The Annals of Statistics
2003, Vol. 31, No. 2, 442–463

ESTIMATION IN A COX REGRESSION MODEL WITH A CHANGE-POINT ACCORDING TO A THRESHOLD IN A COVARIATE

BY ODILE PONS

INRA and University of Paris V

We consider a nonregular Cox model for independent and identically distributed right censored survival times, with a change-point according to the unknown threshold of a covariate. The maximum partial likelihood estimators of the parameters and the estimator of the baseline cumulative hazard are studied. We prove that the estimator of the change-point is n-consistent and the estimator of the regression parameters are $n^{1/2}$-consistent, and we establish the asymptotic distributions of the estimators. The estimators of the regression parameters and of the baseline cumulative hazard are adaptive in the sense that they do not depend on the knowledge of the change-point.

1. Introduction. The proportional hazards regression model introduced by Cox (1972) assumes that conditionally on a vector of covariates Z, the hazard function of a survival time is $\lambda(t \mid Z) = \lambda(t)\exp\{\beta^T Z(t)\}$ at $t \geq 0$, where β is a vector of unknown regression parameters and λ is an unknown and unspecified baseline hazard function. Inference on the regression parameters is based on a partial likelihood and the asymptotic properties of the estimators of β and of the cumulative hazard function gave rise to many papers, among them Cox (1975), Tsiatis (1981), Næs (1982), Bailey (1983) for time-independent covariates, and Andersen and Gill (1982) and Prentice and Self (1983) in a more general set-up. In data analysis, the assumption of proportional hazards is not always relevant in the whole range of a covariate and the covariate may be dichotomized to define new variables satisfying this assumption [Kleinbaum (1996)]. This procedure led to a two-phase Cox model with a change-point according to a threshold that may be fixed or estimated from the data. Several authors also considered a nonregular Cox model involving a two-phase regression on time-dependent covariates, with a change-point at an unknown time [Liang, Self and Liu (1990), Luo, Turnbull and Clark (1997) and Luo (1996)].

The aim of the present paper is to study the asymptotic behavior of the maximum partial likelihood estimator of the parameters in a nonregular Cox model with a change-point according to the unknown threshold of a covariate. Let $Z = (Z_1^T, Z_2^T, Z_3)^T$ be a vector of covariates, where Z_1 and Z_2 are respectively p and q-dimensional left-continuous processes with right-hand limits and Z_3 is a

Received May 1999; revised June 2002.
AMS 2000 subject classifications. 62F12, 62G05, 62M09.
Key words and phrases. Asymptotic distribution, change-point, Cox regression model, hazard function, right censoring.

one-dimensional random variable. We assume that conditionally on Z the hazard rate of a survival time T^0 has the form

$$\lambda_\theta(t \mid Z) = \lambda(t)\exp\{r_\theta(Z(t))\} \tag{1.1}$$

with

$$r_\theta(Z(t)) = \alpha^T Z_1(t) + \beta^T Z_2(t)\mathbb{1}_{\{Z_3 \le \zeta\}} + \gamma^T Z_2(t)\mathbb{1}_{\{Z_3 > \zeta\}},$$

where $\theta = (\zeta, \xi^T)^T$, with $\xi = (\alpha^T, \beta^T, \gamma^T)^T$ the vector of the regression parameters, and λ is an unknown baseline hazard function. Here the regression parameters α, β and γ belong respectively to bounded subsets of $\mathbb{R}^p$ and $\mathbb{R}^q$ and the threshold ζ is a parameter lying in a bounded interval $[\zeta_1, \zeta_2]$ strictly included in the support of Z_3. The true parameter values θ_0 and λ_0 are supposed to be identifiable, that is, θ_0 is such that $\beta_0 \ne \gamma_0$ and a change-point actually occurs at ζ_0. We suppose that the survival time T^0 with hazard function (1.1) may be right-censored at a noninformative censoring time C such that C is independent of T^0 conditionally on Z. We observe the censored time $T = T^0 \wedge C$ and the censoring indicator $\delta = \mathbb{1}_{\{T^0 \le C\}}$.

In the same framework, Luo and Boyett (1997) studied a model where a constant is added to the regression on a covariate Z_1 after a change-point in another variable Z_2, $r_\theta(Z(t)) = \alpha^T Z_1(t) + \beta\mathbb{1}_{\{Z_2 \le \zeta\}}$. They proved the consistency of the maximum partial likelihood estimators and applied the results to a clinical data set of patients with leukemia. Jespersen (1986) studied a test for no change-point in the submodel $r_\theta(Z) = \beta\mathbb{1}_{\{Z \le \zeta\}}$ of (1.1) and investigated risk factors for breast cancer with a threshold in the effect of estrogen receptors. Several other applications of such models may also be found in the literature, for example, a study of the effect of tumor thickness on survival with melanoma in Andersen, Borgan, Gill and Keiding (1993), pages 547–550, and others mentioned by Luo and Boyett (1997). Model (1.1) extends these models by taking into account the smallest value of a variable Z_3 having an interacting effect on covariates Z_2 in a Cox model, as in the linear models with a change in regression coefficients.

Inference will be based on a sample $(T_i, \delta_i, Z_i)_{1 \le i \le n}$ of n independent and identically distributed observations. As in the classical Cox model for i.i.d. individuals, we assume that the variables T_i are observed on a time interval $[0, \tau]$ such that $\Pr(T \ge \tau) > 0$ [Andersen and Gill (1982), Theorem 4.1]. In the model (1.1), θ_0 is estimated by the value $\widehat{\theta}_n$ that maximizes the partial likelihood

$$L_n(\theta) = \prod_{i \le n}\left\{\frac{\exp\{r_\theta(Z_i(T_i))\}}{\sum_j Y_j(T_i)\exp\{r_\theta(Z_j(T_i))\}}\right\}^{\delta_i}$$

where $Y_i(t) = \mathbb{1}_{\{T_i \ge t\}}$ indicates whether individual i is still under observation at t. Let $S_n^{(0)}(t;\theta) = \sum_{i \le n} Y_i(t)\exp\{r_\theta(Z_i(t))\}$. The logarithm of the partial likelihood $l_n = \log L_n$ is written

$$l_n(\theta) = \sum_{i \le n} \delta_i\big\{r_\theta(Z_i(T_i)) - \log S_n^{(0)}(T_i;\theta)\big\}. \tag{1.2}$$

The estimator $\widehat{\theta}_n$ is obtained in the following way: For fixed ζ, let $\widehat{\xi}_n(\zeta) = \arg\max_{\xi\in\Xi} l_n(\zeta,\xi)$ and $l_n(\zeta) = l_n(\zeta, \widehat{\xi}_n(\zeta))$. Then ζ_0 is estimated by $\widehat{\zeta}_n$ which satisfies the relationship

$$\widehat{\zeta}_n = \inf\left\{\zeta \in [\zeta_1, \zeta_2] : \max\{l_n(\zeta^-), l_n(\zeta)\} = \sup_{\zeta\in[\zeta_1,\zeta_2]} l_n(\zeta)\right\},$$

where $l_n(\zeta^-)$ denotes the left-hand limit of l_n at ζ. The maximum likelihood estimator of ξ_0 satisfies $\widehat{\xi}_n = \widehat{\xi}_n(\widehat{\zeta}_n)$, and $\widehat{\theta}_n = (\widehat{\zeta}_n, \widehat{\xi}_n)$. The cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$ is estimated as in Breslow (1972) by

$$\widehat{\Lambda}_n(t) = \int_0^t \frac{d\bar{N}_n(s)}{S_n^{(0)}(s;\widehat{\theta}_n)}.$$

In the two-phase linear regression models with a change-point over time and Gaussian errors, a standard approach consists in indexing the observations according to time, considered as fixed, and in estimating the change-point by the proportion of data in the first phase of the regression [Csörgő and Horváth (1997)]. In such regression models and for Poisson processes with a change-point in the hazard rate [Nguyen, Rogers and Walker (1984) and Kutoyants (1984)], maximum likelihood inference is classically based on random walks which appear in a factorization of the likelihood as a product of terms for individuals in each phase of the model. The Cox model (1.1) involves a nonparametric function λ_0 and θ is estimated by maximization of the partial likelihood L_n which cannot be simply related to random walks because all the individual contributions involve the process $S_n^{(0)}$, and they are therefore all dependent and it is not possible to split (1.2) into terms for individuals with $Z_{3i} \le \zeta_0$ and individuals with $Z_{3i} > \zeta_0$. Here we follow the approach of Ibragimov and Has'minskii (1981) for the parameters of a density with jumps, as in Kutoyants (1998) for change-points in nonhomogeneous Poisson processes.

Assumptions and notation for the asymptotic properties of the estimators are given in the following section. In Section 3 we establish the consistency and the convergence rate of the estimators. In the nonregular model (1.1), the convergence rate derives from the asymptotic behavior of the process $u \mapsto \{l_n(\theta_{n,u}) - l_n(\theta_0)\}$, with $\theta_{n,u} = (\zeta_0 + n^{-1}u_1, \xi_0 + n^{-1/2}u_2)$ for $u = (u_1, u_2)$, u_1 in $\mathbb{R}$, u_2 in $\mathbb{R}^{p+2q}$. We show that it is asymptotically bounded in probability, which entails that $\widehat{\zeta}_n$ is n-consistent and $\widehat{\xi}_n$ is $n^{1/2}$-consistent. Section 4 presents weak convergence results. They are deduced from the limiting distribution of the process $u \mapsto \{l_n(\theta_{n,u}) - l_n(\theta_0)\}$ on compact sets: $n(\widehat{\zeta}_n - \zeta_0)$ converges weakly to the value $\widehat{v}_Q$ where a jump process reaches its maximum and $\widehat{v}_Q$ is a.s. finite, $n^{1/2}(\widehat{\xi}_n - \xi_0)$ is asymptotically Gaussian, $n(\widehat{\zeta}_n - \zeta_0)$ and $n^{1/2}(\widehat{\xi}_n - \xi_0)$ are asymptotically independent, $n^{1/2}(\widehat{\Lambda}_n - \Lambda_0)$ converges weakly to a Gaussian process. Moreover, $n^{1/2}(\widehat{\xi}_n - \xi_0)$ and $n^{1/2}(\widehat{\Lambda}_n - \Lambda_0)$ are adaptive in the sense that their limiting distribution is the same as if ζ_0 were known (Theorems 5 and 6). This result is important for inference on ξ_0 and Λ_0 in practical applications and it allows us to estimate ζ_0 on a grid in $[\zeta_1, \zeta_2]$ with a path of order smaller than n^{-1}. Technical proofs are detailed in Section 5.

