Statistica Sinica Preprint No: SS-2025-0305	
Title	Semiparametric Regression Analysis of Clustered Interval-censored Failure Time Data with Random Change Points and Application to Breast Cancer Study
Manuscript ID	SS-2025-0305
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0305
Complete List of Authors	Yichen Lou, Mingyue Du and Jianguo Sun
Corresponding Authors	Jianguo Sun
E-mails	suncolumbia@163.com
Notice: Accepted author version.	

Semiparametric Regression Analysis of Clustered Interval-censored Failure Time Data with Random Change Points and Application to Breast Cancer Study

Yichen Lou[1], Mingyue Du[2], and Jianguo Sun[3]

[1]School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore.
[2]School of Mathematics, Jilin University, Changchun, China.
[3]Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, China.
[*]Address correspondence to: suncolumbia@163.com

Abstract

Motivated by an international breast cancer study, we investigate regression analysis of clustered interval-censored failure time data in the presence of random change points. Although a large literature has been developed for regression analysis of clustered or interval-censored data, there does not seem to exist an established approach for the situation considered here. Change points can occur in many situations and one example is when a disease risk may shift abruptly once certain biological markers cross critical thresholds. For the problem, we propose a sieve maximum likelihood estimation procedure that can accommodate all three features, clustered structure, interval censoring and random change points. For the implementation of the proposed method, an EM algorithm is developed and the asymptotic properties of the resulting estimators are established. An extensive simulation study is conducted and indicates that the proposed method works well in practical situations. The proposed methodology is applied to the aforementioned breast cancer study.

Keywords: Clustered interval-censored data; EM algorithm; Random change point; Sieve estimation.

1 Introduction

This paper considers regression analysis of clustered interval-censored failure time data in the presence of random change points. A large literature has been established for regression analysis of failure time data when there exists clustering structure, interval censoring or change point, each of which makes the analysis more challenging. However, there does not seem to exist an established approach that is applicable when all three issues exist together.

This study was motivated by a large international breast cancer study conducted in Europe and North America (Huang et al., 2024). Breast cancer is one of the most prevalent malignancies among women and the second leading cause of cancer-related deaths among American women (Siegel et al., 2024). For the study, one goal is to identify risk factors that have significant effects on breast cancer-related events such as death and distant metastasis, meaning that the cancer spreading to distant organs or lymph nodes. Among others, menopause is widely recognized as a critical physiological transition in women's health and has been shown to influence breast cancer risk and progression significantly (Heer et al., 2020). In particular, postmenopausal women have been observed to have a higher risk of breast cancer progression compared to premenopausal women (Stavraky and Emmons, 1974). More details about the study are given below.

Clustered failure time data occur when study subjects come from or can be classified into different groups or clusters and the subjects within each cluster may be correlated (Huang et al., 2019, 2024). One common situation that yields such data is that the subjects within a cluster share or have same characteristics, medical practices, or environmental exposures, and a more specific example of this is siblings within a family. It is well-known that in the presence of cluster structure, the analysis that ignores it could lead to biased estimation and invalid inference, and to deal with it, a commonly used method is the frailty model approach, which uses latent variables to characterize the underlying relationship among the subjects within the clusters (Hanagal, 2011). Among others, Hougaard (2000) provides a comprehensive review of the analysis of clustered failure time data.

Interval-censored failure time data are the most general type of failure time data and a great deal of literature has been established for their analyses (Sun, 2006; Sun and Chen, 2022). By interval-censored data, we usually mean that the failure time of interest is observed only to belong to an interval instead of being observed or known exactly, and it is easy to see that this can naturally happen in, for example, any periodic follow-up study. It is worth noting that the analysis of interval-censored data is much more difficult than that of right-censored data since the former has a much more complicated structure. Many authors have investigated regression analysis of clustered interval-censored data (Lam et al., 2010; Sun, 2006; Yang et al., 2021). However, there is no study considering the situation where there exist change points.

Regression analysis of failure time data with change points has gained substantial attention across disciplines such as biomedicine, clinical research, and precision medicine (He et al., 2018; Lou et al., 2025; Pons, 2003). In clinical studies, for instance, the risk of disease progression may shift abruptly when certain biomarkers exceed critical thresholds, leading to changes in the underlying hazard function. Although many methods have been developed to deal with change points in the analysis of failure time data (Chen et al., 2024; Lee and Lam, 2020; Lee and Wong, 2023), most of the existing methods are for right-censored data except Du et al. (2025). However, Du et al. (2025) did not consider the cluster structure and thus their approach is not applicable to the situation discussed here.

In the following, we will propose a novel frailty-based framework for regression analysis of interval-censored failure time data that can simultaneously take into account both cluster structure and the presence of random change points. For estimation, we will develop a sieve maximum likelihood approach with the use of Bernstein polynomials and for the implementation of the proposed method, an EM algorithm will be provided. The resulting estimators are shown to be consistent and asymptotically normal.

The remainder of this paper is organized as follows. In Section 2, we will first present a frailty-based change point Cox model and then describe the observed likelihood function. Section 3 gives the proposed estimation procedure and the developed EM algorithm. In Section 4, we present some results from a simulation study conducted to assess the empirical performance of the proposed method and they suggest that it works well in practice. Section 5 applies it to the breast cancer study described above and Section 6 concludes with some discussion and remarks.

2 Notation, Models and Likelihood Function

Consider a failure time study and suppose that there are N independent and identically distributed clusters with n_i subjects in the ith cluster for $i = 1, \ldots, N$. For the jth subject in the ith cluster, let T_{ij} denote the failure time of interest and suppose that there exist two vectors of covariates denoted by $\boldsymbol{X}_{ij}$ and $\boldsymbol{Z}_{ij}$ that are p_1-dimensional and p_2-dimensional, respectively, whose meanings will be more specifically described below. Also suppose that there exists a latent variable b_i, and given $(\boldsymbol{X}_{ij}^\top, \boldsymbol{Z}_{ij}^\top, b_i)^\top$, $\{T_{i1}, \ldots, T_{i,n_i}\}$ are mutually independent and the T_{ij}'s follow the change point proportional hazards model

$$\Lambda(t \mid \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, b_i) = \Lambda(t) \exp\left\{ \boldsymbol{X}_{ij}^\top \boldsymbol{\beta} + \boldsymbol{Z}_{ij}^\top \boldsymbol{\alpha} + \boldsymbol{Z}_{ij}^\top \boldsymbol{\gamma}\, \mathbb{I}(U_{ij} > \zeta_{ij}) + b_i \right\} \quad (1)$$

in terms of the cumulative hazard function. In the above, $\Lambda(t)$ denotes an unknown baseline cumulative hazard function, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are vectors of unknown regression parameters, and U_{ij} is a univariate covariate that could be an element of $\boldsymbol{Z}_{ij}$ such that the effect of $\boldsymbol{Z}_{ij}$ on T_{ij} may depend on U_{ij} due to the random change point denoted by the latent variable ζ_{ij} in U_{ij}.

It is easy to see that the frailty b_i's characterizes the correlation among the subjects within the same cluster, and $\boldsymbol{X}_{ij}$ and $\boldsymbol{Z}_{ij}$ represent the covariates with constant and threshold-varying effects, respectively. In practice, the classification of covariates into $\boldsymbol{X}_{ij}$ or $\boldsymbol{Z}_{ij}$ should be guided by substantive knowledge and existing biological evidence regarding the threshold variable U_{ij}. The vector $\boldsymbol{Z}_{ij}$ should be constructed to include the variables whose effects are suspected to vary with the level of U_{ij} such as a treatment effect that may differ before and after a specific biological transition. Standard prognostic factors that are expected to exert a constant influence on risk should be included in $\boldsymbol{X}_{ij}$. In the following, we will assume that the b_i's follow a distribution $f_b(b_i;\,\sigma)$ with mean zero and unknown variance σ^2 and the latent variables ζ_{ij}'s have the density function $g_\zeta(\cdot\,;\boldsymbol{\nu})$ known up to a vector of unknown parameters $\boldsymbol{\nu}$. Also we will focus on the situation where the change point exists and assume that $\boldsymbol{\gamma}$ is non-zero. More comments on this will be given below.

Define $\boldsymbol{\theta} = \big(\boldsymbol{\psi},\, \Lambda\big)$ with $\boldsymbol{\psi} = (\boldsymbol{\beta}^\top,\, \boldsymbol{\alpha}^\top,\, \boldsymbol{\gamma}^\top,\, \boldsymbol{\nu}^\top,\, \sigma)^\top$ and suppose that one observes the interval-censored data given by $\boldsymbol{O} = \{O_i,\, i = 1, \ldots, N\}$ with $O_i = \{L_{ij},\, R_{ij},\, \boldsymbol{X}_{ij},\, \boldsymbol{Z}_{ij},\, U_{ij}\,;\, j = 1, \ldots, n_i\}$ and $L_{ij} < T_{ij} \le R_{ij}$. That is, T_{ij} is known only to belong to the interval $(L_{ij},\, R_{ij}]$. Then, under the independent interval censoring assumption, the likelihood function of $\boldsymbol{\theta}$ can be written as

$$
\begin{aligned}
L(\boldsymbol{\theta}\,;\boldsymbol{O}) &= \prod_{i=1}^{N}\int_{b_i}\left[\prod_{j=1}^{n_i}\int_{\zeta_{ij}}\Big\{S(L_{ij}) - S(R_{ij})\Big\}\, g_\zeta(\zeta_{ij};\,\boldsymbol{\nu})\, d\,\zeta_{ij}\right] f_b(b_i;\,\sigma)\, d\,b_i \\
&= \prod_{i=1}^{N}\int_{b_i}\Bigg(\prod_{j=1}^{n_i}\Bigg[\Big\{S^\sharp_{(1)}(L_{ij}) - S^\sharp_{(1)}(R_{ij})\Big\}\, G_\zeta(U_{ij};\,\boldsymbol{\nu}) \\
&\qquad\qquad + \Big\{S^\sharp_{(2)}(L_{ij}) - S^\sharp_{(2)}(R_{ij})\Big\}\Big\{1 - G_\zeta(U_{ij};\,\boldsymbol{\nu})\Big\}\Bigg]\Bigg)\, f_b(b_i;\,\sigma)\, d\,b_i \qquad (2)\\
&= \prod_{i=1}^{N}\int_{b_i}\left\{\prod_{j=1}^{n_i}\widetilde{L}^\sharp_{ij}(\boldsymbol{\theta}_{-\sigma}\,;\, O_{ij})\, f_b(b_i;\,\sigma)\right\} d\,b_i\,.
\end{aligned}
$$

In the above, $\boldsymbol{\theta}_{-\sigma}$ denotes $\boldsymbol{\theta}$ excluding σ, $G_\zeta(\cdot\,;\boldsymbol{\nu})$ the CDF of the ζ_{ij}'s,

$$
S^\sharp_{(1)}(t) = \exp\left[-\Big\{\Lambda(t)\exp(\boldsymbol{X}_{ij}^\top\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^\top\boldsymbol{\alpha} + \boldsymbol{Z}_{ij}^\top\boldsymbol{\gamma} + b_i)\Big\}\right],
$$

and

$$
S^\sharp_{(2)}(t) = \exp\left[-\Big\{\Lambda(t)\exp(\boldsymbol{X}_{ij}^\top\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^\top\boldsymbol{\alpha} + b_i)\Big\}\right].
$$

In the next section, we will discuss the estimation of the parameters based on the likelihood function above with the focus on $f_b(b_i;\,\sigma)$ being the normal density function. The method applies to other density functions too.

