Research Article

Fuxiao Li*, Zhanshou Chen, and Yanting Xiao

Sequential change-point detection in a multinomial logistic regression model

<https://doi.org/10.1515/math-2020-0037>

received July 6, 2019; accepted May 15, 2020

Abstract: Change-point detection in categorical time series has recently gained attention as statistical models incorporating change-points are common in practice, especially in the area of biomedicine. In this article, we propose a sequential change-point detection procedure based on the partial likelihood score process for the detection of changes in the coefficients of multinomial logistic regression model. The asymptotic results are presented under both the null of no change and the alternative of changes in coefficients. We carry out a Monte Carlo experiment to evaluate the empirical size of the proposed procedure as well as its average run length. We illustrate the method by using data on a DNA sequence. Monte Carlo experiments and real data analysis demonstrate the effectiveness of the proposed procedure.

Keywords: categorical time series, sequential change-point detection, multinomial logistic regression model, partial likelihood score process

MSC 2010: 60F15, 62G30

1 Introduction

Structural change is of central importance in many fields of research and data analysis. In the past few decades, studies of change-point detection have attracted a great deal of attention in fields such as statistics, engineering, economics, climatology and bioscience. For surveys, we refer to the studies by Csörgö and Horváth [1], Perron [2], Gombay [3], Chen and Tian [4], Na et al. [5], Zou et al. [6], Ross [7], Robbins et al. [8], Li et al. [9], Cao et al. [10] and Chen [11]. Most of these studies deal with the classical linear regression model either with independent identical distribution errors or assuming some type of dependence. However, the research of detection of change-point in generalized linear models has received increased attention in recent years. Antoch et al. [12] presented some results on testing for changes in generalized linear models based on overall maximum-type test statistic. Zhou and Liang [13] proposed an estimating procedure for estimating the change-point along with other regression coefficients under the generalized linear model framework.

Categorical time series data are frequently encountered in biomedicine, social science and genetics. As generalized linear regression models for categorical time series allow for parsimonious modeling and incorporation of random time-dependent covariates, Fokianos and Kedem [14] suggested the generalized linear model for categorical time series modeling. For change-point detection in categorical time series, there are retrospective procedures which deal with the detection of a structural change within an observed

* **Corresponding author: Fuxiao Li**, Department of Applied Mathematics, Xi'an University of Technology, Xi'an, Shaanxi 710048, People's Republic of China, e-mail: fx_lmzq@163.com

Zhanshou Chen: Department of Mathematics, Qinghai Normal University, Xining, Qinghai 810008, People's Republic of China, e-mail: lmzqxut@qq.com

Yanting Xiao: Department of Applied Mathematics, Xi'an University of Technology, Xi'an, Shaanxi 710048, People's Republic of China, e-mail: fxxut@qq.com

data set of fixed size, whereas sequential procedures check the stability hypothesis each time a new observation is available (see Horváth et al. [15]). Fokianos et al. [16] provided a new statistical procedure based on the partial likelihood score process for the retrospective detection of change in binary time series. Gombay et al. [17] discussed the retrospective detection of change in categorical time series using the same method. Hudecová [18] studied the change-point problem within the framework of autoregressive models for binary time series. Wang et al. [19] proposed a novel change-point detection procedure motivated by high-dimensional homogeneity tests to estimate the locations of multiple change-points in multinomial data with a large number of categories.

The above study of detection of change is retrospective change-point detection, in regard to sequential change-point detection (on-line monitoring, sequential test or priori test) whereby data are not observed at once, but arrive in a sequential manner – one by one, Xia et al. [20] introduced two procedures to sequentially detect structural change in generalized linear models with assuming independence. Höhle [21] proposed a CUSUM control chart method based on the generalized likelihood ratio statistic for sequential change-point detection in regression models for categorical time series.

The goal of this article is to propose a sequential change-point detection procedure based on the partial likelihood score process for the coefficients of multinomial logistic regression model for categorical time series. Score test for the sequential detection of changes in time series models has been studied by many authors. Related work can be found in the studies by Gombay and Serban [22] and Gombay et al. [23]. Compared with the CUSUM detection method, the alternative value of the parameter of the proposed method does not have to be estimated, and in our current model the parameter estimation is complex. Also, the probability of false alarms of the proposed score test is lower.

In this article, we propose a sequential test statistic based on the partial likelihood score process for the coefficients of multinomial logistic regression model. The asymptotic distribution of test statistic is derived under the null hypothesis and the consistency is proven under the alternative hypothesis. We perform Monte Carlo simulations to explore the finite sample performance of the proposed test statistics in terms of empirical size as well as average run lengths. The results show that the proposed test statistic can make the alarms reliable signals of genuine change, and the price is an increased average run length in detection. A real data example is also provided to examine the efficiency of the proposed procedure.

The rest of the article is organized as follows. Section 2 describes the multinomial logistic regression model. Section 3 contains the proposed procedure, necessary assumptions and their asymptotics. In Section 4, Monte Carlo simulations and real data analysis are conducted to investigate the performance of the proposed procedure. Section 5 concludes the article. All proofs of the theorems are gathered in Appendix.

2 Multinomial logistic regression model

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

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

for $t = 1, 2, \dots$ and $j = 1, \dots, q$. Denote the vector of conditional probabilities given $\mathcal{F}_{t-1}$ by $\pi_t = (\pi_{t1}, \dots, \pi_{tq})'$, where

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

for every t , σ field $\mathcal{F}_{t-1}$ is generated by previous observations and covariates $Y_{t-1}, Y_{t-2}, \dots, Z_{t-1}, Z_{t-2}, \dots$, that is,

$$\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, Y_{t-2}, \dots, Z_{t-1}, Z_{t-2}, \dots\},$$

with $p \times q$ matrices $\{\mathbf{Z}_{t-1}\}$, representing the covariate process. Define

$$Y_{tl} = 1 - \sum_{j=1}^q Y_{tj}, \quad \pi_{tl} = 1 - \sum_{j=1}^q \pi_{tj}.$$

Following the definition of generalized linear models, the vector of conditional probabilities is linked to the covariate process through the following equation:

$$\pi_t(\boldsymbol{\beta}) = \begin{pmatrix} \pi_{t1}(\boldsymbol{\beta}) \\ \pi_{t2}(\boldsymbol{\beta}) \\ \vdots \\ \pi_{tq}(\boldsymbol{\beta}) \end{pmatrix} = h(\mathbf{Z}_{t-1}'\boldsymbol{\beta})$$

with $\boldsymbol{\beta}$, a p -dimensional vector of parameters. The inverse link function h is defined on R^q and takes values in R^q as well.

