RESEARCH ARTICLE

Structural change detection in ordinal time series

Fuxiao Li[1]*, **Mengli Hao**[1], **Lijuan Yang**[2]

1 Department of Applied Mathematics, Xi'an University of Technology, Xi'an, Shaanxi, People's Republic of China, **2** Department of Mathematics and Information Science, Chang'an University, Xi'an, Shaanxi, People's Republic of China

☯ These authors contributed equally to this work.
* fx_lmzq@163.com

Abstract

Change-point detection in health care data has recently obtained considerable attention due to the increased availability of complex data in real-time. In many applications, the observed data is an ordinal time series. Two kinds of test statistics are proposed to detect the structural change of cumulative logistic regression model, which is often used in applications for the analysis of ordinal time series. One is the standardized efficient score vector, the other one is the quadratic form of the efficient score vector with a weight function. Under the null hypothesis, we derive the asymptotic distribution of the two test statistics, and prove the consistency under the alternative hypothesis. We also study the consistency of the change-point estimator, and a binary segmentation procedure is suggested for estimating the locations of possible multiple change-points. Simulation results show that the former statistic performs better when the change-point occurs at the centre of the data, but the latter is preferable when the change-point occurs at the beginning or end of the data. Furthermore, the former statistic could find the reason for rejecting the null hypothesis. Finally, we apply the two test statistics to a group of sleep data, the results show that there exists a structural change in the data.

Introduction

In categorical data analysis, ordinal categorical variables are frequently encountered in many contexts, such as health status (very good, good, so-so, bad, very bad), blood pressure (low, normal, high). The data observed hourly or daily constitutes an ordinal time series. The cumulative logistic regression model is often applied for analyzing the ordinal time series [1]. Sometimes the model may change at some unknown time moments (change-points) while it remains stable between these points. Structural stability is of prime importance in statistical modeling and inference. If the parameters have changed with the observed sample, inferences can be severely biased, and forecasts lose accuracy. Because of the importance of parameter stability, it is necessary to detect the structural change. Studies of structural change detection has been a popular research subject in statistics, see Csörgő and Horváth [2], Bai and Perron [3],

Lee et al. [4], Perron [5], Gombay [6], Wang et al. [7], Chen et al. [8], Baranowski et al. [9], Wang et al. [10], Chen [11] and Liu et al. [12] for reviews of the field.

Structural changes detection in categorical data have been considered as well. Höhle [13] proposed a prospective CUSUM change-point detection procedure to detect a structural change in categorical time series; Wang et al. [10] described a procedure based on high-dimensional homogeneity test to detect and estimate multiple change-point in multinomial data; Plasse and Adams [14] illustrated a multiple change-point detection method for categorical data streams, which could adaptively monitor the category probabilities. As generalized linear regression models for categorical time series allow for parsimonious modeling and incorporation of random time-dependent covariates, Fokianos and Kedem [15] suggested the generalized linear model for categorical time series modeling. For change-point detection in the generalized linear model, Xia et al. [16] introduced two procedures to sequentially detect the structural change in generalized linear models with assuming independence; Hudecová [17] investigated the detection of change in autoregressive models for binary time series; Fokianos et al. [18] provided a statistical procedure based on the partial likelihood score process to detect a structural change in binary logistic regression model; Gombay et al. [19] and Li et al. [20] discussed retrospective change detection and sequential change detection in multinominal logistic regression model.

Score test for detection of changes in time series models has been studied by Gombay and Serban [21], Gombay et al. [22]. The test statistic is usually computationally less demanding than the likelihood ratio test statistic. In this paper, we first propose a test statistic based on the efficient score vector to detect a structural change in cumulative logistic regression model, which extends the change-point detection of Gombay et al. [19]. Simulation shows that the empirical power of the proposed statistic is low when the change-point occurs at the beginning or end of the data. To this end, we propose a new statistic, which is the quadratic form of the efficient score vector and has a weight function. Under the null hypothesis of no change, we derive the asymptotic distribution of the two statistics, and prove the consistency under the alternative hypothesis. We also study the consistency of the change-point estimator, and a binary segmentation procedure is suggested for estimating the locations of possible multiple change-points. Simulation results show that the empirical size of the two statistics is close to the significance level 0.05, and the empirical power is approximate to 1 when the sample size is large. The empirical power of the former statistic is higher when the change-point is located at the centre of the data, but the latter performs better when the change-point is located at the beginning or end of the data. Furthermore, the former statistic could find the reason for rejecting the null hypothesis. Finally, we apply the two statistics to study a group of sleep data, and find a structural change in the data.

The model and hypotheses

Consider a categorical time series $\{Y_t\}$ with m categories, $Y_t = (Y_{t1}, \ldots, Y_{tq})'$, $q = m - 1$,

$$Y_{tj} = \begin{cases} 1, & \text{if the } j \text{ category is observed at time } t, \\ 0, & \text{otherwise} \end{cases}$$

for $t = 1, 2, \ldots, n$ and $j = 1, \ldots, q$, $Y_{tm} = 1 - \sum_{j=1}^{q} Y_{tj}$. The vector of conditional probability $\boldsymbol{\pi}_t = (\pi_{t1}, \ldots, \pi_{tq})'$ is defined by

$$\pi_{tj} = E(Y_{tj}|\mathcal{F}_{t-1}) = P(Y_{tj} = 1|\mathcal{F}_{t-1}), \quad j = 1, \ldots, q,$$

for every t, $\pi_{tm} = 1 - \sum_{j=1}^{q} \pi_{tj}$, where

$$\mathcal{F}_{t-1} = \sigma\{\boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \dots, \boldsymbol{Z}_{t-1}, \boldsymbol{Z}_{t-2}, \dots\},$$

$\{\boldsymbol{Z}_{t-1}\}$ denotes the $p \times q$ covariate matrices.

Define an ordinal time series $\{Y_t\}$, where $Y_t = j$ is equivalent to $Y_{tj} = 1$ for $j = 1, 2, \dots, m$, $t = 1, 2, \dots, n$. Let $\{X_t\}$ be a latent variable time series, where $X_t = -\boldsymbol{\beta}'\boldsymbol{z}_{t-1} + e_t$, $\boldsymbol{\beta} \in \boldsymbol{R}^d$, $\boldsymbol{z}_{t-1}$ is a d-dimensional covariate vector, e_t is a white noise process with continuous cumulative distribution function F. Suppose that $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_m = \infty$ are threshold parameters, such that Y_t satisfies

$$Y_t = j \text{ if } \alpha_{j-1} \le X_t < \alpha_j$$

for $j = 1, 2, \dots, m$. According to the equivalence relation between Y_t and Y_{tj}, we have

$$\begin{aligned} \pi_{tj} &= P(Y_{tj} = 1|\mathcal{F}_{t-1}) = P(Y_t = j|\mathcal{F}_{t-1}) = P(\alpha_{j-1} \le X_t < \alpha_j|\mathcal{F}_{t-1}) \\ &= F(\alpha_j + \boldsymbol{\beta}'\boldsymbol{z}_{t-1}) - F(\alpha_{j-1} + \boldsymbol{\beta}'\boldsymbol{z}_{t-1}), \end{aligned}$$

then

$$P(Y_t \le j|\mathcal{F}_{t-1}) = F(\alpha_j + \boldsymbol{\beta}'\boldsymbol{z}_{t-1}), \ j = 1, 2, \dots, m.$$

If $F(x)$ is the logistic distribution function, then F^{-1} is the logistic link function $\log it(x)$, where logit $(x) = \ln (x/(1-x))$, $0 < x < 1$. Thus we have

$$\log it(P(Y_t \le j|\mathcal{F}_{t-1})) = \alpha_j + \boldsymbol{\beta}'\boldsymbol{z}_{t-1}, \ j = 1, 2, \dots, q,$$

which is called the cumulative logistic regression model.