2. Notation and conditions. Let $(\Omega, \mathcal{F}, P_{\theta,\lambda})_{\theta,\lambda}$ be a family of complete probability spaces provided with a history $\mathbb{F} = (\mathcal{F}_t)_t$, where $\mathcal{F}_t \subseteq \mathcal{F}$ is an increasing and right-continuous filtration such that N and Z are $\mathbb{F}$-adapted. We assume that under $P_{\theta,\lambda}$, T^0 satisfies (1.1), C and Z having the same distribution under all probabilities $P_{\theta,\lambda}$. Under the true parameter values, let $P_0 = P_{\theta_0,\lambda_0}$ and let $\mathbb{E}_0$ be the expectation of the random variables. The processes Z_1 and Z_2 have left-continuous sample paths with right-hand limits, with values in sets $\mathcal{Z}_1 \subset \mathbb{R}^p$ and $\mathcal{Z}_2 \subset \mathbb{R}^q$. The random variable Z_3 has its values in $\mathcal{Z}_3$, a subset of $\mathbb{R}$. For t in $[0, \tau]$, $\theta = (\zeta, \xi^T)^T$ and $k = 0, 1, 2$, we denote

$$\widetilde{Z}(t;\zeta) = \big(Z_1^T(t), Z_2^T(t)\mathbb{1}_{\{Z_3 \le \zeta\}}, Z_2^T(t)\mathbb{1}_{\{Z_3 > \zeta\}}\big)^T,$$

$$S_n^{(k)}(t;\theta) = \sum_i Y_i(t)\widetilde{Z}_i^{\otimes k}(t;\zeta)\exp\{r_\theta(Z_i(t))\},$$

where $x^{\otimes 0} = 1$, $x^{\otimes 1} = x$ and $x^{\otimes 2} = xx^T$, for x in $\mathbb{R}^{p+2q}$. For $1 \le i \le n$, let $N_i(t) = \delta_i \mathbb{1}_{\{T_i \le t\}}$ be the counting process of death for individual i and let $M_i(t) = N_i(t) - \int_0^t Y_i(s)\exp\{r_{\theta_0}(Z_i(s))\}\,d\Lambda_0(s)$, a martingale on $[0, \tau]$. We also denote $\bar{N}_n = \sum_{i \le n} N_i$, $\mathbb{M}_n^{(0)}(t) = n^{-1/2}\{\bar{N}_n(t) - \int_0^t S_n^{(0)}(\theta_0)\,d\Lambda_0\}$ and $\mathbb{M}_n^{(1)}(t) = n^{-1/2}\{\sum_i \int_0^t \widetilde{Z}_i(\zeta_0)\,dN_i - \int_0^t S_n^{(1)}(\theta_0)\,d\Lambda_0\} = n^{-1/2}\sum_i \int_0^t \widetilde{Z}_i(\zeta_0)\,dM_i$.

Adapting the notation given in Andersen and Gill (1982), we define

$$s^{(k)}(t;\theta) = \mathbb{E}_0\big[Y_i(t)\widetilde{Z}_i^{\otimes k}(t;\zeta)\exp\{r_\theta(Z_i(t))\}\big],$$

$$V_n(t;\theta) = \big\{S_n^{(2)}S_n^{(0)-1} - \big[S_n^{(1)}S_n^{(0)-1}\big]^{\otimes 2}\big\}(t;\theta),$$

$$v(t;\theta) = \big\{s^{(2)}s^{(0)-1} - \big[s^{(1)}s^{(0)-1}\big]^{\otimes 2}\big\}(t;\theta),$$

$$I(\theta) = \int_0^\tau v(s;\theta)s^{(0)}(s;\theta_0)\lambda_0(s)\,ds.$$

We denote the first p components of $s^{(1)}$ by $s_1^{(1)}(t;\theta) = \mathbb{E}_0[Y_i(t)Z_{1i}(t) \times \exp\{r_\theta(Z_i(t))\}]$. Let also $s_2^{(1)-}(\theta)$ and $s_2^{(1)+}(\theta)$ be the q-dimensional components of $s^{(1)}$ related to the component Z_2 of Z under restrictions on the location of the variable Z_3 with respect to the parameter ζ,

$$\begin{aligned} s_2^{(1)-}(t;\theta) &\equiv s_2^{(1)-}(t;\zeta,\alpha,\beta) \\ &= \mathbb{E}_0\big[Y_i(t)Z_{2i}(t)\mathbb{1}_{\{Z_{3i} \le \zeta\}}\exp\big\{\alpha^T Z_{1i}(t) + \beta^T Z_{2i}(t)\big\}\big], \\ s_2^{(1)+}(t;\theta) &\equiv s_2^{(1)+}(t;\zeta,\alpha,\gamma) \\ &= \mathbb{E}_0\big[Y_i(t)Z_{2i}(t)\mathbb{1}_{\{Z_{3i} > \zeta\}}\exp\big\{\alpha^T Z_{1i}(t) + \gamma^T Z_{2i}(t)\big\}\big]. \end{aligned}$$

For $\zeta < \zeta'$, let $s_2^{(1)}(]\zeta, \zeta'], \alpha, \beta) = s_2^{(1)-}(\zeta', \alpha, \beta) - s_2^{(1)-}(\zeta, \alpha, \beta)$ and $s_2^{(1)}(]\zeta, \zeta'], \alpha, \gamma) = s_2^{(1)+}(\zeta, \alpha, \gamma) - s_2^{(1)+}(\zeta', \alpha, \gamma)$. Similar notation is used for the processes $S_n^{(k)}$,

$$\begin{aligned}
S_n^{(k)-}(t;\theta) &= \sum_i Y_i(t)\widetilde{Z}_i^{\otimes k}(t;\zeta)\mathbb{1}_{\{Z_{3i}\le\zeta\}}\exp\{\alpha^T Z_{1i}(t) + \beta^T Z_{2i}(t)\},\\
S_n^{(k)+}(t;\theta) &= \sum_i Y_i(t)\widetilde{Z}_i^{\otimes k}(t;\zeta)\mathbb{1}_{\{Z_{3i}>\zeta\}}\exp\{\alpha^T Z_{1i}(t) + \gamma^T Z_{2i}(t)\},\\
S_{1n}^{(1)}(t;\theta) &= \sum_i Y_i(t) Z_{1i}(t)\exp\{r_\theta(Z_i(t)\},\\
S_{2n}^{(1)-}(t;\theta) &= \sum_i Y_i(t) Z_{2i}(t)\mathbb{1}_{\{Z_{3i}\le\zeta\}}\exp\{\alpha^T Z_{1i}(t) + \beta^T Z_{2i}(t)\},\\
S_{2n}^{(1)+}(t;\theta) &= \sum_i Y_i(t) Z_{2i}(t)\mathbb{1}_{\{Z_{3i}>\zeta\}}\exp\{\alpha^T Z_{1i}(t) + \gamma^T Z_{2i}(t)\}, \qquad \text{etc.,}
\end{aligned}$$

and $S_n^{(k)-1}$ denotes the inverse of $S_n^{(k)}$.

Using (1.2), the estimator $\widehat{\theta}_n$ maximizes the process

$$\begin{aligned}
X_n(\theta) &= n^{-1}\{l_n(\theta) - l_n(\theta_0)\}\\
&= n^{-1}\sum_{i\le n}\left\{(r_\theta - r_{\theta_0})(Z_i(T_i)) - \log\frac{S_n^{(0)}(T_i;\theta)}{S_n^{(0)}(T_i;\theta_0)}\right\}
\end{aligned} \tag{2.1}$$

and we define the function

$$\begin{aligned}
X(\theta) = \int_0^\tau \bigg\{&(\xi - \xi_0)^T s^{(1)}(\theta_0) + (\beta - \beta_0)^T s_2^{(1)-}(\zeta \wedge \zeta_0, \alpha_0, \beta_0)\\
&+ (\gamma - \gamma_0)^T s_2^{(1)+}(\zeta \vee \zeta_0, \alpha_0, \gamma_0)\\
&+ (\beta - \gamma_0)^T s_2^{(1)}(]\zeta_0, \zeta], \alpha_0, \gamma_0)\\
&+ (\gamma - \beta_0)^T s_2^{(1)}(]\zeta, \zeta_0], \alpha_0, \beta_0) - s^{(0)}(\theta_0)\log\frac{s^{(0)}(\theta)}{s^{(0)}(\theta_0)}\bigg\}\, d\Lambda_0.
\end{aligned} \tag{2.2}$$

The norms in $\mathbb{R}^{p+2q}$ and in $(\mathbb{R}^{p+2q})^{\otimes 2}$ are denoted $\|\cdot\|$. The asymptotic properties of the estimators will be established under the following conditions:

C1. The variable Z_3 has a density h_3 which is strictly positive, bounded and continuous in a neighborhood of ζ_0, $\sup_{t\in[0,\tau]}\lambda_0(t) < \infty$ and $P_0(T \ge \tau) > 0$.

C2. There exists a neighborhood $V(\zeta_0)$ of ζ_0 such that the variance $\operatorname{Var} \widetilde{Z}(t;\zeta)$ is positive definite on $[0,\tau] \times V(\zeta_0)$,

$$\int_0^\tau \mathbb{E}_0 \inf_\beta \Big[Y(t)\{(\beta_0-\gamma_0)^T Z_2(t)\}^2 e^{\alpha_0^T Z_1(t)+\beta^T Z_2(t)} \mid Z_3=\zeta_0 \Big] \, d\Lambda_0 > 0, \tag{2.3}$$

where the infimum is over β between β_0 and γ_0, and there exists a convex and bounded neighborhood Θ of θ_0 such that for $k=0,1,2$,

$$\mathbb{E}_0 \sup_{t\in[0,1]} \sup_{\theta\in\Theta} \{(\|Z_1(t)\|^k + \|Z_2(t)\|^k) e^{r_\theta(Z(t))}\}^2 < \infty, \tag{2.4}$$

$$\sup_{z\in[\zeta_1,\zeta_2]} \mathbb{E}_0 \Big[\sup_{t\in[0,1]} \sup_{\theta\in\Theta} \{(\|Z_1(t)\|^k + \|Z_2(t)\|^k) e^{r_\theta(Z(t))}\}^j \mid Z_3=z \Big] < \infty, \qquad j=1,2,$$

$$\sup_{z,z'} \sup_{t\in[0,1]} \sup_{\theta\in\Theta} \big| \mathbb{E}_0\{e^{r_\theta(Z(t))} \mid Z_3=z\} - \mathbb{E}_0\{e^{r_\theta(Z(t))} \mid Z_3=z'\} \big| \stackrel{|z-z'|\to 0}{\longrightarrow} 0,$$

where z and z' vary in $[\zeta_1,\zeta_2]$ and both z and z' are either larger than ζ_0 or smaller than ζ_0.

C3. The variables $\sup_{t\in[0,1]} \sup_{\theta\in\Theta} \|n^{-1} S_n^{(k)}(t;\theta) - s^{(k)}(t;\theta)\|$ converge a.s. to zero, $k=0,1,2$.

If Z is a random variable, C3 is satisfied by the Glivenko–Cantelli theorem. If Z_1 or Z_2 are processes, it may be proved by the arguments of Theorem 4.1, Appendix III, in Andersen and Gill (1982).

3. Convergence of the estimators. In this section we establish the consistency and the rate of convergence of $\widehat{\zeta}_n$ and $\widehat{\xi}_n$. Luo and Boyett (1997) proved the consistency in their submodel of (1.1) from a local approximation of the process X_n. Here the proof is based on the uniform convergence of X_n to X and on properties of X in the neighborhood of θ_0. The behavior of X follows from the next lemma which ensures properties similar to those of condition D in Andersen and Gill (1982) and the arguments of its proof are the same as in their Theorem 4.1.

LEMMA 1. *Under conditions* C1–C2, $s^{(0)}$ *is bounded away from zero on* $[0,\tau]\times\Theta$, $s^{(1)}(t;\zeta,\xi)$ *and* $s^{(2)}(t;\zeta,\xi)$ *are the first two partial derivatives of* $s^{(0)}(t;\zeta,\xi)$ *with respect to* ξ, *and the functions* $s^{(k)}$ *are continuous on* Θ, *uniformly in* $t\in[0,\tau]$, *for* $k=0,1,2$, *with* $s^{(k)}(t;\theta') - s^{(k)}(t;\theta) = O(|\zeta-\zeta'| + \|\xi-\xi'\|)$ *uniformly on* $[0,\tau]\times\Theta$, *as* $\|\theta-\theta'\|\to 0$. *Moreover, as* $\|\theta-\theta'\|\to 0$,

$$\begin{aligned} s^{(0)}(\theta') - s^{(0)}(\theta) &= (\xi'-\xi)^T s^{(1)}(\theta) + \tfrac{1}{2}(\xi'-\xi)^T s^{(2)}(\theta)(\xi'-\xi) \\ &\quad + (\zeta'-\zeta)\dot{s}_\zeta^{(0)}(\theta) + o(|\zeta-\zeta'| + \|\xi-\xi'\|^2) \end{aligned}$$

uniformly on $[0,\tau]\times\Theta$, *where* $\dot{s}_\zeta^{(0)}(\theta) = h_3(\zeta)\mathbb{E}_0\{e^{\alpha^T Z_1}(e^{\beta^T Z_2} - e^{\gamma^T Z_2}) \mid Z_3=\zeta\}$.