In this article, we investigate the multinomial logistic regression model, which is frequently employed in the analysis of nominal time series (Agresti [24], Section 9.2). Let $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_q)'$, $\boldsymbol{\beta}_j \in R^d$, $j = 1, \dots, q$, then we have

$$\pi_{tj}(\boldsymbol{\beta}) = \frac{\exp(\boldsymbol{\beta}'_j \mathbf{z}_{t-1})}{1 + \sum_{i=1}^q \exp(\boldsymbol{\beta}'_i \mathbf{z}_{t-1})}, \quad j = 1, \dots, q, \quad (1)$$

where $\mathbf{z}_{t-1}$ is the corresponding d -dimensional vector of stochastic time-dependent covariates independent of j . Obviously,

$$\pi_{tl}(\boldsymbol{\beta}) = \frac{1}{1 + \sum_{i=1}^q \exp(\boldsymbol{\beta}'_i \mathbf{z}_{t-1})}.$$

3 Sequential change-point detection

In this section, we propose a sequential change-point detection procedure for the coefficients of multinomial logistic regression model. Following the general paradigm of Chu et al. [25], sequential change-point detection uses the initial time period of length m to estimate a model, and its goal is to verify that the probability of the detection approaches α under the null of no change and one under the alternative of a change in parameters after the initial time period.

We first assume that there is no change in the regression parameter during the first m observations, i.e.,

$$\boldsymbol{\beta} = \boldsymbol{\beta}_0, \quad 1 \leq t \leq m.$$

We are interested in testing the hypothesis

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0, \quad \text{for all } \pi_t(\boldsymbol{\beta}), t = m+1, m+2, \dots, \kappa m,$$

$$H_A : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0, \quad \text{for all } \pi_t(\boldsymbol{\beta}), t = m+1, m+2, \dots, m+k^*, \text{ and } \boldsymbol{\beta} \neq \boldsymbol{\beta}_0 \text{ for } \pi_t, m+k^* < t \leq \kappa m,$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_q)'$, $\boldsymbol{\beta}_0$ is the true value of the parameter vector $\boldsymbol{\beta}$ and unknown, k^* is the unknown time when change-point occurs in some of the regression coefficients, and κ is some fixed positive integer larger than 1. The sequential change-point detection procedure is based on a detecting statistic Γ_{m+k} and a boundary $g(m, k)$, we stop and reject H_0 at

$$\tau(m) = \begin{cases} \inf \{1 < k < N : \Gamma_{m+k} \geq g(m, k)\}, \\ \infty, & \text{if } \Gamma_{m+k} < g(m, k), \end{cases}$$

where $N = (\kappa - 1)m$, this is also called the “closed-end” procedure [26]. The detector and boundary must be chosen such that

$$\lim_{m \rightarrow \infty} P\{\tau(m) < \infty\} = \alpha, \quad \text{under } H_0, \quad (2)$$

$$\lim_{m \rightarrow \infty} P\{\tau(m) < \infty\} = 1, \quad \text{under } H_A. \quad (3)$$

Condition (2) ensures that the probability of a false alarm is asymptotically bounded by α , while condition (3) means that a change-point is detected with probability approaching 1.

Next, we will state the assumptions on the covariate vector $\mathbf{z}_t$ and parameter vector $\boldsymbol{\beta}$. We denote the i th component of a vector $\mathbf{X}$ as $\mathbf{X}^{(i)}$.

Assumption 1. The process $\{\mathbf{z}_t\}$ is ergodic and stationary in the sense that for all t and all $l \geq 0$, $(\mathbf{z}_t, \mathbf{z}_{t+1}, \dots, \mathbf{z}_l)$ has the same distribution as $(\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_{l-t})$.

Assumption 2. All components of the covariate vector have finite fourth moment, that is, $E|\mathbf{z}^{(i)}|^4 < \infty$, $n \rightarrow \infty$.

Assumption 3. For the true value of the parameter vector $\boldsymbol{\beta}_0$, we have $\boldsymbol{\beta}_0 \in \boldsymbol{\Omega}$, where $\boldsymbol{\Omega} \subset \mathbb{R}^p$ is an open subset.

From Assumptions 1–3, we can obtain the stationarity of various powers of $\{\mathbf{z}_t^{(i)}\}$ (see Fokianos et al. [16]),

$$\begin{aligned} \frac{1}{m} \sum_{t=1}^m \mathbf{z}_t^{(i)} \mathbf{z}_t^{(j)} &\xrightarrow{a.s.} E(\mathbf{z}_t^{(i)} \mathbf{z}_t^{(j)}), \quad n \rightarrow \infty, \\ \frac{1}{m} \sum_{t=1}^m \mathbf{z}_t^{(i)} \mathbf{z}_t^{(j)} \mathbf{z}_t^{(l)} &\xrightarrow{a.s.} E(\mathbf{z}_t^{(i)} \mathbf{z}_t^{(j)} \mathbf{z}_t^{(l)}), \quad n \rightarrow \infty, \end{aligned}$$

for all $i, j, l \in \{1, 2, \dots, d\}$, and

$$\frac{1}{m} \sum_{t=1}^m (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta})))^{(i)} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta})))^{(j)} \xrightarrow{a.s.} E(\mathbf{Z}_{t-1} \boldsymbol{\Sigma}(\boldsymbol{\beta}) \mathbf{Z}_{t-1}^T)^{(i,j)}$$

for all $i, j \in \{1, 2, \dots, qd\}$, where

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}) = (\boldsymbol{\sigma}_{ij})_{q \times q}, \quad \boldsymbol{\sigma}_{ij}(\boldsymbol{\beta}) = \begin{cases} -\pi_{ii}(\boldsymbol{\beta})\pi_{ij}(\boldsymbol{\beta}), & i \neq j, \\ \pi_{ii}(\boldsymbol{\beta})(1 - \pi_{ii}(\boldsymbol{\beta})), & i = j. \end{cases}$$

The estimation of the parameter vector $\boldsymbol{\beta}$ in model (1) follows the partial likelihood methodology described in Kedem and Fokianos [27].

