Statistica Sinica **25** (2015), 1163-1184
doi:http://dx.doi.org/10.5705/ss.2013.046

A LIKELIHOOD RATIO TEST FOR MONOTONE BASELINE HAZARD FUNCTIONS IN THE COX MODEL

Gabriela F. Nane

Delft University of Technology

Abstract: We consider a likelihood ratio method for testing whether a monotone baseline hazard function in the Cox model has a particular value at a fixed point. We derive the asymptotic distribution of the likelihood ratio statistic, which is identical for a nondecreasing and a nonincreasing baseline hazard. The asymptotic distribution of the likelihood ratio test enables, via inversion, the construction of pointwise confidence intervals. Simulations show that these confidence intervals exhibit comparable coverage probabilities but shorter length, on average, than the confidence intervals based on the asymptotic distribution of the nonparametric maximum likelihood estimator of a monotone baseline hazard function.

Key words and phrases: Cox model, likelihood ratio test, nonparametric maximum likelihood estimation, shape constrained estimators.

1. Introduction

In survival analysis, Cox (1972) proportional hazards model is the typical choice to account for the effect of covariates on the lifetime distribution. Its attractiveness resides in its form that allows for efficient estimation of the regression coefficient, while leaving the baseline distribution completely unspecified, see e.g., Efron (1977), Oakes (1977) and Slud (1982). The regression coefficient estimator is the well-known maximum partial likelihood estimator $\hat{\beta}_n$, Cox (1972, 1975). In his discussion of Cox's paper, Breslow proposed a different approach that yields the maximum partial likelihood estimator $\hat{\beta}_n$, as well as Λ_n, the NPMLE of the baseline cumulative hazard function Λ_0. An impressive amount of research rapidly followed Cox's seminal paper, which focused primarily on deriving the (asymptotic) properties of $\hat{\beta}_n$, as well as of the Breslow estimator Λ_n.

Even though the baseline hazard λ_0 is usually left completely unspecified, there are circumstances in which one might be interested in restricting λ_0 qualitatively. This can be done by assuming the baseline hazard to be monotone, for example, as suggested by Cox (1972) himself. Various studies have indicated that a monotonicity constraint can be imposed occasionally on the baseline hazard, which complies in these situations with the medical expertise. For an illustration of a nonincreasing baseline hazard estimator in the study of patients with acute coronary syndrome, see van Geloven et al. (2013).

Lopuhaä and Nane (2013) proposed a nonparametric maximum likelihood estimator and a Grenander type estimator of a monotone baseline hazard function. The Grenander type estimator is defined in terms of slopes of the greatest convex minorant of the Breslow estimator Λ_n. The two estimators have been proven strongly consistent and have been shown to exhibit the same distributional law. Furthermore, at a fixed point x_0, the scaled difference between the maximum likelihood estimator $\hat{\lambda}_n$ and the true baseline hazard λ_0 converges to the distribution of the minimum of two-sided Brownian motion plus a parabola times a constant depending on the underlying parameters. These results adhere to the general nonparametric shape constrained theory and, in particular, prolong naturally the findings of Huang and Wellner (1995) in the case of the random censorship model with no covariates.

Ensuing inference is pursued in this paper, by testing the hypothesis that the underlying monotone baseline hazard has a particular value θ_0, at a fixed point x_0. We use a likelihood ratio test of $H_0 : \lambda_0(x_0) = \theta_0$ versus $H_1 : \lambda_0(x_0) \neq \theta_0$. For the shape restricted problems, this approach was initially employed for monotone distributions in the current status model by Banerjee and Wellner (2001). The authors focused on deriving the limiting distribution of the likelihood ratio test under the null hypothesis, and to obtaining a so-called fixed universal distribution, defined in terms of slopes of the greatest convex minorant of the two-sided Brownian motion plus a parabola. These findings were followed by a stream of research, see, e.g., Banerjee and Wellner (2005), Banerjee (2007), and Banerjee (2008), showing that the likelihood ratio method can be extended in other shape constrained settings.

In this paper, we carry on this research for the monotone baseline hazard function λ_0 in the Cox model. In addition to directly extending the results of Banerjee (2008) in the right censoring model with no covariates, we aim to provide a thorough description of the method and detailed proofs for the results. The likelihood ratio method described here can be applied in other semiparametric models, including extensions of the monotone response models described by Banerjee (2007), such as the partially linear regression and the semiparametric logistic regression model.

Furthermore, based on the likelihood ratio method, we derive confidence sets for $\lambda_0(x_0)$. More specifically, we have that inverting the family of tests can yield, in turn, pointwise confidence intervals for the baseline hazard function. Another approach to pointwise confidence intervals is based on the asymptotic distribution, at a fixed point x_0, of the nonparametric maximum likelihood estimator $\hat{\lambda}_n$, derived by Lopuhaä and Nane (2013). Nonetheless, the method based on the asymptotic distribution entails estimating the nuisance parameter, and more specifically, estimating the derivative of the baseline hazard function $\lambda_0'(x_0)$. This proves to be a bothersome issue, since, to the author's best knowledge, there is no available smooth monotone estimator of the baseline hazard function in the Cox model. One option would be to kernel smooth the NPMLE $\hat{\lambda}_n$, but this would pose such extra difficulties, as an appropriate choice of a bandwidth. For a discussion of these issues in the right censoring model, see Banerjee (2008).

The paper is organized as follows. Section 2 introduces the Cox model, the notations, and the common assumptions. In Section 3, we introduce the likelihood ratio method and characterize the maximum likelihood estimator $\hat{\lambda}_n$ of a nondecreasing baseline hazard function and the estimator $\hat{\lambda}_n^0$, such that $\hat{\lambda}_n^0(x_0) = \theta_0$, for a fixed x_0 in the interior of the support of the baseline distribution. The asymptotic distribution of the likelihood ratio statistic is provided, along with preparatory lemmas, in Section 4. Finally, Section 5 is devoted to constructing pointwise confidence intervals and comparing them, via simulations, with the confidence intervals based on the asymptotic distribution of the NPMLE $\hat{\lambda}_n$.

The proofs of some results are deferred to a supplement, which is available online. Moreover, the Supplement contains the characterization of the estimators for the nonincreasing baseline hazard function λ_0.

2. Definitions and Assumptions

Suppose that the observed data consist of the independent and identically distributed triplets (T_i, Δ_i, Z_i), with $i = 1, \ldots, n$. The event time, denoted by X and commonly referred to as the survival time is subject to random censoring. Thus, $T = \min(X, C)$, where T is the follow-up time and C denotes the censoring time. The indicator $\Delta = \{X \leq C\}$ marks whether the follow-up time is an event or a censoring time. Finally, $Z \in \mathbb{R}^p$ denotes the covariate vector of the observed follow-up time T, which is assumed to be time invariant. The event time X and censoring time C are assumed to be conditionally independent, given the covariate vector Z. Let F be the distribution function of the non-negative random variable X, G the distribution function of the non-negative random variable C, and H the distribution function of T. The distribution function $F(x|z)$ is assumed to be absolutely continuous, with density $f(x|z)$. Similarly, the distribution function $G(c|z)$ is assumed to be absolutely continuous, with density $g(c|z)$. In addition, $F(x|z)$ and $G(c|z)$ share no parameters, thus the censoring mechanism is assumed to be non-informative.

Let $\lambda(x|z)$ be the hazard function of an individual with covariate vector $z \in \mathbb{R}^p$. The Cox model specifies that

$$\lambda\,(x|z) = \lambda_0(x)\,\mathrm{e}^{\beta_0' z}, \tag{2.1}$$

where λ_0 represents the baseline hazard function, that corresponds to $z = 0$, and $\beta_0 \in \mathbb{R}^p$ is the vector of the underlying regression coefficients. Finally, we consider the assumptions that are typically employed when deriving large sample properties of estimators within the Cox model; e.g., see Tsiatis (1981).

(A1) Let τ_F, τ_G and τ_H be the end points of the support of F, G and H respectively. Then,

$$\tau_H = \tau_G < \tau_F \leq \infty.$$

(A2) There exists $\varepsilon > 0$ such that

$$\sup_{|\beta-\beta_0|\leq\varepsilon} \mathbb{E}\left[|Z|^2\, \mathrm{e}^{2\beta' Z}\right] < \infty,$$

where $|\cdot|$ denotes the Euclidean norm.

