

# CUSUM 型统计量中调节参数对变点估计效果的影响分析

谭常春, 江敏

(合肥工业大学经济学院, 安徽合肥 230009)

**摘要:** CUSUM 型变点估计量中的调节参数, 理论上一般假定其取值范围为  $(0, 1)$ , 但在实际数据的变点估计时, 不同的取值往往得到相异的结果. 基于跳跃度变点模型, 利用蒙特卡罗方法, 研究了调节参数的取值对变点估计结果的影响. 模拟结果发现: 在跳跃度较大时, 无论变点真实位置如何, 变点估计值基本不受调节参数取值的影响; 但在跳跃度较小时, 调节参数的取值对变点估计结果有显著影响; 特别在变点真实位置靠近序列其中任意一个端点时, 调节参数取值为 0.5 时, 估计的效果最好; 当变点真实位置靠近中间时, 调节参数的取值越小, 估计效果越好. 在模拟和实例分析基础上, 提出了基于数据驱动的调节参数的选取方法, 使得 CUSUM 型变点估计量更具稳健性.

**关键词:** 变点; CUSUM 型估计量; 调节参数

**中图分类号:** O212.7 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2020.07.008

**2010 Mathematics Subject Classification:** Primary 62G03; Secondary 62C05

**引用格式:** 谭常春, 江敏. CUSUM 型统计量中调节参数对变点估计效果的影响分析[J]. 中国科学技术大学学报, 2020, 50(7): 920-928.

TAN Changchun, JIANG Min. Influence analysis of tuning parameters on the change-point estimation in CUSUM type statistics[J]. Journal of University of Science and Technology of China, 2020, 50(7): 920-928.

## Influence analysis of tuning parameters on the change-point estimation in CUSUM type statistics

TAN Changchun, JIANG Min

(School of Economics, Hefei University of Technology, Hefei 230009, China)

**Abstract:** Generally, the range of tuning parameters in CUSUM type change-point estimation statistic is assumed to be  $(0, 1)$  in theory. But the different values of tuning parameters often lead to the different estimation results in application. Here Monte Carlo method was used to study the influence of tuning parameters on the change-point estimation based on the jump change-point model. It was found that when the jump is large, the change-point estimate is not affected by the value of tuning parameters no matter where the true location of the change-point is. However, the value of tuning parameter has a significant effect on the change-point estimate when the jump is small. Especially, when the true location of change-point is close to one of the two trails, best estimation is obtained with the tuning parameter at 0.5. When the true location of change-point is near the center of sequence, it was observed that the smaller the tuning parameter, the better the estimation. On the basis of simulation and applications, a data-driven method was proposed to select appropriate tuning parameters from a set of possible values, which makes the CUSUM type change-point estimator more robust.

**Key words:** change-point; cumulative sum estimator; tuning parameter

### 0 引言

变点的检验和估计问题一直是统计学中的热门课题之一. 变点问题在医学、经济、金融、气象、人口等领域都有着较为广泛的应用. 当一系列数据之中存在变点, 则在变点前后的子序列必然会呈现不同

的统计学特征, 利用这些特征差异便可以估计出变点位置. Page<sup>[1]</sup> 首次提出了累积和 (cumulative sum, CUSUM) 方法, 针对变点的连续检验问题有着非常好的检验效果.

近年来, CUSUM 方法被许多统计学者进行改进和应用于变点的检测和估计, 并取得了不错的成

收稿日期: 2020-04-24; 修回日期: 2020-06-16

基金项目: 国家社会科学基金一般项目(16BTJ023)资助.

作者简介: 谭常春(通讯作者), 男, 1977年生, 博士/教授. 研究方向: 变点问题统计推断, 面板数据分析, 金融变结构分析.

E-mail: cctan@hfut.edu.cn

果. Giraitis 等<sup>[2]</sup>应用 CUSUM 方法研究了平稳正时间序列的边际分布函数的变点问题. Csorgo 和 Horvath<sup>[3]</sup>出版了关于 CUSUM 型方法研究变点的理论专著. Horvath 和 Kokoszka<sup>[4]</sup>应用 CUSUM 方法研究了序列长相依性对变点估计的影响,并在模拟分析中将调节参数分别设为  $0, (0, 0.7], (0.7, 1], 1$ , 发现调节参数的不同取值对模拟结果有显著影响. Kokoszka 和 Leipus<sup>[5]</sup>在 ARCH 模型的变点检验问题中,建立了调节参数值为 0 的 CUSUM 检验统计量,并研究了该检验统计量的渐近分布. Hariz 和 Welie<sup>[6]</sup>构造了相依序列下调节参数为 0 的 CUSUM 变点估计量,并给出了估计量的收敛速度. Hariz 等<sup>[7]</sup>进一步研究了非平稳序列的 CUSUM 型变点估计量的收敛速度,并运用调节参数为 0 的 CUSUM 估计量模拟了各种非平稳序列的变点估计情况. Nie 等<sup>[8]</sup>建立了修正的 CUSUM 型变点估计量,模拟分析显示调节参数的取值对估计结果有不可忽视的影响,并在独立序列下建议调

节参数取值为  $\frac{1}{3}$ . Fremdt<sup>[9]</sup>提出了改进的 CUSUM

变点检测统计量,发现当调节参数的取值在  $(0, 0.5)$  时,能提高检验的稳定性. Chen 和 Hu<sup>[10]</sup>针对面板数据提出了一种新的 CUSUM 型均值变点估计量,模拟分析发现,变点位置在两边时,调节参数取值越大则估计效果越好,变点位置在中间时,调节参数取值越小则估计效果越好,最终选择的取值为 0.5. Qin 和 Ma<sup>[11]</sup>应用调节参数为 0 的 CUSUM 方法估计方差变点,并分析了变点位置对估计效果的影响. Xu 等<sup>[12]</sup>通过格子点模拟方法选择调节参数,建立了 CUSUM 型方差变点检验统计量,并给出统计量的渐近分布.