The partial likelihood function is

$$\text{PL}(\boldsymbol{\beta}) = \prod_{t=1}^m \prod_{j=1}^l \pi_{tj}^{Y_{tj}}(\boldsymbol{\beta}),$$

so that the partial log-likelihood function is given by

$$l(\boldsymbol{\beta}) = \log \text{PL}(\boldsymbol{\beta}) = \sum_{t=1}^m \sum_{j=1}^l Y_{tj} \log \pi_{tj}(\boldsymbol{\beta}),$$

the partial score process is defined through the partial sum

$$\mathbf{S}_m(\boldsymbol{\beta}) = \nabla l(\boldsymbol{\beta}) = \sum_{t=1}^m \mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta})). \quad (4)$$

The maximum partial likelihood estimator denoted by $\hat{\boldsymbol{\beta}}_m$ is a solution of the score equations $\mathbf{S}_m(\boldsymbol{\beta}) = 0$, and its asymptotic properties have been studied by Kedem and Fokianos [27].

The following results show that the score vector behaves, approximately, as a vector of the Wiener process.

Lemma 3.1. *Under Assumptions 1–3, there exists a Wiener process $(\mathbf{W}(t))_{t \geq 0}$ with covariate matrix $\mathbf{T}$ such that if $\boldsymbol{\beta}$ is the true vector of coefficients, then the score vector admits the following approximation:*

$$\|\mathbf{S}_m(\boldsymbol{\beta}) - \mathbf{W}(m)\| = O(m^{1/2-\lambda}) \quad a.s.,$$

for some constant $\lambda > 0$, $\mathbf{S}_m(\boldsymbol{\beta})$ is defined by (4), and $\mathbf{T}$ is the limit of the sample information matrix

$$\frac{1}{m} \mathbf{S}_m(\boldsymbol{\beta}) \mathbf{S}_m'(\boldsymbol{\beta}) \xrightarrow{a.s.} \mathbf{T}. \quad (5)$$

Let $\hat{\boldsymbol{\beta}}_m$ be the maximum partial likelihood estimator of the true value of the parameter $\boldsymbol{\beta}$.

Lemma 3.2. *Under Assumptions 1–3,*

$$\hat{\boldsymbol{\beta}}_m \xrightarrow{P} \boldsymbol{\beta}, \quad m \rightarrow \infty,$$

and

$$m^{1/2} \mathbf{T}^{1/2} (\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \mathbf{I}), \quad m \rightarrow \infty,$$

where $\mathbf{I} = \mathbf{I}_{p \times p}$ is the identity matrix, $p = qd$, $\mathbf{T}$ is given by (5).

Next, we consider the asymptotics under the null hypothesis and alternative hypothesis.

Theorem 3.1. *Suppose that Assumptions 1–3 and H_0 hold,*

(i) *If $g(s) = c g_1(s)$, $s \in (1, \infty)$, where g_1 is a given continuous real-valued function with $\inf_{1 < s < \infty} g_1(s) > 0$, then*

$$\begin{aligned} \lim_{m \rightarrow \infty} P\{\tau(m) < \infty\} &= \lim_{m \rightarrow \infty} P\left\{ \max_{1 \leq k \leq N} \|\hat{\Gamma}(m, k)\|_{\infty} \geq g\left(\frac{m+k}{m}\right) \right\} \\ &= \lim_{m \rightarrow \infty} P\left\{ \max_{1 \leq k \leq N} \|(\kappa m)^{-1/2} \hat{\mathbf{T}}^{-1/2} \mathbf{S}_{m+k}(\hat{\boldsymbol{\beta}}_m)\|_{\infty} \geq g\left(\frac{m+k}{m}\right) \right\} \\ &= P\left\{ \sup_{1 < s < \kappa} \frac{\|\kappa^{-1/2}(\mathbf{W}(s) - s\mathbf{W}(1))\|_{\infty}}{g_1(s)} \geq c \right\}. \end{aligned}$$

(ii) *If g is a constant, $g(s) \equiv c$, then*

$$\begin{aligned} \lim_{m \rightarrow \infty} P\{\tau(m) < \infty\} &= \lim_{m \rightarrow \infty} P\left\{ \max_{1 \leq k \leq N} \|\hat{\Gamma}(m, k)\|_{\infty} \geq c \right\} \\ &= \lim_{m \rightarrow \infty} P\left\{ \max_{1 \leq k \leq N} \|(\kappa m)^{-1/2} \hat{\mathbf{T}}^{-1/2} \mathbf{S}_{m+k}(\hat{\boldsymbol{\beta}}_m)\|_{\infty} \geq c \right\} \\ &= P\left\{ \sup_{1 < s < \kappa} \|\kappa^{-1/2}(\mathbf{W}(s) - s\mathbf{W}(1))\|_{\infty} \geq c \right\}, \end{aligned}$$

where $N = (\kappa - 1)m$, κ is some fixed positive integer larger than 1,

$$\hat{\Gamma}(m, k) = (\kappa m)^{-1/2} \hat{\mathbf{T}}^{-1/2} \mathbf{S}_{m+k}(\hat{\boldsymbol{\beta}}_m), \quad \hat{\mathbf{T}} = \frac{1}{m} \mathbf{S}_m(\hat{\boldsymbol{\beta}}_m) \mathbf{S}_m'(\hat{\boldsymbol{\beta}}_m),$$

$$\mathbf{S}_{m+k}(\hat{\boldsymbol{\beta}}_m) = \sum_{s=1}^{m+k} \mathbf{Z}_{s-1}(\mathbf{Y}_s - \boldsymbol{\pi}_s(\hat{\boldsymbol{\beta}}_m))$$

$c = c(\alpha)$ is the $(1 - \alpha)$ quantile point of asymptotic distribution, α is the significance level, $\mathbf{W}(s)$ is a $p = qd$ -dimensional vector of the independent Wiener process and $\|\cdot\|_{\infty}$ denotes the maximum norm.

Theorem 3.2. Suppose Assumptions 1–3 and H_A hold, if the coefficient changes from β_0 to β_0^* at $m + k^*$, $\beta_0^{*(j)} = \beta_0^{(j)} + \delta$, $\beta_0^{(j)}$ is the j th component of β_0 , $j \in \{1, 2, \dots, qd\}$, where δ is a constant, $\delta \neq 0$, then we have

$$\max_{1 \leq k \leq N} \|(\kappa m)^{-1/2} \hat{T}^{-1/2} \mathbf{S}_{m+k}(\hat{\beta}_m)\|_{\infty} \xrightarrow{P} \infty$$

as $m \rightarrow \infty$, where N , $\hat{T}^{-1/2}$ and $\mathbf{S}_{m+k}(\hat{\beta}_m)$ are defined in Theorem 3.1.

Proofs of Theorems 3.1 and 3.2 are postponed to Appendix.

