	1.3631	2.5854	0.8349	1.7069	0.5435	1.1670
	25	2.0670	3.6824	1.4007	2.7779	0.9066	1.9292	0.6104	1.3632
	50	2.3336	3.7729	1.8069	3.3357	1.4996	3.0553	1.1281	2.5462

G: gamma estimator. LL: local linear estimator. Model A: exponential survival times with uniform censoring times. Model B: Weibull survival times with Weibull censoring times. The results are based on 1000 replications.

Table 4: Standard deviation of  $L_2$  error for the hazard rate function estimators.

		$n = 125$		$n = 250$		$n = 500$		$n = 1000$	
Model	% cens	G	LL	G	LL	G	LL	G	LL
A	10	0.1408	0.1082	0.1047	0.0887	0.0764	0.0731	0.0491	0.0541
	15	0.2057	0.1673	0.1364	0.1259	0.1010	0.1106	0.0706	0.0829
	25	0.1959	0.1748	0.1094	0.1061	0.0667	0.0704	0.0348	0.0398
	50	0.0517	0.0507	0.0265	0.0274	0.0129	0.0153	0.0056	0.0096
B	10	1.408	1.9243	1.0770	1.6290	0.5891	0.9443	0.3147	0.4828
	15	1.3112	1.8304	0.9431	1.5248	0.5679	0.9390	0.3478	0.5296
	25	1.2356	1.6136	0.8459	1.4410	0.5409	0.9406	0.3821	0.6097
	50	0.9440	1.1802	0.6833	1.0439	0.6357	1.1174	0.5383	1.0419

G: gamma estimator. LL: local linear estimator. Model A: exponential survival times with uniform censoring times. Model B: Weibull survival times with Weibull censoring times. The results are based on 1000 replications.

## Appendix

We start with some notations which will be useful for the proofs in this section.

Let

$$g(t) = \int_0^t (1 - H(s))^{-2} dH^*(s),$$

where

$$H^*(t) = P(X_1 \leq t, \delta_i = 1) = \int_0^t (1 - G(s)) dF(s)$$

which is the distribution of the uncensored observations.

For positive real numbers  $z$  and  $x$ , and  $\delta = 0$  or  $1$ , let

$$\xi(z, \delta, t) = -g(z \wedge t) + (1 - H)^{-1} I(z \leq t, \delta = 1)$$

Note that

$$E(\xi(X_i, \delta, t)) = 0$$

and

$$\text{Cov}(\xi(X_i, \delta, t), \xi(X_i, \delta, s)) = g(s \wedge t).$$

The following lemma plays an important role in the demonstrations.

**Lemma 1.** *Cai (1998a)* Under condition 1, for  $a_n = (\log \log n/n)^{1/2}$  and  $t \leq \tau$

$$\sup_{t \geq 0} |\bar{Y}_n(t) - \bar{H}(t)| = O(a_n) \quad (16)$$

and

$$\sup_{t \geq 0} |\bar{H}_n^*(t) - H^*(t)| \stackrel{a.s.}{=} O(a_n) \quad a.s. \quad (17)$$

where  $\bar{Y}_n(t) = \frac{1}{n} \sum_{i=1}^n I(X_i \geq t)$  and  $H_n^*$  is the empirical estimator of  $H^*$ , namely

$$H_n^*(t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t, \delta_i = 1).$$

Also

$$\Gamma_n(t) - F(t) = \bar{F}(t)(\tilde{\Delta}_n(t) - \Delta(t)) + O(a_n^2), \quad (18)$$

and

$$\tilde{\Delta}_n(t) - \Delta(t) = \frac{1}{n} \sum_{i=1}^n \xi_i(X_i, \delta_i, t) + U_n(t) + O(a_n^2), \quad (19)$$

where

$$U_n(t) = \int_0^t \left( \frac{1}{\bar{Y}_n(s)} - \frac{1}{\bar{H}(s)} d(H_n^*(s) - H^*(s)) \right).$$

■

### Proof of Theorem 1

We start with the expectation of the gamma kernel estimator. We suppose that the bandwidth parameter is small enough. From Chen (2000), we can show that for  $x > 2b$

$$K(\rho(x), b)(t) \equiv K_{x,b}(t) \leq \frac{1}{\sqrt{2\pi}}(x-b)^{-1/2}b^{-1/2}. \quad (20)$$