在目前所有带有调节参数的 CUSUM 型变点检测或估计统计量中,对于调节参数取值的研究,一部分是仅仅假定其取值范围为  $(0, 1)$ ,不对其具体取值进行分析;而另一部分则是根据所研究模型的估计结果,人为选择一个相对合适的值. 这些研究都没有详细讨论过调节参数对 CUSUM 估计量的影响,因此有必要进一步讨论和分析 CUSUM 型变点量中调节参数  $\gamma$  的取值问题.

本文首先简单介绍 CUSUM 变点估计方法;其次分别对独立样本和非独立样本序列,利用蒙特卡罗方法,在不同的调节参数、不同变点位置设置下,探究调节参数取值对 CUSUM 型变点估计量的估计结果的影响;最后提出了一种基于数据驱动的调节参数的选取方法,提高了变点估计量的稳健性,并将其用于美国消费者价格指数数据的变点估计分析.

## 1 CUSUM 型均值变点估计量介绍

考虑如下跳跃均值变点模型:

$$Y_i = \begin{cases} \alpha + X_i, & i = 1, \dots, k_0; \\ \alpha + \delta + X_i, & i = k_0 + 1, \dots, n \end{cases} \quad (1)$$

式中,  $k_0$  是未知的,记  $\theta_0 = \frac{k_0}{n}, 0 < \theta_0 < 1$ ,称  $\theta_0$  或  $k_0$

为变点;  $\{X_i, i = 1, 2, \dots, n\}$  是一随机变量序列,满足  $E(X_i) = 0, E(X_i^2) < \infty; \delta (\neq 0)$  为跳跃度,反映了序列在变点处均值变化的大小. 应用 CUSUM 方法得变点  $k_0$  的估计  $\hat{k}$  为

$$\hat{k} = \min(\arg \max_{1 < k < n} |U_k|) \quad (2)$$

式中,

$$U_k = \left(\frac{k(n-k)}{k}\right)^{1-\gamma} \left(\frac{1}{k} \sum_{i=1}^k Y_i - \frac{1}{n-k} \sum_{i=k+1}^n Y_i\right) \quad (3)$$

$\gamma \in (0, 1)$ , 是调节参数. 类似,变点  $\theta_0$  的估计为

$$\hat{\theta} = \frac{\hat{k}}{n} \quad (4)$$

## 2 调节参数 $\gamma$ 的取值对变点位置估计的影响分析

为了研究调节参数  $\gamma$  对变点估计量的影响性模拟分析,假设变点真实位置  $\theta_0$ . 选择为 0.1, 0.2,  $\dots$ , 0.9, 共 9 种情形,均值跳跃度  $\delta$  选择为 1, 2, 3; 调节参数  $\gamma$  取值为  $(0, 1]$  上的格子点,即分别取值  $\frac{j}{200}, j = 1, 2, \dots, 200$ ; 重复模拟次数  $M = 1000$ ; 样本容量  $n$  选取为 200.

由于序列的相依性可能会对估计结果产生影响,因此模拟分析中,分别考虑独立样本序列和相依样本序列下调节参数  $\gamma$  对变点估计量的影响性分析.

### 2.1 独立样本序列

假定  $X_i, i = 1, 2, \dots, n, i. i. d. \sim N(0, 1)$ , 此处为方便起见,假设方差为 1. 在给定调节参数为  $\gamma$  时,记第  $j$  次重复的,所得变点估计量为  $\hat{\theta}(j, \gamma), j = 1, 2, \dots, M$ ; 取  $\hat{\theta}(j, \gamma)$  的均值作为对应调节参数下的变点估计量,并记为  $\hat{\theta}(\gamma)$ , 即

$$\hat{\theta}(\gamma) = \frac{1}{M} \sum_{j=1}^M \hat{\theta}(j, \gamma) \quad (5)$$

模拟可得给定真实变点位置  $\theta_0$ . 下调节参数  $\gamma$  与变点估计量  $\hat{\theta}(\gamma)$  之间的关系,如图 1 所示.

由图 1(a) 发现,当跳跃度  $\delta = 1$  时,不论变点真实位置  $\theta_0$  如何,当  $\gamma \in (0.7, 1)$  时,变点估计值与真实值之间有较为明显的偏差;当变点位置  $\theta_0$  靠近端点时,仅在  $\gamma = 0.5$  附近时,变点估计值与真实值之间基本没有差异. 当跳跃度  $\delta = 2$  时,如图 1(b) 所示,在  $\gamma \in (0.8, 1)$  时,变点估计值与真实位置的偏差较为明显;当  $\gamma \in (0, 0.2)$ , 变点真实位置  $\theta_0$  靠近端点时,则变点估计值与真实位置有一定的偏差. 由图 1(c) 可得,  $\delta = 3$  时, CUSUM 型变点估计量的估计结果总体较好;仅在  $\gamma \in (0.9, 1)$  时,变点估计值与真实值有较为显著的差异;当  $\gamma \in (0, 0.1)$  时,且变点位置  $\theta_0$  靠近端点,则估计结果有些许偏差.

