

# 基于S-BCUSUM的回归模型系数变点在线检测

王继梅

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2023年2月27日; 录用日期: 2023年4月10日; 发布日期: 2023年4月18日

## 摘要

随着信息技术的发展, 近年来许多领域对在线变点检测方法的需求急剧增加, 在线变点检测使人们在观测到越来越多数据时能够连续检测模型是否仍然不变, 这在实际数据监控中有重要的应用价值。本文提出基于S-BCUSUM (Sequential Backward Cumulative Sum) 监测统计量的封闭式在线变点检测方法用于研究回归模型系数变点的在线监测问题, 在一定条件下, 给出监测统计量在原假设和备择假设下的渐近性质, 并通过渐近分布得到拒绝阈值。模拟结果表明, 所提方法能较好地控制检验水平, 且有较高的功效和较短的平均运行长度。实例分析表明, 所提方法能有效监测北京市PM<sub>2.5</sub>浓度与空气污染物和气象因素之间的动态关系, 这为政府部门的空气质量治理工作提供一定的理论支持, 具有一定的应用前景。

## 关键词

回归模型, 变点, 在线检测, S-BCUSUM

# Online Detection of Regression Model Coefficient Change Points Based on S-BCUSUM

Jimei Wang

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Feb. 27<sup>th</sup>, 2023; accepted: Apr. 10<sup>th</sup>, 2023; published: Apr. 18<sup>th</sup>, 2023

## Abstract

With the development of information technology, the demand for online change point detection methods has increased dramatically in many fields in recent years. Online change point detection enables people to sequentially the hypothesis that the model still holds as more and more data are observed. It's widely used in data monitoring in practice. In this paper, a closed-end online change

point detection method based on the sequential backward cumulative sum statistic is proposed to investigate the online detection problem for the parameter change in linear regression model. Under certain conditions, the asymptotic properties of the monitoring statistic under the null hypothesis and the alternative hypothesis are proved. In addition, rejection threshold is obtained from the asymptotic null distribution. Simulation results show that the proposed method is able to control the test level well, and has higher power and shorter average run length. The example analysis shows that the proposed method can effectively monitor the dynamic relationship between PM<sub>2.5</sub> concentration, air pollutants and meteorological factor in Beijing, which provides a certain theoretical support for the government's air quality control work and has certain application prospects.

## Keywords

Regression Model, Change Point, Online Detection, S-BCUSUM

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

变点指的是模型中某个或某些量起突然变化的点,这种变化通常反映出该模型质的变化。变点问题的研究对象大多是时间序列数据,即按照时间顺序排列而成的数据序列。变点问题最初来源于 Page [1] 在 *Biometrika* 上发表的一篇关于连续抽样检验的文章,文中指出人们通过抽检生产线上的产品以检测产品质量是否出现较大波动,若产品质量变化超出其质量控制范围,则要判断产品何时出现质变,及时预警并修正,以免出现更多的次品。此后,许多统计学者和应用工作者对这一问题进行了深入的研究,取得了一系列显著的成果。尤其是近 30 年以来,变点问题的研究在理论和应用上都取得了重大的进展,具体相关内容可参考文献综述和专著[2] [3] [4] [5] [6]。

回归模型是应用最广泛的一类模型,经典的回归模型假定回归系数在整个样本空间保持不变,但是实际数据的复杂性使得经典回归模型已经不能很好地反映响应变量与协变量之间的关系,因此含变点的回归模型引起了众多学者的重视,当数据为一维或二维的情形时,该模型将退化为均值模型或线性模型。根据抽取样本的类型,变点检测方法可分为离线变点检测(也叫回顾性检验、后验检验)和在线变点检测(也叫序贯监测、在线监测)。随着技术的发展,近年来几个领域对在线变点检测算法的需求急剧增加,如信息技术与网络安全、监测不良健康事件、工业过程监控等。因此,如何快速监测到系统在任意时刻出现的变点并发出预警,避免决策失误、减少损失、降低风险、提高效率成为人们日益关注的话题。

在线变点检测问题最早是由 Chu 等[7]提出的,其主要思想是对一组连续观测的新样本数据进行监测,当监测统计量的值超过给定的极限临界值时停止且记当前时刻为变点,并对现有模型进行调整,否则一直监测下去。在线变点检测算法通常应用监测统计量与边界函数的关系分析模型中存在的变点,根据需要监测的总样本量,该算法可分为封闭式和开放式两种研究类型,前者需要提前设定监测样本量,后者对需要监测的样本量无限制,则可以一直监测到发现变点为止。近年来,研究学者在在线变点检测方面取得了一些成果。Horváth 等[8]在 Chu 等人方法的基础上基于最小二乘估计残差定义了一种 CUSUM (Cumulative Sum)监测函数,用于研究具有独立同分布误差的线性回归模型系数变点在线检测问题,但是该方法在监测距离起始时刻较远的变点时延迟时间较长。为了解决此问题,一些改进的方法相继被提出,如 Horváth 等[9]和 Aue 等[10]提出滑动和(Moving Sum, MOSUM)统计量在线检测位置模型变点;Chen 等

[11]引进一个窗宽参数,将其它时刻出现的变点“移动”到开始时刻,提出一种改进的滑动和统计量(Modified Moving Sum, mMOSUM)来研究线性回归模型结构变点的在线检测问题,但是模拟结果显示该方法功效较低;Fremdt等[12]提出在线检测回归模型系数变点的累积和(Page-CUSUM)统计量,在一定程度上提高了监测的稳定性。上述三种监测统计量的主要区别在于部分和中所含观测样本量不同,其中MOSUM使用的是恒定窗宽,使得部分和中包含的样本量保持不变;mMOSUM通过引入窗宽系数,采用滑动方法改变监测起始时刻,从而使部分和中包含的样本量随着观测样本量的变化而变点;Page-CUSUM是选择所有可能样本中使部分和达到最大的样本量。Kirch等[13]使用估计方程作为统一框架,整合上述三种改进方法,对位置偏移、线性回归和自回归进行开放式在线变点检测。Dette等[14]基于似然比检验提出一种监测 $d$ 维时间序列一般参数变点的封闭式在线检测方法,这些参数可以由经验分布函数的近似线性泛函来估计,证明了统计量的渐近性质,由于引入了自归一化,不需要估计长期方差,但是该方法计算成本昂贵。Kojadinovic等[15]基于经验分布函数提出一种非参数的封闭式变点监测方法,通过将检测器与重采样估计的阈值函数进行比较,使虚警概率在监测周期内保持近似恒定,与Dette等人所提方法的主要区别在于该方法涉及一个加权,允许对最近的观测给予或多或少的重要性。Song[16]提出一种对异常值具有稳健性的开放式在线检测方法来研究时间序列的参数变点问题,并使用DP(Density Power)散度推导了该方法中的检测器和停时。Gösmann等[17]对多元时间序列的参数变点提出了一种新的开放式在线监测方法,与以往的文献中提出比较训练样本计算出的估计量与剩余样本计算出的估计量不同的是,该方法建议在训练样本之后的每个时间点分割样本,然后连续比较来自所有分割点之前和之后的样本估计值,计算差异的最大范数。在开放式情形下,证明了备择假设下的参数变点的一致性。Holmes等[18]提出基于retrospective CUSUM的开放式非参数序列变点监测方法,在为选择的检测器提出合适的阈值函数后,研究了在监测均值变化的特殊情况下,即在平稳性的零假设和相关的备择假设下,方法的渐近有效性。只要变点不靠近监测的起始时刻,该方法的性能会优于Gösmann等人的方法。然而,就目前所知,对于模型中参数变化的在线变点检测方法的相关研究仍然较少,同时在线变点检测问题将是未来的关注热点,算法中存在的功效较低以及平均运行长度较长等问题亟待进一步解决,从而为相关领域的工作者及时发现变点做出相应决策提供方法保障,本文考虑封闭式情形下的在线变点检测方法。