Let $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_q, \boldsymbol{\beta}')'$ be a p-dimensional parameter vector, $p = q + d$. In this paper we wish to test if there exists a structural change in the parameter $\boldsymbol{\theta}$, that is,

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \ t = 1, 2, \dots, n,$$
$$H_A : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \ t = 1, 2, \dots, k^*, \text{ and } \boldsymbol{\theta} = \boldsymbol{\theta}_0^*, \ t = k^* + 1, \dots, n,$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$, k^* denotes the change-point which occurs in some of the parameter $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0 \ne \boldsymbol{\theta}_0^*$, $\boldsymbol{\theta}_0, \boldsymbol{\theta}_0^*$ and k^* are unknown.

Next, we estimate the parameter vector $\boldsymbol{\theta}$ by the partial likelihood method (Fokianos et al. [18]). The partial likelihood function

$$PL(\boldsymbol{\theta}) = \prod_{t=1}^{n} f(\boldsymbol{Y}_t; \boldsymbol{\theta}|\mathcal{F}_{t-1}) = \prod_{t=1}^{n} \prod_{j=1}^{m} \pi_{tj}^{Y_{tj}}(\boldsymbol{\theta})$$

and the partial log-likelihood function

$$l(\boldsymbol{\theta}) = \log PL(\boldsymbol{\theta}) = \sum_{t=1}^{n} \sum_{j=1}^{m} Y_{tj} \log \pi_{tj}(\boldsymbol{\theta}) \tag{1}$$

are defined in Gombay et al. [19]. Denote the partial score vector

$$\begin{aligned} \boldsymbol{S}_n(\boldsymbol{\theta}) &= \nabla l(\boldsymbol{\theta}) = \left(\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1}, \dots, \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_p} \right)' \\ &= \sum_{t=1}^{n} \boldsymbol{Z}_{t-1} \boldsymbol{U}_t(\boldsymbol{\theta})(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta})) = \sum_{t=1}^{n} \boldsymbol{Z}_{t-1} \boldsymbol{D}_t(\boldsymbol{\theta})\Sigma_t^{-1}(\boldsymbol{\theta})(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta})), \end{aligned}$$

where

$$
\boldsymbol{Z}_{t-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ \boldsymbol{z}_{t-1} & \boldsymbol{z}_{t-1} & \cdots & \boldsymbol{z}_{t-1} \end{pmatrix}_{p \times q},
$$

$\boldsymbol{U}_t(\boldsymbol{\theta}) = \boldsymbol{D}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})$, $\boldsymbol{D}_t(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{h}(\boldsymbol{\eta}_t)}{\partial \boldsymbol{\eta}_t'}$, $\boldsymbol{h}(\boldsymbol{\eta}_t) = (h_1(\boldsymbol{\eta}_t), \ldots, h_q(\boldsymbol{\eta}_t))'$, $\boldsymbol{\eta}_t = \boldsymbol{Z}_{t-1}'\boldsymbol{\theta} = (\eta_{t1}, \ldots, \eta_{tq})'$.
$h_1(\boldsymbol{\eta}_t), \ldots, h_q(\boldsymbol{\eta}_t)$ satisfies

$$
\pi_{t1}(\boldsymbol{\theta}) = h_1(\boldsymbol{\eta}_{t1}) = F(\eta_{t1}), \ldots,
$$
$$
\pi_{tj}(\boldsymbol{\theta}) = h_j(\boldsymbol{\eta}_t) = F(\eta_{tj}) - F(\eta_{t(j-1)}), j = 2, \ldots, q,
$$

where $F(x) = \frac{1}{1+\exp(-x)}$. $\Sigma_t(\boldsymbol{\theta})$ is the conditional covariance matrix of $\boldsymbol{Y}_t$ with

$$
\boldsymbol{\Sigma}_t^{(i,j)}(\boldsymbol{\theta}) = \begin{cases} -\pi_{ti}(\boldsymbol{\theta})\pi_{tj}(\boldsymbol{\theta}), & i \neq j, \\ \pi_{ti}(\boldsymbol{\theta})(1 - \pi_{ti}(\theta)), & i = j, \end{cases}
$$

for $i, j = 1, \ldots, q$ [23].

To obtain the existence, consistency and asymptotic normality of the maximum partial likelihood estimator, we give a few assumptions on the the covariate matrices $\{\boldsymbol{Z}_t\}$ and parameter vector $\boldsymbol{\theta}$.

Assumption 1 *The parameter vector $\boldsymbol{\theta} \in \Omega \subseteq \mathbb{R}^p$, where Ω is an open set.*

Assumption 2 *The link function h is twice continuously differentiable, and satisfies $\det(\partial h(\boldsymbol{\eta}_t)/\partial \boldsymbol{\eta}_t) \neq 0$, where $\boldsymbol{\eta}_t = \boldsymbol{Z}_{t-1}'\boldsymbol{\theta}$.*

Assumption 3 *The covariate matrix $\boldsymbol{Z}_{t-1}$ lies almost surely in a non-random compact subset Φ of $\mathbb{R}^{p \times q}$ such that $P\left(\boldsymbol{\lambda}'\left(\sum_{t=1}^{n}\boldsymbol{Z}_{t-1}\boldsymbol{Z}_{t-1}'\right)\boldsymbol{\lambda} > 0\right) = 1$, $\boldsymbol{Z}_{t-1}'\boldsymbol{\theta}$ lies almost surely in the domain H of h for all $\boldsymbol{Z}_{t-1} \in \Phi$ and $\boldsymbol{\theta} \in \Omega$, where $\boldsymbol{\lambda} \in \mathbb{R}^p, \boldsymbol{\lambda} \neq \boldsymbol{0}$.*

Assumptions 1 and 2 ensure that the second derivative of $l(\boldsymbol{\theta})$ is continuous, $\det(\partial \boldsymbol{h}(\boldsymbol{\eta}_t)/\partial \boldsymbol{\eta}_t) \neq 0$ implies that $\boldsymbol{U}_t(\boldsymbol{\theta})$ is not singular (Fokianos and Kedem [24]). From Assumption 3,

$$
\boldsymbol{G}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\boldsymbol{U}_t'(\boldsymbol{\theta})\boldsymbol{Z}_{t-1}'.
$$

is positive definite with probability one [24]. Since the likelihood estimation employs an assumption regarding ergodicity of the joint process $(\boldsymbol{Y}_t^T, \boldsymbol{z}_t^T)^T$ (Fokianos and Truquet [25]), let $\{Y_t\}$ be a time series taking values in a finite set E with cardinal m, and such that

$$
P(Y_t = \omega | Y_{t-1}^-, z) = q(\omega | Y_{t-1}^-, z_{t-1}^-), t \in \mathbb{Z},
$$

where $z = \{z_t, t \in \mathbb{Z}\}$, $Y_{t-1}^- = \{Y_{t-1-j}, j \geq 0\}$, $z_{t-1}^- = \{z_{t-1-j}, j \geq 0\}$, q is a transition kernel. We assume that the applications $(\omega, y, x)| \rightarrow q(\omega | y, x_{t-1}^-)$ are measurable, as applications from $E \times E^m \times \mathcal{D}$ to $(0, 1)$, where $\{x_t, t \in \mathbb{Z}\}$ is a sequence, $x_{t-1}^- = \{x_{t-1-j}, j \geq 0\}$, $\mathcal{D} \in \mathcal{B}(\mathbb{R}^d)^{\otimes \mathbb{Z}}$ is

such that $P(z \in \mathcal{D}) = 1$. Assume that v_1 and v_2 are two probability measures on E, define

$$d_{TV}(v_1, v_2) = \frac{1}{2} \sum_{f \in E} |v_1(f) - v_2(f)|.$$

For $y, y' \in E^m$ and a positive integer s, we write $y \overset{s}{=} y'$ if $y_i = y'_i, 0 \leq i \leq s-1$ (Truquet [26]).