3. The Likelihood Ratio and the Characterization of the Estimators

By definition, $\Lambda(x|z) = -\log(1 - F(x|z))$ is the cumulative hazard function. Thus, from (2.1), it follows that $\Lambda(x|z) = \Lambda_0(x)\exp(\beta_0' z)$, where $\Lambda_0(x) = \int_0^x \lambda_0(u)\mathrm{d}u$ is the baseline cumulative hazard function. Since, for a continuous distribution, $\lambda(t) = f(t)/(1 - F(t))$, for $t \geq 0$, the full likelihood is given by

$$\prod_{i=1}^{n} \{f(T_i \mid Z_i)\,[1 - G(T_i \mid Z_i)]\}^{\Delta_i}\, \{g(T_i \mid Z_i)\,[1 - F(T_i \mid Z_i)]\}^{1-\Delta_i}$$
$$= \prod_{i=1}^{n} \lambda(T_i \mid Z_i)^{\Delta_i} \exp\left[-\Lambda(T_i \mid Z_i)\right] \times \prod_{i=1}^{n} [1 - G(T_i \mid Z_i)]^{\Delta_i}\, g(T_i \mid Z_i)^{1-\Delta_i}.$$

As the censoring mechanism is assumed to be non-informative, and by (2.1), maximizing the full likelihood is the same as maximizing

$$\prod_{i=1}^{n} \lambda(T_i \mid Z_i)^{\Delta_i} \exp\left[-\Lambda(T_i \mid Z_i)\right] = \prod_{i=1}^{n} \left[\lambda_0(T_i)\mathrm{e}^{\beta_0' Z_i}\right]^{\Delta_i} \exp\left[-\mathrm{e}^{\beta_0' Z_i}\Lambda_0(T_i)\right],$$

which yields the following (pseudo) loglikelihood function, written as a function of $\beta \in \mathbb{R}^p$ and λ_0,

$$\sum_{i=1}^{n} \left[\Delta_i \log\lambda_0(T_i) + \Delta_i \beta' Z_i - \mathrm{e}^{\beta' Z_i}\Lambda_0(T_i)\right].$$

Let $T_{(1)} < T_{(2)} < \cdots < T_{(n)}$ be the ordered follow-up times and, for $i = 1, \ldots, n$, let $\Delta_{(i)}$ and $Z_{(i)}$ be the censoring indicator and covariate vector corresponding to $T_{(i)}$. Writing the above (pseudo) likelihood as a function of β and λ_0 gives

$$L_\beta(\lambda_0) = \sum_{i=1}^{n} \left[\Delta_{(i)} \log\lambda_0(T_{(i)}) + \Delta_{(i)}\beta' Z_{(i)} - \mathrm{e}^{\beta' Z_{(i)}} \int_0^{T_{(i)}} \lambda_0(u)\mathrm{d}u\right]. \qquad (3.1)$$

Following the approach of Lopuhaä and Nane (2013), we do not proceed with the joint maximization of (3.1) over β and monotone λ_0. Alternatively, for $\beta \in \mathbb{R}^p$ fixed, we consider maximum likelihood estimation of a monotone baseline hazard function λ_0, and denote the estimator by $\hat{\lambda}_n(x;\beta)$. Subsequently, we replace β by $\hat{\beta}_n$, the maximum partial likelihood estimator, due to its commendable asymptotic properties (see, e.g., Efron (1977), Oakes (1977) and Slud (1982)). The proposed NPMLE is thus $\hat{\lambda}_n(x) = \hat{\lambda}_n(x;\hat{\beta}_n)$ and is referred to as the unconstrained estimator of a monotone λ_0. Furthermore, for $\beta \in \mathbb{R}^p$ fixed, we maximize the loglikelihood function $L_\beta(\lambda_0)$ in (3.1) over the class of all monotone baseline hazard functions, under the null hypothesis $H_0 : \lambda_0(x_0) = \theta_0$, for $x_0 \in (0, \tau_H)$ and $\theta_0 \in (0,\infty)$, fixed. We obtain $\hat{\lambda}_n^0(x;\beta)$ and hence propose $\hat{\lambda}_n^0(x) = \hat{\lambda}_n^0(x;\hat{\beta}_n)$ as the constrained NPMLE.

Replacing β by $\hat{\beta}_n$ in the loglikelihood function (3.1) yields the likelihood ratio statistic for testing $H_0 : \lambda_0(x_0) = \theta_0$,

$$2 \log \xi_n(\theta_0) = 2L_{\hat{\beta}_n}(\hat{\lambda}_n) - 2L_{\hat{\beta}_n}(\hat{\lambda}_n^0). \tag{3.2}$$

Thus, for computing the likelihood ratio statistic, we need to characterize the unconstrained NPMLE $\hat{\lambda}_n$ and the constrained NPMLE $\hat{\lambda}_n^0$ of a monotone baseline hazard function λ_0.

3.1. Nondecreasing baseline hazard

We first consider maximum likelihood estimation of a nondecreasing baseline hazard function λ_0. Both the unconstrained estimator $\hat{\lambda}_n$ and the constrained estimator $\hat{\lambda}_n^0$ are characterized in terms of the processes

$$W_n(\beta, x) = \int \left(\mathrm{e}^{\beta' z} \int_0^x \{u \geq s\} \mathrm{d}s \right) \mathrm{d}P_n(u, \delta, z), \tag{3.3}$$

$$V_n(x) = \int \delta\{u < x\} \mathrm{d}P_n(u, \delta, z), \tag{3.4}$$

with $\beta \in \mathbb{R}^p$ and $x \geq 0$, and where P_n is the empirical measure of the (T_i, Δ_i, Z_i), with $i = 1, \ldots, n$. The characterization of the unconstrained estimator $\hat{\lambda}_n(x;\beta)$ has already been provided in Lemma 1 in Lopuhaä and Nane (2013), which we restate below. We also provide a closed form of the estimator on blocks of indices on which the estimator is constant.

Lemma 1. *Let $T_{(1)} < \ldots < T_{(n)}$ be the ordered follow-up times and consider a fixed $\beta \in \mathbb{R}^p$.*

(i) *Let* W_n *and* V_n *be as in* (3.3) *and* (3.4). *Then the NPMLE* $\hat{\lambda}_n(x;\beta)$ *of a nondecreasing baseline hazard function* λ_0 *is of the form*

$$\hat{\lambda}_n(x;\beta) = \begin{cases} 0 & x < T_{(1)}, \\ \hat{\lambda}_i & T_{(i)} \le x < T_{(i+1)},\ \textit{for } i = 1,\ldots,n-1, \\ \infty & x \ge T_{(n)}, \end{cases}$$

where $\hat{\lambda}_i$ *is the left derivative of the greatest convex minorant* (*GCM*) *at the point* P_i *of the cumulative sum diagram* (*CSD*) *consisting of the points*

$$P_j = \Big(W_n(\beta, T_{(j+1)}) - W_n(\beta, T_{(1)}), V_n(T_{(j+1)}) \Big), \tag{3.5}$$

for $j = 1,\ldots,n-1$ *and* $P_0 = (0,0)$.

(ii) *For* $k \ge 1$, *let* $B_1,\ldots,B_k$ *be blocks of indices such that* $\hat{\lambda}_n(x;\beta)$ *is constant on each block and* $B_1 \cup \ldots \cup B_k = \{1,\ldots,n-1\}$. *Denote by* $v_{nj}(\beta)$ *the value of* $\hat{\lambda}_n(x;\beta)$ *on block* B_j. *Then,*

$$v_{nj}(\beta) = \frac{\sum_{i \in B_j} \Delta_{(i)}}{\sum_{i \in B_j} \left[T_{(i+1)} - T_{(i)} \right] \sum_{l=i+1}^{n} e^{\beta' Z_{(l)}}}. \tag{3.6}$$

The proof can be found in the Supplement. As mentioned beforehand, the proposed unconstrained estimator is thus $\hat{\lambda}_n(x) = \hat{\lambda}_n(x;\hat{\beta}_n)$. Equivalently, on each block of indices B_j, for $j = 1,\ldots,k$, we propose the estimate $\hat{v}_{nj} = v_{nj}(\hat{\beta}_n)$. Under the null hypothesis $H_0 : \lambda_0(x_0) = \theta_0$, the characterization of the constrained maximum likelihood estimator $\hat{\lambda}_n^0$ is provided by the next lemma. The proof of the lemma can be found in the Supplement.

Lemma 2. *Let* $x_0 \in (0, \tau_H)$ *be such that* $T_{(m)} < x_0 < T_{(m+1)}$ *for a given* $1 \le m \le n-1$. *Consider a fixed* $\beta \in \mathbb{R}^p$.