3 Sieve Maximum Likelihood Estimation

Now we discuss the estimation of the parameter $\boldsymbol{\theta}$ and for this, it is natural to maximize the likelihood function $L(\boldsymbol{\theta})$. On the other hand, it is apparent that this is not easy or straightforward due to the unknown function $\Lambda(\cdot)$. To address this, by following Ma et al. (2015) and others, we will employ the sieve approach that first approximates $\Lambda(\cdot)$ by Bernstein polynomials over $\mathcal{I} = [t_l, t_r]$, where t_l and t_r denote the lower and upper bounds of the observation times, respectively.

More specifically, let $\boldsymbol{\Theta}$ denote the parameter space of $\boldsymbol{\theta}$ and define the sieve space $\boldsymbol{\Theta}_n = \{\boldsymbol{\theta}_n = (\boldsymbol{\psi},\, \Lambda_n)\} = \mathcal{B} \otimes \Gamma_n$, where $\mathcal{B} = \{\boldsymbol{\psi} \in \mathbb{R}^d,\, \|\boldsymbol{\psi}\| \le M\}$, d is the dimension of $\boldsymbol{\psi}$, and

$$\Gamma_n = \left\{ \Lambda_n(t) = \sum_{q=0}^{m} \phi_q^* B_q(t, m, t_l, t_r) : \sum_{0\le q\le m} |\phi_q^*| \le M_n,\, 0 \le \phi_0^* \le \phi_1^* \le \cdots \le \phi_m^* \right\}.$$

In the above, $M_n = o(n^a)$ with $0 < a < 1/2$ and $n = \sum n_i$, the ϕ_l^*'s are unknown parameters to be estimated, and

$$B_q(t, m, t_l, t_r) = \binom{m}{q} \left(\frac{t - t_l}{t_r - t_l}\right)^q \left(1 - \frac{t - t_l}{t_r - t_l}\right)^{m-q},$$

where m denotes the degree of the Bernstein polynomials that is usually taken to be $m = O(n^v)$ for some $0 < v < 1/2$.

Define the sieve maximum likelihood estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\psi}},\, \widehat{\Lambda})$ of $\boldsymbol{\theta}$ as the value of $\boldsymbol{\theta}$ that maximizes the likelihood function $L(\boldsymbol{\theta})$ over the sieve space $\boldsymbol{\Theta}_n$. Note that in the definition of $\boldsymbol{\Theta}_n$, we need the constraints $0 \le \phi_0^* \le \phi_1^* \le \cdots \le \phi_m^*$ due to the non-negativity and monotonicity features of Λ, and it can be easily removed by the reparameterizations $\phi_0^* = \exp(\phi_0)$ and $\phi_q^* = \sum_{i=0}^{q} \exp(\phi_i)$ for the ϕ_q^*'s, $1 \le q \le m$. Also note that $\widehat{\boldsymbol{\theta}}$ does not have a closed form and for this, we will develop an EM algorithm.

For the development of the EM algorithm, we will first assume that the frailty b_i's were known or we had the complete data $\boldsymbol{O}^\sharp = \left\{ O_i^\sharp = (O_i,\, b_i),\, i = 1, \ldots, N \right\}$. Then one can easily derive the complete data likelihood function as

$$L^\sharp(\boldsymbol{\theta},\, \boldsymbol{b}) = \prod_{i=1}^{N} \left\{ \prod_{j=1}^{n_i} \widetilde{L}_{ij}^\sharp(\boldsymbol{\theta}_{-\sigma}\,;\, O_{ij}) \right\} f_b(b_i;\, \sigma) := \prod_{i=1}^{N} L_i^\sharp(\boldsymbol{\theta};\, O_i^\sharp)\,,$$

where $\boldsymbol{b} = (b_1, \ldots, b_N)^\top$. It follows that the complete data log-likelihood function has the form

$$\ell^\sharp(\boldsymbol{\theta},\, \mathbf{b}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \widetilde{\ell}_{ij}^\sharp(\boldsymbol{\theta}_{-\sigma}\,;\, O_{ij}) + \sum_{i=1}^{N} \log f_b(b_i;\, \sigma) \, := \widetilde{\ell}_{(1)}^\sharp(\boldsymbol{\theta}_{-\sigma}) + \ell_{(2)}^\sharp(\sigma)\,, \tag{3}$$

where $\widetilde{\ell}_{ij}^\sharp(\boldsymbol{\theta}_{-\sigma}\,;\, O_{ij}) = \log \widetilde{L}_{ij}^\sharp(\boldsymbol{\theta}_{-\sigma}\,;\, O_{ij})$.

For the E-step of the EM algorithm, at the $(k+1)$th iteration, we need to compute the expectation of (3) conditional on the observed data and the current estimates of the parameters, which has the form

$$\begin{aligned} Q(\boldsymbol{\theta} \mid \boldsymbol{O},\, \widehat{\boldsymbol{\theta}}^{(k)}) &= \mathbb{E}\left\{ \ell^{\sharp}(\boldsymbol{\theta},\, \boldsymbol{b}) \mid \boldsymbol{O},\, \widehat{\boldsymbol{\theta}}^{(k)} \right\} \\ &= \sum_{i=1}^{N} \left[\mathbb{E}\left\{ \widetilde{\ell}^{\sharp}_{i(1)}(\boldsymbol{\theta}_{-\sigma}) \mid O_i,\, \widehat{\boldsymbol{\theta}}^{(k)} \right\} + \mathbb{E}\left\{ \ell^{\sharp}_{i(2)}(\sigma) \mid O_i,\, \widehat{\boldsymbol{\theta}}^{(k)} \right\} \right]. \end{aligned} \tag{4}$$

For the expectation above, it is apparent that one needs to evaluate the following types of integrals

$$\mathbb{E}\left\{ \varrho(b_i) \mid O_i,\, \widehat{\boldsymbol{\theta}}^{(k)} \right\} = \int \varrho(b_i)\, \hbar\,(b_i \mid O_i,\, \widehat{\boldsymbol{\theta}}^{(k)})\, d\, b_i = \frac{\int_{-\infty}^{\infty} \varrho(b_i)\, L_i^{\sharp}(\boldsymbol{\theta}^{(k)};\, O_i^{\sharp})\, d\, b_i}{\int_{-\infty}^{\infty} L_i^{\sharp}(\boldsymbol{\theta}^{(k)};\, O_i^{\sharp})\, d\, b_i} \tag{5}$$

for a function $\varrho(b_i)$ of b_i, where $\hbar\,(b_i \mid O_i,\, \widehat{\boldsymbol{\theta}}^{(k)})$ denotes the conditional density function of b_i given the observed data and the k-th iterative estimate of $\boldsymbol{\theta}$. To calculate (5), we propose using the Gauss-Hermite quadrature integral to approximate the integrals in it. Specifically, with the integral in the form $\int \varrho(b)\exp(-b^2)\, d\, b$ with a specified function $\varrho(\cdot)$, we approximate the integral by the weighted sum of $\varrho(\cdot)$ at a number of suitably specified points:

$$\int \varrho(b)\exp(-b^2)\, d\, b \approx \sum_{q=1}^{M} \omega_q\, \varrho(b^{[q]})\,,$$

where for a user-specified positive integer M, $b^{[q]}$ denotes the q-th quadrature point and ω_q the corresponding weight. In particular, the larger M, the more accurate the approximation. In applications, however, the specification of M is usually driven by the trade-off between the approximation accuracy and the computation time. For the numerical studies below, we use $M = 8$. Alternatively, when dealing with other types of probability density functions for the b_i's, one may employ the Monte Carlo method as described in Chen et al. (2009).

For the M-step of the EM algorithm, at the $(k+1)$-th iteration, we need to maximize the conditional expectation (4) with respect to $\boldsymbol{\theta}$ to obtain the updated estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$. More specifically, to obtain the updated estimate of $\boldsymbol{\theta}$, we only need to find the updated estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}_{-\sigma}$ by finding the value that maximizes $\mathbb{E}\{\widetilde{\ell}^{\sharp}_{(1)}(\boldsymbol{\theta}_{-\sigma}) \mid O_i,\, \widehat{\boldsymbol{\theta}}^{(k)}\}$. For the updated estimate of σ, we can calculate it from

$$\widehat{\sigma}^{(k+1)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left\{ b_i^2 \mid O_i,\, \widehat{\boldsymbol{\theta}}^{(k)} \right\}}\,.$$

To check the convergence of the EM algorithm above, one can use various criteria. A simple approach, which is used in the numerical studies below, is to calculate the differences of the values of the likelihood functions between adjacent iterations and declare the convergence if the difference is smaller than a predefined positive number such as 10^{-3}. For the implementation of the method above, it is apparent that one also needs to choose the degree of the Bernstein polynomials m and for this, a common approach is to try different values and compare the results. Our experience suggests that the estimation results are usually robust with respect to m and a small value is generally good enough. The same has been pointed out by others.