4 Simulations and real data analysis

4.1 Simulations

In this section, we considered a multinomial logistic regression model with $l = 3$ categories and length κm , where the history data set $m = 100, 200, 500$ and a monitoring length $N = (\kappa - 1)m$ with $\kappa = 3, 5, 7$. The data were generated according to the model

$$\begin{aligned} \log\left(\frac{\pi_{t1}}{\pi_{t3}}\right) &= \beta_1' z_{t-1} = \beta_{10} + \beta_{11} \cos(2\pi t/12) + \beta_{12} Y_{(t-1)1}, \\ \log\left(\frac{\pi_{t2}}{\pi_{t3}}\right) &= \beta_2' z_{t-1} = \beta_{20} + \beta_{21} \cos(2\pi t/12) + \beta_{22} Y_{(t-1)1}, \end{aligned}$$

with $\beta_1 = (\beta_{10}, \beta_{11}, \beta_{12})' = (0.3, 1.25, 0.5)'$, $\beta_2 = (\beta_{20}, \beta_{21}, \beta_{22})' = (-0.2, -2, -0.75)'$ and $\mathbf{z}_{t-1} = (1, \cos(2\pi t/12), Y_{(t-1)1})$. The null hypothesis is rejected at level α at the first time k such that $\hat{\Gamma}(m, k) > g\left(\frac{m+k}{m}\right)$, let $g_1(s) = \sqrt{s}$, critical values c are determined by

$$P\left\{\sup_{1 \leq s \leq \kappa} \frac{\|\kappa^{-1/2}(\mathbf{W}(s) - s\mathbf{W}(1))\|_{\infty}}{\sqrt{s}} \geq c\right\} = \alpha$$

and

$$P\left\{\sup_{1 \leq s \leq \kappa} \|\mathbf{W}(s) - s\mathbf{W}(1)\|_{\infty} \geq c\right\} = \alpha,$$

which could be obtained by Monte Carlo simulation.

Theorems 3.1 and 3.2 are for the case when all parameters are of interest. In some applications this is not the case. It is possible that some parameters are of special concern, and the others are nuisance parameters. In this simulation, we focused on β_{10} and β_{20} , which means $p = 2$. Table 1 reports the simulated empirical sizes for the proposed procedure based on detecting statistic $\hat{\Gamma}(m, k)$, and the significance level is chosen as 5%. It can be seen that the results are close to the significance level 5% in most cases.

Table 1: Empirical sizes of tests based on $\hat{\Gamma}(m, k)$ for the 5% significance level

m	$g(s) \equiv c$			$g(s) = c\sqrt{s}$		
	$N = 2m$	$N = 4m$	$N = 6m$	$N = 2m$	$N = 4m$	$N = 6m$
100	0.054	0.053	0.054	0.048	0.042	0.056
200	0.062	0.033	0.036	0.042	0.06	0.056
500	0.072	0.038	0.058	0.044	0.054	0.054

Next, we considered the following alternative hypotheses:

$$H_A : \beta_{10} \text{ changes from } 0.3 \text{ to } 1.5 \text{ and } \beta_{20} \text{ changes from } -0.2 \text{ to } -1.5 \text{ at } m + k^*,$$

where $k^* = 0.01N, 0.05N, 0.1N, 0.2N, 0.3N$. H_A means that there exists a change-point in the regression coefficients. To evaluate the performance of the proposed score test, we made a comparison with the CUSUM detection method (Höhle [21]). The threshold of the CUSUM detection method could be obtained by simulation of the run length. In the control state, suppose that the average run length is set to 370, and then the threshold is selected as 1.7. Monte Carlo simulations were performed in R and all simulations were based on 2,000 replications.

Tables 2–5 show the conditional power, the first and third quartiles (Q_1, Q_3), the median Q_2 , the mean (ARL), the maximum of the distribution of the run lengths and the probability of false alarms (P_f). The conditional power is defined as the proportion of signals occurring after $m + k^*$ but before the truncation point κm . The probability of false alarms is obtained as the proportion of signals occurring before the change point $m + k^*$, denoted by P_f . Let $\bar{k} = m + k^*$, $ARL \geq \bar{k}$ is the average run length for runs stopping on or after $\bar{k}$.

From Tables 2–5, it can be seen that the conditional power of score test increases as the sample size increases and decreases as the change-point occurs late. The general relation between power and change-point location in hypothesis testing is that a change-point occurs after a longer period of monitoring and tends to have a lower power. For the boundaries $g(s) \equiv c$ and $g(s) = c\sqrt{s}$, when $m = 100$, the conditional power of the former is higher than that of the latter, and when $m = 200$, the conditional power of two boundaries is almost 1.

Next, we investigate the numerical characteristics of run lengths, Q_1, Q_2, ARL, Q_3 and maximum of the distribution of run lengths increase with the increase in sample size. When the boundary $g(s) \equiv c$, these numerical characteristics are slightly influenced by the change-point location. For instance, when $m = 100, 200$, the ARLs are around 265 and 357, respectively. However, when the boundary $g(s) = c\sqrt{s}$, these numerical characteristics increase as the change-point occurs late. For instance, when $m = 100$, $k^* = 0.01N, 0.3N$, the ARLs are 258.65 and 330.75, respectively. In addition, compared with the boundary $g(s) \equiv c$, the ARLs of test statistic with $g(s) = c\sqrt{s}$ are shorter when the change-point is at the beginning of the series being monitored, but they are longer when the change-point occurs late. Finally, the probability of false alarm P_f in two cases is almost zero in most cases.

Table 6 illustrates the conditional power, Q_1, Q_2, ARL, Q_3 , maximum of the distribution of run lengths and P_f for the CUSUM detection method. It can be seen from Table 6 that the conditional power and ARLs decrease as change-point occurs late. When the change-point is at the beginning of the series being monitored, the condition power is high, the ARLs are short, and the probability of false alarms is low. However, when the change-point occurs after a period of monitoring, the conditional power decreases rapidly, and the probability of false alarms grows rapidly. The monitoring may stop before the change-point $m + k^*$, thus the ARLs are negative when change-point occurs late. In comparison with the proposed score test, the ARLs are shorter, but the condition power is lower, and the probability of false alarms is higher.

To summarize, the score test has good control over its type I errors (empirical size), the price is longer ARL. The CUSUM detection method has shorter ARL, but the probability of false alarms grows rapidly when the change-point occurs late.