本文余下部分的结构如下:在第2节,首先,介绍模型和假设。进一步,构造S-BCUSUM统计量,并给出原假设和备择假设下封闭式变点监测统计量的渐近性质以及估计量的相合性。在第3节,在一系列数值模拟实验下评估监测统计量的有限样本性质。在第4节,通过一组北京市空气中PM<sub>2.5</sub>浓度以及相关影响因素的数据建模来验证所提方法的可行性。

## 2. 变点监测模型

### 2.1. 模型和假设

基于参数设置,考虑如下的线性回归模型:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_t + \varepsilon_t, \quad 1 \leq t < \infty \quad (1)$$

其中 $y_t$ 是响应变量, $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tk})'$ 为 $k \times 1$ 维的协变量, $\boldsymbol{\beta}_t$ 是依赖于时间指标 $t$ 的 $k \times 1$ 维回归系数向量, $\varepsilon_t$ 为不可观测的随机误差项。若式(1)中包含一个常数项,则对于任意的 $t$ ,有 $x_{t1} = 1$ 。

假定前 $n$ 个样本无污染,等价于在观测前 $n$ 个样本时,模型中没有产生变点,即

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}, \quad 1 \leq t \leq n \quad (2)$$

其中 $\boldsymbol{\beta}$ 是一个固定的 $k \times 1$ 维向量,称这 $n$ 个数据为历史数据,从第 $n+1$ 个新观测的样本开始在线检测回归系数 $\boldsymbol{\beta}_t$ 是否发生改变,直到在线检测系统发出警报(产生变点),或者到第 $[mn]$ 个样本结束监测。其中,当 $1 < m < \infty$ 时,定义为封闭式在线检测;当 $m = \infty$ 时,定义为开放式在线检测。

考虑如下的假设检验问题:

$$H_0: \beta_t = \beta, \quad t = n+1, n+2, \dots \quad (3)$$

$$H_1: \beta_t = \beta + n^{-1/2} \mathbf{g}(t/n), \quad t = n+1, n+2, \dots \quad (4)$$

其中  $\mathbf{g}: \mathbf{R} \rightarrow \mathbf{R}^k$  为有界的分段常值函数。若备择假设  $H_1$  成立, 说明参数向量  $\beta_t$  发生了变化, 需对模型中存在的未知变点进行估计。

为了研究检验统计量的渐近性质, 做以下假设。

### 假设 1

(A1) 对于所有的  $t$ , 误差  $\varepsilon_t$  满足  $E(\varepsilon_t) = 0$ ,  $E(\varepsilon_t^2) = \sigma^2 > 0$  以及  $E(|\varepsilon_t|^8) < \infty$ 。

(A2) 对于所有的  $t$ ,  $E(\|\mathbf{x}_t\|^8) < \infty$  成立, 且对任意的  $t > k$ , 样本协方差矩阵  $\hat{\mathbf{C}}_t = t^{-1} \sum_{j=1}^t \mathbf{x}_j \mathbf{x}_j'$  一致正定,

并有  $\hat{\mathbf{C}}_t \xrightarrow{P} \mathbf{C}$ 。其中  $\|\cdot\|$  表示  $\infty$ -范数,  $\mathbf{C}$  是  $k \times k$  的非奇异常数矩阵,  $\xrightarrow{P}$  表示依概率收敛。

(A3) 存在一个正定矩阵  $\mathbf{\Omega}$ , 使得  $n^{-1} \left( \sum_{i=1}^n \mathbf{x}_i \varepsilon_i \right) \left( \sum_{i=1}^n \mathbf{x}_i \varepsilon_i \right)' \xrightarrow{P} \mathbf{\Omega}$ 。

假设 2  $\{\varepsilon_t, \mathcal{F}_t, t \geq 1\}$  是鞅差序列, 其中  $\mathcal{F}_t$  是由  $\{(\mathbf{x}'_{i+1}, \varepsilon_i)', i \leq t\}$  产生的  $\sigma$  领域。

假设 3 边界函数的形式为  $b(r) = \lambda_\alpha d(r)$ , 且  $d(r)$  连续, 存在  $\varepsilon > 0$ , 使得对任意的  $r \geq 0$  有  $d(r) > \varepsilon$ , 其中  $\lambda_\alpha$  为临界值,  $\alpha$  为给定的显著性水平。

## 2.2. 监测统计量

针对式(3)和式(4)中的假设检验问题, 传统的一元 CUSUM 检验考虑根型边界函数

$b_{rad}(r) = \sqrt{r(\log r - \log(\alpha^2))}$  ( $r > 1$ ) 进行在线变点检测, 根据文献可知, 一元 CUSUM 检验的功效受协变量均值与参数变化方向的夹角影响, 因此, 可将一元 CUSUM 在线检测方法推广到多元的情况, 于是多元 CUSUM 监测器定义如下:

$$\mathbf{Q}_{i,n}^{oli} = \mathbf{Q}_n(t/n) - \mathbf{Q}_n(1) = \frac{1}{\tilde{\sigma}_n \sqrt{n}} \hat{\mathbf{C}}_n^{-1/2} \sum_{j=n+1}^t \mathbf{x}_j \tilde{\varepsilon}_j, \quad t > n \quad (5)$$

其中, 检验统计量  $\mathbf{Q}_n(r)$  定义如下:

$$\mathbf{Q}_n(r) = \frac{1}{\tilde{\sigma}_n \sqrt{n}} \hat{\mathbf{C}}_n^{-1/2} \sum_{i=1}^{[rn]} \mathbf{x}_i \tilde{\varepsilon}_i \quad (6)$$

式中  $\tilde{\sigma}_n = (n-k)^{-1} \sum_{i=k+1}^n (\tilde{\varepsilon}_i - \bar{\tilde{\varepsilon}})^2$ , 这里  $\bar{\tilde{\varepsilon}} = (n-k)^{-1} \sum_{i=k+1}^n \tilde{\varepsilon}_i$ ;  $\tilde{\varepsilon}_i$  为递归残差, 即标准化的超前一步预测误差, 它的定义如下:

$$\tilde{\varepsilon}_i = \frac{y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}_{i-1}}{\sqrt{1 + \mathbf{x}'_i (\mathbf{X}'_{i-1} \mathbf{X}_{i-1})^{-1} \mathbf{x}_i}}, \quad i \geq k+1 \quad (7)$$

式中, 当  $i = 1, \dots, k$  时,  $\tilde{\varepsilon}_i = 0$ ,  $\mathbf{X}_{i-1} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{i-1}]'$ ,  $\mathbf{Y}_{i-1} = [y_1, y_2, \dots, y_{i-1}]'$ ,  $\hat{\boldsymbol{\beta}}_{i-1} = (\mathbf{X}'_{i-1} \mathbf{X}_{i-1})^{-1} \mathbf{X}'_{i-1} \mathbf{Y}_{i-1}$ 。令  $0 < m < \infty$ , 对任意的  $r \in [0, m]$ ,  $\mathbf{Q}_n(r)$  是  $k$  折积空间  $D([0, m])^k = D([0, m]) \times \dots \times D([0, m])$  中的元素。在假设 1 的条件下, 多元序列  $\mathbf{x}_t \varepsilon_t$  满足多元泛函中心极限定理[19], 类似地, 也适用于基于递归残差的多元 CUSUM 过程。

