Statistical Methods in Medical Research
XX(X):1–18

DOI: 10.1177/ToBeAssigned
www.sagepub.com/

Survival analysis with a random change-point

Chun Yin Lee[1] and Kin Yau Wong[1,2]

Abstract

Contemporary works in change-point survival models mainly focus on an unknown universal change-point shared by the whole study population. However, in some situations, the change-point is plausibly individual-specific, such as when it corresponds to the telomere length or menopausal age. Also, maximum-likelihood-based inference for the fixed change-point parameter is notoriously complicated. The asymptotic distribution of the maximum likelihood estimator is non-standard, and computationally intensive bootstrap techniques are commonly used to retrieve its sampling distribution. This paper is motivated by a breast cancer study, where the disease-free survival time of the patients is postulated to be regulated by the menopausal age, which is unobserved. As menopausal age varies across patients, a fixed change-point survival model may be inadequate. Therefore, we propose a novel proportional hazards model with a random change-point. We develop a nonparametric maximum likelihood estimation approach and devise a stable EM algorithm to compute the estimators. Because the model is regular, we employ conventional likelihood theory for inference based on the asymptotic normality of the Euclidean parameter estimates, and the variance of the asymptotic distribution can be consistently estimated by a profile-likelihood approach. A simulation study demonstrates the satisfactory finite-sample performance of the proposed methods, which yield small bias and proper coverage probabilities. The methods are applied to the motivating breast cancer study.

Keywords

Breast cancer, EM algorithm, Profile likelihood, Proportional hazards model, Right-censored data.

[1]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong.[2]Hong Kong Polytechnic University Shenzhen Research Institute

Corresponding author:
K.Y. Wong, Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong.
Email: kin-yau.wong@polyu.edu.hk

Prepared using sagej.cls *[Version: 2015/06/09 v1.01]*

1 Introduction

Change-point models are widely applicable in medical studies, where the effect of a covariate on the response variable is not linear but is regulated by the same or another covariate, often called a change-point variable. One motivation for considering a change-point model is to develop a classification method for assigning patients into sub-groups that possess different levels of disease risk according to the change-point variable. This can improve risk prediction and identification of individuals who require intensive treatments. For instance, patients with primary biliary cirrhosis can be classified as early-phase or late-phase according to a change-point parameter (also referred to as a threshold parameter), where the effect of bilirubin level when the level is above the parameter differs from that when the level is below the parameter[1]. Also, the risk of diabetes can be classified as high or low according to a change-point parameter of leukocyte telomere length[2]. For both diseases, the disease risk of a patient changes substantially when a covariate value exceeds a change-point parameter. The change-point parameter is typically unknown and needs to be estimated from data.

The Cox proportional hazards (PH) model[3] has been used extensively to analyze the association between a time-to-event outcome and explanatory variables, with the constant hazard ratio assumption. Generalizations of the Cox model to capture change-point effects have been studied. There are mainly three types of change-point survival models in the literature, namely (i) change-point in time models, where the effect of a time-independent covariate changes when the time since origin exceeds a certain threshold[4–7], (ii) change-points in covariate models with smooth changes, where the effect of a covariate changes when the value of the covariate exceeds one or more thresholds[8–10], and (iii) change-points in covariate models with one or more jumps, where the effects of a set of covariates change when the value of another covariate exceeds the threshold value(s)[11–15]. Model (i) typically aims to accommodate the time lag or fading out of the treatment effects in clinical trials. Model (ii) aims to capture the nonlinear effects of a covariate, serving as an alternative to the polynomial splines, and the inference on the change-point(s) is also of interest. Model (iii) pertains to the regulation of the effects of some covariates based on whether the change-point variable exceeds a threshold value(s). In most contemporary works, the change-point parameter, namely η^*, is an unknown universal quantity shared by all individuals. Under this model, whether the baseline risk of a patient has shifted is completely determined by the observed data, and this facilitates the stratification of patients into different risk levels according to a global reference guideline.

Personalized medicine has garnered considerable attention recently, and we may wish to perform a risk assessment using an individual-specific change-point. Notably, we may be interested in evaluating how likely the baseline risk of a particular patient has shifted. Lam et al.[16] studied the disease-free survival of breast cancer patients based on a piecewise linear covariate effect model and suggested that the change-point of age effects on the survival pattern may be related to the menopausal status of the patients. Along this line, Lee et al.[17] proposed the maximal score and Wald tests for the presence of a change-point in the (disease-free) survival rates of breast cancer patients using the

following hazard function

$$\lambda^*(t)\exp\left\{\boldsymbol{\gamma}^{*T}\boldsymbol{X}+\beta^* Z+\alpha_1^*\mathrm{I}(Z\geq\eta^*)+\alpha_2^*\mathrm{I}(Z\geq\eta^*)(Z-\eta^*)\right\} \quad (1)$$

under slightly different notations, where λ^* is an arbitrary baseline hazard function, η^* is an unknown fixed change-point parameter, and $(\boldsymbol{\gamma}^*, \beta^*, \alpha_1^*, \alpha_2^*)$ are regression coefficients. With Z being the age at diagnosis, they rejected the null hypothesis that $\alpha_1^* = \alpha_2^* = 0$ for the German Breast Cancer dataset[18;19]. A threshold age of $\eta^* = 50$ was detected, plausibly associated with age-at-menopause. They observed that patients diagnosed at an age before the threshold have a better prognosis, whereas those diagnosed at an age after the threshold have poorer survival rates. Intuitively, if the change-point parameter is truly the age at menopause, it should not be treated as fixed, since the age at menopause is highly dependent on family history and genetic predisposition. By contrast, it is more plausible to assume that the menopausal age is random in nature. Random change-point models have been tailored to a diverse field of studies, besides breast cancer study. Under the Bayesian paradigm, Lange et al.[20] modeled the disease progression of the HIV-infected patients by the mean trajectory of the CD4 T-Cell numbers, where the response variable is expected to change smoothly at a random change-point in time. Slate and Turnbull[21] modeled the disease progression of prostate cancer patients by their trajectories of prostate-specific antigen with a random change-point for the onset time. Based on the maximum likelihood approach, Jacqmin-Gadda et al.[22] studied the joint model for repeatedly measured cognitive test scores and the risk of dementia, where a random change-point represents the individual's age at accelerated cognitive decline. However, the contemporary work mainly focused on modeling a longitudinal response variable, instead of a survival outcome variable.

There is a pressing need to develop effective statistical methods for the analysis of survival outcomes with a random change-point. To this end, we propose a novel model formulated based on (1) with η^* treated as random, and provide an effective and reliable inference procedure for parameter estimation. The rest of this article is structured as follows. In Section 2, we first outline the model and the likelihood function and propose a nonparametric maximum likelihood estimator (NPMLE). In Section 3, we devise an EM algorithm for computation of the NPMLE, where a quadrature rule based on the truncated normal distribution is adopted for the E-step, and propose a profile likelihood approach for variance estimation for the Euclidean parameter estimates. In Section 4, we conduct a simulation study to evaluate the finite-sample performance of the proposed methods, and investigate the scenario when the fixed change-point model in (1) is misfitted to the data, which are generated with a random change-point. In Section 5, we demonstrate the practical utility of the proposed methods using the motivating breast cancer dataset. Concluding remarks are given in Section 6.