LEMMA 2. *Under conditions* C1–C3, $\sup_{\theta\in\Theta}|X_n - X|(\theta)$ *converges in probability to zero as* $n \to \infty$.

THEOREM 1. *Under conditions* C1–C3, *there exists a neighborhood* $\mathcal{B}_0$ *of* θ_0 *such that if* $\widehat{\theta}_n$ *lies in* $\mathcal{B}_0$, *then it converges weakly to* θ_0 *as* $n \to \infty$.

PROOF. For every $\theta = (\zeta, \xi^T)^T$ in Θ, the first derivatives of the function X with respect to α, β and γ are zero at θ_0 and the second derivative of the function $X(\theta)$ with respect to ξ, at fixed ζ, is the matrix $-I(\theta)$. The assumptions that λ_0 is bounded and $\operatorname{Var}\widetilde{Z}(t;\zeta)$ is positive definite imply that $I(\theta)$ is positive definite in a neighborhood of θ_0 [Pons and de Turckheim (1988), Lemma 2.2], and therefore the function $\xi \mapsto X(\zeta,\xi)$ is concave for every $(\zeta,\xi^T)^T$ in a neighborhood of θ_0.

Moreover, in a neighborhood of θ_0, X has partial derivatives with respect to ζ, at fixed ξ, $\dot{X}_\zeta^-(\zeta,\xi)$ for $\zeta < \zeta_0$ and $\dot{X}_\zeta^+(\zeta,\xi)$ for $\zeta > \zeta_0$. They are defined by

$$\dot{X}_\zeta^-(\theta) = \int_0^\tau \mathbb{E}_0\biggl[Y\biggl\{(\beta-\gamma)^T Z_2 e^{\alpha_0^T Z_1 + \beta_0^T Z_2} - e^{\alpha^T Z_1}\bigl(e^{\beta^T Z_2} - e^{\gamma^T Z_2}\bigr)\frac{s^{(0)}(\theta_0)}{s^{(0)}(\theta)}\biggr\} \mid Z_3 = \zeta\biggr] h_3(\zeta)\,d\Lambda_0,$$

$$\dot{X}_\zeta^+(\theta) = \int_0^\tau \mathbb{E}_0\biggl[Y\biggl\{(\beta-\gamma)^T Z_2 e^{\alpha_0^T Z_1 + \gamma_0^T Z_2} - e^{\alpha^T Z_1}\bigl(e^{\beta^T Z_2} - e^{\gamma^T Z_2}\bigr)\frac{s^{(0)}(\theta_0)}{s^{(0)}(\theta)}\biggr\} \mid Z_3 = \zeta^+\biggr] h_3(\zeta)\,d\Lambda_0.$$

If θ tends to θ_0 with $\zeta < \zeta_0$, the continuity of $s^{(0)}(t;\theta)$ with respect to θ (Lemma 1) implies that $\dot{X}_\zeta^-(\theta)$ tends to

$$\begin{aligned}\dot{X}_\zeta^-(\theta_0) &= \int_0^\tau \mathbb{E}_0\bigl[Y e^{\alpha_0^T Z_1}\{(\beta_0-\gamma_0)^T Z_2 e^{\beta_0^T Z_2} \\ &\qquad + e^{\gamma_0^T Z_2} - e^{\beta_0^T Z_2}\} \mid Z_3 = \zeta_0\bigr]\,d\Lambda_0 h_3(\zeta_0) \\ &= \tfrac{1}{2}\int_0^\tau \mathbb{E}_0\bigl[Y e^{\alpha_0^T Z_1}\{(\beta_0-\gamma_0)^T Z_2\}^2 e^{\beta_*^T Z_2} \mid Z_3 = \zeta_0\bigr]\,d\Lambda_0 h_3(\zeta_0),\end{aligned} \tag{3.1}$$

where β_* is between β_0 and γ_0. By condition (2.3), $\dot{X}_\zeta^-(\theta_0)$ is strictly positive and therefore $\dot{X}_\zeta^-(\theta)$ is strictly positive in a neighborhood of θ_0. Similarly, if θ tends to θ_0 with $\zeta < \zeta_0$, $\dot{X}_\zeta^+(\theta)$ tends to

$$\begin{aligned}\dot{X}_\zeta^+(\theta_0) &= \int_0^\tau \mathbb{E}_0\bigl[Y e^{\alpha_0^T Z_1}\{(\beta_0-\gamma_0)^T Z_2 e^{\gamma_0^T Z_2} \\ &\qquad + e^{\gamma_0^T Z_2} - e^{\beta_0^T Z_2}\} \mid Z_3 = \zeta_0^+\bigr]\,d\Lambda_0 h_3(\zeta_0)\end{aligned} \tag{3.2}$$

and it is strictly negative. This implies the existence of a neighborhood $\mathcal{B}_0$ of θ_0 where X attains a strict maximum at θ_0 and where X is concave. As X_n converges uniformly to X (Lemma 2), it follows that if $\widehat{\theta}_n$ lies in $\mathcal{B}_0$ then it converges weakly to θ_0 as $n \to \infty$. $\square$

To study the rates of convergence of $\widehat{\zeta}_n$ and $\widehat{\xi}_n$, let $\mathcal{U}_n = \{u = (u_1, u_2^T)^T : u_1 = n(\zeta - \zeta_0), u_2 = n^{1/2}(\xi - \xi_0)$ with $\zeta \in [\zeta_1, \zeta_2], \xi \in \Xi\}$. For $x = (x_1, x_2^T)^T$ with $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}^{p+2q}$, we denote $\rho(x) = (|x_1| + \|x_2\|^2)^{1/2}$ and $V_\varepsilon(\theta_0) = \{\theta \in \Theta : \rho(\theta - \theta_0) < \varepsilon\}$ an ε-neighborhood of θ_0 with respect to ρ, though it is not a norm. For $u = (u_1, u_2^T)^T \in \mathcal{U}_n$, let $\zeta_{n,u} = \zeta_0 + n^{-1}u_1$, $\xi_{n,u} = \xi_0 + n^{-1/2}u_2$ and $\theta_{n,u} = (\zeta_{n,u}, \xi_{n,u}^T)^T$ in Θ, and let $\mathcal{U}_{n,\varepsilon} = \{u \in \mathcal{U}_n : \rho(u) \le n^{1/2}\varepsilon\}$. Let W_n be the partial log-likelihood process defined by

$$W_n(\theta) = n^{1/2}(X_n - X)(\theta), \tag{3.3}$$

with X_n and X given by (2.1) and (2.2). The rates of convergence of $\widehat{\zeta}_n$ and $\widehat{\xi}_n$ will be deduced from the limiting behavior of the process W_n following classical arguments. It relies on the next lemmas, proved in Section 5.

LEMMA 3. *Under conditions* C1–C3, *for every $\varepsilon > 0$ there exists a constant $\kappa > 0$ such that $\mathbb{E}_0 \sup_{\theta \in V_\varepsilon(\theta_0)} |W_n(\theta)| \le \kappa\varepsilon$ as $n \to \infty$.*

LEMMA 4. *Under conditions* C1–C2, *for ε sufficiently small there exists a constant $\kappa_0 > 0$ such that for all θ in $V_\varepsilon(\theta_0)$, $X(\theta) \le -\kappa_0\{\rho(\theta - \theta_0)\}^2$.*

THEOREM 2. *Under conditions* C1–C3, *for $\varepsilon > 0$ sufficiently small, $P_0(\sup_{u \in \mathcal{U}_{n,\varepsilon}, \rho(u) > A} X_n(\theta_{n,u}) \ge 0)$ tends to zero as n and $A \to \infty$, and*

$$\limsup_{n\to\infty, A\to\infty} P_0(n|\widehat{\zeta}_n - \zeta_0| > A) = 0, \qquad \limsup_{n\to\infty, A\to\infty} P_0(n^{1/2}\|\widehat{\xi}_n - \xi_0\| > A) = 0.$$

PROOF. Let $\widehat{u}_n = (n(\widehat{\zeta}_n - \zeta_0), n^{1/2}(\widehat{\xi}_n - \xi_0))$, let $\eta > 0$ and let $\varepsilon > 0$ be sufficiently small to ensure that Lemma 4 holds on $V_\varepsilon(\theta_0)$. From Theorem 1, for all n larger than some integer n_0, $P_0\{\widehat{u}_n \in \mathcal{U}_{n,\varepsilon}\} = P_0\{\widehat{\theta}_n \in V_\varepsilon(\theta_0)\} > 1 - \eta$. Both probabilities $P_0(n|\widehat{\zeta}_n - \zeta_0| > A^2)$ and $P_0(n^{1/2}\|\widehat{\xi}_n - \xi_0\| > A)$ are bounded by $P_0(\rho(\widehat{u}_n) > A) \le P_0(\sup_{u \in \mathcal{U}_{n,\varepsilon}, \rho(u) > A} L_n(\theta_{n,u}) \ge L_n(\theta_0)) + \eta = P_0(\sup_{u \in \mathcal{U}_{n,\varepsilon}, \rho(u) > A} X_n(\theta_{n,u}) \ge 0) + \eta$. The latter probability is finally bounded following the arguments of Theorem 5.1 in Ibragimov and Has'minskii (1981), where $\mathcal{U}_{n,\varepsilon}$ is split into subsets $H_{n,j}$ defined by its intersection with the sets $\{g(j) < \rho(u) \le g(j+1)\}$, $j \in \mathbb{N}$, for a function g such that $\sum_{j; g(j) > A} g(j+1)/g^2(j)$ tends to zero as $A \to \infty$. Then from Lemma 3, $P_0(\sup_{u \in \mathcal{U}_{n,\varepsilon}, \rho(u) > A} X_n(\theta_{n,u}) \ge 0) \le \sum_{j; g(j) > A} P_0(\sup_{H_{n,j}} W_n(\theta_{n,u}) \ge n^{-1/2} g^2(j)\kappa_0)$ and it tends to zero by the Bienaymé–Chebyshev inequality. $\square$

4. Asymptotic distribution of the estimators. Let $A > 0$ and $\mathcal{U}_n^A = \{u \in \mathcal{U}_n; |u_1| + \|u_2\|^2 \le A\}$. The limiting distribution of $(n(\widehat{\zeta}_n - \zeta_0), n^{1/2}(\widehat{\xi}_n - \xi_0))$ will be deduced from Theorem 2 and from the behavior of the restriction of the process $u \mapsto l_n(\theta_{n,u}) - l_n(\theta_0)$ to the compact set $\mathcal{U}_n^A$, for A sufficiently large. We define a process Q_n on $\mathbb{R}$ and a variable $\widetilde{l}_n$ by

$$
\begin{aligned}
Q_n(u_1) = \sum_i \delta_i \bigg\{ &(\gamma_0 - \beta_0)^T Z_{2i}(T_i)\big(\mathbb{1}_{\{\zeta_{nu} < Z_{3i} \le \zeta_0\}} - \mathbb{1}_{\{\zeta_0 < Z_{3i} \le \zeta_{nu}\}}\big) \\
&- \frac{S_n^{(0)}(T_i; \zeta_{nu}, \xi_0) - S_n^{(0)}(T_i; \theta_0)}{S_n^{(0)}(T_i; \theta_0)} \bigg\},
\end{aligned} \tag{4.1}
$$

$$
\widetilde{l}_n = n^{-1/2} \sum_i \int_0^\tau \bigg\{ \widetilde{Z}_i(\zeta_0) - \frac{S_n^{(1)}(\theta_0)}{S_n^{(0)}(\theta_0)} \bigg\} \, dM_i,
$$

