Biometrical Journal **52** (2023) 61, zzz–zzz / DOI: 10.1002/bimj.200100000

A semiparametric two-sample density ratio model with a change point

Jiahui Feng [1], **Kin Yau Wong** [2,3], and **Chun Yin Lee** [*,2]

[1] Department of Statistics and Actuarial Science, Simon Fraser University, Canada

[2] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

[3] Hong Kong Polytechnic University Shenzhen Research Institute

Received zzz, revised zzz, accepted zzz

The logistic regression model for a binary outcome with a continuous covariate can be expressed equivalently as a two-sample density ratio model for the covariate. Utilizing this equivalence, we study a change-point logistic regression model within the corresponding density ratio modeling framework. We investigate estimation and inference methods for the density ratio model and develop maximal score-type tests to detect the presence of a change point. In contrast to existing work, the density ratio modeling framework facilitates the development of a natural Kolmogorov–Smirnov type test to assess the validity of the logistic model assumptions. A simulation study is conducted to evaluate the finite-sample performance of the proposed tests and estimation methods. We illustrate the proposed approach using a mother-to-child HIV-1 transmission dataset and an oral cancer dataset.

Key words: Biased sampling; Empirical likelihood; Goodness-of-fit; Logistic regression model; Score test.

1 Introduction

The logistic regression model is widely used in case-control studies, where subjects are sampled separately from both the event and non-event groups. With a single continuous covariate, the logistic regression model can be reformulated as a two-sample semiparametric density ratio model. Here, the density of the covariate in the non-event group is modeled nonparametrically, while the density in the event group is linked to the baseline density through an exponential tilt. In the literature, the density ratio model can be viewed as a special case of a biased sampling model (Vardi, 1982, 1985), particularly when the target population is not directly observable. In such instances, the exponential tilt can be manipulated as a weight function indexed by some unknown parameters.

In practice, the logistic regression model is often employed without a thorough verification of its underlying assumptions. A merit of the density ratio model formulation is that it leads to a natural approach to testing the logistic model assumptions. Specifically, it allows for the examination of whether the density ratio adheres to the exponential tilt structure, without presupposing any specific relationships between the covariate densities of the two groups. Qin and Zhang (1997) explored a Kolmogorov–Smirnov test for the logistic model assumptions, where the test measures the discrepancy between the estimated distribution function of the covariate under the density ratio model and the empirical distribution function. They proposed a bootstrap method to estimate the null distribution of the test statistic. To circumvent the demanding computational requirements of the bootstrap approach, Zhang (1999) introduced a goodness-of-fit test based on the Nikulin–Rao–Robson–Moore statistic, and Zhang (2001) developed an information matrix test based on White (1982) to detect model misspecifications; both approaches yield χ^2-distributed

*Corresponding author: Chun Yin Lee, Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, email: james-chun-yin.lee@polyu.edu.hk

test statistics under the null hypothesis. Bondell (2007) proposed a test statistic assessing differences between the characteristic functions estimated from non-event data and full data, where the densities in the characteristic functions are estimated via the kernel method.

The change-point model offers a valuable alternative to traditional models when a significant shift occurs in the distribution of the variable of interest. It effectively describes a nonlinear relationship between the response and explanatory variables. In the literature, the change-point is also known as a threshold or break-point (Molinari et al., 2001; Muggeo, 2003). This paper focuses on the density ratio change-point model, where a covariate impacts the outcome variable only if its value surpasses a certain unknown threshold. Such a model has garnered significant interest in medical research. For example, in mother-to-child HIV-1 transmission studies, a change point is typically assumed for the influence of an immune response biomarker, conferring a protective effect only when the level of neutralizing antibodies exceeds a specific threshold (Permar et al., 2015). In this context, Fong et al. (2015) introduced maximum score-type and likelihood-ratio test statistics to detect the presence of a change point under the logistic regression model. Lee (2021) suggested using the maximal change in the area under the receiver operating characteristic curve as an alternative test statistic. Additionally, Zhao et al. (2014) analyzed data from the Strong Heart Family Study, finding that subjects in the lowest quartile of leukocyte telomere length have a significantly higher diabetes incidence risk compared with those in the highest quartile. However, this risk ratio appears insignificant for subjects with leukocyte telomere lengths in the second and third quartiles, indicating that a harmful effect may only be triggered when the leukocyte telomere length falls below a certain level. Reliable detection and precise estimation of change-point effects are crucial as they enhance clinicians' understanding of disease mechanisms and aid in identifying high-risk individuals who may benefit from intensive treatments.

In this paper, we explore a two-sample semiparametric density ratio model with a change point. We introduce novel statistical tests to assess the assumptions of the logistic regression model and to detect a change point in the density ratio. The remainder of this paper is structured as follows. Section 2 details the density ratio change-point model and introduces an empirical likelihood approach for parameter estimation. We also present the maximal score-type test for the null hypothesis of the absence of change points. To retrieve the distributions of the proposed test statistics under the null hypothesis, we employ Monte Carlo and bootstrap methods. Furthermore, we utilize a Kolmogorov–Smirnov type test to evaluate the model's goodness-of-fit. In Section 3, we conduct extensive simulation studies to assess the finite-sample performance of the proposed tests and estimation methods. Section 4 demonstrates the application of these methods using datasets from mother-to-child HIV-1 transmission and oral cancer studies. The paper is concluded in Section 5, while technical details are provided in the Appendix.

2 Methods

2.1 Model and estimation

Let Y be a binary outcome variable and V be a univariate explanatory variable. We assume a logistic regression model with a change point:

$$P(Y=1 \mid V) = \frac{\exp\{\gamma^* + \alpha(V-\eta)_+\}}{1+\exp\{\gamma^* + \alpha(V-\eta)_+\}}, \tag{1}$$

where $x_+ = x\mathrm{I}(x \geq 0)$, with $\mathrm{I}(\cdot)$ being the indicator function; γ^* and α are unknown regression parameters, and η is an unknown change-point parameter. For $V < \eta$, V does not have an effect on Y. If $V \geq \eta$, then a unit increase in V is associated with an increase in α in the log-odds ratio of Y. If η is equal to the lower limit of the support of V, then the model reduces to the standard logistic regression model. Let f and g be the density functions of V, given $Y = 0$ and $Y = 1$, respectively. Under model (1) and using Bayes' rule, we obtain the following semiparametric density ratio change-point model:

$$g(v) = \exp\{\gamma + \alpha(v-\eta)_+\}\, f(v), \tag{2}$$

where $\gamma = -\log\left[\int \exp\{\alpha(v-\eta)_+\} f(v)\,\mathrm{d}v\right]$, such that the resulting g is a proper density function. The constraint on γ implies that if $\alpha = 0$, then no change point exists and the distributions of two groups become identical (i.e., $\gamma = 0$ and $g = f$). Here, we do not assume a parametric form for f.