Also

$$\int dK_{x,b}(t) = O(b^{-1}). \quad (21)$$

We establish the bias for  $x > 2b$ . Using inequalities (20) and (18) and integration by parts

$$\begin{aligned} E(\hat{f}_b(x)) &= E\left(\int K_{x,b}(t)d\Gamma_n(t)\right) \\ &= -E\left(\int \Gamma_n(t)dK_{x,b}(t)\right) \\ &= -\int FdK_{x,b}(t) + E\left(\int \bar{F}(t)(\tilde{\Delta}_n(t) - \Delta_n(t))dK_{x,b}(t)\right) + O(a_n^2b^{-1/2}) \\ &= \int K_{x,b}(t)f(t)dt + E\left(\int \bar{F}(t)(\tilde{\Delta}_n(t) - \Delta_n(t))dK_{x,b}(t)\right) + O(a_n^2b^{-1/2}) \\ &= \int K_{x,b}(t)f(t)dt + E\left(\int \bar{F}(t)U_n(t)dK_{x,b}(t)\right) + O(a_n^2b^{-1/2}) \\ &= f(x) + I + II + O(a_n^2b^{-1/2}) \end{aligned} \quad (22)$$

From Chen (2000),

$$I \equiv \int K_{x,b}(t)f(t)dt - f(x) \begin{cases} \xi_b(x)f'(x)b + o(b) & \text{if } x \in [0, 2b) \\ \frac{1}{2}xf''(x)b + o(b) & \text{if } x \geq 2b. \end{cases} \quad (23)$$

where  $\xi_b(x) = (1-x)\{\rho(x) - x/b\}/\{1 + b\rho(x) - x\}$ . Now we show that  $II = O(a_n^2b^{-1/2})$ .

In fact,

$$\begin{aligned} \int \bar{F}(t)U_n(t)dK_{x,b}(t) &= \int f(t)U_n(t)K_{x,b}(t) - \int \bar{F}(t)K_{x,b}(t)dU_n(t) \\ &= II' + II'' \end{aligned}$$

On the one hand From Cai (1998b),  $|U_n(t)| \stackrel{a.s.}{=} O(n^{-1/2}(\log(n))^{-\lambda})$  for some  $\lambda > 0$ . Therefore

$$II' \stackrel{a.s.}{=} O(n^{-1/2}(\log(n))^{-\lambda}) = o(n^{-1/2}b^{-1/4})$$

On the another hand, using (16), (17) and (20)

$$\begin{aligned} II'' &= \int \bar{F}(t)K_{x,b}(t) \left( \frac{\bar{H}(s) - \bar{Y}_n(s)}{\bar{Y}_n(s)\bar{H}(s)} d(H_n^*(s) - H^*(s)) \right) \\ &\stackrel{a.s.}{=} O(a_n^2 b^{-1/2}). \end{aligned}$$

Using second condition on the bandwidth parameter we have  $O(a_n^2 b^{-1/2}) = o(n^{-1/2} b^{-1/4})$ . Now, for the boundary region  $x \leq 2b$ , we follow the same arguments but now instead of using inequality (20), we use inequality (21). Hence, the bias of the gamma kernel estimator is

$$E(\hat{f}_b(x) - f(x)) = \begin{cases} \xi_b(x)f'(x)b + o(b) + o(n^{-1/2}b^{-1/2}) & \text{if } x \in [0, 2b) \\ \frac{1}{2}xf''(x)b + o(b) + o(n^{-1/2}b^{-1/4}) & \text{if } x \geq 2b. \end{cases} \quad (24)$$

The integrated squared bias is

$$\int (E(\hat{f}_b(x) - f(x)))^2 = b^2 \int \frac{1}{4}(xf''(x))^2 dx + o(b^2) + o(n^{-1}b^{-1/2}). \quad (25)$$