Assumption 4 *The d-dimensional covariate vector $\{z_{t-1}\}$ is stationary and ergodic.*

Assumption 5 *Setting for $s \geq 0$,*

$$b_s = \sup \left\{ d_{TV}(q(\cdot|y, x_t^-), q(\cdot|y', x_t^-)) : (y, y', x) \in E^m \times E^m \times \mathcal{D}, t \in Z, y \overset{s}{=} y' \right\},$$

we have $b_0 < 1$ and $\displaystyle\sum_{s \geq 0} b_s < \infty$.

Assumptions 4 and 5 guarantees that $\left(Y_t^T, z_t^T\right)^T$ is stationary and ergodic [26]. Assumptions 1–5 are required to obtain consistency and asymptotic normality of the maximum likelihood estimator. However, existence of moments for the covariate process is still required to study large sample properties of the maximum likelihood estimator [25]. So we have

Assumption 6 $E|z_{t-1}^{(i)}|^4 < \infty, i = 1, 2, \cdots, d$, *where $z_{t-1}^{(i)}, 1 \leq i \leq d$ are components of vector z_{t-1}.*

The proposed testing procedure

Based on the partial likelihood score process, a test statistic is defined by

$$W_1 = \max_{1 \leq k \leq n} n^{-1/2} \hat{T}_n^{-1/2} S_k(\hat{\boldsymbol{\theta}}_n)$$

where $\hat{T}_n = \frac{1}{n} S_n(\hat{\boldsymbol{\theta}}_n) S'_n(\hat{\boldsymbol{\theta}}_n)$, $\hat{\boldsymbol{\theta}}_n$ is the maximum partial likelihood estimator of θ, which can be obtained by maximizing the partial log-likelihood function (1) (see Fokianos and Kedem [23]).

Under the null hypothesis of no change, we derive the asymptotic distribution of the proposed test statistic.

Theorem 1 *If Assumptions 1–6 and H_0 hold, then we have*

$$W_1 = \max_{1 \leq k \leq n} n^{-1/2} \hat{T}_n^{-1/2} S_k(\hat{\boldsymbol{\theta}}_n) \overset{d}{\to} \sup_{0 < t < 1} B(t),$$

where $\hat{T}_n = \frac{1}{n} S_n(\hat{\boldsymbol{\theta}}_n) S'_n(\hat{\boldsymbol{\theta}}_n)$, $B(t)$ is a p-dimensional vector of independent Brownian bridge, $\overset{d}{\to}$ means convergence in distribution.

Proof: Since $S_n(\hat{\boldsymbol{\theta}}_n) = 0$, we can write

$$n^{-1/2}(S_k(\hat{\boldsymbol{\theta}}_n)) = n^{-1/2}\left(S_k(\hat{\boldsymbol{\theta}}_n) - \frac{k}{n} S_n(\hat{\theta}_n)\right),$$

let $S_k^{(i)}$ denote the i-th element of S_k, $i = 1, 2, \ldots, p$, $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$, then we have

$$n^{-1/2}S_k^{(i)}\left(\hat{\boldsymbol{\theta}}_n\right) = n^{-1/2}\left(S_k^{(i)}(\hat{\boldsymbol{\theta}}_n) - \frac{k}{n}S_n^{(i)}(\hat{\boldsymbol{\theta}}_n)\right)$$

$$= n^{-1/2}\left(S_k^{(i)}(\boldsymbol{\theta}_0) - \frac{k}{n}S_n^{(i)}(\boldsymbol{\theta}_0)\right) - n^{1/2}\sum_{j=1}^{p}(\hat{\boldsymbol{\theta}}_n^{(j)} - \boldsymbol{\theta}_0^{(j)})$$

$$\times \frac{1}{n}\left(\sum_{t=1}^{k}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(i)}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(j)}\right.$$

$$\left. - \frac{k}{n}\sum_{t=1}^{n}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(i)}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(j)}\right) + E_{kn}^{(i)}.$$

Next, it is similar to the proof of Proposition 3 in Gombay et al. [19], we can prove that

$$\max_{1 \le k \le n}\frac{1}{n}\left|\sum_{t=1}^{k}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(i)}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(j)}\right.$$

$$\left. - \frac{k}{n}\sum_{t=1}^{n}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(i)}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(j)}\right| = o_P(1)$$

By Theorem 4.1 of Fokianos and Kedem [23], we get

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n^{(i)} - \boldsymbol{\theta}_0^{(i)}) = O_P(1).$$

The error terms $E_{kn}^{(i)}$ have higher orders of products of $(\hat{\boldsymbol{\theta}}_n^{(i)} - \boldsymbol{\theta}_0^{(i)})$, it can be shown that $E_{kn}^{(i)} = o_P(1)$. According to Proposition 1 (Gombay et al. [19]) and Slutsky's theorem, we get

$$\max_{1 \le k \le n} n^{-1/2}\hat{\boldsymbol{T}}_n^{-1/2}\boldsymbol{S}_k(\hat{\boldsymbol{\theta}}_n) \xrightarrow{d} \sup_{0 < t < 1} B(t)$$

as $n \to \infty$.

Remark 1 When using the above test, if there exists some i, $1 \le i \le p$,

$$\max_{1 \le k \le n} n^{-1/2}\left|(\hat{T}^{-1/2}\boldsymbol{S}_k(\hat{\boldsymbol{\theta}}_n))^{(i)}\right| \ge C(\alpha^*),$$

the null hypothesis is rejected and a change-point occurs, $\alpha^* = 1 - (1 - \alpha)^{1/p}$. Let $B(u)$ be a one-dimensional Brownian bridge, Csörgö and Révész [27] suggested that $C(\alpha^*)$ could be obtained by

$$P\left(\sup_{0 \le u \le 1}|B(u)| \ge x\right) = \sum_{k \ne 0}(-1)^{k+1}\exp(-2k^2x^2).$$

Simulation shows that W_1 has poor performance at the boundaries. In particular, the limiting Brownian bridge is tied down at $t = 0$ and $t = 1$ (meaning $B(0) = B(1) = 0$), and hampers the ability of the test to detect the structural change occurring near the beginning or end of the data. Many authors address this problem by adding a weight function [28]. Therefore, we construct a new test statistic

$$W_2 = \max_{l < \frac{k}{n} < h}\frac{n^{-1}\boldsymbol{\xi}_k'\boldsymbol{\xi}_k}{k/n(1 - k/n)},$$

which is the quadratic form of the efficient score vector and has a weight function, where

$$\xi_k = \left(\left(\hat{T}_n^{-1/2} S_k(\hat{\theta}_n) \right)^{(i_1)}, \ldots, \left(\hat{T}_n^{-1/2} S_k(\hat{\theta}_n) \right)^{(i_j)} \right)',$$

$\{i_1, i_2, \ldots, i_j\} \subset \{1, 2, \ldots, p\}, j = 1, 2, \ldots, p, 0 < l < h < 1.$

Theorem 2 *If Assumptions 1–6 and H_0 hold, then we have*

$$W_2 = \max_{l < \frac{k}{n} < h} \frac{n^{-1} \xi_k' \xi_k}{k/n(1 - k/n)} \xrightarrow{d} \sup_{l < t < h} \frac{\sum_{i=1}^{j} B_i^2(t)}{t(1-t)}$$

for each $0 < l < h < 1$, $B_i(t)$, $i = 1, \ldots, j$ are independent one-dimensional Brownian bridges.