2 Model and estimation

Consider a study that comprises n independent subjects. For the ith subject ($i = 1, \ldots, n$), let T_i be the failure time, $\boldsymbol{X}_i$ be a vector of covariates, Z_i be a continuous change-point variable, and η_i be an individual-specific random change-point. Based on

the formulation in (1), we assume that the hazard function of T_i conditional on the covariates and η_i is

$$\lambda(t|\boldsymbol{X}_i, Z_i, \eta_i) = \lambda(t) \exp\left\{\boldsymbol{\gamma}^T \boldsymbol{X}_i + \beta Z_i + \alpha_1 \mathrm{I}(Z_i \geq \eta_i) + \alpha_2 \mathrm{I}(Z_i \geq \eta_i)(Z - \eta_i)\right\} \quad (2)$$

where λ is an unspecified baseline hazard function, and $(\boldsymbol{\gamma}, \beta, \alpha_1, \alpha_2)$ are regression parameters. We assume that $\eta_i \sim \mathrm{N}(\mu, \sigma^2)$. Model (2) implies that as Z_i increases beyond the threshold η_i, the effect of a unit change in Z_i changes from β to $(\beta + \alpha_2)$, and also the hazard changes by a multiplicative factor of e^{α_1}. This is analogous to a piecewise linear model with two partitions but with a vertical jump at a random point η_i.

Suppose that the subjects are followed until they experience the event or are right-censored. For the ith subject, we use C_i to denote the censoring time, $\Delta_i \equiv \mathrm{I}(T_i \leq C_i)$ to denote the event indicator, and $Y_i \equiv \min(T_i, C_i)$ to denote the observed failure or censoring time. Then, the random sample consists of $\mathcal{O} = \{Y_i, \Delta_i, \boldsymbol{X}_i, Z_i\}_{i=1,\ldots,n}$. We assume that T and C are independent given $(\boldsymbol{X}, Z)$. In particular, this requires that C is independent of η given the covariates. Let $\boldsymbol{\theta} = (\Lambda, \boldsymbol{\xi})$, where $\Lambda(t) \equiv \int_0^t \lambda(u)\,\mathrm{d}u$ is the baseline cumulative hazard function, $\boldsymbol{\xi} \equiv (\boldsymbol{\zeta}, \mu, \sigma)$ is the collection of all Euclidean parameters, and $\boldsymbol{\zeta} \equiv (\boldsymbol{\gamma}, \beta, \alpha_1, \alpha_2)$ is the set of regression parameters in the survival model. Let $h_i(\boldsymbol{\zeta}; \eta) = \boldsymbol{\gamma}^T \boldsymbol{X}_i + \beta Z_i + \alpha_1 \mathrm{I}(Z_i \geq \eta) + \alpha_2 \mathrm{I}(Z_i \geq \eta)(Z_i - \eta)$. The observed likelihood is

$$\begin{aligned} L^{obs}(\boldsymbol{\theta}|\mathcal{O}) \quad = \quad & \prod_{i=1}^{n} \int \Big[\lambda(Y_i) \exp\left\{h_i(\boldsymbol{\zeta}; \eta_i)\right\}\Big]^{\Delta_i} \\ & \times \exp\Big[-\Lambda(Y_i) \exp\left\{h_i(\boldsymbol{\zeta}; \eta_i)\right\}\Big] \phi(\eta_i; \mu, \sigma)\,\mathrm{d}\eta_i \end{aligned} \quad (3)$$

where $\phi(.; \mu, \sigma)$ is the normal density with mean μ and variance σ^2. Note that the maximizer of (3) does not exist due to the infinite-dimensional parameter Λ. We adopt a nonparametric maximum likelihood estimation (NPMLE) approach, where we set Λ as a step function that jumps only at the observed failure times and replaces $\lambda(Y_i)$ by the jump size of Λ at Y_i. Let $(\widehat{\Lambda}, \widehat{\boldsymbol{\xi}})$ be the maximizer of the resulting nonparametric likelihood function.

Under some mild regularity conditions, the proposed model is identifiable; details on model identifiability are given in the Appendix. In addition, the proposed model falls under the general framework described in Zeng and Lin[23]. The estimators are strongly consistent and asymptotically Gaussian, whereas the estimators of the Euclidean parameters achieve the semiparametric efficiency bound. Because the derivations of the theoretical properties of the estimators are largely similar to existing works[24–26], we omit them here. There is an interesting contrast with the fixed change-point model. The conventional fixed change-point model is irregular, and sophisticated techniques (such as the m-out-of-n bootstrap) have to be employed for inference. When we allow the change-point to be random, the model becomes regular, the maximum likelihood estimators are $\sqrt{n}$-consistent, and standard likelihood approaches that involve inversion of the information are applicable.

3 Computation of the NPMLE

Since η can be thought of as a missing variable, we propose to use the EM algorithm[27] to compute the NPMLE. The complete-data log-likelihood based on $\{Y_i, \Delta_i, \boldsymbol{X}_i, Z_i, \eta_i\}_{i=1,\ldots,n}$ is

$$\sum_{i=1}^{n} \Delta_i \log \Lambda\{Y_i\} + \Delta_i h_i(\boldsymbol{\zeta}; \eta_i) - \Lambda(Y_i) \exp\{h_i(\boldsymbol{\zeta}; \eta_i)\} + \log \phi(\eta_i; \mu, \sigma),$$

where $\Lambda\{t\}$ is the jump size of Λ at t. In the E-step of the dth iteration ($d = 1, 2, \ldots$), we need to evaluate terms of the form $\widehat{\mathrm{E}}[\eta_i^a \mathrm{I}(Z_i \geq \eta_i)^b \exp\{c h_i(\boldsymbol{\zeta}; \eta_i)\}]$ for $a = 0, 1, 2$, $b = 0, 1$ and $c = 0, 1$, where $\widehat{\mathrm{E}}(\cdot)$ denotes the expectation taken with respect to the conditional density of η_i given $\mathcal{O}_i$ evaluated at the current estimate, namely $\boldsymbol{\theta}^{(d)}$. Specifically, for any function g of η_i, we have

$$\widehat{\mathrm{E}}\{g(\eta_i)\} = \frac{\int g(\eta_i) f_i^*(\eta_i, \mathcal{O}_i; \boldsymbol{\theta}^{(d)}) \mathrm{d}\eta_i}{\int f_i^*(\eta_i, \mathcal{O}_i; \boldsymbol{\theta}^{(d)}) \mathrm{d}\eta_i},$$

where

$$f_i^*(\eta, \mathcal{O}_i; \boldsymbol{\theta}) = \exp\Big\{\Delta_i h_i(\boldsymbol{\zeta}; \eta)\Big\} \exp\Big[- \Lambda(Y_i) \exp\{h_i(\boldsymbol{\zeta}; \eta)\} \Big] \phi(\eta; \mu, \sigma).$$

The above expectations do not have closed-form expressions but can be approximated by numerical integration. Since the integrand is not smooth at $\eta_i = Z_i$, the usual Gauss–Hermite quadrature rule may perform poorly even with a large number of knots, and the adaptive Gaussian quadrature rule[28] is not applicable since the functions inside the integrals are not unimodal. To overcome these difficulties, we perform integration over $\eta_i < Z_i$ and $\eta_i > Z_i$ separately. Note that on each partition, the integration can be thought of as an expectation under a truncated normal distribution, and the expectation can be approximated using the quadrature rule of Golub and Welsch[29]; see also Burkardt[30]. In particular, for $X \sim \mathrm{TN}(\mu, \sigma^2, a, b)$ and a function g, where $\mathrm{TN}(\mu, \sigma^2, a, b)$ denotes the $\mathrm{N}(\mu, \sigma^2)$ distribution truncated to be within (a, b) with $-\infty \leq a < b \leq \infty$, we approximate $\mathrm{E}\{g(X)\}$ using the following algorithm:

1. Compute $\varsigma_k \equiv \mathrm{E}(X^k)$ for $k = 0, \ldots, 2m$, where $X \sim \mathrm{TN}(\mu, \sigma^2, a, b)$, and m is a positive integer.