THEOREM 3. *Under conditions* C1–C3, *the following approximation holds uniformly on* $\mathcal{U}_n^A$, *for every* $A > 0$, *as* $n \to \infty$:

$$
l_n(\theta_{n,u}) - l_n(\theta_0) = Q_n(u_1) + u_2^T \widetilde{l}_n - \tfrac{1}{2} u_2^T I(\theta_0) u_2 + o_p(1).
$$

The proof of Theorem 3 is given in Section 5. We now study the weak convergence of Q_n as a random variable on the space D of right-continuous functions with left-hand limits on $\mathbb{R}$ endowed with the Skorohod topology, and on its restriction to the space D_A of right-continuous functions with left-hand limit functions on $[-A, A]$, for any $A > 0$. The process Q_n is written as the difference $Q_n = Q_n^+ - Q_n^-$, where Q_n^+ and Q_n^- are defined by $Q_n^+ = 0$ on $\mathbb{R}_-$, $Q_n^- = 0$ on $\mathbb{R}_+$,

$$
\begin{aligned}
Q_n^+(v) = \sum_i \delta_i \bigg\{ &(\beta_0 - \gamma_0)^T Z_{2i}(T_i) \mathbb{1}_{\{\zeta_0 < Z_{3i} \le \zeta_0 + n^{-1} v\}} \\
&- \frac{\sum_j Y_j(T_i) e^{\alpha_0^T Z_{1j}(T_i)} (e^{\beta_0^T Z_{2j}(T_i)} - e^{\gamma_0^T Z_{2j}(T_i)}) \mathbb{1}_{\{\zeta_0 < Z_{3j} \le \zeta_0 + n^{-1} v\}}}{S_n^{(0)}(\theta_0)} \bigg\} \\
&\qquad \text{if } v > 0,
\end{aligned}
$$

$$
\begin{aligned}
Q_n^-(v) = \sum_i \delta_i \bigg\{ &(\gamma_0 - \beta_0)^T Z_{2i}(T_i) \mathbb{1}_{\{\zeta_0 + n^{-1} v < Z_{3i} \le \zeta_0\}} \\
&- \frac{\sum_j Y_j(T_i) e^{\alpha_0^T Z_{1j}(T_i)} (e^{\beta_0^T Z_{2j}(T_i)} - e^{\gamma_0^T Z_{2j}(T_i)}) \mathbb{1}_{\{\zeta_0 + n^{-1} v < Z_{3j} \le \zeta_0\}}}{S_n^{(0)}(T_i; \theta_0)} \bigg\} \\
&\qquad \text{if } v < 0.
\end{aligned}
$$

In order to describe the asymptotic distribution of Q_n, let ν^+ and ν^- be the real jump processes such that $\nu^+ = 0$ on $\mathbb{R}_-$, $\nu^- = 0$ on $\mathbb{R}_+$, $\nu^+(s)$ is a Poisson variable with parameter $sh_3(\zeta_0)$ on $\mathbb{R}_+$ and $\nu^-(s)$ is a Poisson variable with parameter $-sh_3(\zeta_0)$ on $\mathbb{R}_-$. Let $(V_k^+)_{k\geq 1}$ and $(V_k^-)_{k\geq 1}$ be independent sequences of i.i.d. random variables with characteristic functions

$$\begin{aligned}\varphi^+(t) &= \mathbb{E}_0\big(e^{itV_k^+}\big)\\ &= \mathbb{E}_0\Big[e^{it\{\delta(\beta_0-\gamma_0)^T Z_2(T)-\int_0^\tau Ye^{\alpha_0^T Z_1}(e^{\beta_0^T Z_2}-e^{\gamma_0^T Z_2})\,d\Lambda_0\}}\mid Z_3=\zeta_0^+\Big],\\ \varphi^-(t) &= \mathbb{E}_0\big(e^{itV_k^-}\big)\\ &= \mathbb{E}_0\Big[e^{it\{\delta(\gamma_0-\beta_0)^T Z_2(T)-\int_0^\tau Ye^{\alpha_0^T Z_1}(e^{\beta_0^T Z_2}-e^{\gamma_0^T Z_2})\,d\Lambda_0\}}\mid Z_3=\zeta_0\Big]\end{aligned} \tag{4.2}$$

and let $V_0^+ = V_0^- = 0$; $(V_k^+)_{k\geq 1}$ and $(V_k^-)_{k\geq 1}$ are supposed to be independent of ν^+ and ν^-.

Let $Q = Q^+ - Q^-$ be the right-continuous jump process defined on $\mathbb{R}$ by

$$Q^+(s) = \sum_{0\leq j\leq \nu^+(s)} V_k^+,\quad Q^-(s) = \sum_{0\leq j\leq \nu^-(s)} V_k^-, \tag{4.3}$$

and let $\widehat{v}_Q = \inf\{v; Q(v) = \arg\max Q\}$ be the maximum value of Q.

LEMMA 5. *The process Q has independent increments, $Q^+ = 0$ on $\mathbb{R}_-$, $Q^- = 0$ on $\mathbb{R}_+$ and the variables $Q^+(s)$ and $Q^-(s)$ have the characteristic functions $\phi_s^+(t) = \exp[sh_3(\zeta_0)\{\varphi^+(t) - 1\}]$ for s in $\mathbb{R}_+$ and $\phi_s^-(t) = \exp[-sh_3(\zeta_0)\{\varphi^-(t) - 1\}]$ for s in $\mathbb{R}_-$. Moreover, $\widehat{v}_Q$ is a.s. a finite random time.*

THEOREM 4. *Under conditions* C1–C3, *the variable $\widetilde{l}_n$ converges weakly to a Gaussian variable $\mathcal{N}(0, I(\theta_0))$, the process Q_n converges weakly to Q in D_A, for every $A > 0$, and they are asymptotically independent.*

PROOF. As in Theorem 4.1 of Andersen and Gill (1982), the variable $\widetilde{l}_n$ in (4.2) converges weakly to a Gaussian variable $\mathcal{N}(0, I(\theta_0))$. For the convergence of Q_n, we may restrict our attention to Q_n^+ and the proof extends to (Q_n^+, Q_n^-) since the processes Q_n^+ and Q_n^- are independent and similarly defined. To prove the weak convergence of the finite dimensional distributions of Q_n^+, we shall prove their tightness and the convergence of their characteristic functions. Let $J \in \mathbb{N}$, let $0 = v_0 < v_1 < \cdots < v_J \leq A$ be an increasing sequence and $I_{nj} =]\zeta_0 + n^{-1}v_{j-1}, \zeta_0 + n^{-1}v_j]$ and let $q_1, \dots, q_J$ be constants. The variable $\Sigma_n = \sum_{j\leq J} q_j\{Q_n^+(v_j) - Q_n^+(v_{j-1})\}$ is the sum of the n variables $\eta_{n,i} = \sum_{j\leq J} q_j \eta_{nj,i}$, where $\eta_{nj,i} = \eta_{nj,i}^{(1)} + \eta_{nj,i}^{(2)}$,

$$\eta_{nj,i}^{(1)} = \mathbb{1}_{I_{nj}}(Z_{3i})\bigg\{\delta_i(\beta_0-\gamma_0)^T Z_{2i}(T_i) - \int_0^\tau \phi_i\, d\Lambda_0\bigg\},$$

$$\eta_{nj,i}^{(2)} = \mathbb{1}_{I_{nj}}(Z_{3i})\int_0^\tau \phi_i\,\{S_n^{(0)-1}(\theta_0)\, d\bar{N}_n - d\Lambda_0\}$$

with $\phi_i = Y_i e^{\alpha_0^T Z_{1i}}(e^{\beta_0^T Z_{2i}} - e^{\gamma_0^T Z_{2i}})$. Since the intervals I_{nj} and I_{nl} are disjoint, $\eta_{nj,i}^{(\ell)}\eta_{nl,i}^{(\ell)} = 0$ if $j \neq l$, for $\ell = 1, 2$. Let $\Sigma_n^{(1)} = \sum_{i\le n}\sum_{j\le J} q_j \eta_{nj,i}^{(1)}$ and $\Sigma_n^{(2)} = \sum_{i\le n}\sum_{j\le J} q_j \eta_{nj,i}^{(2)}$. By the martingale property $\mathbb{E}_0\Sigma_n^{(2)} = 0$, and by C3 and (2.4),

$$\begin{aligned}
\mathbb{E}_0\{\eta_{nj,i}^{(2)}\}^2 &= \mathbb{E}_0 \mathbb{1}_{I_{nj}}(Z_{3i}) \int_0^\tau \phi_i^2 S_n^{(0)-1}(\theta_0)\, d\Lambda_0 \\
&= n^{-2}(v_j - v_{j-1}) h_3(\zeta_0) \\
&\quad \times \int_0^\tau \mathbb{E}_0(\phi_i^2 \mid Z_{3i} = \zeta_0^+) s^{(0)-1}(\theta_0)\, d\Lambda_0 + o(n^{-2}), \\
\mathbb{E}_0\{\eta_{nj,i}^{(2)}\eta_{nk,l}^{(2)}\} &= n^{-3}(v_j - v_{j-1})(v_k - v_{k-1}) h_3^2(\zeta_0) \\
&\quad \times \int_0^\tau \mathbb{E}_0^2(\phi_i \mid Z_{3i} = \zeta_0^+) s^{(0)-1}(\theta_0)\, d\Lambda_0 + o(n^{-3}).
\end{aligned}$$

Therefore $\mathbb{E}_0\{\Sigma_n^{(2)}\}^2 = O(n^{-1})$ and $\Sigma_n^{(2)}$ converges to zero in probability. The variable $\Sigma_n^{(1)}$ is the sum of the n i.i.d. variables $\eta_{n,i}^{(1)} = \sum_{j\le J} q_j \eta_{nj,i}^{(1)}$. Its mean and its variance are $m_n = \sum_j q_j m_{nj}$ and $\sigma_n^2 = \sum_j q_j^2 \mathbb{E}_0\{\eta_{nj,i}^{(1)}\}^2 - m_n^2$, with

$$\begin{aligned}
m_{nj} &= \int_0^\tau \mathbb{E}_0\big[Y \mathbb{1}_{I_{nj}}(Z_3)\{(\beta_0 - \gamma_0)^T Z_2 e^{r_{\theta_0}(Z)} - \phi\}\big]\, d\Lambda_0 \\
&= n^{-1}(v_j - v_{j-1}) m_1 + o(n^{-1}), \\
\mathbb{E}_0\{\eta_{nj,i}^{(1)}\}^2 &= \mathbb{E}_0\bigg[\mathbb{1}_{I_{nj}}(Z_{3i})\bigg\{\delta_i(\beta_0 - \gamma_0)^T Z_{2i}(T_i) - \int_0^\tau \phi_i\, d\Lambda_0\bigg\}^2\bigg] \\
&\le 2n^{-1}(v_j - v_{j-1}) m_2 + o(n^{-1})
\end{aligned}$$

for constants m_1 and m_2 depending only on the distributions under P_0. Then the sequence of the distributions of Σ_n, $n \ge 1$, is tight since, for all $K > 0$,

$$\begin{aligned}
P(|\Sigma_n| > K) &\le 2K^{-2}\big[\mathbb{E}_0\{\Sigma_n^{(1)}\}^2 + \mathbb{E}_0\{\Sigma_n^{(2)}\}^2\big] \\
&\le 2K^{-2}\big(n\sigma_n^2 + n^2 m_n^2 + o(1)\big) = O(K^{-2}).
\end{aligned}$$

As $\Sigma_n^{(2)}$ converges to zero in probability, Σ_n and $\Sigma_n^{(1)}$ have the same limiting distribution if they converge. The characteristic function of $\Sigma_n^{(1)}$ is $\varphi_n(s) = (\mathbb{E}_0 \exp\sum_j isq_j\eta_{nj,k}^{(1)})^n$. Since the intervals I_{nj} do not overlap, for each k there is at most one index j such that $\eta_{nj,k} \neq 0$. Then using the equality $e^{\sum_j a_j} - 1 = \sum_j (e^{a_j} - 1)$ for a sum where only one term a_j is different from zero,

$\mathbb{E}_0 \exp\{\sum_{j\le J} isq_j\eta^{(1)}_{nj,k}\} = 1 + \sum_{j\le J} n^{-1}\{(v_j - v_{j-1})\varphi(s,q_j) + o(1)\}$ with

$$\varphi(s,q_j) = h_3(\zeta_0)$$
$$\times \left\{\mathbb{E}_0\left(\exp\left[isq_j\left\{\delta(\beta_0-\gamma_0)^T Z_2(T) - \int_0^\tau \phi\, d\Lambda_0\right\}\right] \,\middle|\, Z_3 = \zeta_0^+\right) - 1\right\},$$

and $\varphi_n(s)$ converges to $\varphi(s) = \exp\{\sum_{j\le J}(v_j - v_{j-1})\varphi(s,q_j)\}$. It follows that the finite-dimensional distributions of Q_n^+ converge weakly to those of Q^+ defined by (4.3).

