Statistics and Its Interface Volume 9 (2016) 113–122

# Detecting change point in linear regression using jackknife empirical likelihood*

Xinqi Wu, Sanguo Zhang, Qingzhao Zhang, and Shuangge Ma†

Data generated in quite a few examples can be described using a linear regression model with a change point. In this paper for such a model, we develop a nonparametric method based on the jackknife empirical likelihood (JEL) to detect the change in regression coefficients. Under mild conditions, we show that the null distribution of the JEL ratio test statistic is asymptotically Gumbel. The test and the estimator of change point are shown to be consistent under the alternative hypothesis. Simulation suggests that the proposed method is computationally much more affordable than the alternative based on empirical likelihood. We also demonstrate the proposed method using two real datasets.

AMS 2000 subject classifications: 62J05, 62G10.
Keywords and phrases: Change point, Jackknife empirical likelihood, Jackknife pseudo-values.

## 1. INTRODUCTION

Change-point problems are encountered in many disciplines such as economics, finance, medicine, geology and so on. Statistically, a change point is a place or time point such that the observations follow one distribution up to that point and a different distribution afterwards. The change-point problem is two-fold: the first is to decide if there is any change which is often formulated as a hypothesis testing problem, and the second is to locate the change point which is often formulated as an estimation problem. With parametric distributions, research on change point has been summarized in [4, 6].

This study addresses the change point problem in linear regression. In this line of research, [8] used a union and intersection approach to test changes. The researches [2, 3] estimated multiple structural changes in a linear regression model by least squares. [1] suggested to apply the permutation and bootstrap methods in change-point analysis. Their idea was later pursued by others (for a survey, see [9]). The consistency and limiting distribution of the maximum likelihood estimators of the underlying parameters in a two-phase linear regression model were estimated by [13]. Recently, [10] proposed a bootstrap method for change point detection in linear regression under the sequential settings.

Different from most of the existing ones, this study adopts the jackknife empirical likelihood (JEL) as the main technique. Although not having a long history, the empirical likelihood (EL) technique has been shown to be very useful. Briefly, [18] proposed the EL technique as an alternative to the bootstrap for constructing confidence regions. The EL technique is employed to the detection of change point in the mean and in a linear model, respectively by [25] and [24]. A new EL ratio statistic for change point detection in segmented linear regression was developed by [14]. However, the methods based on the EL technique for change point detection have computational limitations. Such methods proceed by the two-sample EL [11], which demands setting up two probability vectors. The solution to the optimization problem is the saddle point [19]. Optimizing the EL can thus be difficult and time-consuming. To address the computational problem, we apply the JEL [12], as opposed to EL, for change point detection in a linear regression model. The key to our procedure is to turn the statistic of interest into a one-sample statistic by using the jackknife pseudo-values, which makes our procedure easier to implement. In addition to the computational advantage, the proposed procedure also enjoys the much desired consistency properties.

This study has two main contributions. The first is a novel JEL-based method for change point detection in a linear regression model. The proposed method has significant computational advantages over the EL-based methods. Second, there is no available theoretical tool ready to analyze the proposed method. The establishment of consistency properties is highly nontrivial with the test statistic constructed based on dependent variables. The rest of the article is organized as follows. Section 2 describes the change point detection procedure based on JEL. Section 3 establishes the asymptotic distribution of the test statistic under the null and consistency of test. Simulation and data analyses are conducted in Section 4 to illustrate performance of the proposed test. The article concludes with discussions in Section 5. All proofs are given in the Appendix.

*We thank the editor and reviewer for careful review and insightful comments. This study was partly supported by the China Postdoctoral Science Foundation (2014M550799), National Natural Science Foundation of China (11401561, 71471152, 71201139, 71301162) and National Social Science Foundation of China (13&ZD148, 13CTJ001).
†Corresponding author.

## 2. DETECTION OF CHANGE POINT USING JEL

### 2.1 Jackknife empirical likelihood method

For the integrity of this article, we first outline the JEL approach as proposed by [12]. Let $Z_1, ..., Z_n$ be independent (but not necessarily identically distributed) random variables (r.v.'s). Let

$$T_n = T(Z_1, \cdots, Z_n)$$

be a consistent estimator for a parameter of interest, denoted by $\delta$. Define the jackknife pseudo-values by

$$V_i = nT_n - (n-1)T_{n-1}^{(-i)},$$

where $T_{n-1}^{(-i)} := T(Z_1, \cdots, Z_{i-1}, Z_{i+1}, \cdots, Z_n)$ is the statistic $T$ computed on the sample of $n-1$ variables generated from the original dataset by deleting the $i$th variable. The pseudo-value $V_i$'s ($1 \le i \le n$) are asymptotically independent under mild conditions. Therefore, we can apply [18]'s empirical likelihood method to the jackknife pseudo-value $V_i$'s and proceed as follows.

Let $\mathbf{p} = (p_1, \cdots, p_n)$ with $\sum_{i=1}^n p_i = 1$ and $p_i \ge 0$ for $1 \le i \le n$. Then the JEL for $\delta$, evaluated at $\delta_0$, is given by

$$L(\delta_0) = \sup\left\{\prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i V_i = \delta_0\right\}.$$

It is clear that $\prod_{i=1}^n p_i$, subject to $\sum_{i=1}^n p_i = 1$, attains its maximum $n^{-n}$ at $p_i = n^{-1}$. So the JEL ratio at $\delta_0$ is defined as

$$R(\delta_0) = \frac{L(\delta_0)}{n^{-n}} = \sup\left\{\prod_{i=1}^n np_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i V_i = \delta_0\right\}.$$

It is shown by [12] that the Wilks' theorem holds for $-2\log R(\delta_0)$ under weak assumptions when $T_n$ is a one- or two-sample $U$-statistic.

### 2.2 Change-point problem in the linear regression model

Let $y_1, y_2, \cdots, y_n$ be a sequence of independent r.v.'s in $R$. Assume the following linear regression model which has at most one change point at location $k_0$, which is unknown,

$$y_i = \begin{cases} x_i^\top \beta + e_i, & 1 \le i \le k_0, \\ x_i^\top \beta^* + e_i, & k_0 < i \le n. \end{cases}$$

where $x_i \in R^p$, $\beta$ and $\beta^*$ are unknown regression coefficients, $e_i$'s are independently identically distributed (i.i.d) random variables with $Ee_i = 0,\ Ee_i^2 = \sigma^2 > 0$. The analysis goal is to test the null hypothesis of no change in the regression coefficients against the alternative hypothesis of change at an unknown location. Let $\delta = \beta^* - \beta$, then we test:

$$H_0: \ \delta = 0 \ \leftrightarrow \ H_1: \ \delta \neq 0.$$

The role the linear regression model with a change point plays in the change point literature is similar to what the simple linear regression plays in regression. It is the basis of more complicated change-point regression problems. In addition, multiple published studies, as referred to in this paper, have shown that this model has important practical applications. We also conjecture that the method developed in this study for change-point linear regression and its theoretical tools can be extended to other more complicated regression models.

### 2.3 Jackknife empirical log-likelihood ratio test

When describing the proposed method, we focus on the scenario with at most one change point. In data analysis when multiple change points are suspected, we can apply [23]'s binary segmentation method, which proceeds as follows. First, determine if there is at least one change point. If there is none, then the null hypothesis is accepted. If one change point is detected, then it naturally divides the original sequence of random variables into two subsequences. For each subsequence, apply the proposed method. This process continues until no more change points can be found in any of the subsequences.