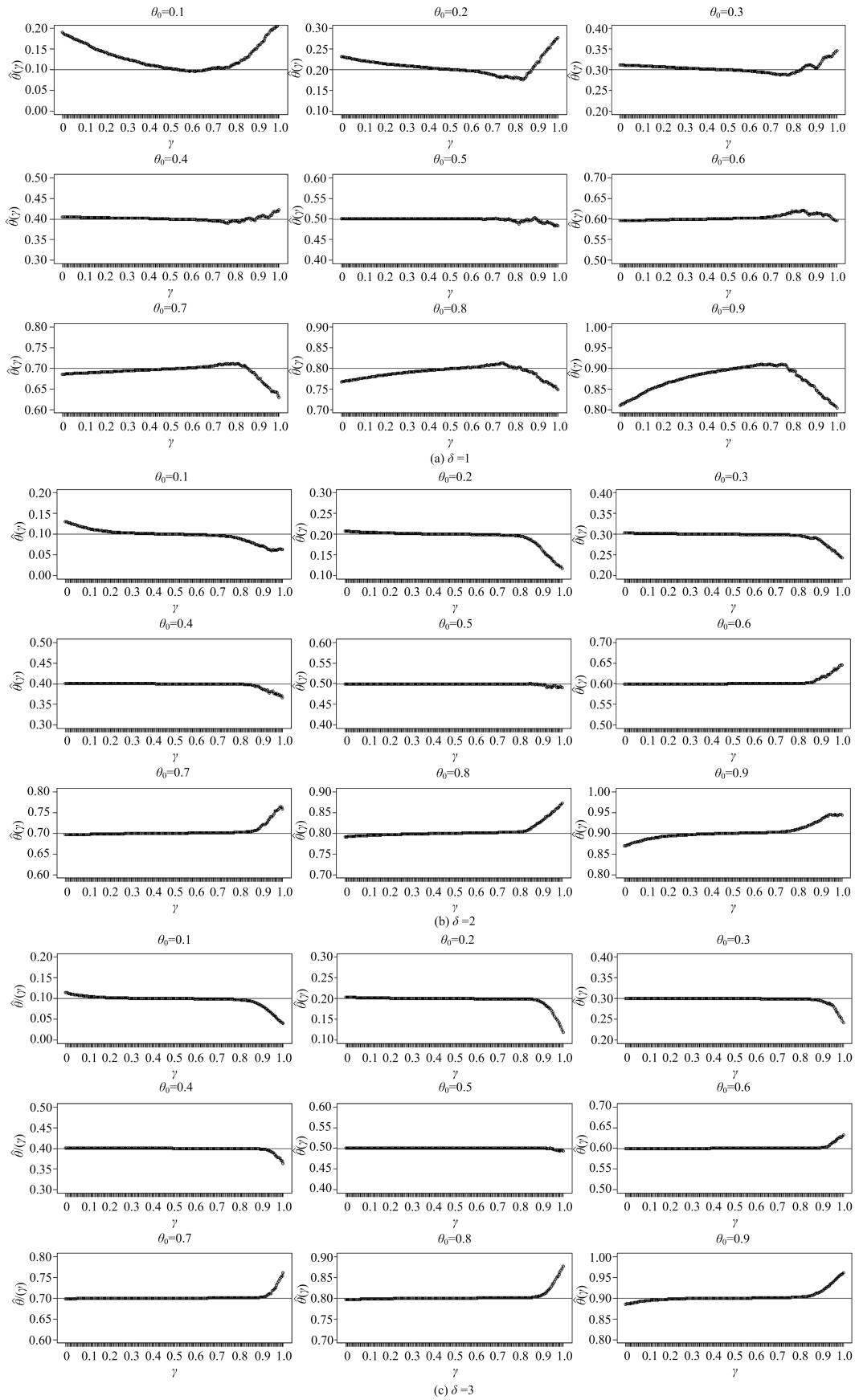


图 1 独立样本下  $\hat{\theta}(\gamma)$  与  $\gamma$  之间的关系

Fig. 1 The relation between  $\hat{\theta}(\gamma)$  and  $\gamma$  for independent sequence

综合上述不同跳跃度下的模拟结果,可得如下结论:变点真实位置  $\theta_0$  靠近端点时,令调节参数  $\gamma$  取值 0.5,则 CUSUM 型变点估计量的效果最好;变点位置  $\theta_0$  靠近中间时,调节参数  $\gamma$  的取值在 (0, 0.7) 之间,较为合适。

另外,上述模拟结果是基于多次估计的均值(点估计)来评估估计量的准确性,为了更全面评价调节参数  $\gamma$  取值对变点估计量的影响,下面基于区间估计来进一步分析.选定区间  $[\theta_0 - 0.05, \theta_0 + 0.05]$  作为参考区间,计算  $M$  次重复模拟中调节参数  $\gamma$  取不同值情况下变点估计值  $\hat{\theta}(j, \gamma)$ ,  $j = 1, 2, \dots, M$ ,落在  $[\theta_0 - 0.05, \theta_0 + 0.05]$  中的频率  $R(\gamma)$ ,

则  $R(\gamma)$  与  $\gamma$  的关系图 2 所示,其中,

$$R(\gamma) = \frac{\#\{\hat{\theta}(j, \gamma) \in [\theta_0 - 0.05, \theta_0 + 0.05]\}}{M}, \quad j = 1, 2, \dots, M \quad (6)$$

式中,  $\#\{\}$  表示  $M$  次重复模拟中估计值落入区间中的频数.由图 2 可发现,变点位置  $\theta_0$  固定时,调节参数  $\gamma$  取值对变点估计量有显著的影响.当变点真实值  $\theta_0$  靠近端点时( $\theta_0 = 0.1, 0.2$  或者  $0.8, 0.9$ ),  $\gamma$  取值越接近 0.5 时,变点估计的准确率越高;而当  $\theta_0$  靠近中间时( $\theta_0 = 0.5$ ),  $\gamma$  取值越接近 1,变点估计准确率越低。