The conclusion of Theorem 2 can be deduced directly from Theorem 1. To obtain the critial values of the asymptotic distribution, Csörgö and Horváth [2] used a result of Vostrikova [29] to show that

$$P \left\{ \sup_{l < t < h} \frac{\sum_{i=1}^{j} B_i^2(t)}{t(1-t)} \geq x \right\} = \frac{x^{j/2} e^{-x/2}}{2^{j/2} \Gamma(j/2)} \left\{ \left(1 - \frac{j}{x} \right) \log \frac{(1-l)h}{l(1-h)} + \frac{4}{x} + O \left(\frac{1}{x^2} \right) \right\}$$

as $x \to \infty$. For example, when $\alpha = 0.05$, $l = 0.05$, $h = 0.95$, $j = 2$, the critical value $C(\alpha) = 13.1$.

Under the alternative hypothesis, there exists a structural change in the model, then we will prove the consistency of the two statistics.

Theorem 3 *Suppose Assumptions 1–6 and H_A hold, if the coefficient changes from θ_0 to θ_0^* at k^*, $\theta_0^{*(j)} = \theta_0^{(j)} + \delta$, $\theta_0^{(j)}$ is the jth component of θ_0, $j \in \{1, 2, \ldots, p\}$, where δ is a constant, $\delta \neq 0$, then we have*

(i).

$$W_1 = \max_{1 \leq k \leq n} n^{-1/2} \left\| \hat{T}_n^{-1/2} S_k(\hat{\theta}_n) \right\| \xrightarrow{P} \infty;$$

(ii).

$$W_2 = \max_{l < \frac{k}{n} < h} \frac{n^{-1} \xi_k' \xi_k}{k/n(1 - k/n)} \xrightarrow{P} \infty,$$

where $0 < l < h < 1$, $\|\cdot\|$ denotes the Euclidean norm of a vector, $\xrightarrow{P}$ means convergence in probability.

Proof: Under the alternative hypothesis $\theta = \theta_0$, $t = 1, 2, \ldots, k^*$, $\theta = \theta_0^*$, $t = k^* + 1, \ldots, n$. Suppose that the coefficient changes from θ_0 to θ_0^* at k^*, $\theta_0^{*(j)} = \theta_0^{(j)} + \delta$, $\theta_0^{(j)}$ is the j-th component of θ_0, $1 < j < p$, where δ is a constant, $\delta \neq 0$.

When $k^* < k < n$,

$$n^{-1/2} S_k(\hat{\theta}_n) = n^{-1/2} (S_{1k}(\hat{\theta}_n) + S_{2k}(\hat{\theta}_n)),$$

where $S_{1k} = \sum_{t=1}^{k^*} Z_{t-1} U_t(\hat{\theta}_n)(Y_t - \pi_t(\hat{\theta}_n))$, $S_{2k} = \sum_{t=k^*+1}^{k} Z_{t-1} U_t(\hat{\theta}_n)(Y_t - \pi_t(\hat{\theta}_n))$. For the ith

component of $S_{1k}(\hat{\boldsymbol{\theta}}_n)$, $1 < i < p$, we have

$$n^{-1/2}S_{1k}^{(i)}(\hat{\boldsymbol{\theta}}_n) = n^{-1/2}S_{1k}^{(i)}(\boldsymbol{\theta}_0) + n^{1/2}\sum_{j=1}^{p}(\hat{\boldsymbol{\theta}}_n^{(j)} - \boldsymbol{\theta}_0^{(j)})$$

$$\times\frac{1}{n}\sum_{t=1}^{k^*}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y} - \pi_t(\boldsymbol{\theta}_0)))^{(i)}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y} - \pi_t(\boldsymbol{\theta}_0)))^{(j)} + E_{1k}^{(i)}$$

$$n^{-1/2}S_{2k}^{(i)}(\hat{\boldsymbol{\theta}}_n) = n^{-1/2}S_{2k}^{(i)}(\boldsymbol{\theta}_0^*) + n^{1/2}\sum_{j=1}^{p}(\hat{\boldsymbol{\theta}}_n^{(j)} - \boldsymbol{\theta}_0^{*(j)})$$

$$\times\frac{1}{n}\sum_{t=k^*+1}^{k}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0^*)(\boldsymbol{Y} - \pi_t(\boldsymbol{\theta}_0^*)))^{(i)}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0^*)(\boldsymbol{Y} - \pi_t(\boldsymbol{\theta}_0^*)))^{(j)} + E_{2k}^{(i)}$$

where $E_{1k}^{(i)}$ has two orders of products of $(\hat{\boldsymbol{\theta}}_n^{(j)} - \boldsymbol{\theta}_0^{(j)})$, $E_{2k}^{(i)}$ has two orders of products of $(\hat{\boldsymbol{\theta}}_n^{(j)} - \boldsymbol{\theta}_0^{*(j)})$. By Theorem 1 we have

$$n^{-1/2}S_{2k}^{(i)}(\boldsymbol{\theta}_0^*) = O_P(1)$$

as $n \to \infty$. Following Assumptions 1–6, we conclude that

$$\frac{1}{n}\sum_{t=k^*+1}^{k}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0^*)(\boldsymbol{Y} - \pi_t(\boldsymbol{\theta}_0^*)))^{(i)}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0^*)(\boldsymbol{Y} - \pi_t(\boldsymbol{\theta}_0^*)))^{(j)} = O_P(1).$$

Since $\delta \neq 0$, we have

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n^{(j)} - \boldsymbol{\theta}_0^{*(j)}) = n^{1/2}(\hat{\boldsymbol{\theta}}_n^{(j)} - \boldsymbol{\theta}_0^{(j)} - \delta) \xrightarrow{P} \infty$$

as $n \to \infty$. When $1 < k < k^*$, the proof is similar. The proof of (ii) is similar to the proof of (i).

Once the null hypothesis is rejected, indicating there may exist a change-point, then we locate the change-point position by

$$\hat{k}^* = \min\left\{k : \max_{1 \le k \le n} n^{-1/2}\left|(\hat{T}_n^{-1/2}S_k(\hat{\theta}_n))^{(i)}\right|, 1 \le i \le p\right\}. \tag{2}$$

The following theorem shows that the change-point estimator $\hat{k}^*$ is consistent for the true change-point k^*, as $n \to \infty$.