Let $\tau_k = k/n$,

$$y_{1,k} = (y_1, \cdots, y_k)^\top, \ x_{1,k} = (x_1, \cdots, x_k)^\top,$$
$$y_{2,k} = (y_{k+1}, \cdots, y_n)^\top, \ x_{2,k} = (x_{k+1}, \cdots, x_n)^\top.$$

Then the statistic

$$T_{n,k} = (x_{2,k}^\top x_{2,k})^{-1} x_{2,k}^\top y_{2,k} - (x_{1,k}^\top x_{1,k})^{-1} x_{1,k}^\top y_{1,k}$$

is a consistent estimator of the parameter $\delta$ under $H_0$. The corresponding jackknife pseudo-values are defined by

$$V_{i,k} = nT_{n,k} - (n-1)T_{n-1,k}^{(-i)},$$

where $T_{n-1,k}^{(-i)}$ is computed on the sample of $n-1$ variables generated from the original dataset by deleting the $i$th data point. More specifically, for $1 \le i \le k$

$$T_{n-1,k}^{(-i)} = \left(x_{2,k}^\top x_{2,k}\right)^{-1} x_{2,k}^\top y_{2,k} - \left(x_{1,k}^\top x_{1,k} - x_i x_i^\top\right)^{-1}\left(x_{1,k}^\top y_{1,k} - x_i y_i\right),$$

and for $k < i \le n$,

$$T_{n-1,k}^{(-i)} = \left(x_{2,k}^\top x_{2,k} - x_i x_i^\top\right)^{-1}\left(x_{2,k}^\top y_{2,k} - x_i y_i\right) - \left(x_{1,k}^\top x_{1,k}\right)^{-1} x_{1,k}^\top y_{1,k}.$$

Let $(p_1, \cdots, p_n)$ be the probability vector with $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0$ for $1 \leq i \leq n$. If the change occurs at $k$, the jackknife empirical log-likelihood ratio (JELR) test statistic is defined as

$$-2\log\Lambda_k = -2\sup\left\{\sum_{i=1}^n \log np_i \,\middle|\, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i V_{i,k} = 0\right\}.$$

We reject $H_0$ for a sufficiently large value of $\max_{1<k<n}\{-2\log\Lambda_k\}$. The Lagrange multiplier method leads to

$$p_i = \frac{1}{n}\frac{1}{1+\lambda^\top V_{i,k}}.$$

After plugging $p_i$'s back into $-2\log\Lambda_k$, and according to the dual theory, the nonparametric JELR test statistic can be rewritten as

$$-2\log\Lambda_k = 2\sup\left\{\sum_{i=1}^n \log(1+\lambda^\top V_{i,k})\right\} \overset{\text{def}}{=} 2\sup l_E(\lambda).$$

Furthermore, let

$$Q_n(\lambda) = \frac{1}{n}\sum_{i=1}^n \frac{V_{i,k}}{1+\lambda^\top V_{i,k}}.$$

Then the optimal $\lambda$, denoted as $\tilde{\lambda}$, satisfies that $Q_n(\tilde{\lambda}) = 0$. Since $k$ is unknown, it is natural to use the maximum JELR statistic

$$Z_n^* = \max_{1<k<n}\{-2\log\Lambda_k\}.$$

We reject the null hypothesis with a significantly large value of $Z_n^*$. Note that if $k$ or $n-k$ is too small, the estimator of jackknife empirical log-likelihood $\tilde{\lambda}$ may not exist. That is, our test may not be able to detect the change point if it occurs close to either end. Therefore, we further propose the trimmed likelihood ratio statistic as

$$Z_n = \max_{k\in\Theta_{nk}}\{Z_{n,k}\},$$

where $\Theta_{nk} = \{k : C_1 \leq k \leq n - C_2\}$. The maximum of $Z_{n,k}$ is attained at $\hat{k}$, and the corresponding estimator is $\hat{\tau}_k = \frac{\hat{k}}{n}$. The number $C_i$ $(i = 1, 2)$ can be chosen relatively arbitrarily, which was pointed by [22]. In our numerical study, we choose $C_i = 2[n^{\frac{1}{2}}]$ where $[x]$ is the largest integer no larger than $x$.

## 3. STATISTICAL PROPERTIES

In this section, we outline the main results for the asymptotic distribution under the null and consistency properties under the alternative. We first assume the following condition:

$(C)$ For a fixed $\delta = \beta^* - \beta \neq 0$, there exists a constant $c_0 > 0$ such that

$$\begin{aligned} c_0 \;\leq\; & \sup_\lambda \{\tau_0 E\log[1+\lambda^\top(\delta - \tau_0^{-1}A^{-1}xe)] \\ & + (1-\tau_0)E\log\{1+\lambda^\top[\delta-(1-\tau_0)^{-1}A^{-1}xe]\}\} \\ & < \infty, \end{aligned}$$

where $\tau_{k_0} = k_0/n \to \tau_0 \in (0,1)$ as $n \to \infty$ and $A = E(x_1x_1^\top)$ (for more details, see Appendix). This condition is mild and similar to that in [17].

**Theorem 1.** *Suppose that for $i = 1, \cdots, n$, $(x_i^\top, e_i)^\top$'s are i.i.d, $x_i, e_i$ are mutually independent with $E\|x_i\|^4 < \infty, Ee_i^4 < \infty$, and $E(x_ix_i^\top)$ is positive-definite. If the null hypothesis $H_0$ is true, then*

$$P(A(\log(u(n)))(Z_n)^{\frac{1}{2}} \leq x + D_p(\log(u(n)))) \longrightarrow \exp(-e^{-x}),$$

*as $n \to \infty$ for all $x$, where $A(x) = (2\log x)^{\frac{1}{2}}, D_p(x) = 2\log x + \frac{p}{2}\log\log x - \log\Gamma(\frac{p}{2}), u(n) = \frac{n^2+(2[n^{\frac{1}{2}}])^2-2n[n^{\frac{1}{2}}]}{(2[n^{\frac{1}{2}}])^2}$.*

Theorem 1 shows that under the null hypothesis, the asymptotic distribution of the JELR test statistic is the Gumbel extreme value distribution. This result is similar to that for the traditional union-intersection (UI) test (see [6, 8]).

**Theorem 2.** *Suppose that the conditions of Theorem 1 and condition $(C)$ hold. If $H_1$ is true, then there exists a constant $c > 0$ such that*

$$P(Z_n > cn) \to 1.$$

*That is, the JELR test is consistent.*

**Theorem 3.** *Suppose that the conditions of Theorem 1 and condition $(C)$ hold, $H_1$ is true, and $\tau_{k_0} = \frac{k_0}{n} \to \tau_0 \in (0,1)$ as $n \to \infty$. Consider $\hat{\tau}_k = \frac{\hat{k}}{n}$ where $\hat{k}$ is the estimator of the change point location. We have $\hat{\tau}_k \to \tau_0$ in probability as $n \to \infty$.*

Theorem 2 and 3 establish that under the alternative hypothesis, the JELR enjoys the test and estimation consistency properties. The proofs are provided in Appendix.

## 4. NUMERICAL STUDY

### 4.1 Simulation

We conduct simulation to examine the finite-sample behaviors of the JELR test under different settings and compare with the UI test and the empirical log-likelihood ratio (ELR) test, which proceeds by the two-sample empirical log-likelihood ratio [24].