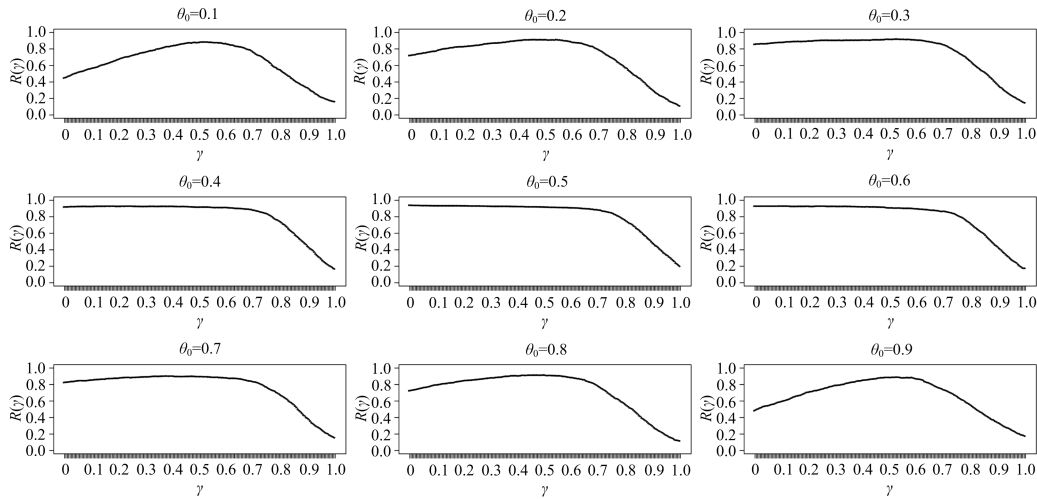


图 2 独立样本下变点估计的准确率  $R(\gamma)$  与  $\gamma$  的关系;  $\delta = 1$   
 Fig. 2 The relation between  $R(\gamma)$  and  $\gamma$  with  $\delta = 1$  for independent sequence

### 2.2 相依样本序列

在相依样本序列下,假设  $X_i, i = 1, 2, \dots, n$ , 满足较为常见的一阶自回归 AR(1) 过程,即

$$X_i = \phi X_{i-1} + \epsilon_i, \quad i = 1, 2, \dots, n \quad (7)$$

式中,  $\epsilon_i$  i. i. d.  $\sim N(0, 1)$ . 为满足平稳性条件,假定系数  $\phi \in (0, 1)$ , 此处令  $\phi = 0.9$ , 即

$$X_i = 0.9 X_{i-1} + \epsilon_i, \quad i = 1, 2, \dots, n \quad (8)$$

类似独立样本序列下的变点估计量模拟计算,可得到  $\hat{\theta}(\gamma)$  与  $\gamma, R(\gamma)$  与  $\gamma$  之间的关系,如图 3 和 4 所示。

由图 3(a) 可知,当跳跃度  $\delta = 1$  时,在  $\gamma \in (0.7, 1)$  时,变点估计值与实际位置的偏差较为显著;当变点位置  $\theta_0$  靠近端点时,无法较好地估计出变点位置.当  $\delta = 2$  时,如图 3(b) 所示,当  $\gamma \in (0.7, 1)$  时,变点估计值与实际位置的偏差较为明显;当变点位置  $\theta_0$  靠近端点时,仅在  $\gamma = 0.5$  附近时,有较好的估计.由图 3(c) 可知,当  $\delta = 3$ , 即跳跃度较大时,估计的结果总体良好;仅在  $\gamma \in (0.8, 1)$  时,变点估计值与真实位置有较为显著的差异.由图 4 可以看出,在相依样本序列下调节参数  $\gamma$  对准确率  $R(\gamma)$  的影响与独立样本序列下有类似的结论。

综合图 1~4,应用 CUSUM 方法估计相依样本序列的变点时,调节参数  $\gamma$  对变点估计值的影响与独立样本序列的情形总体上是类似的.当变点真实位置  $\theta_0$  靠近端点时,调节参数  $\gamma$  取值为 0.5,估计

的效果最好;变点位置  $\theta_0$  位于中间时,调节参数  $\gamma$  取值越接近 1,变点估计的效果越差。

结合独立样本序列和相依样本序列的模拟结果,可以得到:应用 CUSUM 方法估计变点时,当均值跳跃度  $\delta$  较大时,调节参数  $\gamma$  的取值对于估计结果的影响可以忽略;但是在均值跳跃度  $\delta$  较小时,调节参数  $\gamma$  的取值对估计结果有较显著的影响;变点位置  $\theta_0$  靠近端点时,调节参数  $\gamma$  的取值在 0.5 附近时,则估计效果最好;变点位置  $\theta_0$  靠近中间时,调节参数  $\gamma$  的取值不宜过大,在 (0, 0.7) 内较为合适。

## 3 基于数据驱动的调节参数 $\gamma$ 的选取方法

### 3.1 算法介绍

虽然变点真实位置  $\theta_0$  对调节参数  $\gamma$  的最优选取是有影响的,但是在实际应用中,变点真实位置  $\theta_0$  是未知的.因此需要寻找一种仅基于样本数据来选取恰当调节参数  $\gamma$  的方法,以使得 CUSUM 型变点估计量的估计效果更加稳健.受到 Tan 等<sup>[13]</sup> 启发,本文提出基于数据驱动的调节参数选取算法。

#### 算法 3.1

第一步:设  $Y_1, Y_2, \dots, Y_{k_0}, Y_{k_0+1}, \dots, Y_n$  是带有变点  $k_0$  的样本序列,记  $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_l\}$  为调节参数  $\gamma$  的可能取值集.给定  $\gamma_i$ ,得到一个变点估计量,