Let $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}_0, \Lambda_0)$ denote the true value of $\boldsymbol{\theta}$. We now establish the asymptotic properties of the proposed estimators in the following theorems with the proofs sketched in the Appendix.

Theorem 1. *Suppose that the regularity conditions 1 - 7 given in the Appendix hold. Then the estimator $\widehat{\boldsymbol{\theta}}$ is consistent and we have that*

$$d\,(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_p\left(n^{-(1-v)/2} + n^{-rv/2}\right),$$

where $v \in (0,1)$ such that $m = o(n^v)$ and r is defined in Condition 3.

Theorem 2. *Suppose that the regularity conditions 1 - 7 given in the Appendix hold with $r > 2$ in Condition 3. Then it can be shown that $\sqrt{n}\,(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)$ converges in distribution to a zero-mean multivariate normal distribution.*

To make the inference about $\boldsymbol{\psi}$, it is apparent that one needs to estimate the covariance matrix of $\widehat{\boldsymbol{\psi}}$. For this and following Yang et al. (2021) and others, we suggest to employ the profile likelihood approach and the numerical studies below indicate that it performs well.

4 A Simulation Study

In this section, we present some results obtained from a simulation study conducted to assess the empirical performance of the sieve maximum likelihood estimation procedure proposed in the previous sections. In the study, we considered a two-dimensional vector of covariates $\boldsymbol{X}_{ij} = (X_{ij,1}, X_{ij,2})^\top$ with $X_{ij,1} \sim$ Uniform$(-1,1)$ and $X_{ij,2} \sim$ Normal$(0,1)$ and a three-dimensional vector of covariates $\boldsymbol{Z}_{ij} = (Z_{ij,1}, Z_{ij,2}, Z_{ij,3})^\top$ with $Z_{ij,1} \sim$ Normal$(0.5,1)$, $Z_{ij,2} \sim$ Bernoulli(0.6) and $Z_{ij,3} \sim$ Uniform$(-2,2)$. The covariates U_{ij}'s were generated from Uniform$(-2,2)$, and for the generation of the change points ζ_{ij}'s, we considered two situations, the normal distribution Normal$(0.2, 0.5^2)$ and the exponential distribution with the hazard of 0.8. For the latent variables b_i's, they were generated from the normal distribution Normal$(0, 0.5^2)$. Given the $\boldsymbol{X}_{ij}$'s, $\boldsymbol{Z}_{ij}$'s, U_{ij}'s, ζ_{ij}'s and b_i's, the failure times of interest T_{ij}'s were generated under model (1) with $\Lambda(t) = 0.5\,t^{1.5}$.

For the generation of the observed data, to mimic a follow-up study and for each subject, we first generated a sequence of examination times with the gap between consecutive examination times following the uniform distribution over $(0.1, 0.3)$ and the maximum observation time to be 3. Then for subject (ij), L_{ij} was taken to be the last observation time that is smaller than T_{ij} and R_{ij} the first observation time that is greater than T_{ij}. For the cluster size n_i, we considered two situations. One, Scenario I, is to set n_i to be fixed at either 10 or 20 for all clusters and the other, Scenario II, is to generate n_i from the discrete uniform distribution over the set $\{6, \ldots, 15\}$ or $\{16, \ldots, 25\}$. Note that both configurations yielded similar overall sample sizes with the average cluster size, denoted by $\widetilde{n}$, approximately equal to 10 or 20. The results reported below are based on $N = 50$ or 100 clusters and $m = 4$ with 500 replications.

[Tables 1 – 2 near here.]

Tables 1 – 2 present the estimation results for all parameters given by the proposed approach with $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (0.3, -0.5)^\top$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^\top = (-0.6, 0.8, 0.3)^\top$ and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)^\top = (1.2, -0.5, 0.2)^\top$ under the normal and exponential change points, respectively. They include the estimated bias (Bias) given by the average of the estimates minus the true value, the average of the estimated standard errors (SSE), the sample standard errors (SEE), and the 95% empirical coverage probability (CP). The results suggest that the proposed estimators seem to be unbiased and the estimated standard errors provide a reasonable approximation to the true variability. Furthermore, the empirical coverage probabilities suggest that the normal approximation to the distribution of the proposed estimates is appropriate, and as expected, the results got better when either the number of clusters or the sample size within each cluster increased. In addition, the results seem to be consistent with respect to the within-cluster sample sizes, which are either fixed or random.

[Table 3 near here.]

In practice, one question of interest may be the importance of taking into account the cluster structure. To see this, we repeated the study above with $\sigma = 0.8$ and under Scenario I and applied the method, referred to as the Naive method, that ignores the cluster structure or assumes $b_i = 0$ for all i in addition to the proposed method. The results are given in Table 3 and it is apparent that the Naive method not only can yield biased estimates but also may result in misleading inference.

[Table 4 near here.]

Following a reviewer's suggestion, we further investigated the robustness of the proposed method with respect to the distribution of the latent variable. For this, we repeated the study above that yielded Table 1 with $\widetilde{n} = 10$ and $N = 50$ under both scenarios except generating the latent variables from (A) a mixed distribution or (B) a Cox-type model. For the former, it was assumed that $\zeta_{ij} \sim \text{Normal}(0.2, 0.5^2)$ when $\zeta_{ij} \leq 0$ and ζ_{ij} follows an exponential distribution with hazard 0.2 when $\zeta_{ij} > 0$. For the latter, we generated the ζ_{ij}'s from the Cox-type model with the hazard function $\Lambda_\zeta(\zeta_{ij}) = 5\,\zeta_{ij}$. For the latent variable b_i's, under case (A), we generated them such that $b_i \sim \text{Normal}(-0.25, 0.5^2)$ when $b_i \leq 0$ and $b_i \sim \text{Normal}(0.25, 0.5^2)$ when $b_i > 0$. Under case (B), we first generated the b_i^*'s from the Cox-type model with the hazard function $\Lambda_b(b_i^*) = (b_i^*)^5$ and set $b_i = \log b_i^* + 0.1155$. Note that the mixed distribution in (A) is more complex than the simple parametric distribution used previously, while the Cox-type model in (B) represents a commonly used semiparametric specification. The results are presented in Table 4 and one can see that both the performance and the conclusions are similar to those given in Table 1. In other words, the proposed method seems to be reasonably robust to the misspecification of the latent-variable distribution.

5 An Application

Now we apply the sieve maximum likelihood estimation procedure proposed in the previous sections to the international breast cancer study described above with the focus on the distant metastasis-free survival. The distant metastasis, defined as cancer spread to distant organs or lymph nodes as mentioned above, plays a central role in disease progression and is essential for predicting tumor dissemination risk. Also it is a key prognostic indicator closely linked to mortality and approximately 30% of women diagnosed with early-stage breast cancer eventually develop distant metastasis (Redig and McAllister, 2013). Furthermore, since the patients were examined yearly, only interval-censored data are available on the distant metastasis-free survival. For breast cancer patients, among other possible influential factors, menopause represents a critical biological transition such as substantial hormonal and physiological changes (Dall and Britt, 2017; Heer et al., 2020). The main objective here is to investigate the effects of several influential or risk factors on the development of disease progression with particular interest in the role of menopausal transition.

For the risk factors, following Huang et al. (2024), we will consider three prognostic factors, estrogen receptor status (ER), tumor size (SIZE), and histological grade (GRADE). More specifically, the factor ER indicates the sensitivity to hormone therapy and takes 1 if the tumor is responsive to the therapy that suppresses estrogen receptor expression and 0 otherwise. The factor SIZE is a binary variable coded as 1 if the tumor is larger than 2 cm and 0 otherwise, and the factor GRADE reflects the tumor aggressiveness, categorized as 1 (low), 2 (intermediate), or 3 (high). To describe the effect of the factor GRADE more accurately, we use two binary variables, GRADE2 and GRADE3, to represent it. For the application of the proposed method, we set the covariate $\boldsymbol{Z}$ to be the factor ER, the covariate $\boldsymbol{X}$ including all of the other factors, and U to be the age at breast cancer diagnosis with the change point ζ representing the age at which major hormonal changes occur. In the following, we will focus on the data of 2,026 patients from 21 sub-studies after excluding the patients with missing values in the relevant factors by treating sub-studies as clusters. The original study includes 52 sub-studies.

[Table 5 near here.]

Table 5 presents the analysis results given by the proposed approach. They include the estimated effect (EST), the estimated standard error (SD), and the p-value (PVAL) for testing no effect for each of the risk factors. They suggest that all four risk factors seem to have had significant effects on the development of disease progression. More specifically, the analysis indicates that the patient with a tumor that is larger than 2 cm or the intermediate or high tumor aggressiveness had a relatively significantly higher risk of developing disease progression. Furthermore, the effect of the factor ER seems to have a significant change at around age 46. Before the change, it was positively related to the risk of developing disease progression but negatively associated with the risk after the change. A possible reason for this is that premenopausal women usually have higher levels of endogenous estrogen, which may stimulate tumor growth in ER-positive breast cancers, while after menopause, declining estrogen levels may mitigate the effect, potentially reducing disease progression in ER-positive patients.

For comparison, we also applied the Naive method considered in the previous section that ignores the cluster structure and include the obtained results in Table 5. One can see from the table that a major difference from the analysis above is that the Naive method did not identify the significant effects of the risk factors SIZE and GRADE2 possibly due to not taking into account the cluster structure or the association within the cluster. To further look at the results above, we performed a simple analysis that ignored the existence of both cluster structure and change point and give the analysis result again in Table 5. As expected, it suggests that the factor ER still had a significant effect and was negatively associated with the risk of developing disease progression. This is consistent with Huang et al. (2024), who fitted the data to a mean residual life model and did not consider the existence of the change point.