2. Construct the $(m+1)$ by $(m+1)$ moment matrix $\boldsymbol{M} = (\varsigma_{j+k-2})_{j,k=1,\ldots,m+1}$, and obtain the upper triangular Cholesky factor $\boldsymbol{R}$ of $\boldsymbol{M}$.

3. Using $\boldsymbol{R}$, construct the m by m symmetric tridiagonal matrix $\boldsymbol{J}$ as defined in Formulae 2.2 and 4.3 in Golub and Welsch[29].

4. Let $x_1, \ldots, x_m$ be the eigenvalues of $\boldsymbol{J}$ and $w_1, \ldots, w_m$ be the squares of the first entries of the corresponding eigenvectors. We approximate $\mathrm{E}\{g(X)\}$ by $\sum_{j=1}^{m} w_j g(x_j)$.

In general, a larger m results in higher approximation precision. Although the moment matrix $\boldsymbol{M}$ is positive semidefinite by construction, numerical instability may arise when the truncation is far away from the center of the normal distribution, resulting in a non-positive semidefinite $\boldsymbol{M}$. In the implementation, if the numerical evaluation of the moment matrix is not positive semidefinite with the specified m, then we set the number of nodes to be the largest integer smaller than m that yields a positive semidefinite $\boldsymbol{M}$. The proposed random change-point model includes the fixed change-point model as the limiting case with $\sigma = 0$. Computationally, we cannot allow the estimator of σ to approach the boundary value of 0 (we set a small lower bound of 0.05 instead), as the numerical integration in the E-step is highly unstable for a very small value of σ.

In the M-step, we update $(\Lambda, \boldsymbol{\zeta})$ by maximizing the expected complete-data log-likelihood. In particular, $\boldsymbol{\zeta}^{(d+1)}$ can be computed by maximizing the following partial-likelihood-type function:

$$\ell(\boldsymbol{\zeta}) = \sum_{i=1}^{n} \Delta_i \left[\widehat{\mathrm{E}}\{h_i(\boldsymbol{\zeta}; \eta_i)\} - \log \left\{ \sum_{j=1}^{n} \mathrm{I}(Y_j \geq Y_i) \widehat{\mathrm{E}} \left[\exp\{h_j(\boldsymbol{\zeta}; \eta_j)\} \right] \right\} \right].$$

To improve computational efficiency, we set $\boldsymbol{\zeta}^{(d+1)}$ to be the output from a one-step Newton Raphson algorithm on $\ell(\boldsymbol{\zeta})$. Then, we update Λ by

$$\Lambda^{(d+1)}(t) = \sum_{i=1}^{n} \frac{\Delta_i \mathrm{I}(Y_i \leq t)}{\sum_{j=1}^{n} \mathrm{I}(Y_j \geq Y_i) \widehat{\mathrm{E}} \left[\exp\{h_j(\boldsymbol{\zeta}^{(d+1)}; \eta_j)\} \right]}.$$

Lastly, we update μ and σ^2 by maximizing $\sum_{i=1}^{n} \widehat{\mathrm{E}} \left\{ \log \phi(\eta_i; \mu, \sigma) \right\}$, with closed-form solutions $\mu^{(d+1)} = n^{-1} \sum_{i=1}^{n} \widehat{\mathrm{E}}(\eta_i)$ and $\sigma^{2(d+1)} = n^{-1} \sum_{i=1}^{n} \widehat{\mathrm{E}}(\eta_i^2) - \left\{ n^{-1} \sum_{i=1}^{n} \widehat{\mathrm{E}}(\eta_i) \right\}^2$. The EM algorithm is computationally efficient because (i) in the E-step, only a one-dimensional integration is involved; and (ii) in the M-step, the high-dimensional parameter Λ is updated with an explicit formulation that does not require the inversion of a large matrix, whereas the low-dimensional parameter $\boldsymbol{\zeta}$ is updated via the Newton–Raphson algorithm.

The conventional EM algorithm iterates between the E- and M-steps, but it may converge very slowly in our application and tends to stop early before the observed likelihood function reaches the maximum, as the difference in consecutive parameter vector update may be small in the intermediate steps. To enhance the speed of convergence, we adopt the accelerated EM algorithm proposed by Varadhan and Roland[31]. Let $\boldsymbol{s}(\boldsymbol{\theta})$ be the updated parameter vector obtained from a regular EM step with the initial parameter vector value set at $\boldsymbol{\theta}$. Given that the current estimate is $\boldsymbol{\theta}^{(k)}$, $k = 1, 2, \ldots$, a step of the accelerated EM algorithm comprises the following four steps:

1. Compute $\boldsymbol{\theta}_1 = \boldsymbol{s}(\boldsymbol{\theta}^{(k)})$.
2. Compute $\boldsymbol{\theta}_2 = \boldsymbol{s}(\boldsymbol{\theta}_1)$.

3. Compute $\boldsymbol{r} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}^{(k)}$, $\boldsymbol{v} = \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 - \boldsymbol{r}$, and $a = -\|\boldsymbol{r}\|_2/\|\boldsymbol{v}\|_2$.

4. Update the vector of parameter estimates by $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{s}(\boldsymbol{\theta}^{(k)} - 2a\boldsymbol{r} + a^2\boldsymbol{v})$.

We terminate the accelerated EM algorithm when the maximum absolute difference between two consecutive estimates of $\boldsymbol{\theta}$ is less than a small threshold.

The likelihood function involves the high-dimensional parameter Λ, and thus the computation of the inverse of the observed Fisher information matrix can be computationally intensive. Alternatively, we propose to estimate the covariance matrix for the Euclidean parameter estimator $\widehat{\boldsymbol{\xi}}$ based on the profile likelihood approach[32]. Define the following profile log-likelihood:

$$\mathrm{pl}(\boldsymbol{\xi}) = \max_{\Lambda \in \mathcal{B}_\Lambda} \log L^{obs}(\Lambda; \boldsymbol{\xi}),$$

where $\mathcal{B}_\Lambda$ is the set of step functions with nonnegative jumps at the observed distinct failure times, and this function can be computed by updating only Λ in the EM algorithm with $\boldsymbol{\xi}$ fixed. Then, the covariance matrix estimate is the inverse of

$$\sum_{i=1}^{n} \left[\left\{ \frac{\partial}{\partial \boldsymbol{\xi}} \mathrm{pl}_i(\boldsymbol{\xi}) \bigg|_{\boldsymbol{\xi}=\widehat{\boldsymbol{\xi}}} \right\}^{\otimes 2} \right],$$

where pl_i is the contribution to pl from the ith subject, and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$. For $k = 1, \ldots, p$ where $p = \dim(\boldsymbol{\xi})$, the kth element in the vector $\partial \mathrm{pl}_i(\boldsymbol{\xi})/\partial \boldsymbol{\xi}$ can be approximated numerically by the central numerical difference:

$$\frac{\mathrm{pl}_i(\boldsymbol{\xi} + h_n e_k) - \mathrm{pl}_i(\boldsymbol{\xi} - h_n e_k)}{2h_n},$$

where e_k is the kth canonical vector in $\mathcal{R}^p$ and h_n is a pre-specified perturbation constant. Setting $\widehat{\Lambda}$ as the initial value, the abridged EM algorithm typically converges in just a few iterations.