(i) *For* $i = 1,\ldots,m$, *let* $\hat{\lambda}_i^L$ *be the left derivative of the GCM at the point* P_i^L *of the CSD consisting of the points* $P_j^L = P_j$, *for* $j = 1,\ldots,m$, *with* P_j *defined in* (3.5) *and* $P_0^L = (0,0)$. *For* $i = m+1,\ldots,n-1$, *let* $\hat{\lambda}_i^R$ *be the left derivative of the GCM at the point* P_i^R *of the CSD consisting of the points* $P_j^R = P_j$, *for* $j = m,\ldots,n-1$, *with* P_j *defined in* (3.5). *Then, for* $\theta_0 \in (0,\infty)$, *the NPMLE* $\hat{\lambda}_n^0(x;\beta)$ *of a nondecreasing baseline hazard function* λ_0, *under the null hypothesis* $H_0 : \lambda_0 = \theta_0$, *is of the form*

$$\hat{\lambda}_n^0(x;\beta) = \begin{cases} 0 & x < T_{(1)}, \\ \hat{\lambda}_i^0 & T_{(i)} \le x < T_{(i+1)},\ \textit{for } i \in \{1,\ldots,n-1\} \setminus \{m\}, \\ \hat{\lambda}_m^0 & T_{(m)} \le x < x_0, \\ \theta_0 & x_0 \le x < T_{(m+1)}, \\ \infty & x \ge T_{(n)}, \end{cases} \tag{3.7}$$

where $\hat{\lambda}_i^0 = \min(\hat{\lambda}_i^L, \theta_0)$ *for* $i = 1, \ldots, m$, *and* $\hat{\lambda}_i^0 = \max(\hat{\lambda}_i^R, \theta_0)$ *for* $i = m+1, \ldots, n-1$.

(ii) *For* $k \geq 1$, *let* $B_1^0, \ldots, B_k^0$ *be blocks of indices such that* $\hat{\lambda}_n^0(x;\beta)$ *is constant on each block and* $B_1^0 \cup \ldots \cup B_k^0 = \{1, \ldots, n-1\}$. *Then, there is one block, say* B_r^0, *on which* $\hat{\lambda}_n^0(x;\beta)$ *is equal to* θ_0, *and one block, say* B_p^0, *that contains* m. *On all other blocks* B_j^0, *denote by* $v_{nj}^0(\beta)$ *the value of* $\hat{\lambda}_n^0(x;\beta)$ *on block* B_j^0. *Then,*

$$v_{nj}^0(\beta) = \frac{\sum_{i \in B_j^0} \Delta_{(i)}}{\sum_{i \in B_j^0} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^n e^{\beta' Z_{(l)}}}, \tag{3.8}$$

for $j = 1, \ldots, p-1, p+1, \ldots, k$. *On the block* B_p^0 *that contains* m,

$$v_{np}^0(\beta) = \frac{\sum_{i \in B_p^0} \Delta_{(i)}}{\sum_{i \in B_p^0 \setminus \{m\}} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^n e^{\beta' Z_{(l)}} + [x_0 - T_{(m)}] \sum_{l=m+1}^n e^{\beta' Z_{(l)}}}. \tag{3.9}$$

Similar to the unconstrained estimator, we propose $\hat{\lambda}_n^0(x) = \hat{\lambda}_n^0(x; \hat{\beta}_n)$ as the constrained estimator and $\hat{v}_{nj}^0 = v_{nj}^0(\hat{\beta}_n)$, where $\hat{\beta}_n$ is the maximum partial likelihood estimator.

Remark 1. As already pointed out by Lopuhaä and Nane (2013), if we take all covariates equal to zero, the characterization of the unconstrained estimator differs slightly from the characterization of the nondecreasing hazard estimator in the ordinary random censorship model provided by Huang and Wellner (1995). Correspondingly, the characterizations in Lemma 1 and 2, with all $Z_l \equiv 0$ differ from the characterizations provided by Banerjee (2008) in the right censored model. Although the estimators in Banerjee (2008) do not maximize the (pseudo) loglikelihood function in (3.1) (in the absence of covariates and under the null hypothesis) over nondecreasing λ_0, the asymptotic distribution of the likelihood ratio test based on these estimators coincide with our proposed distribution in the case of no covariates.

Using the notations in Banerjee (2008), let $\text{slogcm}(f, I)$ be the left-hand slope of the greatest convex minorant of the restriction of the real-valued function f to the interval I. Denote by $\text{slogcm}(f) = \text{slogcm}(f, \mathbb{R})$. Moreover, let

$$\text{slogcm}^0(f) = \min\left(\text{slogcm}(f, (-\infty, 0]), 0\right) 1_{(-\infty,0]} + \max\left(\text{slogcm}(f, (0, \infty)), 0\right) 1_{(0,\infty)}.$$

For positive constants a and b, define

$$X_{a,b}(t) = a\mathbb{W}(t) + bt^2, \tag{3.10}$$

where $\mathbb{W}$ is a standard two-sided Brownian motion originating from zero. Let

$$g_{a,b}(t) = \text{slogcm}(X_{a,b})(t), \tag{3.11}$$

the left-hand slope of the GCM $G_{a,b}$ of the process $X_{a,b}$, at point t. The constrained analogous is defined as follows: for $t \leq 0$, construct the GCM of $X_{a,b}$, denoted by $G^L_{a,b}$, and take its left-hand slopes at point t, denoted by $D_L(X_{a,b})(t)$. When the slopes exceed zero, replace them by zero. In the same manner, for $t > 0$, denote the GCM of $X_{a,b}$ by $G^R_{a,b}$ and its slopes at point t by $D_R(X_{a,b})(t)$. Replace the slopes by zero when they decrease below zero. This slope process is denoted by $g^0_{a,b}$, and

$$g^0_{a,b}(t) = \begin{cases} \min\left(D_L(X_{a,b})(t), 0\right) & t < 0, \\ 0 & t = 0, \\ \max\left(D_R(X_{a,b})(t), 0\right) & t > 0. \end{cases} \tag{3.12}$$

Note that for $t \leq 0$, there exists, almost surely $s < 0$ such that $D_L(X_{a,b})(s)$ is strictly positive for any point greater than or equal to s and the left derivative at s is non-positive. Equivalently, for $t > 0$ there exists almost surely $s > 0$ such that $D_R(X_{a,b})(s)$ is strictly negative for any point smaller than or equal to s and the left derivative at s is non-negative. In addition, observe that $g^0_{a,b}(t) = \text{slogcm}^0(X_{a,b})(t)$, as defined and characterized by Banerjee and Wellner (2001).

The characterization of the unconstrained and constrained estimators for nonincreasing baseline hazard functions is similar to the nondecreasing case and can be found in the Supplement.

4. The Limit Distribution

Let $B_{loc}(\mathbb{R})$ be the space of all locally bounded real functions on $\mathbb{R}$, equipped with the topology of uniform convergence on compact sets. Take $\mathbb{C}_{min}(\mathbb{R})$ to be the subset of $B_{loc}(\mathbb{R})$ consisting of continuous functions f for which $f(t) \to \infty$ when $|t| \to \infty$, and f has a unique minimum. Let $\mathcal{L}$ be the space of locally square integrable real-valued functions on $\mathbb{R}$, equipped with the topology of L_2 convergence on compact sets.

For a generic follow-up time T, consider $H^{uc}(x) = \mathbb{P}(T \leq x, \Delta = 1)$, the subdistribution function of the uncensored observations. Moreover, let

$$\Phi(\beta, x) = \int \{u \geq x\}\, \mathrm{e}^{\beta' z} \mathrm{d}P(u, \delta, z), \tag{4.1}$$

for $\beta \in \mathbb{R}^p$ and $x \in \mathbb{R}$, where P is the underlying probability measure corresponding to the distribution of (T, Δ, Z). For a fixed point $x_0 \in (0, \tau_H)$, define the processes

$$X_n(x) = n^{1/3}\left(\hat{\lambda}_n(x_0 + n^{-1/3}x) - \theta_0\right),$$
$$Y_n(x) = n^{1/3}\left(\hat{\lambda}_n^0(x_0 + n^{-1/3}x) - \theta_0\right). \tag{4.2}$$

Our result gives the joint asymptotic distribution of these processes. Its proof is deferred to the online Supplement.

Lemma 3. *Assume* (A1) *and* (A2) *and let* $x_0 \in (0, \tau_H)$. *Suppose that* λ_0 *is nondecreasing on* $[0, \infty)$ *and continuously differentiable in a neighborhood of* x_0, *with* $\lambda_0(x_0) \neq 0$ *and* $\lambda_0'(x_0) > 0$, *and assume that the functions* $x \mapsto \Phi(\beta_0, x)$ *and* $H^{uc}(x)$ *defined at* (4.1) *are continuously differentiable in a neighborhood of* x_0. *If the density of the follow-up times is continuous and bounded away from zero in a neighborhood of* x_0, *and*

$$a = \sqrt{\frac{\lambda_0(x_0)}{\Phi(\beta_0, x_0)}} \qquad \text{and} \qquad b = \frac{1}{2}\lambda_0'(x_0), \tag{4.3}$$

then (X_n, Y_n) *converge jointly to* $(g_{a,b}, g_{a,b}^0)$, *in* $\mathcal{L} \times \mathcal{L}$, *where the processes* $g_{a,b}$ *and* $g_{a,b}^0$ *have been defined in* (3.11) *and* (3.12).