Consider the following regression model

$$y_i = \begin{cases} 0.5x_{i1} - x_{i2} + e_i, & 1 \leq i \leq k_0, \\ 1.5x_{i1} - 0.5x_{i2} + e_i, & k_0 + 1 \leq i \leq n. \end{cases}$$

The distributions of $x_{i1}$ and $x_{i2}$ are standard normal and standard uniform respectively. Consider the following four

Table 1. *Simulation under different sample sizes and error distributions: pseudo coverage accuracy in each cell*

| | | Normal | | | Exp | | | $\chi^2$ | | | $t$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $k_0$ | UI | ELR | JELR | UI | ELR | JELR | UI | ELR | JELR | UI | ELR | JELR |
| 50 | 0 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| | 15 | 0.549 | 0.493 | 0.464 | 0.172 | 0.466 | 0.478 | 0.361 | 0.55 | 0.517 | 0.274 | 0.552 | 0.549 |
| | 20 | 0.678 | 0.604 | 0.608 | 0.235 | 0.541 | 0.603 | 0.472 | 0.724 | 0.717 | 0.359 | 0.638 | 0.64 |
| | 25 | 0.743 | 0.684 | 0.675 | 0.338 | 0.634 | 0.659 | 0.503 | 0.744 | 0.735 | 0.367 | 0.641 | 0.654 |
| 100 | 0 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| | 30 | 0.962 | 0.943 | 0.919 | 0.676 | 0.865 | 0.842 | 0.768 | 0.848 | 0.827 | 0.883 | 0.906 | 0.923 |
| | 40 | 0.963 | 0.947 | 0.922 | 0.788 | 0.893 | 0.921 | 0.875 | 0.911 | 0.883 | 0.938 | 0.949 | 0.952 |
| | 50 | 0.969 | 0.947 | 0.929 | 0.833 | 0.917 | 0.938 | 0.927 | 0.941 | 0.942 | 0.948 | 0.958 | 0.958 |
| 200 | 0 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| | 60 | 0.991 | 0.991 | 0.994 | 0.981 | 0.97 | 0.972 | 0.992 | 0.981 | 0.979 | 0.992 | 0.989 | 0.987 |
| | 80 | 0.998 | 0.996 | 0.996 | 0.989 | 0.983 | 0.985 | 0.998 | 0.992 | 0.991 | 0.997 | 0.995 | 0.995 |
| | 100 | 0.999 | 0.999 | 0.999 | 0.99 | 0.988 | 0.988 | 0.999 | 0.999 | 0.999 | 0.998 | 0.998 | 0.997 |

Table 2. *Simulation study under different sample sizes and error distributions: true coverage accuracy in each cell*

| | | Normal | | | Exp | | | $\chi^2$ | | | $t$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $k_0$ | UI | ELR | JELR | UI | ELR | JELR | UI | ELR | JELR | UI | ELR | JELR |
| 50 | 0 | 0.036 | 0.133 | 0.055 | 0.108 | 0.131 | 0.043 | 0.09 | 0.128 | 0.043 | 0.1 | 0.112 | 0.031 |
| | 15 | 0.481 | 0.7 | 0.484 | 0.402 | 0.631 | 0.444 | 0.541 | 0.736 | 0.488 | 0.447 | 0.685 | 0.476 |
| | 20 | 0.605 | 0.809 | 0.639 | 0.48 | 0.71 | 0.579 | 0.648 | 0.86 | 0.685 | 0.546 | 0.758 | 0.556 |
| | 25 | 0.68 | 0.862 | 0.701 | 0.618 | 0.78 | 0.631 | 0.678 | 0.865 | 0.713 | 0.544 | 0.767 | 0.581 |
| 100 | 0 | 0.016 | 0.085 | 0.033 | 0.082 | 0.101 | 0.035 | 0.066 | 0.103 | 0.042 | 0.056 | 0.094 | 0.028 |
| | 30 | 0.911 | 0.969 | 0.89 | 0.793 | 0.915 | 0.811 | 0.82 | 0.91 | 0.809 | 0.895 | 0.948 | 0.889 |
| | 40 | 0.914 | 0.971 | 0.893 | 0.877 | 0.945 | 0.904 | 0.903 | 0.953 | 0.871 | 0.942 | 0.968 | 0.931 |
| | 50 | 0.922 | 0.971 | 0.897 | 0.907 | 0.956 | 0.921 | 0.95 | 0.973 | 0.935 | 0.953 | 0.974 | 0.935 |
| 200 | 0 | 0.01 | 0.068 | 0.029 | 0.085 | 0.093 | 0.042 | 0.057 | 0.093 | 0.041 | 0.056 | 0.078 | 0.026 |
| | 60 | 0.999 | 0.999 | 0.999 | 0.993 | 0.984 | 0.966 | 0.993 | 0.989 | 0.975 | 0.993 | 0.992 | 0.981 |
| | 80 | 0.999 | 0.999 | 0.999 | 0.994 | 0.992 | 0.982 | 0.999 | 0.995 | 0.988 | 0.998 | 0.997 | 0.994 |
| | 100 | 1 | 1 | 1 | 0.995 | 0.992 | 0.987 | 0.999 | 0.999 | 0.998 | 0.998 | 0.998 | 0.996 |

distributions for the random error: *i)* $e_i \sim N(0,1)$, *ii)* $e_i \sim \exp(1.0)-1.0$; *iii)* $e_i \sim \frac{1}{2\sqrt{2}}(\chi^2(4.0)-4.0)$; *iv)* $e_i \sim \frac{1}{\sqrt{2}}t(4.0)$. All error distributions have mean zero and variance one. Consider sample size $n = 50, 100$ and 200. Set $\alpha = 0.05$. Multiple change point locations are considered (Table 1). Under each setting, we simulate 5,000 replicates and compute summary statistics.

In the first set of analysis, we choose the 95% percentile values of the test statistics with $k_0 = 0$ from the 5,000 simulated replicates as the critical values. That is, with $k_0 = 0$, the pseudo coverage accuracy is exactly 0.05. This approach can ensure more precise false probabilities. For the other three $k_0$ values, the pseudo coverage accuracy is computed based on the calculated critical values. Table 1 shows that when the sample size is not large ($n = 50,\ 100$), under the normal random error distribution, the JELR and ELR methods have a slight disadvantage compared to UI. However, under the other three error distributions with $n = 50,\ 100$, both perform more efficiently and more powerfully than UI in detecting the existence of change point. Although JELR and ELR have similar performance, JELR is computationally more efficient. Specifically, in our simulation, the computer time of ELR is about 2 times more than that of JELR. When the sample size is large ($n = 200$), all three methods have satisfactory performance.

In the second set of analysis, we use the theoretical critical value obtained based on Theorem 1 to calculate the true coverage accuracy. The results are shown in Table 2. We observe that the JEL method is able to control the type I error while the UI and EL methods fail, although these two methods mostly have higher power. We also examine the estimate $\hat{\tau}_k$. The results on bias and standard deviation are shown in Table 3. All three methods have reasonable performance. Under all simulation scenarios, the proposed method does not show any systematic bias.

## 4.2 Analysis of the NASA dataset

The first dataset is from a NASA calibration application and has been analyzed in [15]. The dataset is available from the authors. Consider the linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i,$$

where $y_i$ and $x_i$ are the force balance and axial force component respectively, and $e_i$ is normally distributed. We apply