Theorem 4 *Let k^* be the true position of change-point under the alternative hypothesis H_A and $\hat{k}^*$ be the estimate of k^* given by* (2). *Under Assumptions 1–6, then $\hat{k}^*$ is consistent to k^*, as $n \to \infty$.*

Proof: First we note that

$$n^{-1/2}S_k^{(i)}\left(\hat{\boldsymbol{\theta}}_n\right) = n^{-1/2}\left(S_k^{(i)}(\hat{\boldsymbol{\theta}}_n) - \frac{k}{n}S_n^{(i)}(\hat{\boldsymbol{\theta}}_n)\right)$$

$$= n^{-1/2}\left(S_k^{(i)}(\boldsymbol{\theta}_0) - \frac{k}{n}S_n^{(i)}(\boldsymbol{\theta}_0)\right) - n^{1/2}\sum_{j=1}^{q+d}(\hat{\boldsymbol{\theta}}_n^{(j)} - \boldsymbol{\theta}_0^{(j)})$$

$$\times\frac{1}{n}\left(\sum_{t=1}^{k}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(i)}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(j)}\right.$$

$$\left. - \frac{k}{n}\sum_{t=1}^{n}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(i)}(\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(j)}\right) + E_{kn}^{(i)},$$

where $i = 1, 2, \ldots, p$. Since

$$S_k^{(i)}(\boldsymbol{\theta}_0) - \frac{k}{n} S_n^{(i)}(\boldsymbol{\theta}_0) = \frac{k(n-k)}{n} \left(\frac{S_k^{(i)}(\boldsymbol{\theta}_0)}{k} - \frac{S_{n-k}^{(i)}(\boldsymbol{\theta}_0)}{n-k} \right),$$

where $S_{n-k}(\boldsymbol{\theta}_0) = \sum\limits_{t=k+1}^{n} \boldsymbol{Z}_{t-1} \boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0))$. And because

$$E(Y_{tj}|\mathcal{F}_{t-1}) = \pi_{tj}(\theta_0), t = 1, 2, \cdots, k^*, j = 1, 2, \cdots, m-1,$$

$$E(Y_{tj}|\mathcal{F}_{t-1}) = \pi_{tj}(\theta_0^*), t = k^* + 1, k^* + 2, \cdots, n, j = 1, 2, \cdots, m-1,$$

Therefore $E\left(S_k^{(i)}(\boldsymbol{\theta}_0) - \frac{k}{n} S_n^{(i)}(\boldsymbol{\theta}_0) \right)$ increases as $k = 1, 2, \cdots, k^*$, and decrease as $k = k^* + 1, k^* + 2, \cdots, n$, then we take (2) as the change-point estimator.

By the proof of Theorem 1, we have

$$-n^{1/2} \sum_{j=1}^{p} (\hat{\boldsymbol{\theta}}_n^{(j)} - \boldsymbol{\theta}_0^{(j)})$$

$$\times \frac{1}{n} \left(\sum_{t=1}^{k} (\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(i)} (\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(j)} \right.$$

$$\left. - \frac{k}{n} \sum_{t=1}^{n} (\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(i)} (\boldsymbol{Z}_{t-1}\boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(j)} \right) = o_P(1)$$

and $E_{kn}^{(i)} = o_P(1)$. By Theorem 2 of Gombay [6], to prove (2) it is enough to show that

$$\lim_{K \to \infty} \limsup_{n \to \infty} P \left\{ \max_{1 < k \le k^* - K} f_k \ge \max_{k^* - K < k \le k^* + K} f_k \right\} = 0, \tag{3}$$

and

$$\lim_{K \to \infty} \limsup_{n \to \infty} P \left\{ \max_{k^* + K < k \le n} f_k \ge \max_{k^* - K < k \le k^* + K} f_k \right\} = 0, \tag{4}$$

where $f_k = n^{-1/2} \left(S_k^{(i)}(\hat{\boldsymbol{\theta}}_n) - \frac{k}{n} S_n^{(i)}(\hat{\boldsymbol{\theta}}_n) \right)$. To prove (3), assume that there exists a constant K, $K < k^*$,

$$P \left\{ \max_{1 < k \le k^* - K} n^{-1/2} \left(S_k^{(i)}(\hat{\boldsymbol{\theta}}_n) - \frac{k}{n} S_n^{(i)}(\hat{\boldsymbol{\theta}}_n) \right) \right.$$

$$\left. > \max_{k^* - K < k \le k^* + K} n^{-1/2} \left(S_k^{(i)}(\hat{\boldsymbol{\theta}}_n) - \frac{k}{n} S_n^{(i)}(\hat{\boldsymbol{\theta}}_n) \right) \right\}$$

$$\le P \left\{ \max_{1 < k \le k^* - K} n^{-1/2} \left(S_k^{(i)}(\hat{\boldsymbol{\theta}}_n) - \frac{k}{n} S_n^{(i)}(\hat{\boldsymbol{\theta}}_n) \right) \right.$$

$$\left. > \max_{k^* - K < k \le k^*} n^{-1/2} \left(S_k^{(i)}(\hat{\boldsymbol{\theta}}_n) - \frac{k}{n} S_n^{(i)}(\hat{\boldsymbol{\theta}}_n) \right) \right\}$$

$$= P \left\{ \exists k : 1 < k \le k^* - K, \max_{1 \le r \le k^* - k} n^{-1/2} \left(S_r^{(i)}(\hat{\boldsymbol{\theta}}_n) - \frac{r}{n} S_n^{(i)}(\hat{\boldsymbol{\theta}}_n) \right) \le 0 \right\}$$

$$\le P \left\{ n^{-1/2} \left(S_K^{(i)}(\hat{\boldsymbol{\theta}}_n) - \frac{K}{n} S_n^{(i)}(\hat{\boldsymbol{\theta}}_n) \right) < 0 \right\}$$

$$= P \left\{ n^{-1/2} \left(S_K^{(i)}(\boldsymbol{\theta}_0) - \frac{K}{n} S_n^{(i)}(\boldsymbol{\theta}_0) \right) + S_K^{*(i)} < 0 \right\},$$

where

$$S_K^{*(i)} = -n^{1/2} \sum_{j=1}^{p} (\hat{\boldsymbol{\theta}}_n^{(j)} - \boldsymbol{\theta}_0^{(j)})$$

$$\times \frac{1}{n} \left(\sum_{t=1}^{k} (\boldsymbol{Z}_{t-1} \boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(i)} (\boldsymbol{Z}_{t-1} \boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(j)} \right.$$

$$\left. - \frac{k}{n} \sum_{t=1}^{n} (\boldsymbol{Z}_{t-1} \boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(i)} (\boldsymbol{Z}_{t-1} \boldsymbol{U}_t(\boldsymbol{\theta}_0)(\boldsymbol{Y}_t - \pi_t(\boldsymbol{\theta}_0)))^{(j)} \right) + E_{kn}^{(i)} = o_p(1).$$

By Theorem 1, choosing $\delta > 0$ arbitrarily

$$\lim_{n \to \infty} P \left\{ n^{-1/2} \left(\boldsymbol{S}_K^{(i)}(\boldsymbol{\theta}_0) - \frac{K}{n} \boldsymbol{S}_n^{(i)}(\boldsymbol{\theta}_0) \right) + \boldsymbol{S}_K^{*(i)} < 0 \right\} < \delta.$$

if K is large enough, so (3) is proven. The proof of (4) is the same by symmetry.

