Econometric Theory, **27**, 2011, 178–198.
doi:10.1017/S0266466610000162

EMPIRICAL LIKELIHOOD CONFIDENCE INTERVALS FOR DEPENDENT DURATION DATA

ANOUAR EL GHOUCH AND INGRID VAN KEILEGOM
Université catholique de Louvain

IAN W. MCKEAGUE
Columbia University

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 block size is also discussed.

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, for example, may be right censored and are typically correlated. Such dependent censored data occur, for example, when study participants belong to clusters (e.g., month of unemployment, job type, neighborhood, school), with members of the same cluster having correlated risk of the event of interest. Chen, Dahl, and Khan (2005) discuss an example involving the number of weeks that an individual would like to collect unemployment benefits. In a duration analysis of migration dynamics, Bijwaard (2004) tracked migration to the Netherlands between 1995 and 2003. This dynamic process is right censored for individuals who emigrate from the Netherlands during the period of the study. See also Brooks, Faff, and Fry (2001) and Franses and Paap (2002) for more studies related to this area. Other examples can be found in medical follow-up studies, epidemiology, and reliability. See Eriksson and Adell

The research was supported by Belgian IAP research network grant nr. P5/24. The authors are grateful to Eckhard Liebsher for his help with mixing data and also to Peter Hall for some helpful discussions. Address correspondence to Ingrid Van Keilegom, Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium; e-mail: Ingrid.Vankeilegom@uclouvain.be.

 © Cambridge University Press 2010

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

(1994) and Ying and Wei (1994) for some concrete examples. Conventional analyses of censored 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 $\delta_i = I(X_i \le Y_i)$ 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

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

where $Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)}$ are the order statistics of Z_i and $\delta_{(1)}, \ldots, \delta_{(n)}$ are the corresponding δ_i. As usual, if the last observation is censored, then we consider $\hat{F}(t)$ to be 1 for t larger than the maximum censoring time.

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 $\xi(t) = I(t \le t_0)$, then $\theta = F(t_0)$, and if $\xi(t) = t$, then $\theta = E(X)$. We refer to Stute and Wang (1993) for other examples.

For independent and identically distributed (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

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

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 we apply the assumption below.

Assumption A1.

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

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) using 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. All proofs are collected in the Appendix.

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

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

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

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 ϕ-mixing. In the sequel we use Assumption A.2, which follows.

Assumption A2. Assume $\{X_i\}$ 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:

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

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

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

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

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

$$\textit{where } \eta_i := \eta_i(F, G) = U_i + \gamma_1(Z_i)(1 - \delta_i) - \gamma_2(Z_i). \tag{3}$$

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 a CLT for α-mixing sequences; see, for example, Rio (2000), and note that β-mixing implies α-mixing.

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

COROLLARY 1. *Under the assumptions of Theorem 1,*

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

in distribution, with $\sigma_\eta^2 = \mathrm{Var}(\eta_1) + 2\sum_{i>1}\mathrm{Cov}(\eta_1, \eta_i)$.

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

Corollary 1 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 that is available for a general ξ. To motivate our approach, note that

$$\sigma_\eta^2 = \lim_{n\to\infty} \mathrm{Var}\left(n^{-1/2}\sum_{i=1}^{n}\eta_i\right),$$

and η_i is absolutely regular. Given the success of the moving-block jackknife (BJ) for variance estimation with dependent data (see Künsch, 1989; 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 $L := L(n) = n - l + 1$ 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 Künsch, 1989, Thm. 3.4), and converges to σ_η^2 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 $\hat{F}$ 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

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

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

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

3. EMPIRICAL LIKELIHOOD

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

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

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

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

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 $\mathcal{L}_n(\theta)$, we need the lemma below.

LEMMA 1. *Under the assumptions of Theorem 1,*

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

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

The next theorem provides the asymptotic distribution of $\mathcal{L}_n(\theta)$.