**定理 1** 假设 1 和假设 2 成立。

(A1) 若原假设成立, 则对于任意的  $0 < m < \infty$ , 当  $n \rightarrow \infty$  时, 有

$$Q_n(r) \Rightarrow B^{(k)}(r), \quad r \in [0, m] \tag{8}$$

其中  $B^{(k)}(r)$  是一个  $k$  维的标准布朗运动。

(A2) 若备择假设成立, 则对于任意的  $0 < m < \infty$ , 当  $n \rightarrow \infty$  时, 有

$$Q_n(r) \Rightarrow B^{(k)}(r) + C^{1/2}h(r), \quad r \in [0, m] \tag{9}$$

其中,  $h(r) = \frac{1}{\sigma} \int_0^r g(z) dz - \frac{1}{\sigma} \int_0^r \int_0^z \frac{1}{z} g(v) dv dz$ 。若  $\|Q_{t,n}^{onli}\|$  在  $t > n$  中至少有一次大于边界函数

$b_t = \lambda_\alpha d((t-n)/n)$ , 拒绝原假设。对于  $1 < m < \infty$ , 相应的最大统计量为

$$Q_{n,m} = \max_{t=n+1, \dots, [mn]} \frac{\|Q_{t,n}^{onli}\|}{d((t-n)/n)} \tag{10}$$

根据定理 1、假设 3 和连续映射定理可得:

(A1) 若原假设成立, 则当  $n \rightarrow \infty$  时, 有

$$Q_{n,m} \xrightarrow{d} \sup_{0 < r < m-1} \frac{\|B^{(k)}(r)\|}{d(r)} \tag{11}$$

(A2) 若备择假设成立, 则当  $n \rightarrow \infty$  时, 有

$$Q_{n,m} \xrightarrow{d} \sup_{0 < r < m-1} \frac{\|B^{(k)}(r) + C^{1/2}(h(r+1) - h(1))\|}{d(r)} \tag{12}$$

为了提高变点监测性能, 引入 S-BCUSUM 统计量, 从而得到更优的在线变点检测方法, 令

$$BQ_n(t) = \max_{s=1, \dots, t} \frac{\|Q_n(t/n) - Q_n((s-1)/n)\|}{d((t-s+1)/n)} \tag{13}$$

S-BCUSUM 的基本思想是在每个时刻  $t=1, \dots, n$  处有序地计算该统计量, 得到序列  $BQ_n(1), \dots, BQ_n(n)$ , S-BCUSUM 统计量是这些统计量序列中的最大值。该序列尤为重要的特征是它对于  $t$  时刻的  $\sigma$  代数流可测, 使得  $BQ_n(t)$  适用于进行在线变点检测。因此, S-BCUSUM 监测器定义如下:

$$SBQ_{s,t,n} = \frac{1}{\hat{\sigma}_n \sqrt{n}} \hat{C}_n^{-1/2} \sum_{j=s}^t x_j \tilde{\varepsilon}_j, \quad 1 \leq s \leq t, t \in \mathbf{N} \tag{14}$$

在满足  $s \leq t$  的条件下, 根据  $s$  和  $t$  的取值不同, 式(14)会形成一个三角数组结构。

**假设 4** 二维边界函数的形式为  $b(r,s) = \lambda_\alpha d(r,s)$ , 且  $d(r,s)$  连续, 存在  $\varepsilon > 0$ , 使得对任意的  $0 \leq s \leq r$  有  $d(r,s) > \varepsilon$ 。

当  $t > n$  时, 式(14)对于  $t$  时刻的  $\sigma$  代数流可测, 进而可用于在线变点检测, 其封闭式变点监测统计量 S-BCUSUM 定义如下:

$$SBQ_{n,m} = \max_{t=n+1, \dots, [mn]} BQ_n(t) = \max_{t=n+1, \dots, [mn]} \max_{s=n+1, \dots, t} \frac{\|SBQ_{s,t,n}\|}{d(t/n, (s-1)/n)} \tag{15}$$

$BQ_n(t)$  统计量具备序贯特性, 将所有  $BQ_n(t)$  堆叠起来产生一个新的最大值, 同时相应的极限分布中会有 2 个上确界, S-BCUSUM 通过多种方式采用递归残差, 使得  $BQ_n(t)$  在不同  $t$  的取值下所构成的集合与 CUSUM 检验相比包含了更多元素。Dette 等[14]已经提出类似的想法, 不过作者是基于似然比检验进行变点检测, 其检测器由似然比统计量的三角形数组结构的最大值给出, 这也导致了双最大统计量。

如果对于某个  $s$  和  $t$  ( $n < s \leq t$ ), 有  $\|\mathbf{SBQ}_{s,t,n}\|$  大于  $b_{s,t} = \lambda_\alpha d(t/n, (s-1)/n)$  或者式(15)大于  $\lambda_\alpha$ , 则拒绝原假设。本文二维边界函数中的  $d(r, s)$  采用  $1+2(r-s)$ 。

根据定理 1、假设 4 和连续映射定理可得:

(A1) 若原假设成立, 则当  $n \rightarrow \infty$  时, 有

$$SBQ_{n,m} \xrightarrow{d} \sup_{0 < r < m-1} \sup_{10 < s < r} \frac{\|\mathbf{B}^{(k)}(r) - \mathbf{B}^{(k)}(s)\|}{d(r, s)}, \quad 1 < m < \infty \quad (16)$$

(A2) 若备择假设成立, 则当  $n \rightarrow \infty$  时, 有

$$SBQ_{n,m} \xrightarrow{d} \sup_{0 < r < m-1} \sup_{10 < s < r} \frac{\|\mathbf{B}^{(k)}(r) - \mathbf{B}^{(k)}(s) + C^{1/2}(\mathbf{h}(r+1) - \mathbf{h}(s+1))\|}{d(r, s)} \quad (17)$$

为了说明 S-BCUSUM 检验的优势, 研究 CUSUM 和 S-BCUSUM 检验在备择假设下的渐近功效性质, 不失一般性, 考虑  $\beta_t = \beta + n^{-1/2} \mathbf{g}(t/n)$  的单变点模型, 式中  $\mathbf{g}(r) = cI_{\{r \geq \tau^*\}}$ ,  $c \in \mathbf{R}^k$  以及  $\tau^*$  表示变点位置,  $I_{\{r \geq \tau^*\}}$  是示性函数, 如果  $r \geq \tau^*$ , 那么  $I_{\{r \geq \tau^*\}} = 1$ , 否则  $I_{\{r < \tau^*\}} = 0$ 。由  $\mathbf{h}(r)$  的表达式可知

$$\mathbf{h}(r) = c\sigma^{-1} \left( \int_{\tau^*}^r dz - \int_0^r \int_{\tau^*}^z \frac{1}{z} dv dz \right) = c\sigma^{-1} \tau^* (\log r - \log \tau^*) I_{\{r \geq \tau^*\}} \quad (18)$$

图 1 展示了 CUSUM 和 S-BCUSUM 检验在  $m=2$  的渐近功效曲线。实验循环 100,000 次, 基于 1000 个等距网格点近似极限分布中的布朗运动, 采用调整检验水平为 5% 的临界值。图 1(a)~(f) 的功效结果表明, 除了  $\tau^* = 1.1$  外, 当变点位置位于其他位置时, S-BCUSUM 检验相比 CUSUM 检验有更好的变点检测性能, 随着变点位置越靠后, 其优势显著提升。从图 1(e)中可看出, 对于固定的  $c/\sigma$ , 整体来看, 变点位置的变化对于 S-BCUSUM 检验的影响较弱一些。