To prove the weak convergence of the process Q_n^+ in the Skorohod topology on D_A, it remains to prove its tightness. Let $0 \le v_1 \le v \le v_2 \le A$. Since the intervals $I'_{n1} =]\zeta_0 + n^{-1}v_1, \zeta_0 + n^{-1}v]$ and $I'_{n2} =]\zeta_0 + n^{-1}v, \zeta_0 + n^{-1}v_2]$ are disjoint,

$$\mathbb{E}_0|Q_n^+(v) - Q_n^+(v_1)|\,|Q_n^+(v_2) - Q_n^+(v)|$$
$$\le \sum_{i\ne j} \mathbb{E}_0 \mathbb{1}_{I'_{n1}}(Z_{3i})\mathbb{1}_{I'_{n2}}(Z_{3j})$$
$$\times \left\{\delta_i|(\beta_0-\gamma_0)^T Z_{2i}(T_i)| + \left|\sum_k \int_0^\tau \delta_k \phi_i(T_k) S_n^{(0)-1}(T_k;\theta_0)\right|\right\}$$

and it is bounded by $(v_2 - v_1)^2$ times a constant for every n by similar arguments as above. Hence the process Q_n^+ satisfies the D-tightness criterion (15.21) of Billingsley (1968), and then the processes Q_n converge weakly to Q. Finally, Q_n and $\widetilde{l}_n$ are asymptotically independent because any linear combination $a\Sigma_n^{(1)} + b^T\widetilde{l}_n$ converges weakly to $a\sum_{j\le J} q_j\{Q^+(v_j) - Q^+(v_{j-1})\} + b^T\mathcal{N}(0, I(\theta_0))$, since the variable $n^{-1/2}\sum_{i\le n}\sum_{j\le J} q_j \mathbb{1}_{I_{nj}}(Z_{3i}) \int_0^\tau\{\widetilde{Z}_i(\zeta_0) - S_n^{(1)}(\theta_0)S_n^{(0)-1}(\theta_0)\}\, dM_i$ tends to zero in probability. $\square$

THEOREM 5. *Under conditions* C1–C3, $n(\widehat{\zeta}_n - \zeta_0)$ *and* $n^{1/2}(\widehat{\xi}_n - \xi_0)$ *are asymptotically independent*, $n(\widehat{\zeta}_n - \zeta_0) = \arg\max_{u_1} Q_n(u_1) + o_p(1)$ *and it converges weakly to* $\widehat{v}_Q$, *and* $n^{1/2}(\widehat{\xi}_n - \xi_0) = I(\theta_0)^{-1}\widetilde{l}_n + o_p(1)$ *and converges weakly to a Gaussian variable* $\mathcal{N}(0, I(\theta_0)^{-1})$.

PROOF. Let $\widehat{u}_n = (n(\widehat{\zeta}_n - \zeta_0), n^{1/2}(\widehat{\xi}_n - \xi_0)^T)^T$. For every $x \in \mathbb{R}$ and $y \in \mathbb{R}^{p+2q}$,

$$P_0(\widehat{u}_n < (x, y^T)^T)$$
$$= P_0\left\{\left(\arg\max_{u_1} Q_n(u_1^-) \vee Q_n(u_1), I(\theta_0)^{-1}\widetilde{l}_n^T\right)^T + o_p(1) < (x, y^T)^T\right\}$$

with a uniform o_p on $\mathcal{U}_n^A$ for every $A \ge (|x| + \|y\|^2)^{1/2}$, by Theorem 3. The asymptotic independence of Q_n and $\widetilde{l}_n$ and their weak convergence (Theorem 4) entail that $P_0(\widehat{u}_n < (x, y^T)^T)$ tends to $P_0(\widehat{v}_Q < x)\, P_0(G_0 < y)$, where G_0 is a Gaussian variable $\mathcal{N}(0, I(\theta_0)^{-1})$. Using this convergence and Theorem 2, for every $\varepsilon > 0$, there exist n_0 and A_0 such that for all $n \geq n_0$, $P_0(\|\widehat{u}_{2n}\| \geq A_0) \leq \varepsilon/3$, $|P_0(\widehat{u}_{1n} < x, \|\widehat{u}_{2n}\| < A_0) - P_0(\widehat{v}_Q < x)\, P_0(\|G_0\| < A_0)| \leq \varepsilon/3$, $P_0(\|G_0\| \geq A_0) \leq \varepsilon/3$, and hence

$$\begin{aligned} &|P_0(\widehat{u}_{1n} < x) - P_0(\widehat{v}_Q < x)| \\ &\quad \leq |P_0(\widehat{u}_{1n} < x, \|\widehat{u}_{2n}\| < A_0) - P_0(\widehat{v}_Q < x)\, P_0(\|G_0\| < A_0)| \\ &\qquad + P_0(\widehat{v}_Q < x)\, P_0(\|G_0\| \geq A_0) + P_0(\widehat{u}_{1n} < x, \|\widehat{u}_{2n}\| \geq A_0) \leq \varepsilon, \end{aligned}$$

so $\widehat{u}_{1n}$ converges weakly to $\widehat{v}_Q$. By the same arguments, $\widehat{u}_{2n}$ converges weakly to G_0. Moreover, they are asymptotically independent, and on the set $\{\rho(\widehat{u}_n) < A_0\}$ with probability larger than $1 - \varepsilon$, we have $\widehat{u}_{1n} = \arg\max_{u_1} Q_n(u_1^-) \vee Q_n(u_1) + o_p(1)$ and $\widehat{u}_{2n} = I(\theta_0)^{-1}\widetilde{l}_n + o_p(1)$. $\square$

REMARK 1. As proved in Lemma 5, $\widehat{v}_Q$ is a.s. finite and by Theorem 2, it is sufficient to consider the distribution of Q on compacts to build asymptotic confidence intervals for ζ_0. However, the distribution of Q depends on the unknown parameter θ_0 and it seems difficult to use the conditional characteristic functions (4.2). Bootstrap confidence intervals with a resampling of the individuals could be considered but their asymptotic behavior will not be studied here.

REMARK 2. If ζ_0 were known, the maximum partial likelihood estimator of ξ_0 would have the same asymptotic distribution as $n^{1/2}(\widehat{\xi}_n - \xi_0)$ in Theorem 5 and it would be an efficient estimator of ξ_0. With ζ_0 unknown, $\widehat{\xi}_n$ is thus an adaptive estimator of ξ_0.

The weak convergence of $n^{1/2}(\widehat{\Lambda}_n - \Lambda_0)$ may be established using the approach of Andersen and Gill (1982). Its asymptotic behavior follows from Theorem 5 and from the next result, which is the same as if ζ_0 were known. From Theorem 6 and Remark 2, the limit distribution of $n^{1/2}(\widehat{\Lambda}_n - \Lambda_0)$ does not depend on knowledge of ξ_0.

THEOREM 6. *Under conditions* C1–C3, *the process defined for $t \in [0, \tau]$ by*

$$n^{1/2}(\widehat{\Lambda}_n - \Lambda_0)(t) + n^{1/2}(\widehat{\xi}_n - \xi_0)^T \int_0^t \frac{s^{(1)}}{s^{(0)}}(\theta_0)\, d\Lambda_0 \tag{4.4}$$

converges weakly to a centered Gaussian process with covariance $\int_0^{s\wedge t} s^{(0)-1}(\theta_0)\, d\Lambda_0$ *at s and t in* $[0, \tau]$, *and it is asymptotically independent of* $n^{1/2}(\widehat{\xi}_n - \xi_0)$.

PROOF. By definition of the predictable compensator of $\bar{N}_n$,

$$n^{1/2}(\widehat{\Lambda}_n - \Lambda_0)(t) = \int_0^t \frac{d\mathbb{M}_n^{(0)}}{n^{-1}S_n^{(0)}(\widehat{\theta}_n)} - \int_0^t \frac{n^{-1/2}\{S_n^{(0)}(\widehat{\theta}_n) - S_n^{(0)}(\theta_0)\}}{n^{-1}S_n^{(0)}(\widehat{\theta}_n)}\, d\Lambda_0.$$

The first term in the right-hand side is the integral of the left-continuous process $nS_n^{(0)-1}(\widehat{\theta}_n)$ with respect to the martingale $\mathbb{M}_n^{(0)}$ and it converges weakly to a centered Gaussian process with covariance $\int_0^{s\wedge t} s^{(0)-1}(\theta_0)\, d\Lambda_0$ by Rebolledo's (1980) convergence theorem. The asymptotic equivalence of the second term and $n^{1/2}(\widehat{\xi}_n - \xi_0)^T \int_0^t s^{(1)}(\theta_0)s^{(0)-1}(\theta_0)\, d\Lambda_0$ is obtained from the expansion

$$\begin{aligned} &n^{-1/2}\{S_n^{(0)}(\widehat{\theta}_n) - S_n^{(0)}(\theta_0)\} \\ &\qquad = n^{1/2}(\widehat{\xi}_n - \xi_0)^T n^{-1} S_n^{(1)}(\widehat{\zeta}_n, \xi_n^*) + n^{-1/2}\{S_n^{(0)}(\widehat{\zeta}_n, \xi_0) - S_n^{(0)}(\theta_0)\}, \end{aligned}$$

with ξ_n^* between $\widehat{\xi}_n$ and ξ_0. From Condition C3, Lemma 1 and Theorem 1, $\sup_{t\in[0,\tau]} \|n^{-1}S_n^{(0)}(t;\widehat{\theta}_n) - s^{(0)}(t;\theta_0)\|$ and $\sup_{t\in[0,\tau]} \sup_{\xi\in]\widehat{\xi}_n,\xi_0[\cup]\xi_0,\widehat{\xi}_n[} \|n^{-1} \times S_n^{(1)}(t;\widehat{\zeta}_n,\xi) - s^{(1)}(t;\theta_0)\|$ tend to zero in probability. Moreover,

$$\begin{aligned} &n^{-1/2}\{S_n^{(0)}(\widehat{\zeta}_n, \xi_0) - S_n^{(0)}(\theta_0)\} \\ &\qquad = n^{-1/2}\sum_i Y_i e^{\alpha_0^T Z_{1i}}\big(e^{\beta_0^T Z_{2i}} - e^{\gamma_0^T Z_{2i}}\big)\big(\mathbb{1}_{\{\zeta_0 < Z_{3i} \le \widehat{\zeta}_n\}} - \mathbb{1}_{\{\widehat{\zeta}_n < Z_{3i} \le \zeta_0\}}\big), \end{aligned}$$

