I N S T I T U T D E

S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N

P A P E R

0521

EMPIRICAL LIKELIHOOD CONFIDENCE INTERVALS FOR DEPENDENT DURATION DATA

A. EL GHOUCH, I. VAN KEILEGOM and I.W. McKEAGUE

http://www.stat.ucl.ac.be

Empirical Likelihood Confidence Intervals for Dependent Duration Data

Anouar El Ghouch*, Ingrid Van Keilegom* & Ian W. McKeague†

Abstract

Three types of confidence intervals are developed for a general class of functionals of a survival distribution based on censored dependent data. The confidence intervals are constructed via asymptotic normality (Wald's method), the empirical likelihood (EL) method, and the blockwise EL method in which sample means over blocks of observations are used in place of the original data. Asymptotic results are derived to accurately calibrate the various procedures and their performance is evaluated in a simulation study. The problem of the choice of the blocksize is also discussed.

Key words: blocking, strong mixing, censoring, Kaplan-Meier integral.

1 Introduction

Dependent censored data arise in economic duration analysis, in which event times (duration or survival times) are correlated, and the observation of the event may be prevented by the occurrence of an earlier competing event (censoring). Observations on duration of unemployment e.g., may be right censored and are typically correlated. Such dependent censored data occurs, for example, when study participants belong to clusters (e.g., month of unemployment, job type, neighborhood, school), with members of the same

*Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium. Email: elghouch@stat.ucl.ac.be, vankeilegom@stat.ucl.ac.be. Research supported by Belgian IAP research network grant nr. P5/24.

†Department of Biostatistics, Columbia University, 722 West 168th St., 6th Floor, NY 10032, USA. Email: im2131@columbia.edu.

cluster having correlated risk of the event of interest. Chen et al. (2005) discuss an example involving number of weeks that an individual would like to collect unemployment benefits. Other examples can be found in medical follow-up studies, epidemiology and reliability. See Eriksson and Adell (1994) and Ying and Wei (1994) for some concrete examples. Conventional analyses of such data assume that study participants are randomly sampled from the population, which can produce misleadingly narrow interval estimates of survival probabilities. In the present paper we allow dependence between individuals and construct more suitable confidence intervals for a general class of functionals of the survival distribution.

Let $X_1, X_2, \ldots$ (survival times) and $Y_1, Y_2, \ldots$ (censoring times) be two independent, strictly stationary, sequences of random variables on the real line with marginal distribution functions (df) F and G, respectively. The dependence along each sequence is assumed to diminish geometrically (see assumption **A2** below). Under the censoring model, instead of observing X_i, we observe the pair (Z_i, δ_i), $i = 1, \ldots, n$, where $Z_i = \min(X_i, Y_i)$ and $\delta_i = I(X_i \leq Y_i)$ with $I(\cdot)$ the indicator function. Let $H(t) = 1 - (1 - F(t))(1 - G(t))$ be the df of Z_i, which we assume to be continuous. Let $\hat{F}$ and $\hat{G}$ denote the Kaplan–Meier (KM) estimators of F and G, respectively, that is

$$1 - \hat{F}(t) = \prod_{Z_{(i)} \leq t} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}} \text{ and } 1 - \hat{G}(t) = \prod_{Z_{(i)} \leq t} \left(\frac{n-i}{n-i+1} \right)^{1-\delta_{(i)}},$$

where $Z_{(1)} \leq Z_{(2)} \leq \ldots \leq Z_{(n)}$ are the order statistics of Z_i and $\delta_{(1)}, \ldots, \delta_{(n)}$ are the corresponding δ_i.

We are interested in constructing a nonparametric confidence interval (CI) for a parameter of the form

$$\theta = \theta(F) = \int \xi(t) dF(t), \tag{1}$$

where ξ is some given measurable function (see assumption **A1** below). Various parameters of interest can be written in the form of (1). For example, if $\xi(t) = I(t \leq t_0)$, then $\theta = F(t_0)$ and if $\xi(t) = t$, then $\theta = \mathbb{E}(X)$. We refer to Stute and Wang (1993) for other examples.

For i.i.d. complete (uncensored) data, the central limit theorem (CLT) for the sample mean $n^{-1} \sum_{i=1}^{n} \xi(X_i)$ can be used to provide a Wald-type CI for θ. For censored data, such a fundamental result did not exist until Stute (1995) obtained a CLT for functionals of the form $\int \xi d\hat{F}$. A consistent estimator of the limiting variance of $\int \xi d\hat{F}$ was proposed by Stute (1996), so a Wald-type CI can be found for θ. Wald-type CIs are centered on the point estimate

and calibrated easily given asymptotic normality (AN), but they have several drawbacks: their small sample properties can be unsatisfactory and they may include values outside the natural range of the parameter. Improved CIs can be obtained using the empirical likelihood (EL) approach of Owen (1988). A discussion of the advantages of the EL method over classical methods (based on a normal approximation and the bootstrap) can be found in Hall and La Scala (1990) and Owen (2001). It is important to note that EL was originally introduced by Thomas and Grunkemeier (1975) to construct CIs for survival probabilities, but the idea cannot be easily adapted to general functionals of the form (1). Recently, Wang and Jing (2001) used a plug-in version of EL to find a CI for θ in the case of independent censored data. In the case of dependent censored data, however, only Wald-type CIs are available, and only when ξ is an indicator function, see, e.g., Cai (2001).

Throughout we restrict attention to functionals $\theta(F)$ for which
ASSUMPTION **A1**:

$$\xi(t) = 0 \text{ for all } t > T, \text{ for some } T < \tau := \inf\{t : H(t) = 1\}.$$

The truncation imposed on ξ means, for example, that instead of the survival mean, $\int t\, dF(t)$, we get the truncated mean, $\int_{-\infty}^{T} t\, dF(t)$. However, as Gijbels and Veraverbeke (1991) explain it, the truncated functional is very often not too different from the complete (untruncated) functional if T is taken sufficiently large. In practice, T can be taken as the last observed survival time.

Our first goal is to establish the asymptotic normality of $\hat{\theta} := \int \xi d\hat{F}$ via a representation of the KM integral in terms of the partial sum of a stationary β-mixing sequence plus an asymptotically negligible remainder term. For the proof of this result, we adapt to our setting the approach of Stute (1995), which is only valid for i.i.d. data. Our second goal is the construction of EL-based CIs for θ. This will be done in two ways: (1) adjusting the EL statistic to have an asymptotic χ^2-distribution, and (2) blockwise empirical likelihood (BEL), which is a version of EL based on data blocking techniques proposed by Kitamura (1997) in the context of weakly dependent processes; here the blockwise log-likelihood ratio is adjusted to have an asymptotic χ^2-distribution. The adjusted (B)EL has the same advantages as standard (B)EL over Wald-type CIs.