Next calculate the variance of the gamma kernel estimator. Using equations (19) and (18), the gamma kernel estimator for density function can rewrite as follows:

$$\hat{f}_b(x) = \beta(x) + \sigma_n(x) + r_n(x) + O({}_n^2 b^{-1/2}),$$

where

$$\begin{aligned} \beta(x) &= \int fK_{x,b}(t)dt, \\ \sigma_n(x) &= \frac{1}{n} \sum_{i=1}^n \int \bar{F}(t)\xi_i(X_i, \delta_i, t)dK_{x,b}(t) \end{aligned}$$

and

$$r_n(x) = - \left( \int \bar{F}(t)U_n(t)dK_{x,b}(t) \right).$$

We start with the variance of  $\sigma_n(x)$ , since the others terms are negligible of the variance of  $\hat{f}_b(x)$ .

$$\begin{aligned} Var(\sigma_n(x)) &= n^{-2} \sum_{i=1}^n \int \int \bar{F}(t)\bar{F}(s)cov(\xi_i(X_i, \delta_i, t), \xi_i(X_i, \delta_i, s))dk(t)dk(s) \\ &= n^{-1} \int \int \bar{F}(s)\bar{F}(t)g(t \wedge s)dk(t)dk(s). \end{aligned}$$

Using integration by parts, the first integral is

$$\begin{aligned}
\int \bar{F}(t)g(t \wedge s)dk(t) &= \int_0^s \bar{F}(t)g(t)dk(t) + \int_s^\tau \bar{F}(t)g(s)dk(t) \\
&= - \int_0^s (\bar{F}(t)g(t))'k(t)dt + [\bar{F}(t)g(t)k(t)]_0^s \\
&\quad + -g(s) \int_s^\tau \bar{F}(t)'k(t)dt + g(s)[\bar{F}(t)k(t)]_s^\tau \\
&= - \int_0^s (\bar{F}(t)g(t))'k(t)dt - g(s) \int_s^\tau \bar{F}(t)'k(t)dt
\end{aligned}$$

Then

$$\begin{aligned}
nVar(\sigma_n(x)) &= - \int \bar{F}(s) \int_0^s (\bar{F}(t)g(t))'k(t)dt dk(s) - \int \bar{F}(s)g(s) \int_s^\tau \bar{F}(t)'k(t)dk(s)dt \\
&= - \int (\bar{F}(t)g(t))'k(t) \int_t^\tau \bar{F}(s)dk(s)dt - \int \bar{F}(t)'k(t) \int_0^t \bar{F}(s)g(s)dk(s)dt \\
&= - \int \bar{F}(t)g'(t)k(t) \int_t^\tau \bar{F}(s)dk(s)dt - \int \bar{F}'(t)g(t)k(t) \int_t^\tau \bar{F}(s)dk(s)dt \\
&\quad - \int \bar{F}(t)'k(t) \int_0^t \bar{F}(s)g(s)dk(s)dt \\
&= I + II + III.
\end{aligned}$$

By integration by parts, we can see that

$$I = \int \bar{F}^2(t)g'(t)k^2(t)dt + O(1),$$

$$II = \int \bar{F}'(t)\bar{F}(t)g(t)k^2(t)dt + O(1),$$

and

$$III = - \int \bar{F}'(t)\bar{F}(t)g(t)k^2(t)dt + O(1),$$

Therefore and using the fact that

$$g'(t) = f(t)/[\bar{G}(t)\bar{F}(t)^2]$$

we get

$$\begin{aligned}
nVar(\sigma_n(x)) &= \int \bar{F}^2(t)g'(t)k^2(t)dt + O(1) \\
&= \int f(t)/[\bar{G}(t)]k^2(t)dt \\
&= B_b(x)E(f(\eta_x)\bar{G}^{-1}(\eta_x))
\end{aligned}$$

where  $\eta_x$  is a gamma( $2x/b + 1, b/2$ ) random variable and

$$B_b(x) = \frac{b^{-1}\Gamma(2x/b + 1)}{2^{2x/b+1}\Gamma^2(x/b + 1)}$$

From Chen (2000), for a small value of  $b$ ,

$$B_b(x) \sim \begin{cases} \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}b^{-1} & \text{if } x/b \rightarrow \kappa \\ \frac{1}{\sqrt{2\pi}}b^{-1/2}x^{-1/2} & \text{if } x/b \rightarrow \infty \end{cases}$$

Thefore

$$nVar(\sigma_n(x)) \sim \begin{cases} \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}b^{-1}f(x)/\bar{G}(x) + o(b^{-1}) & \text{if } x/b \rightarrow \kappa \\ \frac{1}{\sqrt{2\pi}}b^{-1/2}x^{-1/2}f(x)/\bar{G}(x) + o(b^{-1/2}) & \text{if } x/b \rightarrow \infty \end{cases}$$

■

To prove Theorem 1, we need the following lemma.