Consider a case-control study with n_0 and n_1 numbers of non-events and events, respectively, and let $n = n_0 + n_1$. Let $(X_1, \ldots, X_{n_0})$ and $(Z_1, \ldots, Z_{n_1})$ be the independent realizations of V for the non-event and event groups, respectively. Let $\boldsymbol{\zeta} = (\gamma, \alpha)$ and $r(v; \boldsymbol{\zeta}, \eta) = \exp\{\gamma + \alpha(v-\eta)_+\}$. The likelihood function is given by

$$\prod_{i=1}^{n_0} f(X_i) \prod_{j=1}^{n_1} f(Z_j) r(Z_j; \boldsymbol{\zeta}, \eta) = \prod_{i=1}^{n} f(V_i) \prod_{j=1}^{n_1} r(Z_j; \boldsymbol{\zeta}, \eta),$$

where $(V_1, \ldots, V_n) = (X_1, \ldots, X_{n_0}, Z_1, \ldots, Z_{n_1})$. We employ an empirical likelihood approach (Owen, 1988) for the estimation and treat $F(x) \equiv \int_{-\infty}^{x} f(t)\,\mathrm{d}t$ as a step function that jumps at the observed values of V. Let the jump sizes of F at V_i be p_i for $i = 1, \ldots, n$. We require $p_i \geq 0$,

$$\sum_{i=1}^{n} p_i = 1, \qquad \text{and} \qquad \sum_{i=1}^{n} p_i r(V_i; \boldsymbol{\zeta}, \eta) = 1. \tag{3}$$

The log-empirical likelihood can be written as

$$\ell(\boldsymbol{\zeta}, \eta, F) = \sum_{i=1}^{n} \log p_i + \sum_{j=1}^{n_1} \log r(Z_j; \boldsymbol{\zeta}, \eta). \tag{4}$$

We denote $(\widehat{\boldsymbol{\zeta}}, \widehat{\eta}, \widehat{F})$ to be the maximizer of (4).

We propose a grid-search approach to compute the empirical likelihood estimator. For any fixed value of η, we can maximize $\ell(\boldsymbol{\zeta}, \eta, F)$ using the conventional empirical likelihood maximization procedure below. By the Lagrange multiplier method, the values of p_i's that maximize $\ell(\boldsymbol{\zeta}, \eta, F)$ at fixed $(\boldsymbol{\zeta}, \eta)$ are

$$p_i = \frac{1}{n_0 + n_1 r(V_i; \boldsymbol{\zeta}, \eta)}, \qquad i = 1, \ldots, n.$$

Substituting the values of p_i's into the log-empirical likelihood, we obtain the following profile log-likelihood function:

$$\ell(\boldsymbol{\zeta}, \eta) = -\sum_{i=1}^{n} \log\{n_0 + n_1 r(V_i; \boldsymbol{\zeta}, \eta)\} + \sum_{j=1}^{n_1} \log r(Z_j; \boldsymbol{\zeta}, \eta).$$

The first-order derivatives of $\ell(\boldsymbol{\zeta}, \eta)$ with respect to $\boldsymbol{\zeta}$ are given by

$$\begin{aligned} S_\gamma(\boldsymbol{\zeta}, \eta) = \frac{\mathrm{d}}{\mathrm{d}\gamma}\ell(\boldsymbol{\zeta}, \eta) &= n_1 - \sum_{i=1}^{n} \frac{n_1 r(V_i; \boldsymbol{\zeta}, \eta)}{n_0 + n_1 r(V_i; \boldsymbol{\zeta}, \eta)} \quad \text{and} \\ S_\alpha(\boldsymbol{\zeta}, \eta) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\ell(\boldsymbol{\zeta}, \eta) &= \sum_{j=1}^{n_1} (Z_j - \eta)_+ - \sum_{i=1}^{n} \frac{n_1 (V_i - \eta)_+ r(V_i; \boldsymbol{\zeta}, \eta)}{n_0 + n_1 r(V_i; \boldsymbol{\zeta}, \eta)}. \end{aligned}$$

The maximizer of $\ell(\boldsymbol{\zeta}, \eta)$ at fixed η, namely $\widetilde{\boldsymbol{\zeta}}(\eta)$, can be computed by simultaneously solving $S_\gamma(\boldsymbol{\zeta}, \eta) = 0$ and $S_\alpha(\boldsymbol{\zeta}, \eta) = 0$ via the Newton–Raphson method. To obtain the estimator of η, we repeat the above procedure over a grid of η, denoted by $\eta_1, \ldots, \eta_M$. We compute the estimator of η as $\widehat{\eta} = \arg\max_{\eta \in \{\eta_1, \ldots, \eta_M\}} \ell\{\widetilde{\boldsymbol{\zeta}}(\eta), \eta\}$, and the estimator of $\boldsymbol{\zeta}$ is $\widehat{\boldsymbol{\zeta}} = \widetilde{\boldsymbol{\zeta}}(\widehat{\eta})$. Substituting $\widehat{\boldsymbol{\zeta}}$ into the expression of p_i, we obtain

$$\widehat{p}_i = \frac{1}{n_0 + n_1 r(V_i; \widehat{\boldsymbol{\zeta}}, \widehat{\eta})}, \qquad i = 1, \ldots, n.$$

The corresponding estimators for F and G are given by

$$\widehat{F}(x) = \sum_{i=1}^{n} \frac{\mathrm{I}(V_i \le x)}{n_0 + n_1 r(V_i; \widehat{\boldsymbol{\zeta}}, \widehat{\eta})} \qquad \text{and} \qquad \widehat{G}(x) = \sum_{i=1}^{n} \frac{r(V_i; \widehat{\boldsymbol{\zeta}}, \widehat{\eta})\mathrm{I}(V_i \le x)}{n_0 + n_1 r(V_i; \widehat{\boldsymbol{\zeta}}, \widehat{\eta})}.$$

2.2 Maximal score type test

A major interest in studying the density ratio change-point model in (2) is to test for the hypothesis

$$H_0 : \alpha = 0 \qquad \text{against} \qquad H_1 : \alpha \neq 0. \tag{5}$$

Note that a standard procedure, such as the score test or the likelihood-ratio test, is not applicable, because η is non-identifiable under the null hypothesis. We propose the following maximal score test for (5); the test statistic is

$$W_n = \sup_{\eta \in \mathcal{B}} \Big| S_\alpha(\mathbf{0}, \eta) \Big|,$$

where $\mathcal{B}$ is the parameter space of η. The score statistic $S_\alpha(\mathbf{0}, \eta)$ can be expressed as

$$S_\alpha(\mathbf{0}, \eta) = \frac{1}{\sqrt{n}} \left\{ \sum_{j=1}^{n_1} (Z_j - \eta)_+ - \frac{n_1}{n} \sum_{i=1}^{n} (V_i - \eta)_+ \right\} \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i (V_i - \eta)_+,$$

where $w_i = -n_1/n$ for $i = 1, \ldots, n_0$ and $w_i = n_0/n$ for $i = n_0 + 1, \ldots, n$. The score statistic compares the sum of $(V - \eta)_+$ within the group with $Y = 1$ and a re-scaled sum of $(V - \eta)_+$ from the pooled sample. Clearly, the expectation of the score statistic $S_\alpha(\mathbf{0}, \eta)$ is zero for any η under H_0. If $\alpha > 0$, we expect that the first sum is larger than the second sum, which yields a positive statistic at the true value of η, and vice versa. Alternatively, we can consider the maximal normalized score test for (5) based on the test statistic

$$W_n^* = \sup_{\eta \in \mathcal{B}} \left| \frac{S_\alpha(\mathbf{0}, \eta)}{\widehat{\sigma}_\alpha(\eta, \eta)} \right|,$$

where $\widehat{\sigma}_\alpha^2(\eta, \eta') = n^{-1} \sum_{i=1}^{n} \{S_{\alpha,i}(\mathbf{0}, \eta) - \overline{S}_{\alpha,i}(\mathbf{0}, \eta)\}\{S_{\alpha,i}(\mathbf{0}, \eta') - \overline{S}_{\alpha,i}(\mathbf{0}, \eta')\}$ pertains to the variance of $S_\alpha(\mathbf{0}, \eta)$, with $S_{\alpha,i}(\mathbf{0}, \eta) = w_i(V_i - \eta)_+$ and $\overline{S}_{\alpha,i}(\mathbf{0}, \eta) = n^{-1} \sum_{i=1}^{n} S_{\alpha,i}(\mathbf{0}, \eta)$, for $i = 1, \ldots, n$. We have the following theoretical result, whose proof is given in the Appendix.