The paper is organized as follows. After introducing a β-mixing (absolutely regular) condition, in Section 2 we develop an asymptotic representation of the KM integral. The asymptotic normality of $\hat{\theta}$ is obtained as a corollary. The problem of estimating the limiting variance is also discussed. In Sections 3 and 4 we use the EL approach, with and without blocking, to

construct CIs for θ. The performance of the three methods (Wald (AN), EL and BEL) is compared via simulation in Section 5. In Section 6, we develop a way of selecting the block size and assess its performance numerically.

2 Asymptotic representation of the KM integral

In this section we establish the basic result that is used to construct the various confidence intervals.

We first define a suitable measure of dependence. For any two σ-fields $\mathcal{A}$ and $\mathcal{B}$ in a given probability space, let

$$\beta(\mathcal{A},\mathcal{B}) = \frac{1}{2}\sup\sum_{i=1}^{I}\sum_{j=1}^{J}\left|\mathbb{P}(A_i\cap B_j)-\mathbb{P}(A_i)\mathbb{P}(B_j)\right|,$$

where the supremum is over all finite $\mathcal{A}$-partitions $(A_1,\ldots,A_I)$ and all finite $\mathcal{B}$-partitions $(B_1,\ldots,B_J)$. A strictly stationary sequence $\{T_k,\ k\in\mathbb{Z}\}$ is absolutely regular (or β-mixing) if

$$\beta(n) := \beta\left(\mathcal{F}_{-\infty}^{0},\mathcal{F}_{n}^{\infty}\right)\xrightarrow[n\to\infty]{} 0,$$

where $\mathcal{F}_J^L$ denotes the σ-field generated by the family $\{T_k,\ J\le k\le L\}$. For the properties of this and other strong mixing conditions we refer the reader to Bradley (1986) and Doukhan (1994). Among the various mixing conditions available in the literature, β-mixing is relatively weak; it is more restrictive than α-mixing, but weaker than ρ-mixing. In the sequel we use

Assumption **A2**: $\{X_i\}$ and $\{Y_i\}$ are strictly stationary and absolutely regular, and there exists $\nu>3$ such that both β-mixing coefficients satisfy

$$\beta(n) = O(n^{-\nu}). \tag{2}$$

Along with the assumption that $\{X_i\}$ and $\{Y_i\}$ are independent, this implies that the sequence of (X_i,Y_i) is absolutely regular with β-mixing coefficient satisfying (2). Hence (Z_i,δ_i) satisfies the same property. We are now ready to state our main result, for which we need the following notation:

$$U_i = \frac{\xi(Z_i)\delta_i}{1-G(Z_i)} \equiv \xi(Z_i)\gamma_0(Z_i)\delta_i,$$

$$\gamma_1(t) = (1-H(t))^{-1}\int_{t+}^{\infty}\xi(x)dF(x),\ \text{and}$$

$$\gamma_2(t) = \int_{-\infty}^{t-}\frac{\gamma_1(y)}{1-G(y)}dG(y).$$

Theorem 1 *If* **A1** *and* **A2** *hold, and* $\int |\xi(t)|^p dF(t) < \infty$, *for some* $p \geq 3$, *then*

$$\hat{\theta} := \int \xi(t) d\hat{F}(t) = n^{-1} \sum_{i=1}^{n} \eta_i + o_P(n^{-1/2}),$$

where $\eta_i \equiv \eta_i(F, G) = U_i + \gamma_1(Z_i)(1 - \delta_i) - \gamma_2(Z_i)$. (3)

Note that the sequence $\{\eta_i\}$ is strictly stationary and absolutely regular, with β-mixing coefficient satisfying (2). The following corollary is a direct application of the CLT for strongly mixing sequences; see, for example, Rio (2000).

Corollary 2 *Under the assumptions of Theorem 1,*

$$n^{1/2} \left(\hat{\theta} - \theta \right) \longrightarrow \mathcal{N} \left(0, \, \sigma_\eta^2 \right)$$

in distribution, with $\sigma_\eta^2 = \mathbb{V}ar(\eta_1) + 2 \sum_{i>1} \mathbb{C}ov(\eta_1, \eta_i)$.

Note that $\sigma_\eta^2 < \infty$, but it can be zero. To avoid the uninteresting case, in the sequel we assume that $\sigma_\eta^2 > 0$.

Proof of Theorem 1.
As mentioned in the Introduction, we can adapt the approach of Stute (1995), so many details will be omitted. Let H^0 and H^1 be the true unknown sub-df of the censored and uncensored observation respectively, that is $H^q(t) = \mathbb{P}\left(Z_i \leq t, \, \delta_i = q\right), \; q = 0, 1$. For any (sub-)df Q, we denote by Q_n the corresponding empirical (sub-)df. By Lemma 2.1 in Stute (1995), we write

$$\int \xi d\hat{F} = n^{-1} \sum_{i=1}^{n} U_i + R_{n1} + R_{n2} + S_n, \text{ where}$$

(I) $R_{n1} = n^{-1} \sum_{i=1}^{n} U_i B_{in}$, with

$$B_{in} := n \int_{-\infty}^{Z_i -} \ln \left[1 + \frac{1}{n(1 - H_n(t))} \right] dH_n^0(t) - \int_{-\infty}^{Z_i -} \frac{dH_n^0(t)}{1 - H_n(t)}.$$

It can be shown that $|B_{in}| \leq \dfrac{1}{2n} \dfrac{H_n^0(T)}{(1 - H_n(T))^2}$, for all $Z_i \leq T$.

So, by the SLLN (ergodicity) of strongly mixing sequences, we obtain that

$$R_{n1} = O(n^{-1}) \text{ a.s.}$$

(II) $R_{n2} = \dfrac{1}{2n}\sum_{i=1}^{n} \xi(Z_i)\delta_i e^{\Delta_{in}}(B_{in}+C_{in})^2$ with,

$C_{in} := \displaystyle\int_{-\infty}^{Z_i-} \frac{dH_n^0(t)}{1-H_n(t)} - \int_{-\infty}^{Z_i-} \frac{dH^0(t)}{1-H(t)}$ and

Δ_{in} is between the two terms

$n\displaystyle\int_{-\infty}^{Z_i-} \ln\left[1+\frac{1}{n(1-H_n(t))}\right] dH_n^0(t)$ and $\displaystyle\int_{-\infty}^{Z_i-} \frac{dH^0(t)}{1-H(t)}$.