denoted $n^{-1/2}\sum_i Y_i\phi_i(\mathbb{1}_{\{\zeta_0<Z_{3i}\le\widehat{\zeta}_n\}} - \mathbb{1}_{\{\widehat{\zeta}_n<Z_{3i}\le\zeta_0\}})$. From Theorem 5, for every $\varepsilon > 0$, there exist A and n_0 such that for $n \ge n_0$, $P_0(n|\widehat{\zeta}_n - \zeta_0| > A) \le \varepsilon/2$. Let $\Omega_{nA} = \{n|\widehat{\zeta}_n - \zeta_0| \le A\}$. For every $\eta > 0$, $P_0(\sup_t n^{-1/2}|\sum_i Y_i(t)\phi_i(t) \times \mathbb{1}_{\{\zeta_0<Z_{3i}\le\widehat{\zeta}_n\}}| > \eta)$ is smaller than

$$\begin{aligned} &P_0\left(\sup_t n^{-1/2}\left|\sum_i Y_i(t)\phi_i(t)\mathbb{1}_{\{\zeta_0<Z_{3i}\le\widehat{\zeta}_n\}}\mathbb{1}_{\Omega_{nA}}\right| > \eta\right) + \frac{\varepsilon}{2} \\ &\qquad \le \frac{1}{\eta^2}\mathbb{E}_0\left\{\sup_t Y(t)\phi^2(t)\mathbb{1}_{\{\zeta_0<Z_3\le\zeta_0+n^{-1}A\}}\right\} \\ &\qquad\quad + \frac{n-1}{\eta^2}\left[\mathbb{E}_0\left\{\sup_t Y(t)|\phi|(t)\mathbb{1}_{\{\zeta_0<Z_3\le\zeta_0+n^{-1}A\}}\right\}\right]^2 + \frac{\varepsilon}{2} \\ &\qquad \le \frac{Ah_3(\zeta_0)}{n\eta^2}\mathbb{E}_0\left\{\sup_t |\phi|^2(t) \mid Z_3 = \zeta_0^+\right\} \\ &\qquad\quad + \frac{(n-1)(Ah_3(\zeta_0))^2}{n^2\eta^2}\left[\mathbb{E}_0\left\{\sup_t |\phi|(t) \mid Z_3 = \zeta_0^+\right\}\right]^2 + \frac{\varepsilon}{2}, \end{aligned}$$

which is smaller than ε for n large enough and the same result holds for $n^{-1/2}\sum_i Y_i\phi_i \mathbb{1}_{\{\widehat{\zeta}_n<Z_{3i}\le\zeta_0\}}$. Therefore the process $n^{-1/2}\{S_n^{(0)}(\widehat{\zeta}_n,\xi_0) - S_n^{(0)}(\theta_0)\}$ tends to zero in probability uniformly on $[0,\tau]$, and (4.4) is uniformly approximated by $\int_0^t nS_n^{(0)-1}(\widehat{\theta}_n)\,d\mathbb{M}_n^{(0)}$.

The asymptotic independence of (4.4) and $n^{1/2}(\widehat{\xi}_n-\xi_0)$ is a consequence of the approximation $n^{1/2}(\widehat{\xi}_n-\xi_0)=I(\theta_0)^{-1}\widetilde{l}_n+o_p(1)$ since $\widetilde{l}_n$ and the local martingale $\int_0^{\cdot} nS_n^{(0)-1}(\widehat{\theta}_n)\,d\mathbb{M}_n^{(0)}$ are asymptotically Gaussian with mean zero and they satisfy $\mathbb{E}_0\widetilde{l}_n\int_0^t S_n^{(0)-1}(\widehat{\theta}_n)\,d\mathbb{M}_n^{(0)}=0$ for all t in $[0,\tau]$. $\square$

REMARK 3. As $\widehat{\xi}_n$ and $\widehat{\Lambda}_n$ are adaptive with respect to ζ_0, asymptotic confidence intervals for the components of ξ_0 and for Λ_0 are the same as in the regular Cox model with a change-point at a known time ζ_0. This enables one to use the standard software for survival data analysis by a maximization of the partial likelihood $L_n(a_k,\xi)$ with respect to the parameter ξ for successive values a_k on a grid in $[\zeta_1,\zeta_2]$, with a path of order $o(n^{-1})$. The maximization of $L_n(a_k,\cdot)$ provides an estimator $\widehat{\xi}_{k,n}$ for ξ_0 and $\widehat{\zeta}_n$ can be approximated by the value $\widetilde{\zeta}_n$ that maximizes the sequence $(\widehat{L}_n(a_k))_k=(L_n(a_k,\widehat{\xi}_{k,n}))_k$. Then $\widehat{\xi}_n$ is approximated by the value of $\widetilde{\xi}_n$ associated with $\widetilde{\zeta}_n$ and $\widehat{\Lambda}_n$ is approximated by the Breslow estimator $\widetilde{\Lambda}_n$ calculated with $S_n^{(0)}(\widetilde{\zeta}_n,\widetilde{\xi}_n)$. Under the above conditions, they have the same asymptotic behavior as $\widehat{\zeta}_n$, $\widehat{\xi}_n$ and $\widehat{\Lambda}_n$, described in Theorems 5 and 6.

5. Proofs of results. The proofs are based on functional convergences of empirical processes which are established in a preliminary lemma. We denote $U_i=\delta_i Z_{2i}(T_i)$, $P_n^{\delta,T}$ and P_n^{U,Z_3} the empirical distributions of the variables $(\delta_i,T_i)_{i\le n}$ and $(U_i,Z_{3i})_{i\le n}$, respectively, and $P_0^{\delta,T}$ and P_0^{U,Z_3} their distributions under P_0. Let also $\nu_n^{\delta,T}$ and ν_n^{U,Z_3} be the related empirical processes and ν_n^t be the empirical process associated with the variables $(Y_i(t),Z_i(t))$, $1\le i\le n$. We consider functional families defined by

$$\varphi_\theta(d,t)=d\,\log\{s^{(0)}(t;\theta)s^{(0)-1}(t;\theta_0)\},\qquad d\in\{0,1\},\ t\in[0,\tau],$$

$$f^+_{\zeta,j}(u,z)=u_j\mathbb{1}_{\{\zeta<z\le\zeta_0\}},\qquad f^-_{\zeta,j}(u,z)=u_j\mathbb{1}_{\{\zeta<z\le\zeta_0\}},$$

$$z,u=(u_j)_{1\le j\le q}\in\mathbb{R}^q,$$

$$\psi_{1,t,\theta}(y,z)=y\left\{\frac{e^{\alpha^Tz_1+\beta^Tz_2}}{s^{(0)}(t;\theta)}-\frac{e^{\alpha_0^Tz_1+\beta_0^Tz_2}}{s^{(0)}(t;\theta_0)}\right\}\mathbb{1}_{\{z_3\le\zeta_0\}},$$

$$\psi_{2,t,\theta}(y,z)=y\left\{\frac{e^{\alpha^Tz_1+\gamma^Tz_2}}{s^{(0)}(t;\theta)}-\frac{e^{\alpha_0^Tz_1+\gamma_0^Tz_2}}{s^{(0)}(t;\theta_0)}\right\}\mathbb{1}_{\{z_3>\zeta_0\}},$$

$$\psi_{3,t,\theta}(y,z) = y\left\{\frac{e^{\alpha^T z_1+\gamma^T z_2}}{s^{(0)}(t;\theta)} - \frac{e^{\alpha_0^T z_1+\beta_0^T z_2}}{s^{(0)}(t;\theta_0)}\right\}\mathbb{1}_{\{\zeta<z_3\le\zeta_0\}},$$

$$\psi_{4,t,\theta}(y,z) = y\left\{\frac{e^{\alpha^T z_1+\beta^T z_2}}{s^{(0)}(t;\theta)} - \frac{e^{\alpha_0^T z_1+\gamma_0^T z_2}}{s^{(0)}(t;\theta_0)}\right\}\mathbb{1}_{\{\zeta_0<z_3\le\zeta\}}$$

for $y \in \{0,1\}$ and $z = (z_1, z_2, z_3)$ with $z_j \in \mathcal{Z}_j$, $\mathcal{F}_\varepsilon^+ = \{f_{\zeta,j}; \zeta_0 < \zeta \le \zeta_0 + \varepsilon^2, 1 \le j \le q\}$ and $\mathcal{F}_\varepsilon^- = \{f_{\zeta,j} : \zeta_0 - \varepsilon^2 \le \zeta < \zeta_0, 1 \le j \le q\}$.

LEMMA 6. *Under conditions* C1–C3, $\sup_\theta |(P_n^{\delta,T} - P_0^{\delta,T})(\varphi_\theta)|$, $\sup_{\zeta,j} |(P_n^{U,Z_3} - P_0^{U,Z_3})(f_{\zeta,j}^+)|$ *and* $\sup_{\zeta,j} |(P_n^{U,Z_3} - P_0^{U,Z_3})(f_{\zeta,j}^-)|$ *converge in probability to zero. For every* n, $\mathbb{E}_0 \sup_{\mathcal{F}_\varepsilon^+} |\nu_n(f^+)|$, $\mathbb{E}_0 \sup_{\mathcal{F}_\varepsilon^-} |\nu_n(f^-)|$, $\mathbb{E}_0 \sup_{\theta\in V_\varepsilon(\theta_0)} |\nu_n^{\delta,T}(\varphi_\theta)|$ *and* $\sup_{t\in[0,\tau]} \mathbb{E}_0 \sup_{\theta\in V_\varepsilon(\theta_0)} |\nu_n(k,t;\theta)|$, $k = 1,\dots,4$, *are bounded by* ε *times a constant.*

PROOF. The first two convergences are consequences of the Glivenko–Cantelli theorem for uniformly continuous and integrable functions $(\varphi_\theta)_{\theta\in\Theta}$ and for the Vapnik–Cervonenkis class $(]\zeta,\zeta_0])_{\zeta\in[\zeta_1,\zeta_2]}$ and $(]\zeta_0,\zeta])_{\zeta\in[\zeta_1,\zeta_2]}$. The $L_2(P_0)$ norm of the envelope function of $\mathcal{F}_\varepsilon^+$ is less than

$$\mathbb{E}_0 \sup_{\zeta\in V_{\varepsilon^2}(\zeta_0)} \|U_i\| \mathbb{1}_{\{\zeta_0<Z_{3i}\le\zeta\}} \le \left\{\mathbb{E}_0 \mathbb{1}_{\{\zeta_0<Z_{3i}\le\zeta_0+\varepsilon^2\}} \int_0^\tau \|Z_{2i}\|^2\, dN_i\right\}^{1/2} = O(\varepsilon).$$