If we consider detecting multiple structural changes in the sequences, we can employ the binary segmentation method [30]. First use the single change test. If H_0 is rejected, then find $\hat{k}^*(1)$ by (2). Next divide the sample into two subsamples $\{Y_t, 1 \le t \le \hat{k}^*(1)\}$ and $\{Y_t, \hat{k}^*(1) > t \le n\}$, and test both subsamples for further changes. One continues this segmentation procedure until no subsamples contain further change-points.

Simulation

To evaluate the finite sample performance of the proposed two test statistics (W_1 and W_2), we first simulate an ordinal time series $\{Y_t\}$ with $m = 3$ categories and length $n = 100, 200, 500, 1000$. The data are generated by

$$\ln \frac{P(Y_t \le 1)}{P(Y_t > 1)} = \alpha_1 - \beta_1 \cos(2\pi t/12) - \beta_2 Y_{(t-1)1} - \beta_3 Y_{(t-1)2},$$

$$\ln \frac{P(Y_t \le 2)}{P(Y_t > 2)} = \alpha_2 - \beta_1 \cos(2\pi t/12) - \beta_2 Y_{(t-1)1} - \beta_3 Y_{(t-1)2},$$

where $\alpha_1 = -0.5$, $\alpha_2 = 0.2$, $(\beta_1, \beta_2, \beta_3)' = (2, 0.5, 1)'$, then the parameter vector $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3)'$. All simulation results are based on 1000 replications at the 0.05 significance level.

Suppose that we are only interested in α_1 and β_1, the others are nuisance parameters. Table 1 shows the empirical size of the two statistics under the null hypothesis H_0. W_1^1 and W_1^2 denote the empirical size of W_1 when testing for change in each of α_1 and β_1, respectively. W_1 and W_2 denote the empirical size of W_1 and W_2 when testing for change in both α_1 and β_1, respectively.

It can be seen from Table 1 that the empirical size increases as the historical sample size n increases. When the sample size $n = 1000$, the empirical size of W_1 and W_2 is close to the

Table 1. The empirical size of W_1 and W_2 under the null hypothesis H_0.

n	100	200	500	1000
W_1^1	0.013	0.014	0.018	0.019
W_1^2	0.01	0.018	0.02	0.014
W_1	0.038	0.027	0.046	0.037
W_2	0.038	0.045	0.033	0.047

Table 2. The empirical power of W_1 and W_2 under the alternative hypothesis $H_A^{(1)}$.

	$k^* = 0.1n$				$k^* = 0.5n$				$k^* = 0.8n$			
n	100	200	500	1000	100	200	500	1000	100	200	500	1000
W_1^1	0.016	0.016	0.032	0.057	0.055	0.092	0.292	0.6	0.027	0.04	0.094	0.208
W_1^2	0.01	0.028	0.089	0.14	0.069	0.186	0.502	0.859	0.027	0.042	0.112	0.33
W_1	0.037	0.059	0.102	0.215	0.117	0.264	0.66	0.946	0.029	0.077	0.213	0.454
W_2	0.084	0.153	0.3	0.541	0.068	0.168	0.569	0.924	0.03	0.064	0.259	0.643

Table 3. The empirical power of W_1 and W_2 under the alternative hypothesis $H_A^{(2)}$.

	$k^* = 0.1n$				$k^* = 0.5n$				$k^* = 0.8n$			
n	100	200	500	1000	100	200	500	1000	100	200	500	1000
W_1^1	0.018	0.021	0.052	0.086	0.084	0.192	0.541	0.857	0.031	0.084	0.184	0.495
W_1^2	0.012	0.017	0.018	0.026	0.016	0.01	0.028	0.028	0.016	0.022	0.026	0.019
W_1	0.028	0.044	0.082	0.098	0.104	0.201	0.545	0.873	0.053	0.067	0.215	0.498
W_2	0.029	0.047	0.086	0.222	0.064	0.127	0.394	0.805	0.064	0.095	0.252	0.565

significance level 0.05. In addition, based on the relation between the probability of type I errors when detecting α_1 or β_1 and the overall probability of type I errors, that is W_1^1, W_1^2 and W_1 should satisfy $1 - W_1 = (1 - W_1^1)(1 - W_1^2)$. The results show that $1 - W_1 \approx (1 - W_1^1)(1 - W_1^2)$, which confirms the above inference.

Under the alternative hypothesis H_A, we consider the following three different situations:

$$H_A^{(1)} : \alpha_1 \text{ changes from } -0.5 \text{ to } -1 \text{ at } k^*, \ \beta_1 \text{ changes from } 2 \text{ to } 3 \text{ at } k^*,$$

$$H_A^{(2)} : \alpha_1 \text{ changes from } -0.5 \text{ to } -1 \text{ at } k^*,$$

$$H_A^{(3)} : \beta_1 \text{ changes from } 2 \text{ to } 3 \text{ at } k^*.$$

where $k^* = 0.1n, \ldots, 0.9n$. Tables 2–4 summarize the empirical power of W_1 and W_2 under the alternative hypotheses $H_A^{(1)}$, $H_A^{(2)}$ and $H_A^{(3)}$ when $k^* = 0.1n, 0.5n, 0.8n$. W_1^1 and W_1^2 denote the empirical power when testing for change in each of α_1 and β_1, respectively. W_1 and W_2 denote the empirical power of W_1 and W_2 when testing for change in both α_1 and β_1. From the simulation results, it can be seen that the empirical power of the two statistics increases with the sample size n, and is close to 1 when $n = 1000$. In addition, The empirical power of the two statistics varies according to different change-point locations, and reaches maximum when $k^* = 0.5n$. Fig 1 describes the empirical power of the two statistics when $k^* = 0.1n, \ldots, 0.9n$. It is

Table 4. The empirical power of W_1 and W_2 under the alternative hypothesis $H_A^{(3)}$.

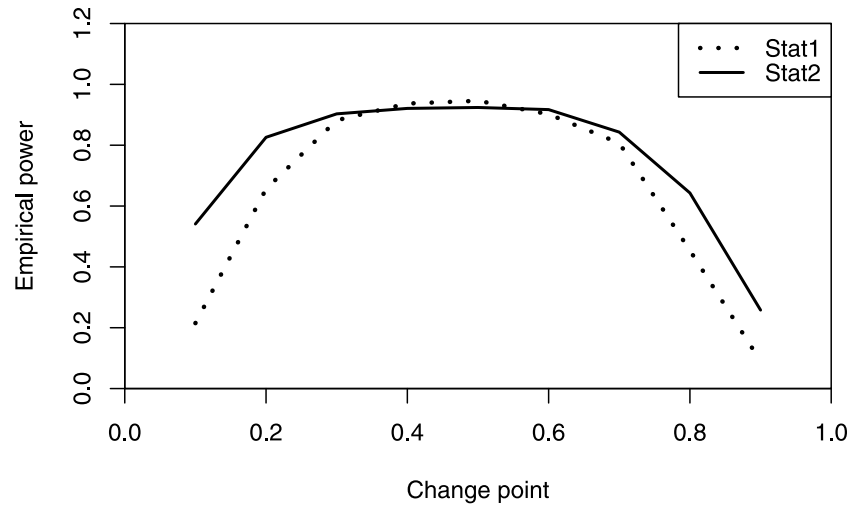|	$k^* = 0.1n$				$k^* = 0.5n$				$k^* = 0.8n$			
n	100	200	500	1000	100	200	500	1000	100	200	500	1000
W_1^1	0.006	0.017	0.019	0.029	0.013	0.026	0.029	0.037	0.011	0.015	0.02	0.029
W_1^2	0.02	0.031	0.077	0.158	0.059	0.19	0.599	0.901	0.014	0.039	0.13	0.415
W_1	0.032	0.059	0.097	0.203	0.078	0.231	0.615	0.921	0.037	0.061	0.178	0.433
W_2	0.1	0.142	0.265	0.436	0.052	0.142	0.461	0.813	0.022	0.036	0.145	0.429

Fig 1. The empirical power of W_1 and W_2 under $H_A^{(1)'}$ when $k^* = 100, 200, \ldots, 900$, $n = 1000$.

showed that the empirical power of W_1 is higher than that of W_2 when the change-point is located at the centre of the data, but W_2 performs better when the change-point is located at the beginning or end of the data.