**Lemma 2.** *Under condition 1, we have, away from zero*

$$n^{1/2}b^{1/4}|\hat{f}_b(x) - E(\hat{f}_b(x)) - (\hat{f}_b^*(x) - E(\hat{f}_b^*(x)))| \stackrel{a.s.}{=} O(b^{1/4}(\log \log n)^{1/2}) + \frac{\log \log n}{n^{1/2}b^{1/4}} \quad (26)$$

and at the boundary region we have

$$n^{1/2}b^{1/4}|\hat{f}_b(x) - E(\hat{f}_b(x)) - (\hat{f}_b^*(x) - E(\hat{f}_b^*(x)))| \stackrel{a.s.}{=} O(b^{1/4}(\log \log n)^{1/2}) + \frac{\log \log n}{n^{1/2}b^{1/4}} \quad (27)$$

where

$$\hat{f}_b^*(x) = \int \frac{1}{\bar{G}(t)}K(x, b)(t)dH_n^*(t)$$

### Proof of Lemma 2

We prove the lemma only for  $x$  away from zero, we follow the almost the same steps for the boundary region. From Lemma 1, (20) and using integration by part, we get

$$\begin{aligned} \hat{f}_b(x) - E(\hat{f}_b(x)) &\stackrel{a.s.}{=} - \int \bar{F}(t)(\Delta_n(t) - \Delta(t))dk(x, b)(t) + O(a_n^2b^{-1/2}) \\ &\stackrel{a.s.}{=} \int \bar{F}(t)k(x, b)(t) d(\Delta_n(t) - \Delta(t)) + O(a_n) + O(a_n^2b^{-1/2}) \\ &\stackrel{a.s.}{=} (\hat{f}_b^*(x) - E(\hat{f}_b^*(x))) + e_n(x) + O(a_n) + O(a_n^2b^{-1/2}) \end{aligned}$$

where

$$e_n(x) = \int \frac{\bar{H}(t) - \bar{Y}_n(t)}{\bar{Y}_n(t)\bar{G}(t)} k(x, b)(t) dH_n^*(t).$$

Using 16, 17 and 20, we have

$$|e_n(x)| \leq \sup_t |\bar{H}(t) - \bar{Y}_n(t)| \int \frac{k(x, b)(t)}{\bar{Y}_n(t)\bar{G}(t)} dH_n^*(t) \quad (28)$$

$$\stackrel{a.s.}{\equiv} O(a_n) \int k(x, b)(t) dH_n^*(t) \quad (29)$$

$$\stackrel{a.s.}{\equiv} O(a_n) + O(a_n^2 b^{-1/2}) \quad (30)$$

which implies that  $n^{1/2} b^{1/4} |e_n(x)| = O(b^{1/4} (\log \log(n))^{1/2}) + O(n^{-1/2} b^{-1/4} \log \log(n))$

■

### Proof of Proposition 1

From Lemma 2 and using Theorem 2 in Bouezmarni and Rombouts (2006) which states the asymptotic normality of gamma kernel estimator of the density function under  $\alpha$ -mixing condition. Concerning the hazard rate function, note that the asymptotic normality of gamma kernel estimator of the hazard rate function under  $\alpha$ -mixing condition is straightforward from Bouezmarni and Rombouts (2006).

■

### Proof of Proposition 2

We begin with the classical composition

$$|\hat{f}(x) - f(x)| \leq |\hat{f}(x) - E(\hat{f}(x))| + |E(\hat{f}(x)) - f(x)| \quad (31)$$

From (22) and condition on the bandwidth parameter that,

$$|E(\hat{f}(x)) - f(x)| = \left| \int K_{x,b}(t) f(t) dt - f(x) \right| + o(1) \quad (32)$$

and by continuity the first term converges to zero.