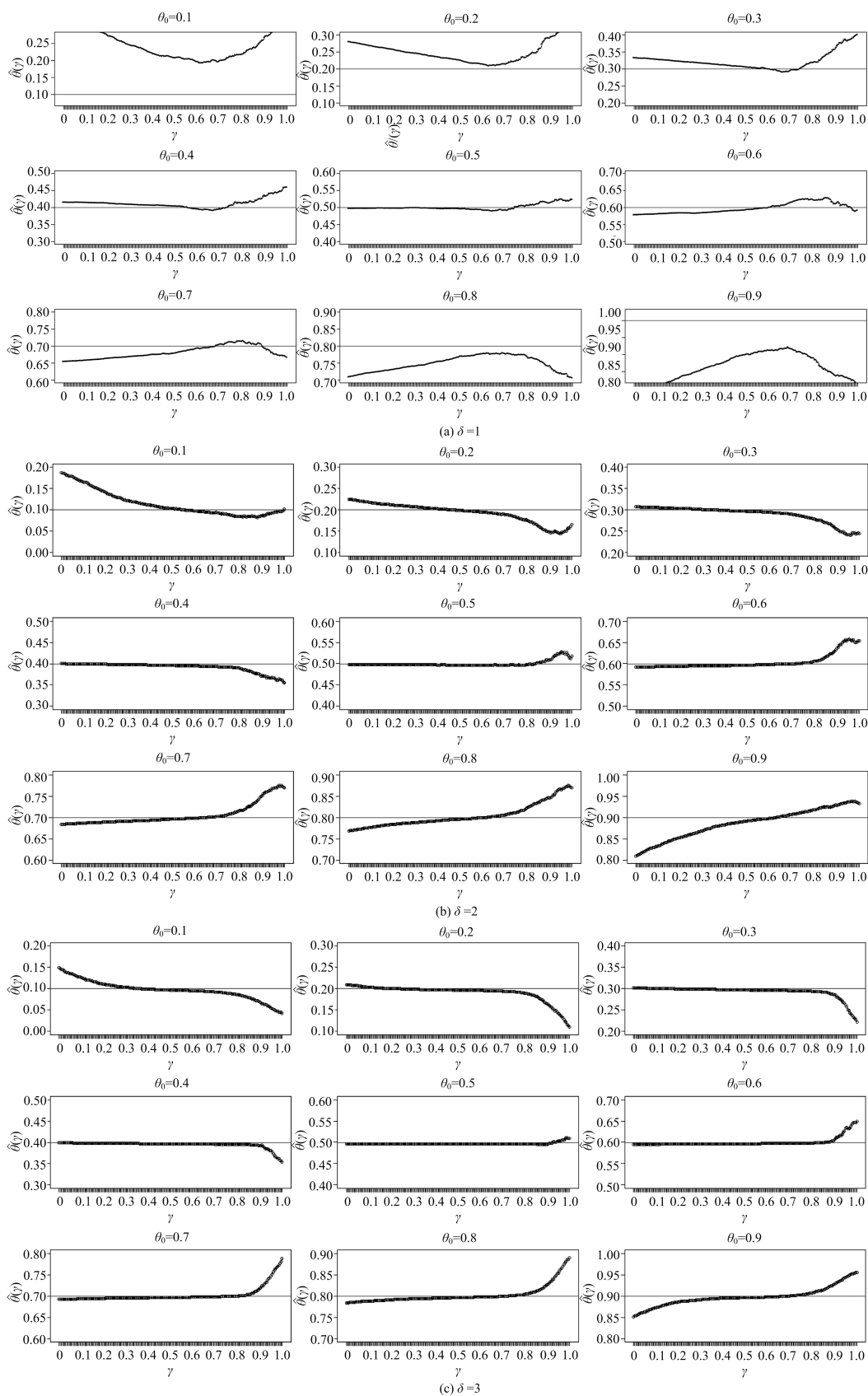


图 3 相依序列下  $\hat{\theta}(\gamma)$  与  $\gamma$  之间的关系

Fig. 3 The relation between  $\hat{\theta}(\gamma)$  and  $\gamma$  for dependent sequence

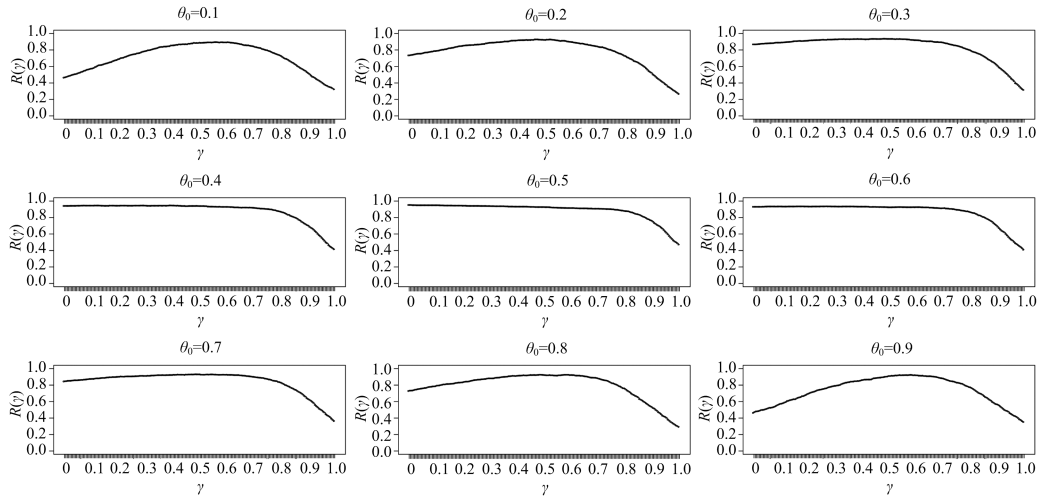


图 4 相依序列下变点估计准确率  $R(\gamma)$  与  $\gamma$  的关系;  $\delta=2$   
 Fig. 4 The relation between  $R(\gamma)$  and  $\gamma$  with  $\delta=2$  for dependent sequence

记为  $\hat{k}_{\gamma_i}$ , 有

$$\hat{k}_{\gamma_i} = \arg \max_{1 \leq k < n} |U_k|, 1 \leq i \leq l \quad (9)$$

第二步: 计算  $\hat{k}_{\gamma_1}, \hat{k}_{\gamma_2}, \dots, \hat{k}_{\gamma_l}$  的均值  $\bar{k}$ :

$$\bar{k} = \frac{1}{l} \sum_{i=1}^l \hat{k}_{\gamma_i} \quad (10)$$

第三步: 定义

$$\gamma_s = \arg \min_{\gamma_i} |\hat{k}_{\gamma_i} - \bar{k}| \quad (11)$$

即为基于数据驱动方法选取的合适的调节参数。

**定理 3.1** 假定序列  $Y_1, Y_2, \dots, Y_{k_0}, Y_{k_0+1}, \dots, Y_n$  为来自模型(1)的样本序列, 其中  $k_0$  为变点, 记  $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_l\}$  为调节参数  $\gamma$  的可能取值集, 则至少存在一个  $\gamma^* \in \Gamma$ , 使得

$$|\hat{k}_{\gamma_s} - k_0| \leq |\hat{k}_{\gamma^*} - k_0| \quad (12)$$

**证明** 不妨假设变点  $k_0 \geq \bar{k}$ , 则有

$$|\hat{k}_{\gamma_s} - k_0| = |\hat{k}_{\gamma_s} - \bar{k} + \bar{k} - k_0| \leq |\hat{k}_{\gamma_s} - \bar{k}| + |\bar{k} - k_0|,$$