6 Discussion and Concluding Remarks

This paper considered regression analysis of clustered interval-censored failure time data in the presence of random change points. Although many methods have been developed for analyzing interval-censored data with either change points or cluster structures, the proposed approach is the first, to our knowledge, that can simultaneously accommodate or deal with both features. For the problem, we presented a change point Cox model and proposed a sieve estimation procedure that employs Bernstein polynomials to approximate the unknown function. In addition, an EM algorithm was developed and the asymptotic properties of the proposed estimators were established. A simulation study was conducted and showed that the proposed method performed well in practical settings. Finally we applied the method to a breast cancer study that motivated this investigation and provided some new insights.

It is worth noting that the proposed method adopts a Cox-type model that is known to have certain limitations such as the proportional hazards assumption and may not always provide an adequate fit to the data. To address these limitations, alternative models such as change point additive hazards or accelerated failure time models could be considered, and the inference procedures similar to those given above could be developed. Furthermore, while Bernstein polynomials were used to approximate unknown functions due to their natural fit and appealing mathematical properties, other smooth functions such as splines could also be employed.

Several further directions for research may be considered. One is that in the above, we have assumed that both covariates and their effects are time-independent but in practice time-dependent covariates or effects could exist. It would be useful to generalize the proposed method to accommodate the latter situation. Also the proposed method has assumed that the censoring mechanism or the process generating the censoring intervals is non-informative. As discussed in the literature (Du and Zhao, 2024; Lou et al., 2024), this may not hold sometimes and instead one may face informative interval censoring. It is apparent that it would be helpful to have methods to allow it but the generalization of the proposed method to accommodate it would not be easy. To address this, one could generalize the proposed method by employing a joint modeling framework. Specifically, one might define a shared frailty model where a common latent variable simultaneously affects both the hazard of the failure time and the intensity of the observation process. Alternatively, one could model the dependence between the failure time and the observation process using copula functions. While conceptually feasible, these extensions would significantly increase the computational complexity as they would require integrating over additional latent structures alongside the random change points.

Acknowledgments

The authors wish to thank the Co-Editor, Dr. John Stufken, an Associate Editor, and a reviewer for their many insightful and valuable comments and suggestions that greatly improved the article. The research was partially supported by the Scientific Research Project of the Education Department of Jilin Province (Grant No. JJKH20261483KJ to Du).

References

Chen, M.-H., Tong, X., and Sun, J. (2009). A frailty model approach for regression analysis of multivariate current status data. *Statistics in Medicine*, 28(27):3424–3436.

Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, 101(475):1228–1240.

Chen, X., Ping, Y., and Sun, J. (2024). Efficient estimation of cox model with random change point. *Statistics in Medicine*, 43(6):1213–1226.

Dall, G. V. and Britt, K. L. (2017). Estrogen effects on the mammary gland in early and late life and breast cancer risk. *Frontiers in Oncology*, 7:110.

Du, M., Lou, Y., and Sun, J. (2025). Estimation and variable selection for interval-censored failure time data with random change point and application to breast cancer study. *Journal of the American Statistical Association*, 0(0):1–12.

Du, M. and Zhao, X. (2024). A conditional approach for regression analysis of case k interval-censored failure time data with informative censoring. *Computational Statistics & Data Analysis*, 198:107991.

Hanagal, D. D. (2011). *Modeling Survival Data using Frailty Models*. Springer.

He, X., Lin, H., and Tu, D. (2018). A single-index threshold cox proportional hazard model for identifying a treatment-sensitive subset based on multiple biomarkers. *Statistics in Medicine*, 37:3267–3279.

Heer, E., Harper, A., Escandor, N., Sung, H., McCormack, V., and Fidler-Benaoudia, M. M. (2020). Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study. *The Lancet Global Health*, 8(8):e1027–e1037.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer.

Huang, R., Sun, L., and Xiang, L. (2024). Conditional quasi-likelihood inference for mean residual life regression with clustered failure time data. *Scandinavian Journal of Statistics*, 51(4):1685–1706.

Huang, R., Xiang, L., and Ha, I. D. (2019). Frailty proportional mean residual life regression for clustered survival data: A hierarchical quasi-likelihood method. *Statistics in Medicine*, 38(24):4854–4870.

Lam, K., Xu, Y., and Cheung, T.-L. (2010). A multiple imputation approach for clustered interval-censored survival data. *Statistics in Medicine*, 29(6):680–693.

Lee, C. Y. and Lam, K. (2020). Survival analysis with change-points in covariate effects. *Statistical Methods in Medical Research*, 29(11):3235–3248.

Lee, C. Y. and Wong, K. Y. (2023). Survival analysis with a random change-point. *Statistical Methods in Medical Research*, 32(11):2083–2095.

Lorentz, G. G. (1986). *Bernstein Polynomials*. New York: Chelsea Publishing Co.

Lou, Y., Du, M., and Song, X. (2025). Regression analysis of interval-censored failure time data with change points and a cured subgroup. *Biometrics*, 81(3):ujaf100.

Lou, Y., Sun, J., and Wang, P. (2024). Semiparametric cure regression models with informative case k interval-censored failure time data. *Statistica Sinica*.

Ma, L., Hu, T., and Sun, J. (2015). Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika*, 102(3):731–738.

Pons, O. (2003). Estimation in a cox regression model with a change-point according to a threshold in a covariate. *The Annals of Statistics*, 31(2):442–463.

Redig, A. J. and McAllister, S. S. (2013). Breast cancer as a systemic disease: a view of metastasis. *Journal of Internal Medicine*, 274(2):113–126.

Shen, X. (1997). On methods of sieves and penalization. *The Annals of Statistics*, 25(6):2555–2591.

Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *The Annals of Statistics*, 24(2):580–615.

Siegel, R. L., Giaquinto, A. N., and Jemal, A. (2024). Cancer statistics, 2024. *CA: A Cancer Journal for Clinicians*.

Stavraky, K. and Emmons, S. (1974). Breast cancer in premenopausal and postmenopausal women. *Journal of the National Cancer Institute*, 53(3):647–654.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer.

Sun, J. and Chen, D.-G. (2022). *Emerging Topics in Modeling Interval-Censored Survival Data*. Springer.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer Science & Business Media.

Yang, D., Du, M., and Sun, J. (2021). Semiparametric regression analysis of clustered interval-censored failure time data with a cured subgroup. *Statistics in Medicine*, 40(30):6918–6930.

Zhou, Q., Hu, T., and Sun, J. (2017). A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *Journal of the American Statistical Association*, 112(518):664–672.

Appendix: Proofs of Asymptotic Properties.

In this appendix, we will sketch the proofs of the asymptotic properties of $\widehat{\boldsymbol{\theta}}$ given in Section 3 using the empirical process theory, and for this, we will first define some more notation and give some regularity conditions. For any $\widehat{\boldsymbol{\theta}}^{(1)} = (\widehat{\boldsymbol{\psi}}^{(1)}, \widehat{\Lambda}^{(1)})$ and $\widehat{\boldsymbol{\theta}}^{(2)} = (\widehat{\boldsymbol{\psi}}^{(2)}, \widehat{\Lambda}^{(2)})$, define the distance

$$d(\widehat{\boldsymbol{\theta}}^{(1)}, \widehat{\boldsymbol{\theta}}^{(2)}) = \left\| \widehat{\boldsymbol{\psi}}^{(1)} - \widehat{\boldsymbol{\psi}}^{(2)} \right\| + \left\| \widehat{\Lambda}^{(1)} - \widehat{\Lambda}^{(2)} \right\|_2 ,$$

and the $L_2(\mathbb{P})$ norm $\|f\|_2 = (\int |f|^2 \, d\,\mathbb{P})^{1/2}$ for a function f with probability measure $\mathbb{P}$. Also let $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}_0, \Lambda_0)$ denote the true value of $\boldsymbol{\theta}$. Following the notation used in van der Vaart and Wellner (1996), throughout the appendix, we denote $\mathbb{P}f = \int_{\mathcal{X}} f(o) d\,\mathbb{P}(o)$ and $\mathbb{P}_n f = n^{-1} \sum_{k=1}^{n} f(O_k)$, which is the empirical process indexed by the function f evaluated as O_k, the sample points of individual k, with $\mathcal{X}$ denoting the sample space. Define $n = \sum_{i=1}^{N} n_i$. To avoid confusion, we add a subscript n to all the estimators in the following, for example, the estimators $\widehat{\boldsymbol{\theta}}_n$, $\widehat{\boldsymbol{\psi}}_n$ and $\widehat{\Lambda}_n$.

Let the letter C represent a constant, which does not necessarily represent the same value each time here and in the proof. For the asymptotic properties, we need the following regularity conditions.