For the variation term we use lemma 2 and the almost convergence of the variation term of gamma kernel estimator for dependent uncensored data established by Bouezmarni and Rombouts (2006)

■

## References

- BOUEZMARNI, T., AND J. ROMBOUTS (2006): “Nonparametric Density Estimation for Positive Time Serie,” *CORE discussion paper 2006/85*.
- BOUEZMARNI, T., AND O. SCAILLET (2003): “Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data,” *Econometric Theory*, 21, 390–412.
- BRESLOW, N., AND J. CROWLEY (1974): “A Large Sample Study of the Life Table and Product-Limit Estimates Under Random Censorship,” *Annals of Statistics*, 2, 437–443.
- CAI, Z. (1998a): “Asymptotic properties of Kaplan-Meier estimator for censored dependent data,” *Statistics and Probability Letters*, 37, 381–389.
- (1998b): “Kernel density and hazard rate estimation for censored dependent data,” *Journal of Multivariate Analysis*, 67, 23–34.
- (2001): “Estimating a Distribution Function for Censored Time Series Data,” *Journal of Multivariate Analysis*, 78, 299–318.
- CAI, Z., AND G. ROUSSAS (1998): “Kaplan Meier estimator under association,” *Journal of Multivariate Analysis*, 67, 318–348.
- CHEN, S. (2000): “Probability Density Functions Estimation Using Gamma Kernels,” *Annals of the Institute of Statistical Mathematics*, 52, 471–480.
- DOUKHAN, P. (1994): *Mixing: Properties and Examples*, vol. 85. Springer-Verlag, New York.
- JONES, M. (1993): “Simple Boundary Correction for Kernel Density Estimation,” *Statistical Computing*, 3, 135–146.
- JONES, M., AND P. FOSTER (1996): “A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation,” *Statistica Sinica*, 6, 1005–1013.
- KAPLAN, E., AND P. MEIER (1958): “Nonparametric estimation from incomplete observations,” *Journal of the American Statistical Association*, 53, 457–481.

- LEJEUNE, M., AND P. SARDA (1992): “Smooth Estimators of Distribution and Density Functions,” *Computational Statistics and Data Analysis*, 14, 457–471.
- LEONENKO, N., AND L. SAKHNO (2001): “On the Kaplan-Meier estimator for dependent sequences,” *Statistical Inference for Stochastic Processes*, 4, 17–40.
- LIEBSCHER, E. (2002): “Kernel density and hazard rate estimation for censored data under  $\alpha$ -mixing condition,” *Probability Theory and Related Fields*, 67, 318–348.
- LO, S., Y. MACK, AND J. WANG (1989): “Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator,” *Probability Theory and Related Fields*, 80, 473–473.
- LO, S., AND K. SINGH (1985): “The Product-Limit Estimator and the Bootstrap: Some Asymptotic Representations,” *Probability Theory and Related Fields*, 71, 318–348.
- MARRON, J., AND D. RUPPERT (1994): “Transformations to reduce boundary bias in kernel density estimation,” *Journal of the Royal Statistical Society, Series B*, 56, 653–671.
- MIELNICZUK, J. (1986): “Some Asymptotic Properties of Kernel Estimators of a Density Function in Case of Censored Data,” *Annals of Statistics*, 14, 766–773.
- MÜLLER, H. (1991): “Smooth Optimum Kernel Estimators near Endpoints,” *Biometrika*, 78, 521–530.
- SCAILLET, O. (2004): “Density estimation using inverse and reciprocal inverse Gaussian kernels,” *Journal of Nonparametric Statistics*, 16, 217–226.
- SCHUSTER, E. (1985): “Incorporating Support Constraints into Nonparametric Estimators of Densities,” *Communications in Statistics - Theory and Methods*, 14, 1123–1136.
- STUTE, W., AND J. WANG (1993): “A Strong Law under Random Censorship,” *Annals of Statistics*, 21, 1591–1607.
- TANNER, M., AND W. WONG (1983): “The Estimation of the Hazard Function from Randomly Censored Data by the Kernel Method,” *Annals of Statistics*, 11, 989–993.
- WANG, J. (1987): “A Note on the Uniform Consistency of the Kaplan-Meier Estimator,” *Annals of Statistics*, 15, 1313–1316.

YING, Z., AND L. WEI (1994): "The Kaplan-Meier Estimate for Dependent Failure Time Observations," *Journal of Multivariate Analysis*, 50, 17–29.