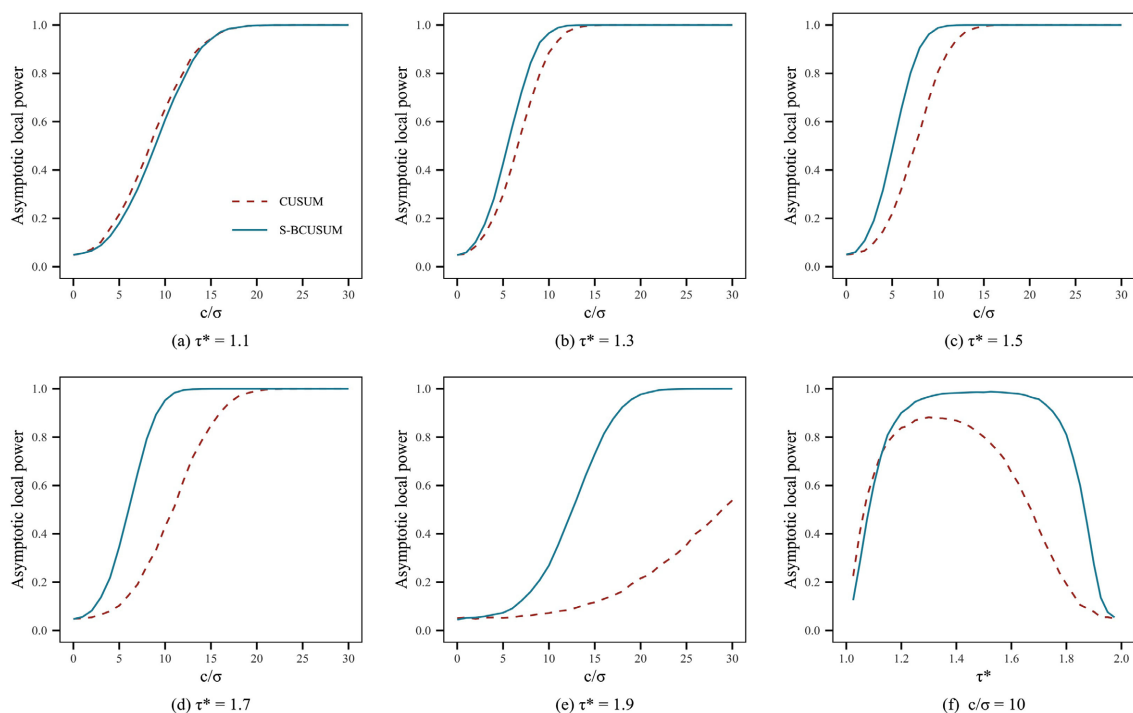


Figure 1. Asymptotic power curves of CUSUM and S-BCUSUM

图 1. CUSUM 和 S-BCUSUM 的渐近功效曲线

然而，对于变点监测统计量性能度量，比功效更重要的评价指标是平均运行长度，因为在监测时间足够长的情况下，任何一个固定的非平凡备择假设成立时，都能检测到变点。有学者基于 OLS (Ordinary Least Squares) 残差的 CUSUM 检验研究在线变点检测问题时，推导了停时的渐近分布[20] [21]，在前人研究的基础上，由本节定义的两个监测统计量  $Q_{n,m}$  和  $SBQ_{n,m}$  及其相应的显著性水平为  $\alpha$  的临界值  $\lambda_{\alpha,m,1}$  和  $\lambda_{\alpha,m,2}$  定义停时：

$$\tau_{n,m,1} = \min \left\{ n+1 \leq t \leq [mn] \left| \frac{\|Q_{t,n}^{oli}\|}{d((t-n)/n)\lambda_{\alpha,m,1}} \geq 1 \right. \right\} \tag{19}$$

$$\tau_{n,m,2} = \min \left\{ n+1 \leq t \leq [mn] \left| \max_{n < s \leq t} \frac{\|SBQ_{s,t,n}\|}{d(t/n, (s-1)/n)\lambda_{\alpha,m,2}} \geq 1 \right. \right\} \tag{20}$$

在前面考虑  $\beta_t = \beta + n^{-1/2} cI_{\{r \geq \tau^*\}}$  的模型下，当  $n \rightarrow \infty$  时，相对停时满足：

$$\frac{\tau_{n,m,1}}{n} \xrightarrow{d} \inf \left\{ 1 < r < m \left| \frac{\|B(r) - B(1) + h(r) - h(1)\|}{(1+2(r-1))\lambda_{\alpha,m,1}} \geq 1 \right. \right\} \triangleq \tau_{n,Q} \tag{21}$$

$$\frac{\tau_{n,m,2}}{n} \xrightarrow{d} \inf \left\{ 1 < r < m \left| \sup_{1 < s < r} \frac{\|B(r) - B(s) + h(r) - h(s)\|}{(1+2(r-s))\lambda_{\alpha,m,2}} \geq 1 \right. \right\} \triangleq \tau_{n,SBQ} \tag{22}$$

由上式可获得渐近平均运行长度，即  $E(\tau_n | \tau^* \leq \tau_n \leq m) - \tau^*$ ，其中  $\tau_n \in \{\tau_{n,Q}, \tau_{n,SBQ}\}$ 。

图 2 展示了三种方法在  $m = 4$  以及变点取不同值时的渐近平均运行长度，其中 CSW 方法与 CUSUM 方法的区别在于，它是采用根型边界函数的一元 CUSUM 检验。从图 2(a)和图 2(b)可以看出，当真实变点  $\tau^*$  分别固定为 1.5 和 3 时，随着结构突变  $c/\sigma$  的增加，所有方法的平均运行长度都在减小，其中 S-BCUSUM 都表现较优；从图 2(c)可看出，当固定  $c/\sigma$  时，变点位置发生改变，S-BCUSUM 的平均运行长度都是较低的且基本上恒定不变，说明该方法对变点位置的依赖性较弱，反之，除靠近监测范围的两端点外，变点位置越靠后，CUSUM 和 CSW 方法的平均运行长度均呈现上升趋势。

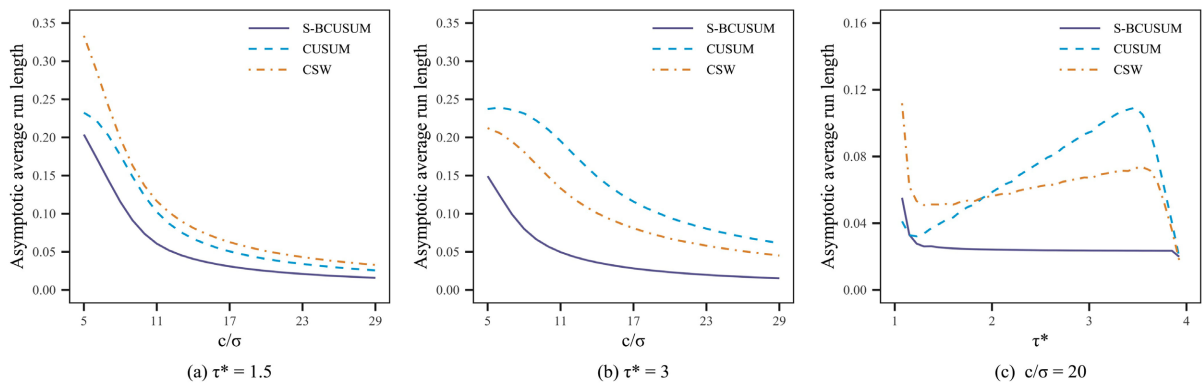


Figure 2. Asymptotic average run length  
 图 2. 渐近平均运行长度

S-BCUSUM 方法监测到模型存在变点后，需估计变点的具体位置，不失一般性，仍考虑  $\beta_t = \beta + \delta I_{\{r \geq \tau^*\}}$  的含变点模型，其中  $\delta \neq 0$ 。假设  $n_d$  为监测统计量首次超过边界函数时所对应的时间，由于真实变点往往接近于停时，使得采用传统的极大似然估计量其估计误差很大[22]，考虑

$$\hat{\tau}_{onl} = \arg \max_{t=n+1, \dots, n_d} \frac{\|BQ_{t, n_d}\|}{n\sqrt{n_d - t + 1}} \quad (23)$$