By making use of results in Lopuhaä and Nane (2013), a completely similar result holds in the nonincreasing setting, which is stated in the Supplement.

Lemma 4. *Let* $x_0 \in (0, \tau_H)$ *fixed and let* $\bar{D}_n$ *be the set on which the unconstrained NPMLE* $\hat{\lambda}_n$, *defined in Lemma* 1, *differs from constrained NPMLE* $\hat{\lambda}_n^0$, *defined in Lemma* 2. *Then, for any* $\varepsilon > 0$, *there exists* $k_\varepsilon > 0$ *such that*

$$\liminf_{n\to\infty} P\left(\bar{D}_n \subset [x_0 - n^{-1/3}k_\varepsilon, x_0 + n^{-1/3}k_\varepsilon]\right) \geq 1 - \varepsilon.$$

Proof. The proof of this fact follows by the reasoning in the proof of Lemma 2.6 in Banerjee (2006), preprint for Banerjee (2007).

Lemma 5. *Consider the processes* X_n *and* Y_n *defined in* (4.2). *Then, for every* $\varepsilon > 0$ *and* $k > 0$, *there exists an* $M > 0$ *such that*

$$\limsup_{n\to\infty} P\Big(\sup_{x\in[-k,k]} |X_n(x)| > M\Big) \leq \varepsilon,$$
$$\limsup_{n\to\infty} P\Big(\sup_{x\in[-k,k]} |Y_n(x)| > M\Big) \leq \varepsilon.$$

Proof. The monotonicity of the processes X_n and Y_n yields that

$$\sup_{x\in[-k,k]} |X_n(x)| = \max\left\{|X_n(-k)|, |X_n(k)|\right\},$$

$$\sup_{x\in[-k,k]} |Y_n(x)| = \max\{|Y_n(-k)|, |Y_n(k)|\}.$$

Assume $|X_n(k)|$ to be the maximum in the above display. Since for fixed k, $X_n(k) \xrightarrow{d} g_{a,b}(k)$, with a and b defined in (4.3), it results that the processes X_n and Y_n in (4.2) are, with high probability, uniformly bounded.

The limiting distribution of the likelihood ratio statistic of a nondecreasing baseline hazard function λ_0 is now supplied.

Theorem 1. *Suppose* (A1) *and* (A2) *hold and let* $x_0 \in (0, \tau_H)$. *Assume that* λ_0 *is nondecreasing on* $[0,\infty)$ *and continuously differentiable in a neighborhood of* x_0, *with* $\lambda_0(x_0) \neq 0$ *and* $\lambda_0'(x_0) > 0$, *and that* $H^{uc}(x)$ *and* $x \to \Phi(\beta_0, x)$, *defined at* (4.1) *are continuously differentiable in a neighborhood of* x_0. *Let* $2\log\xi_n(\theta_0)$ *be the likelihood ratio statistic for testing* $H_0 : \lambda_0(x_0) = \theta_0$, *as at* (3.2). *Then,*

$$2\log\xi_n(\theta_0) \xrightarrow{d} \mathbb{D},$$

where $\mathbb{D} = \int \left[(g_{1,1}(u))^2 - (g_{1,1}^0(u))^2\right] \mathrm{d}u$, *with* $g_{1,1}$ *and* $g_{1,1}^0$ *defined in* (3.11) *and* (3.12).

Proof. The likelihood ratio statistic $2\log\xi_n(\theta_0) = 2L_{\hat\beta_n}(\hat\lambda_n) - 2L_{\hat\beta_n}(\hat\lambda_n^0)$ can be expressed as

$$\begin{aligned}
2\log\xi_n(\theta_0) =& 2\sum_{i=1}^{n-1} \Delta_{(i)} \log\hat\lambda_n(T_{(i)}) - 2\sum_{i=1}^{n-1} \Delta_{(i)} \log\hat\lambda_n^0(T_{(i)}) \\
& - 2\sum_{\substack{i=1\\ i\neq m}}^{n-1} \left[T_{(i+1)} - T_{(i)}\right]\left[\hat\lambda_n(T_{(i)}) - \hat\lambda_n^0(T_{(i)})\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \\
& - 2\left[T_{(m+1)} - x_0\right]\left[\hat\lambda_n(T_{(m)}) - \theta_0\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \\
& - 2\left[x_0 - T_{(m)}\right]\left[\hat\lambda_n(T_{(m)}) - \hat\lambda_n^0(T_{(m)})\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}}.
\end{aligned}$$

For more details, please refer to the Supplement (eq. (S1.1) and (S2.1)). Let

$$S_n = 2\sum_{i=1}^{n-1} \Delta_{(i)} \log\hat\lambda_n(T_{(i)}) - 2\sum_{i=1}^{n-1} \Delta_{(i)} \log\hat\lambda_n^0(T_{(i)}), \tag{4.4}$$

and denote by D_n, the set of indices i on which $\hat\lambda_n(T_{(i)})$ differs from $\hat\lambda_n^0(T_{(i)})$. Hence, expanding both terms of S_n around $\lambda_0(x_0) = \theta_0$, we get

$$S_n = 2\sum_{i\in D_n} \Delta_{(i)} \frac{\hat\lambda_n(T_{(i)}) - \theta_0}{\theta_0} - 2\sum_{i\in D_n} \Delta_{(i)} \frac{\hat\lambda_n^0(T_{(i)}) - \theta_0}{\theta_0}$$

$$-\sum_{i\in D_n}\Delta_{(i)}\frac{\left[\hat\lambda_n(T_{(i)})-\theta_0\right]^2}{\theta_0^2}+\sum_{i\in D_n}\Delta_{(i)}\frac{\left[\hat\lambda_n^0(T_{(i)})-\theta_0\right]^2}{\theta_0^2}+R_n,$$

with

$$\begin{aligned}R_n&=\frac{1}{3}\sum_{i\in D_n}\Delta_{(i)}\frac{\left[\hat\lambda_n(T_{(i)})-\theta_0\right]^3}{\left[\hat\lambda_n^*(T_{(i)})\right]^3}-\frac{1}{3}\sum_{i\in D_n}\Delta_{(i)}\frac{\left[\hat\lambda_n^0(T_{(i)})-\theta_0\right]^3}{\left[\hat\lambda_n^{0*}(T_{(i)})\right]^3}\\&=R_{n,1}-R_{n,2},\end{aligned}$$

where $\hat\lambda_n^*(T_{(i)})$ is a point between $\hat\lambda_n(T_{(i)})$ and θ_0 and $\hat\lambda_n^{0*}(T_{(i)})$ is a point between $\hat\lambda_n^0(T_{(i)})$ and θ_0. We want to show that $R_{n,1}$ and $R_{n,2}$, hence R_n converge to zero, in probability. As for the $R_{n,1}$ term, it can be inferred that

$$|R_{n,1}|\le\frac{1}{3}\int\delta\{u\in\bar D_n\}\frac{\left|n^{1/3}\left(\hat\lambda_n(u)-\theta_0\right)\right|^3}{\left|\hat\lambda_n^*(u)\right|^3}\mathrm{d}P_n(u,\delta,z),$$

where $\bar D_n$ is the time interval on which $\hat\lambda_n$ differs from $\hat\lambda_n^0$. Choose $\varepsilon>0$ and $\gamma>0$ and, for $x_0\in(0,\tau_H)$ fixed and $k_\varepsilon>0$, denote by $I_n=[x_0-n^{-1/3}k_\varepsilon,x_0+n^{-1/3}k_\varepsilon]$. We can write $R_{n,1}=R_{n,1}\{\bar D_n\subset I_n\}+R_{n,1}\{\bar D_n\not\subset I_n\}$. Since, by Lemma 4,

$$\mathbb{P}(|R_{n,1}\{\bar D_n\not\subset I_n\}|>\gamma)\le\mathbb{P}(\bar D_n\not\subset I_n)<\varepsilon,$$

we further focus on bounding $|R_{n,1}\{\bar D_n\subset I_n\}|$. By Lemmas 4 and 5, there exists $k_\varepsilon>0$ such that $\sup_{x\in[-k_\varepsilon,k_\varepsilon]}\left|\hat\lambda_n(x_0+n^{-1/3}x)-\theta_0\right|$ is $\mathcal{O}_p(n^{-1/3})$. Furthermore, since

$$\sup_{x\in[-k_\varepsilon,k_\varepsilon]}\left|\hat\lambda_n^*(x_0+n^{-1/3}x)-\theta_0\right|\le\sup_{x\in[-k_\varepsilon,k_\varepsilon]}\left|\hat\lambda_n(x_0+n^{-1/3}x)-\theta_0\right|,$$

it results that, for $u\in\bar D_n$, $\left|n^{1/3}\left(\hat\lambda_n(u)-\theta_0\right)\right|^3$ is uniformly bounded and $\left|\hat\lambda_n^*(u)\right|^3$ is uniformly bounded away from zero. It then follows that there exists $M>0$ such that

$$\begin{aligned}|R_{n,1}|\le&M\int\delta\{x_0-k_\varepsilon n^{-1/3}\le u\le x_0+k_\varepsilon n^{-1/3}\}\mathrm{d}\left(P_n-P\right)(u,\delta,z)\\&+M\int\delta\{x_0-k_\varepsilon n^{-1/3}\le u\le x_0+k_\varepsilon n^{-1/3}\}\mathrm{d}P(u,\delta,z)+o_p(1).\end{aligned}$$