Theorem 2.1 *Assume that the ratio n_i/n converges to a positive constant ρ_i as $n \to \infty$ for $i = 0, 1$, the support of V includes the interval $\mathcal{B} \equiv (c_l, c_r)$, and the second moment of V exists. Under H_0, $S_\alpha(\mathbf{0}, \eta)/\widehat{\sigma}_\alpha(\eta, \eta)$ converges weakly to a Gaussian process $G(\eta)$ with mean 0.*

Therefore, the maximal score test statistic W_n and the maximal normalized score test statistic W_n^* converge to the supremum of Gaussian processes, where the variances can be consistently estimated.

2.3 Computation of the test statistics

We consider a prespecified sequence of change points $(\eta_1, \ldots, \eta_M)$ and approximate the distribution of $\big\{S_\alpha(\mathbf{0}, \eta_1), \ldots, S_\alpha(\mathbf{0}, \eta_M)\big\}$ based on a multivariate normal distribution with mean $\mathbf{0}$ and variance $\widehat{\boldsymbol{\Sigma}}$. The variance matrix $\widehat{\boldsymbol{\Sigma}}$ is an $(M \times M)$-matrix with the (j, k)th element as $\widehat{\sigma}_\alpha^2(\eta_j, \eta_k)$ for $j, k = 1, \ldots, M$. We propose the following Monte Carlo method to approximate the null distributions for W_n and W_n^*.

1. Generate K sets of standard multivariate normal random variables, namely $\boldsymbol{e}^{(k)}$, each with length M for $k = 1, \ldots, K$.

2. Calculate $W_n^{(k)} \equiv \sup \left| \widehat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{e}^{(k)} \right|$ and $W_n^{*(k)} \equiv \sup \left| \widehat{\boldsymbol{V}}_\alpha^{-1/2} \widehat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{e}^{(k)} \right|$ for each k, where $\widehat{\boldsymbol{V}}_\alpha = \operatorname{diag}\left\{ \widehat{\sigma}_\alpha^2(\eta_1, \eta_1), \ldots, \widehat{\sigma}_\alpha^2(\eta_M, \eta_M) \right\}$.

3. The empirical $100(1-c_0)$th percentiles of $\{W_n^{(1)}, \ldots, W_n^{(K)}\}$ and $\{W_n^{*(1)}, \ldots, W_n^{*(K)}\}$ provide the critical values for W_n and W_n^* at the significance level c_0, respectively.

Since the distribution of W_n (and W_n^*) is completely characterized by the non-parametric distribution function F under the null hypothesis, the bootstrap approach offers a natural alternative to the Monte Carlo method. To generate a bootstrap sample, we draw n observations from $(v_1, \ldots, v_n)$ in the original dataset independently and with replacement. We subsequently assign the first n_0 sampled values to the group with $Y = 0$ and the remaining n_1 sampled values to the group with $Y = 1$. Then, we compute W_n for the bootstrap sample. This sampling procedure is repeated B times to produce a sequence $W_n^{(1)}, \ldots, W_n^{(B)}$. The $100(1-c_0)$th percentile of this sequence can be used as the critical value for the test at the significance level c_0. A similar bootstrap procedure is employed for W_n^*. The finite-sample performance of these two methods will be compared in Section 3.

2.4 m-out-of-n bootstrap for the statistical inference of η

The asymptotic distribution of $\widehat{\eta}$ is complicated and mathematically intractable. The classical bootstrap (Efron and Tibshirani, 1986), which draws a sample of size n with replacement from a dataset of n samples, may be invalid under problems with non-differentiable objective functions or non-smooth statistics (Dümbgen, 1993; Shao, 1994). To resolve this problem, we propose utilizing the m-out-of-n bootstrap method (Bickel et al., 1997; Lee, 1999) to perform statistical inference for $\widehat{\eta}$, where m observations ($m < n$) are sampled with replacement from a dataset of size n, with $m \to \infty$ and $m/n \to 0$. This method has been previously applied to make the inference of the change-point parameter based on the proportional hazards model for right-censored data (Xu et al., 2014; Deng et al., 2017).

Following the work of Bickel and Sakov (2008), Lee (2021) suggested selecting the optimal value of m, denoted as m^*, for the inference of the change-point parameter in logistic regression models. We adopt a similar algorithm to determine m^* here with the following detailed procedure.

1. Generate a sequence of m given by $m_j = [r^j n]$ for $j = 1, 2, \ldots$ and $r \in (0, 1)$, where $[a]$ is the greatest integer less than or equal to a. For each m_j, calculate the empirical bootstrap distribution for the change-point estimator

$$\widetilde{F}_{m_j,n}(x) = N^{-1} \sum_{i=1}^{N} \mathrm{I}\left(m_j^{1/2} (\widetilde{\eta}_{m_j}^{(i)} - \widehat{\eta}) \le x \right),$$

   where N is the number of bootstrap samples, $\widetilde{\eta}_{m_j}^{(i)}$ is the change-point estimate obtained based on the ith bootstrap sample of size m_j, and $\widehat{\eta}$ is the change-point estimate obtained based on the original sample of size n.

2. Set $m^* = \arg\min_{m_j} \sup_x \left| \widetilde{F}_{m_j,n}(x) - \widetilde{F}_{m_{j+1},n}(x) \right|$. If more than one m_j achieves the minimum value, then we choose the largest-valued m_j.

3. The desired m-out-of-n bootstrap estimators are $\widetilde{\eta}_{m^*}^{(i)}, i = 1, \ldots, N$. Denote Q_ψ as the $100 \times \psi$th percentile of the sampled deviations $\widetilde{\eta}_{m^*}^{(i)} - \widehat{\eta}, i = 1, \ldots, N$. The 95% equal-tailed confidence interval for η is constructed according to

$$\left[\widehat{\eta} + (m^*/n)^{1/2} Q_{0.025}, \widehat{\eta} + (m^*/n)^{1/2} Q_{0.975} \right], \tag{6}$$

   where $(m^*/n)^{1/2}$ is the adjustment factor for over-estimated variance when $m^* < n$.