THEOREM 2. *Under the assumptions of Theorem 1,*

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

Remark. If G were known, then it is easy to see that $n^{-1/2}\sum_{i=1}^{n}(U_i - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_U^2)$, with $\sigma_U^2 = \mathrm{Var}(U_1) + 2\sum_{i>1}\mathrm{Cov}(U_1, U_{i+1})$. On the other hand, $n^{-1}\sum_{i=1}^{n}(U_i - \theta)^2$ converges in probability to $\mathrm{Var}(U_1)$ instead of the desired σ_U^2, as in the i.i.d. case. In other words, $n^{-1}\sum_{i=1}^{n}(U_i - \theta)^2$ is a "wrong metric"

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

for $n^{-1/2}\sum_{i=1}^{n}(U_i-\theta)$. This implies that, for uncensored weakly dependent data (namely, U_i, $i=1,\ldots,n$), in order to obtain a χ_1^2 limiting distribution, we must adjust the EL statistic by a factor of $\sigma_U^{-2}\mathrm{Var}(U_1)$. However, when censoring is present, we have to work with the V_i's instead of the U_i's, and $V_1,\ldots,V_n$ are no longer stationary mixing random variables, due to the estimation of G by the KM estimator $\hat{G}$. In that case, $n^{-1}\sum_{i=1}^{n}(V_i-\theta)^2$ still converges to $\mathrm{Var}(U_1)$, but the limit of $n^{-1/2}\sum_{i=1}^{n}(V_i-\theta)$ is now $\mathcal{N}\left(0,\sigma_\eta^2\right)$. Thus a correction for the adjusted factor is needed.

Since $\hat{\theta}$ is a consistent estimator of θ, from Lemma 1(ii), we can consistently estimate $\mathrm{Var}(U_1)$ 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:

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

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 $b:=b_n$ 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 $\{\bar{V}_{i,b}\}_{1\le i\le N}$. The estimated blockwise EL ratio at θ is

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

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 $\lambda_{n,b}$, respectively.

We need the lemma below to prove Theorem 3, which gives the asymptotic distribution of $\mathcal{L}_{n,b}(\theta)$.

LEMMA 2. *Under the assumptions of Theorem 1,*

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

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

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

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

THEOREM 3. *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* $\sigma_U^2 = \mathrm{Var}(U_1) + 2\sum_{i>1}\mathrm{Cov}(U_1, U_i)$.

We omit the proof of Theorem 3, as it follows the same steps as the proof of Theorem 2.

The proposed BEL confidence interval is given by

$$\left\{\theta : r_n \frac{bN^{-1}\sum_{i=1}^{N}\left(\overline{V}_{i,b} - N^{-1}\sum_{i=1}^{N}\overline{V}_{i,b}\right)^2}{\hat{\sigma}^2_{\hat{\eta},l}}\mathcal{L}_{n,b}(\theta) \le \chi_1^2(\alpha)\right\}. \qquad \text{(BEL)}$$

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.
- Except for the last part of the proof of Theorem 1 involving U-statistic theory, where we need the β-mixing assumption to apply Lemma 2 of Arcones (1998), all the proofs and results of this paper still work under the strong mixing condition (i.e., α-mixing) without any modification.

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, and BEL). Two functionals of the survival function are investigated:

- $\xi(x) = I(x \le t)$; i.e., $\theta = F(t)$ the df at a given t.
- $\xi(x) = xI(x \le \tau)$; i.e., $\theta = \int_{-\infty}^{\tau} x dF(x)$ 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 Cai, 2001, eqn. (10)).

To generate our data, we first consider an autoregressive moving average (ARMA) time series of the form

$$Z_t = \sum_i \alpha_i Z_{t-i} + \sum_i \gamma_i \epsilon_{t-i} + \epsilon_t,$$

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

where the ϵ_t are i.i.d. $\mathcal{N}(0,1)$. By an appropriate choice of α_i's and γ_i's, the resulting Z_t is a strictly stationary Gaussian process and absolutely regular, with $\beta(n) \to 0$ at an exponential rate (see Pham and Tran, 1985; Bougerol and Picard, 1992). Then, in order to get a survival (censoring) time that is β-mixing and with a given distribution $F(G)$, we use the probability integral transform method (see Hoel, Port, and Stone, 1971). We carried out simulations for numerous combinations of survival and censoring distributions, but for the sake of brevity we will only report our results in two cases.