**定理 2** 令  $\beta_t = \beta + \delta I_{\{t/n \geq \tau^*\}}$ , 其中  $\delta \neq 0$ , 且满足假设 1, 则对于  $\tau^* \in (1, n_d/n]$ , 当  $n \rightarrow \infty$  时, 有  $\hat{\tau}_{onl} \xrightarrow{P} \tau^*$ 。

定理 2 根据  $h(r)$  的表达式、定理 1 及连续映射定理可得, 该定理表明估计量  $\hat{\tau}_{onl}$  是  $\tau^*$  的相合估计。

### 3. 模拟研究

本节通过数值模拟讨论基于 S-BCUSUM 的在线检测方法的有限样本性质, 主要分析不同变点位置、历史样本量以及监测样本量对监测结果的影响。数据由三种情形下的模型产生:

$$y_t = u_t + \varepsilon_t, \quad t = 1, \dots, n \quad (\text{情形 1})$$

$$y_t = 1 + u_t z_t + \varepsilon_t, \quad t = 1, \dots, n \quad (\text{情形 2})$$

$$y_t = u_t + 0.5y_{t-1} + \varepsilon_t, \quad t = 1, \dots, n \quad (\text{情形 3})$$

其中  $u_t = 0.9I_{\{t/n \geq \tau^*\}}$ ,  $z_t = (1 + 0.5L)e_t$ , 这里  $L$  是滞后算子,  $e_t$  与  $\varepsilon_t$  独立且服从于标准正态分布。对于情形 1 和情形 2 考虑整体结构突变检验, 对于情形 3 考虑部分结构突变检验, 其中  $H = (1, 0)'$ 。

通过 100,000 次重复实验得到  $SBQ_{n,m}$  在原假设下的极限分布的  $1-\alpha$  分位数作为临界值, 其中标准布朗运动利用 50,000 个独立正态随机变量来逼近, 根据  $\alpha$ 、 $k$  及  $m$  的不同取值, 得到的临界值如表 1 所示。

**Table 1.** Critical values for S-BCUSUM

**表 1.** S-BCUSUM 的临界值

$k$	$\alpha$	$m$							
		1.3	1.5	1.7	1.9	2	4	6	10
1	0.10	0.878	0.988	1.052	1.095	1.112	1.261	1.317	1.393
	0.05	0.962	1.067	1.134	1.181	1.198	1.337	1.392	1.464
	0.01	1.118	1.236	1.306	1.349	1.359	1.485	1.529	1.589
	0.10	0.956	1.069	1.136	1.175	1.197	1.344	1.398	1.464
2	0.05	1.036	1.146	1.214	1.257	1.279	1.409	1.472	1.525
	0.01	1.202	1.316	1.383	1.430	1.448	1.565	1.604	1.668
	0.10	1.003	1.116	1.181	1.227	1.245	1.381	1.445	1.502
3	0.05	1.068	1.200	1.262	1.309	1.322	1.456	1.516	1.563
	0.01	1.249	1.367	1.430	1.473	1.488	1.604	1.638	1.694
	0.10	1.034	1.145	1.215	1.261	1.278	1.411	1.471	1.529
4	0.05	1.098	1.231	1.289	1.331	1.348	1.484	1.540	1.590
	0.01	1.276	1.394	1.448	1.488	1.515	1.621	1.672	1.720
	0.10	1.053	1.176	1.242	1.283	1.302	1.433	1.491	1.547
5	0.05	1.120	1.253	1.316	1.357	1.374	1.511	1.559	1.605
	0.01	1.288	1.407	1.490	1.530	1.547	1.649	1.683	1.739
	0.10	1.070	1.202	1.262	1.306	1.320	1.452	1.508	1.566



Continued

6	0.05	1.142	1.273	1.339	1.379	1.395	1.527	1.573	1.618
	0.01	1.303	1.437	1.524	1.557	1.573	1.661	1.695	1.753
	0.10	1.089	1.218	1.276	1.318	1.336	1.473	1.527	1.576
7	0.05	1.156	1.285	1.358	1.396	1.412	1.545	1.585	1.633
	0.01	1.305	1.446	1.535	1.569	1.578	1.677	1.720	1.767
	0.10	1.102	1.230	1.291	1.332	1.350	1.484	1.538	1.588
8	0.05	1.169	1.304	1.371	1.410	1.424	1.553	1.596	1.645
	0.01	1.316	1.454	1.540	1.577	1.581	1.687	1.729	1.778
	0.10	1.113	1.244	1.306	1.345	1.363	1.495	1.544	1.595
9	0.05	1.185	1.317	1.385	1.422	1.434	1.560	1.602	1.653
	0.01	1.327	1.461	1.544	1.577	1.590	1.691	1.739	1.796

假设历史样本量  $n = 50, 200, 500$ ，显著性水平取  $\alpha = 0.05$ ，对于每个历史样本量  $n$ ，从  $n + 1$  开始监测，直至  $[mn]$  停止，总样本量为  $2n$  和  $4n$ ，表 2 列出了原假设下两个变点监测统计量在不同情形下的检验水平。从表中可看出大部分检验水平都接近显著性水平 0.05，当历史样本量  $n$  不变时，随着  $m$  的增加检验水平有少许增大，意味着误报率在增大，当其他变量保持相同，历史样本量  $n$  逐渐增加时，检验水平基本呈现减少趋势，说明了历史样本量越多对监测结果的影响越小，总体来看，S-BCUSUM 能更好地控制检验水平，以下所有模拟实验重复次数均设为 5000 次。

Table 2. Test level  
表 2. 检验水平

情形	统计量	$n = 50$		$n = 200$		$n = 500$	
		$m = 2$	$m = 4$	$m = 2$	$m = 4$	$m = 2$	$m = 4$
情形 1	CUSUM	0.053	0.057	0.047	0.058	0.044	0.050
	S-BCUSUM	0.044	0.056	0.044	0.055	0.042	0.050
情形 2	CUSUM	0.094	0.102	0.059	0.059	0.052	0.054
	S-BCUSUM	0.078	0.092	0.056	0.060	0.049	0.053
情形 3	CUSUM	0.065	0.071	0.049	0.053	0.048	0.051
	S-BCUSUM	0.053	0.075	0.045	0.056	0.045	0.053

考虑不同变点位置、历史样本量以及监测样本量对变点监测效果的影响，表 3 给出了备择假设下两个变点监测统计量在不同情形下的平均运行长度。从表中可看出，当其他参数不变，随着变点时刻的推迟 CUSUM 方法的平均运行长度明显增加，只有变点位置离监测起始时刻较近时，它才有良好的监测效果，而 S-BCUSUM 方法的平均运行长度相对较低且波动幅度较小，说明 S-BCUSUM 受变点位置的影响较弱。当历史样本量和监测样本量都增加时，平均运行长度随之变长，S-BCUSUM 方法的平均运行长度增长速度不是特别明显。无论是哪种情形的数据，S-BCUSUM 方法的监测效果基本上都优于 CUSUM 方法，尤其是监测周期较长和数据情况较为复杂时。

**Table 3.** Average run length  
**表 3.** 平均运行长度

	情形 1			情形 2		情形 3	
	变点	CUSUM	S-BCUSUM	CUSUM	S-BCUSUM	CUSUM	S-BCUSUM
$n = 200$ $m = 2$	1.1	23.5	23.3	23.4	23.2	24.8	24.9
	1.3	31.3	22.4	30.4	22.4	34.6	23.3
	1.5	38.9	22.1	37.3	22.2	43.3	22.6
	1.7	38.2	21.9	35.6	21.8	37.9	22.1
$n = 300$ $m = 2$	1.1	27.1	26.6	26.9	26.2	28.4	27.9
	1.3	36.2	25.5	35.0	25.4	39.2	26.5
	1.5	44.8	25.3	43.3	25.2	50.0	25.8
	1.7	51.1	25.2	47.8	24.8	52.3	25.5
$n = 500$ $m = 4$	2.2	93.6	30.1	89.1	29.8	103.9	30.3
	2.6	115.5	30.3	109.7	29.9	128.5	30.0
	3.0	137.6	30.3	130.7	29.9	154.0	30.0
	3.4	158.3	30.4	150.9	29.9	175.3	29.8