By the expansion

$$\frac{1}{1-H_n} = -\frac{1-H_n}{(1-H)^2} + \frac{2}{1-H} + \frac{(H_n-H)^2}{(1-H)^2(1-H_n)}, \tag{4}$$

and applying the LIL for empirical (sub-)df (Theorem 3.2 of Cai and Roussas (1992)), we obtain

$$C_{in} = O\left(\sqrt{\frac{\log\log n}{n}}\right) \text{ a.s.}$$

On the other hand, it is easily seen that

$$\Delta_{in} \le \frac{H^0(T)}{1-H(T)} \text{ a.s. for all } Z_i \le T.$$

From these two inequalities and using the SLLN, we obtain

$$R_{n2} = O(n^{-1}\log\log n) \text{ a.s.}$$

(III) $S_n = n^{-1}\displaystyle\sum_{i=1}^{n} U_i C_{in}$.

From (4) we expand S_n into

$$S_n = -\int\int \frac{I(y<x)\xi(x)\gamma_0(x)}{1-H(y)} dH^0(y)dH_n^1(x) + 2S_{n1} - S_{n2} + R_{n3}, \tag{5}$$

where

- $R_{n3} = \displaystyle\int\int \xi(x)\gamma_0(x)I(y<x)\frac{(H_n(y)-H(y))^2}{(1-H(y))^2(1-H_n(y))} dH_n^0(y)dH_n^1(x)$
  $= O(n^{-1}\log\log n)$ a.s.

The last equality is a direct application of the LIL and SLLN.

- $S_{n1} = \int\int \frac{I(y<x)\xi(x)\gamma_0(x)}{1-H(y)} dH_n^0(y) dH_n^1(x).$

This is a V-statistic of degree two of the bivariate β-mixing process (Z_i, δ_i). Re-expressing S_{n1} as an U-statistic, using the Hoeffding decomposition and then applying Lemma 3 in Arcones (1998) (see also Arcones (1995)) we get that

$$\begin{aligned} S_{n1} &= \int\int \frac{I(y<x)\xi(x)\gamma_0(x)}{1-H(y)} \Big[dH^0(y) dH_n^1(x) \\ &\quad + dH_n^0(y) dH^1(x) - dH^0(y) dH^1(x) \Big] + o_P\left(n^{-1/2}\right). \end{aligned} \tag{6}$$

- $S_{n2} = \int\int\int \frac{I(y<t, y<x)\xi(x)\gamma_0(x)}{(1-H(y))^2} dH_n(t) dH_n^0(y) dH_n^1(x).$

This is a V-statistic of degree three of the bivariate β-mixing process (Z_i, δ_i). By the same reasoning as for S_{n1}, we obtain

$$\begin{aligned} S_{n2} &= \int\int\int \frac{I(y<t, y<x)\xi(x)\gamma_0(x)}{(1-H(y))^2} \Big[dH(t) dH_n^0(y) dH_n^1(x) \\ &\quad + dH(t) dH_n^0(y) dH^1(x) + dH(t) dH^0(y) dH_n^1(x) \\ &\quad - 2 dH(t) dH^0(y) dH^1(x) \Big] + o_P\left(n^{-1/2}\right). \end{aligned} \tag{7}$$

Substituting (6) and (7) into (5), and making some simplifications, completes the proof. □

Corollary 2 would allow us to construct Wald-type (AN) confidence limits for θ if the limiting variance σ_η^2 were known. Unfortunately, this is not the case and an estimator of σ_η^2 is indeed needed. In the case that ξ is the indicator function, Cai (2001) gave the exact expression of σ_η^2 and, using some blocking and plug-in techniques, he proposed a consistent estimator for this quantity. In our case we need an estimator which is available for a general ξ. To motivate our approach, note that

$$\sigma_\eta^2 = \lim_{n\to\infty} \mathbb{V}ar\left(n^{-1/2} \sum_{i=1}^{n} \eta_i \right).$$

and η_i is absolutely regular. Given the success of the moving-block jackknife (BJ) for variance estimation with dependent data (see Künsch (1989) and

Liu and Singh (1992)), it is natural to apply this procedure in our case. Let the block size $l = l(n)$ satisfy $l \to \infty$ and $l/n \to 0$. The BJ estimator of σ_η^2 is

$$\hat{\sigma}_{\eta,l}^2 = lL^{-1} \sum_{i=1}^{L} \left(\bar{\eta}_i^l - L^{-1} \sum_{i=1}^{L} \bar{\eta}_i^l \right)^2 ,$$

where $L \equiv L(n) = n - l + 1$ and $\bar{\eta}_i^l = l^{-1} \sum_{j=i}^{i+l-1} \eta_j$.

Here $\hat{\sigma}_{\eta,l}^2$ coincides with the moving-block bootstrap variance estimate, see Theorem 3.4 in Künsch (1989), and converges to $\sigma^2(\eta)$ under very weak conditions (see Radulović (1996)) that are clearly fulfilled in our case. However, this estimator cannot be used in practice since it depends on the unknown survival and censoring df's. To overcome this problem, we suggest plugging-in $\hat{F}$ and $\hat{G}$ into (3) to get $\hat{\eta}_i \equiv \eta_i(\hat{F}, \hat{G})$ and then substituting $\hat{\eta}_i$ in the formula for $\hat{\sigma}_{\eta,l}^2$ to obtain $\hat{\sigma}_{\hat{\eta},l}^2$. The numerical performance of this approach is studied in Section 5. The proposed CI is

$$\hat{\theta} \pm \frac{\hat{\sigma}_{\hat{\eta},l}}{\sqrt{n}} z_{\alpha/2} \qquad \textbf{(AN)}$$

where z_α is the upper α-quantile of the standard normal distribution.

3 Empirical likelihood

It is easy to check that $\mathbb{E}(U_i) = \theta$, hence following Owen's (1988) idea, we can define the likelihood ratio function of θ by

$$\tilde{R}(\theta) = \max \prod_{i=1}^{n} n\tilde{p}_i \quad \text{subject to} \quad \theta = \sum_{i=1}^{n} \tilde{p}_i U_i \text{ and } \sum_{i=1}^{n} \tilde{p}_i = 1.$$

Since the definition of U_i involves the unknown df G, it is natural to replace it by $\hat{G}$, cf. Wang and Jing (2001) in the i.i.d. case. The *estimated* likelihood ratio is then defined by

$$R(\theta) = \max \prod_{i=1}^{n} np_i \quad \text{subject to} \quad \theta = \sum_{i=1}^{n} p_i V_i \text{ and } \sum_{i=1}^{n} p_i = 1,$$

where

$$V_i = \frac{\xi(Z_i)\delta_i}{1 - \hat{G}(Z_i-)}.$$

By a standard Lagrange-multiplier argument, we obtain the following expression for the log-likelihood function:

$$\mathcal{L}_n(\theta) = -2\log R(\theta) = 2\sum_{i=1}^{n} \log\left(1 + \lambda_n(V_i - \theta)\right),$$

where λ_n is the solution of the equation $\sum_{i=1}^{n} \frac{V_i - \theta}{1+\lambda_n(V_i-\theta)} = 0$.