Simulation 1
The survival distribution is the standard exponential, and censoring is uniform on $[0, c]$. The value of c is determined to achieve some prespecified censoring rate. The data were generated either from an MA(3), with $(\gamma_1, \gamma_2, \gamma_3) = (4.5, -3.1, 2.7)$, say Model 1, or from an ARMA(3, 3), with $(\alpha_1, \alpha_2, \alpha_3) = (1.7, -1.3, 0.45)$ and $(\gamma_1, \gamma_2, \gamma_3) = (4.5, -3.1, 2.7)$, say Model 2. Clearly, the dependence under Model 2 is stronger than the dependence under Model 1. This can be seen from the autocorrection function (ACF) of each model (see Figure 1). The sample size is $n = 300$.

Simulation 2
The survival time was generated from the bimodal density

$$f = 0.8 f_1 + 0.2 f_2,$$

where f_1 is the density of $\exp(Z/2)$, with Z being $\mathcal{N}(0,1)$, and f_2 is the density of $\mathcal{N}(2, 0.17)$. The censoring is exponential, with a mean value that was chosen according to the desired percentage of censoring. We simulate our data from an AR(1), with $\gamma_1 = -0.8$. The sample size is $n = 150$.

To calculate our 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 also need 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 1,500 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 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 systematically gives 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 poorly, especially with the AN approach. In this case there is a considerable improvement with the BEL method.

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

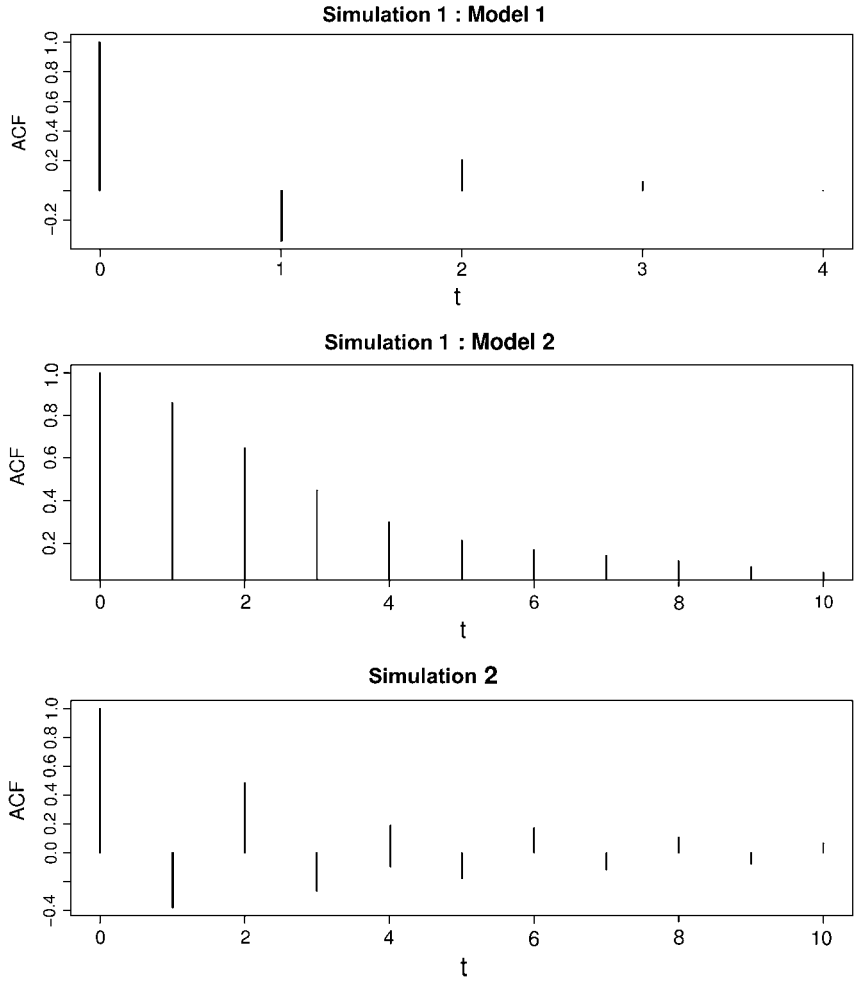


FIGURE 1. Estimated autocorrelation function of the survival time.