For $\mathcal{F}_\varepsilon = \mathcal{F}_\varepsilon^+$ or $\mathcal{F}_\varepsilon^-$, the bound of $\mathbb{E}_0 \sup_{\mathcal{F}_\varepsilon} |\nu_n(f)|$ is a consequence of Theorem 2.14.1 in van der Vaart and Wellner (1996). For the functions φ_θ and for every $t \in [0,\tau]$, θ and θ' in $V_\varepsilon(\theta_0)$, $\varphi_{\theta'}(1,t) - \varphi_\theta(1,t) = \{(\xi' - \xi)^T s^{(1)}(t;\theta) + (\zeta' - \zeta)\dot{s}_\zeta^{(0)}(t;\theta)\} s^{(0)-1}(t;\theta) + \frac{1}{2}(\xi' - \xi)^T v(t;\theta)(\xi' - \xi) + o(\varepsilon^2)$ by Lemma 1, where $s^{(1)}$, $\dot{s}_\zeta^{(0)}$, $s^{(0)-1}$ and v are uniformly bounded. The family $\{\varphi_\theta : \theta \in V_\varepsilon(\theta_0)\}$ has therefore an envelope function with an $L_2(P_0^{\delta,T})$-norm of order ε and its $L_2(P_0^{\delta,T})$-bracketing integral $J_{[]}(1, L_2(P_0^{\delta,T}))$ is finite by Theorem 2.7.11 in van der Vaart and Wellner (1996). The bound of $\mathbb{E}_0 \sup_{\theta\in V_\varepsilon(\theta_0)} |\nu_n^{\delta,T}(\varphi_\theta)|$ is a consequence of their Theorem 2.14.2. Similar arguments hold for the classes of functions $\Psi_{k,t} = \{\psi_{k,t,\theta} : \theta \in V_\varepsilon(\theta_0)\}$: For $k = 1,2$ and for every $t \in [0,\tau]$, the functions $\psi_{k,t,\theta}$ are continuously differentiable with respect to θ and their derivatives are uniformly square integrable on $[0,\tau] \times V_\varepsilon(\theta_0)$, and for every $t \in [0,\tau]$, the functions $\psi_{k,t,\theta}$ are continuously differentiable with respect to θ and their derivatives are uniformly square integrable on $[0,\tau] \times V_\varepsilon(\theta_0)$, by Lemma 1. For $k = 3,4$, the functions $\psi_{k,t,\theta}$ are the product of the indicator function $\mathbb{1}_{]\zeta,\zeta_0]}$, with $\zeta \in [\zeta_0 - \varepsilon^2, \zeta_0[$, and of a continuously differentiable function with respect to θ having uniformly square integrable derivatives on $[0,\tau] \times V_\varepsilon(\theta_0)$. Moreover, $\Psi_{k,t}$ has a finite L_2-bracketing integral which does not depend on t. $\square$

PROOF OF LEMMA 2. The process X_n is written

$$
\begin{aligned}
X_n(\theta) = (\xi - \xi_0)^T &\left\{ n^{-1/2}\mathbb{M}_n^{(1)}(\tau) + \int_0^\tau n^{-1} S_n^{(1)}(\theta_0)\, d\Lambda_0 \right\} \\
&- \int_0^\tau \log \frac{S_n^{(0)}(\theta)}{S_n^{(0)}(\theta_0)} n^{-1}\, d\bar{N}_n \\
&+ (\beta - \gamma)^T \left\{ n^{-1} \sum_{i \le n} \delta_i Z_{2i}(T_i) \mathbb{1}_{\{\zeta_0 < Z_{3i} \le \zeta\}} \right. \\
&\qquad \left. - n^{-1} \sum_{i \le n} \delta_i Z_{2i}(T_i) \mathbb{1}_{\{\zeta < Z_{3i} \le \zeta_0\}} \right\}.
\end{aligned} \tag{5.1}
$$

Under C3, $\int_0^\tau n^{-1} S_n^{(1)}(\theta_0)\, d\Lambda_0$ converges to $\int_0^\tau s^{(1)}(\theta_0)\, d\Lambda_0$ and the variable $n^{-1/2}\mathbb{M}_n^{(1)}(\tau)$ converges in probability to zero since $\mathbb{E}_0\{\mathbb{M}_n^{(1)}(\tau)\}^2 = \mathbb{E}_0 \int_0^\tau n^{-1} \times S_n^{(2)}(\theta_0)\, d\Lambda_0$ is bounded. The convergence of the last three terms in (5.1) is a consequence of Lemma 6 and C3. $\square$

PROOF OF LEMMA 3. The process W_n is written $W_{1n} - W_{2n}$ where

$$
\begin{aligned}
W_{1n}(\theta) = n^{-1/2} \sum_i &\big[r_\theta(Z_i(T_i)) - r_{\theta_0}(Z_i(T_i)) \\
&\quad - \mathbb{E}_0\{ r_\theta(Z_i(T_i)) - r_{\theta_0}(Z_i(T_i)) \} \big],
\end{aligned}
$$

$$
W_{2n}(\theta) = n^{-1/2} \sum_i \left[\log \frac{S_n^{(0)}(T_i;\theta)}{S_n^{(0)}(T_i;\theta_0)} - \int_0^\tau \log \left\{ \frac{s^{(0)}(\theta)}{s^{(0)}(\theta_0)} \right\} s^{(0)}(\theta_0)\, d\Lambda_0 \right].
$$

Let $\mathbb{G}_n^{(1)} = n^{1/2}(n^{-1} S_n^{(1)} - s^{(1)})$ be the empirical processes associated with $S_n^{(1)}$. We shall bound successively the supremum of each term in a neighborhood of θ_0. First, $W_{1n}(\theta) = n^{-1/2} \sum_{i \le n} [(\alpha - \alpha_0)^T \int_0^\tau \{Z_{1i}\, dN_i - s_1^{(1)}(\theta_0)\, d\Lambda_0\} + (\beta - \beta_0)^T \int_0^\tau \{Z_{2i} \mathbb{1}_{\{Z_{3i} \le \zeta_0\}}\, dN_i - s_2^{(1)-}(\zeta_0, \alpha_0, \beta_0)\, d\Lambda_0\} + (\gamma - \gamma_0)^T \int_0^\tau \{Z_{2i} \times \mathbb{1}_{\{Z_{3i} > \zeta_0\}}\, dN_i - s_2^{(1)+}(\zeta_0, \alpha_0, \gamma_0)\, d\Lambda_0\} + (\beta - \gamma)^T \int_0^\tau \{Z_{2i} \mathbb{1}_{\{\zeta_0 < Z_{3i} \le \zeta\}}\, dN_i - s_2^{(1)}(]\zeta_0, \zeta], \alpha_0, \gamma_0)\, d\Lambda_0\} + (\gamma - \beta)^T \int_0^\tau \{Z_{2i} \mathbb{1}_{\{\zeta < Z_{3i} \le \zeta_0\}}\, dN_i - s_2^{(1)}(]\zeta, \zeta_0], \alpha_0, \beta_0)\, d\Lambda_0\}]$. The sum of the first three terms in this expression is $(\xi - \xi_0)^T \times [n^{-1/2} \sum_{i \le n} \int_0^\tau \{\widetilde{Z}_i(\zeta_0)\, dN_i - s^{(1)}(\theta_0)\, d\Lambda_0\}]$ and

$$
\begin{aligned}
\mathbb{E}_0 &\left\| n^{-1/2} \sum_{i \le n} \int_0^\tau \{\widetilde{Z}_i(\zeta_0)\, dN_i - s^{(1)}(\theta_0)\, d\Lambda_0\} \right\|^2 \\
&\le \mathbb{E}_0 \|\mathbb{M}_n^{(1)}(\tau)\|^2 + \mathbb{E}_0 \left\| \int_0^\tau \mathbb{G}_n^{(1)}(\theta_0)\, d\Lambda_0 \right\|^2
\end{aligned}
$$

which is bounded by $n^{-1}\mathbb{E}_0 \int_0^\tau \|S_n^{(2)}\| \, d\Lambda_0 + n^{-1}\mathbb{E}_0 \int_0^\tau \|S_n^{(2)}(s; 2\theta_0)\| \, \lambda_0^2(s)\, ds$. Then for every n, the mean of the supremum of the first three terms in W_{1n} is $O(\varepsilon)$. For the fourth term in the expression of $W_{1n}(\theta)$,

$$\mathbb{E}_0 \sup_{\zeta \in V_{\varepsilon^2}(\zeta_0)} \left\| n^{-1/2} \sum_{i \le n} \int_0^\tau \{ Z_{2i} \mathbb{1}_{\{\zeta_0 < Z_{3i} \le \zeta\}} \, dN_i - s_2^{(1)}(]\zeta_0, \zeta], \alpha_0, \gamma_0) \, d\Lambda_0 \} \right\|$$

$$\le q^{1/2} \mathbb{E}_0 \sup_{f \in \mathcal{F}_\varepsilon^+} |\nu_n^{U, Z_3}(f)|$$

and a similar bound holds for last term, so they are $O(\varepsilon)$ by Lemma 6. Moreover,

$$W_{2n}(\theta) = \nu_n^{\delta, T}(\varphi_\theta) + n^{-1/2} \sum_i \delta_i \left\{ \log \frac{n^{-1} S_n^{(0)}(T_i; \theta)}{s^{(0)}(T_i; \theta)} - \log \frac{n^{-1} S_n^{(0)}(T_i; \theta_0)}{s^{(0)}(T_i; \theta_0)} \right\}.$$

By a Taylor expansion as $n \to \infty$, the second term is uniformly approximated by

$$n^{-3/2} \sum_{i,j} \delta_i \left\{ \frac{Y_j(T_i) e^{r_\theta(Z_j(T_i))}}{s^{(0)}(T_i; \theta)} - \frac{Y_j(T_i) e^{r_{\theta_0}(Z_j(T_i))}}{s^{(0)}(T_i; \theta_0)} \right\} \{1 + o_{\mathrm{a.s.}}(1)\}$$

where $n^{-3/2} \sum_i \mathbb{E}_0 \sup_{\theta \in V_\varepsilon(\theta_0)} \delta_i \{ e^{r_\theta(Z_i(T_i))} s^{(0)-1}(T_i; \theta) + e^{r_{\theta_0}(Z_i(T_i))} s^{(0)-1}(T_i; \theta_0) \}$ $= o(1)$ and

$$\mathbb{E}_0 \sup_{\theta \in V_\varepsilon(\theta_0)} n^{-3/2} \sum_{i \ne j} \delta_i \left\{ \frac{Y_j(T_i) e^{r_\theta(Z_j(T_i))}}{s^{(0)}(T_i; \theta)} - \frac{Y_j(T_i) e^{r_{\theta_0}(Z_j(T_i))}}{s^{(0)}(T_i; \theta_0)} \right\}$$

$$\le \mathbb{E}_0 \int_0^\tau \sup_{\theta \in V_\varepsilon(\theta_0)} n^{-1/2} \sum_j \left\{ \frac{Y_j(t) e^{r_\theta(Z_j(t))}}{s^{(0)}(t; \theta)} - \frac{Y_j(t) e^{r_{\theta_0}(Z_j(t))}}{s^{(0)}(t; \theta_0)} \right\}$$

$$\times s^{(0)}(t; \theta_0) \, d\Lambda_0(t)$$

by an integration conditionally on (Y_j, Z_j), $1 \le j \le n$. Splitting the integrand in the right-hand side of this inequality into a sum of four terms according to the location of the variables Z_{3j} with respect to ζ and ζ_0, we see that the last expression is bounded by $b_1 + \cdots + b_4 = O(\varepsilon)$ from Lemma 6, with

$$b_k = \int_0^\tau \left\{ \mathbb{E}_0 \sup_{\theta \in V_\varepsilon(\theta_0)} \nu_n^t(\psi_{k,t,\theta}) \right\} s^{(0)}(t; \theta_0) \, d\Lambda_0(t). \qquad \square$$

PROOF OF LEMMA 4. Since $X(\theta_0) = \dot{X}_\xi(\theta_0) = 0$, by a Taylor expansion for ε sufficiently small and for θ in $V_\varepsilon(\theta_0)$,

$$X(\theta) = -|\zeta - \zeta_0| \dot{X}_\zeta^-(\theta_0) - \tfrac{1}{2}(\xi - \xi_0)^T I(\theta^*)(\xi - \xi_0)$$
$$+ o(|\zeta - \zeta_0|) \qquad \text{if } \zeta < \zeta_0,$$

$$X(\theta) = |\zeta - \zeta_0| \dot{X}_\zeta^+(\theta_0) - \tfrac{1}{2}(\xi - \xi_0)^T I(\theta^*)(\xi - \xi_0) \\ + o(|\zeta - \zeta_0|) \qquad \text{if } \zeta > \zeta_0,$$

where θ^* is between θ and θ_0 and with (3.1) and (3.2). The matrix $I(\theta^*)$ is positive definite for all θ^* in a neighborhood of θ_0 and by Lemma 1 $\|I(\theta) - I(\theta_0)\|$ tends to zero as $\rho(\theta - \theta_0) \to 0$. Moreover $\dot{X}_\zeta^-(\theta_0)$ is strictly positive if $\zeta < \zeta_0$ and strictly negative if $\zeta > \zeta_0$ (cf. proof of Theorem 1). The result follows if ε is sufficiently small. $\square$