In simulation for Table 2 both α_1 and β_1 change, whereas in Tables 3 and 4 only α_1 and β_1 changes at different change-points. Tables 3 and 4 indicate that most power stems from the parameter that is changed, which means W_1 that could not only detect change in parameters, but also find the reason for rejecting the null hypothesis.

Application to real data

To illustrate the applicability of our results, we use 1000 sleep data (Y_t) collected from the sleep state measurements of a newborn infant sampled every 30 seconds (Fokianos and Kedem [23]). The sleep states are classified as follows: (1) quiet sleep, (2) indeterminate sleep, (3) active sleep, (4) awake (Fig 2). According to the newborn's sleep pattern, the sleep states have the following order: "(4)" < "(1)" < "(2)" < "(3)", which means $\{Y_t\}$ is an ordinal time series. One goal of analyzing these data is to establish a correct model, and predict the sleep state based on the covariate information. Refer to example 6.3 of [23], $\boldsymbol{Y}_{t-1} = (Y_{(t-1)1}, Y_{(t-1)2}, Y_{(t-1)3})'$ is a significant predictor, which can be considered as a covariate. Then these data could be modeled by a cumulative logistic regression model

$$\ln \frac{P(Y_t \leq 1)}{P(Y_t > 1)} = \alpha_1 + \beta_1 Y_{(t-1)1} + \beta_2 Y_{(t-1)2} + \beta_3 Y_{(t-1)3},$$

$$\ln \frac{P(Y_t \leq 2)}{P(Y_t > 2)} = \alpha_2 + \beta_1 Y_{(t-1)1} + \beta_2 Y_{(t-1)2} + \beta_3 Y_{(t-1)3},$$

$$\ln \frac{P(Y_t \leq 3)}{P(Y_t > 3)} = \alpha_3 + \beta_1 Y_{(t-1)1} + \beta_2 Y_{(t-1)2} + \beta_3 Y_{(t-1)3},$$

where $\alpha_1 = -14.722$, $\alpha_2 = -10.389$, $\alpha_3 = -4.078$, $\beta_1 = 18.663$, $\beta_2 = 12.173$, $\beta_3 = 7.566$.

Let $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3)'$, then testing whether there exists a structural change in $\boldsymbol{\theta}$, the result finds that a structural change occurs in $\boldsymbol{\theta}$ by computing the test statistics W_1 and $W2$.
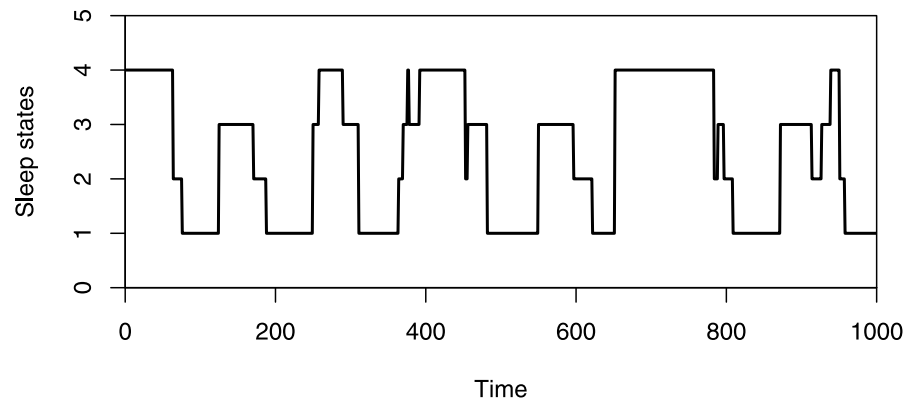
Fig 2. 1000 sleep data (Y_t) collected from the sleep state measurements of a newborn infant sampled every 30 seconds.

After this, using W_1 to check which parameter occurs a structural change, the result shows that there exists a structural change in α_2 at 596. Specifically, the maximum of W_1 is 3.446, and the critical value is 1.35 when $p = 1$, $\alpha = 0.05$, which gives a significant result (Fig 3). Re-estimate the parameters based on the first 596 samples and the last 404 samples, we have $\hat{\alpha}_2 = -10.799$ for the former and $\hat{\alpha}_2 = -8.57$ for the latter. We obtain AIC = 1646.65 for the adjusted model, and AIC = 1652.89 when assuming there is no change-point, which means to improve the model in some extent, so that we can make accurate predictions.

Concluding remark

Cumulative logistic regression model is a generalized linear model, and has a wide application in health care. In this paper, two test statistics based on the efficient score vector are proposed to detect the structural change of cumulative logistic regression model. Under the null hypothesis of no change, we derive the asymptotic distribution of the two test statistics, and prove the consistency under the alternative hypothesis. Furthermore, we prove the consistency of the
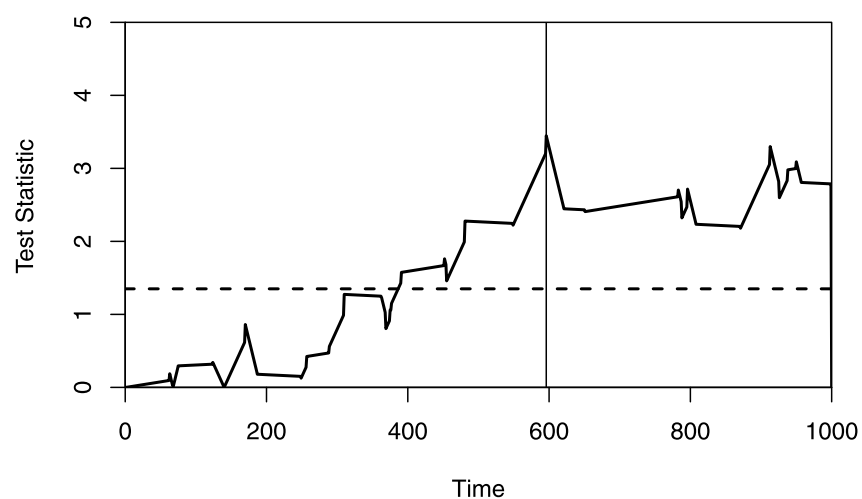

Fig 3. The value of W_1 when testing for α_2, the critical value at $\alpha = 0.05$ is 1.35, and the location of change-point is 596.

change-point estimator, and a binary segmentation procedure is provided for estimating the locations of possible multiple change-points. The finite sample performance is investigate by a monte carlo simulation, the results shows that the empirical size of the two statistics is close to the significance level 0.05, and the empirical power is approximate to 1 when the sample size is large. From the empirical power of view, the two test statistics have different advantages when the change-point occurs at different locations. Furthermore, the proposed statistic W_1 could find the reason for rejecting the null hypothesis. Finally we apply the two test statistics to study 1000 sleep data collected from the sleep state measurements of a newborn infant sampled every 30 seconds, the results shows there exists a structural change in the model.