To study the asymptotic behavior of $\mathcal{L}_n$, we need the following lemma.

Lemma 3 *Under the assumptions of Theorem 1,*

(i) $\max_{1\le i\le n} |V_i| = O_P(n^{1/p})$,

(ii) $n^{-1}\sum_{i=1}^{n} (V_i - \theta)^2 \xrightarrow{P} \mathbb{V}ar(U_1)$.

Proof.

(i) Note that $V_i = U_i \dfrac{1 - G(Z_i)}{1 - \hat{G}(Z_i)}$. Since $\mathbf{E}|U_1|^p < \infty$, by Markov's inequality, $\max_{1\le i\le n} |U_i| = O_P(n^{1/p})$. The result follows from

$$\sup_{t\le T} \frac{1 - G(t)}{1 - \hat{G}(t)} = O_P(1),$$

which can be seen as a consequence of the fact that

$$\sup_{t\le T} \frac{|\hat{G}(t) - G(t)|}{1 - \hat{G}(t)} = O_P\left(\sqrt{\frac{\log\log n}{n}}\right). \tag{8}$$

This last equality can be shown in the same way as Cai (2001) did in the proof of his Theorem 2.

(ii) We write

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n} (V_i - \theta)^2 \\ &= n^{-1}\sum_{i=1}^{n} (V_i - U_i)^2 + n^{-1}\sum_{i=1}^{n} (U_i - \theta)^2 + 2n^{-1}\sum_{i=1}^{n} (U_i - \theta)(V_i - U_i). \end{aligned}$$

Observe that

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} (V_i - U_i)^2 &= n^{-1}\sum_{i=1}^{n} \left(\frac{\hat{G}(Z_i) - G(Z_i)}{1 - \hat{G}(Z_i)} U_i\right)^2 \\ &\le \left(\sup_{t\le T} \frac{|\hat{G}(t) - G(t)|}{1 - \hat{G}(t)}\right)^2 n^{-1}\sum_{i=1}^{n} U_i^2 = O_P\left(\frac{\log\log n}{n}\right). \end{aligned}$$

So using the SLLN and the Cauchy–Schwarz inequality, the result is obtained. □

Theorem 4 *Under the assumptions of Theorem 1,*

$$\sigma_\eta^{-2}\mathbb{V}ar(U_1)\mathcal{L}_n(\theta) \xrightarrow{d} \chi_1^2.$$

Proof.
We only give the main steps of the proof, and refer the reader to Owen (1988) for more details. First note that $\hat{\theta} = \bar{V}_n = n^{-1}\sum_{i=1}^n V_i$, see, e.g., Shorack and Wellner (1986, (13), (9) and (11) on pg. 295). Hence, Corollary 2 implies that

$$n^{-1/2}\sum_{i=1}^n (V_i - \theta) \xrightarrow{d} \mathcal{N}\left(0, \sigma_\eta^2\right). \tag{9}$$

From the definition of λ_n and using Lemma 3 and (9), one can check that

$$\lambda_n = O_P\left(n^{-1/2}\right). \tag{10}$$

This together with Lemma 3(i) implies that

$$\lambda_n = \frac{\sum_{i=1}^n (V_i - \theta)}{\sum_{i=1}^n (V_i - \theta)^2} + o_P\left(n^{-1/2}\right). \tag{11}$$

Using a Taylor expansion of $\mathcal{L}_n$, together with (10) and Lemma 3, yields

$$\mathcal{L}_n(\theta) = 2\lambda_n \sum_{i=1}^n (V_i - \theta) - \lambda_n^2 \sum_{i=1}^n (V_i - \theta)^2 + o_P(1). \tag{12}$$

Substituting (11) into (12), using again (9) and (10) together with Lemma 3, we get

$$\mathcal{L}_n(\theta) = \frac{\left(\sum_{i=1}^n (V_i - \theta)\right)^2}{\sum_{i=1}^n (V_i - \theta)^2} + o_P(1)$$

which leads to the result. □

Since $\hat{\theta}$ is a consistent estimator of θ, from Lemma 3(ii), we can consistently estimate $\mathbb{V}ar(U_1)$ by $n^{-1}\sum_{i=1}^n \left(V_i - \bar{V}_n\right)^2$. As a consequence, we propose the following EL confidence interval for θ:

$$\left\{\mu : \frac{n^{-1}\sum_{i=1}^n (V_i - \bar{V}_n)^2}{\hat{\sigma}^2_{\hat{\eta},l}} \mathcal{L}_n(\mu) \le \chi_1^2(\alpha)\right\} \qquad \textbf{(EL)}$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of the χ_1^2 distribution.

4 Blockwise empirical likelihood

In this section we construct an EL profile ratio for θ based on observational blocks, as proposed by Kitamura (1997).

Let the block size $b \equiv b_n$ satisfy $b \to \infty$ and $bn^{1/p-1/2} \to 0$. For $i = 1, \ldots, N := n - b + 1$ we denote by $\bar{V}_{i,b}$ the sample mean of the block $(V_i, \ldots, V_{i+b-1})$. Instead of assigning mass to each single observation, here we assign a mass $\{p_i\}_{1\le i\le N}$ to each block sample mean $\left\{\bar{V}_{i,b}\right\}_{1\le i\le N}$. The estimated blockwise EL ratio at θ is

$$R^b(\theta) = \max \prod_{i=1}^{N} N p_i \quad \text{subject to} \quad \theta = \sum_{i=1}^{N} p_i \bar{V}_{i,b} \text{ and } \sum_{i=1}^{N} p_i = 1,$$

which yields the log-likelihood function

$$\mathcal{L}_{n,b}(\theta) = 2 \sum_{i=1}^{N} \log\left(1 + \lambda_{n,b}(\bar{V}_{i,b} - \theta)\right),$$

where $\lambda_{n,b}$ is the solution of the equation $\sum_{i=1}^{N} \frac{\bar{V}_{i,b}-\theta}{1+\lambda_{n,b}(\bar{V}_{i,b}-\theta)} = 0$. When no confusion is possible, we will write $\bar{V}_i$ and λ, instead of $\bar{V}_{i,b}$ and $\lambda_{n,b}$.