PROOF OF LEMMA 5. Let $\mu = h_3(\zeta_0)$. For s in $\mathbb{R}_+$,

$$\phi_s^+(t) = \mathbb{E}_0\big[\mathbb{E}_0\{e^{itQ^+(s)} \mid \nu^+(s)\}\big] = e^{-\mu s} \sum_{j \geq 0} \frac{(\mu s)^j}{j!} \mathbb{E}_0 e^{it \sum_{0 \leq k \leq j} V_k^+}$$

$$= e^{-\mu s} \sum_{j \geq 0} \frac{(\mu s \varphi^+(t))^j}{j!}$$

from the independence assumptions, and the proof is similar for ϕ_s^-. From (4.2) and by the mean value theorem, there exist x_1 and x_2 lying strictly between β_0 and γ_0 such that

$$\mathbb{E}_0 V_k^+ = \int_0^\tau \mathbb{E}_0\Big\{(\beta_0 - \gamma_0)^T Z_2 Y e^{\alpha_0^T Z_1 + \gamma_0^T Z_2} \\ - Y e^{\alpha_0^T Z_1}\big(e^{\beta_0^T Z_2} - e^{\gamma_0^T Z_2}\big) \mid Z_3 = \zeta_0^+\Big\}\, d\Lambda_0$$

$$= (\beta_0 - \gamma_0)^T \bigg[\int_0^\tau \mathbb{E}_0\Big\{Z_2^{\otimes 2} Y e^{\alpha_0^T Z_1 + x_2^T Z_2} \mid Z_3 = \zeta_0^+\Big\}\, d\Lambda_0\bigg](\gamma_0 - x_1)$$

and therefore $\mathbb{E}_0 V_k^+$ is strictly negative. By the same arguments, $\mathbb{E}_0 V_k^-$ is strictly positive, and the sums $\sum_{j \geq 0} V_k^+$ and $\sum_{j \geq 0} -V_k^-$ converge a.s. to $-\infty$ and the maximum value $\widehat{v}_Q$ of the process Q is a.s. finite. $\square$

PROOF OF THEOREM 3. Let $A > 0$, $u = (u_1, u_2) \in \mathcal{U}_n^A$, u_1 in $\mathbb{R}$ and u_2 in $\mathbb{R}^{p+2q}$, and let $\theta_{n,u} = (\zeta_{n,u}, \xi_{n,u}^T)^T$ with $\zeta_{n,u} = \zeta_0 + n^{-1} u_1$ and $\xi_{n,u} = \xi_0 + n^{-1/2} u_2$. For $1 \leq i \leq n$,

$$\big(r_{\theta_{nu}} - r_{\theta_0}\big)(Z_i(T_i)) \\ = n^{-1/2}\big\{u_2^T \widetilde{Z}_i(T_i; \zeta_{nu}) \\ + (\gamma_0 - \beta_0)^T Z_{2i}(T_i)\big(\mathbb{1}_{\{\zeta_{nu} < Z_{3i} \leq \zeta_0\}} - \mathbb{1}_{\{\zeta_0 < Z_{3i} \leq \zeta_{nu}\}}\big)\big\}.$$

Using C3 and the continuity of the functions $s^{(k)}$, a Taylor expansion for ξ_{nu} close to ξ_0 gives

$$S_n^{(0)}(\theta_{nu}) = S_n^{(0)}(\zeta_{nu}, \xi_0) + n^{-1/2} u_2^T S_n^{(1)}(\zeta_{nu}, \xi_0) + \tfrac{1}{2} n^{-1} u_2^T S_n^{(2)}(\zeta_{nu}, \xi_0) u_2 + o_{\text{a.s.}}(1),$$

$$\log \frac{S_n^{(0)}(\theta_{nu})}{S_n^{(0)}(\theta_0)} = \frac{S_n^{(0)}(\zeta_{nu}, \xi_0) - S_n^{(0)}(\theta_0)}{S_n^{(0)}(\theta_0)} + n^{-1/2} u_2^T \frac{S_n^{(1)}(\zeta_{nu}, \xi_0)}{S_n^{(0)}(\theta_0)} + \frac{n^{-1}}{2} u_2^T V_n(\zeta_{nu}, \xi_0) u_2 + o_{\text{a.s.}}(1)$$

uniformly on $\mathcal{U}_n^A$ as $n \to \infty$. By the uniform convergence of $n^{-1}\bar{N}_n$ to $\int_0^{\cdot} s^{(0)}(\theta_0)\, d\Lambda_0$, we obtain $l_n(\theta_{n,u}) - l_n(\theta_0) = u_2^T C_n(u) - \frac{1}{2} u_2^T I(\theta_0) u_2 + Q_n(u_1) + o_p(1)$ uniformly on $\mathcal{U}_n^A$, where

$$\begin{aligned} C_n(u) &= n^{-1/2} \sum_i \int_0^\tau \left\{ \widetilde{Z}_i(\zeta_{nu}) - \frac{S_n^{(1)}(\zeta_{nu}, \xi_0)}{S_n^{(0)}(\theta_0)} \right\} dN_i \\ &= n^{-1/2} \sum_i \int_0^\tau \left\{ \widetilde{Z}_i(\zeta_{nu}) - \frac{S_n^{(1)}(\zeta_{nu}, \xi_0)}{S_n^{(0)}(\theta_0)} \right\} dM_i \\ &\quad + n^{-1/2} \int_0^\tau \Big\{ S_n^{(1)}(]\zeta_0, \zeta_{nu}], \alpha_0, \gamma_0) - S_n^{(1)}(]\zeta_0, \zeta_{nu}], \alpha_0, \beta_0) \\ &\qquad + S_n^{(1)}(]\zeta_{nu}, \zeta_0], \alpha_0, \beta_0) - S_n^{(1)}(]\zeta_{nu}, \zeta_0], \alpha_0, \gamma_0) \Big\}\, d\Lambda_0. \end{aligned}$$

Let $a_{1i}(u_1) = \int_0^\tau \{\widetilde{Z}_i(\zeta_{nu}) - \widetilde{Z}_i(\zeta_0)\}\, dM_i$ and $a_{2i}(u_1) = \int_0^\tau \{S_n^{(1)}(\zeta_{nu}, \xi_0) - S_n^{(1)}(\theta_0)\} S_n^{(0)-1}(\theta_0)\, dM_i$, $1 \le i \le n$. The variables $a_{1i}(u_1)$ and $\sum_i a_{2i}(u_1)$ have mean zero as integrals of predictable processes with respect to M_i and $\sum_i M_i$, respectively. Let $\mathbf{0}_q$ be the zero q-dimensional vector. Then

$$\begin{aligned} &\widetilde{Z}_i(t; \zeta_{nu}) - \widetilde{Z}_i(t; \zeta_0) \\ &\quad = \big(\mathbf{0}_p, Z_{2i}^T(t), -Z_{2i}^T(t)\big)^T \big(\mathbb{1}_{\{\zeta_0 < Z_{3i} \le \zeta_{nu}\}} - \mathbb{1}_{\{\zeta_{nu} < Z_{3i} \le \zeta_0\}}\big), \\ &S_n^{(1)}(t; \zeta_{nu}, \xi_0) - S_n^{(1)}(t; \theta_0) \\ &\quad = \sum_i Y_i(t) e^{\alpha_0^T Z_{1i}(t)} \big(\mathbb{1}_{\{\zeta_0 < Z_{3i} \le \zeta_{nu}\}} - \mathbb{1}_{\{\zeta_{nu} < Z_{3i} \le \zeta_0\}}\big) \\ &\qquad \times \big\{\big(Z_{1i}^T(t), Z_{2i}^T(t), \mathbf{0}_q^T\big)^T\big\} e^{\beta_0^T Z_{2i}(t)} - (Z_{1i}^T, \mathbf{0}_q^T, Z_{2i}^T)^T e^{\gamma_0^T Z_{2i}(t)}, \end{aligned}$$

for $\ell = 1, 2$, $\mathbb{E}_0 \sup_{|u_1| \le A} \|a_{\ell i}(u_1)\|^2 = O(n^{-1})$; therefore $\mathbb{E}_0 \sup_{|u_1| \le A} n^{-1/2} \times \sum_i a_{\ell i}(u_1) = O(n^{-1})$, since the supremum is over the intervals $]\zeta_0, \zeta_{nu}]$ and $]\zeta_{nu}, \zeta_0]$, and the result follows. $\square$

REFERENCES

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120.

BAILEY, K. R. (1983). The asymptotic joint distribution of regression and survival parameter estimates in the Cox regression model. *Ann. Statist.* **11** 39–48.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BRESLOW, N. E. (1972). Discussion of "Regression model and life-tables," by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* **34** 216–217.

COX, D. R. (1972). Regression model and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, New York.

IBRAGIMOV, I. and HAS'MINSKII, R. (1981). *Statistical Estimation*: *Asymptotic Theory*. Springer, New York.

JESPERSEN, N. C. B. (1986). Dichotomizing a continuous covariate in the Cox model. Research Report 86/02, Statistical Research Unit, Univ. Copenhagen.

KLEINBAUM, D. G. (1996). *Survival Analysis*: *A Self-Learning Text*. Springer, New York.

KUTOYANTS, Y. A. (1984). *Parameter Estimation for Stochastic Processes*. Heldermann, Berlin.

KUTOYANTS, Y. A. (1998). *Statistical Inference for Spatial Poisson Processes. Lecture Notes in Statist.* **134**. Springer, New York.

LENGLART, E. (1977). Relation de domination entre deux processus. *Ann. Inst. H. Poincaré Sect. B* (*N. S.*) **13** 171–179.

LIANG, K.-Y., SELF, S. and LIU, X. (1990). The Cox proportional hazards model with change point: An epidemiologic application. *Biometrics* **46** 783–793.

LUO, X. (1996). The asymptotic distribution of MLE of treatment lag threshold. *J. Statist. Plann. Inference* **53** 33–61.

LUO, X. and BOYETT, J. (1997). Estimation of a threshold parameter in Cox regression. *Comm. Statist. Theory Methods* **26** 2329–2346.

LUO, X., TURNBULL, B. and CLARK, L. (1997). Likelihood ratio tests for a change point with survival data. *Biometrika* **84** 555–565.

MATTHEWS, D. E., FAREWELL, V. T. and PYKE, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. *Ann. Statist.* **13** 583–591.

NÆS, T. (1982). The asymptotic distribution of the estimator for the regression parameter in Cox's regression model. *Scand. J. Statist*. **9** 107–115.

NGUYEN, H. T., ROGERS, G. S. and WALKER, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika* **71** 299–304.

POLLARD, D. (1989). Asymptotics via empirical processes (with discussion). *Statist. Sci.* **4** 341–366.

PONS, O. (2002). Estimation in a Cox regression model with a change-point at an unknown time. *Statistics* **36** 101–124.

PONS, O. and DE TURCKHEIM, E. (1988). Cox's periodic regression model. *Ann. Statist.* **16** 678–693.

PRENTICE, R. L. and SELF, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 804–813.

REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete* **51** 269–286.

TSIATIS, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9** 93–108.

VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

YAO, Y.-C. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Comm. Statist. Theory Methods* **15** 2455–2466.

LABORATOIRE DE STATISTIQUE
INRA
BIOMÉTRIE
78352 JOUY-EN-JOSAS CEDEX
FRANCE
E-MAIL: Odile.Pons@jouy.inra.fr