Condition 1. *The domain of the covariates $\boldsymbol{X}$, $\boldsymbol{Z}$ and U is a bounded subset of $\mathbb{R}^{\widetilde{d}}$, where $\widetilde{d} = p_1 + p_2 + 1$. Also $\mathbb{E}(\boldsymbol{X}\boldsymbol{X}^\top)$ and $\mathbb{E}(\boldsymbol{Z}\boldsymbol{Z}^\top)$ are positive definite.*

Condition 2. *There exists a positive value e such that $\mathbb{P}(R_{ij} - L_{ij} \geq e \mid \boldsymbol{X}, \boldsymbol{Z}, U) = 1$ for subject j in ith cluster. Also the union of the support of $\{L_{ij}, R_{ij} : j = 1, \ldots, n_i \, ; \, i = 1, \ldots, N\}$ is contained in an interval $[t_l, t_r]$, where $0 < t_l < t_r < \infty$.*

Condition 3. *The parameter space for $\boldsymbol{\psi}$ is compact convex and the true parameter $\boldsymbol{\psi}_0$ is an interior point. In addition, the parameter space of Λ is a collection of nonnegative and increasing continuous functions, and the true function Λ_0 is continuously differentiable up to order r in $[t_l, t_r]$ and satisfies $C^{-1} < \Lambda_0(t_l) < \Lambda_0(t_r) < C$ for some positive constant C.*

Condition 4. *The cluster size N is bounded by a positive constant and is independent of both $\{T_{ij} : j = 1, \ldots, n_i \, , \, i = 1, \ldots, N\}$ and the latent variable b conditional on the covariates.*

Condition 5. *Conditional on U, the density distribution of ζ has a strictly bounded and positive density everywhere and has third-order bounded derivatives with respect to $\boldsymbol{\nu}$.*

Condition 6. *For the single observation data and any $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)} \in \boldsymbol{\Theta}$, if $\ell(\boldsymbol{\theta}^{(1)} ; \boldsymbol{O}) = \ell(\boldsymbol{\theta}^{(2)} ; \boldsymbol{O})$, then $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(2)}$, where $\ell(\boldsymbol{\theta} ; \boldsymbol{O}) = \log L(\boldsymbol{\theta} ; \boldsymbol{O})$ denotes the log-likelihood function given* *in Section 2. Also for every* $\boldsymbol{\theta}$ *in a neighborhood of* $\boldsymbol{\theta}_0$*, we have that* $\mathbb{P}\{\ell(\boldsymbol{\theta}\,;\boldsymbol{O}) - \ell(\boldsymbol{\theta}_0\,;\boldsymbol{O})\} \preceq -d^2(\boldsymbol{\theta},\boldsymbol{\theta}_0)$*, where* $\preceq$ *means 'smaller than, up to a constant'.*

Condition 7. *The matrix* $\boldsymbol{I}(\boldsymbol{\psi}_0)$ *defined in the proof of the asymptotic normality is nonsingular.*

Proof of the Consistency. The consistency of the proposed estimator $\widehat{\boldsymbol{\theta}}_n$ can be established by following similar arguments in Theorems 1 and 2 in Zhou et al. (2017). To this end, let $\ell(\boldsymbol{\theta})$ denote the log-likelihood function based on a single observation O, and $M(\boldsymbol{\theta}) = -\ell(\boldsymbol{\theta})$. Define $K_\epsilon = \{\boldsymbol{\theta} : d\,(\boldsymbol{\theta},\boldsymbol{\theta}_0) \geq \epsilon, \boldsymbol{\theta} \in \Theta_n\}$ for $\epsilon > 0$ and

$$\xi_{1n} = \sup_{\theta\in\Theta_n} |\mathbb{P}_n M(\boldsymbol{\theta}) - \mathbb{P}M(\boldsymbol{\theta})|\,, \quad \xi_{2n} = \mathbb{P}_n M(\boldsymbol{\theta}_0) - \mathbb{P}M(\boldsymbol{\theta}_0)\,.$$

Consider the class of functions $\mathcal{L}_n = \{H(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta_n\}$ and based on Lemma 1 of Zhou et al. (2017), we have that the covering number of $\mathcal{L}_n$ satisfies $N(\epsilon, \mathcal{L}_n, L_1(\mathbb{P}_n)) \leq CM_n^{(m+1)}\epsilon^{-\{d+(m+1)\}}$. Thus, one can show that $\inf_{K_\epsilon} \mathbb{P}M(\boldsymbol{\theta}) = \inf_{K_\epsilon}\{\mathbb{P}M(\boldsymbol{\theta}) - \mathbb{P}_n M(\boldsymbol{\theta}) + \mathbb{P}_n M(\boldsymbol{\theta})\} \leq \xi_{1n} + \inf_{K_\epsilon} \mathbb{P}_n M(\boldsymbol{\theta})$. If $\widehat{\boldsymbol{\theta}}_n \in K_\epsilon$, then $\inf_{K_\epsilon} \mathbb{P}_n M(\boldsymbol{\theta}) \leq \mathbb{P}_n M(\boldsymbol{\theta}_0) = \xi_{2n} + \mathbb{P}M(\boldsymbol{\theta}_0)$. Let $\delta_\epsilon = \inf_{K_\epsilon} \mathbb{P}M(\boldsymbol{\theta}) - \mathbb{P}M(\boldsymbol{\theta}_0)$, and one can prove $\delta_\epsilon > 0$. Thus we have $\inf_{K_\epsilon} \mathbb{P}M(\boldsymbol{\theta}) \leq \xi_{1n} + \xi_{2n} + \mathbb{P}M(\boldsymbol{\theta}_0) := \xi_n + \mathbb{P}M(\boldsymbol{\theta}_0)$ and $\xi_n \geq \delta_\epsilon$, which yields $\{\widehat{\boldsymbol{\theta}}_n \in K_\epsilon\} \subset \{\xi_n \geq \delta_\epsilon\}$. Based on Lemma 2 of Zhou et al. (2017) and the strong law of large numbers, we have that $\xi_{1n} \to 0$ and $\xi_{2n} \to 0$ almost surely. Thus,

$$\cup_{k=1}^{\infty} \cap_{n=k}^{\infty} \left\{\widehat{\boldsymbol{\theta}}_n \in K_\epsilon\right\} \subset \cup_{k=1}^{\infty} \cap_{n=k}^{\infty} \{\xi_n \geq \delta_\epsilon\}\,,$$

which implies that $d\,(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) \to 0$ a.s.

□

Proof of the Convergence Rate. Now we discuss the convergence rate of $\widehat{\boldsymbol{\theta}}_n$. First note by Theorem 1.6.2 of Lorentz (1986), if $m = o(n^v)$, there exist a Bernstein polynomial Λ_n such that $\|\Lambda_0 - \Lambda_n\|_\infty = O(n^{-rv/2})$. For any ϱ, define the class $\mathcal{F}_\varrho = \{\ell(\boldsymbol{\theta}_{0,n}) - \ell(\boldsymbol{\theta}_0) : \boldsymbol{\theta} \in \boldsymbol{\Theta}_n,\ d\,(\boldsymbol{\theta},\boldsymbol{\theta}_{0,n}) < \varrho\}$, where $\boldsymbol{\theta}_{0,n} = (\boldsymbol{\psi}_0,\ \Lambda_n)$. Thus, following the similar calculation in pp 597 of Shen and Wong (1994), one can establish that the entropy with bracketing

$$\log N_{[]}(\epsilon, \mathcal{F}_\varrho, \|\cdot\|_2) \leq CN \log(\varrho/\epsilon)$$

where $N = m+1$, $N_{[]}(\epsilon, \mathcal{F}_\varrho, \|\cdot\|_2)$ is the bracketing number of the class $\mathcal{F}_\varrho$. Moreover, we have that $\|\ell(\boldsymbol{\theta}_{0,n}) - \ell(\boldsymbol{\theta})\|_2^2 \leq C\varrho^2$ for any $\ell(\boldsymbol{\theta}_{0,n}) - \ell(\boldsymbol{\theta}) \in \mathcal{F}_\varrho$. Also, based on Conditions 1 − 6, $\mathcal{F}_\varrho$ is uniformly bounded, and by following Lemma 3.4.2 of van der Vaart and Wellner (1996), we have that

$$\mathbb{E}_{\mathbb{P}}\|n^{1/2}(\mathbb{P}_n - \mathbb{P})\|_{\mathcal{F}_\varrho} \leq C J_\varrho(\epsilon, \mathcal{F}_\varrho, \|\cdot\|_2)\left\{1 + \frac{J_\varrho(\epsilon, \mathcal{F}_\varrho, \|\cdot\|_2)}{\sqrt{n}\varrho^2}\right\},$$

where $J_\varrho(\epsilon, \mathcal{F}_\varrho, \|\cdot\|_2) \leq C N^{1/2}\varrho$. By setting $\phi_n(\varrho) = C(N^{1/2}\varrho + N/\sqrt{n})$, one can be easily verified that $\phi_n(\varrho)/\varrho$ decreases in ϱ and $r_n^2\phi_n(1/r_n) = r_n N^{1/2} + r_n^2 N/n^{1/2} < 2C\sqrt{n}$, where $r_n = N^{-1/2}n^{1/2} = O(n^{(1-v)/2})$ with $0 < v < 0.5$. Hence by Theorem 3.4.1 of van der Vaart and Wellner (1996), we have that $n^{(1-v)/2}\, d\,(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{0,n}) = O_p(1)$. Combined the above results, we can conclude that $d\,(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p(n^{-(1-v)/2} + n^{-rv/2})$. □

Proof of the Asymptotic Normality of $\widehat{\boldsymbol{\psi}}_n$. We denote V as the linear span of $\Theta - \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is the true value of parameter $\boldsymbol{\theta}$. Let $\ell(\boldsymbol{\theta}; O)$ be the log-likelihood for a sample of size one and $\tau_n = n^{-(1-v)/2} + n^{-rv/2}$. For any $\boldsymbol{\theta} \in \Theta$ with $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = O(\tau_n)$, define the first-order directional derivative of $\ell(\boldsymbol{\theta})$ at the direction $\boldsymbol{\varsigma} \in V$ as

$$\dot{\ell}(\boldsymbol{\theta})[\boldsymbol{\varsigma}] = \frac{\partial\, \ell(\boldsymbol{\theta} + k\boldsymbol{\varsigma})}{\partial\, k}\Big|_{k=0}$$

and the second-order directional derivative as

$$\ddot{\ell}(\boldsymbol{\theta})[\boldsymbol{\varsigma}, \widetilde{\boldsymbol{\varsigma}}] = \frac{\partial^2 \ell(\boldsymbol{\theta} + k\boldsymbol{\varsigma} + \widetilde{k}\widetilde{\boldsymbol{\varsigma}})}{\partial\, k\, \partial\, \widetilde{k}}\Big|_{\widetilde{k}=k=0}.$$

Also, define the Fisher inner product on the space V as $< \boldsymbol{\varsigma}, \widetilde{\boldsymbol{\varsigma}} >= \mathbb{P}\,\{\dot{\ell}(\boldsymbol{\theta})[\boldsymbol{\varsigma}]\, \dot{\ell}(\boldsymbol{\theta})[\widetilde{\boldsymbol{\varsigma}}]\}$ and the Fisher norm for $\boldsymbol{\varsigma} \in V$ as $\|\boldsymbol{\varsigma}\|^{1/2} =< \boldsymbol{\varsigma}, \boldsymbol{\varsigma} >$. Let $\overline{V}$ be the closed linear span of V under the Fisher norm. Then $(\overline{V}, \|\cdot\|)$ is a Hilbert space.