Note also that the performance of the CIs depends also on the censoring rate. Due essentially to the estimated variance, the CIs become too large as the censoring rate increases. This explains the fact that in Table 1 we observe better coverage accuracy for data with high censoring rate. Generally, the coverage probability gets closer to the nominal coverage as we pass from AN to EL and from EL to BEL. The advantage of using the (B)EL approach becomes clear in Simulation 2 (see Table 2), where we get better results by using the (B)EL method. This might be due to the fact that the sample size is relatively small.

For the truncated mean, first note that we have taken a different value of τ for the different censoring rates ($\tau = F^{-1}(0.79)$ and $\tau = F^{-1}(0.65)$ for 25% and

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

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

			AN		EL		BEL	
t	% cens	Var	Cai	BJ	Cai	BJ	Cai	BJ
Model 1								
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	0.951
		length	*0.117*	*0.117*	*0.117*	*0.118*	*0.121*	*0.120*
	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*
Model 2								
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*

50% censoring, respectively). This is natural, since we cannot hope to do very well with high censoring. From Tables 3 and 4 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.

Remark. Another way to estimate σ_η^2 is to make use of the Newey and West (1987) heteroskedastic and autocorrelation consistent (HAC) estimator of the variance that can be written as

$$\hat{\gamma}_{\hat{\eta},0} + 2\sum_{j=1}^{l}(1 - j/(1+l))\hat{\gamma}_{\hat{\eta},j},$$

where $\hat{\gamma}_{\hat{\eta},j}$, $j = 0, \ldots, l$, denote the empirical autocovariances of the process $\hat{\eta}_t$ and l is a user-chosen truncation lag. All the simulations done here were also performed using the HAC variance estimator with an l that ranges from 1 to 35. For the mean lifetime in both Simulations 1 and 2, the results are very much the

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

TABLE 2. (Simulation 2) 95% confidence interval for $F(x)$ at $x = F^{-1}(t)$ (best fixed block (lag) size results)

			AN		EL		BEL	
t	% cens	Var	HAC	BJ	HAC	BJ	HAC	BJ
0.2	25	coverage	0.910	0.925	0.917	0.936	0.922	0.940
		length	*0.130*	*0.135*	*0.129*	*0.135*	*0.123*	*0.135*
0.5	25	coverage	0.936	0.956	0.937	0.956	0.944	0.960
		length	*0.159*	*0.176*	*0.158*	*0.175*	*0.166*	*0.178*
	50	coverage	0.929	0.941	0.933	0.947	0.937	0.950
		length	*0.197*	*0.205*	*0.196*	*0.204*	*0.202*	*0.211*
	70	coverage	0.931	0.931	0.937	0.937	0.944	0.943
		length	*0.260*	*0.261*	*0.260*	*0.260*	*0.280*	*0.275*
0.7	25	coverage	0.926	0.930	0.928	0.932	0.934	0.940
		length	*0.165*	*0.168*	*0.164*	*0.167*	*0.172*	*0.172*

TABLE 3. (Simulation 1) 95% confidence intervals for $\int_0^\tau t\,dF(t)$ (best fixed block (lag) size results)

	% cens	25			50		
		AN	EL	BEL	AN	EL	BEL
Var = BJ							
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*
Var = HAC							
Model 1	coverage	0.945	0.950	0.952	0.936	0.943	0.950
	length	*0.123*	*0.123*	*0.123*	*0.104*	*0.104*	*0.105*
Model 2	coverage	0.914	0.924	0.928	0.922	0.925	0.932
	length	*0.170*	*0.170*	*0.171*	*0.135*	*0.133*	*0.136*

same as those obtained via the BJ procedure; see Tables 3 and 4. This is also the case for the survival function in Simulation 1 (for brevity, the results are not shown here). However, in Simulation 2 the BJ procedure generally leads, to better results, as can be seen from Table 2.