Table 3. Simulation study: bias (standard deviation) of $\hat{\tau}_k$ under the alternative hypothesis

| | | Normal | | | Exp | | | $\chi^2$ | | | $t$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $\tau_{k_0}$ | UI | ELR | JELR | UI | ELR | JELR | UI | ELR | JELR | UI | ELR | JELR |
| 50 | 0.3 | 0.03 | 0.069 | 0.078 | 0.048 | 0.079 | 0.078 | 0.017 | 0.061 | 0.059 | 0.085 | 0.088 | 0.116 |
| | | (0.17) | (0.117) | (0.133) | (0.203) | (0.13) | (0.118) | (0.192) | (0.128) | (0.127) | (0.205) | (0.129) | (0.13) |
| | 0.4 | 0.013 | 0.026 | 0.04 | 0.034 | 0.051 | 0.033 | 0.031 | 0.046 | 0.053 | 0.043 | 0.048 | 0.065 |
| | | (0.136) | (0.099) | (0.114) | (0.185) | (0.107) | (0.091) | (0.16) | (0.121) | (0.127) | (0.178) | (0.118) | (0.114) |
| | 0.5 | -0.014 | -0.008 | -0.013 | 0.015 | 0.024 | 0.004 | -0.001 | 0.011 | -0.002 | -0.028 | -0.03 | -0.02 |
| | | (0.118) | (0.095) | (0.111) | (0.169) | (0.104) | (0.097) | (0.156) | (0.116) | (0.125) | (0.178) | (0.118) | (0.115) |
| 100 | 0.3 | 0.05 | 0.051 | 0.087 | 0.015 | 0.002 | 0.027 | 0.018 | 0.032 | 0.032 | 0.018 | 0.041 | 0.037 |
| | | (0.088) | (0.103) | (0.124) | (0.135) | (0.093) | (0.094) | (0.108) | (0.1) | (0.099) | (0.088) | (0.112) | (0.105) |
| | 0.4 | 0.004 | 0.001 | 0.029 | 0.007 | -0.002 | 0.005 | 0.01 | 0.012 | -0.012 | 0.017 | 0.031 | 0.01 |
| | | (0.088) | (0.107) | (0.128) | (0.093) | (0.101) | (0.09) | (0.09) | (0.099) | (0.096) | (0.079) | (0.11) | (0.107) |
| | 0.5 | -0.001 | -0.004 | 0.001 | 0.001 | -0.005 | -0.005 | 0.001 | 0.004 | 0.004 | 0.009 | 0.025 | 0.001 |
| | | (0.092) | (0.109) | (0.125) | (0.087) | (0.103) | (0.092) | (0.076) | (0.099) | (0.105) | (0.085) | (0.111) | (0.113) |
| 200 | 0.3 | 0.001 | -0.008 | 0.021 | -0.005 | -0.003 | 0.008 | -0.002 | -0.001 | 0.009 | -0.006 | -0.01 | -0.005 |
| | | (0.028) | (0.049) | (0.067) | (0.054) | (0.091) | (0.093) | (0.046) | (0.075) | (0.071) | (0.042) | (0.074) | (0.073) |
| | 0.4 | -0.009 | -0.014 | 0.021 | 0.009 | 0.003 | 0.014 | -0.022 | -0.022 | -0.016 | 0.007 | -0.003 | 0.003 |
| | | (0.027) | (0.049) | (0.059) | (0.052) | (0.1) | (0.1) | (0.038) | (0.077) | (0.075) | (0.034) | (0.072) | (0.065) |
| | 0.5 | 0.009 | 0.008 | 0.004 | -0.003 | -0.014 | -0.013 | 0.009 | 0.022 | 0.013 | -0.007 | 0.001 | -0.002 |
| | | (0.026) | (0.043) | (0.056) | (0.051) | (0.111) | (0.099) | (0.03) | (0.076) | (0.068) | (0.036) | (0.076) | (0.072) |

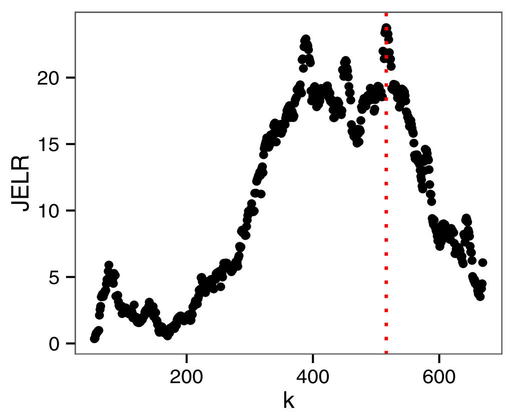


Figure 1. JELR values of the NASA example. The dashed line corresponds to the estimated location of the maximum JELR at $\hat{k} = 516$.

the JELR procedure along with the binary segmentation method to this dataset. One change point with estimated location $\hat{k} = 516$ is identified. The test statistic $Z_n = 23.76$ has p-value 0.006 based on Theorem 1. The estimated location is also in accordance with the traditional UI method. With our method, the JELR values are plotted in Figure 1.

### 4.3 Analysis of the discount rate dataset

This dataset has been analyzed in [7], where the sample period goes from 1973 to 1989, and there are a total of 56 observations. [2] adopted the following linear regression model to describe the relationship between the change in the discount rate for the $i$th observation ($\Delta DR_i$) and the change in the market interest rate ($\Delta TB_i$)

Table 4. Analysis of the discount rate data: test statistics and break points

| Sample | Data | $n_{obs}$ | JELR | p-value | $\hat{k}$ |
|---|---|---|---|---|---|
| $[1, 56]$ | 1/15/73−2/24/89 | 56 | 11.284 | 0.059 | 27 |
| $[1, 27]$ | 1/15/73−9/19/79 | 27 | 5.908 | 0.120 | – |
| $[28, 56]$ | 10/9/79−2/24/89 | 29 | 14.343 | 0.050 | 38 |

$$\Delta TB_i = \alpha + \beta \Delta DR_i + e_i.$$

He applied the sup-Wald type test and detected breaks at positions 28, 38 and 42 at the 10% nominal size. Here we apply the JELR procedure along with the binary segmentation method and present the results in Table 4, where the p-values are calculated based on Theorem 1.

We also detect a break at position 27, which has test statistic $Z_n = 11.284$ with p-value 0.059. Using the binary segmentation method, we continue and detect another break at position 3, which has test statistic $Z_n = 14.343$ with p-value 0.050. As the remaining sample sizes of subsequences are small, we stop the process.

## 5. DISCUSSION

In this article, we have developed a new method for detecting change point(s) in linear regression. The first major contribution is the adoption of the jackknife empirical likelihood technique, which differs from many of the existing studies. The second contribution is the rigorous establishment of statistical properties, especially including the Gumbel extreme value distribution under the null and the test and estimation consistency under the alternative. Simulation shows reasonable performance of the proposed method.

Compared with the direct competitor, the EL method, the JEL method has lower computational cost and can better control type I error. Following strategies in the literature, we have chosen to first study the linear regression model. This study may serve as the basis for further methodological development for the generalized linear models and Cox survival model. However, the theoretical development is expected to be highly nontrivial and postponed to future studies.

## APPENDIX

In the following, the convergence of random variables involved is the convergence in probability, and the convergence of distribution involved is the convergence in distribution. For simplicity of notation, we use "$\to$" for both types of convergence and denote "$\xrightarrow{\text{a.s.}}$" as the almost sure convergence of a sequence of random variables. $\|\cdot\|$ denotes the Frobenius norm of a matrix. $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix. For symmetric matrices $M_1$ and $M_2$, if $M_1 - M_2$ is (semi-)positive definite matrix, we denote $(M_1 \geq M_2) M_1 > M_2$. Define $\tau_k = k/n$ and let $\varepsilon_k = \min\{\tau_k, 1-\tau_k\}$, $m = n\varepsilon_k$. By the law of large number, $\frac{1}{k} x_{1,k}^\top x_{1,k} \to A$, $\frac{1}{n-k} x_{2,k}^\top x_{2,k} \to A$ as $k \to \infty$ and $n-k \to \infty$, where $A = E(x_1 x_1^\top)$ is positive definite.

**Lemma 1.** *Define $S_k = \frac{1}{n}\sum_{i=1}^n V_{i,k} V_{i,k}^\top$. Under the conditions of Theorem 1 and null hypothesis, we have*

1) $\max_{1\leq i\leq n} \|V_{i,k}\| = o_p(\sqrt{\frac{n}{\varepsilon_k}})$;

2) $\bar{V}_k \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n V_{i,k} = O_p(m^{-\frac{1}{2}})$;

3) $S_k = \frac{1}{\tau_k(1-\tau_k)}(\sigma^2 A^{-1} + O_p(m^{-\frac{1}{2}})) = O_p(\varepsilon_k^{-1})$;

4) $\frac{1}{n}\sum_{i=1}^n ||V_{i,k}||^3 = O_p(\varepsilon_k^{-2})$.