因为对任意的  $\gamma_i \in \Gamma, 1 \leq i \leq l$ , 由式(11)中  $\gamma_s$  的定义可知,

$$|\hat{k}_{\gamma_s} - \bar{k}| \leq |\hat{k}_{\gamma_i} - \bar{k}|, 1 \leq i \leq l,$$

故有

$$|\hat{k}_{\gamma_s} - k_0| \leq |\hat{k}_{\gamma_i} - \bar{k}| + |\bar{k} - k_0|, 1 \leq i \leq l.$$

又因为在  $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_l\}$  中至少存在一个  $\gamma^* \neq$

$\gamma_s$ , 使得  $\hat{k}_{\gamma^*} \leq \min(\hat{k}_{\gamma_s}, \bar{k})$ , 因此

$$|\hat{k}_{\gamma_s} - k_0| \leq |\hat{k}_{\gamma^*} - \bar{k}| + |\bar{k} - k_0| \leq |\hat{k}_{\gamma^*} - k_0| + |\bar{k} - k_0| = |\hat{k}_{\gamma^*} - k_0| \quad (13)$$

对  $k_0 < \bar{k}$  情形, 类似可证明得式(13)。

### 3.2 模拟分析

为了验证基于数据驱动方法选择的调节参数  $\gamma_s$  的稳健性, 通过模拟分析, 比较在调节参数  $\gamma$  取不同值以及  $\gamma = \gamma_s$  的情况下 CUSUM 变点估计量的准确率。

在独立样本序列假设下, 类似图 2 中的准确率计算, 调节参数  $\gamma$  分别取值为  $\gamma = 0.1, 0.2, \dots, 0.9, 1$ , 选定  $[\theta_0 - a, \theta_0 + a]$  为参考区间, 其中  $a = 0.025, 0.05, 0.075$ , 模拟重复次数  $M = 1000$ , 计算变点估计值  $\hat{\theta}(j, \gamma), j = 1, 2, \dots, M$ , 落入各参考区间的频率, 如表 1 所示。从表 1 可以看出, 基于数据驱动方法选择的调节参数  $\gamma_s$  可能不是最佳的, 但在大多数情形下都是较为合适的, 不会出现估计效果较差的情况。对于相依样本序列情形, 在其他参数设置不变情况下, 假定均值跳跃度  $\delta = 2$ , 准确率计算结果见表 2, 由表 2 可以发现, 相依序列下的变点估计量的准确率与独立样本序列的准确率有类似结论, 即不论真实变点位置如何, 由基于数据驱动的方法选择出的参数  $\gamma_s$  是较为合适的, 能有效保证变点估计量的稳健性。

表 1 独立样本序列下变点估计值的准确率;  $\delta=1$

Tab. 1 The accuracy rate of change-point estimation with  $\delta=1$  for independent sequence

$\theta_0$	$a$	$\gamma$											$\gamma_s$
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	
0.1	0.025	0.376	0.472	0.555	0.664	0.743	0.779	0.764	0.683	0.557	0.403	0.226	0.734
	0.05	0.45	0.554	0.672	0.781	0.863	0.9	0.904	0.843	0.735	0.529	0.316	0.841
	0.075	0.549	0.644	0.746	0.845	0.911	0.946	0.953	0.929	0.853	0.688	0.479	0.897
0.2	0.025	0.591	0.663	0.717	0.776	0.794	0.793	0.759	0.708	0.591	0.4	0.166	0.688
	0.05	0.761	0.812	0.863	0.89	0.915	0.92	0.896	0.851	0.743	0.51	0.251	0.841
	0.075	0.824	0.869	0.913	0.937	0.96	0.966	0.952	0.917	0.813	0.572	0.292	0.9

续表 1

$\theta_0$	$a$	$\gamma$											
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	$\gamma_s$
0.3	0.025	0.738	0.765	0.798	0.807	0.81	0.796	0.772	0.735	0.657	0.473	0.242	0.732
	0.05	0.851	0.876	0.894	0.911	0.924	0.922	0.913	0.87	0.795	0.619	0.34	0.861
	15	0.92	0.939	0.951	0.958	0.967	0.969	0.962	0.925	0.855	0.672	0.369	0.915
0.4	0.025	0.832	0.833	0.836	0.84	0.838	0.833	0.812	0.795	0.733	0.6	0.346	0.784
	0.05	0.934	0.936	0.934	0.933	0.93	0.924	0.912	0.89	0.844	0.692	0.376	0.903
	0.075	0.965	0.964	0.965	0.965	0.964	0.962	0.952	0.93	0.885	0.733	0.401	0.935
0.5	0.025	0.827	0.82	0.814	0.811	0.805	0.798	0.792	0.786	0.762	0.634	0.374	0.799
	0.05	0.956	0.953	0.946	0.94	0.934	0.929	0.927	0.912	0.875	0.729	0.428	0.928
	0.075	0.987	0.985	0.981	0.977	0.972	0.968	0.964	0.951	0.914	0.759	0.442	0.964
0.6	0.025	0.802	0.808	0.808	0.803	0.806	0.804	0.799	0.782	0.736	0.594	0.347	0.779
	0.05	0.927	0.935	0.935	0.934	0.926	0.927	0.92	0.907	0.87	0.722	0.436	0.889
	0.075	0.969	0.971	0.969	0.968	0.962	0.962	0.959	0.943	0.909	0.759	0.456	0.926
0.7	0.025	0.686	0.718	0.744	0.775	0.787	0.789	0.784	0.759	0.677	0.525	0.292	0.738
	0.05	0.845	0.868	0.886	0.903	0.911	0.909	0.907	0.891	0.822	0.633	0.36	0.839
	0.075	0.912	0.922	0.934	0.95	0.959	0.959	0.958	0.937	0.87	0.674	0.385	0.892
0.8	0.025	0.534	0.61	0.679	0.733	0.779	0.787	0.791	0.75	0.639	0.451	0.227	0.72
	0.05	0.716	0.775	0.831	0.873	0.911	0.921	0.916	0.887	0.784	0.554	0.297	0.855
	0.075	0.81	0.853	0.892	0.924	0.952	0.97	0.966	0.943	0.839	0.609	0.323	0.908
0.9	0.025	0.343	0.45	0.566	0.661	0.745	0.793	0.79	0.73	0.593	0.41	0.237	0.73
	0.05	0.468	0.593	0.704	0.791	0.861	0.897	0.919	0.878	0.747	0.553	0.333	0.869
	0.075	0.544	0.659	0.768	0.848	0.908	0.939	0.956	0.942	0.858	0.694	0.484	0.916