Theorem 5 *Under the assumptions of Theorem 1,*

$$r_n \sigma_U^2 \sigma_\eta^{-2} \mathcal{L}_{n,b}(\theta) \xrightarrow{d} \chi_1^2,$$

where $r_n = N^{-1}n/b$ *and* $\sigma_U^2 = \mathbb{V}ar(U_1) + 2\sum_{i>1} \mathbb{C}ov(U_1, U_i)$.

To prove this theorem we need the following lemma.

Lemma 6 *Under the assumptions of Theorem 1,*

(i) $n^{1/2} N^{-1} \sum_{i=1}^{N} (\bar{V}_i - \theta) \xrightarrow{d} N(0, \sigma_\eta^2)$,

(ii) $\max_{1\le i\le N} \left|\bar{V}_i\right| = O_P\left(n^{1/p}\right)$,

(iii) $bN^{-1} \sum_{i=1}^{N} \left(\bar{V}_i - \theta\right)^2 \xrightarrow{p} \sigma_U^2$.

Proof.
(i) Note that $\sum_{i=1}^{N}\left(\bar{V}_i-\theta\right)=\sum_{i=1}^{n}(V_i-\theta)-\hat{K}_n$, with

$$\hat{K}_n=b^{-1}\sum_{j=1}^{b}(b-j)(V_j-\theta)+b^{-1}\sum_{j=1}^{b}(b-j)(V_{n-j+1}-\theta)=\hat{K}_n^1+\hat{K}_n^2 \text{ (say) .}$$

$\hat{K}_n^1$ may be written as

$$\hat{K}_n^1=b^{-1}\sum_{j=1}^{b}(b-j)(U_j-\theta)+b^{-1}\sum_{j=1}^{b}(b-j)(V_j-U_j)=K_n^1+I_n^1 \text{ (say) .}$$

Clearly $\mathbb{E}(K_n^1)=0$, and by stationarity

$$\begin{aligned}b^2\mathbb{V}ar(K_n^1)&=\sum_{j=1}^{b}(b-j)^2\mathbb{V}ar(U_1)+2\sum_{i=1}^{b-1}\sum_{j=1}^{b-i}(b-i)(b-i-j)\mathbb{C}ov(U_1,U_{j+1})\\&\leq b^3\mathbb{V}ar(U_1)+2b^3\sum_{i=1}^{b}\left|\mathbb{C}ov\left(U_1,U_{i+1}\right)\right|.\end{aligned}$$

By Davydov's inequality (see for example Theorem 3 in Doukhan (1994)),

$$\sum_{i=1}^{b}\left|\mathbb{C}ov\left(U_1,U_{i+1}\right)\right|=O\left(\sum_{n\geq 1}\beta(n)^{1-2/p}\right)=O(1).$$

So, $\mathbb{V}ar\left(n^{-1/2}K_n^1\right)=O(n^{-1}b)$, and hence $K_n^1=o_P(n^{1/2})$. On the other hand,

$$|I_n^1|\leq\sup_{t\leq T}\frac{|\hat{G}(t)-G(t)|}{1-\hat{G}(t)}\sum_{j=1}^{b}|U_j|=o_P(n^{1/2}).$$

We deduce that $\hat{K}_n^1=o_P(n^{1/2})$. Following the same procedure, it can be shown that $\hat{K}_n^2=o_P(n^{1/2})$, and hence this is also the case for $\hat{K}_n$. To conclude the proof, it suffices to apply (9) and the fact that $n/N\to 1$.

(ii) From the definition of $\bar{V}_i$ it is easy to check that $\max_{1\leq i\leq N}\left|\bar{V}_i\right|\leq\max_{1\leq i\leq n}|V_i|$. So the result follows directly by Lemma 3(i).

(iii) We write

$$\begin{aligned}bN^{-1}\sum_{i=1}^{N}\left(\bar{V}_i-\theta\right)^2=bN^{-1}\sum_{i=1}^{N}\left(\bar{V}_i-\bar{U}_i\right)^2+bN^{-1}\sum_{i=1}^{N}\left(\bar{U}_i-\theta\right)^2+\\2bN^{-1}\sum_{i=1}^{N}\left(\bar{U}_i-\theta\right)\left(\bar{V}_i-\bar{U}_i\right),\end{aligned}$$

with $\bar{U}_i = b^{-1}\sum_{j=i}^{i+b-1} U_j$. Observe that

$$bN^{-1}\sum_{i=1}^{N}\left(\bar{V}_i - \bar{U}_i\right)^2 \leq \left(\sup_{t\leq T}\frac{|\hat{G}(t)-G(t)|}{1-\hat{G}(t)}\right)^2 bN^{-1}\sum_{i=1}^{N}\left(b^{-1}\sum_{j=i}^{i+b-1}|U_j|\right)^2. \tag{13}$$

Clearly, for a fixed n, $\{(b^{-1}\sum_{j=i}^{i+b-1}|U_j|)^2,\ i\geq 1\}$ is stationary, so using Minkowski's inequality it follows that

$$\mathbb{E}\left[N^{-1}\sum_{i=1}^{N}\left(b^{-1}\sum_{j=i}^{i+b-1}|U_j|\right)^2\right] \leq \mathbb{E}(U_1^2) < \infty.$$

This together with (8) implies that the left hand side of (13) converges to 0. Now, by Lemma 1 in Radulović (1996) one can easily check that $bN^{-1}\sum_{i=1}^{N}\left(\bar{U}_i-\theta\right)^2 \xrightarrow{P} \sigma_U^2$. Finally use the Cauchy–Schwarz inequality to complete the proof. □

We omit the proof of Theorem 5 as it follows the same steps as the proof of Theorem 4.

The proposed BEL confidence interval is given by

$$\left\{\mu:\ r_n\frac{bN^{-1}\sum_{i=1}^{N}\left(\overline{V}_{i,b}-N^{-1}\sum_{i=1}^{N}\overline{V}_{i,b}\right)^2}{\hat{\sigma}^2_{\hat{\eta},l}}\mathcal{L}_{n,b}(\mu)\leq\chi^2_1(\alpha)\right\}. \qquad \textbf{(BEL)}$$

REMARK: If we choose a fixed $b=1$, we obtain exactly the EL confidence interval. So, one can consider the classical EL (without blocking) as a particular case of the BEL.

5 Numerical study

In this section we present a simulation study in order to compare, for the case of finite samples, the performance of the three proposed confidence intervals (AN, EL, BEL). Two functionals of the survival function are investigated:

- $\xi(x) = I(x\leq t)$, i.e. $\theta = F(t)$ the df at a given t.
- $\xi(x) = xI(x\leq\tau)$, i.e. $\theta = \int_{-\infty}^{\tau} xdF(x)$ the truncated mean at a given τ.