Define the smooth functional of $\boldsymbol{\theta}$ as $\varrho(\boldsymbol{\theta}) = \boldsymbol{u}^\top\boldsymbol{\psi}$, where $\boldsymbol{u}$ is any vector of q dimension with $\|\boldsymbol{u}\| \leq 1$. For any $\boldsymbol{\varsigma} \in V$, denote

$$\dot{\varrho}(\boldsymbol{\theta}_0)[\boldsymbol{\varsigma}] = \frac{\partial\, \varrho(\boldsymbol{\theta}_0 + k\boldsymbol{\varsigma})}{\partial\, k}\Big|_{k=0}.$$

Note that $\varrho(\boldsymbol{\theta}) - \varrho(\boldsymbol{\theta}_0) = \dot{\varrho}(\boldsymbol{\theta}_0)[\boldsymbol{\theta} - \boldsymbol{\theta}_0]$. It follows from the Riesz representation theorem that there exists $\boldsymbol{\varsigma}^* \in \overline{V}$ such that $\dot{\varrho}(\boldsymbol{\theta}_0)[\boldsymbol{\varsigma}] =< \boldsymbol{\varsigma}^*, \boldsymbol{\varsigma} >$ for all $\boldsymbol{\varsigma} \in \overline{V}$ and $\|\boldsymbol{\varsigma}^*\|^2 = \|\dot{\varrho}(\boldsymbol{\theta}_0)\|$. Thus it follows from the Cramér-Wold device and the formula $\boldsymbol{u}^\top(\widehat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) = \varrho(\widehat{\boldsymbol{\theta}}_n) - \varrho(\boldsymbol{\theta}_0) = \dot{\varrho}(\boldsymbol{\theta}_0)[\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0] =< \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0,\, \boldsymbol{\varsigma}^* >$, also using similar arguments of Theorem 1 of Shen (1997), we have that $\sqrt{n} < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0,\, \boldsymbol{\varsigma}^* > \to N(\mathbf{0},\, \boldsymbol{u}^\top\boldsymbol{\Sigma}\boldsymbol{u})$ in distribution. In the following, we will first prove that $\sqrt{n} < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0,\, \boldsymbol{\varsigma}^* > \to N(\mathbf{0},\, \|\boldsymbol{\varsigma}^*\|^2)$ in distribution and then $\|\boldsymbol{\varsigma}^*\|^2 = \boldsymbol{u}^\top\boldsymbol{\Sigma}\boldsymbol{u}$.

To prove the former result, first note that for any $\boldsymbol{\varsigma}^* \in \boldsymbol{\Theta}_0$, by Theorem 1.6.2 of Lorentz (1986), there exists $\Pi_n\boldsymbol{\varsigma}^* \in \boldsymbol{\Theta}_n$ such that $\|\Pi_n\boldsymbol{\varsigma}^* - \boldsymbol{\varsigma}^*\| = o(1)$ and $\delta_n\|\Pi_n\boldsymbol{\varsigma}^* - \boldsymbol{\varsigma}^*\| = o(n^{-1/2})$. Let ϵ_n be any positive sequence satisfying $\epsilon_n = o(n^{-1/2})$. Also define $\wp[\boldsymbol{\theta}-\boldsymbol{\theta}_0] = \ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}_0) - \dot{\ell}(\boldsymbol{\theta})[\boldsymbol{\theta}-\boldsymbol{\theta}_0]$. By definition of $\boldsymbol{\theta}$, we have

$$\begin{aligned}
0 &\le \mathbb{P}_n\Big\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_n \pm \epsilon_n \Pi_n \boldsymbol{\varsigma}^*)\Big\} \\
&= \Big(\mathbb{P}_n - \mathbb{P}\Big)\Big\{\ell(\widehat{\boldsymbol{\theta}}_n) - \ell(\widehat{\boldsymbol{\theta}}_n \pm \epsilon_n \Pi_n \boldsymbol{\varsigma}^*)\Big\} + \mathbb{P}\Big\{\ell(\widehat{\boldsymbol{\theta}}_n) - \ell(\widehat{\boldsymbol{\theta}}_n \pm \epsilon_n \Pi_n \boldsymbol{\varsigma}^*)\Big\} \\
&= \mp\, \epsilon_n \mathbb{P}_n \dot{\ell}(\boldsymbol{\theta}_0)[\boldsymbol{\varsigma}^*] \mp \epsilon_n \mathbb{P}_n \dot{\ell}(\boldsymbol{\theta}_0)[\Pi_n \boldsymbol{\varsigma}^* - \boldsymbol{\varsigma}^*] + \mathbb{P}\Big\{\wp[\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0] - \wp[\widehat{\boldsymbol{\theta}}_n \pm \epsilon_n \Pi_n \boldsymbol{\varsigma}^* - \boldsymbol{\theta}_0]\Big\} \\
&\quad + \Big(\mathbb{P}_n - \mathbb{P}\Big)\Big\{\wp[\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0] - \wp[\widehat{\boldsymbol{\theta}}_n \pm \epsilon_n \Pi_n \boldsymbol{\varsigma}^* - \boldsymbol{\theta}_0]\Big\} \\
&= \mp\, \epsilon_n \mathbb{P}_n \dot{\ell}(\boldsymbol{\theta}_0)[\boldsymbol{\varsigma}^*] \mp I_1 + I_2 + I_3\,.
\end{aligned}$$

Following the same arguments as on page 9 of the supplementary materials of Zhou et al. (2017), we obtain that

$$I_1 = \epsilon_n o_p(n^{-1/2})\,, \qquad I_2 = \epsilon_n o_p(n^{-1/2})\,,$$

and

$$I_3 = \pm \epsilon_n < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0,\, \boldsymbol{\varsigma}^* > + \epsilon_n o_p(n^{-1/2})\,.$$

Together with $\mathbb{P}\,\dot{\ell}(\boldsymbol{\theta}_0)[\boldsymbol{\varsigma}^*] = 0$, we can obtain that

$$\begin{aligned}
0 &\le \mathbb{P}_n\Big\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}} \pm \epsilon_n \Pi_n \boldsymbol{\varsigma}^*)\Big\} \\
&= \mp \epsilon_n \Big(\mathbb{P}_n - \mathbb{P}\Big)\Big\{\dot{\ell}(\boldsymbol{\theta}_0)[\boldsymbol{\varsigma}^*]\Big\} \pm \epsilon_n < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \boldsymbol{\varsigma}^* > + \epsilon_n o_p(n^{-1/2})\,.
\end{aligned}$$

Therefore, we obtain that $\sqrt{n} < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \boldsymbol{\varsigma}^* >= \sqrt{n}\Big(\mathbb{P}_n - \mathbb{P}\Big)\Big\{\dot{\ell}(\boldsymbol{\theta}_0)[\boldsymbol{\varsigma}^*]\Big\} + o_p(1) \to N(\mathbf{0}, \|\boldsymbol{\varsigma}^*\|^2)$, where the asymptotic normality is guaranteed by the central limits theorem and the asymptotic variance being equal to $\|\boldsymbol{\varsigma}^*\|^2 = \|\dot{\ell}(\boldsymbol{\theta}_0)[\boldsymbol{\varsigma}^*]\|^2$.

Now we will prove that $\|\boldsymbol{\varsigma}^*\|^2 = \boldsymbol{u}^\top \boldsymbol{\Sigma} \boldsymbol{u}$. For each component ψ_j of $\boldsymbol{\psi}$, we denote $\boldsymbol{u}_j^*$ and $\widetilde{\boldsymbol{u}}_j^*$ as the solution to $\inf_{\boldsymbol{u}} \mathbb{E}\,\{\ell_{\boldsymbol{\psi}} \boldsymbol{e}_j - \ell_{\boldsymbol{u}_j}[\boldsymbol{u}_j]\}^2$, where $\ell_{\boldsymbol{\psi}}$ is the partial derivative of the log-likelihood about $\boldsymbol{\psi}$, $\boldsymbol{e}_j$ is a q dimensional vector of zeros except the j-th element equal to 1, and $\ell_{\boldsymbol{u}_j}[\boldsymbol{u}_j]$ is the directional derivatives with respect to Λ, and can be calculated as directional derivative defined above. Following the calculations of Chen et al. (2006), we have that

$$\|\boldsymbol{\varsigma}^*\|^2 = \|\dot{\varrho}(\boldsymbol{\theta}_0)\| = \sup_{\varsigma \in \overline{V}:\, \|\varsigma\| > 0} \frac{|\dot{\varrho}(\boldsymbol{\theta}_0)[\varsigma]|}{\|\varsigma\|} = \boldsymbol{u}^\top \boldsymbol{\Sigma} \boldsymbol{u}\,,$$

and $\sqrt{n}(\widehat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) \to N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = [\mathbb{E}\,(S_{\boldsymbol{\psi}} S_{\boldsymbol{\psi}})^\top]^{-1} = [\boldsymbol{I}(\boldsymbol{\psi}_0)]^{-1}$, $S_{\boldsymbol{\psi}} = \ell_{\boldsymbol{\psi}} - \ell_{\boldsymbol{u}^*}[\boldsymbol{u}^*]$ and $\boldsymbol{I}(\boldsymbol{\psi}_0)$ is the Fisher information for $\boldsymbol{\psi}$. This completes the proof.

□

Table 1: Simulation results with the normal change point.