表 2 相依样本序列下变点估计量的准确率;  $\delta=2$ Fig. 2 The accuracy rate of change-point estimation with  $\delta=2$  for dependent sequence

$\theta_0$	$a$	$\gamma$											
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	$\gamma_s$
0.1	0.025	0.312	0.412	0.527	0.636	0.716	0.757	0.719	0.591	0.405	0.211	0.088	0.64
	0.05	0.44	0.55	0.653	0.764	0.843	0.881	0.856	0.736	0.53	0.303	0.143	0.765
	0.075	0.515	0.631	0.74	0.82	0.895	0.927	0.926	0.825	0.66	0.44	0.264	0.85
0.2	0.025	0.557	0.625	0.679	0.717	0.756	0.768	0.736	0.67	0.472	0.236	0.089	0.662
	0.05	0.719	0.782	0.832	0.868	0.9	0.901	0.876	0.802	0.573	0.295	0.116	0.796
	0.075	0.804	0.865	0.908	0.932	0.948	0.957	0.938	0.848	0.634	0.362	0.153	0.879
0.3	0.025	0.707	0.739	0.757	0.773	0.776	0.776	0.765	0.716	0.582	0.337	0.139	0.708
	0.05	0.866	0.887	0.903	0.915	0.924	0.927	0.909	0.851	0.701	0.411	0.169	0.849
	15	0.915	0.932	0.946	0.955	0.958	0.961	0.942	0.893	0.719	0.42	0.169	0.877
0.4	0.025	0.768	0.77	0.775	0.772	0.774	0.772	0.757	0.724	0.613	0.364	0.135	0.745
	0.05	0.923	0.926	0.934	0.934	0.93	0.924	0.912	0.873	0.741	0.447	0.163	0.894
	0.075	0.968	0.97	0.969	0.968	0.965	0.967	0.953	0.919	0.766	0.441	0.169	0.932
0.5	0.025	0.791	0.786	0.777	0.776	0.764	0.758	0.751	0.731	0.613	0.357	0.161	0.752
	0.05	0.947	0.942	0.935	0.93	0.921	0.917	0.909	0.889	0.737	0.432	0.19	0.912
	0.075	0.978	0.976	0.973	0.972	0.969	0.964	0.948	0.925	0.794	0.476	0.2	0.951
0.6	0.025	0.767	0.774	0.775	0.782	0.776	0.769	0.759	0.739	0.622	0.387	0.154	0.743
	0.05	0.915	0.921	0.922	0.925	0.921	0.919	0.911	0.876	0.747	0.467	0.18	0.878
	0.075	0.96	0.959	0.965	0.964	0.962	0.957	0.947	0.921	0.753	0.446	0.172	0.936
0.7	0.025	0.698	0.731	0.76	0.773	0.779	0.781	0.765	0.728	0.588	0.328	0.129	0.727
	0.05	0.839	0.871	0.897	0.911	0.922	0.926	0.899	0.847	0.691	0.388	0.155	0.853
	0.075	0.905	0.935	0.947	0.957	0.967	0.964	0.96	0.907	0.735	0.443	0.178	0.891

续表 2

$\theta_0$	$a$	$\gamma$											
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	$\gamma_s$
0.8	0.025	0.566	0.641	0.686	0.732	0.754	0.773	0.743	0.663	0.488	0.26	0.107	0.669
	0.05	0.713	0.782	0.828	0.874	0.904	0.919	0.884	0.802	0.595	0.331	0.134	0.811
	0.075	0.79	0.845	0.894	0.932	0.956	0.961	0.947	0.874	0.664	0.356	0.135	0.863
0.9	0.025	0.329	0.424	0.523	0.638	0.69	0.719	0.694	0.606	0.421	0.238	0.118	0.639
	0.05	0.449	0.548	0.658	0.767	0.831	0.867	0.845	0.748	0.549	0.333	0.186	0.781
	0.075	0.512	0.64	0.75	0.838	0.89	0.935	0.931	0.82	0.649	0.428	0.25	0.841

3.3 实证分析

消费者价格指数 (consumer price index, CPI) 用来表示通货膨胀的程度, 是反映一国经济发展情况的一个非常重要的经济指标. 将美国 1970 年至 1990 年共 21 年的平均 CPI<sup>[14]</sup> 转换为环比增长率数据, 如图 5 所示. 在不同的调节参数取值的情况下, 应用 CUSUM 方法估计该数据序列的均值变点位置, 结果如表 3 所示. 由表 3 的估计结果可以看到, 调节参数  $\gamma$  的选取对变点估计的结果会有影响, 在调节参数  $\gamma$  取值较小时, CUSUM 方法估计得到的变点位置为 1981 年, 而在调节参数  $\gamma$  取值较大时, 估计得到的变点位置为 1982 年.