4 Simulation Studies

We conduct a simulation study to evaluate the finite-sample performance of the proposed methods. The covariates X and Z are generated independently from $\mathrm{N}(0, 1)$ and $\mathrm{Unif}(0, 4)$, respectively. The survival times are generated according to model (2) with $\Lambda(t) = t^2$. Two scenarios, namely Scenarios I and II, are considered, with regression parameters set to be $(\gamma, \beta, \alpha_1, \alpha_2) = (0.5, -1, 2, 1.5)$ and $(-0.5, 1, -2, -1)$, respectively. The major difference between Scenarios I and II lies in the change-point pattern. For Scenario I, the effect of Z on the log-hazard function is negative for $Z < \eta$ (because $\beta < 0$), while the effect of Z is positive for $Z \geq \eta$ (because $\beta + \alpha_2 > 0$). For Scenario II, the effect of Z is positive for $Z < \eta$ and is zero for $Z \geq \eta$. The value of α_1 represents a vertical shift in the log-hazard function when Z passes the

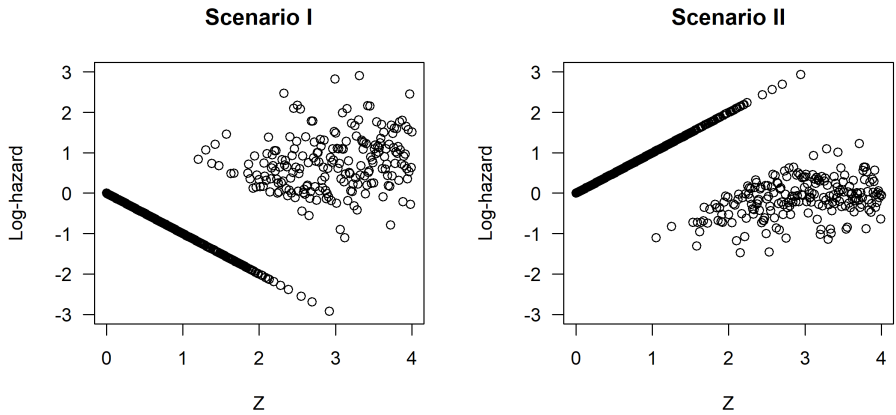

Figure 1. Plot of the effects of Z on the log-hazard function in the simulation study, based on one set of realizations of the change-point η for each scenario with $\mu = 2$ and $\sigma = 0.5$.

change point η. For each scenario, we consider two values for the mean of the change-point η_i's, specifically $\mu = 1.5$ and $\mu = 2$, and set the standard deviation of η_i's to be $\sigma = 0.5$. Figure 1 illustrates the change-point patterns for $\mu = 2$ in the two scenarios. For $\mu = 1.5$ and $\mu = 2$, about 63% and 50% of subjects have Z exceeding the corresponding change-point η, respectively. The censoring time follows $\mathrm{Unif}(0, 5)$, and the censoring proportions vary between 10% and 20% in the two scenarios. We consider sample sizes of $n = 500, 1000$, and 2000.

We set $m = 10$ nodes in applying the quadrature rule for truncated normal distributions. The convergence criterion of the EM algorithm is set to be 10^{-3}. The perturbation constant h_n for variance estimation is set to be $\mathcal{K}n^{-1}$ where $\mathcal{K} = 5$; our empirical results show that the standard error estimate differs only in the third decimal places when $\mathcal{K}$ increases from 1 to 10.

Table 1 summarizes the parameter estimates obtained based on 1000 replicates for each scenario. In both scenarios, the proposed estimator is virtually unbiased, and the averaged standard error estimate closely resembles the empirical standard deviation, suggesting that the profile likelihood approach provides a good approximation to the variance-covariance matrix. Also, the confidence intervals generated based on the normal approximation provide proper coverage probabilities, which are close to the 95% nominal level, for all Euclidean parameters. As expected, the parameter estimates become more accurate as the sample size n increases. The EM algorithm always converges in the estimation procedures.

To investigate the performance of the fixed change-point model when the change-point is random, we fit model (1) to the previous sets of simulated data. A grid search method, considering a profile of the change-point parameter η^*, is adopted. To avoid edge effects, we partition the interval $[0.5, 3.5]$ equally with grid size 0.01 to search for the optimal

Table 1. Main simulation results

Scenario	Par	True	$n = 500$				$n = 1000$				$n = 2000$			
			Bias	ESD	ESE	EC	Bias	ESD	ESE	EC	Bias	ESD	ESE	EC
I	γ	0.5	−0.002	0.077	0.077	0.95	0.000	0.052	0.053	0.95	0.001	0.035	0.037	0.96
	β	−1	−0.008	0.247	0.243	0.95	−0.006	0.167	0.163	0.95	−0.005	0.110	0.113	0.96
	α_1	2	0.020	0.312	0.310	0.95	0.017	0.219	0.214	0.94	0.014	0.155	0.150	0.94
	α_2	1.5	−0.012	0.342	0.322	0.94	−0.006	0.220	0.219	0.95	0.001	0.147	0.152	0.97
	μ	1.5	0.009	0.118	0.109	0.93	0.003	0.077	0.076	0.95	0.001	0.051	0.054	0.96
	σ	0.5	−0.020	0.088	0.085	0.95	−0.008	0.059	0.058	0.95	−0.005	0.040	0.040	0.95
I	γ	0.5	0.002	0.072	0.073	0.95	0.005	0.051	0.051	0.95	0.000	0.035	0.035	0.96
	β	−1	−0.009	0.189	0.185	0.94	−0.009	0.122	0.126	0.96	−0.003	0.088	0.088	0.95
	α_1	2	0.015	0.344	0.317	0.96	0.028	0.222	0.219	0.94	0.008	0.167	0.154	0.93
	α_2	1.5	−0.001	0.311	0.305	0.95	0.002	0.210	0.210	0.94	−0.006	0.151	0.147	0.94
	μ	2	−0.002	0.124	0.112	0.93	0.002	0.077	0.076	0.94	0.000	0.056	0.054	0.94
	σ	0.5	−0.012	0.078	0.079	0.96	−0.009	0.054	0.053	0.94	−0.006	0.038	0.038	0.94
II	γ	−0.5	0.001	0.072	0.070	0.94	0.001	0.051	0.049	0.94	−0.001	0.035	0.035	0.96
	β	1	−0.026	0.250	0.233	0.93	−0.013	0.166	0.163	0.94	−0.006	0.115	0.115	0.95
	α_1	−2	−0.011	0.275	0.270	0.95	−0.017	0.185	0.188	0.95	−0.006	0.133	0.132	0.95
	α_2	−1	0.040	0.309	0.289	0.94	0.029	0.204	0.201	0.94	0.013	0.143	0.141	0.94
	μ	1.5	0.012	0.111	0.101	0.92	0.005	0.074	0.070	0.93	0.001	0.048	0.050	0.95
	σ	0.5	−0.014	0.106	0.095	0.94	−0.009	0.070	0.065	0.93	−0.004	0.047	0.046	0.93
II	γ	−0.5	−0.005	0.068	0.066	0.94	0.000	0.046	0.046	0.94	0.001	0.032	0.032	0.96
	β	1	−0.003	0.174	0.161	0.94	−0.005	0.121	0.113	0.92	−0.006	0.079	0.080	0.95
	α_1	−2	−0.029	0.263	0.267	0.96	−0.012	0.181	0.186	0.96	−0.006	0.130	0.130	0.95
	α_2	−1	0.033	0.277	0.254	0.93	0.020	0.185	0.177	0.94	0.013	0.124	0.124	0.95
	μ	2	0.003	0.108	0.096	0.91	0.003	0.071	0.068	0.94	0.003	0.048	0.048	0.95
	σ	0.5	−0.018	0.088	0.083	0.94	−0.010	0.059	0.058	0.95	−0.006	0.042	0.040	0.94

Par, parameter; ESD, empirical standard deviation; ESE, estimated standard error; EC, empirical coverage with 95% nominal level.

change-point parameter value that maximizes the likelihood function. When the change-point is fixed and the survival times are right-censored, the partial likelihood approach can be easily implemented to obtain the maximum likelihood estimates of the regression parameters. Table 2 provides the bias and empirical standard deviation of the maximum likelihood estimator. Here, we regard the value of μ as the true value of η^*. Overall, the estimator is severely biased and does not resemble the true value of the parameter of interest, including the change-point unrelated parameter γ. Therefore, if the underlying model contains a random rather than a fixed change-point, assuming a fixed change-point can be problematic and may give misleading results for almost all regression parameters.