*Proof.* For a fixed $k$, we first define some notations. Let

$$\Delta_i = \begin{cases} x_i^\top (x_{1,k}^\top x_{1,k})^{-1} x_i, & 1 \leq i \leq k, \\ x_i^\top (x_{2,k}^\top x_{2,k})^{-1} x_i, & k < i \leq n. \end{cases}$$

When $1 \leq i \leq k$, we denote $e_i = y_i - x_i^\top \beta$, $u_i = x_i^\top(\hat\beta_k - \beta)$, $r_i = y_i - x_i^\top \hat\beta_k = e_i - u_i$ and $\xi_i = (x_{1,k}^\top x_{1,k})^{-\frac{1}{2}} x_i r_i$, where $\hat\beta_k = (x_{1,k}^\top x_{1,k})^{-1} x_{1,k}^\top y_{1,k}$ is the least square estimate of $\beta$. Similarly, when $k+1 \leq j \leq n$, we also can defined $e_j$, $u_j^*$, $r_j^*$ and $\xi_j^*$, where $\hat\beta_k^* = (x_{2,k}^\top x_{2,k})^{-1} x_{2,k}^\top y_{2,k}$ is the least square estimate of $\beta^*$. Let $\sigma^2 = Ee_i^2 = Ee_i^{*2}$. We introduce the following results which were established by [16].

1. $\max_{1\leq i\leq k} x_i^\top (x_{1,k}^\top x_{1,k})^{-1} x_i \xrightarrow{\text{a.s.}} 0$ as $k \to \infty$, $\max_{k+1\leq i\leq n} x_i^\top (x_{2,k}^\top x_{2,k})^{-1} x_i \xrightarrow{\text{a.s.}} 0$ as $n-k \to \infty$. (See Lemma 2.3 in [16].)
2. $T_{n,k} - T_{n-1,k}^{-i} = \begin{cases} -\frac{(x_{1,k}^\top x_{1,k})^{-1} x_i (y_i - x_i^\top \hat\beta_k)}{1 - x_i^\top (x_{1,k}^\top x_{1,k})^{-1} x_i}, & 1 \leq i \leq k, \\ \frac{(x_{2,k}^\top x_{2,k})^{-1} x_i (y_i - x_i^\top \hat\beta_k^*)}{1 - x_i^\top (x_{2,k}^\top x_{2,k})^{-1} x_i}, & k < i \leq n. \end{cases}$ (See Lemma 3.2 in [16].)
3. $\frac{1}{k}\sum_{i=1}^k r_i^2 x_i x_i^\top \to \sigma^2 A$ as $k \to \infty$, $\frac{1}{n-k}\sum_{i=k+1}^n r_i^{*2} x_i x_i^\top \to \sigma^2 A$ as $n-k \to \infty$. (See Lemma 3.4 in [16].)

It is not hard to deduce that, under the null hypothesis ($\delta = 0$),

$$\begin{aligned} T_{n,k} &= (x_{2,k}^\top x_{2,k})^{-1} x_{2,k}^\top y_{2,k} - (x_{1,k}^\top x_{1,k})^{-1} x_{1,k}^\top y_{1,k} \\ &= (x_{2,k}^\top x_{2,k})^{-1} x_{2,k}^\top e_{2,k} - (x_{1,k}^\top x_{1,k})^{-1} x_{1,k}^\top e_{1,k}, \end{aligned}$$

where $e_{1,k} = (e_1, \cdots, e_k)^\top$ and $e_{2,k} = (e_{k+1}, \cdots, e_n)^\top$. For any given $x_{1,k}$ and $x_{2,k}$, we can derive that

$$\sqrt{m}(\sigma^2 A^{-1})^{-\frac{1}{2}} T_{n,k} \to N(0, I_p).$$

1). Obviously,

$$\sqrt{\frac{\varepsilon_k}{n}} V_{i,k} = \sqrt{\frac{\varepsilon_k}{n}} T_{n,k} + \sqrt{\frac{\varepsilon_k}{n}} (n-1)(T_{n,k} - T_{n-1,k}^{-i}).$$

The order of the first part is $o_p(1)$. For the second part, $\|\sqrt{n\varepsilon_k}(T_{n,k} - T_{n-1,k}^{-i})\| = O_p(\|\sqrt{n\varepsilon_k}(x_{1,k}^\top x_{1,k})^{-\frac{1}{2}} \xi_i\|) = O_p(\|\xi_i\|)$. It is sufficient to prove that $\max_{1\leq i\leq k} ||\xi_i|| = o_p(1)$. Since $\|\xi_i\| = \Delta_i^{\frac{1}{2}} |r_i| \leq |e_i| \Delta_i^{\frac{1}{2}} + |u_i| \Delta_i^{\frac{1}{2}}$ and $E(x_1 x_1^\top) < \infty$, we have

$$\max_{1\leq i\leq k} |e_i| \Delta_i^{\frac{1}{2}} \leq \sqrt{\lambda_{\max}\{(\frac{x_{1,k}^\top x_{1,k}}{k})^{-1}\}} \max_{1\leq i\leq k} \frac{\|x_i\| |e_i|}{\sqrt{k}} = o_p(1),$$

and

$$\begin{aligned} \max_{1\leq i\leq k} |u_i| \Delta_i^{\frac{1}{2}} &\leq \max_{1\leq i\leq k} \|x_i\| \Delta_i^{\frac{1}{2}} \|\hat\beta_k - \beta\| \\ &\leq \sqrt{\lambda_{\max}((\frac{x_{1,k}^\top x_{1,k}}{k})^{-1})} \|\sqrt{k}(\hat\beta_k - \beta)\| \max_{1\leq i\leq k} \frac{\|x_i\|^2}{k} \\ &= O_p(1) \max_{1\leq i\leq k} \frac{\|x_i\|^2}{k} = o_p(1). \end{aligned}$$

Similarly, $\max_{k+1\leq i\leq n} ||\xi_i^*|| = o_p(1)$. Thus, the first conclusion of Lemma 1 is derived.

2). Since

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n V_{i,k} &= T_{n,k} + \frac{n-1}{n} \sum_{i=1}^k (T_{n,k} - T_{n-1,k}^{-i}) \\ &\quad + \frac{n-1}{n} \sum_{i=k+1}^n (T_{n,k} - T_{n-1,k}^{-i}), \end{aligned}$$

it is sufficient to prove that the orders of the last two terms are $O_p(k^{-\frac{1}{2}})$ and $O_p((n-k)^{-\frac{1}{2}})$, respectively. In fact,

$$\sqrt{k}\sum_{i=1}^{k}(T_{n,k}-T_{n-1,k}^{-i})$$
$$= -(\frac{x_{1,k}^{\top}x_{1,k}}{k})^{-\frac{1}{2}}\sum_{i=1}^{k}\frac{(x_{1,k}^{\top}x_{1,k})^{-\frac{1}{2}}x_ir_i}{1-\Delta_i}=O_p(\|\sum_{i=1}^{k}\xi_i\|).$$

Since $Ex_1x_1^{\top}<\infty$, $\sum_{i=1}^{k}x_ie_i/\sqrt{k}=O_p(1)$ and $\sqrt{k}(\hat{\beta}_k-\beta)=O_p(1)$, we have

$$\sum_{i=1}^{k}\xi_i = \sum_{i=1}^{k}(x_{1,k}^{\top}x_{1,k})^{-\frac{1}{2}}x_ie_i - \sum_{i=1}^{k}(x_{1,k}^{\top}x_{1,k})^{-\frac{1}{2}}x_ix_i^{\top}(\hat{\beta}_k-\beta)=O_p(1).$$

Similarly, we can show that $\sqrt{n-k}\sum_{i=k+1}^{n}(T_{n,k}-T_{n-1,k}^{-i})=O_p(1)$. Therefore, $\frac{1}{n}\sum_{i=1}^{n}V_{i,k}=O_p(m^{-\frac{1}{2}})$.