When ξ is the indicator function, we will also compare the performance of the BJ estimator for σ_η with Cai's estimator (see formula (2.13) in Cai (2001)). Simulations have been carried out for several Weibull and Uniform distributions for the survival and censoring distribution. Since the results were quite similar, here we will only show the case when the survival distribution is the standard exponential, $1-F(t) = \exp(-t)$, $t > 0$ and censoring is uniform on $[0, c]$, $G(t) = t/c$, $0 < t < c$. The value of c is determined to achieve some prespecified censoring rate (25%, 50% and 75%). To generate our data, we consider an ARMA(p, q) time series of the form

$$Z_t = \sum_{i=1}^{p} \alpha_i Z_{t-i} + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i} + \epsilon_t,$$

where the ϵ_t are i.i.d. $\mathcal{N}(0, 1)$. Two different models are chosen:

- **Model 1**: MA(3), $(\gamma_1, \gamma_2, \gamma_3) = (4.5, -3.1, 2.7)$.
- **Model 2**: ARMA(3, 3), $(\alpha_1, \alpha_2, \alpha_3) = (1.7, -1.3, 0.45)$, $(\gamma_1, \gamma_2, \gamma_3) = (4.5, -3.1, 2.7)$.

Clearly the dependence under Model 2 is stronger than the dependence under Model 1. This can be seen from the theoretical auto-correlation-function of each model (see Figure 1 below). The resulting process Z_t is strictly stationary and β-mixing, with $\beta(n) \to 0$ at an exponential rate (see Pham and Tran (1985) and Bougerol and Picard (1992)). The marginal distribution of Z_t is normal with mean 0 and variance $\sigma^2 \approx 38.15$ for Model 1 and $\sigma^2 \approx 104.67$ for Model 2. From each model, we generate independently two samples, Z_t^1 and Z_t^2, of size $n = 300$ and then we take $X_t = F^{-1}(\Phi(Z_t^1/\sigma))$ and $Y_t = G^{-1}(\Phi(Z_t^2/\sigma))$, where Φ is the df of a $\mathcal{N}(0, 1)$. Of course the process X_t (Y_t) is strictly stationary β-mixing with df F (G). To calculate the three CIs we need a block size l for the estimated asymptotic variance. In this study the value of l ranges from $l = 1$ to $l = 35$. Moreover, for the blockwise EL we need also the block size b. In this case, for each fixed value of l, b ranges from $b = 2$ to $b = 25$. For each scenario, the empirical coverage probability and the mean length are calculated over 1500 simulated confidence intervals. The results are summarized in Tables 1 and 2 for the distribution function and Table 3 for the truncated mean. Each entry in the table represents the best result (minimum coverage error, as the first criteria, and minimum length) obtained over all possible (fixed) values of l and b.

For $\theta = F(t)$, from Table 1 and 2 we observe that the coverage probabilities and lengths of all the CIs found using the Cai and BJ variance estimators are quite close, but the BJ procedure gives systematically slightly better coverage probability. The coverage error and the length of all CIs increases as

the degree of dependence in the data increases. For Model 2, except for the case where $t = 0.5$ and 70% censoring, the CIs undercover. In general we get better results at middle time points. At early time points, the CIs perform quite poor especially with the AN approach. In this case there is a considerable improvement with BEL method. Note also that the performance of the CIs depends also on the censoring rate. Generally, the length of the CI increases as the censoring rate increases, but the coverage accuracy also increases. Finally, the coverage probability gets close to the nominal coverage as we pass from AN to EL and from EL to BEL.

For the truncated mean, first note that we have taken a different value of τ for the different censoring rates ($\tau = 0.8$ and $\tau = 0.65$ for 25% and 50% censoring, respectively). This is natural since we cannot hope to do very well with high censoring. From Table 3 we observe that the results are quite similar to those for $F(t)$. In particular, BEL still does the best and under Model 2 the performance of none of the CIs is very satisfactory.

Another objective of our simulations was to study the effect of the blocksize. Table 4 provides an illustration of this. The performance of the CIs depends rather critically on the choice of the blocksize. Typically, choosing an inappropriate block length leads to under coverage, although we did find some cases (not shown here) of over coverage. In summary, two things have become clear. First, BEL appears to be more sensitive to the choice of l (the blocksize of the asymptotic variance estimator) than to the choice of b (the blocksize of the blockwise EL). With a 'good' value of l, one may obtain a reasonable result even if the choice of b is 'not so good.' Second, around the optimal value of l and/or b there is a tolerable range of blocksizes within which the results are close to optimal.

			AN		EL		BEL	
		Var	Cai	BJ	Cai	BJ	Cai	BJ
t	%cens							
0.2	25	coverage	0.926	0.927	0.932	0.933	0.938	0.940
		length	*0.085*	*0.085*	*0.085*	*0.085*	*0.087*	*0.087*
0.5	25	coverage	0.933	0.935	0.934	0.935	0.950	0.950
		length	*0.102*	*0.103*	*0.102*	*0.102*	*0.107*	*0.108*
	50	coverage	0.944	0.947	0.946	0.948	0.950	0.950
		length	*0.117*	*0.117*	*0.117*	*0.118*	*0.121*	*0.118*
	70	coverage	0.938	0.946	0.942	0.950	0.943	0.950
		length	*0.181*	*0.188*	*0.184*	*0.192*	*0.186*	*0.192*
0.7	25	coverage	0.932	0.935	0.935	0.937	0.937	0.941
		length	*0.103*	*0.104*	*0.103*	*0.104*	*0.104*	*0.105*

Table 1: **Model** 1. 95% *confidence interval for* $F(x)$ *at* $x = F^{-1}(t)$ *(best fixed block size results).*

			AN		EL		BEL	
		Var	Cai	BJ	Cai	BJ	Cai	BJ
t	%cens							
0.2	25	coverage	0.866	0.867	0.882	0.883	0.892	0.893
		length	*0.179*	*0.180*	*0.179*	*0.179*	*0.180*	*0.180*
0.5	25	coverage	0.889	0.892	0.898	0.900	0.900	0.910
		length	*0.243*	*0.243*	*0.239*	*0.240*	*0.242*	*0.241*
	50	coverage	0.914	0.916	0.915	0.917	0.918	0.920
		length	*0.253*	*0.255*	*0.251*	*0.253*	*0.258*	*0.255*
	70	coverage	0.946	0.950	0.943	0.950	0.946	0.951
		length	*0.300*	*0.311*	*0.315*	*0.320*	*0.300*	*0.300*
0.7	25	coverage	0.887	0.891	0.887	0.891	0.891	0.900
		length	*0.222*	*0.225*	*0.222*	*0.222*	*0.221*	*0.224*