It is also of interest to evaluate the finite-sample performance of the proposed estimator when the change point is indeed fixed. We re-consider scenario I with $\mu = 2$ and $\sigma = 0$, and the results are provided in the supplemental material. We observe that the estimator of the regression parameters is unbiased with averaged standard error aligned closely with the empirical standard deviation. However, the standard error of the estimator of σ is overestimated and the coverage probability for μ is inflated. These might be due to the fact that the proposed estimator of σ cannot be sufficiently close to the true value 0.

To mimic the setting of the real data, we perform an additional simulation study with a binary covariate and with the censoring time C generated from $U(0, 2)$ (yielding a censoring proportion of about 50%). The simulation settings and results are presented in the supplemental material. We can see that the estimators are virtually unbiased, the averaged standard errors closely match the empirical standard derivations, and the confidence intervals yield proper coverage probabilities.

Table 2. Estimation results based on the fixed change-point model in (1)

Scenario	Par	True	$n = 500$		$n = 1000$		$n = 2000$	
			Bias	ESD	Bias	ESD	Bias	ESD
I	γ^*	0.5	−0.138	0.059	−0.145	0.040	−0.146	0.028
	β^*	−1	0.635	0.394	0.658	0.256	0.660	0.212
	α_1^*	2	−1.232	0.388	−1.343	0.313	−1.425	0.258
	α_2^*	1.5	−0.672	0.478	−0.654	0.331	−0.622	0.279
	η^*	1.5	0.157	0.401	0.110	0.346	0.074	0.315
I	γ^*	0.5	−0.127	0.060	−0.128	0.041	−0.133	0.030
	β^*	−1	0.437	0.232	0.417	0.161	0.398	0.130
	α_1^*	2	−1.355	0.588	−1.471	0.468	−1.592	0.387
	α_2^*	1.5	−0.361	0.357	−0.281	0.255	−0.225	0.193
	η^*	2	−0.084	0.437	−0.139	0.362	−0.210	0.321
II	γ^*	−0.5	0.104	0.058	0.109	0.041	0.111	0.028
	β^*	1	−0.589	0.485	−0.676	0.327	−0.750	0.265
	α_1^*	−2	1.190	0.357	1.256	0.186	1.346	0.129
	α_2^*	−1	0.496	0.582	0.594	0.397	0.647	0.320
	η^*	1.5	−0.066	0.470	−0.134	0.322	−0.162	0.253
II	γ^*	−0.5	0.099	0.058	0.106	0.041	0.110	0.028
	β^*	1	−0.374	0.355	−0.413	0.228	−0.446	0.167
	α_1^*	−2	1.146	0.319	1.254	0.236	1.332	0.154
	α_2^*	−1	0.240	0.424	0.244	0.301	0.250	0.227
	η^*	2	−0.322	0.323	−0.350	0.276	−0.371	0.222

Par, parameter; ESD, empirical standard deviation.

5 Application

The German Breast Cancer Study Group (GBSG) conducted a randomized clinical trial between 1983 and 1989[18;19] and the data consisted of the disease-free survival times of 686 primary node-positive breast cancer patients. Lee et al.[17] analyzed this motivating dataset and tested for the presence of change-point(s) in age effects, but the change-point(s) were assumed to be unknown constant(s). The authors demonstrated the presence of a change-point in model (1) with its estimate located at age 50 using the maximal score and Wald tests. In a similar setting, Lee and Lam[10] proposed a sequential maximal likelihood ratio testing approach for the presence of multiple change-points in the model related to (1) and showed that there were two estimated (fixed) change-points in age located before and after the menopausal age. In this paper, we re-analyze the data using model (2), assuming that the change-point is individual-specific and random in nature. In the model, we assume that the multiplicative effect of the age variable on the baseline hazard changes as age passes a subject-specific change-point η, whereas the effects of other covariates are constant.

By the end of the study, 299 of the patients experienced recurrence or death, and the rest were right-censored. Various baseline covariates, treated as $\boldsymbol{X}$, were present in the data, including tumor grade, hormonal treatment indicator, and the number of

positive nodes. We dichotomize tumor grade into grade I and grade II/III, and, following the research work of Schumacher et al.[18], categorize the number of positive nodes into $(0, 3]$, $(3, 9]$ and $[9, \infty]$. The age at diagnosis is treated as Z in the model. Seven patients aged at or below 30 are excluded from the analysis as in the work of Lee et al.[17]. The resulting data set contains 679 observations, and the age at diagnosis ranges from 31 to 80.

We set the convergence criterion to be 10^{-5}, initial values for $\boldsymbol{\xi}$ and Λ to be $\boldsymbol{\xi}^{(0)} = \mathbf{0}$ and $\Lambda^{(0)}\{Y_i\} = n^{-1}$ for each Y_i where $\Delta_i = 1$, respectively, and consider three sets of initial values, namely $(45, 10)$, $(50, 10)$ and $(55, 10)$ for (μ, σ), in the EM algorithm. For all three sets of initial values, the algorithm converges to the same set of estimates. The estimated standard errors and 95% confidence intervals are computed based on $\mathcal{K} = 5$ as in the simulation studies. The results are summarized in Table 3. All the change-point related regression parameters $(\beta, \alpha_1, \alpha_2)$ are shown to be significant. The change-point distribution is centered at 49.55 which is almost identical to 50 as reported in Lee et al.[17]. However, the change-point we considered here is random and has an estimated standard deviation of 1.169, whereas the change-point estimate in Lee et al.[17] is fixed without variation.

For further illustration of the results, we obtain the posterior expectation $\widehat{\mathrm{E}}(\eta_i)$ for $i = 1, \ldots, 679$ which is a by-product of the EM algorithm upon convergence, and plot the predicted age effects of the subjects on the log-hazard function in Figure 2. We can interpret the diagram in the following manner: (i) the effect of age comprises two (lower and upper) regimes that characterize whether a change has occurred; (ii) it is almost certain that individuals from the age group $(30, 48]$ belong to the lower regime; (iii) For those aged 49 or above, the risk shifts from the lower to the upper regime with a jump of magnitude α_1 at age $\widehat{\mathrm{E}}(\eta_i)$; and (iv) patients with their age at diagnosis surpassing $\widehat{\mathrm{E}}(\eta_i)$ stay in the upper regime, such that this group may deserve more attention in the therapy. Our findings are largely similar to the pattern obtained in the research work of Adami et al.[33] where they analyzed the survival of $57,068$ patients diagnosed with breast cancer; they reported that women aged 45 to 49 have the best prognosis, but the relative survival rate declined markedly after the age of 49. The medically diagnosed menopausal statuses at the study entry of the patients are also available in the dataset. To further explore how the latent change-point variable could potentially be associated with the menopausal status of the patients, we compare their age at diagnosis with $\widehat{\mathrm{E}}(\eta_i)$'s. In particular, 235 out of 283 peri-menopausal patients (83.0%) have their age at diagnosis smaller than $\widehat{\mathrm{E}}(\eta_i)$, whereas 370 out of 396 post-menopausal patients (93.4%) have their age at diagnosis larger than or equal to $\widehat{\mathrm{E}}(\eta_i)$. These findings suggest that the change in the baseline risk pattern may be attributed to the change in menopausal statuses, induced by a random threshold in age, thus being related to the changes in the hormonal environment of a patient.