Supporting information

S1 Table. Simulation data for Table 1.
(XLS)

S2 Table. Simulation data for Table 2.
(XLS)

S3 Table. Simulation data for Table 3.
(XLS)

S4 Table. Simulation data for Table 4.
(XLS)

S5 Table. Simulation data for Fig 1.
(XLS)

S6 Table. 1000 sleep data for Fig 2.
(XLS)

S7 Table. Simulation data for Fig 3.
(XLS)

Author Contributions

Conceptualization: Fuxiao Li, Mengli Hao, Lijuan Yang.

Data curation: Fuxiao Li, Mengli Hao, Lijuan Yang.

Formal analysis: Fuxiao Li.

Funding acquisition: Fuxiao Li.

Investigation: Fuxiao Li, Mengli Hao, Lijuan Yang.

Methodology: Fuxiao Li.

Project administration: Fuxiao Li.

Resources: Fuxiao Li.

Software: Fuxiao Li.

Supervision: Fuxiao Li.

Validation: Fuxiao Li, Mengli Hao, Lijuan Yang.

Visualization: Fuxiao Li, Mengli Hao, Lijuan Yang.

Writing – original draft: Fuxiao Li.

Writing – review & editing: Fuxiao Li, Mengli Hao, Lijuan Yang.

References

1. McCullagh P. Regression Models for Ordinal Data. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 1980; 42: 109–142.

2. Csörgö M, Horváth L. Limit theorems in change-point analysis. New York: John Wiley & Sons Inc, 1997.

3. Bai J, Perron P. Estimating and testing linear models with multiple structural changes. Econometrica. 1998; 66: 47–78. https://doi.org/10.2307/2998540

4. Lee S, Ha J, Na O, Na S. The Cusum Test for Parameter Change in Time Series Models. Scandinavian Journal of Statistics. 2003; 30: 781–796. https://doi.org/10.1111/1467-9469.00364

5. Perron P. Dealing with structural breaks, in: Mills TC, Patterson K, Palgrave handbook of econometrics econometric theory: vol 1. Springer, New York, 2006; pp. 278–352.

6. Gombay E. Change detection in autoregressive time series. Journal of Multivariate Analysis. 2008; 99 (3): 451–464. https://doi.org/10.1016/j.jmva.2007.01.003

7. Wang Y, Wu C, Ji Z, Wang B, Liang Y. Non-Parametric Change-Point Method for Differential Gene Expression Detection. PLoS ONE. 2011; 6(5): e20060. https://doi.org/10.1371/journal.pone.0020060

8. Chen Z, Jin Z, Tian Z, et al. Bootstrap testing multiple changes in persistence for a heavy-tailed sequence. Computational Statistics & Data Analysis. 2012; 55(7): 2303–2316. https://doi.org/10.1016/j.csda.2012.01.011

9. Baranowski R, Chen Y, Fryzlewicz P. Narrowestoverthreshold detection of multiple change points and changepointlike features. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2016; 81: 649–672. https://doi.org/10.1111/rssb.12322

10. Wang G, Zou C, Yin G. Change-point detection in multinomial data with a large number of categories. Annals of Statistics. 2018; 46(5): 2020–2044. https://doi.org/10.1214/17-AOS1610

11. Chen H. Sequential change-point detection based on nearest neighbors. The Annals of Statistics. 2019; 47(3): 1381–1407. https://doi.org/10.1214/18-AOS1718

12. Liu B, Zhou C, Zhang X, Liu Y. A unified data-adaptive framework for high dimensional change point detection. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2020; 82(4): 933–963. https://doi.org/10.1111/rssb.12375

13. Höhle M. Online change-point detection in categorical time series, In: Thomas K, Gerhard T,statistical modelling and regression structures. Heidelberg: Physica-Verlag. 2010; pages 377–397.

14. Plasse J, Adams NM. Multiple change point detection in categorical data streams. Statistics and Computing. 2019; 29: 1109–1125. https://doi.org/10.1007/s11222-019-09858-0

15. Fokianos K, Kedem B. Regression model for time series analysis. Hoboken: John Wiley & Sons. 2002.

16. Xia Z, Guo P, Zhao W. Monitoring Structural Changes in Generalized Linear Models. Communication in Statistics- Theory & Methods. 2009; 38(11): 1927–1947. https://doi.org/10.1080/03610920802549910

17. Hudecová Š. Structural changes in autoregressive models for binary time series. Journal of Statistical Planning and Inference. 2013; 143(10): 1744–1752. https://doi.org/10.1016/j.jspi.2013.05.009

18. Fokianos K, Gombay E, Hussein A. Retrospective change detection for binary time series models. Journal of Statistical Planning and Inference. 2014; 145: 102–112. https://doi.org/10.1016/j.jspi.2013.08.017

19. Gombay E, Li F, Yu H. Retrospective change detection in categorical time series. Communications in Statistics-Simulation and Computation. 2017; 46 (14): 6831–6845.

20. Li F, Chen Z, Xiao Y. Sequential change-point detection in a multinomial logistic regression model. Open Mathematics, 2020; 18(1):807–819. https://doi.org/10.1515/math-2020-0037

21. Gombay E, Serban D. Monitoring parameter change in AR(p) time series models. Journal of Multivariate Analysis. 2009; 100(4): 715–725.

22. Gombay E, Hussein AA, Steiner SH. Monitoring binary outcomes using risk-adjusted charts: a comparative study. Statistics in Medicine. 2011; 30(23): 2815–2826. https://doi.org/10.1002/sim.4305

23. Fokianos K, Kedem B. Regression theory for categorical time series. Statistics Science. 2003; 18(3): 357–376. https://doi.org/10.1214/ss/1076102425

24. Fokianos K, Kedem B. Prediction and classification of non-stationary categorical time series. Journal of Multivariate Analysis. 1998; 67: 277–296. https://doi.org/10.1006/jmva.1998.1765

25. Fokianos K, Truquet L. On categorical time series models with covariates. Stochastic Process & Their Applications. 2019; 129(9): 3446–3462. https://doi.org/10.1016/j.spa.2018.09.012

26. Truquet L. Coupling and perturbation techniques for categorical time series. Bernoulli. 2020; 26(4): 3249–3279. https://doi.org/10.3150/20-BEJ1225

27. Csörgö M, Révész P. Strong Approximations in Probability and Statistics. New York: Academic Press, 1981.

28. Robbins M, Gallagher C, Lund R, et al. Mean shift testing in correlated data. Journal of Time Series Analysis, 2011; 32(5): 498–511. https://doi.org/10.1111/j.1467-9892.2010.00707.x

29. Vostrikova LY. Detection of Disorder in a Wiener Process. Theory of Probability and Its Applications. 1981; 26: 356–362. https://doi.org/10.1137/1126034

30. Xu M, Zhong P S, Wang W. Detecting variance change-points for blocked time series and dependent panel data. Journal of Business & Economic Statistics. 2016, 34(2): 213–226. https://doi.org/10.1080/07350015.2015.1026438