		Scenario I				Scenario II			
$(N, \tilde{n})$	Parameters	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
(50, 10)	$\beta_1 = 0.3$	0.0070	0.0921	0.1098	0.978	0.0079	0.0979	0.1076	0.958
	$\beta_2 = -0.5$	-0.0133	0.0621	0.0651	0.948	-0.0142	0.0593	0.0638	0.954
	$\alpha_1 = -0.6$	-0.0191	0.1018	0.0978	0.932	-0.0245	0.0973	0.0962	0.942
	$\alpha_2 = 0.8$	0.0126	0.1468	0.1345	0.930	0.0223	0.1366	0.1317	0.944
	$\alpha_3 = 0.3$	0.0099	0.0703	0.0774	0.966	0.0058	0.0701	0.0755	0.962
	$\gamma_1 = 1.2$	0.0287	0.1482	0.1380	0.918	0.0494	0.1434	0.1371	0.916
	$\gamma_2 = -0.5$	-0.0074	0.1730	0.1893	0.962	-0.0203	0.1704	0.1853	0.956
	$\gamma_3 = 0.2$	-0.0056	0.1051	0.1214	0.968	0.0009	0.1128	0.1193	0.970
	$\nu_\mu = 0.2$	-0.0193	0.1514	0.1686	0.938	-0.0061	0.1545	0.1537	0.916
	$\nu_\sigma = 0.5$	-0.0444	0.1882	0.2221	0.928	-0.0256	0.1859	0.2095	0.934
	$\sigma = 0.5$	0.0016	0.0857	0.0930	0.976	-0.0065	0.0792	0.0924	0.974
(50, 20)	$\beta_1 = 0.3$	0.0000	0.0668	0.0742	0.968	0.0063	0.0658	0.0738	0.968
	$\beta_2 = -0.5$	-0.0027	0.0410	0.0440	0.960	-0.0063	0.0420	0.0427	0.952
	$\alpha_1 = -0.6$	-0.0075	0.0645	0.0664	0.954	-0.0104	0.0660	0.0662	0.936
	$\alpha_2 = 0.8$	0.0159	0.0927	0.0960	0.950	0.0048	0.0984	0.0953	0.932
	$\alpha_3 = 0.3$	0.0061	0.0420	0.0529	0.982	0.0052	0.0471	0.0526	0.972
	$\gamma_1 = 1.2$	0.0189	0.0966	0.0951	0.934	0.0187	0.0979	0.0936	0.922
	$\gamma_2 = -0.5$	-0.0106	0.1181	0.1293	0.960	-0.0039	0.1107	0.1272	0.976
	$\gamma_3 = 0.2$	-0.0017	0.0707	0.0835	0.976	-0.0025	0.0745	0.0831	0.968
	$\nu_\mu = 0.2$	-0.0007	0.1035	0.1117	0.974	-0.0107	0.1045	0.1099	0.942
	$\nu_\sigma = 0.5$	-0.0171	0.1204	0.1332	0.960	-0.0174	0.1227	0.1322	0.960
	$\sigma = 0.5$	-0.0025	0.0654	0.0726	0.956	-0.0048	0.0731	0.0720	0.924
(100, 10)	$\beta_1 = 0.3$	0.0034	0.0657	0.0709	0.974	-0.0010	0.0653	0.0695	0.956
	$\beta_2 = -0.5$	-0.0022	0.0434	0.0416	0.930	-0.0055	0.0398	0.0411	0.952
	$\alpha_1 = -0.6$	-0.0107	0.0665	0.0637	0.944	-0.0097	0.0644	0.0621	0.932
	$\alpha_2 = 0.8$	0.0005	0.0938	0.0871	0.940	0.0059	0.0966	0.0859	0.918
	$\alpha_3 = 0.3$	0.0027	0.0463	0.0505	0.974	0.0025	0.0478	0.0492	0.962
	$\gamma_1 = 1.2$	0.0123	0.1016	0.0907	0.920	0.0168	0.0946	0.0880	0.924
	$\gamma_2 = -0.5$	-0.0033	0.1169	0.1230	0.958	-0.0048	0.1152	0.1207	0.966
	$\gamma_3 = 0.2$	0.0001	0.0723	0.0788	0.974	0.0042	0.0745	0.0775	0.968
	$\nu_\mu = 0.2$	-0.0068	0.1040	0.1064	0.942	-0.0077	0.0998	0.1046	0.942
	$\nu_\sigma = 0.5$	-0.0141	0.1218	0.1300	0.954	-0.0042	0.1150	0.1207	0.940
	$\sigma = 0.5$	-0.0084	0.0599	0.0599	0.948	-0.0021	0.0564	0.0592	0.958

Table 2: Simulation results with the exponential change point.

		Scenario I				Scenario II			
$(N, \widetilde{n})$	Parameters	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
(50, 10)	$\beta_1 = 0.3$	0.0059	0.0943	0.1089	0.982	0.0046	0.0970	0.1062	0.952
	$\beta_2 = -0.5$	-0.0102	0.0628	0.0641	0.948	-0.0082	0.0626	0.0630	0.950
	$\alpha_1 = -0.6$	-0.0158	0.0789	0.0818	0.952	-0.0090	0.0777	0.0792	0.946
	$\alpha_2 = 0.8$	0.0078	0.1309	0.1184	0.930	0.0188	0.1345	0.1173	0.918
	$\alpha_3 = 0.3$	0.0072	0.0617	0.0677	0.962	0.0060	0.0624	0.0670	0.952
	$\gamma_1 = 1.2$	0.0348	0.1755	0.1627	0.910	0.0276	0.1781	0.1610	0.916
	$\gamma_2 = -0.5$	-0.0015	0.1964	0.2148	0.956	-0.0065	0.2007	0.2146	0.954
	$\gamma_3 = 0.2$	0.0036	0.1215	0.1403	0.976	0.0134	0.1291	0.1391	0.954
	$\nu_\mu = 0.8$	-0.0024	0.2565	0.2634	0.948	0.0225	0.2535	0.2675	0.946
	$\sigma = 0.5$	-0.0040	0.0801	0.0914	0.972	-0.0057	0.0838	0.0916	0.966
(50, 20)	$\beta_1 = 0.3$	0.0051	0.0679	0.0740	0.964	0.0053	0.0665	0.0732	0.964
	$\beta_2 = -0.5$	-0.0043	0.0416	0.0438	0.960	-0.0041	0.0422	0.0437	0.950
	$\alpha_1 = -0.6$	-0.0041	0.0529	0.0554	0.976	-0.0060	0.0531	0.0553	0.954
	$\alpha_2 = 0.8$	0.0073	0.0867	0.0861	0.946	0.0048	0.0911	0.0849	0.924
	$\alpha_3 = 0.3$	0.0040	0.0409	0.0465	0.972	0.0056	0.0420	0.0459	0.968
	$\gamma_1 = 1.2$	0.0175	0.1149	0.1096	0.928	0.0111	0.1094	0.1088	0.940
	$\gamma_2 = -0.5$	0.0047	0.1365	0.1463	0.960	-0.0011	0.1243	0.1439	0.978
	$\gamma_3 = 0.2$	0.0024	0.0816	0.0959	0.970	-0.0040	0.0827	0.0950	0.972
	$\nu_\mu = 0.8$	-0.0044	0.1519	0.1748	0.976	0.0055	0.1635	0.1738	0.966
	$\sigma = 0.5$	-0.0058	0.0670	0.0722	0.952	-0.0089	0.0724	0.0709	0.926
(100, 10)	$\beta_1 = 0.3$	0.0016	0.0642	0.0707	0.986	0.0051	0.0668	0.0693	0.952
	$\beta_2 = -0.5$	-0.0001	0.0539	0.0416	0.922	-0.0054	0.0399	0.0410	0.948
	$\alpha_1 = -0.6$	-0.0053	0.0656	0.0530	0.950	-0.0046	0.0537	0.0518	0.946
	$\alpha_2 = 0.8$	-0.0082	0.0973	0.0778	0.922	0.0029	0.0891	0.0771	0.918
	$\alpha_3 = 0.3$	-0.0009	0.0444	0.0446	0.964	0.0031	0.0416	0.0434	0.968
	$\gamma_1 = 1.2$	0.0040	0.1349	0.1067	0.932	0.0112	0.1222	0.1033	0.920
	$\gamma_2 = -0.5$	0.0002	0.1339	0.1404	0.964	-0.0038	0.1336	0.1371	0.966
	$\gamma_3 = 0.2$	0.0060	0.0871	0.0910	0.956	0.0046	0.0890	0.0896	0.962
	$\nu_\mu = 0.8$	0.0052	0.1651	0.1690	0.958	0.0068	0.1637	0.1649	0.952
	$\sigma = 0.5$	-0.0130	0.0669	0.0595	0.940	-0.0063	0.0550	0.0594	0.964

Table 3: Simulation results on the comparison with the naive method.