For comparison, we fit the Cox PH model with covariates age at diagnosis, menopausal status, their interaction term, and other covariates in the random change-point model; menopausal age is not observed in the data. Note that this model is different from the proposed model even if η_i truly corresponds to the menopausal age, because our

Table 3. Application to GBSG data based on the random change-point model

Covariate	Estimate	Standard error	95% Confidence interval
Tumor grade (II/III)	0.739	0.251	(0.247, 1.230)
Hormonal treatment	-0.478	0.129	(-0.730, -0.226)
Positive node $\in (3, 9]$	0.742	0.135	(0.478, 1.006)
Positive node > 9	1.405	0.165	(1.082, 1.727)
Age	-0.068	0.020	(-0.107, -0.028)
$\mathrm{I}(\mathrm{Age} > \eta)$	0.898	0.289	(0.331, 1.465)
$\mathrm{I}(\mathrm{Age} > \eta)\times(\mathrm{Age}-\eta)$	0.061	0.025	(0.013, 0.109)
μ	49.55	0.994	(47.60, 51.50)
σ	1.169	1.041	(0.204, 6.694)

Table 4. Application to GBSG data based on the Cox PH model with $\widehat{\mu} = 49.55$

Covariate	Estimate	Standard error	95% Confidence interval
Tumor grade (II/III)	0.760	0.244	(0.281, 1.239)
Hormonal treatment	-0.429	0.130	(-0.683, -0.175)
Positive node $\in (3, 9]$	0.741	0.136	(0.474, 1.007)
Positive node > 9	1.349	0.154	(1.048, 1.651)
Age	-0.031	0.017	(-0.064, 0.001)
Post-menopausal	0.180	0.194	(-0.201, 0.561)
Post-menopausal$\times$(Age$-\widehat{\mu}$)	0.045	0.020	(0.006, 0.084)

model includes the latent menopausal age, whereas the Cox PH model includes just the menopausal status. The results based on the Cox PH model are reported in Table 4. We subtract the age by $\widehat{\mu}$ in the interaction term to make an easier comparison with the results of our model. This subtraction will not affect the model's predictive accuracy as any constant terms associated with α_2 are absorbed by α_1. One can see that the effects of all covariates share the same signs as in the proposed model. In particular, the effect of age is also negative before menopause and changes to positive after menopause. We use a cross-validation method to evaluate the predictive performance of the two models. The original data ($n = 679$) are randomly split into a training set and a testing set with the ratio $2 : 1$, which corresponds to 453 and 226 observations, respectively. Then, use the training data to fit the random change-point model and the Cox PH model and evaluate the estimated risk scores of the subjects in the testing set using the fitted models; for the random change-point model, we replace the random effect η_i by $\widehat{\mathrm{E}}(\eta_i)$ for $i = 1, \ldots, 226$. We estimate the C-index of the estimated risk scores in the testing data, using the estimator of Uno et al.[34]. We repeat the splitting procedure 100 times. The proposed model yields a larger C-index than the Cox PH model 73 times, and the average C-index values of the proposed and Cox PH model are 0.678 and 0.666, respectively. This suggests that the random change-point model has higher predictive power than the Cox PH model for the survival outcomes of the breast cancer patients.

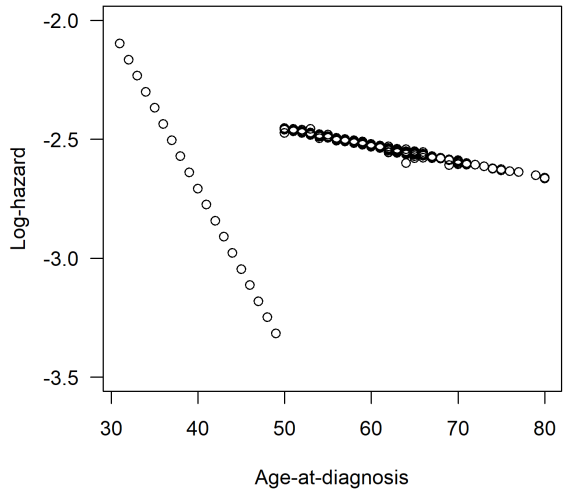

Figure 2. The fitted age effects for the GBSG data, with the change-point η_i predicted by the posterior expectation of η_i.

6 Conclusion

The change-point model serves as an important alternative to the linear model in biomedical studies when the effect of a covariate on the response variable is indeed non-linear. The traditional fixed universal change-point model assumes that the change-point parameter is fixed. The method is less appealing when conducting statistical inference for the change-point estimator because its asymptotic distribution is mathematically intractable. Motivated by the GBSG study, we consider the plausibly individual-specific nature of the change-point and develop a novel survival model with a random change-point. The maximum likelihood estimator of the Euclidean parameter is shown to be consistent and asymptotically normal, so the confidence intervals for the parameters can be easily obtained based on normal approximation.

The proposed model is similar to a fixed change-point model with measurement error on the change-point variable Z. In particular, let $\eta_i = \mu + \epsilon_i$, where ϵ_i's are independent and identically distributed according to a zero mean normal distribution with variance σ^2. If we let $Z_i^* = Z_i - \epsilon_i$, the conditional hazard function in (2) for the ith individual ($i = 1, \ldots, n$) can be written as

$$\lambda(t) \exp\left\{\boldsymbol{\gamma}^T \boldsymbol{X}_i + \beta Z_i^* + \beta\epsilon_i + \alpha_1 \mathrm{I}(Z_i^* \geq \mu) + \alpha_2 \mathrm{I}(Z_i^* \geq \mu)(Z_i^* - \mu)\right\}.$$

Here, we can think of Z^* as the true covariate and Z as the observed value with an additive measurement error of $-\epsilon$. If there is no main effect of Z (that is, $\beta = 0$), then the

proposed model coincides with a fixed change-point model with measurement error on the change-point covariate. When $\beta \neq 0$, the two models differ by an individual-specific log-normal frailty term with parameters 0 and $\beta^2\sigma^2$.

In clinical trials, a typical phenomenon is that not all individuals are susceptible to the event of interest, and the non-susceptible individuals are typically regarded as cured. For instance, a breast cancer patient can be free from recurrence and death caused by breast cancer. In this case, a cure rate model with a change-point in the survival component may be more suitable in describing the underlying survival mechanism of the patients. For right-censored data, Zhao et al.[35] considered the mixture cure model assuming a Cox PH model with a change-point for the latency component, and they estimated the parameters based on the Bayesian paradigm. Wang et al.[36] recently studied a similar mixture cure model under the classical approach where the change-point parameter is assumed to be fixed. Following their work, the proposed random change-point model can be extended naturally to accommodate a cure fraction in the population. Presumably, with a covariate-dependent/independent cure rate parameter, the EM algorithm stated in Section 2 can be reused with a modification made for the presence of an additional cure status latent variable.

In this paper, we assume that the change-point of each subject follows a common normal distribution. In the future, we may consider a population with a subgroup of subjects with fixed change-point. In particular, we define a latent binary variable U, such that the change-point is $(1-U)\eta_0 + U\eta$, where η_0 is a fixed unknown parameter, and η is as defined in this paper. In this model, subjects with $U = 0$ has a fixed change-point at η_0, whereas subjects with $U = 1$ has a random change-point η. The proposed EM algorithm can be extended to incorporate the additional latent variable U.

For ease of computation and interpretation, we assume a normal distribution for the change-point, but this parametric assumption could be restrictive. Alternatively, we may consider a more flexible parametric class of distributions or treat it as a nonparametric component in the model. For the latter, nonparametric approaches, such as kernel density estimation method or polynomial splines approximation, can be used. However, this typically requires a much larger sample size for accurate approximation, and the numerical stability of the estimation procedures for this type of model is in question.