3). Since $\frac{1}{n}\sum_{i=1}^{n}V_{i,k}V_{i,k}^{\top}=\frac{1}{n}\sum_{i=1}^{k}V_{i,k}V_{i,k}^{\top}+\frac{1}{n}\sum_{i=k+1}^{n}V_{i,k}V_{i,k}^{\top}$, we will prove that $\frac{1}{n}\sum_{i=1}^{k}V_{i,k}V_{i,k}^{\top}=O_p(\frac{n}{k})$ and $\frac{1}{n}\sum_{i=k+1}^{n}V_{i,k}V_{i,k}^{\top}=O_p(\frac{n}{n-k})$. In fact,

$$\frac{1}{n}\sum_{i=1}^{k}V_{i,k}V_{i,k}^{\top} = \frac{k}{n}T_{n,k}T_{n,k}^{\top}+\frac{(n-1)^2}{n}\sum_{i=1}^{k}(T_{n,k}-T_{n,k}^{-i})(T_{n,k}-T_{n,k}^{-i})^{\top} + \frac{n-1}{n}\sum_{i=1}^{k}T_{n,k}(T_{n,k}-T_{n,k}^{-i})^{\top} + \frac{n-1}{n}\sum_{i=1}^{k}(T_{n,k}-T_{n,k}^{-i})T_{n,k}^{\top}.$$

It is not hard to derive that the order of the first term is $n^{-1}kT_{n,k}T_{n,k}^{\top}=n^{-1}kO_p(m^{-1})$, and the order of the last two terms is $O_p(m^{-\frac{1}{2}})O_p(k^{-\frac{1}{2}})$. Then we plug $T_{n,k}-T_{n,k}^{-i}=-\frac{(x_{1,k}^{\top}x_{1,k})^{-1}x_ir_i}{1-\Delta_i}=O_p(\|(x_{1,k}^{\top}x_{1,k})^{-1}x_ir_i\|)$ into the second term when $1\le i\le k$ and obtain that

$$\sum_{i=1}^{k}(T_{n,k}-T_{n,k}^{-i})(T_{n,k}-T_{n,k}^{-i})^{\top} = \sum_{i=1}^{k}(x_{1,k}^{\top}x_{1,k})^{-1}r_i^2x_ix_i^{\top}(x_{1,k}^{\top}x_{1,k})^{-1} = \frac{1}{k}(A^{-1}+O_p(k^{-\frac{1}{2}}))(\frac{1}{k}\sum_{i=1}^{k}r_i^2x_ix_i^{\top})(A^{-1}+O_p(k^{-\frac{1}{2}})). \tag{1}$$

Since $e_i$, $x_i$ are independent and $E||x||^4<\infty$, we have $\sqrt{k}(\hat{\beta}_k-\beta)=O_p(1)$,

$$\frac{1}{k}\sum_{i=1}^{k}\|e_ix_ix_i^{\top}\|^2=O_p(1), \qquad \max_{1\le i\le k}\frac{\|x_i\|^2}{\sqrt{k}}=O_p(1).$$

Noticing that $\frac{1}{k}\sum_{i=1}^{k}e_i^2x_ix_i^{\top}=\sigma^2A+O_p(k^{-\frac{1}{2}})$,

$$\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}e_iu_ix_ix_i^{\top}\|^2\le\frac{1}{k}\sum_{i=1}^{k}\|e_ix_ix_i^{\top}\|^2\sum_{i=1}^{k}u_i^2=O_p(1),$$

and

$$\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}u_i^2x_ix_i^{\top}\|\le\frac{1}{k}\|\sum_{i=1}^{k}x_ix_i^{\top}\|\max_{1\le i\le k}\sqrt{k}u_i^2 \le \frac{1}{k}\|\sum_{i=1}^{k}x_ix_i^{\top}\|\sqrt{k}\max_{1\le i\le k}\|x_i\|^2\|\hat{\beta}_k-\beta\|^2=o_p(1).$$

In addition,

$$\frac{1}{k}\sum_{i=1}^{k}r_i^2x_ix_i^{\top} = \frac{1}{k}\sum_{i=1}^{k}e_i^2x_ix_i^{\top}-\frac{2}{k}\sum_{i=1}^{k}e_iu_ix_ix_i^{\top} + \frac{1}{k}\sum_{i=1}^{k}u_i^2x_ix_i^{\top},$$

we have $\frac{1}{k}\sum_{i=1}^{k}r_i^2x_ix_i^{\top}=\sigma^2A+O_p(k^{-\frac{1}{2}})$, which, combined with (1), implies

$$\frac{(n-1)^2}{n}\sum_{i=1}^{k}(T_{n,k}-T_{n,k}^{-i})(T_{n,k}-T_{n,k}^{-i})^{\top} = \frac{n}{k}[\sigma^2A^{-1}+O_p(k^{-\frac{1}{2}})].$$

Similarly, it can be shown

$$\frac{(n-1)^2}{n}\sum_{i=k+1}^{n}(T_{n,k}-T_{n,k}^{-i})(T_{n,k}-T_{n,k}^{-i})^{\top} = \frac{n}{n-k}[\sigma^2A^{-1}+O_p((n-k)^{-\frac{1}{2}})].$$

Thus, we can reach the conclusion that

$$\frac{1}{n}\sum_{i=1}^{n}V_{i,k}V_{i,k}^{\top}=\frac{1}{\tau_k(1-\tau_k)}[\sigma^2A^{-1}+O_p(m^{-\frac{1}{2}})]=O_p(\varepsilon_k^{-1}).$$

4). The proof is similar to that above. □

**Lemma 2.** Under the conditions of Theorem 1,

$$\tilde{\lambda}=\varepsilon_kO_p(m^{-\frac{1}{2}}) \text{ and } \tilde{\lambda}=S_k^{-1}\bar{V}_k+\varepsilon_kO_p(m^{-1}).$$

*Proof.* The proof is similar to that of Theorem 3.2 of [19]. □

**Lemma 3.** *Under the conditions of Theorem 1 and null hypothesis, we have*

$$-2\log\Lambda_k=n\bar{V}_k^{\top}S_k^{-1}\bar{V}_k+O_p(\|\frac{\tilde{\lambda}}{\varepsilon_k}\|).$$

*Proof.* For convenience, we write $\tilde{\lambda}=S_k^{-1}\bar{V}_k+\beta$. According to Lemma 2, $\beta=\varepsilon_kO_p(m^{-1})$. Let $\gamma_i=\tilde{\lambda}^{\top}V_{i,k}$, then

$\max_{1\le i\le n}|\gamma_i| = o_p(1)$ follows from Lemma 1 and Lemma 2. We may expand $\log(1+\gamma_i) = \gamma_i - \frac{1}{2}\gamma_i^2 + \eta_i$, where we have $P(|\eta_i| \le B|\gamma_i|^3), 1 \le i \le n) \to 1$ as $n \to \infty$, for some finite $B > 0$. Therefore,

$$\begin{aligned} -2\log\Lambda_k &= 2\sum_{i=1}^{n}\log(1+\tilde{\lambda}^\top V_{i,k}) \\ &= n\bar{V}_k^\top S_k^{-1}\bar{V}_k - n\beta^\top S_k\beta + 2\sum_{i=1}^{n}\eta_i. \end{aligned}$$

For the last two terms, $n\beta^\top S_k\beta = O_p(m^{-1})$ and $\sum_{i=1}^n \eta_i = O_p(\|\frac{\tilde{\lambda}}{\varepsilon_k}\|)$ follow from Lemma 1 and Lemma 2. Lemma 3 can be concluded by combining these results. □