#### 4. 实例分析

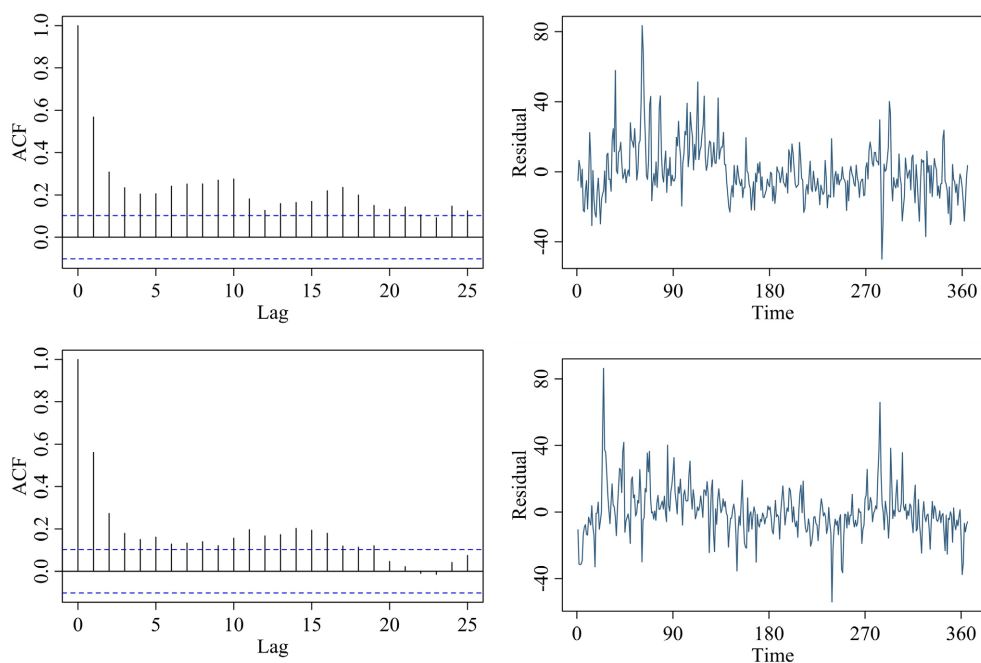
随着全球经济的快速发展,能源、交通规模显著扩大,城市化进程显著加快,人类活动对生态环境的影响日益增强,最明显的问题是空气污染。空气质量问题一直是我国的重要环境问题之一,我国从改革开放以来仅用了三十多年就完成了很多欧美国家需要百年才能完成的发展目标,但是这样高速的发展使得我国的空气质量一直不容乐观,空气质量问题会危害人们的健康,比如增加呼吸道、肺部疾病等诸多疾病的发病率,会带来雾霾等天气异常状况,严重影响人们的生活,比如阻碍人们出行,也会给我国经济造成巨大损失。本文探究北京市的 $PM_{2.5}$ 浓度(环境空气中空气动力学当量直径小于等于 $2.5\ \mu m$ 的颗粒物,也称细颗粒物)与空气污染物和气象因素之间的关系,空气中 $PM_{2.5}$ 的浓度越高,意味着空气污染就越严重。一方面, $PM_{2.5}$ 作为反映空气质量的主要因素,其质量轻,体积小,活性强,极易附带有毒、有害物质(如重金属、微生物等),并且能够长时间存在于大气之中,凭借体积小和易于随空气扩散的特性可以轻松进入人体呼吸道深处,给人体带来极大危害。自2013年北京雾霾事件以来,北京民众对 $PM_{2.5}$ 的敏感程度较高,对 $PM_{2.5}$ 影响因素的许多研究通常仅考虑其他几种污染物或者仅考虑气象因素,很少同时考虑这两方面因素的综合影响。另一方面,北京市空气质量问题一直以来都是人们普遍关注的话题,北京作为我国的经济、政治、文化中心,也是我国最早进入工业化的城市之一,其空气质量问题不仅影响着当地居民的健康和生活,还关系着我国在国际社会上的国际形象,因此,坚持推进北京空气质量持续改善尤为重要。利用本文所提出的在线变点检测方法监测 $PM_{2.5}$ 时间序列的变点,探究 $PM_{2.5}$ 浓度与空气污染物和气象因素之间的动态关系,分析导致变点产生的主要原因,揭示北京空气质量的变化特征,为日后相关研究和空气质量治理工作提供理论支持。

以2019年1月1日至2020年12月31日北京市五种空气污染物和一种气象因素为研究对象,即细颗粒物( $PM_{2.5}$ ,  $\mu g/m^3$ )、二氧化硫( $SO_2$ ,  $\mu g/m^3$ )、一氧化碳( $CO$ ,  $\mu g/m^3$ )、二氧化氮( $NO_2$ ,  $\mu g/m^3$ )、臭氧( $O_3$ ,  $\mu g/m^3$ )和风速(Wind speed, m/s),时间跨度为731天,运用变点监测模型综合分析这两方面因素对 $PM_{2.5}$ 的影响。其中, $PM_{2.5}$ 、 $SO_2$ 、 $CO$ 、 $NO_2$ 和 $O_3$ 的小时浓度数据来自中国空气质量在线监测分析平台,按采集粒度

24 小时统计得到逐日浓度数据。气象因素对空气质量存在显著影响, 但气象数据收集相对比较困难, 且有研究显示对  $PM_{2.5}$  起重要作用的是风速, 因此本文尚未考虑其他气象因素。风速数据来源于美国国家海洋和大气管理局(National Oceanic and Atmospheric Administration, NOAA)下设的国家环境信息中心(National Centers for Environmental Information, NCEI), 网址为

<https://www.ncei.noaa.gov/data/global-summary-of-the-day/archive/>, 获取的原始数据是气象观测站点的数据, 每个站点记录了逐日的气象指标, 首先, 基于气象站点的经纬度及其逐日平均风速数值, 采用反距离权重法插值得到全国范围的逐日平均风速栅格图, 再基于全国地级市的行政边界数据和风速栅格图, 统计出北京市逐日平均风速数值。

将 2019 年 1 月 1 日至 2019 年 12 月 31 日  $PM_{2.5}$  逐日数据作为响应变量,  $SO_2$ 、 $CO$ 、 $NO_2$ 、 $O_3$  和 Wind speed 逐日数据作为协变量, 建立线性回归模型。从该模型中提取残差序列, 画出相关系数和时序图(图 3 上方), 由图可知, 残差序列的自相关系数长期位于零轴的一边, 且序列高度相关, 说明随着时间的变化,  $PM_{2.5}$  与五种因素之间的关系是动态变化的, 如果使用平稳模型会忽略数据的动态结构, 可能导致统计推断的偏差, 该数据符合本文模型的特征, 因此使用本文的方法对这组数据进行在线变点检测。2020 年 1 月 1 日至 2020 年 12 月 31 日的的数据也有类似的结论(图 3 下方), 这里不再赘述。



