

Moving Ratio Test for Multiple Mean Changes in Long Memory Time Series with Applications

Guolong Fu*, Mingcan He

Department of Mathematics and Statistics, Qinghai Normal University, Xining Qinghai
Email: *1070064380@qq.com

Received: Mar. 12th, 2018; accepted: Mar. 28th, 2018; published: Apr. 4th, 2018

Abstract

In this paper, by introducing a bandwidth parameter, a moving ratio type statistic is proposed to detect multiple mean changes in long memory time series. The null distribution of the test statistic and its consistency under the alternative hypothesis are proved. Simulations show that the proposed new method can effectively solve the problem of testing failure when using the single change point test method to test multiple change points. Finally, the practicability of the method is illustrated by two groups of actual data.

Keywords

Long Memory Time Series, Mean Change Point, Moving Ratio Statistics

长记忆时间序列均值多变点滑动比检验及应用

付国龙*, 何明灿

青海师范大学数学与统计学院, 青海 西宁
Email: *1070064380@qq.com

收稿日期: 2018年3月12日; 录用日期: 2018年3月28日; 发布日期: 2018年4月4日

摘要

本文通过引进一个窗宽参数, 提出了一种滑动比率型统计量来检验长记忆时间序中的列均值多变点。在

*通讯作者。

原假设下给出了检验统计量的极限分布, 在备择假设下证明了检验方法的一致性。模拟结果表明, 提出的新方法能够有效解决单变点检验方法在数据中存在多个变点时检验失效的问题, 最后通过两组实际数据说明了该方法的实用性。

关键词

长记忆时间序列, 均值多变点, 滑动比统计量

Copyright © 2018 by authors and Hans Publishers Inc.

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

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



Open Access

1. 引言

在过去几十年里, 检验各类模型和数据中的变点问题一直是统计学, 计量经济学和一些应用学科中的热点问题。均值变点作为一类最常见的变点, 文献中已有大量研究, 常见的均值变点检验方法主要有CUSUM方法、似然比方法、最小二乘法、Wald检验等。但现有的大部分文献主要研究的是独立序列或短期相依序列中的均值变点问题。越来越多的证据表明, 许多气象、经济、金融等时间序列具有长记忆性, 无法用传统的短记忆时间序列模型刻画, 这使得检验长记忆时间序列模型中的均值变点问题也变的非常重要。

Hidalgo 和 Robinson [1]在假定变点位置已知的情况下, 提出了一种检验长记忆时间序列均值变点的Wald方法, Wright [2]在变点位置未知的情况下, 研究了长记忆时间序列中的均值变点, 并提出了一种CUSUM检验方法和一种Sup-Wald检验方法。Lazarova [3]研究了线性回归模型含有长记忆时间序列误差项时结构变点的检验问题, Wang [4]研究了无限滑动平均时间序列均值变点的检验和估计问题, Shao [5]通过提出一种简单的比率型统计量继续研究长记忆时间序列均值变点的检验问题, 这种比率型检验方法的有点是不需要估计尺度参数等冗余参数, 在实际应用中非常方便。赵文芝等[6], 马健琦等[7]提出了一些新的比率型统计量来检验长记忆时间序列中的均值变点, 模拟结果表明这些新比率统计量具有比原比率型统计量更高的检验势。

上述关于长记忆时间序列均值变点检验的文献均考虑的是单变点问题, 由于数据中存在变点时, 不一定只有一个变点, 而基于单变点问题设计的检验方法在多变点检验问题中通常不适用, 所以研究长记忆时间序列均值多变点的检验问题也很重要。如今也有许多检验多变点问题的文献, 如Kejriwal [8]利用Wald检验方法研究了时间序列中持久性均值多变点问题, Bai [9]用似然比检验方法检验了短记忆时间序列中的均值多变点问题, 秦瑞兵等[10]提出了检验独立随机序列均值多变点的非参数方法, Chen [11] [12]等利用滑动比率方法研究了时间序列中持久性多变点的检验和估计问题。

比率型检验方法由于不需要估计尺度参数, 所以在应用中非常方便, 但现有的基于比率型方法检验长记忆时间序列均值变点的研究都只考虑的是单变点问题, 本文试图将比率型方法推广到均值多变点的检验问题当中去。推广的基本思想是通过引进一个窗宽参数, 利用滑动的方法比较当前时刻前后一段时间内的样本差异, 并决策是否存在变点。由于通过窗宽可以进入统计量中的样本中至多只有一个变点, 所以可以避免多变点改变统计量不单调的问题, 从而达到检测多变点问题的最终目的。第4节的模拟结果表明, 本文采用的改进方法行之有效。