Another objective of our simulations was to study the effect of the block size. Table 5 provides an illustration of this. The performance of the CIs depends rather critically on the choice of the block size. Typically, choosing an inappropriate block length leads to undercoverage, although we did find some cases (not shown

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

TABLE 4. (Simulation 2) 95% confidence intervals for $\int_0^\tau t\,dF(t)$ (best fixed block (lag) size results)

	% cens	25			50		
		AN	EL	BEL	AN	EL	BEL
Var = BJ	coverage	0.928	0.930	0.932	0.914	0.924	0.930
	length	*0.231*	*0.231*	*0.233*	*0.244*	*0.245*	*0.254*
Var = HAC	coverage	0.930	0.931	0.931	0.914	0.923	0.931
	length	*0.231*	*0.230*	*0.233*	*0.244*	*0.245*	*0.254*

TABLE 5. (Simulation 1, Model 2) 95% confidence intervals for $F(x)$ at $x = F^{-1}(0.5)$ for different value of l and b using the BJ variance estimator

		EL	BEL				
	AN	$b = 1$	$b = 5$	$b = 10$	$b = 15$	$b = 20$	$b = 25$
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

here) of overcoverage. In summary, two things have become clear. First, BEL appears to be more sensitive to the choice of l (the block size of the asymptotic variance estimator) than to the choice of b (the block size 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 block sizes within which the results are close to optimal.

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

TABLE 6. (Simulation 1) 95% confidence intervals using the data-driven procedure and the BJ variance estimator

				Model 1		Model 2	
	t	% cens		AN	BEL	AN	BEL
Distribution function							
	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*
Truncated 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*

6. BLOCK SIZE CHOICE

In practice we need to choose a block length to compute any of our confidence intervals. However, it is known that choosing a block size 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 block size 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, Romano, and Wolf (1997). We give preference to that procedure for its simplicity. It does not in fact require any bootstrapping or subsampling. The main idea behind this method is to select a block size 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 block size. In other words, we will look for a (l^*, b^*) around which small changes will be observed in the CIs. This idea translates into the algorithm below.

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

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 CI and denote it by $I_{l,b} = \left[I_{l,b}^{low}, I_{l,b}^{up}\right]$.
3. For each fixed value (l, b) calculate $VI_{l,b}$, which is the sum of the standard deviation of $\left\{I_{l-k,b}^{low}, \ldots, I_{l+k,b}^{low}, I_{l,b-k}^{low}, \ldots, I_{l,b+k}^{low}\right\}$ and the standard deviation of $\left\{I_{l-k,b}^{up}, \ldots, I_{l+k,b}^{up}, I_{l,b-k}^{up}, \ldots, I_{l,b+k}^{up}\right\}$.
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 1,500 times, and the results are shown in Table 6 for the df and the truncated mean (here we only show the results for Simulation 1 with the BJ variance estimator). By comparing these results with those of Tables 1 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.

7. CONCLUSION

In this paper we have studied three methods of constructing confidence intervals for a general class of smooth functionals of survival distributions under β-mixing and censoring conditions. The first uses the classical Wald approach (AN), and extends results of Cai (2001) beyond the case of survival functions evaluated at a fixed point. The others (EL and BEL) are based on empirical likelihood methodology. We have proposed a new estimator of the asymptotic variance based on the block jackknife, and we have compared its performance with two other procedures. We have also shown that standard EL can be adapted to dependent data and gives reasonable results, and that the two EL methods perform better than the classical Wald approach. Blockwise EL, which requires the selection of two block sizes instead of one, was found to outperform standard EL. Finally, we investigated the problem of selecting the block size(s) in numerical examples, and we devised an algorithm to provide an automatic selection. Further research is needed to understand how to select optimal block size(s) in this setting.

REFERENCES

Arcones, M. (1995) On the central limit theorem for U-statistics under absolute regularity. *Statistics and Probability Letters* 24, 245–249.

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

Arcones, M. (1998) The law of large numbers for U-statistics under absolute regularity. *Electronic Communications in Probability* 3, 13–19.

Bijwaard, G. (2004) Dynamic economic aspects of migration. *Medium Econometrische Toepassingen*, 26–30.