2.5 Goodness-of-fit test

The parameter estimation and hypothesis testing methods described in Sections 2.1 and 2.2 are derived under the assumption that the density ratio change-point model in (2) is correct. The validity of this model in applications needs to be verified. Following the approach of Qin and Zhang (1997), we propose a goodness-of-fit test utilizing a Kolmogorov–Smirnov type test statistic:

$$T_n = \sqrt{n} \sup_x \left| \widehat{F}(x) - \overline{F}(x) \right|,$$

where $\overline{F}(x)$ is the empirical distribution function computed for the sample $(x_1, \ldots, x_{n_0})$. Clearly, if model (2) accurately fits the data, $\widehat{F}$ is anticipated to be a reasonable estimator of F, and the discrepancy between $\widehat{F}$ and $\overline{F}$ should be small. When model (2) is incorrectly specified, the value of T_n is expected to be large. Provided that $\overline{G}$ is the empirical distribution function of the sample $(z_1, \ldots, z_{n_1})$, an alternative approach to constructing the test statistic is to replace $\widehat{F}$ and $\overline{F}$ by $\widehat{G}$ and $\overline{G}$, respectively. The derivation of the asymptotic distribution of T_n is beyond the scope of this study. To obtain the reference distribution of T_n, samples $\{x_1^*, \ldots, x_{n_0}^*\}$ are drawn independently and with replacement from $\widehat{F}(x)$, and independent of x_i^*, $\{z_1^*, \ldots, z_{n_1}^*\}$ are drawn independently and with replacement from $\widehat{G}(x)$, respectively. For each bootstrap sample, we define $\{v_1^*, \ldots, v_n^*\} \equiv \{x_1^*, \ldots, x_{n_0}^*, z_1^*, \ldots, z_{n_1}^*\}$, and compute the test statistic, namely T_n^*. Therefore, we can empirically obtain the critical values of T_n by repeating the bootstrap procedure a sufficient number of times.

3 Simulation studies

We conduct a large-scale simulation study to evaluate the finite-sample performance of the proposed methods. We set f to be the density function of a standard normal random variable. The number of replicates is set at 10,000 for each scenario. For the estimation of the change-point parameter, we consider $M = 50$ candidates, which are equally-spaced intervals between the 10th and 90th empirical quantiles of V.

First, we assess the size and power of the maximal score tests for the hypotheses (5). According to the density ratio change-point model in (2),

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \text{ and } g(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\alpha)^2 + \frac{1}{2}\alpha^2 + \gamma - \alpha\eta\right\} & \text{if } x > \eta, \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2 + \gamma\right) & \text{if } x \le \eta. \end{cases} \tag{7}$$

Let Φ be the cumulative distribution function for a standard normal random variable. For $g(x)$ being a proper density function, the following must be satisfied:

$$\exp\left(\frac{1}{2}\alpha^2 + \gamma - \alpha\eta\right)\Big\{1 - \Phi(\eta - \alpha)\Big\} + \exp(\gamma)\Phi(\eta) = 1, \tag{8}$$

such that the value of γ is deterministic when the values of η and α are assigned. Let $TN(\mu, \sigma^2, a, b)$ denote the $N(\mu, \sigma^2)$ distribution truncated within (a, b). To sample X from density g, we introduce an independent binary latent variable U with probability

$$P(U = 1) = \frac{\Phi(\eta)}{\exp\left(\frac{1}{2}\alpha^2 - \alpha\eta\right)\Big\{1 - \Phi(\eta - \alpha)\Big\} + \Phi(\eta)}.$$

If $X^* \sim TN(\alpha, 1, \eta, \infty)$ when $U = 0$, and $X^* \sim TN(0, 1, -\infty, \eta)$ when $U = 1$, then X^* marginally has the same distribution as X. We consider four sets of sample sizes $(n_0, n_1) = (50, 50), (50, 100), (100, 50)$

and $(100, 100)$. Under the null hypothesis, $\alpha = \gamma = 0$, and η is undefined. Under the alternative hypothesis, we set $\alpha = 0.5$ and 1, and $\eta = -0.5, 0, 0.5$, and $-\infty$, and γ is determined by equation (8). To approximate the null distributions of W_n and W_n^*, we adopt $K = 500,000$ for the Monte Carlo method and $B = 1000$ for the bootstrap method, as detailed in Section 2.3. We present the results under both the null and alternative hypotheses in Table 1. Our simulation results show that the rejection probabilities of both the maximal score and maximal normalized score tests under the null hypothesis closely align with the corresponding nominal significance levels, namely $c_0 = 0.10$ and 0.05. Under the alternative models, the powers of the two tests under both methods are comparable. As expected, the power of the test increases with the magnitude of α and the sample size n. No substantial difference is found between the results obtained via Monte Carlo and bootstrap methods. To assess the computational efficiency of the bootstrap and Monte Carlo methods, we conducted 100 simulation runs on a MacBook Pro equipped with an M2 chip and 16GB memory. Under the scenario $(n_0, n_1) = (50, 50)$, $\alpha = 0.5$, and $\eta = 0$, the bootstrap method took 142 seconds, while the Monte Carlo method required 286 seconds. However, as the sample size increased to $(n_0, n_1) = (500, 500)$, the running times adjusted, with the bootstrap method taking 339 seconds and the Monte Carlo method taking 290 seconds. Given the relatively stable computation time of the Monte Carlo method with varying sample sizes, despite K being substantially larger than B, we recommend the Monte Carlo method over the bootstrap approach.

Second, we study the estimation of the parameters in model (2) and statistical inference of $\widehat{\eta}$. We set the sample sizes $(n_0, n_1) = (500, 500), (500, 1000), (1000, 500)$, and $(1000, 1000)$. Based on (7), we set the parameters $\alpha = 1$ and $\eta = -0.5, 0$, and 0.5. Table 2 reports the bias and empirical variance of the parameter estimates. The results indicate that the proposed estimators of γ, α, and η are virtually unbiased, and the empirical variance decreases as the sample size increases. Table 3 reports the coverage probability of the m-out-of-n bootstrap approach for the nominal 95% confidence interval for η, where we additionally consider the case $\alpha = -1$ and 1.5, as suggested by a reviewer. The sequence of candidates for optimal m is set to be $m_j = [0.75^j n]$ with $j = 1, \ldots, 5$ and $n = n_0 + n_1$. The number of bootstrap samples used for each m_j is set to be $N = 200$. The simulation results show that the m-out-of-n bootstrap provides proper coverage to the change-point parameter η irrespective of the effect size and the true value of η.

Lastly, we examine the finite-sample performance of the goodness-of-fit test. We consider four sets of sample sizes $(n_0, n_1) = (200, 200), (200, 400), (400, 200)$, and $(400, 400)$. We consider model (7) as the null model in this simulation and set $\alpha = -1, 0$ and 1. When $\alpha = 0$, the change-point parameter η is undefined. For $\alpha = -1$ or 1, the values of η are set at $-0.5, 0$, and 0.5. We consider three alternative models that do not possess the structure in model (2), namely $g_1^*(x) = \exp(\gamma - 0.5x^2)f(x)$, $g_2^*(x) = \exp\{\gamma + 2\sin(x)\}f(x)$, and $g_3^*(x) = \exp\left(\gamma + x_+ - 0.5x^2\right)f(x)$. Here, we set the number of replications as 1000 and we adopt grid size $d = 0.1$ within the interval $(-1.5, 1.5)$ in the search of η. We generated 2000 independent combined bootstrap samples based on the estimated distribution functions $\widehat{F}$ and $\widehat{G}$. Table 4 demonstrates that the rejection probabilities of T_n are consistent with the corresponding nominal significance levels under the null hypothesis. Furthermore, as shown in Table 5, our proposed method demonstrates reasonable power under the alternative models.