2. 长记忆模型和检验统计量

令

$$(1-B)^d(Y_t - \mu) = u_t, \quad t=1, \dots, n$$

这里 B 是置后算子, μ 是 Y_t 的无条件均值, $\{u_t\}_{t \in \mathbb{Z}}$ 是均值为 0, 具有有限方差的独立同分布的平稳序列, n 为样本容量, 当参数 $d \neq 0$ 时, 称时间序列 Y_t 是长记忆时间序列, 记为 $Y_t \sim I(d)$. 参数 d 称为长记忆参数, 当 $d \in (-1/2, 1/2)$ 时, Y_t 为平稳序列, 而当 $d \in (1/2, 3/2)$ 时, Y_t 为非平稳长记忆序列. 长记忆时间序列 Y_t 可分解为如下形式:

$$Y_t = \sum_{j=0}^{\infty} \pi_j(d) u_{t-j}, \quad \pi_j(d) = \frac{\Gamma(d+j)}{j! \Gamma(d)}, \quad j=1, 2, \dots, \quad \pi_0(d) = 1.$$

其中 $\Gamma(\cdot)$ 是 gamma 函数.

设 $X_t = \mu_t + Y_t$ 其中 $Y_t \sim I(d)$, $d \in [0, 1/2)$, 本文考虑平稳长记忆时间序列均值多变点的检验问题, 即考虑如下假设检验问题:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_n$$

$$H_1: \mu_1 = \dots = \mu_{k_1-1} \neq \mu_{k_1} = \dots = \mu_{k_1+k_2-1} \neq \mu_{k_1+k_2} = \dots \\ = \mu_{k_1+k_2+\dots+k_{m-1}+1} \neq \mu_{k_1+k_2+\dots+k_m} \neq \mu_{k_1+k_2+\dots+k_n}$$

即在备择假设下长记忆时间序列 X_t 的均值:

$$\mu_t = \begin{cases} \mu_{k_1}, & t=1, 2, \dots, k_1^* \\ \mu_{k_1+k_2}, & t=k_1^*+1, \dots, k_2^* \\ \vdots & \vdots \\ \mu_{k_1+k_2+\dots+k_m}, & t=k_m^*+1, \dots, n \end{cases}$$

其中 $\mu_{k_1}, \mu_{k_1+k_2}, \dots, \mu_{k_1+k_2+\dots+k_m}$ 不全相等, 若定义

$$\mu_t = \begin{cases} \mu, & t=1, 2, \dots, k_1^* \\ \mu + \Delta_1, & t=k_1^*+1, \dots, k_2^* \\ \vdots & \vdots \\ \mu + \Delta_m, & t=k_m^*+1, \dots, n \end{cases}$$

则上述假设检验问题可转换为:

$$H_0: \Delta_1 = \Delta_2 = \dots = \Delta_m = 0;$$

$H_1: \Delta_1, \Delta_2, \dots, \Delta_m$ 至少有一个不为零.

这里 m 是变点的个数, 当 $m=1$ 时, Shao [5] 提出了如下比率统计量

$$G_n = \frac{\sqrt{n} |\bar{X}_{1,k} - \bar{X}_{k+1,n}|}{(n)^{-1} \left\{ \sum_{t=1}^k S_t^2(1,k) + \sum_{t=k+1}^n S_t^2(k+1,n) \right\}^{1/2}}, \quad 1 \leq k \leq n \quad (1)$$

其中

$$S_t(j,k) = \sum_{i=j}^t (X_i - \bar{X}_{j,k}), \quad \bar{X}_{j,k} = \frac{1}{k-j+1} \sum_{t=j}^k X_t, \quad t=j, \dots, k, \quad 1 \leq j < k \leq n.$$

上述检验统计量的优点是不需要估计样本方差, 所以在实际应用中非常方便, 但该统计量只能检验单个均值变点, 为了使该方法在多变点问题中仍然有效, 本文提出如下修正统计量:

$$H_L(k) = \frac{\sqrt{2L} |\bar{X}_{k-L+1,k} - \bar{X}_{k+1,k+L}|}{(2L)^{-1} \left\{ \sum_{t=k-L+1}^k S_t^2(k-L+1, k) + \sum_{t=k+1}^{k+L} S_t^2(k+1, k+L) \right\}^{1/2}} \quad (2)$$

其中 $k-L \leq k \leq k+L$, 若定义窗宽 $h \in (0, 1/2)$, 则 $L = \lfloor nh \rfloor$, $n = 1, 2, \dots$, $\lfloor \cdot \rfloor$ 表示取整。为推导该统计量的极限分布, 需做如下假设:

假设 1: 在备择假设下, 两个均值变点之间最少有 $2L \leq n$ 个样本。

假设 2: 当 $n \rightarrow \infty$ 时, 存在一个 $d \in (0, 1/2)$, 使得

$$n^{-(1/2+d)} \sum_{t=1}^{\lfloor nr \rfloor} \{X_t - E(X_t)\} \Rightarrow C_d W_d(r), \quad r \in [0, 1]$$

其中, C_d 为一个常数, $W_d(\cdot)$ 表示分数布朗运动。

假设 1 保证变点可以被区分, 这在多变点检验问题中是基本假设。假设 2 同 Shao [5] 中的假设 1, 是长记忆时间序列所满足的中心极限定理。

3. 检验统计量的渐近性质

本节将给出检验统计量(2)在原假设下的极限分布, 并证明检验方法在备择假设下的一致性。

定理 1: 在原假设 H_0 成立的条件下, 当 $n \rightarrow \infty$ 时, 有

$$H_L(k) = \frac{\sqrt{2} |2W_d(r) - W_d(r-1/2) - W_d(r+1/2)|}{\int_{r-1/2}^r \xi_1^2(s, r) d_s + \int_r^{r+1/2} \xi_2^2(s, r) d_s}$$

其中

$$\xi_1(s, r) = W_d(s) - (2s - 2r + 1)W_d(r) + 2(s - r)W_d(r - 1/2)$$

$$\xi_2(s, r) = W_d(s) - 2(s - r)W_d(r + 1/2) + (2s - 2r - 1)W_d(r)$$

证明: 令 $k = \lfloor 2Lr \rfloor$, $t = \lfloor 2Ls \rfloor$ 则由假设 2 有:

$$\begin{aligned} 2^{-1/2-d} L^{1/2-d} (\bar{X}_{k-L+1,k} - X_{k+1,k+L}) &= (2L)^{-1/2-d} \left[\sum_{t=k-L+1}^k (X_t - EX_t) - \sum_{t=k+1}^{k+L} (X_t - EX_t) \right] \\ &\Rightarrow C_d [2W_d(r) - W_d(r-1/2) - W_d(r+1/2)] \\ (2L)^{-1/2-d} S_t(k-L+1, k) &= (2L)^{-1/2-d} \sum_{i=k-L+1}^t \left(X_i - \frac{1}{L} \sum_{j=k-L+1}^k X_j \right) \\ &= (2L)^{-1/2-d} \left[\sum_{i=k-L+1}^t (X_i - EX_i) - \frac{t-k+L}{L} \sum_{j=k-L+1}^k (X_j - EX_j) \right] \\ &\Rightarrow C_d [W_d(s) - (2s - 2r + 1)W_d(r) + 2(s - r)W_d(r - 1/2)] \triangleq C_d \xi_1(s, r) \end{aligned}$$

同理可得

$$(2L)^{-1/2-d} S_t(k+1, k+L) = (2L)^{-1/2-d} \left[\sum_{i=k+1}^t (X_i - EX_i) - \frac{t-k}{L} \sum_{j=k+1}^{k+L} (X_j - EX_j) \right]$$

$$\Rightarrow C_d [W_d(s) - 2(s-r)W_d(r+1/2) + (2s-2r-1)W_d(r)] \triangleq C_d \xi_2(s, r)$$

因此

$$\begin{aligned} H_L(k) &= \frac{\sqrt{2L} |\bar{X}_{k-L+1,k} - \bar{X}_{k+1,k+L}|}{(2L)^{-1} \left\{ \sum_{t=k-L+1}^k S_t^2(k-L+1, k) + \sum_{t=k+1}^{k+L} S_t^2(k+1, k+L) \right\}^{1/2}} \\ &= \frac{2^{1+d} L^d \left| 2^{-1/2-d} L^{1/2-d} (\bar{X}_{k-L+1,k} - \bar{X}_{k+1,k+L}) \right|}{2^{1/2+d} L^d \left\{ (2L)^{-1} \sum_{t=k-L+1}^k \left[(2L)^{-1/2-d} S_t(k-L+1, k) \right]^2 + (2L)^{-1} \sum_{t=k+1}^{k+L} \left[(2L)^{-1/2-d} S_t(k+1, k+L) \right]^2 \right\}^{1/2}} \\ &\Rightarrow \frac{\sqrt{2} |2W_d(r) - W_d(r-1/2) - W_d(r+1/2)|}{\int_{r-1/2}^r \xi_1^2(s, r) d_s + \int_r^{r+1/2} \xi_2^2(s, r) d_s} \text{证毕。} \end{aligned}$$