Table 2: **Model** 2. 95% *confidence interval for* $F(x)$ *at* $x = F^{-1}(t)$ *(best fixed block size results).*

	%cens	25, $\tau = 0.8$			50, $\tau = 0.65$		
		AN	EL	BEL	AN	EL	BEL
Model 1.	coverage	0.944	0.950	0.952	0.930	0.943	0.950
	length	*0.123*	*0.123*	*0.123*	*0.104*	*0.104*	*0.106*
Model 2.	coverage	0.913	0.922	0.930	0.922	0.924	0.932
	length	*0.170*	*0.168*	*0.170*	*0.135*	*0.132*	*0.136*

Table 3: 95% *confidence intervals for* $\int_0^\tau tdF(t)$ *(best fixed block size results).*

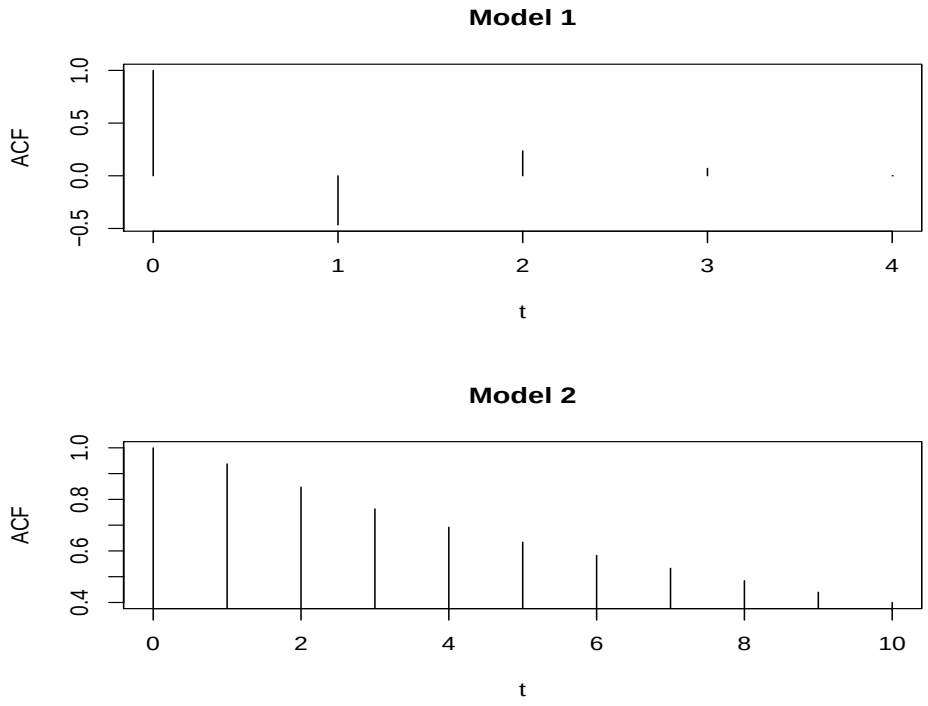

Figure 1: Theoretical auto-correlation-function of Model 1 and Model 2.

Model 1	AN	EL			BEL		
		b=1	b=5	b=10	b=15	b=20	b=25
25% of censoring							
l=5	0.934	0.935	0.941	0.943	0.942	0.944	0.947
l=10	0.933	0.934	0.940	0.939	0.939	0.943	0.950
l=15	0.926	0.925	0.930	0.932	0.935	0.936	0.940
l=20	0.920	0.919	0.923	0.925	0.930	0.927	0.930
l=25	0.917	0.917	0.913	0.920	0.924	0.925	0.923
l=30	0.911	0.911	0.910	0.913	0.918	0.915	0.915
l=35	0.904	0.905	0.908	0.910	0.912	0.911	0.911
50% of censoring							
l=5	0.947	0.948	0.948	0.948	0.950	0.948	0.948
l=10	0.942	0.943	0.945	0.946	0.950	0.948	0.948
l=15	0.937	0.937	0.940	0.944	0.946	0.948	0.946
l=20	0.932	0.930	0.935	0.935	0.941	0.942	0.941
l=25	0.923	0.924	0.924	0.928	0.931	0.934	0.932
l=30	0.918	0.917	0.918	0.925	0.929	0.929	0.928
l=35	0.915	0.915	0.917	0.921	0.925	0.925	0.922

Model 2	AN	EL			BEL		
		b=1	b=5	b=10	b=15	b=20	b=25
25% of censoring							
l=5	0.814	0.820	0.820	0.820	0.824	0.822	0.821
l=10	0.867	0.872	0.880	0.880	0.880	0.879	0.876
l=15	0.889	0.896	0.897	0.896	0.890	0.893	0.893
l=20	0.892	0.899	0.899	0.898	0.896	0.895	0.895
l=25	0.890	0.898	0.900	0.897	0.895	0.893	0.891
l=30	0.886	0.893	0.895	0.896	0.893	0.891	0.890
l=35	0.884	0.889	0.894	0.894	0.891	0.887	0.885
50% of censoring							
l=5	0.863	0.864	0.868	0.868	0.865	0.863	0.863
l=10	0.900	0.909	0.900	0.906	0.903	0.905	0.904
l=15	0.912	0.914	0.920	0.913	0.916	0.916	0.913
l=20	0.916	0.917	0.914	0.915	0.918	0.916	0.916
l=25	0.910	0.914	0.914	0.915	0.913	0.917	0.917
l=30	0.914	0.914	0.914	0.915	0.913	0.914	0.915
l=35	0.915	0.912	0.911	0.915	0.914	0.912	0.913

Table 4: 95% *confidence intervals for* $F(x)$ *at* $x = F^{-1}(0.5)$ *for different value of* l *and* b *using the BJ variance estimator.*

6 Blocksize choice

In practice we need to choose a block length to compute any of our confidence intervals. However, it is known that choosing a blocksize is not an easy task in inference with dependent data. For more discussion of this issue we refer to Politis and White (2004), Zvingelis (2001) and the references given in those papers. To the best of our knowledge, there are no guidelines in the literature about how to select a blocksize in the case of censored data. Here we propose to select l and b by a *data-driven* procedure, using an idea from subsampling theory due to Politis et al. (1997). We give preference to that procedure for its simplicity. It does in fact not require any bootstrapping or subsampling. The main idea behind this method is to select a blocksize in a suitable range. For any value of (l, b) in this range, one may hope to get a CI $I_{l,b}$ almost close to the best possible obtained by using the optimal blocksize. In other words we will look for a (l^*, b^*) around which small changes will be observed in the confidence intervals. This idea translates into the following algorithm.