4 Application

4.1 Mother-to-child-transmission dataset

We apply the proposed methods to analyze the mother-to-child HIV-1 transmission dataset as the first example, where the event of interest is the transmission of HIV-1 from infected mothers to newborns. The main objective is to investigate the effects of an immunological biomarker of the mothers on the incidence risk. The dataset includes 236 subjects, with $n_0 = 157$ non-events and $n_1 = 79$ events. The immunological biomarker $NAb\_SF162LS$, denoted by V^*, is a continuous variable that quantifies the neutralization activity against HIV-1 viruses. Fong et al. (2015) reported that $NAb\_SF162LS$ influences

Table 1 Empirical rejection probabilities of the maximal score and maximal normalized score tests under the null and alternative hypotheses.

(α, η)	(n_0, n_1)	Bootstrap				Monte Carlo			
		W_n		W_n^*		W_n		W_n^*	
		$c_0 = 0.10$	$c_0 = 0.05$	$c_0 = 0.10$	$c_0 = 0.05$	$c_0 = 0.10$	$c_0 = 0.05$	$c_0 = 0.10$	$c_0 = 0.05$
$(0, \eta)$	(50, 50)	0.103	0.051	0.099	0.050	0.102	0.050	0.096	0.042
	(50, 100)	0.103	0.053	0.103	0.051	0.103	0.051	0.100	0.048
	(100, 50)	0.098	0.050	0.099	0.050	0.097	0.048	0.095	0.046
	(100, 100)	0.102	0.050	0.095	0.049	0.098	0.049	0.092	0.044
$(0.5, 0)$	(50, 50)	0.452	0.330	0.459	0.339	0.448	0.328	0.447	0.318
	(50, 100)	0.536	0.403	0.518	0.383	0.536	0.402	0.511	0.370
	(100, 50)	0.568	0.443	0.600	0.481	0.567	0.440	0.593	0.468
	(100, 100)	0.695	0.575	0.706	0.588	0.695	0.570	0.702	0.577
$(0.5, -0.5)$	(50, 50)	0.629	0.509	0.591	0.474	0.629	0.506	0.583	0.448
	(50, 100)	0.746	0.625	0.681	0.554	0.745	0.626	0.675	0.542
	(100, 50)	0.757	0.648	0.737	0.635	0.756	0.645	0.733	0.621
	(100, 100)	0.877	0.805	0.855	0.773	0.876	0.804	0.852	0.766
$(0.5, 0.5)$	(50, 50)	0.254	0.163	0.287	0.196	0.250	0.162	0.280	0.182
	(50, 100)	0.285	0.186	0.309	0.190	0.286	0.183	0.303	0.181
	(100, 50)	0.321	0.219	0.391	0.286	0.320	0.215	0.383	0.277
	(100, 100)	0.408	0.288	0.475	0.356	0.406	0.285	0.472	0.343
$(0.5, -\infty)$	(50, 50)	0.778	0.676	0.699	0.595	0.777	0.673	0.684	0.562
	(50, 100)	0.873	0.789	0.800	0.701	0.873	0.790	0.796	0.690
	(100, 50)	0.885	0.810	0.834	0.751	0.873	0.790	0.796	0.690
	(100, 100)	0.968	0.932	0.940	0.895	0.968	0.931	0.938	0.890
$(1, 0)$	(50, 50)	0.968	0.933	0.969	0.937	0.968	0.933	0.965	0.926
	(50, 100)	0.993	0.980	0.990	0.977	0.993	0.980	0.989	0.975
	(100, 50)	0.992	0.981	0.993	0.985	0.992	0.980	0.992	0.984
	(100, 100)	1.000	0.998	1.000	0.999	0.999	0.998	1.000	0.998
$(1, -0.5)$	(50, 50)	0.996	0.992	0.994	0.987	0.996	0.992	0.993	0.984
	(50, 100)	1.000	0.999	0.999	0.998	1.000	0.999	0.999	0.997
	(100, 50)	1.000	0.999	0.999	0.998	1.000	0.999	0.999	0.997
	(100, 100)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$(1, 0.5)$	(50, 50)	0.771	0.658	0.826	0.734	0.771	0.656	0.819	0.714
	(50, 100)	0.873	0.777	0.896	0.814	0.872	0.777	0.892	0.805
	(100, 50)	0.877	0.799	0.927	0.883	0.875	0.798	0.924	0.880
	(100, 100)	0.963	0.923	0.980	0.962	0.963	0.921	0.979	0.958
$(1, -\infty)$	(50, 50)	0.999	0.998	0.998	0.995	0.999	0.998	0.998	0.994
	(50, 100)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(100, 50)	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.999
	(100, 100)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2 Parameter estimation under the density ratio change-point model.

(α, η)	(n_0, n_1)	γ		α		η	
		$\mathrm{Bias}(\widehat{\gamma})$	$\mathrm{Var}(\widehat{\gamma})$	$\mathrm{Bias}(\widehat{\alpha})$	$\mathrm{Var}(\widehat{\alpha})$	$\mathrm{Bias}(\widehat{\eta})$	$\mathrm{Var}(\widehat{\eta})$
(1, 0)	(500, 500)	−0.014	0.012	0.023	0.025	−0.014	0.063
	(500, 1000)	−0.008	0.006	0.018	0.019	−0.008	0.040
	(1000, 500)	−0.010	0.009	0.014	0.015	−0.008	0.042
	(1000, 1000)	−0.005	0.005	0.010	0.011	−0.005	0.025
(1, −0.5)	(500, 500)	−0.024	0.027	0.016	0.013	−0.019	0.067
	(500, 1000)	−0.013	0.015	0.012	0.010	−0.009	0.041
	(1000, 500)	−0.023	0.026	0.010	0.009	−0.019	0.056
	(1000, 1000)	−0.013	0.014	0.007	0.006	−0.010	0.033
(1, 0.5)	(500, 500)	−0.008	0.004	0.046	0.062	−0.010	0.072
	(500, 1000)	−0.005	0.003	0.039	0.050	−0.006	0.050
	(1000, 500)	−0.005	0.003	0.026	0.033	−0.005	0.046
	(1000, 1000)	−0.003	0.002	0.022	0.025	−0.001	0.029

Table 3 Coverage probability of the nominal 95% confidence interval for η based on the m-out-of-n bootstrap under the alternative hypothesis with $\alpha = -1, 1$, and 1.5.

α	η	(n_0, n_1)			
		(500, 500)	(500, 1000)	(1000, 500)	(1000, 1000)
−1	0	0.966	0.966	0.968	0.966
	−0.5	0.966	0.961	0.961	0.957
	0.5	0.952	0.959	0.965	0.964
1	0	0.958	0.960	0.959	0.957
	−0.5	0.960	0.961	0.953	0.954
	0.5	0.966	0.965	0.964	0.964
1.5	0	0.954	0.957	0.955	0.954
	−0.5	0.953	0.945	0.951	0.947
	0.5	0.961	0.961	0.958	0.955

the incidence risk only if its value exceeds a specific change point. This finding prompts us to re-examine the dataset using the density ratio change-point model.