Table 2: Conditional powers, average of run lengths and probability of false alarms for score test with $g(s) \equiv c$ under H_A when $m = 100, N = 400$

k^*	Power	Q_1	Q_2	ARL	Q_3	Max	$ARL \geq \bar{k}$	P_f
0.01N	0.922	219	263	271.25	321	396	271.25	0.000
0.05N	0.891	218	269	268.93	320	380	268.93	0.000
0.10N	0.855	214	265	265.98	322	360	265.98	0.000
0.20N	0.694	212	268	257.71	320	320	257.71	0.000
0.30N	0.569	206	264	237.63	280	280	239.23	0.001

Table 3: Conditional powers, average of run lengths and probability of false alarms for score test with $g(s) = c\sqrt{s}$ under H_A when $m = 100, N = 400$

k^*	Power	Q_1	Q_2	ARL	Q_3	Max	$ARL \geq \bar{k}$	P_t
$0.01N$	0.858	191	250.5	258.65	337	399	258.65	0.000
$0.05N$	0.808	204	276	277.5	359	395	277.5	0.000
$0.10N$	0.756	211	283	285.28	377	390	285.28	0.000
$0.20N$	0.598	270	344	317.76	380	380	318.83	0.004
$0.30N$	0.440	300	370	330.75	370	370	330.24	0.002

Table 4: Conditional powers, average of run lengths and probability of false alarms for score test with $g(s) \equiv c$ under H_A when $m = 200, N = 800$

k^*	Power	Q_1	Q_2	ARL	Q_3	Max	$ARL \geq \bar{k}$	P_t
$0.01N$	1.000	306	351	358.33	400	735	358.33	0.000
$0.05N$	1.000	304	349.5	357.6	409	728	357.6	0.000
$0.10N$	1.000	300	350	356.04	408	668	356.04	0.000
$0.20N$	0.993	294	352.5	360.56	424	640	360.56	0.000
$0.30N$	0.958	283	348	356.70	424	560	356.7	0.000

Table 5: Conditional powers, average of run lengths and probability of false alarms for score test with $g(s) = c\sqrt{s}$ under H_A when $m = 200, N = 800$

k^*	Power	Q_1	Q_2	ARL	Q_3	Max	$ARL \geq \bar{k}$	P_t
$0.01N$	1.000	177	214	225.25	259	617	225.25	0.000
$0.05N$	1.000	206	250.5	259.74	301	628	259.74	0.000
$0.10N$	1.000	242	291	301.05	348	605	301.05	0.000
$0.20N$	0.996	309	386	382.35	440	760	382.35	0.000
$0.30N$	0.976	385	451	463.37	525	740	464.63	0.004

Table 6: Conditional powers, average of run lengths and probability of false alarms for CUSUM detection method under H_A when $m = 100, N = 400$

k^*	Power	Q_1	Q_2	ARL	Q_3	Max	$ARL \geq \bar{k}$	P_t
$0.01N$	0.886	47.8	118	153.64	244	396	153.79	0.01
$0.05N$	0.651	-3	59	109.21	184	389	150.71	0.261
$0.10N$	0.476	-17.2	2	56.22	94.2	360	125.96	0.481
$0.20N$	0.276	-56	-32	-1.215	16	320	103.81	0.704
$0.30N$	0.194	-95	-68	-42.91	-21	280	77.3	0.795

4.2 Application to real data

In this section, we applied the proposed procedure to a DNA sequence, which was obtained from the gene BNRF1 of the Epstein-Barr Virus (Shumway and Stoffer [28]), and contains $n = 1,197$ observations. Let A, G, C and T denote adenine, guanine, cytosine and thymine nucleotides, respectively, then a sequence of letters A, G, C and T can be viewed as a nominal categorical time series. Fokianos and Kedem (2002) [27] argued in detail that the most appropriate model for these data is the multinomial logistic model with covariate vector $\mathbf{z}_{t-1} = (1, Y_{(t-1)1}, Y_{(t-1)2}, Y_{(t-1)3}, Y_{(t-3)1}, Y_{(t-3)2}, Y_{(t-3)3})'$ and link function

Figure 1: Regression coefficient β_{34} is monitored at $\alpha = 0.1$ level of significance.

$$\begin{aligned}\log it(\pi_{t1}) &= \log \frac{\pi_{t1}}{1 - \pi_{t1} - \pi_{t2} - \pi_{t3}} \\ &= \beta_{10} + \beta_{11} Y_{(t-1)1} + \beta_{12} Y_{(t-1)2} + \beta_{13} Y_{(t-1)3} + \beta_{14} Y_{(t-3)1} + \beta_{15} Y_{(t-3)2} + \beta_{16} Y_{(t-3)3}, \\ \log it(\pi_{t2}) &= \log \frac{\pi_{t2}}{1 - \pi_{t1} - \pi_{t2} - \pi_{t3}} \\ &= \beta_{20} + \beta_{21} Y_{(t-1)1} + \beta_{22} Y_{(t-1)2} + \beta_{23} Y_{(t-1)3} + \beta_{24} Y_{(t-3)1} + \beta_{25} Y_{(t-3)2} + \beta_{26} Y_{(t-3)3}, \\ \log it(\pi_{t3}) &= \log \frac{\pi_{t3}}{1 - \pi_{t1} - \pi_{t2} - \pi_{t3}} \\ &= \beta_{30} + \beta_{31} Y_{(t-1)1} + \beta_{32} Y_{(t-1)2} + \beta_{33} Y_{(t-1)3} + \beta_{34} Y_{(t-3)1} + \beta_{35} Y_{(t-3)2} + \beta_{36} Y_{(t-3)3},\end{aligned}$$

where β has 21 components. Let the first 400 samples be the historical sample, then monitor the remaining sample one by one (401–1,197). The proposed score test statistic was computed and graphically reported in Figure 1. The results show that there exists a structural change, which occurs at 738 for β_{34} .

To evaluate the validity of the above results, we analyzed the data by the retrospective change point detection procedure (Gombay et al. [17]), and the results show that there exists a structural change at 701. It is also demonstrated that there exists a certain delay, but the proposed score test is effective.

5 Concluding remarks

In this article, we propose a sequential change-point detection procedure based on the partial likelihood score process to detect the coefficients of multinomial logistic regression model. The asymptotic properties of score test statistic are derived under both the null of no change and the alternative of changes in coefficients. To evaluate the finite sample performance of the proposed score test statistic, we conduct a Monte Carlo simulation; simulation results show that the score test method can signal a genuine change reliably, and the price is an increased average run length in detection.