Chebyshev's inequality provides that the first term on the right-hand side is $\mathcal{O}_p(n^{-2/3})$. As the function H^{uc} defined above (4.1) is assumed to be continuously differentiable in a neighborhood of x_0, the second term on the right-hand side is

$\mathcal{O}_p(n^{-1/3})$. We can conclude that $R_{n,1} = o_p(1)$. Similarly, by using Lemmas 4 and 5, it can be shown that $R_{n,2} = o_p(1)$. Thus $2\log\xi_n(\theta_0) = A_n - B_n + o_p(1)$, where

$$\begin{aligned} A_n = \frac{2}{\theta_0} \sum_{i\in D_n} \Delta_{(i)} \left[\hat\lambda_n(T_{(i)}) - \hat\lambda_n^0(T_{(i)})\right] \\ -2 \sum_{i\in D_n\setminus\{m\}} \left[T_{(i+1)} - T_{(i)}\right] \left[\hat\lambda_n(T_{(i)}) - \hat\lambda_n^0(T_{(i)})\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \\ -2\left[T_{(m+1)} - x_0\right] \left[\hat\lambda_n(T_{(m)}) - \theta_0\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \\ -2\left[x_0 - T_{(m)}\right] \left[\hat\lambda_n(T_{(m)}) - \hat\lambda_n^0(T_{(m)})\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}}, \end{aligned} \tag{4.5}$$

and

$$B_n = \frac{1}{\theta_0^2} \sum_{i\in D_n} \Delta_{(i)} \left\{ \left[\hat\lambda_n(T_{(i)}) - \theta_0\right]^2 - \left[\hat\lambda_n^0(T_{(i)}) - \theta_0\right]^2 \right\}. \tag{4.6}$$

Hence, A_n can be written as $A_n = A_{n1} - A_{n2}$, where

$$\begin{aligned} A_{n1} &= \frac{2}{\theta_0} \sum_{i\in D_n} \left[\hat\lambda_n(T_{(i)}) - \theta_0\right] \left\{ \Delta_{(i)} - \theta_0 \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \right\}, \\ A_{n2} &= \frac{2}{\theta_0} \sum_{i\in D_n\setminus\{m\}} \left[\hat\lambda_n^0(T_{(i)}) - \theta_0\right] \left\{ \Delta_{(i)} - \theta_0 \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \right\} \\ &\quad + \frac{2}{\theta_0} \left[\hat\lambda_n^0(T_{(m)}) - \theta_0\right] \left\{ \Delta_{(m)} - \theta_0 \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \right\}. \end{aligned}$$

For the term A_{n1}, partition the set of indices D_n into s consecutive blocks of indices $B_1, \dots, B_s$, such that $\hat\lambda_n$ is constant on each block. Denote by $\hat v_{nj}$ the unconstrained estimator $\hat\lambda_n(T_{(i)})$, for each $i \in B_j$, with $j = 1, \dots, s$. By (3.6), it follows that

$$\begin{aligned} A_{n1} &= \frac{2}{\theta_0} \sum_{j=1}^{s} \sum_{i\in B_j} (\hat v_{nj} - \theta_0) \left\{ \Delta_{(i)} - \theta_0 \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \right\} \\ &= \frac{2}{\theta_0} \sum_{j=1}^{s} (\hat v_{nj} - \theta_0) \left\{ \sum_{i\in B_j} \Delta_{(i)} - \theta_0 \sum_{i\in B_j} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \right\} \\ &= \frac{2}{\theta_0} \sum_{j=1}^{s} (\hat v_{nj} - \theta_0)^2 \sum_{i\in B_j} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \end{aligned}$$

$$= \frac{2}{\theta_0} n \sum_{i \in D_n} \left[\hat{\lambda}_n(T_{(i)}) - \theta_0\right]^2 \frac{1}{n} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}}.$$

Define

$$\Phi_n(\beta, x) = \int \{u \geq x\} \, \mathrm{e}^{\beta' z} \mathrm{d}P_n(u, \delta, z), \tag{4.7}$$

and note that

$$\int_{[T_{(i)}, T_{(i+1)})} \Phi_n(\hat{\beta}_n, u) \mathrm{d}u = \frac{1}{n} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}},$$

for each $i = 1, \ldots, n-1$. The term A_{n1} can then be written as

$$A_{n1} = \frac{2}{\theta_0} n \int \left\{u \in \bar{D}_n\right\} \left[\hat{\lambda}_n(u) - \theta_0\right]^2 \Phi_n(\hat{\beta}_n, u) \mathrm{d}u,$$

where $\bar{D}_n$ is the interval on which $\hat{\lambda}_n$ and $\hat{\lambda}_n^0$ differ. Similarly, for the term A_{n2}, partition D_n into q consecutive blocks of indices $B_1^0, \ldots, B_q^0$, such that the constrained estimator $\hat{\lambda}_n^0$ is constant on each block. There is one block, say B_r^0, on which the constrained estimator is θ_0, and one block, say B_p^0, that contains m. On all other blocks B_j^0, denote by $\hat{v}_{nj}^0$ the constrained estimator $\hat{\lambda}_n^0(T_{(i)})$, for each $i \in B_j^0$. Then

$$\begin{aligned} A_{n2} =& \frac{2}{\theta_0} \sum_{\substack{j=1 \\ j \neq r,p}}^{q} \sum_{i \in B_j^0} \left(\hat{v}_{nj}^0 - \theta_0\right) \left\{ \Delta_{(i)} - \theta_0 \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\ &+ \frac{2}{\theta_0} \sum_{i \in B_p^0 \setminus \{m\}} \left(\hat{v}_{np}^0 - \theta_0\right) \left\{ \Delta_{(i)} - \theta_0 \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\ &+ \frac{2}{\theta_0} \left(\hat{v}_{np}^0 - \theta_0\right) \left\{ \Delta_{(m)} - \theta_0 \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\ =& \frac{2}{\theta_0} \sum_{\substack{j=1 \\ j \neq r,p}}^{q} \left(\hat{v}_{nj}^0 - \theta_0\right) \left\{ \sum_{i \in B_j^0} \Delta_{(i)} - \theta_0 \sum_{i \in B_j^0} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\} \\ &+ \frac{2}{\theta_0} \left(\hat{v}_{np}^0 - \theta_0\right) \Bigg\{ \sum_{i \in B_p^0} \Delta_{(i)} - \theta_0 \Bigg[\sum_{i \in B_p^0 \setminus \{m\}} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \\ &\qquad + \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \Bigg] \Bigg\}. \end{aligned}$$

By (3.8) and (3.9),

$$A_{n2} = \frac{2}{\theta_0} \sum_{\substack{j=1 \\ j \neq r,p}}^{q} \left(\hat{v}_{nj}^0 - \theta_0\right)^2 \sum_{i \in B_j^0} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}}$$

$$+ \frac{2}{\theta_0} \left(\hat{v}_{np}^0 - \theta_0\right)^2 \left\{ \sum_{i \in B_p^0 \setminus \{m\}} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right.$$

$$\left. + \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}} \right\}$$

$$= \frac{2}{\theta_0} n \sum_{i \in D_n \setminus \{m\}} \left[\hat{\lambda}_n^0(T_{(i)}) - \theta_0\right]^2 \frac{1}{n} \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}}$$

$$+ \frac{2}{\theta_0} n \left[\hat{\lambda}_n^0(T_{(m)}) - \theta_0\right]^2 \frac{1}{n} \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}}.$$