**Figure 3.** Autocorrelation coefficients and time series plots of residuals (top: 2019; bottom: 2020)

**图 3.** 残差的自相关系数和时序图(上: 2019; 下: 2020)

通过分析响应变量  $PM_{2.5}$  与协变量之间的动态关系, 来说明本文方法的有效性。设显著性水平为 0.05, 历史样本量为 36, 考虑变点监测统计量 CUSUM 和 S-BCUSUM 分别对 2019 年和 2020 年的数据进行在线变点检测。首先, 取 2019 年 3 月 6 日至 2019 年 4 月 10 日的数据作为第一阶段监测的历史样本, 从 2019 年 4 月 11 日开始进行监测, 在 2019 年 5 月 18 日时监测统计量 S-BCUSUM 取值大于相应的临界值, 从而拒绝原假设, 即序列存在变点, 发出报警, 监测持续到 2019 年 6 月 15 日时, 监测统计量 CUSUM 拒绝原假设, 至此第一阶段监测结束。接着, 取 2019 年 6 月 16 日至 2019 年 7 月 21 日的数据作为第二

阶段监测的历史样本,从2019年7月22日开始进行监测,在2019年10月13日S-BCUSUM拒绝原假设,继续监测到2019年10月20日,CUSUM拒绝原假设,于是结束第二阶段监测。最后,通过第三阶段监测,两种方法均未拒绝原假设。类似地,对2020年的数据进行在线变点检测,结果发现S-BCUSUM分别在2020年4月27日、2020年7月19日和2020年10月9日监测到变点,CUSUM在2020年5月29日和2020年8月26日监测到变点,在第三个监测阶段该方法并未监测到变点,其中三个监测阶段的平稳段分别是2020年2月23日至2020年3月29日、2020年5月30日至2020年7月4日和2020年8月27日至2020年10月1日。上述分析的结果如图4所示,浅绿色的阴影区域表示被用作平稳段的观测值,垂直虚线表示S-BCUSUM监测到的变点位置,垂直点线表示CUSUM监测到的变点位置,由图可知,S-BCUSUM方法能够尽快且准确地监测到变点,验证了理论分析,也表明使用S-BCUSUM方法可以节省宝贵的时间。从这两年的变点检测结果来看,发现变点的位置大部分集中在每年的4~5月和10~11月,具有季节性变化特征,每年3月到9月对应北京的春夏季节,而每年10月到第二年2月对应北京的秋冬季节,说明不同季节下的协变量对 $PM_{2.5}$ 浓度的影响有所不同。针对2020年S-BCUSUM监测到的变点7月19日,查阅相关资料发现,6月中旬北京市终结56天无本地新增确诊病例,随后新发地疫情爆发,为应对疫情北京市突发公共卫生事件应急响应级别由三级调升至二级,通过集中隔离、关闭娱乐场所、禁止聚集性活动、交通出行管制、市民非必要不出京、暂停学生到校等措施,使得各空气污染源受到不同程度的控制,进而影响空气质量的波动。

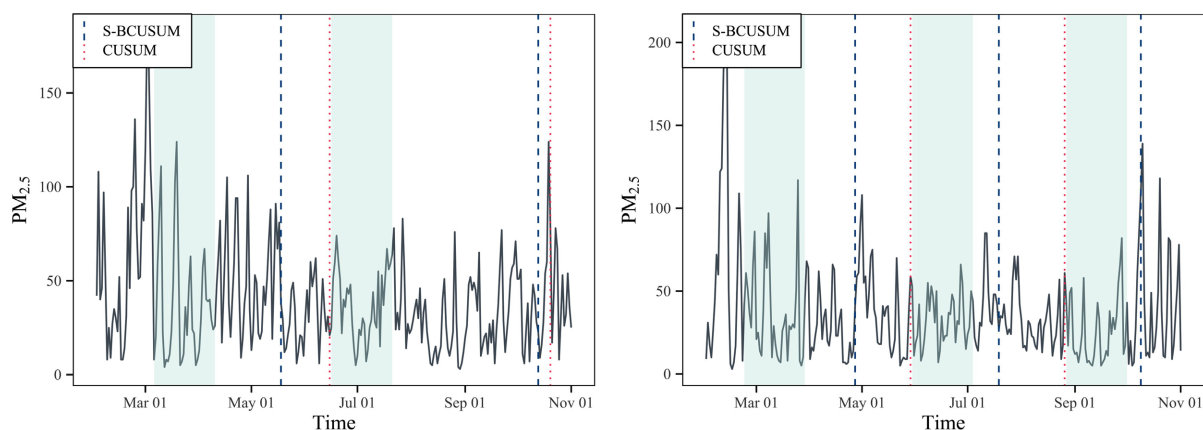


Figure 4. Online detection results of change points in air quality model (left: 2019; right: 2020)

图4. 空气质量模型变点在线检测结果(左:2019;右:2020)

根据五个监测阶段的历史数据构建的模型,分别对其残差序列进行平稳性检验,结果如图5所示,由图可知,Ljung-Box(LB)和 $\tilde{Q}$ 统计量[23]长期位于临界值下方,意味着残差不存在显著的序列相关性,因此各监测阶段的历史样本是合理的。

综上所述,实证分析结果验证了基于S-BCUSUM的在线变点检测方法的有效性和可用性,通过对新的观测数据进行变点监测,及时判断是否存在变点,发出预警,同时,导致北京市雾霾问题的主要因素 $PM_{2.5}$ 与其他观测指标之间的动态模型描述实际空气质量问题更贴切、更准确,有利于减少后续统计推断的偏差,在一定程度上反映出了空气质量数据在描述性统计分析时体现的变化特征。相比于CUSUM方法,本文所提出的方法可以较为准确地监测到变点且延迟短。监测到的变点位置能够较好解释引起空气质量波动的原因,根据其原因的总结,针对北京市空气质量治理工作有如下建议:政府部门要加大对民众环境保护工作的宣传,提高民众的环保意识;鼓励民众选乘公共交通出行,制定激励政策鼓励民众以