			Proposed method				Naive method			
Distribution	$(N, \widetilde{n})$	Parameters	Bias	SSE	SEE	CP	Bias	SSE	SEE	CP
Normal	(50, 10)	$\beta_1 = 0.3$	0.0092	0.1087	0.1127	0.948	-0.0646	0.0951	0.0901	0.874
		$\beta_2 = -0.5$	-0.0064	0.0644	0.0662	0.936	0.1006	0.0623	0.0546	0.514
		$\alpha_1 = -0.6$	-0.0119	0.1055	0.1004	0.936	0.1110	0.1090	0.0908	0.664
		$\alpha_2 = 0.8$	0.0152	0.1592	0.1422	0.928	-0.1756	0.1447	0.1339	0.712
		$\alpha_3 = 0.3$	0.0045	0.0706	0.0804	0.970	-0.0651	0.0729	0.0665	0.808
		$\gamma_1 = 1.2$	0.0341	0.1577	0.1430	0.912	-0.2105	0.1605	0.1346	0.544
		$\gamma_2 = -0.5$	-0.0112	0.1896	0.1958	0.962	0.1005	0.1777	0.1676	0.888
		$\gamma_3 = 0.2$	0.0094	0.1114	0.1260	0.976	-0.0349	0.1198	0.1073	0.912
		$\nu_\mu = 0.2$	-0.0028	0.1680	0.1693	0.910	-0.0039	0.2306	0.1889	0.846
		$\nu_\sigma = 0.5$	-0.0196	0.1871	0.2280	0.940	0.0049	0.2816	7.4068	0.828
	(100, 10)	$\beta_1 = 0.3$	0.0042	0.0709	0.0726	0.958	-0.0680	0.0653	0.0629	0.796
		$\beta_2 = -0.5$	-0.0087	0.0485	0.0430	0.950	0.1086	0.0405	0.0381	0.194
		$\alpha_1 = -0.6$	-0.0085	0.0731	0.0651	0.932	0.1271	0.0688	0.0627	0.466
		$\alpha_2 = 0.8$	0.0102	0.1125	0.0928	0.908	-0.1790	0.1000	0.0827	0.500
		$\alpha_3 = 0.3$	0.0180	0.0510	0.0511	0.956	-0.0726	0.0477	0.0455	0.598
		$\gamma_1 = 1.2$	-0.0051	0.1168	0.0930	0.908	-0.2409	0.1088	0.0943	0.334
		$\gamma_2 = -0.5$	0.0073	0.1295	0.1267	0.952	0.1082	0.1258	0.1152	0.800
		$\gamma_3 = 0.2$	0.0031	0.0777	0.0816	0.958	-0.0328	0.0796	0.0741	0.910
		$\nu_\mu = 0.2$	-0.0086	0.1029	0.1073	0.936	0.0179	0.1447	0.1258	0.900
		$\nu_\sigma = 0.5$	0.0058	0.1272	0.1292	0.942	0.0153	0.1921	0.4246	0.860
Exponential	(50, 10)	$\beta_1 = 0.3$	0.0022	0.1093	0.1119	0.946	-0.0705	0.1045	0.0903	0.852
		$\beta_2 = -0.5$	-0.0033	0.0712	0.0654	0.944	0.1071	0.0588	0.0544	0.506
		$\alpha_1 = -0.6$	-0.0037	0.0877	0.0831	0.944	0.1418	0.0778	0.0695	0.474
		$\alpha_2 = 0.8$	0.0079	0.1488	0.1275	0.910	-0.1874	0.1297	0.1207	0.656
		$\alpha_3 = 0.3$	0.0049	0.0677	0.0698	0.960	-0.0676	0.0646	0.0570	0.746
		$\gamma_1 = 1.2$	0.0269	0.1976	0.1700	0.908	-0.2072	0.1924	0.1629	0.690
		$\gamma_2 = -0.5$	0.0042	0.2130	0.2243	0.954	0.1345	0.2155	0.1962	0.852
		$\gamma_3 = 0.2$	0.0044	0.1326	0.1447	0.964	-0.0247	0.1379	0.1289	0.934
		$\nu_\mu = 0.8$	0.0179	0.2640	0.2759	0.952	0.1110	0.3567	0.2915	0.868
	(100, 10)	$\beta_1 = 0.3$	0.0031	0.0718	0.0725	0.958	-0.0690	0.0630	0.0632	0.796
		$\beta_2 = -0.5$	-0.0031	0.0549	0.0426	0.932	0.1102	0.0409	0.0380	0.200
		$\alpha_1 = -0.6$	-0.0083	0.0681	0.0543	0.940	0.1340	0.0549	0.0493	0.252
		$\alpha_2 = 0.8$	-0.0035	0.1046	0.0841	0.910	-0.1923	0.0824	0.0848	0.384
		$\alpha_3 = 0.3$	0.0007	0.0461	0.0457	0.950	-0.0708	0.0404	0.0401	0.564
		$\gamma_1 = 1.2$	0.0131	0.1426	0.1093	0.924	-0.2333	0.1265	0.1141	0.456
		$\gamma_2 = -0.5$	-0.0014	0.1433	0.1451	0.952	0.1233	0.1423	0.1328	0.812
		$\gamma_3 = 0.2$	0.0051	0.0908	0.0937	0.954	-0.0258	0.0945	0.0881	0.922
		$\nu_\mu = 0.8$	-0.0001	0.1681	0.1719	0.944	0.0626	0.2398	0.2057	0.900

Table 4: Simulation results with misspecified distributions.

Violation	Case	Parameters	Scenario I Bias	Scenario I SSE	Scenario I SEE	Scenario I CP	Scenario II Bias	Scenario II SSE	Scenario II SEE	Scenario II CP
ζ_{ij}'s	(A)	$\beta_1 = 0.3$	0.0014	0.0954	0.1073	0.970	0.0031	0.0954	0.1057	0.962
		$\beta_2 = -0.5$	-0.0081	0.0572	0.0638	0.970	-0.0016	0.0586	0.0624	0.962
		$\alpha_1 = -0.6$	-0.0178	0.0970	0.0997	0.952	-0.0154	0.0987	0.0988	0.944
		$\alpha_2 = 0.8$	0.0183	0.1493	0.1392	0.918	0.0211	0.1404	0.1361	0.932
		$\alpha_3 = 0.3$	0.0077	0.0744	0.0811	0.970	0.0017	0.0724	0.0794	0.958
		$\gamma_1 = 1.2$	0.0352	0.1345	0.1321	0.940	0.0259	0.1276	0.1310	0.954
		$\gamma_2 = -0.5$	-0.0117	0.1669	0.1805	0.960	-0.0088	0.1602	0.1770	0.954
		$\gamma_3 = 0.2$	0.0032	0.1031	0.1162	0.974	0.0123	0.1024	0.1136	0.954
		$\sigma = 0.5$	-0.0072	0.0829	0.0919	0.962	-0.0156	0.0783	0.0924	0.980
	(B)	$\beta_1 = 0.3$	0.0044	0.0902	0.1080	0.980	0.0062	0.0957	0.1045	0.966
		$\beta_2 = -0.5$	-0.0115	0.0595	0.0636	0.964	-0.0108	0.0601	0.0624	0.958
		$\alpha_1 = -0.6$	-0.0235	0.0873	0.0892	0.946	-0.0212	0.0817	0.0873	0.958
		$\alpha_2 = 0.8$	0.0150	0.1415	0.1274	0.928	0.0184	0.1320	0.1254	0.928
		$\alpha_3 = 0.3$	0.0068	0.0644	0.0739	0.972	0.0064	0.0680	0.0725	0.960
		$\gamma_1 = 1.2$	0.0355	0.1220	0.1276	0.944	0.0309	0.1229	0.1260	0.940
		$\gamma_2 = -0.5$	-0.0118	0.1581	0.1743	0.950	-0.0093	0.1598	0.1706	0.952
		$\gamma_3 = 0.2$	-0.0028	0.0946	0.1118	0.974	-0.0001	0.0991	0.1088	0.952
		$\sigma = 0.5$	-0.0026	0.0799	0.0908	0.970	-0.0044	0.0831	0.0906	0.974
b_i's	(A)	$\beta_1 = 0.3$	0.0061	0.0975	0.1116	0.962	0.0124	0.0942	0.1077	0.970
		$\beta_2 = -0.5$	-0.0075	0.0625	0.0656	0.952	-0.0099	0.0608	0.0647	0.954
		$\alpha_1 = -0.6$	-0.0232	0.0984	0.1003	0.946	-0.0172	0.0947	0.0974	0.960
		$\alpha_2 = 0.8$	0.0164	0.1553	0.1408	0.924	0.0134	0.1438	0.1363	0.930
		$\alpha_3 = 0.3$	0.0049	0.0730	0.0796	0.966	-0.0013	0.0683	0.0774	0.978
		$\gamma_1 = 1.2$	0.0326	0.1480	0.1408	0.922	0.0310	0.1441	0.1395	0.928
		$\gamma_2 = -0.5$	-0.0108	0.1803	0.1925	0.952	-0.0077	0.1744	0.1882	0.956
		$\gamma_3 = 0.2$	0.0092	0.1126	0.1244	0.970	0.0122	0.1080	0.1216	0.958
		$\nu_\mu = 0.2$	-0.0236	0.1584	0.1585	0.928	-0.0127	0.1640	0.1609	0.910
		$\nu_\sigma = 0.5$	-0.0342	0.2019	0.2269	0.924	-0.0373	0.1956	0.2356	0.940
	(B)	$\beta_1 = 0.3$	0.0095	0.0930	0.1066	0.968	0.0056	0.0879	0.1041	0.978
		$\beta_2 = -0.5$	-0.0159	0.0609	0.0627	0.952	-0.0136	0.0598	0.0622	0.952
		$\alpha_1 = -0.6$	-0.0265	0.0976	0.0953	0.942	-0.0206	0.0948	0.0924	0.936
		$\alpha_2 = 0.8$	0.0261	0.1405	0.1250	0.920	0.0150	0.1404	0.1227	0.912
		$\alpha_3 = 0.3$	0.0133	0.0685	0.0753	0.960	0.0069	0.0692	0.0740	0.952
		$\gamma_1 = 1.2$	0.0463	0.1402	0.1341	0.934	0.0398	0.1417	0.1327	0.924
		$\gamma_2 = -0.5$	-0.0164	0.1622	0.1824	0.960	-0.0110	0.1668	0.1809	0.964
		$\gamma_3 = 0.2$	-0.0002	0.1046	0.1180	0.968	0.0051	0.1039	0.1155	0.956
		$\nu_\mu = 0.2$	-0.0171	0.1486	0.1609	0.932	-0.0024	0.1537	0.1497	0.920
		$\nu_\sigma = 0.5$	-0.0321	0.1755	0.2021	0.928	-0.0312	0.1594	0.1827	0.954

Table 5: Analysis results of the breast cancer study.

	Proposed method			Naive method			Simple method		
FACTOR	EST	SD	PVAL	EST	SD	PVAL	EST	SD	PVAL
SIZE	0.2802	0.1226	0.0223	0.1635	0.0920	0.0754	0.0629	0.0846	0.4670
GRADE2	0.3398	0.1398	0.0151	0.2000	0.1251	0.1098	0.1739	0.1154	0.1319
GRADE3	0.7113	0.1363	0.0000	0.6531	0.1234	0.0000	0.6133	0.1137	0.0000
ER	1.4104	0.2517	0.0000	1.4207	0.1986	0.0000	-0.6374	0.1013	0.0000
ER (γ)	-2.8345	0.2438	0.0000	-2.9324	0.1755	0.0000	—	—	—
ν_μ	46.329	2.5789	0.0000	47.401	1.5416	0.0000	—	—	—
ν_σ	7.0864	0.3237	0.0000	6.9009	0.1180	0.0000	—	—	—
σ	0.4483	0.1075	0.0000	—	—	—	—	—	—