In Monte Carlo simulation, we find that the proposed score test is suitable for the case that the amount of structural change is relatively large and is not sensitive for small deviations. Thus in our simulation, we assume that β_{10} changes from 0.3 to 1.5 and β_{20} changes from -0.2 to -1.5 under the alternative hypotheses. When the amount of structural change is small, the conditional power is lower, and the ARL is longer, so how to establish a more effective test statistic is the next research subject.

Acknowledgement(s): The authors are grateful to the editors and referees for useful suggestions and helpful comments for improving the article. This work was supported by the National Nature Science

Foundation of China (Program No. 11801438), the Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2018JQ1089), Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No.18JK0552) and the Innovation Capability Support Program of Shaanxi (Program No. 2020PT-023).

Conflict of Interest: The authors report no potential conflict of interest.

References

- [1] Miklós Csörgö and Lajos Horváth, *Limit Theorems in Change-Point Analysis*, John Wiley & Sons Inc., New York, 1997.
- [2] Pierre Perron, *Dealing with structural breaks*, Palgrave Handbook of Econometrics **1** (2006), 278–352, DOI: 10.1016/j.gfj.2006.04.004.
- [3] Edit Gombay, *Change detection in autoregressive time series*, J. Multivariate Anal. **99** (2008), no. 3, 451–464, DOI: 10.1016/j.jmva.2007.01.003.
- [4] Zhanshou Chen and Zheng Tian, *Modified procedures for change point monitoring in linear models*, Math. Comput. Simulat. **81** (2010), no. 1, 62–75, DOI: 10.1016/j.matcom.2010.06.021.
- [5] Okyoung Na, Youngmi Lee, and Sangyeol Lee, *Monitoring parameter change in time series models*, Stat. Methods Appl. **20** (2011), no. 2, 171–199, DOI: 10.1007/s10260-011-0162-3.
- [6] Changliang Zou, Guosheng Yin, Long Feng, and Zhaojun Wang, *Nonparametric maximum likelihood approach to multiple change-point problems*, Ann. Stat. **42** (2014), no. 3, 970–1002, DOI: 10.1214/14-aos1210.
- [7] J. Ross Gordon, *Parametric and nonparametric sequential change detection in R: The cpm package*, J. Stat. Softw. **66** (2015), no. 3, 1–20, DOI: 10.18637/jss.v066.i03.
- [8] Michael W. Robbins, Colin M. Gallagher, and Robert B. Lund, *A general regression change point test for time series data*, J. Am. Stat. Assoc. **111** (2016), no. 514, 670–683, DOI: 10.1080/01621459.2015.1029130.
- [9] Fuxiao Li, Zheng Tian, and Zhanshou Chen, *Monitoring distributional changes of squared residuals in GARCH models*, Comm. Statist. Theory Methods **46** (2017), no. 1, 354–372, DOI: 10.1080/03610926.2014.995819.
- [10] Yang Cao, Liyan Xie, Yao Xie, and Huan Xu, *Sequential change-point detection via online convex optimization*, Entropy **20** (2018), no. 2, 108, DOI: 10.3390/e20020108.
- [11] Hao Chen, *Sequential change-point detection based on nearest neighbors*, Ann. Stat. **47** (2019), no. 3, 1381–1407, DOI: 10.1214/18-aos1718.
- [12] Jaromír Antoch, Gérard Gregoire, and Daniela Jarušková, *Detection of structural changes in generalized linear models*, Stat. Probabil. Lett. **69** (2004), no. 3, 315–332, DOI: 10.1016/j.spl.2004.06.028.
- [13] Hongling Zhou and Kung-Yee Liang, *On estimating the change point in generalized linear models*, IMS Collections Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen **1** (2008), 305–320, DOI: 10.1214/193940307000000239.
- [14] Konstantinos Fokianos and Benjamin Kedem, *Regression theory for categorical time series*, Stat. Sci. **18** (2003), no. 3, 357–376, DOI: 10.1214/ss/1076102425.
- [15] Lajos Horváth, Marie Hušková, Piotr Kokoszka, and Josef Steinebach, *Monitoring changes in linear models*, J. Stat. Plan. Infer. **126** (2004), no. 1, 225–251, DOI: 10.1016/j.jspi.2003.07.014.
- [16] Konstantinos Fokianos, Edit Gombay, and Abdulkadir Hussein, *Retrospective change detection for binary time series models*, J. Stat. Plan. Infer. **145** (2014), 102–112, DOI: 10.1016/j.jspi.2013.08.017.
- [17] Edit Gombay, Fuxiao Li, and Hao Yu, *Retrospective change detection in categorical time series*, Comm. Statist. Simulation Comput. **46** (2017), no. 14, 6831–6845, DOI: 10.1080/03610926.2015.1137595.
- [18] Šárka Hudecová, *Structural changes in autoregressive models for binary time series*, J. Stat. Plan. Infer. **143** (2013), no. 10, 1744–1752, DOI: 10.1016/j.jspi.2013.05.009.
- [19] Guanghui Wang, Changliang Zou, and Guosheng Yin, *Change-point detection in multinomial data with a large number of categories*, Ann. Stat. **46** (2018), no. 5, 2020–2044, DOI: 10.1214/17-AOS1610.
- [20] Zhiming Xia, Pengjiang Guo, and Wenzhi Zhao, *Monitoring structural changes in generalized linear models*, Comm. Statist. Theory Methods **38** (2009), no. 11, 1927–1947, DOI: 10.1080/03610920802549910.
- [21] Michael Höhle, *Online change-point detection in categorical time series*, in: T. Kneib, G. Tutz (eds), Statistical Modelling and Regression Structures, Physica, Springer-Verlag Berlin Heidelberg, 2010, pp. 377–397, DOI: 10.1007/978-3-7908-2413-1_20.
- [22] Edit Gombay and Daniel Serban, *Monitoring parameter change in AR(p) time series models*, J. Multivariate Anal. **100** (2009), no. 4, 715–725, DOI: 10.1016/j.jmva.2008.08.005.
- [23] Edit Gombay, Abdulkadir A. Hussein, and Stefan H. Steiner, *Monitoring binary outcomes using risk-adjusted charts: a comparative study*, Stat. Med. **30** (2011), no. 23, 2815–2826, DOI: 10.1002/sim.4305.
- [24] Alan Agresti, *Categorical Data Analysis*, Wiley, New York, 1990.