Acknowledgements

The authors would like to thank the Editor, Associate Editor, and two reviewers for their valuable comments and suggestions that greatly improved the quality of this work.

Appendix — Model Identifiability

Let Λ_0 be the true value of Λ and τ be the end-of-study time. The following conditions are required for model identifiability.

(C1) The function Λ_0 is strictly increasing and continuous on $[0, \tau]$ with $\Lambda_0(0) = 0$.

(C2) The variable Z is continuous, and the support of the conditional distribution of Z given any value of $\boldsymbol{X}$ is $\mathbb{R}$.

(C3) If there exist a number a and a vector $\boldsymbol{b}$ of appropriate dimension such that $a + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X} = 0$ with probability 1, then $a = 0$ and $\boldsymbol{b} = \boldsymbol{0}$.

(C4) The censoring time C satisfies $P(C \geq \tau) = P(C = \tau) > \epsilon$ for some positive constant ϵ.

Under conditions (C1)–(C4), the parametric components of model (2) are identifiable, and $\Lambda(t)$ is identifiable over $t \in [0, \tau]$.

To prove model identifiability, we consider a single observation and drop the subscript i. By the continuity of Λ_0 and condition (C4), we can set the survival time to be right censored at any time point within $[0, \tau]$ when establishing identifiability. Suppose that there exist two sets of parameters, $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$, such that the survival probabilities at $t \in [0, \tau]$ evaluated at the two sets of parameters are equal almost surely, i.e.,

$$\begin{aligned} &\int e^{-\Lambda(t)e^{\boldsymbol{\gamma}^T\boldsymbol{X}+\beta Z+\alpha_1 \mathrm{I}(Z\geq\eta)+\alpha_2 \mathrm{I}(Z\geq\eta)(Z-\eta)}} \phi(\eta;\mu,\sigma)\,\mathrm{d}\eta \\ =&\int e^{-\widetilde{\Lambda}(t)e^{\widetilde{\boldsymbol{\gamma}}^T\boldsymbol{X}+\widetilde{\beta} Z+\widetilde{\alpha}_1 \mathrm{I}(Z\geq\eta)+\widetilde{\alpha}_2 \mathrm{I}(Z\geq\eta)(Z-\eta)}} \phi(\eta;\widetilde{\mu},\widetilde{\sigma})\,\mathrm{d}\eta \end{aligned} \tag{4}$$

for all $(\boldsymbol{X}, Z)$ and $t \in [0, \tau]$. It suffices to show that (4) implies $\boldsymbol{\gamma} = \widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}, \alpha_1 = \widetilde{\alpha}_1, \alpha_2 = \widetilde{\alpha}_2, \mu = \widetilde{\mu}, \sigma = \widetilde{\sigma}$ and $\Lambda(t) = \widetilde{\Lambda}(t)$ for $t \in [0, \tau]$. The left-hand side of (4) is

$$\begin{aligned} &\int_{-\infty}^{Z} e^{-\Lambda(t)e^{\boldsymbol{\gamma}^T\boldsymbol{X}+(\beta+\alpha_2)Z+\alpha_1-\alpha_2\eta}} \phi(\eta;\mu,\sigma)\,\mathrm{d}\eta + P(\eta \geq Z)e^{-\Lambda(t)e^{\boldsymbol{\gamma}^T\boldsymbol{X}+\beta Z}} \\ =&e^{-\Lambda(t)e^{\boldsymbol{\gamma}^T\boldsymbol{X}+\beta Z}} + O(e^{-Z^2/(2\sigma^2)}), \end{aligned} \tag{5}$$

where the second term is finite as $Z \to -\infty$. Differentiation with respect to t and setting $t = 0$ yield $-\lambda(0)e^{\boldsymbol{\gamma}^T\boldsymbol{X}+\beta Z} + O(e^{-Z^2/(2\sigma^2)})$. Hence, we have

$$\lambda(0)e^{\boldsymbol{\gamma}^T\boldsymbol{X}} = \widetilde{\lambda}(0)e^{\widetilde{\boldsymbol{\gamma}}^T\boldsymbol{X}}e^{(\widetilde{\beta}-\beta)Z} + \frac{O(e^{-Z^2/(2\sigma^2)}) + O(e^{-Z^2/(2\widetilde{\sigma}^2)})}{e^{\beta Z}}$$

Considering $Z \to -\infty$, we conclude that $\lambda(0) = \widetilde{\lambda}(0), \beta = \widetilde{\beta}$ and $\boldsymbol{\gamma} = \widetilde{\boldsymbol{\gamma}}$.

Next, differentiating (5) with respect to t and setting $t = 0$ yield

$$-\lambda(0)e^{\boldsymbol{\gamma}^T\boldsymbol{X}+\beta Z}\left\{1 + e^{\alpha_1+\alpha_2(Z-\mu)+\frac{1}{2}\sigma^2\alpha_2^2}\Phi\left(\frac{Z-\mu}{\sigma}+\sigma\alpha_2\right) - \Phi\left(\frac{Z-\mu}{\sigma}\right)\right\},$$

where Φ denotes the distribution function of a standard normal random variable. Therefore,

$$\begin{aligned} &e^{\alpha_1+\alpha_2(Z-\mu)+\frac{1}{2}\sigma^2\alpha_2^2}\Phi\left(\frac{Z-\mu}{\sigma}+\sigma\alpha_2\right) - \Phi\left(\frac{Z-\mu}{\sigma}\right) \\ =&e^{\widetilde{\alpha}_1+\widetilde{\alpha}_2(Z-\widetilde{\mu})+\frac{1}{2}\widetilde{\sigma}^2\widetilde{\alpha}_2^2}\Phi\left(\frac{Z-\widetilde{\mu}}{\widetilde{\sigma}}+\widetilde{\sigma}\widetilde{\alpha}_2\right) - \Phi\left(\frac{Z-\widetilde{\mu}}{\widetilde{\sigma}}\right). \end{aligned} \tag{6}$$

We obtain $\alpha_2 = \widetilde{\alpha}_2$ by considering $Z \to \infty$. Differentiating (6) with respect to Z yields

$$\begin{aligned} &\alpha_2 e^{\alpha_1+\alpha_2(Z-\mu)+\frac{1}{2}\sigma^2\alpha_2^2}\Phi\left(\frac{Z-\mu}{\sigma}+\sigma\alpha_2\right) - \frac{1}{\sigma}\phi\left(\frac{Z-\mu}{\sigma}\right) \\ &+ \frac{1}{\sigma}e^{\alpha_1+\alpha_2(Z-\mu)+\frac{1}{2}\sigma^2\alpha_2^2}\phi\left(\frac{Z-\mu}{\sigma}+\sigma\alpha_2\right) \\ =\ &\widetilde{\alpha}_2 e^{\widetilde{\alpha}_1+\widetilde{\alpha}_2(Z-\widetilde{\mu})+\frac{1}{2}\widetilde{\sigma}^2\widetilde{\alpha}_2^2}\Phi\left(\frac{Z-\widetilde{\mu}}{\widetilde{\sigma}}+\widetilde{\sigma}\widetilde{\alpha}_2\right) - \frac{1}{\widetilde{\sigma}}\phi\left(\frac{Z-\widetilde{\mu}}{\widetilde{\sigma}}\right) \\ &+ \frac{1}{\widetilde{\sigma}}e^{\widetilde{\alpha}_1+\widetilde{\alpha}_2(Z-\widetilde{\mu})+\frac{1}{2}\widetilde{\sigma}^2\widetilde{\alpha}_2^2}\phi\left(\frac{Z-\widetilde{\mu}}{\widetilde{\sigma}}+\widetilde{\sigma}\widetilde{\alpha}_2\right). \end{aligned} \tag{7}$$

Note that

$$\Phi\left(\frac{Z-\mu}{\sigma}+\sigma\alpha_2\right) = h(Z)e^{-\frac{Z^2}{2\sigma^2}+(\frac{\mu}{\sigma^2}-\alpha_2)Z}$$

for some $h(Z) = O(1/Z)$ as $Z \to -\infty$. Suppose that $\sigma^2 < \widetilde{\sigma}^2$. Dividing both sides of (7) by $e^{-Z^2/(2\sigma^2)}$ and setting $Z \to -\infty$, the left-hand side of (7) goes to ∞ but the right-hand side remains finite. Therefore, $\sigma^2 < \widetilde{\sigma}^2$ is impossible; we can similarly rule out $\sigma^2 > \widetilde{\sigma}^2$ and conclude that $\sigma^2 = \widetilde{\sigma}^2$.