**Lemma 4.** *Assume the conditions of Theorem 1 and null hypothesis. Denote* $U_{nk} = \{\frac{k}{n} : \frac{T}{n} \le \frac{k}{n} \le (1-\frac{T}{n})\}$. *For all* $\eta > 0$, *we can find finite* $C = C(\eta), T_0 = T_0(\eta), N = N(\eta)$, *such that when* $T > T_0, n > N$,

$$P(\max_{\frac{k}{n}\in U_{nk}}(\frac{m}{\log\log m})^{\frac{1}{2}} \parallel \frac{\tilde{\lambda}}{\varepsilon_k} \parallel > C) \le \eta$$

*and*

$$P(n^{-\frac{1}{2}}\max_{\frac{k}{n}\in U_{nk}} m \parallel \frac{\tilde{\lambda}}{\varepsilon_k} \parallel > C) \le \eta.$$

*Proof.* The proof is similar to that of Lemma 1.2.2 of [6]. □

**Lemma 5.** *Under the conditions of Theorem 1 and null hypothesis, for all* $0 \le \alpha < \frac{1}{2}$,

$$n^\alpha \max_{k\in\Theta_{nk}}[\tau_k(1-\tau_k)]^\alpha| - 2\log\Lambda_k - R_k| = O_p(1),$$

$$\max_{k\in\Theta_{nk}}[\tau_k(1-\tau_k)]^\alpha| - 2\log\Lambda_k - R_k| = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{3}{2}}),$$

*where* $\Theta_{nk} = \{k : C_1 \le k \le n - C_2\}, R_k = n\bar{V}_k^\top Cov(\sqrt{n}\bar{V}_k)^{-1}\bar{V}_k$.

*Proof.* With Lemma 1, we can prove that $Cov(\sqrt{n}\bar{V}_k)^{-1}S_k - I_p = O_p(m^{-\frac{1}{2}})$. Using Lemma 3, we can derive that

$$-2\log\Lambda_k = R_k + O_p(||\frac{\tilde{\lambda}}{\varepsilon_k}||).$$

Then applying Lemma 4 and following Lemma 1.1.1 of [6] finish the proof. □

**Lemma 6.** *Let* $k_0$ *be the true position of the change point. Under the conditions of Theorem 1 and alternative hypothesis, if* $k \le k_0$, *then we have*

1. $\hat{\beta}_k \xrightarrow{\text{a.s.}} \beta$ *and* $\hat{\beta}_k^* \xrightarrow{\text{a.s.}} \beta + \rho_k\delta$, *which imply that* $T_{n,k} \xrightarrow{\text{a.s.}} \rho_k\delta$, *where* $\rho_k = \frac{n-k_0}{n-k}$;
2. $\tau_k V_{i,k} \xrightarrow{\text{a.s.}} \tau_k\rho_k\delta - A^{-1}x_ie_i$ *for* $i \le k$, $(1-\tau_k)V_{i,k} \xrightarrow{\text{a.s.}}$ $$\begin{cases}(1-\tau_k)\rho_k\delta + A^{-1}x_i(e_i - x_i^\top\delta\rho_k), & k < i \le k_0,\\ (1-\tau_k)\rho_k\delta + A^{-1}x_i[e_i + x_i^\top\delta(1-\rho_k)], & k_0 < i \le n.\end{cases}$$

*Proof.* By the law of large numbers, we can easily prove the lemma. □

*Proof of Theorem 1.* Theorem 1 can be proved using Lemma 5 and similar arguments as in the proof of Theorem 1.3.1 of [6]. One difference is that our theorem is obtained from Theorem A.3.4 as opposed to Corollary A.3.1 of [6], as we need to derive the null distribution of $Z_n$ not $Z_n^*$. □

*Proof of Theorem 2.* Under the alternative hypothesis, by Lemma 6 we have

$$\begin{aligned} & \sup_\lambda \frac{1}{n}\sum_{i=1}^{n}\log(1+\lambda^\top V_{i,k_0}) \\ \xrightarrow{\text{a.s.}} & \sup_\lambda\Big\{\frac{1}{n}\sum_{i=1}^{k_0}\log\left[1+\lambda^\top(\delta-\tau_{k_0}^{-1}A^{-1}x_ie_i)\right] \\ & + \frac{1}{n}\sum_{i=k_0+1}^{n}\log\left[1+\lambda^\top(\delta+(1-\tau_{k_0})^{-1}A^{-1}x_ie_i)\right]\Big\} \\ \xrightarrow{\text{a.s.}} & \sup_\lambda\big\{\tau_0E\ \log\left[1+\lambda^\top(\delta-\tau_0^{-1}A^{-1}xe)\right] \\ & + (1-\tau_0)E\ \log\left[1+\lambda^\top(\delta+(1-\tau_0)^{-1}A^{-1}xe)\right]\big\} \\ \ge & \ c_0, \end{aligned}$$

where the last inequality is due to condition $(C)$. Then the conclusion is derived. □

*Proof of Theorem 3.* We only need to prove that for arbitrary small $\eta > 0$ with $\frac{\tau_{k_0}}{2} > \eta$ and $|\frac{k_0-k}{n}| \ge \eta$, $-2\log\Lambda_k$ can not arrive at its maximum with probability approaching to 1. Thus, by the definition of $\hat{k}$, we have $|\frac{k_0-\hat{k}}{n}| \le \eta$ with probability approaching to 1. Since $\eta$ is arbitrary, Theorem 3 is proved.

Without loss of generality, suppose that $k < k_0$, and $\frac{k_0-k}{n} \ge \eta$. Applying Lemma 6, we have

$$\begin{aligned} \frac{-2\log\Lambda_k}{2n} &= \sup_\lambda \frac{1}{n}\sum_{i=1}^{n}\log(1+\lambda^\top V_{i,k}) \\ =\ & \sup_\lambda\Big\{\frac{1}{n}\sum_{i=1}^{k}\log(\tau_k+\lambda^\top\tau_kV_{i,k}) \\ & + \frac{1}{n}\sum_{i=k+1}^{n}\log\left[(1-\tau_k)+\lambda^\top(1-\tau_k)V_{i,k}\right] \\ & - \tau_k\log\tau_k - (1-\tau_k)\log(1-\tau_k)\Big\} \\ \xrightarrow{\text{a.s.}}\ & \sup_\lambda\Big\{\tau_kE\ \log\left[1+\lambda^\top(\rho_k\delta-\tau_k^{-1}A^{-1}xe)\right] \\ & + (\tau_{k_0}-\tau_k)E\ \log\left[1+\lambda^\top(\rho_k\delta + (1-\tau_k)^{-1}A^{-1}x(e-x^\top\delta\rho_k))\right] \\ & + (1-\tau_{k_0})E\ \log\left[1+\lambda^\top(\rho_k\delta + (1-\tau_k)^{-1}A^{-1}x(e+x^\top\delta(1-\rho_k))\right]\Big\}. \end{aligned}$$

Denote

$$(2)\quad H = \tau_k \log\left[1+\lambda^\top(\rho_k\delta - \tau_k^{-1}A^{-1}xe)\right] + (\tau_{k_0}-\tau_k)\log\left[1+\lambda^\top(\rho_k\delta + (1-\tau_k)^{-1}A^{-1}x(e-x^\top\delta\rho_k))\right] + (1-\tau_{k_0})\log\left[1+\lambda^\top(\rho_k\delta + (1-\tau_k)^{-1}A^{-1}x(e+x^\top\delta(1-\rho_k)))\right] := H_1+H_2+H_3.$$