地热等新能源代替烧煤供暖、以新能源车代替汽油车；政府部门制定政策限制大气污染物排放，保持对植树造林工作的重视，控制城市施工面积。

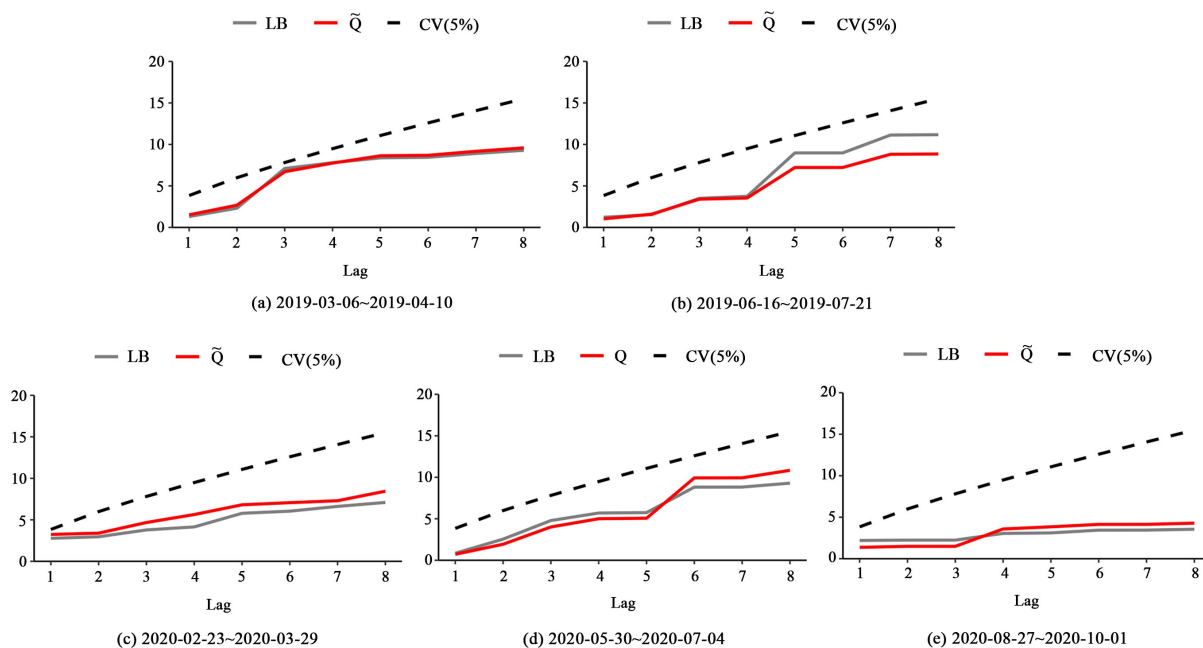


Figure 5. Residual stationarity test

图 5. 残差平稳性检验

## 5. 结论

本文主要针对线性回归模型的在线变点检测方法进行相关探究，对于目前变点检测方法存在的局限性，在前人研究的基础上，探究了基于 S-BCUSUM 统计量的变点监测和估计的统计性质及有限样本行为。首先，构造了封闭式变点监测统计量 S-BCUSUM，进一步，推导出原假设和备择假设下该统计量的极限性质及估计量的一致性。模拟研究了三种情形下基于 S-BCUSUM 的在线检测方法相比于 CUSUM 方法其监测结果受不同变点位置、历史样本量及监测样本量的影响情况，通过模拟统计量的渐近分布得到其拒绝域的临界值，与 CUSUM 相比 S-BCUSUM 在整体上能更好地控制检验水平，历史样本量和监测样本量会对方法的检验水平有相反的影响，且在渐近理论和有限样本下 S-BCUSUM 较 CUSUM 有效提高了功效并缩短了监测的平均运行长度，变点位置越靠后其优势会有显著提升。通过实例分析表明，所提方法能够有效监测 PM2.5 浓度与空气污染物和气象因素之间的关系在时间维度上的差异性，其结果反映出北京市季节性变化特征，且新冠疫情期间的关系模型存在显著差异。这为政府部门的空气质量治理工作提供一定的参考依据，具有一定的现实意义和应用价值。

## 参考文献

- [1] Page, E.S. (1954) Continuous Inspection Schemes. *Biometrika*, **41**, 100-115. <https://doi.org/10.1093/biomet/41.1-2.100>
- [2] Truong, C., Oudre, L. and Vayatis, N. (2020) Selective Review of Offline Change Point Detection Methods. *Signal Processing*, **167**, Article ID: 107299. <https://doi.org/10.1016/j.sigpro.2019.107299>
- [3] Aminikhanghahi, S. and Cook, D.J. (2017) A Survey of Methods for Time Series Change Point Detection. *Knowledge and Information Systems*, **51**, 339-367. <https://doi.org/10.1007/s10115-016-0987-z>
- [4] Shi, X., Gallagher, C., Lund, R., et al. (2022) A Comparison of Single and Multiple Changepoint Techniques for Time

Series Data. *Computational Statistics & Data Analysis*, **170**, Article ID: 107433.

<https://doi.org/10.1016/j.csda.2022.107433>

- [5] 陈希孺. 变点统计分析简介[J]. 数理统计与管理, 1991, 10(1): 55-58.
- [6] Wang, C.S.H. and Xie, Y.M. (2015) Structural Change and Monitoring Tests. In: Lee, CF. and Lee, J., Eds., *Handbook of Financial Econometrics and Statistics*, Springer, New York, 873-902. [https://doi.org/10.1007/978-1-4614-7750-1\\_31](https://doi.org/10.1007/978-1-4614-7750-1_31)
- [7] Chu, C.S.J., Stinchcombe, M. and White, H. (1996) Monitoring Structural Change. *Econometrica*, **6**, 1045-1065. <https://doi.org/10.2307/2171955>
- [8] Horváth, L., Hušková, M., Kokoszka, P. and Steinebach, J. (2004) Monitoring Changes in Linear Models. *Journal of Statistical Planning and Inference*, **126**, 225-251. <https://doi.org/10.1016/j.jspi.2003.07.014>
- [9] Horváth, L., Kühn, M. and Steinebach, J. (2008) On the Performance of the Fluctuation Test for Structural Change. *Sequential Analysis*, **27**, 126-140. <https://doi.org/10.1080/07474940801989087>
- [10] Aue, A., Horváth, L., Kühn, M., et al. (2012) On the Reaction Time of Moving Sum Detectors. *Journal of Statistical Planning and Inference*, **142**, 2271-2288. <https://doi.org/10.1016/j.jspi.2012.02.053>
- [11] Chen, Z. and Tian, Z. (2010) Modified Procedures for Change Point Monitoring in Linear Models. *Mathematics and Computers in Simulation*, **81**, 62-75. <https://doi.org/10.1016/j.matcom.2010.06.021>
- [12] Fremdt, S. (2015) Page's Sequential Procedure for Change-Point Detection in Time Series Regression. *Statistics*, **49**, 128-155. <https://doi.org/10.1080/02331888.2013.870568>
- [13] Kirch, C. and Weber, S. (2018) Modified Sequential Change Point Procedures Based on Estimating Functions. *The Electronic Journal of Statistics*, **12**, 1579-1613. <https://doi.org/10.1214/18-EJS1431>
- [14] Dette, H. and Gösmann, J. (2020) A Likelihood Ratio Approach to Sequential Change Point Detection for a General Class of Parameters. *Journal of the American Statistical Association*, **115**, 1361-1377. <https://doi.org/10.1080/01621459.2019.1630562>
- [15] Kojadinovic, I. and Verdier, G. (2021) Nonparametric Sequential Change-Point Detection for Multivariate Time Series Based on Empirical Distribution Functions. *The Electronic Journal of Statistics*, **15**, 773-829. <https://doi.org/10.1214/21-EJS1798>
- [16] Song, J. (2021) Sequential Change Point Test in the Presence of Outliers: The Density Power Divergence Based Approach. *The Electronic Journal of Statistics*, **15**, 3504-3550. <https://doi.org/10.1214/21-EJS1798>
- [17] Gösmann, J., Kley, T. and Dette, H. (2021) A New Approach for Open-End Sequential Change Point Monitoring. *Journal of Time Series Analysis*, **42**, 63-84. <https://doi.org/10.1111/jtsa.12555>
- [18] Holmes, M. and Kojadinovic, I. (2021) Open-End Nonparametric Sequential Change-Point Detection Based on the Retrospective CUSUM Statistic. *The Electronic Journal of Statistics*, **15**, 2288-2335. <https://doi.org/10.1214/21-EJS1840>
- [19] Phillips, P.C.B. and Durlauf, S.N. (1986) Multiple Time Series Regression with Integrated Processes. *The Review of Economic Studies*, **53**, 473-495. <https://doi.org/10.2307/2297602>
- [20] Aue, A. and Horváth, L. (2004) Delay Time in Sequential Detection of Change. *Statistics & Probability Letters*, **67**, 221-231. <https://doi.org/10.1016/j.spl.2004.01.002>
- [21] Aue, A., Horváth, L. and Reimherr, M.L. (2009) Delay Times of Sequential Procedures for Multiple Time Series Regression Models. *Journal of Econometrics*, **149**, 174-190. <https://doi.org/10.1016/j.jeconom.2008.12.018>
- [22] Casini, A. and Perron, P. (2021) Continuous Record Laplace-Based Inference about the Break Date in Structural Change Models. *Journal of Econometrics*, **224**, 3-21. <https://doi.org/10.1016/j.jeconom.2020.05.020>
- [23] Dalla, V., Giraitis, L. and Phillips, P.C.B. (2022) Robust Tests for White Noise and Cross-Correlation. *Econometric Theory*, **38**, 913-941. <https://doi.org/10.1017/S0266466620000341>