We suppose without loss of generality that $\mu \le \tilde{\mu}$. Dividing (7) on both sides by $e^{\frac{1}{\sigma^2}\widetilde{\mu}Z}$, we obtain

$$\{c_1 h(Z) + c_2\}e^{\frac{1}{\sigma^2}(\mu-\widetilde{\mu})Z} = c_3\widetilde{h}(Z) + c_4$$

where c_1, c_2, c_3 and c_4 are constant with respect to Z, $c_1 > 0$, and $\widetilde{h}(Z) = O(1/Z)$. If $\mu \neq \widetilde{\mu}$, then the left-hand side goes to infinity and the right-hand side remains finite as $Z \to -\infty$, so we conclude that $\mu = \widetilde{\mu}$. Now, it is easy to see that $\alpha_1 = \widetilde{\alpha}_1$ follows from (6) and $\Lambda(t) = \widetilde{\Lambda}(t)$ for $t \in [0, \tau]$ follows from (4).

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was partially supported by the Central Guided Local Science and Technology Development Funds for Research Laboratories (No. 2021Szvup145).

Supplemental material

The R codes associated with this article can be found on Github: `https://github.com/lcyjames/RCPsurv`. It provides the codes to implement the proposed methods and reproduce the simulation results. The data set analyzed in this paper is publicly available from the `R` package `TH.data`. Supplemental material that contains extra simulation results is available online.

References

1. Shapiro J, Smith H and Schaffner F. Serum bilirubin: a prognostic factor in primary biliary cirrhosis. *Gut* 1979; 20: 137–140.
2. Willeit P, Raschenberger J, Heydon EE et al. Leucocyte telomere length and risk of type 2 diabetes mellitus: new prospective cohort study and literature-based meta-analysis. *PloS One* 2014; 9: e112483.
3. Cox DR. Regression models and life-tables. *J R Stat Soc B* 1972; 34: 187–220.
4. Liang KY, Self SG and Liu X. The Cox proportional hazards model with change point: An epidemiologic application. *Biometrics* 1990; 46: 783–793.
5. Pons O. Estimation in a Cox regression model with a change-point at an unknown time. *Statistics* 2002; 36: 101–124.
6. Liu M, Lu W and Shao Y. A Monte Carlo approach for change-point detection in the Cox proportional hazards model. *Stat Med* 2008; 27: 3894–3909.
7. He P, Fang L and Su Z. A sequential testing approach to detecting multiple change points in the proportional hazards model. *Stat Med* 2013; 32: 1239–1245.
8. Gandy A and Jensen U. On goodness-of-fit tests for Aalen's additive risk model. *Scand J Stat* 2005; 32: 425–445.
9. Jensen U and Lütkebohmert C. A Cox-type regression model with change-points in the covariates. *Lifetime Data Anal* 2008; 14: 267–285.
10. Lee CY and Lam KF. Survival analysis with change-points in covariate effects. *Stat Methods Med Res* 2020; 29: 3235–3248.
11. Luo X and Boyett JM. Estimations of a threshold parameter in Cox regression. *Commun Stat Theory* 1997; 26: 2329–2346.
12. Pons O. Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *Ann Stat* 2003; 31: 442–463.
13. Deng Y, Zeng D, Zhao J et al. Proportional hazards model with a change point for clustered event data. *Biometrics* 2017; 73: 835–845.
14. Li J and Jin B. Multi-threshold accelerated failure time model. *Ann Stat* 2018; 46(6A): 2657–2682.
15. Wang B, Li J and Wang X. Multi-threshold proportional hazards model and subgroup identification. *Stat Med* 2022; 41(29): 5715–5737.
16. Lam KF, Xu J and Xue H. Estimation of age effect with change-points on survival of cancer patients. *Stat Med* 2018; 37: 1732–1743.
17. Lee CY, Chen X and Lam KF. Testing for change-point in the covariate effects based on the Cox regression model. *Stat Med* 2020; 39: 1473–1488.
18. Schumacher M, Bastert G, Bojar H et al. Randomized 2 x 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. german breast cancer study group. *J Clin Oncol* 1994; 12: 2086–2093.
19. Schmoor C, Olschewski M and Schumacher M. Randomized and non-randomized patients in clinical trials: experiences with comprehensive cohort studies. *Stat Med* 1996; 15: 263–271.
20. Lange N, Carlin BP and Gelfand AE. Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. *J Am Stat Assoc* 1992; 87: 615–626.

21. Slate EH and Turnbull BW. Statistical models for longitudinal biomarkers of disease onset. *Stat Med* 2000; 19: 617–637.
22. Jacqmin-Gadda H, Commenges D and Dartigues JF. Random changepoint model for joint modeling of cognitive decline and dementia. *Biometrics* 2006; 62: 254–260.
23. Zeng D and Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data. *J R Stat Soc B* 2007; 69: 507–564.
24. Zeng D and Lin DY. Semiparametric transformation models with random effects for recurrent events. *J Am Stat Assoc* 2007; 102: 167–180.
25. Zeng D, Lin DY and Lin X. Semiparametric transformation models with random effects for clustered failure time data. *Stat Sinica* 2008; 18: 355–377.
26. Wong KY, Zeng D and Lin DY. Efficient estimation for semiparametric structural equation models with censored data. *J Am Stat Assoc* 2018; 113: 893–905.
27. Dempster A, Laird N and Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* 1977; 39: 1–38.
28. Liu Q and Pierce DA. A note on Gauss–Hermite quadrature. *Biometrika* 1994; 81: 624–629.
29. Golub GH and Welsch JH. Calculation of Gauss Quadrature rules. *Math Comput* 1969; 23: 221–230.
30. Burkardt J. The truncated normal distribution. *Department of Scientific Computing Website, Florida State University* 2014; 1: 35.
31. Varadhan R and Roland C. Simple and globally convergent methods for accelerating the convergence of any EM algorithm. *Scand J Stat* 2008; 35: 335–353.
32. Murphy SA and Van der Vaart AW. On profile likelihood. *J Am Stat Assoc* 2000; 95: 449–465.
33. Adami HO, Malker B, Holmberg L et al. The relation between survival and age at diagnosis in breast cancer. *New Engl J Med* 1986; 315: 559–563.
34. Uno H, Cai T, Pencina MJ et al. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011; 30: 1105–1117.
35. Zhao L, Feng D, Bellile EL et al. Bayesian random threshold estimation in a Cox proportional hazards cure model. *Stat Med* 2014; 33(4): 650–661.
36. Wang B, Li J and Wang X. Change point detection in Cox proportional hazards mixture cure model. *Stat Methods Med Res* 2021; 30: 440–457.