We apply a monotonic transformation to the variable V^*, such that $V = \log(1 + V^*)$, which ranges from 1.592 to 2.713. We set η_1 and η_M as the 10th and 90th percentiles of the observed values of V, respectively, to avoid edge effects where numerical instability may arise, and $\eta_2, \ldots, \eta_{M-1}$ are obtained with a grid size $d = 0.01$. To assess the validity of the density ratio change-point model, we first implement the goodness-of-fit test as detailed in Section 2.5. Based on 10,000 bootstrap samples, the p-value for the goodness-of-fit test is 0.173, indicating that the observed sample conforms to the density ratio change-point model. Subsequently, we conduct maximal score and normalized score tests to detect the presence of a change point, using the same set of grids for η as in the goodness-of-fit test. The results are summarized in Table 6. Both tests reject the null hypothesis of no change point, with p-values less than 0.01, where the reference distributions of W_n and W_n^* are obtained via Monte Carlo and bootstrap methods (with $K = 500,000$ and $B = 10,000$). For comparative purposes, we also implement the maximal likelihood

Table 4 Empirical rejection probabilities of the goodness-of-fit test when the true model is a density ratio change-point model.

α	(n_0, n_1)	$\eta = 0$		$\eta = -0.5$		$\eta = 0.5$	
		$c_0 = 0.10$	$c_0 = 0.05$	$c_0 = 0.10$	$c_0 = 0.05$	$c_0 = 0.10$	$c_0 = 0.05$
0	$(200, 200)$	0.097	0.047	–		–	
	$(200, 400)$	0.094	0.040	–		–	
	$(400, 200)$	0.072	0.035	–		–	
	$(400, 400)$	0.125	0.059	–		–	
-1	$(200, 200)$	0.089	0.041	0.088	0.042	0.108	0.044
	$(200, 400)$	0.091	0.044	0.078	0.032	0.114	0.064
	$(400, 200)$	0.102	0.043	0.093	0.040	0.119	0.067
	$(400, 400)$	0.091	0.038	0.092	0.039	0.104	0.049
1	$(200, 200)$	0.091	0.039	0.108	0.045	0.092	0.047
	$(200, 400)$	0.100	0.045	0.102	0.046	0.100	0.053
	$(400, 200)$	0.082	0.040	0.086	0.044	0.083	0.040
	$(400, 400)$	0.096	0.053	0.103	0.055	0.101	0.044

Table 5 Empirical rejection probabilities of the goodness-of-fit test under the alternative hypothesis.

(n_0, n_1)	$g_1^*(x)$		$g_2^*(x)$		$g_3^*(x)$	
	$c_0 = 0.10$	$c_0 = 0.05$	$c_0 = 0.10$	$c_0 = 0.05$	$c_0 = 0.10$	$c_0 = 0.05$
$(200, 200)$	0.811	0.689	0.756	0.670	0.635	0.473
$(200, 400)$	0.894	0.825	0.779	0.656	0.696	0.538
$(400, 200)$	0.911	0.843	0.906	0.845	0.768	0.614
$(400, 400)$	0.970	0.937	0.941	0.894	0.881	0.769

ratio test proposed by Fong et al. (2015). The p-value of this test in the logistic regression model with a change-point effect is 0.002, which aligns with our test results.

In model (2), the estimated values (standard errors) for γ, α, and η are 0.523 (0.198), -3.960 (0.660), and 2.125 (0.076), respectively, where the standard errors are obtained based on the m-out-of-n bootstrap with $N = 1000$, $d = 0.01$, and $m_j = [0.75^j n]$ for $j = 1, \ldots, 5$. The confidence interval for η can be constructed via the m-out-of-n bootstrap method described in Section 2.4. The 95% confidence interval for α can be constructed analogously by $\left[\widehat{\alpha} + (m^*/n)^{1/2} Q_{\alpha, 0.025}, \widehat{\alpha} + (m^*/n)^{1/2} Q_{\alpha, 0.975}\right]$, where $Q_{\alpha,\psi}$ denotes the ψth quantile of the deviations between the estimates of α in sample of size m^* and $\widehat{\alpha}$; the number of m^* is the same as in the m-out-of-n bootstrap for η. The 95% confidence intervals for α and η are $[-4.542, -2.703]$ and $[2.007, 2.298]$, respectively. Hence, the risk of HIV-1 transmission starts to drop significantly when $NAb\_SF162LS$ is higher than 7.37 with a 95% confidence interval $[6.441, 8.954]$, via a backward transformation of $\exp(\cdot) - 1$ of the confidence interval of η.

Figure 1a displays the maximized log-likelihood values across different η values. Figure 1b illustrates the empirical cumulative distribution functions $\overline{F}$ and $\overline{G}$ for the non-event and event groups, respectively, along with the estimated distribution functions $\widehat{F}$ and $\widehat{G}$, based on the proposed model evaluated at the estimated parameter values. One can observe that the estimated distribution function closely resembles the empirical one in their respective groups, and there is a significant discrepancy between $\overline{F}$ and $\overline{G}$ (also $\widehat{F}$ and $\widehat{G}$).

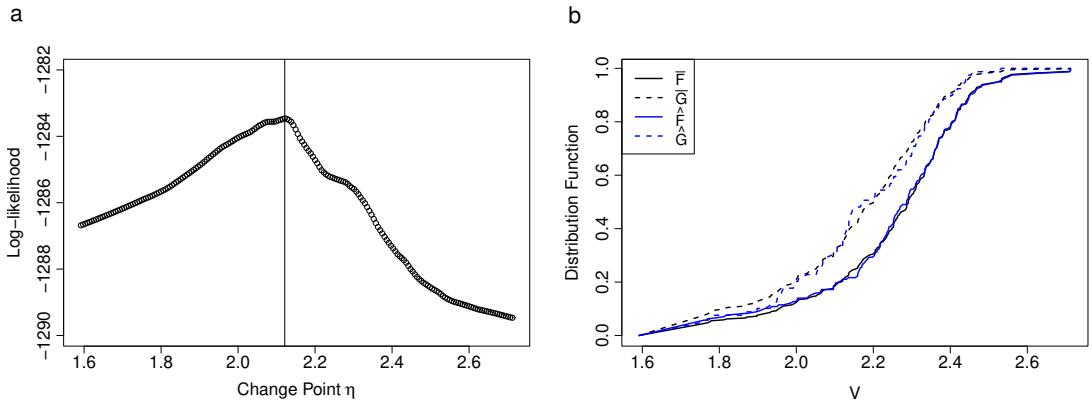

Figure 1 Results for the mother-to-child HIV transmission dataset: **a** The log-empirical likelihoods maximized at different candidate change points. **b** The empirical and estimated cumulative distribution functions of the event and non-event groups.

Table 6 Test results for the presence of a change point for the mother-to-child HIV transmission dataset.

Method	Test	Statistic	Critical values			p-value
			$c_0 = 0.10$	$c_0 = 0.05$	$c_0 = 0.01$	
Monte Carlo	W_n	0.259	0.140	0.166	0.218	0.002
	W_n^*	3.367	2.077	2.372	2.958	0.003
Bootstrap	W_n	0.259	0.140	0.166	0.217	0.003
	W_n^*	3.367	2.063	2.346	2.940	0.002

4.2 Oral cancer dataset

In the second example, we apply our methods to the oral cancer dataset of African Americans collected from the National Cancer Institute's Oral Cancer Study (Day et al., 1993). Previously analyzed by Rosenberg et al. (2003) using a class of semiparametric generalized linear models, their findings suggested that alcohol consumption has a nonlinear effect on the risk of oral cancer incidence. Here, we fit the data using the proposed density ratio change-point model. The dataset includes $n_0 = 203$ subjects without oral cancer (non-event) and $n_1 = 194$ subjects with oral cancer (event). We form the change-point variable by $V = \log(1 + V^*)$, ranging from 0 to 4.949, where V^* is the weekly alcohol consumption variable given in the original dataset. We apply the goodness-of-fit test with a grid size of $d = 0.05$ and $10,000$ bootstrap samples. We obtain a p-value of 0.545 indicating that the observed data are consistent with the proposed model. Then, we apply the maximal score type tests for the presence of a change point in the effect of alcohol consumption on oral cancer risk. As shown in Table 7, both score tests under two re-sampling methods (with $K = 500,000$ and $B = 10,000$) reject the null hypothesis of no change point with p-values less than 0.001. The maximal likelihood ratio test under a logistic regression model with a change-point effect also rejects the null hypothesis with a p-value < 0.001.