As $\hat{\lambda}_n^0(x) = \hat{\lambda}_n^0(T_{(m)})$ on the interval $[T_{(m)}, x_0)$ and $\hat{\lambda}_n^0(x) = \theta_0$ on the interval $[x_0, T_{(m+1)})$, one has

$$\int_{T_{(m)}}^{T_{(m+1)}} \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \Phi_n(\hat{\beta}_n, u) \mathrm{d}u$$

$$= \int_{T_{(m)}}^{x_0} \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \Phi_n(\hat{\beta}_n, u) \mathrm{d}u + \int_{x_0}^{T_{(m+1)}} \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \Phi_n(\hat{\beta}_n, u) \mathrm{d}u$$

$$= \frac{1}{n} \left[\hat{\lambda}_n^0(T_{(m)}) - \theta_0\right]^2 \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat{\beta}_n' Z_{(l)}}.$$

This leads to

$$A_{n2} = \frac{2}{\theta_0} n \int \left\{u \in \bar{D}_n\right\} \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \Phi_n(\hat{\beta}_n, u) \mathrm{d}u,$$

so that A_n in (4.5) can be written as

$$A_n = \frac{2}{\theta_0} n \int \left\{u \in \bar{D}_n\right\} \left\{ \left[\hat{\lambda}_n(u) - \theta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \right\} \Phi_n(\hat{\beta}_n, u) \mathrm{d}u.$$

In a similar manner, B_n in (4.6) can be expressed as

$$B_n = \frac{1}{\theta_0^2} n \int \left\{u \in \bar{D}_n\right\} \left\{ \left[\hat{\lambda}_n(u) - \theta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \right\} \mathrm{d}V_n(u),$$

by (3.4), and by noting that for every $i = 1, \dots, n-1$,

$$\int_{[T_{(i)},T_{(i+1)})} \mathrm{d}V_n(u) = V_n(T_{(i+1)}) - V_n(T_{(i)}) = \frac{1}{n}\Delta_{(i)}.$$

Concluding,

$$\begin{aligned} 2\log\xi_n(\theta_0) = \frac{2}{\theta_0} n \int \left\{u \in \bar{D}_n\right\} \left\{ \left[\hat{\lambda}_n(u) - \theta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \right\} \Phi_n(\hat{\beta}_n, u)\mathrm{d}u \\ - \frac{1}{\theta_0^2} n \int \left\{u \in \bar{D}_n\right\} \left\{ \left[\hat{\lambda}_n(u) - \theta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \right\} \mathrm{d}V_n(u) + o_p(1). \end{aligned}$$

Let $V(x) = \int \delta\{u < x\}\mathrm{d}P(u, \delta, z)$, and see that, in fact, $V(x) = H^{uc}(x)$, where H^{uc} has been defined above (4.1). Thus,

$$\begin{aligned} 2\log\xi_n(\theta_0) = \frac{2}{\theta_0} n \int \left\{u \in \bar{D}_n\right\} \left\{ \left[\hat{\lambda}_n(u) - \theta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \right\} \Phi(\beta_0, u)\mathrm{d}u \\ - \frac{1}{\theta_0^2} n \int \left\{u \in \bar{D}_n\right\} \left\{ \left[\hat{\lambda}_n(u) - \theta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \right\} \mathrm{d}V(u) \\ + \bar{R}_n + o_p(1), \end{aligned}$$

where $\bar{R}_n = \bar{R}_{n1} - \bar{R}_{n2}$, with

$$\begin{aligned} \bar{R}_{n1} &= \frac{2}{\theta_0} n \int \left\{u \in \bar{D}_n\right\} \left\{ \left[\hat{\lambda}_n(u) - \theta_0\right]^2 \right. \\ &\qquad \left. - \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \right\} \Big(\Phi_n(\hat{\beta}_n, u) - \Phi(\beta_0, u)\Big)\,\mathrm{d}u, \\ \bar{R}_{n2} &= \frac{1}{\theta_0^2} n \int \left\{u \in \bar{D}_n\right\} \left\{ \left[\hat{\lambda}_n(u) - \theta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \right\} \mathrm{d}\left(V_n(u) - V(u)\right). \end{aligned}$$

The aim is to show that $\bar{R}_{n1}$ and $\bar{R}_{n2}$, and thus $\bar{R}_n$ is $o_p(1)$. The term $\bar{R}_{n1}$ can be written as

$$\begin{aligned} \frac{2}{\theta_0} n^{1/3} \int \left\{u \in \bar{D}_n\right\} \left\{ \left[n^{1/3}\left(\hat{\lambda}_n(u) - \theta_0\right)\right]^2 \right. \\ \left. - \left[n^{1/3}\left(\hat{\lambda}_n^0(u) - \theta_0\right)\right]^2 \right\} \Big(\Phi_n(\hat{\beta}_n, u) - \Phi(\beta_0, u)\Big)\,\mathrm{d}u. \end{aligned}$$

Lemma 4 in Lopuhaä and Nane (2013) provides that

$$\sup_{x\in\mathbb{R}} \left|\Phi_n(\hat{\beta}_n, x) - \Phi(\beta_0, x)\right| \to 0,$$

with probability one. From Lemma 5 and since $\int\{u \in \bar{D}_n\}\mathrm{d}u \le 2k_\varepsilon n^{-1/3}$, by Lemma 4 and by using similar arguments as for the term $R_{n,1}$, we can conclude that $\bar{R}_{n1}$ is $o_p(1)$. Analogously,

$$\bar{R}_{n2} = \frac{1}{\theta_0^2} n^{1/3} \int \{u \in \bar{D}_n\} \Big\{ \Big[n^{1/3}\Big(\hat{\lambda}_n(u) - \theta_0\Big)\Big]^2 - \Big[n^{1/3}\Big(\hat{\lambda}_n^0(u) - \theta_0\Big)\Big]^2 \Big\}\delta \mathrm{d}(P_n - P)(u,\delta,z).$$

Once more, by Lemmas 4 and 5, there exists $M_2 > 0$ such that

$$|\bar{R}_{n2}| \le \frac{M_2^2}{\theta_0^2} n^{1/3} \int \delta \left\{u \in \bar{D}_n\right\} \mathrm{d}(P_n - P)(u,\delta,z),$$

with arbitrarily large probability. Chebyshev's inequality along with the same reasoning as for the term $R_{n,1}$ provides that $\bar{R}_{n2} = o_p(1)$. Hence,

$$\begin{aligned} 2\log\xi_n(\theta_0) =& \frac{2}{\theta_0} n \int \left\{u \in \bar{D}_n\right\} \left\{ \left[\hat{\lambda}_n(u) - \theta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \right\} \Phi(\beta_0,u)\mathrm{d}(u) \\ &- \frac{1}{\theta_0^2} n \int \left\{u \in \bar{D}_n\right\} \left\{ \left[\hat{\lambda}_n(u) -!\theta_0\right]^2 - \left[\hat{\lambda}_n^0(u) - \theta_0\right]^2 \right\} \mathrm{d}V(u) + o_p(1). \end{aligned}$$

Consider the change of variable $x = n^{1/3}(u - x_0)$ and let $\widetilde{D}_n = n^{1/3}\left(\bar{D}_n - x_0\right)$. This yields that

$$\begin{aligned} 2\log\xi_n(\theta_0) =& \frac{2}{\theta_0} \int \{x \in \widetilde{D}_n\} \left[X_n^2(x) - Y_n^2(x)\right] \Phi(\beta_0, x_0 + n^{-1/3}x)\mathrm{d}x \\ &- \frac{1}{\theta_0^2} \int \{x \in \widetilde{D}_n\} \left[X_n^2(x) - Y_n^2(x)^2\right] V'(x_0 + n^{-1/3}x)\mathrm{d}x + o_p(1) \\ =& \frac{2}{\theta_0} \Phi(\beta_0, x_0) \int \{x \in \widetilde{D}_n\} \left[X_n^2(x) - Y_n^2(x)\right] \mathrm{d}x \\ &- \frac{1}{\theta_0^2} V'(x_0) \int \{x \in \widetilde{D}_n\} \left[X_n^2(x) - Y_n^2(x)\right] \mathrm{d}x + o_p(1). \end{aligned}$$

As inferred in Lopuhaä and Nane (2013),

$$\lambda_0(x) = \frac{\mathrm{d}V(x)/\mathrm{d}x}{\Phi(\beta_0, x)},$$

which gives that

$$2\log\xi_n(\theta_0) = \frac{1}{\theta_0} \Phi(\beta_0, x_0) \int \{x \in \widetilde{D}_n\} \left[X_n^2(x) - Y_n^2(x)\right] \mathrm{d}x + o_p(1).$$