- [25] Chia-Shang James Chu, Maxwell Stinchcombe, and Halbert White, *Monitoring structural change*, *Econometrica* **64** (1996), no. 5, 1045–1065, DOI: 10.2307/2171955.
- [26] Zdeněk Hlávka, Marie Hušková, Claudia Kirch, and Simos G. Meitanis, *Monitoring changes in the error distribution of autoregressive models based on Fourier methods*, *Test* **21** (2012), 605–634, DOI: 10.1007/s11749-011-0265-z.
- [27] Konstantinos Fokianos and Benjamin Kedem, *Regression model for time series analysis*, John Wiley & Sons, Hoboken, 2002.
- [28] Robert H. Shumway and David S. Stoffer, *Time Series Analysis and its Applications*, Springer, New York, 2002.

Appendix

Proofs of Lemmas 3.1 and 3.2 are similar to Propositions 1 and 2 of Gombay et al. [17].

Proof of Theorem 3.1. The proof can be done along the lines of retrospective change point detection as in Gombay et al. [17].

First, since $\mathbf{S}_m(\hat{\boldsymbol{\beta}}_m) = \mathbf{0}$, we can write

$$(\kappa m)^{-1/2}(\mathbf{S}_k(\hat{\boldsymbol{\beta}}_m)) = (\kappa m)^{-1/2}\left(\mathbf{S}_k(\hat{\boldsymbol{\beta}}_m) - \frac{k}{m}\mathbf{S}_m(\hat{\boldsymbol{\beta}}_m)\right).$$

For the i th component of the score vector, $\mathbf{S}_k^{(i)}$, $i = 1, \dots, qd$, with $\boldsymbol{\beta}_0$ denoting the true value of the parameter, we have

$$\begin{aligned} (\kappa m)^{-1/2}\mathbf{S}_k^{(i)}(\hat{\boldsymbol{\beta}}_m) &= (\kappa m)^{-1/2}\left(\mathbf{S}_k^{(i)}(\hat{\boldsymbol{\beta}}_m) - \frac{k}{m}\mathbf{S}_m^{(i)}(\hat{\boldsymbol{\beta}}_m)\right) = (\kappa m)^{-1/2}\left(\mathbf{S}_k^{(i)}(\boldsymbol{\beta}_0) - \frac{k}{m}\mathbf{S}_m^{(i)}(\boldsymbol{\beta}_0)\right) \\ &\quad + (\kappa m)^{1/2}\sum_{j=1}^{qd}(\hat{\boldsymbol{\beta}}_m^{(j)} - \boldsymbol{\beta}_0^{(j)})\frac{1}{\kappa m}\left(\sum_{t=1}^k(\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)}(\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)}\right. \\ &\quad \left. - \frac{k}{m}\sum_{t=1}^m(\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)}(\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)}\right) + \mathbf{E}_{k,m}^{(i)}, \end{aligned}$$

where $\mathbf{E}_{k,m}^{(i)}$ have two orders of products of $(\hat{\boldsymbol{\beta}}_m^{(j)} - \boldsymbol{\beta}_0^{(j)})$.

Next, we show that

$$\begin{aligned} \max_{m+1 \leq k \leq \kappa m} \frac{1}{\kappa m} \left| \sum_{t=1}^k (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)}(\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} \right. \\ \left. - \frac{k}{m} \sum_{t=1}^m (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)}(\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} \right| = o_p(1). \end{aligned} \quad (6)$$

By Assumptions 1 and 2, we have as $k \rightarrow \infty$,

$$\frac{1}{k} \sum_{t=1}^k ((\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)} \mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} - \mathbf{E}(\mathbf{Z}_{t-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)\mathbf{Z}_{t-1}^T)^{(i,j)}) \xrightarrow{a.s.} 0,$$

then for every $\varepsilon > 0$,

$$\lim_{k \rightarrow \infty} P \left\{ \left| \frac{1}{k} \sum_{t=1}^k ((\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)}(\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} - \mathbf{E}(\mathbf{Z}_{t-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)\mathbf{Z}_{t-1}^T)^{(i,j)}) \right| > \varepsilon \right\} = 0,$$

hence, as $m \rightarrow \infty$,

$$\max_{1 \leq k \leq \kappa m} \left| \frac{1}{k} \sum_{t=1}^k ((\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)}(\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} - \mathbf{E}(\mathbf{Z}_{t-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)\mathbf{Z}_{t-1}^T)^{(i,j)}) \right| = o_p(1).$$

Based on this, we have as $m \rightarrow \infty$,

$$\max_{m+1 \leq k \leq km / \log(km)} \frac{k}{km} \left| \frac{1}{k} \sum_{t=1}^k (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} - \mathbf{E}(\mathbf{Z}_{t-1} \boldsymbol{\Sigma}(\boldsymbol{\beta}_0) \mathbf{Z}_{t-1}^T)^{(i,j)} \right| = o_P(1),$$

and

$$\max_{km / \log(km) \leq k \leq km} \frac{k}{km} \left| \frac{1}{k} \sum_{t=1}^k ((\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} - \mathbf{E}(\mathbf{Z}_{t-1} \boldsymbol{\Sigma}(\boldsymbol{\beta}_0) \mathbf{Z}_{t-1}^T)^{(i,j)}) \right| = o_P(1).$$

Hence, we get that

$$\max_{m+1 \leq k \leq km} \frac{1}{km} \left| \sum_{t=1}^k ((\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} - \mathbf{E}(\mathbf{Z}_{t-1} \boldsymbol{\Sigma}(\boldsymbol{\beta}_0) \mathbf{Z}_{t-1}^T)^{(i,j)}) \right| = o_P(1).$$

Following the proof of Proposition 3 (see Fokianos et al. [16])

$$\max_{1 \leq k \leq m} \frac{k}{m} \left| \frac{1}{k} \sum_{t=1}^m ((\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} - \mathbf{E}(\mathbf{Z}_{t-1} \boldsymbol{\Sigma}(\boldsymbol{\beta}_0) \mathbf{Z}_{t-1}^T)^{(i,j)}) \right| = o_P(1),$$

then we have

$$\begin{aligned} & \max_{m+1 \leq k \leq km} \frac{1}{km} \frac{k}{m} \left| \sum_{t=1}^m ((\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} - \mathbf{E}(\mathbf{Z}_{t-1} \boldsymbol{\Sigma}(\boldsymbol{\beta}_0) \mathbf{Z}_{t-1}^T)^{(i,j)}) \right| \\ &= \max_{m+1 \leq k \leq km} \frac{m}{km} \frac{k}{m} \left| \frac{1}{m} \sum_{t=1}^m ((\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} - \mathbf{E}(\mathbf{Z}_{t-1} \boldsymbol{\Sigma}(\boldsymbol{\beta}_0) \mathbf{Z}_{t-1}^T)^{(i,j)}) \right| = o_P(1). \end{aligned}$$

Combining the above with $m^{1/2}(\hat{\boldsymbol{\beta}}_m^{(i)} - \hat{\boldsymbol{\beta}}_0^{(i)}) = o_P(1)$ from Lemma 3.2, we get (6). The error terms $\mathbf{E}_{k,m}^{(i)}$ have higher orders of products of $(\hat{\boldsymbol{\beta}}_m^{(j)} - \boldsymbol{\beta}_0^{(j)})$, and it can be shown that they are $o_P(1)$ by Lemma 3.2.