The estimated values (standard errors) for γ, α, and η are, respectively, -1.039 (0.217), 1.167 (0.278), and 2.150 (0.574), where the standard errors are obtained based on the m-out-of-n bootstrap with $N = 1000$, $d = 0.01$ and $m_j = [0.75^j n]$ for $j = 1, \ldots, 5$. The 95% confidence intervals for α and η are $[0.758, 1.831]$ and $[0.576, 2.750]$, respectively. The fitted model suggests that the risk of oral cancer increases when alcohol consumption exceeds 7.585 oz drinks per week with a 95% confidence interval $[0.779, 14.636]$ reconverted by $\exp(\cdot) - 1$. The plots of the profile log-likelihood and the empirical and

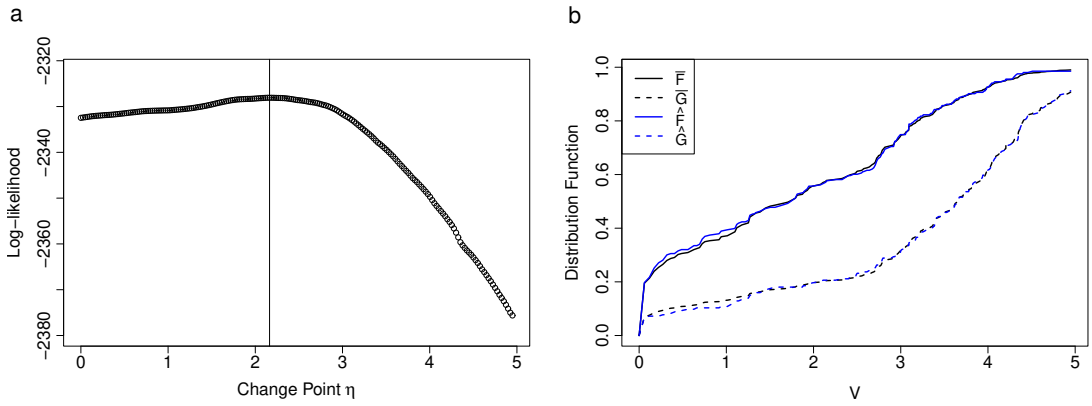

Figure 2 Results for the oral cancer dataset: **a** The log-empirical likelihoods maximized at different candidate change points. **b** The empirical and estimated cumulative distribution functions of the event and non-event groups.

estimated distribution functions are shown in Figure 2. Again, a large discrepancy is observed between the empirical distribution functions of the two groups, and the estimated distribution functions can provide a good approximation to the empirical ones.

Table 7 Test results for the presence of a change point for the oral cancer dataset.

Method	Test	Statistic	Critical values			p-value
			$c_0 = 0.10$	$c_0 = 0.05$	$c_0 = 0.01$	
Monte Carlo	W_n	7.391	1.358	1.616	2.120	< 0.001
	W_n^*	9.393	2.087	2.387	2.974	< 0.001
Bootstrap	W_n	7.391	1.366	1.631	2.172	< 0.001
	W_n^*	9.393	2.094	2.391	2.949	< 0.001

5 Concluding remarks

In this study, we address the challenge of detecting a change point under the two-sample density ratio model, which stems from the logistic regression model for case-control data. This model allows us to verify the validity of the logistic regression model by employing a Kolmogorov–Smirnov type test statistic. We introduce both the maximal score and normalized score tests to detect the presence of a change point in the covariate effects and we establish the asymptotic properties of the score statistics. A major motivation for considering a density ratio change-point model over the existing change-point models, such as those described by Fong et al. (2015), is its ability to scrutinize the underlying assumptions of logistic regression models via a goodness-of-fit test. This enhancement may improve the accuracy and interpretability of the research findings, which is particularly relevant in the pharmaceutical industry where logistic regression models are frequently assumed without empirical validation.

Although the proposed methods are developed based on the popular model with zero covariate effect on the response variable before the change point, they can be easily generalized to accommodate a model with a non-zero baseline covariate effect. This model can be specified as

$$g(v) = \exp\{\gamma + \alpha_1 v + \alpha_2 (v - \eta)_+\} f(v),$$

where α_1 and $(\alpha_1 + \alpha_2)$ denote the effects of V on outcome Y before and after change point, respectively.

Our work can be extended in several directions. First, while the logistic regression model is commonly used for binary outcomes in practice, the S-shaped assumption of the logistic link function may be restrictive. An alternative approach could involve a binary outcome model with an unknown nonparametric link function to model the association between the change-point covariate and the response variable, such as those described by Lee et al. (2024). This class of flexible models, which includes the probit and complementary log-log models as special cases, can generally provide a better fit for the data. Second, when two covariates are present, namely W and Z, we may be interested in the following change-point model:

$$g(w, z) = \exp\{\gamma + \beta_1 w + \beta_2 w \, \mathrm{I}(z \geq \eta)\} \, f(w, z),$$

where the ratio of the density of W between the two groups is influenced by the change-point variable Z. Such a model can be particularly useful in medical studies, for example, in diabetes research where Z could represent telomere length and W could be a continuous covariate, such as age. Third, we may consider the proportional likelihood ratio model (Luo and Tsai, 2012) to accommodate both linear and binary outcomes. In this model, the density of Y given V may be specified as

$$f(y \mid v) = \frac{\exp\{\alpha y (v - \eta)_+\} h(y)}{\int \exp\{\alpha y (v - \eta)_+\} \mathrm{d}H(y)},$$

where H represents an unspecified baseline distribution function with respect to a certain dominating measure, with h being the corresponding density function. Here, the parameter α is a generalization of the log odds ratio to encompass an arbitrary outcome Y. Lastly, in epidemiological studies of disease progression, we are often interested in the analysis of the survival times of subjects in two or more groups, where the variable concerned (such as time to death) is either left truncated or censored. While the Cox proportional hazards model has been adopted in change-point analysis for survival data, as demonstrated in recent studies by Lee et al. (2020) and Lee and Wong (2023), a notable gap in research utilizing the density ratio model remains. The proposed methods can be extended for analyzing time-to-event data, particularly in settings where the assumptions of the Cox model may not hold, providing new insights into epidemiological research.

Acknowledgements This research was supported by a start-up research grant from The Hong Kong Polytechnic University (P0034889) and the GuangDong Basic and Applied Basic Research Foundation (Project No. 2021A1515110048).

Conflict of Interest

The authors have declared no conflict of interest.