Thus

$$2\log\xi_n(\theta_0) = \frac{1}{a^2} \int \{x \in \widetilde{D}_n\} \left[X_n^2(x) - Y_n^2(x)\right] \mathrm{d}x + o_p(1),$$

where a has been defined in (4.3). From Lemma 4, for every $\varepsilon > 0$, we can find an interval $[-k_\varepsilon, k_\varepsilon]$ such that $\mathbb{P}(\widetilde{D}_n \subset [-k_\varepsilon, k_\varepsilon]) > 1 - \varepsilon$, for n sufficiently large. In order to prove the theorem, we apply Lemma 4.2 in Prakasa Rao (1969), by taking

$$\begin{aligned} Q_n =& \frac{1}{a^2} \int \{x \in \widetilde{D}_n\} \left[X_n^2(x) - Y_n^2(x)\right] \mathrm{d}x, \\ Q_{n\varepsilon} =& \frac{1}{a^2} \int \{x \in [-k_\varepsilon, k_\varepsilon]\} \left[X_n^2(x) - Y_n^2(x)\right] \mathrm{d}x, \\ Q_\varepsilon =& \frac{1}{a^2} \int \{x \in [-k_\varepsilon, k_\varepsilon]\} \left[(g_{a,b}(x))^2 - \left(g^0_{a,b}(x)\right)^2\right] \mathrm{d}x, \\ Q =& \frac{1}{a^2} \int \{x \in D_{a,b}\} \left[(g_{a,b}(x))^2 - \left(g^0_{a,b}(x)\right)^2\right] \mathrm{d}x, \end{aligned}$$

where $D_{a,b}$ denotes the set on which $g_{a,b}$ and $g^0_{a,b}$ differ. Condition (i) in Lemma 4.2 of Prakasa Rao follows by Lemma 4. In addition, Lemmas 4 and 3 yield condition (ii), since for every $\varepsilon > 0$, we can find $k_\varepsilon > 0$ such that $\mathbb{P}(D_{a,b} \subset [-k_\varepsilon, k_\varepsilon]) > 1 - \varepsilon$. The third condition follows, for every fixed ε, by Lemma 3 and the Continuous Mapping Theorem. Thus (X_n, Y_n) converges to $(g_{a,b}, g^0_{a,b})$ as a process in $\mathcal{L} \times \mathcal{L}$ and $(f, g) \mapsto \int \{x \in [-c, c]\}(f^2(x) - g^2(x))\mathrm{d}x$ is a continuous function defined on $\mathcal{L} \times \mathcal{L}$ with values in $\mathbb{R}$. Conclusively,

$$\begin{aligned} \frac{1}{a^2} \int \left[X_n^2(x) - Y_n^2(x)\right] \{x \in \widetilde{D}_n\} \mathrm{d}x &\xrightarrow{d} \frac{1}{a^2} \int \left[(g_{a,b}(x))^2 - \left(g^0_{a,b}(x)\right)^2\right] \{x \in D_{a,b}\}\, \mathrm{d}x \\ &\overset{d}{=} \int \left[(g_{1,1}(x))^2 - \left(g^0_{1,1}(x)\right)^2\right] \{x \in D_{1,1}\}\, \mathrm{d}x \end{aligned}$$

by the Continuous Mapping Theorem and by Brownian scaling, as derived in Banerjee and Wellner (2001). This completes the proof.

The asymptotic distribution of the likelihood ratio statistic in the nonincreasing baseline hazard setting can be derived completely analogous and is stated in the supplement.

Remark 2. The same limiting distribution $\mathbb{D}$ is obtained for the loglikleihood ratio statistic in the absence of covariates in Banerjee (2008), as well as in other censoring frameworks, as derived by Banerjee and Wellner (2001). In fact, it has been shown in Banerjee (2007) that the same holds true for a wide class of monotone response models. This distribution differs from the usual χ^2_1 distribution obtained in the regular parametric setting. It is noteworthy that $\mathbb{D}$ does not depend on any of the parameters of the underlying model, and this property turns out to be particularly useful in constructing confidence intervals for the parameters of interest.

5. Pointwise Confidence Intervals via Simulations

Once having derived the asymptotic distribution of the likelihood ratio statistic, the practical application at hand is to construct, for fixed $x_0 \in (0, \tau_H)$, pointwise confidence intervals.

To compute the likelihood ratio statistic for a nondecreasing λ_0, suppose that $\lambda_0(x_0) = \theta$, for fixed $\theta \in (0, \infty)$ and let m such that $T_{(m)} < x_0 < T_{(m+1)}$. Then,

$$\begin{aligned}
2\log \xi_n(\theta) = {} & 2\sum_{i=1}^{n-1}\left\{\Delta_{(i)} \log \hat\lambda_n(T_{(i)}) - \hat\lambda_n(T_{(i)}) \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}}\right\} \\
& -2\sum_{i=1}^{m-1}\left\{\Delta_{(i)} \log \hat\lambda_n^0(T_{(i)}) - \hat\lambda_n^0(T_{(i)}) \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}}\right\} \\
& -2\Big\{\Delta_{(m)} \log \hat\lambda_n^0(T_{(m)}) - \hat\lambda_n^0(T_{(m)}) \left[x_0 - T_{(m)}\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \\
& -\theta\left[T_{(m+1)} - x_0\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}}\Big\} \\
& -2\sum_{i=m+1}^{n-1}\left\{\Delta_{(i)} \log \hat\lambda_n^0(T_{(i)}) - \hat\lambda_n^0(T_{(i)}) \left[T_{(i+1)} - T_{(i)}\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}}\right\}.
\end{aligned}$$

Write

$$\begin{aligned}
2\log \xi_n(\theta) = {} & 2\sum_{i=1}^{n-1} \Delta_{(i)}[\log \hat\lambda_n(T_{(i)}) - \log \hat\lambda_n^0(T_{(i)})] \\
& -2\sum_{\substack{i=1\\ i\neq m}}^{n-1} \left[T_{(i+1)} - T_{(i)}\right]\left[\hat\lambda_n(T_{(i)}) - \hat\lambda_n^0(T_{(i)})\right] \sum_{l=i+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \\
& -2\left[T_{(m+1)} - x_0\right]\left[\hat\lambda_n(T_{(m)}) - \theta\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}} \\
& -2\left[x_0 - T_{(m)}\right]\left[\hat\lambda_n(T_{(m)}) - \hat\lambda_n^0(T_{(m)})\right] \sum_{l=m+1}^{n} \mathrm{e}^{\hat\beta_n' Z_{(l)}}.
\end{aligned}$$

The characterization of the estimators $\hat\lambda_n$ and $\hat\lambda_n^0$ in Lemmas 1 and 2 is then sufficient to compute the statistic.

Let $2\log \xi_n(\theta)$ denote the likelihood ratio statistic defined in (3.2), for testing $H_0 : \lambda_0(x_0) = \theta$ versus $H_1 : \lambda_0(x_0) \neq \theta$. A $1 - \alpha$ confidence interval is obtained by inverting $2\log \xi_n(\theta)$ for different values of θ:

$$C_{n,\alpha}^1 \equiv \{\theta : 2\log \xi_n(\theta) \le q(\mathbb{D}, 1-\alpha)\},$$

where $q(\mathbb{D}, 1-\alpha)$ is the $(1-\alpha)^{th}$ quantile of the distribution $\mathbb{D}$. Quantiles of $\mathbb{D}$, based on discrete approximations of Brownian motion, are provided in Banerjee and Wellner (2005), and we use $q(\mathbb{D}, 0.95) = 2.286922$. The parameter θ is chosen to take values on a fine grid between 0 and 6. It can be shown immediately that, for large enough n, the coverage probability of $C^1_{n,\alpha}$ is approximately $1-\alpha$.

Furthermore, given the covariate vectors $z_0, z \in \mathbb{R}^p$, we can write

$$\lambda(x|z) = \lambda_0(x)\mathrm{e}^{\beta_0' z} = \lambda_0(x)\mathrm{e}^{\beta_0' z_0}\mathrm{e}^{\beta_0'(z-z_0)}.$$

If we consider now the covariate vector $\tilde{Z} = Z - z_0$, then, according to the Cox model, the hazard function of an individual with covariate vector $\tilde{z} \in \mathbb{R}^p$ can be written as

$$\lambda(x|\tilde{z}) = \tilde{\lambda}_0(x)\mathrm{e}^{\beta_0'\tilde{z}},$$

where $\tilde{\lambda}_0$ is the baseline function that corresponds to $\tilde{z} = 0$. The baseline hazard function $\tilde{\lambda}_0$ is, in fact,

$$\tilde{\lambda}_0(x) = \lambda(x|\tilde{z} = 0) = \lambda(x|z - z_0 = 0) = \lambda(x|z = z_0) = \lambda_0(x)\mathrm{e}^{\beta_0' z_0},$$

the hazard function of an individual with covariate vector z_0. Hence testing whether $\tilde{\lambda}_0$ has a particular value θ_0 at a fixed point x_0 is equivalent to testing that $\lambda(x_0|z_0) = \theta_0$. Therefore, the likelihood ratio method presented in this paper can also be used for constructing confidence intervals for the hazard function, given a covariate vector z_0 and a fixed point x_0.