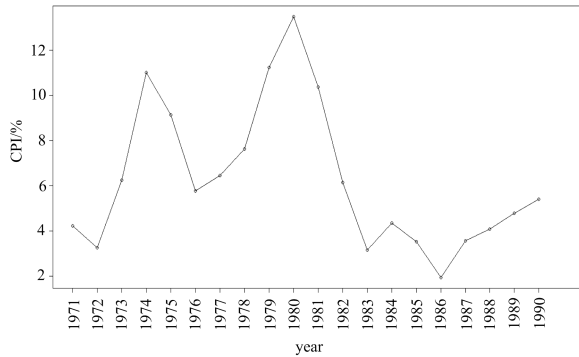


图 5 美国 1971 年至 1990 年 CPI 环比增长率数据  
Fig. 5 Year-to-year growth rate of US annual average CPI from 1971 to 1990

表 3 不同调节参数情况下的变点估计结果  
Tab. 3 Change-point estimations under different values of tuning parameter

year	$\hat{k}$	$\gamma$
1981	11	0
1981	11	0.1
1981	11	0.2
1982	12	0.3
1982	12	0.4
1982	12	0.5
1982	12	0.6
1982	12	0.7
1982	12	0.8
1982	12	0.9

为检验本文提出的基于数据驱动的调节参数选择算法在实际应用中的有效性, 将其应用到上述美国 CPI 变点估计问题中. 令调节参数  $\gamma$  的可能取

值集为  $\gamma = \left\{ \frac{j}{200}, j = 1, 2, \dots, 200 \right\}$ , 应用所提出的基于数据驱动的调节参数  $\gamma$  的选取方法, 选定 1971 年至 1990 年美国年度平均 CPI 环比增长率数据所适用的调节参数  $\gamma_s = 0.24$ , 并计算 CUSUM 型变点估计量(2), 得出变点估计位置为 1982 年.

王伟<sup>[15]</sup> 利用基于 M 估计的线性回归模型检验 1970 年至 1990 年美国年度平均 CPI 的变点时, 将 1981 年作为变点. 我们应用 CUSUM 方法给出美国 CPI 的变点估计值为 1982 年, 更符合实际. 因为美国在 20 世纪 70 年代通货膨胀严重, 出现了经济萧条的情况, 1980 年至 1982 年期间经历了第二次经济危机, 后来经过一系列的宏观调控, 1982 年结束了经济危机, 扭转了通货膨胀的态势, 1983 年恢复了经济增长, 所以我们认为 1982 年为美国消费者价格指数的变点更合适.

4 结论

CUSUM 方法作为变点检测和估计最常用的非参数方法之一, 其调节参数取值的选择一直备受关注. 本文分别对独立样本和相依样本序列, 通过蒙特卡罗模拟方法探究了调节参数取值对 CUSUM 型变点估计量的估计效果的影响, 并提出了一种基于数据驱动的调节参数选取法, 证明了该方法的稳健性, 这为带有调节参数的 CUSUM 方法在变点问题统计推断中的应用提供了理论支撑. 实证分析显示, 基于数据驱动方法给出的调节参数取值在实际数据分析中具有明显的应用价值. 但是模拟结果同时发现, 基于数据驱动方法给出的调节参数取值并不是最优的, 如何基于数据驱动方法给出一个最优的调节参数取值, 这将是后续需要解决的问题.

参考文献 (References)

[1] PAGE E S. Continuous inspection schemes [J]. Biometrika, 1954, 41: 100-115.  
 [2] GIRAITIS L, LEIPUS R, SURGAILIS D. The change-point problem for dependent observations [J]. Statistical Planning and Inference, 1996, 53 (3): 297-310.  
 [3] CSORGO M, HORWLTH L. Limit Theorems in Change-Point Analysis [M]. Chichester, UK: John Wiley and Sons Ltd, 1997.  
 [4] HORVATH L, KOKOSZKA P. The effect of long-range dependence on change-point estimators [J]. Statistical Planning and Inference, 1997, 64 (1):



- 57-81.
- [ 5 ] KOKOSZKA P, LEIPUS R. Testing for parameter changes in ARCH models [ J ]. Lithuanian Mathematical Journal, 1999, 39(2): 182-195.
- [ 6 ] HARIZ S B, WYLIE J J. Rates of convergence for the change-point estimator for long-range dependent sequences[J]. Statistics and Probability Letters, 2005, 73(2): 155-164.
- [ 7 ] HARIZ S B, WYLIE J J, ZHANG Q. Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences[J]. The Annals of Statistics, 2007, 35(4): 1802-1826.
- [ 8 ] NIE W L, HARIZ S B, WYLIE J J, et al. Change-point detection for long-range dependent sequences in a general setting[J]. Nonlinear Analysis, 2009, 71(12): 2398-2405.
- [ 9 ] FREMDT S. Asymptotic distribution of the delay time in Page's sequential procedure[J]. Statistical Planning and Inference, 2013, 145: 74-91.
- [10] CHEN Z H, HU Y J. Cumulative sum estimator for change-point in panel data [ J ]. Statistical Papers, 2017, 58: 707-728.
- [11] QIN R B, MA J J. An efficient algorithm to estimate the change in variance[J]. Economics Letters, 2018, 168:15-17.
- [12] XU M, WU Y, JIN B. Detection of a change-point in variance by a weighted sum of powers of variances test [J]. Journal of Applied Statistics, 2019, 46 (4): 664-679.
- [13] TAN C C, SHI X P, WU Y H. On nonparametric change point estimator based on empirical characteristic functions[J]. Science China Mathematics, 2016, 59 (12): 2463-2484.
- [14] Federal Reserve Bank of St. Louis. Consumer price index for all urban consumers; All items in U. S. city average (CPIAUCSL) [ EB/OL ]. [ 2020-04-01 ]. <https://fred.stlouisfed.org/series/CPIAUCSL>.
- [15] 王伟. 基于 M 估计的线性回归模型均值变点检测[D]. 南京:东南大学, 2011.