Appendix

Proof of Theorem 2.1. Note that

$$\begin{aligned} S_\alpha(\mathbf{0}, \eta) &= \frac{1}{\sqrt{n}} \left\{ \sum_{j=1}^{n_1} (Z_j - \eta)_+ - \frac{n_1}{n} \sum_{i=1}^{n} (V_i - \eta)_+ \right\} \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n_1} \frac{n_0}{n} \left[(Z_j - \eta)_+ - \mathrm{E}\{(Z - \eta)_+\} \right] - \frac{1}{\sqrt{n}} \sum_{i=1}^{n_0} \frac{n_1}{n} \left[(X_j - \eta)_+ - \mathrm{E}\{(X - \eta)_+\} \right]. \end{aligned} \tag{9}$$

We first show that the first term on the right-hand side of the second equality of (9) converges weakly to a mean zero Gaussian process. Let $\mathcal{H}$ be the class of functions $\{h(Z) = \mathrm{I}(Z \geq \eta)(Z - \eta) : \eta \in \mathcal{B}\}$. Then

$H(z) = \mathrm{I}(z \geq c_l)(z - c_l)$ is an envelope for $\mathcal{H}$. The uniform covering number is

$$\sup_Q N(\epsilon \|H\|_{Q,2}, \mathcal{H}, L_2(Q)) = O\Big(\frac{1}{\epsilon}\Big),$$

where the supremum is taken over all finitely discrete probability measures Q with $\|H\|_{Q,2} > 0$. Thus

$$J(1, \mathcal{H}, L_2) = \int_0^1 \sqrt{\log \sup_Q N(\epsilon \|H\|_{Q,2}, \mathcal{H}, L_2(Q))}\, \mathrm{d}\epsilon < \infty.$$

We have that $\mathcal{H}$ is a Donsker class by Theorem 2.5 of Kosorok (2008). Thus the first term on the right-hand side of the second equality of (9) converges weakly to a mean zero Gaussian process with covariance function evaluated at η, η' as $\rho_0^2 \rho_1 \mathrm{E}\big(\big[(Z-\eta)_+ - \mathrm{E}\{(Z-\eta)_+\}\big]\big[(Z-\eta')_+ - \mathrm{E}\{(Z-\eta')_+\}\big]\big)$. Similarly, the second term on the right-hand side of the second equality of (9) converges weakly to a mean zero Gaussian process with covariance $\rho_0 \rho_1^2 \mathrm{E}\big(\big[(X-\eta)_+ - \mathrm{E}\{(X-\eta)_+\}\big]\big[(X-\eta')_+ - \mathrm{E}\{(X-\eta')_+\}\big]\big)$. Combining the above results, the score statistic $S_\alpha(\mathbf{0}, \eta)$ converges weakly to a Gaussian process with mean zero and covariance function

$$\sigma_\alpha^2(\eta, \eta') = \rho_0 \rho_1 \mathrm{E}\big(\big[(V-\eta)_+ - \mathrm{E}\{(V-\eta)_+\}\big]\big[(V-\eta')_+ - \mathrm{E}\{(V-\eta')_+\}\big]\big)$$

for every $\eta, \eta' \in \mathcal{B}$. To have the desired result, it suffices to show that

$$\sup_{\eta \in \mathcal{B}} |\widehat{\sigma}_\alpha^2(\eta, \eta') - \sigma_\alpha^2(\eta, \eta')| = o_p(1).$$

Let $\mathcal{K}$ be the class of functions $\{k(Z) = \mathrm{I}(Z \geq \eta)(Z - \eta)\mathrm{I}(Z \geq \eta')(Z - \eta') : \eta, \eta' \in \mathcal{B}\}$. We can show that $\mathcal{K}$ is Glivenko-Cantelli analogously. The desired result follows. □

References

Bickel, P. J., Götze, F., and van Zwet, W. R. (1997). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica*, 7:1–31.

Bickel, P. J. and Sakov, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica*, 18(3):967–985.

Bondell, H. D. (2007). Testing goodness-of-fit in logistic case-control studies. *Biometrika*, 94(2):487–495.

Day, G. L., Blot, W. J., Austin, D. F., Bernstein, L., Greenberg, R. S., Preston-Martin, S., Schoenberg, J. B., Winn, D. M., McLaughlin, J. K., and Fraumeni Jr, J. F. (1993). Racial differences in risk of oral and pharyngeal cancer: alcohol, tobacco, and other determinants. *JNCI: Journal of the National Cancer Institute*, 85(6):465–473.

Deng, Y., Zeng, D., Zhao, J., and Cai, J. (2017). Proportional hazards model with a change point for clustered event data. *Biometrics*, 73(3):835–845.

Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*, 95:125–140.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, pages 54–75.

Fong, Y., Di, C., and Permar, S. (2015). Change point testing in logistic regression models with interaction term. *Statistics in Medicine*, 34(9):1483–1494.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer.

Lee, C. Y. (2021). Nested logistic regression models and ΔAUC applications: change-point analysis. *Statistical Methods in Medical Research*, 30(7):1654–1666.

Lee, C. Y., Chen, X., and Lam, K. F. (2020). Testing for change-point in the covariate effects based on the Cox regression model. *Statistics in Medicine*, 39(10):1473–1488.

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

Lee, C. Y., Wong, K. Y., and Bandyopadhyay, D. (2024). Partly linear single-index cure models with a nonparametric incidence link function. *Statistical Methods in Medical Research*, 33(3):498–514.

Lee, S. M. (1999). On a class of m out of n bootstrap confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(4):901–911.

Luo, X. and Tsai, W. Y. (2012). A proportional likelihood ratio model. *Biometrika*, 99(1):211–222.

Molinari, N., Daurès, J.-P., and Durand, J.-F. (2001). Regression splines for threshold selection in survival data analysis. *Statistics in Medicine*, 20(2):237–247.

Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, 22(19):3055–3071.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Permar, S. R., Fong, Y., Vandergrift, N., Fouda, G. G., Gilbert, P., Parks, R., Jaeger, F. H., Pollara, J., Martelli, A., Liebl, B. E., et al. (2015). Maternal HIV-1 envelope–specific antibody responses and reduced risk of perinatal transmission. *The Journal of Clinical Investigation*, 125(7):2702–2706.

Qin, J. and Zhang, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 84(3):609–618.

Rosenberg, P. S., Katki, H., Swanson, C. A., Brown, L. M., Wacholder, S., and Hoover, R. N. (2003). Quantifying epidemiologic risk factors using non-parametric regression: model selection remains the greatest challenge. *Statistics in Medicine*, 22(21):3369–3381.

Shao, J. (1994). Bootstrap sample size in nonregular cases. *Proceedings of the American Mathematical Society*, 122(4):1251–1262.

Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *The Annals of Statistics*, 10(2):616–620.

Vardi, Y. (1985). Empirical distributions in selection bias models. *The Annals of Statistics*, 13(1):178–203.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, 50(1):1–25.

Xu, G., Sen, B., and Ying, Z. (2014). Bootstrapping a change-point Cox model for survival data. *Electronic journal of statistics*, 8(1):1345.

Zhang, B. (1999). A chi-squared goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 86(3):531–539.

Zhang, B. (2001). An information matrix test for logistic regression models based on case-control data. *Biometrika*, 88(4):921–932.

Zhao, J., Zhu, Y., Lin, J., Matsuguchi, T., Blackburn, E., Zhang, Y., Cole, S. A., Best, L. G., Lee, E. T., and Howard, B. V. (2014). Short leukocyte telomere length predicts risk of diabetes in American Indians: The Strong Heart Family Study. *Diabetes*, 63(1):354–362.