From Lemmas 3.1 and 3.2, we get as $m \rightarrow \infty$

$$\max_{m+1 \leq k \leq km} (\kappa m)^{-1/2} \mathbf{T}^{-1/2} \left(\mathbf{S}_k(\boldsymbol{\beta}_0) - \frac{k}{m} \mathbf{S}_m(\boldsymbol{\beta}_0) \right) \xrightarrow{D} \sup_{1 < s < \kappa} \kappa^{-1/2} (\mathbf{W}(s) - s\mathbf{W}(1)).$$

Finally, we replace $\mathbf{T}$ with $\hat{\mathbf{T}}$, and the results hold true by using Slutsky's theorem. $\square$

Proof of Theorem 3.2. Under the alternative hypothesis, $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, for all $\boldsymbol{\pi}_t(\boldsymbol{\beta})$, $t = m+1, m+2, \dots, m+k^*$, and $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ for $\boldsymbol{\pi}_t$, $m+k^* < t \leq km$. Suppose the coefficient changes from $\boldsymbol{\beta}_0$ to $\boldsymbol{\beta}_0^*$ at $m+k^*$, $\boldsymbol{\beta}_0^{*(j)} = \boldsymbol{\beta}_0^{(j)} + \delta$, $\boldsymbol{\beta}_0^{(j)}$ is the j th component of $\boldsymbol{\beta}_0$, $j = 1, \dots, qd$, where δ is a constant, $\delta \neq 0$. $\square$

For $m+k^* < k \leq km$, we can write

$$\max_{m+1 \leq k \leq km} (\kappa m)^{-1/2} \mathbf{S}_k(\hat{\boldsymbol{\beta}}_m) \geq (\kappa m)^{-1/2} \mathbf{S}_k(\hat{\boldsymbol{\beta}}_m) = (\kappa m)^{-1/2} (\mathbf{S}_{k_1}(\hat{\boldsymbol{\beta}}_m) + \mathbf{S}_{k_2}(\hat{\boldsymbol{\beta}}_m)),$$

where

$$\mathbf{S}_{k_1}(\hat{\boldsymbol{\beta}}_m) = \sum_{t=1}^{m+k^*-1} \mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\hat{\boldsymbol{\beta}}_m)), \mathbf{S}_{k_2}(\hat{\boldsymbol{\beta}}_m) = \sum_{t=m+k^*}^k \mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\hat{\boldsymbol{\beta}}_m)).$$

For the i th component of $\mathbf{S}_{k_1}(\hat{\boldsymbol{\beta}}_m)$, $i = 1, \dots, qd$, we have

$$\begin{aligned}
& (\kappa m)^{-1/2} \mathbf{S}_{k_1}^{(i)}(\hat{\boldsymbol{\beta}}_m) \\
&= (\kappa m)^{-1/2} (\mathbf{S}_{k_1}^{(i)}(\boldsymbol{\beta}_0)) + (\kappa m)^{1/2} \sum_{j=1}^{qd} (\hat{\boldsymbol{\beta}}_m^{(j)} - \boldsymbol{\beta}_0^{(j)}) (\kappa m)^{-1} \sum_{t=1}^{m+k^*-1} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(i)} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0)))^{(j)} + \mathbf{E}_{k_1, m}^{(i)}, \\
& (\kappa m)^{-1/2} \mathbf{S}_{k_2}^{(i)}(\hat{\boldsymbol{\beta}}_m) \\
&= (\kappa m)^{-1/2} (\mathbf{S}_{k_2}^{(i)}(\boldsymbol{\beta}_0^*)) + (\kappa m)^{1/2} \sum_{j=1}^{qd} (\hat{\boldsymbol{\beta}}_m^{(j)} - \boldsymbol{\beta}_0^{*(j)}) (\kappa m)^{-1} \sum_{t=m+k^*}^k (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0^*)))^{(i)} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0^*)))^{(j)} + \mathbf{E}_{k_2, m}^{(i)},
\end{aligned}$$

where $\mathbf{E}_{k_1, m}^{(i)}$ has two orders of products of $(\hat{\boldsymbol{\beta}}_m^{(j)} - \boldsymbol{\beta}_0^{(j)})$, $\mathbf{E}_{k_2, m}^{(i)}$ has two orders of products of $(\hat{\boldsymbol{\beta}}_m^{(j)} - \boldsymbol{\beta}_0^{*(j)})$.

As $m \rightarrow \infty$, by Theorem 3.1 we have

$$(\kappa m)^{-1/2} \mathbf{S}_{k_1}^{(i)}(\hat{\boldsymbol{\beta}}_m) = o_P(1), \quad (7)$$

$$(\kappa m)^{-1/2} \mathbf{S}_{k_2}^{(i)}(\hat{\boldsymbol{\beta}}_m) = o_P(1), \quad (8)$$

following Assumptions 1–3, we conclude

$$(\kappa m)^{-1} \sum_{t=m+k^*}^k (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0^*)))^{(i)} (\mathbf{Z}_{t-1}(\mathbf{Y}_t - \boldsymbol{\pi}_t(\boldsymbol{\beta}_0^*)))^{(j)} = o_P(1). \quad (9)$$

Since $\delta \neq 0$ and δ is a constant, as $m \rightarrow \infty$ we have

$$(\kappa m)^{1/2} (\hat{\boldsymbol{\beta}}_m^{(j)} - \boldsymbol{\beta}_0^{*(j)}) = (\kappa m)^{1/2} (\hat{\boldsymbol{\beta}}_m^{(j)} - \boldsymbol{\beta}_0^{(j)} - \boldsymbol{\delta}) \xrightarrow{P} \infty \quad (10)$$

by Lemma 3.2. Putting together (7), (8), (9) and (10), the proof of Theorem 3.4 is complete.