Bougerol, P. & N. Picard (1992) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1730.

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

Brooks, R., R. Faff, & T. Fry (2001) Garch modelling of individual stock data: The impact of censoring, firm size and trading volume, institutions and money. *Journal of International Financial Markets, Institutions and Money* 11, 215–222.

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

Cai, Z. & G. Roussas (1992) Uniform strong estimation under α-mixing, with rates. *Statistics and Probability Letters* 15, 47–55.

Chen, S., G. Dahl, & S. Khan (2005) Nonparametric identification and estimation of a censored location-scale regression model. *Journal of the American Statistical Association* 100, 212–221.

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

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

Franses, P. & R. Paap (2002) Censored latent effects autoregression, with an application to US unemployment. *Journal of Applied Econometrics* 17, 347–366.

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

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

Hoel, P., S. Port, & C. Stone (1971) *Introduction to Probability Theory.* Houghton Mifflin.

Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent processes. *Annals of Statistics* 25, 2084–2102.

Künsch, H. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.

Liu, R. & K. Singh (1992) Moving blocks jackknife and bootstrap capture weak dependence. In R. Lepage & L. Billard (eds.), *Exploring the Limits of Bootstrap*, pp. 225–248. Wiley.

Newey, W. & K. West (1987) A simple positive definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–705.

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

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

Pham, T. & L. Tran (1985) Some mixing properties of time series models. *Stochastic Processes and their Applications* 19, 297–303.

Politis, D., J. Romano, & M. Wolf (1997) Subsampling for heteroskedastic time series. *Journal of Econometrics* 81, 281–317.

Politis, D. & 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. *Statistics and Probability Letters* 28, 65–72.

Rio, E. (2000) *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*. Springer-Verlag.

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

Stute, W. (1995) The central limit theorem under random censorship. *Annals of Statistics* 23, 422–439.

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

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

Stute, W. & J.-L. Wang (1993) The strong law under random censorship. *Annals of Statistics* 21, 1591–1607.

Thomas, D. & G. Grunkemeier (1975) Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association* 70, 865–871.

Wang, Q. & B. Jing (2001) Empirical likelihood for a class of functionals of survival distribution with censored data. *Annals of the Institute of Statistical Mathematics* 53, 517–527.

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

Zvingelis, J. (2001) On bootstrap coverage probability with dependent data. In D.E.A Giles (ed.), *Computer-Aided Econometrics*. Marcel Dekker.

APPENDIX

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 observations, respectively, that is $H^q(t) = \mathbb{P}(Z_i \le t, \delta_i = q)$, $q = 0, 1$. For any (sub-)df Q, we denote by Q_n the corresponding empirical (sub-)df. The representation for $\int \xi d\hat{F}$ given below is from Lemma 2.1 in Stute (1995). It is based on a purely mathematical derivation, without making use of any probabilistic or statistical arguments, so it continues to hold in the present context.

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

where the last three terms are defined below.

(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 $|B_{in}| \le \frac{1}{2n} \frac{H_n^0(T)}{(1 - H_n(T))^2}$, for all $Z_i \le T$.

So, by the Strong law of large numbers (SLLN) (ergodicity) of strongly mixing sequences, we obtain

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

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

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

Δ_{in} is between the two terms

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

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

By the expansion

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

and applying the law of integrated logarithms (LIL) for empirical (sub-)df (Thm. 3.2 of Cai and Roussas, 1992), we obtain

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

On the other hand, it is easily seen that

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

From these two inequalities and using the SLLN, we obtain

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

(III) $S_n = n^{-1}\sum_{i=1}^n U_i C_{in}$.
From (A.1) we expand S_n into

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

where

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

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 a U-statistic, using the Hoeffding decomposition, and then applying Lemma 3 in Arcones (1998) (see also Arcones, 1995), we get that

$$S_{n1} = \int\int \frac{I(y<x)\xi(x)\gamma_0(x)}{1-H(y)} \Big[dH^0(y)\, dH_n^1(x) + dH_n^0(y)\, dH^1(x) - dH^0(y)\, dH^1(x) \Big] + o_P\left(n^{-1/2}\right). \tag{A.3}$$