Then $\frac{-2\log\Lambda_k}{2n} \xrightarrow{\text{a.s.}} \sup_\lambda E(H)$. Since $\log(\cdot)$ is a strictly concave function, we have

$$(3)\quad H_1 \le \tau_{k_0}\log[1+\lambda^\top\rho_k(\delta-\tau_{k_0}^{-1}A^{-1}xe)] - (\tau_{k_0}-\tau_k)\log[1+\lambda^\top(\rho_k\delta+(1-\tau_k)^{-1}A^{-1}xe)]$$

and similarly,

$$(4)\quad H_2+H_3 \le (1-\tau_k)\log\{1+\lambda^\top[\rho_k\delta+(1-\tau_k)^{-1}A^{-1}xe]\}.$$

Recall that $\rho_k=(n-k_0)/(n-k)$. Combining (2), (3), (4) and the fact $\rho_k^{-1}(1-\tau_k)^{-1}=(1-\tau_{k_0})^{-1}$, we have

$$H \le \tau_{k_0}\log[1+\lambda^\top\rho_k(\delta-\tau_{k_0}A^{-1}xe)] + (1-\tau_{k_0})\log\{1+\lambda^\top\rho_k[\delta+(1-\tau_{k_0})^{-1}A^{-1}xe].$$

Suppose that the expectation of the left-hand-side in the above inequality attains its maximum at $\lambda^*$. If $\lambda^*=0$, it's obvious that $\frac{-2\log\Lambda_{k_0}}{2n} > \frac{-2\log\Lambda_k}{2n}$ *a.s.* If $\lambda^*\ne 0$, since $\lambda^{*\tau}A^{-1}xe$ is non-degenerate, we have

$$\begin{aligned}
&\frac{-2\log\Lambda_k}{2n}\\
< \;& \tau_{k_0}E\log[1+\lambda^{*\tau}\rho_k(\delta-\tau_{k_0}A^{-1}xe)] + (1-\tau_{k_0})E\log\{1+\lambda^{*\tau}\rho_k[\delta+(1-\tau_{k_0})^{-1}A^{-1}xe]\}\\
\le \;& \sup_\lambda\Big\{\tau_{k_0}E\log[1+\lambda^\top\rho_k(\delta-\tau_{k_0}A^{-1}xe)] + (1-\tau_{k_0})E\log\{1+\lambda^\top\rho_k[\delta+(1-\tau_{k_0})^{-1}A^{-1}xe]\}\Big\}\\
= \;& \sup_\lambda\Big\{\tau_{k_0}E\log[1+\lambda^\top(\delta-\tau_{k_0}A^{-1}xe)] + (1-\tau_{k_0})E\log\{1+\lambda^\top[\delta+(1-\tau_{k_0})^{-1}A^{-1}xe]\}\Big\}\\
\xleftarrow{\text{a.s.}} \;& \frac{-2\log\Lambda_{k_0}}{2n},
\end{aligned}$$

which implies that

$$\frac{-2\log\Lambda_{k_0}}{2n} > \frac{-2\log\Lambda_k}{2n} \quad a.s.$$

Thus, in such a situation, $-2\log\Lambda_k$ cannot arrive at its maximum with probability approaching to 1. Based on the above discussions, we can conclude the result. □

*Received 27 October 2014*

## REFERENCES


[1] Antoch, J. and Hušková, M. (2001). Permutation tests for change point analysis. *Stat. Probab. Lett.* **53** 37–46. MR1843339

[2] Bai, J. S. (1997). Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics* **79** 551–563.

[3] Bai, J. S. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121

[4] Chen, J. and Gupta, A. K. (2000). *Parametric Statistical Change Point Analysis*. Boston: Birkhäuser. MR1761850

[5] Chernoff, H. and Zacks, S. (1964). Estimating the Current Mean of a Normal Distribution Which is Subject to Changes in Time. *Annals of Mathematical Statistics* **35** 999–1018. MR0179874

[6] Csörgó, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. New York: John Wiley & Sons. MR2743035

[7] Dueker, M. J. (1992). The Response of Market Interest Rates to Discount Rate Changes. *Federal Reserve Bank of St. Louis Review* **74** 78–91.

[8] Hawkins, D. L. (1989). A U-I approach to retrospective testing for shifting parameters in a linear model. *Commun. Statist. Theor. Meth.* **18** 3117–3134. MR1033153

[9] Hušková, M. (2004). Permutation principle and bootstrap in change point analysis. *Fields Inst. Commun.* **44** 273–291. MR2106860

[10] Hušková, M. and Kirch, C. (2012). Bootstrapping sequential change-point tests for linear regression. *Metrika* **75** 673–708. MR2946654

[11] Jing, B. (1995). Two-sample empirical likelihood method. *Statist. Probab. Lett.* **24** 315–319. MR1353889

[12] Jing, B. Y., Yuan, J. Q. and Zhou, W. (2009). Jackknife Empirical Likelihood. *Journal of the American Statistical Association* **104** 1224–1232. MR2562010

[13] Koul, H. and Qian, L. (2002). Asymptotics of maximum likelihood estimator in a two-phase linear regression model. C. R. Rao 80th birthday felicitation volume, Part II. *J. Statist. Plann. Inference* **108** 99–119. MR1947394

[14] Liu, Z. and Qian, L. (2010). Changepoint Estimation in a Segmented Linear Regression via Empirical Likelihood. *Comm. Statist. Simulation Comput.* **39** 85–100. MR2784554

[15] Mahmoud, M. A., Parker, P. A., Woodall, W. H. and Hawkins, D. M. (2007). A change point method for linear profile data. *Qual. Reliab. Eng. Int.* **23** 247–268.

[16] Miller, R. G. (1974). An Unbalanced Jackknife. *Annals of Statistics* **2** 880–891. MR0356353

[17] Otsu, T. (2010). On Bahadur efficiency of empirical likelihood. *Journal of Econometrics* **157** 248–256. MR2661598

[18] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. MR0946049

[19] Owen, A. B. (2001). *Empirical Likelihood*. New York: Chapman & Hall/CRC.

[20] Page E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–235. MR0088850

[21] Page E. S. (1955). A test for a chance in a parameter occurring at an unknown point. *Biometrika* **42** 523–527. MR0072412

[22] Perron, P. and Vogelsang, T. J. (1992). Testing for a unit root in a time series with a changing mean: corrections and extensions. *J. Bus. Econ. Statist.* **10** 467–470.

[23] Vostrikova, L. J.(1981). Detecting “disorder” in multidimensional random processes. *Soviet Mathematics Doklady* **24** 55–59.

[24] Wu, X. Q., Zhang, S. G. and Ning, W. (2013). Empirical Likelihood Ratio Based Test for Change Point Detection in Linear Regression Model. *To appear in Acta Mathematicae Applicatae Sinica.*

[25] Zou, C., Liu, Y., Qin, P. and Wang, Z. (2007). Empirical likelihood ratio test for the change-point problem. *Statistics & Probability Letters* **77** 374–382. MR2339041

Xinqi Wu
School of Mathematical Sciences
University of Chinese Academy of Sciences
China
Key Laboratory of Big Data Mining
and Knowledge Management
Chinese Academy of Sciences
China
E-mail address: wuxinqi09@mails.ucas.ac.cn

Sanguo Zhang
School of Mathematical Sciences
University of Chinese Academy of Sciences
China
Key Laboratory of Big Data Mining
and Knowledge Management
Chinese Academy of Sciences
China
E-mail address: sgzhang@ucas.ac.cn

Qingzhao Zhang
School of Mathematical Sciences
University of Chinese Academy of Sciences
China
Key Laboratory of Big Data Mining
and Knowledge Management
Chinese Academy of Sciences
China
E-mail address: qzzhang@ucas.ac.cn

Shuangge Ma
Department of Biostatistics
Yale University
New Haven, CT 06520
USA
E-mail address: shuangge.ma@yale.edu