ALGORITHM:

1. Fix intervals $[l_{small}, l_{big}]$ and $[b_{small}, b_{big}]$ in which l^* and b^* will be determined.

2. For each (l, b) from a grid $\{l_{small}, \ldots, l_{big}\} * \{b_{small}, \ldots, b_{big}\}$, compute the confidence interval and denote it by $I_{l,b} = [I^{low}_{l,b}, I^{up}_{l,b}]$.

3. For each fixed value (l, b) calculate $VI_{l,b}$, which is the sum of the standard deviation of $\left\{I^{low}_{l-k,b}, \ldots, I^{low}_{l+k,b}, I^{low}_{l,b-k}, \ldots, I^{low}_{l,b+k}\right\}$ and the standard deviation of $\left\{I^{up}_{l-k,b}, \ldots, I^{up}_{l+k,b}, I^{up}_{l,b-k}, \ldots, I^{up}_{l,b+k}\right\}$.

4. Choose (l^*, b^*) corresponding to the smallest value of $(l + b)^s VI_{l,b}$, for some fixed s.

This is a generalization of the original algorithm of Politis et al. (1997) in the sense that the data-driven procedure is used to choose l and b simultaneously. Note that we also multiply the volatility index $VI_{l,b}$ by $(l+b)^s$ with typically $s = 1$ or $s = 2$ in order to avoid selecting a large value of l and b. However, even with $s = 0$ the algorithm still gives reasonable results. For the simulation, we take $l_{small} = 1$, $l_{big} = 35$, $b_{small} = 1$, $b_{big} = 25$, $k = 2$ for AN and $k = 1$ for BEL. For each scenario, this procedure was replicated 1500 times and the results are shown in Table 5 for the df and the truncated mean (only results based on BJ variance estimator are shown). By comparing these results with those of Table 1, 2 and 3 we can observe that the difference in

the coverage probability is about 5% on average for the df and also for the truncated mean.

				Model 1		**Model 2**	
				AN	BEL	AN	BEL
Distr. funct.	**t**	**%cens**					
	0.2	25	coverage	0.924	0.928	0.868	0.893
			length	*0.086*	*0.086*	*0.183*	*0.183*
	0.5	25	coverage	0.927	0.938	0.894	0.902
			length	*0.103*	*0.104*	*0.246*	*0.244*
		50	coverage	0.937	0.946	0.915	0.918
			length	*0.118*	*0.119*	*0.257*	*0.256*
		70	coverage	0.934	0.943	0.955	0.952
			length	*0.181*	*0.192*	*0.315*	*0.319*
	0.7	25	coverage	0.930	0.933	0.891	0.900
			length	*0.105*	*0.105*	*0.228*	*0.227*
Trunc. mean	τ	**%cens**					
	0.8	25	coverage	0.946	0.950	0.911	0.920
			length	*0.122*	*0.123*	*0.168*	*0.170*
	0.65	50	coverage	0.936	0.940	0.924	0.930
			length	*0.104*	*0.104*	*0.136*	*0.138*

Table 5: 95% *confidence intervals using the data-driven procedure and the BJ variance estimator.*

Acknowledgments

The authors are grateful to Eckhard Liebsher for his help with mixing data and also to Peter Hall for some helpful discussions.

References

Arcones, M. A. (1995). On the central limit theorem for U-statistics under absolute regularity. *Statist. Probab. Lett. 24*, 245–249.

Arcones, M. A. (1998). The law of large numbers for U-statistics under absolute regularity. *Electron. Comm. Probab. 3*, 13–19 (electronic).

Bougerol, P. and N. Picard (1992). Strict stationarity of generalized autoregressive processes. *Ann. Probab. 20*.

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in Probability and Statistics: A Survey of Recent Results*, pp. 165–192. Birkhuser, Boston (eds. E. Eberlein and M. S. Taqqu).

Cai, Z. (2001). Estimating a distribution function for censored time series data. *J. Multivariate Anal. 78*, 299–318.

Cai, Z. W. and G. G. Roussas (1992). Uniform strong estimation under α-mixing, with rates. *Statist. Probab. Lett. 15*, 47–55.

Chen, S., G. B. Dahl, and S. Khan (2005). Nonparametric identification and estimation of a censored location-scale regression model. *J. Amer. Statist. Assoc. 100*, 212–221.

Doukhan, P. (1994). *Mixing: Properties and examples.* Lecture Notes in Statistics. Springer-Verlag.

Eriksson, B. and R. Adell (1994). On the analysis of life tables for dependent observations. *Statistics in Medicine 13*, 43–51.

Gijbels, I. and N. Veraverbeke (1991). Almost sure asymptotic representation for a class of functionals of the Kaplan-Meier estimator. *Ann. Statist. 19*, 1457–1470.

Hall, P. and B. La Scala (1990). Methodology and algorithms of empirical likelihood. *International statistical review 58*, 109–127.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist. 25*, 2084–2102.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist. 17*, 1217–1241.

Liu, R. Y. and K. Singh (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap (East Lansing, MI, 1990)*, pp. 225–248. Wiley.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*, 237–249.

Owen, A. B. (2001). *Empirical Likelihood.* Chapman and Hall/CRC.

Pham, T. D. and L. T. Tran (1985). Some mixing properties of time series models. *Stochastic Process. Appl. 19*, 297–303.

Politis, D. N., J. P. Romano, and M. Wolf (1997). Subsampling for heteroskedastic time series. *J. Econometrics 81*.

Politis, D. N. and H. White (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews 23*, 53–70.

Radulović, D. (1996). The bootstrap of the mean for strong mixing sequences under minimal conditions. *Statist. Probab. Lett. 28*.

Rio, E. (2000). *Theorie asymptotique des processus aleatoires faiblement dependants*. Springer-Verlag.

Shorack, G. R. and J. A. Wellner (1986). *Empirical Processes with Applications to Statistics*. Wiley.

Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist. 23*, 422–439.

Stute, W. (1996). The jackknife estimate of variance of a Kaplan-Meier integral. *Ann. Statist. 24*, 2679–2704.

Stute, W. and J.-L. Wang (1993). The strong law under random censorship. *Ann. Statist. 21*, 1591–1607.

Thomas, D. R. and G. L. Grunkemeier (1975). Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc. 70*, 865–871.

Wang, Q. and B. Jing (2001). Empirical likelihood for a class of functionals of survival distribution with censored data. *Ann. Inst. Statist. Math. 53*, 517–527.

Ying, Z. and L. J. Wei (1994). The Kaplan-Meier estimate for dependent failure time observations. *J. of Multivariate Analysis 50*, 17–29.

Zvingelis, J. J. (2001). On bootstrap coverage probability with dependent data. In *Computer-Aided Econometrics*.