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

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

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

$$S_{n2} = \int\int\int \frac{I(y<t, y<x)\xi(x)\gamma_0(x)}{(1-H(y))^2} \times \Big[dH(t)\, dH_n^0(y)\, dH_n^1(x) + dH(t)\, dH_n^0(y)\, dH^1(x) + dH(t)\, dH^0(y)\, dH_n^1(x) - 2\, dH(t)\, dH^0(y)\, dH^1(x) \Big] + o_P\left(n^{-1/2}\right). \tag{A.4}$$

Substituting (A.3) and (A.4) into (A.2), and making some simplifications completes the proof. ■

Proof of Lemma 1.

(i) Note that $V_i = U_i \frac{1-G(Z_i)}{1-\hat{G}(Z_i)}$. Since $\mathrm{E}|U_1|^p < \infty$, 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

$$\sup_{t\le T} \frac{|\hat{G}(t)-G(t)|}{1-\hat{G}(t)} = O_P\left(\sqrt{\frac{\log\log n}{n}}\right). \tag{A.5}$$

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

(ii) We write

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

Observe that

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

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

Proof of Theorem 2. 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), p. 295). Hence Corollary 1 implies that

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

$$n^{-1/2}\sum_{i=1}^{n}(V_i-\theta)\xrightarrow{d}\mathcal{N}\left(0,\sigma_\eta^2\right). \tag{A.6}$$

From the definition of λ_n and using Lemma 1 and (A.6), one can check that

$$\lambda_n = O_P\left(n^{-1/2}\right). \tag{A.7}$$

This together with Lemma 1(i) implies that

$$\lambda_n = \frac{\sum_{i=1}^{n}(V_i-\theta)}{\sum_{i=1}^{n}(V_i-\theta)^2} + o_P\left(n^{-1/2}\right). \tag{A.8}$$

Using a Taylor expansion of $\mathcal{L}_n$, together with (A.7) and Lemma 1, yields

$$\mathcal{L}_n(\theta) = 2\lambda_n\sum_{i=1}^{n}(V_i-\theta) - \lambda_n^2\sum_{i=1}^{n}(V_i-\theta)^2 + o_P(1). \tag{A.9}$$

Substituting (A.8) into (A.9), using again (A.6) and (A.7) together with Lemma 1, 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. ■

Proof of Lemma 2. (i) Note that $\sum_{i=1}^{N}\left(\bar{V}_i-\theta\right) = \sum_{i=1}^{n}(V_i-\theta) - \hat{K}_n$, with, say,

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

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

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

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

$$\begin{aligned} b^2\mathrm{Var}(K_n^1) &= \sum_{j=1}^{b}(b-j)^2\mathrm{Var}(U_1) + 2\sum_{i=1}^{b-1}\sum_{j=1}^{b-i}(b-i)(b-i-j)\mathrm{Cov}(U_1,U_{j+1}) \\ &\le b^3\mathrm{Var}(U_1) + 2b^3\sum_{i=1}^{b}\left|\mathrm{Cov}\left(U_1,U_{i+1}\right)\right|. \end{aligned}$$

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

$$\sum_{i=1}^{b}\left|\mathrm{Cov}\left(U_1,U_{i+1}\right)\right| = O\left(\sum_{n\ge1}\beta(n)^{1-2/p}\right) = O(1).$$

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press

So, $\mathrm{Var}\left(n^{-1/2}K_n^1\right) = O(n^{-1}b)$, 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 (A.6) 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 1(i).

(iii) We write

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

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

$$bN^{-1}\sum_{i=1}^{N}\left(\bar{V}_i - \bar{U}_i\right)^2 \leq \left(\sup_{t\leq T}\frac{|\hat{G}(t) - G(t)|}{1-\hat{G}(t)}\right)^2 bN^{-1}\sum_{i=1}^{N}\left(b^{-1}\sum_{j=i}^{i+b-1}|U_j|\right)^2. \quad \textbf{(A.10)}$$

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

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

This together with (A.5) implies that the left-hand side of (A.10) 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. ■

https://doi.org/10.1017/S0266466610000162 Published online by Cambridge University Press