Pointwise confidence intervals for $\lambda_0(x_0)$ can also be constructed based on the asymptotic distribution of the NPMLE $\hat{\lambda}_n$. According to Theorem 2 in Lopuhaä and Nane (2013), for fixed x_0,

$$n^{1/3}\left(\hat{\lambda}_n(x_0) - \lambda_0(x_0)\right) \xrightarrow{d} \left(\frac{4\lambda_0(x_0)\lambda_0'(x_0)}{\Phi(\beta_0, x_0)}\right)^{1/3} \operatorname*{argmin}_{x\in\mathbb{R}}\{\mathbb{W}(t) + t^2\} \equiv C(x_0)\mathbb{Z},$$

where $\mathbb{W}$ is standard two-sided Brownian motion starting from zero, and the constant $C(x_0)$ depends on x_0 and on the underlying parameters. An estimator $\hat{C}_n(x_0)$ of $C(x_0)$ will then yield an $1-\alpha$ confidence interval for $\lambda_0(x_0)$:

$$C^2_{n,\alpha} \equiv \left[\hat{\lambda}_n(x_0) - n^{-1/3}\hat{C}_n(x_0)q(\mathbb{Z}, 1-\frac{\alpha}{2}), \hat{\lambda}_n(x_0) + n^{-1/3}\hat{C}_n(x_0)q(\mathbb{Z}, 1-\frac{\alpha}{2})\right],$$

where $q(\mathbb{Z}, 1-\alpha/2)$ is the $(1-\alpha/2)^{th}$ quantile of the distribution $\mathbb{Z}$. These quantiles have been computed in Groeneboom and Wellner (2001), and we use $q(\mathbb{Z}, 0.975) = 0.998181$.

For simulation purposes, we propose

$$\hat{C}_n(x_0) = \left(\frac{4\hat{\lambda}_n(x_0)\hat{\lambda}_n'(x_0)}{\Phi_n(\hat{\beta}_n, x_0)}\right)^{1/3},$$

where $\Phi_n(\beta, x)$ has been defined in (4.7), and $\hat{\beta}_n$ is the maximum partial likelihood estimator. Lemma 4 in Lopuhaä and Nane (2013) ensures that $\Phi_n(\hat{\beta}_n, \cdot)$ is a strong uniform consistent estimator of $\Phi(\beta_0, \cdot)$. As an estimate for $\lambda_0'(x_0)$, we chose the numerical derivative of $\hat{\lambda}_n$ on the interval that contains x_0, the slope of the segment $[\hat{\lambda}_n(T_{(m)}), \hat{\lambda}_n(T_{(m+1)})]$.

For the performance analysis, we constructed and compared, from simulated data, the confidence intervals $C^1_{n,\alpha}$ and $C^2_{n,\alpha}$, for $\alpha = 0.05$ and various n. As the baseline hazard function was assumed to be nondecreasing, we chose a Weibull baseline distribution function for the event times, with shape parameter 2 and scale parameter 1. For simplicity, we took the covariate as single-valued and uniformly $(0,1)$ distribute, with $\beta_0 = 0.5$. Given the covariate, the censoring times were assumed to be uniformly $(0,1)$ distributed. We chose $x_0 = \sqrt{\log 2}$, the median of the baseline distribution of the event times. For each chosen sample size, we generated 1,000 replicates and computed the empirical coverage and the average length of the corresponding confidence intervals.

Since we were simulating from a Weibull distribution with shape parameter 2 and scale parameter 1, and hence the true baseline hazard function λ_0 and its derivative were known, as well as the true underlying regression coefficient, we could also consider a confidence interval $\bar{C}^2_{n,\alpha}$ given by

$$\bar{C}^2_{n,\alpha} \equiv \left[\hat{\lambda}_n(x_0) - n^{-1/3} C_0(x_0) q(\mathbb{Z}, 1 - \frac{\alpha}{2}), \hat{\lambda}_n(x_0) + n^{-1/3} C_0(x_0) q(\mathbb{Z}, 1 - \frac{\alpha}{2})\right],$$

where C_0 is a deterministic function given by

$$C_0(x_0) = \left(\frac{4\lambda_0(x_0)\lambda_0'(x_0)}{\Phi(\beta_0, x_0)}\right)^{1/3}.$$

Table 1 shows the performance, for various sample sizes, of the confidence interval $C^1_{n,0.05}$ based on the likelihood ratio method (LR), the confidence interval $C^2_{n,0.05}$, based on the asymptotic distribution (AD) of the scaled differences between the NPMLE $\hat{\lambda}_n$ and the true baseline hazard at a fixed point, as well as the confidence interval $\bar{C}^2_{n,0.05}$ based on the Weibull distribution (TD).

For each sample size, the likelihood ratio method gave, on average, shorter pointwise confidence intervals in comparison with the confidence intervals based on the asymptotic distribution of the NPMLE estimator $\hat{\lambda}_n$. Moreover, the confidence intervals based on the likelihood ratio exhibit comparable coverage probabilities with the confidence intervals $C^2_{n,0.05}$, based on the asymptotic distribution. As expected, the highest coverage rate was attained by the confidence intervals $\bar{C}^2_{n,0.05}$, for all sample sizes. Furthermore, they gave confidence intervals with the shortest length, on average. For the largest sample sizes, of 5,000 observations, the likelihood ratio method yielded comparable confidence intervals with

Table 1. Simulation results for constructing 95% pointwise confidence intervals using the likelihood ratio method $C^1_{n,0.05}$ (LR) or the asymptotic distribution $C^2_{n,0.05}$ (AD) and $\bar{C}^2_{n,0.05}$ (TD), in terms of average length (AL) and empirical coverage (CP).

	LR		AD		TD	
n	AL	CP	AL	CP	AL	CP
50	4.275	0.917	5.203	0.932	1.506	0.964
100	3.837	0.923	4.838	0.941	1.317	0.953
200	3.009	0.931	4.605	0.947	1.247	0.947
500	2.734	0.947	3.372	0.948	0.961	0.964
1000	1.454	0.942	2.259	0.940	0.713	0.957
5000	0.879	0.945	1.768	0.952	0.546	0.953

$\bar{C}^2_{n,0.05}$ in terms of interval length, on average, as well as for empirical coverage. The simulations are readily extendable to more than one covariate, and similar findings are obtained.

Acknowledgements

The author thanks Mouli Banerjee for the idea of this paper, and is grateful to her PhD supervisor, Rik Lopuhaä, for his comments and guidance. The author thanks the associate editor and the referee for their comments and suggestions, which have led to an improved version of the paper.

References

Banerjee, M. (2006). Likelihood based inference for monotone response models. Preprint. Available at `www.stat.lsa.umich.edu/~moulib/wilks3.pdf`.

Banerjee, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35**, 931-956.

Banerjee, M. (2008). Estimating monotone, unimodal and U-shaped failure rates using asymptotic pivots. *Statist. Sinica* **18**, 467-492.

Banerjee, M. and Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699-1731.

Banerjee, M. and Wellner, J. A. (2005). Score statistics for current status data: comparisons with likelihood ratio and Wald statistics. *Int. J. Biostatistics* **1** Art. 3, 29 pp. (electronic).

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

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269-276.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72**, 557-565.

van Geloven, N., Martin, I. Damman, P. de Winter, R. J., Thijssen, J. G. and Lopuhaä, H. P. (2013). Estimation of a decreasing hazard of patients with myocardial infarction. *Statist. Medicine* **32**, 1223-1238.

Groeneboom, P. and Wellner, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10**, 388-400.

Huang, J. and Wellner, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Statist.* **22**, 3-33.

Lopuhaä, H. P. and Nane, G. F. (2013). Shape constrained nonparametric estimators of the baseline distribution in Cox proportional hazards model. *Scand. J. Statist.* **40**(3), 619-646.

Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika* **64**, 441-448.

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31**, 23-36.

Slud, E. V. (1982). Consistency and efficiency of inferences with the partial likelihood *Biometrika* **69**, 547-552.

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

Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands.

E-mail: g.f.nane@tudelft.nl

(Received March 2013; accepted July 2014)