定理 2: 若假设 1 成立, 则当数据中至少存在一个均值变点 $k^* \in [\lfloor nh \rfloor, \lfloor n(1-h) \rfloor]$ 时, 有

$$\lim_{L \rightarrow \infty} H_L(k) \rightarrow_p \infty.$$

证明: 由假设 1, 任意两个变点间至少有 $2L$ 个样本, 而统计量 $H_L(k)$ 只用到了当前时刻前后共 $2L$ 个样本, 即这些样本中至多只有一个变点, 因此只需要证明变点个数 $m=1$ 的情况即可。不妨记变点时刻为 $k^* = \lfloor 2Lr^* \rfloor$, 则, 当 $k = k^*$ 时,

$$H_L(k) = \frac{\sqrt{2L} |\bar{X}_{k^*-L+1,k^*} - \bar{X}_{k^*+1,k^*+L}|}{(2L)^{-1} \left\{ \sum_{t=k^*-L+1}^{k^*} S_t^2(k^*-L+1, k^*) + \sum_{t=k^*+1}^{k^*+L} S_t^2(k^*+1, k^*+L) \right\}^{1/2}}$$

由于 $S_t(k^*-L+1, k^*)$ 和 $S_t(k^*+1, k^*+L)$ 分别只用到了变点 k^* 之前的 L 个样本和变点之后的 L 个样本, 所以由定理 1 的证明有

$$S_t(k^*-L+1, k^*) = O_p\left((2L)^{1/2+d}\right)$$

$$S_t(k^*+1, k^*+L) = O_p\left((2L)^{1/2+d}\right)$$

从而统计量 $H_L(k)$ 的分母为 $O_p\left((2L)^{1/2+d}\right)$, 若记 EX_t 在变点 k^* 时刻由常数 μ 变为 $\mu + \Delta$, $\Delta \neq 0$, 另一方面, 统计量 $H_L(k)$ 的分子

$$\begin{aligned} & \sqrt{2L} |\bar{X}_{k^*-L+1,k^*} - \bar{X}_{k^*+1,k^*+L}| \\ & \geq \sqrt{2L} \left| \frac{1}{L} \sum_{t=k^*-L+1}^{k^*} (X_t - \mu) - \frac{1}{L} \sum_{t=k^*+1}^{k^*+L} (X_t - \mu - \Delta) \right| + \sqrt{2L} |L\Delta| \\ & = I_1 + I_2 \end{aligned}$$

其中

$$\begin{aligned} 2^{-(1+d)} L^{-d} I_1 &= 2^{-(1+d)} L^{-d} \sqrt{2L} \left| \frac{1}{L} \sum_{t=k^*-L+1}^{k^*} (X_t - \mu) - \frac{1}{L} \sum_{t=k^*+1}^{k^*+L} (X_t - \mu - \Delta) \right| \\ &= \left| (2L)^{-1/2-d} \sum_{t=k^*-L+1}^{k^*} (X_t - \mu) - (2L)^{-1/2-d} \sum_{t=k^*+1}^{k^*+L} (X_t - \mu - \Delta) \right| \end{aligned}$$

$$\Rightarrow C_d [2W_d(r^*) - W_d(r^* - 1/2) - W_d(r^* + 1/2)]$$

即 $I_1 = O_p(2^{1+d}L^d)$ 。而 $I_2 = \sqrt{2L}|L\Delta| = O_p(L^2)$ 。因此, 当 $n \rightarrow \infty$, 即 $L \rightarrow \infty$ 时, 有 $H_L(k) \rightarrow \infty$ 。证毕。

4. 数值模拟

本节基于 ARFIMA(0, d, 0) 模型通过模拟实验研究本文提出的统计量(2)的有限样本性质, 并和 Shao 的统计量(1)作比较。在统计量(1)中参数 τ 分别取 0.1 和 0.2, 修正后的统计量(2)中的窗宽参数 h 同样取 0.1 和 0.2, 所有的结果都是利用 R 语言模拟 10000 次得到的。本文所用到的临界值是由数据生成过程 ARFIMA(0, d, 0) 生成 10000 组容量为 10000 的随机序列计算所得。在模拟实验中, 分别取样本容量为 100 和 400, 长记忆参数取 0、0.2 和 0.4。

表 1 给出了原假设下两种检验统计量在 5% 检验水平下的经验水平, 可以看出两种方法在所有参数组合下都能较好的控制经验水平。比较而言, 本文提出的修正统计量在样本量较小时相对保守一些, 长记忆参数 d 对经验水平的影响不明显。

为检验两种统计量的检验势, 分别考虑数据中存在 $m = 1, 2$ 和 3 个变点的情况, 跳跃度 Δ 分别取 1 和 2, 变点位置 k^* 分别取 0.25、0.5 和 0.8。表 2 和表 3 分别列出了两种检验统计量在变点个数 $m = 1$ 时的

Table 1. Empirical size of the two statistics at the 5% test level

表 1. 两种统计量在 5% 检验水平下的经验水平

d	H_L				G_n			
	n = 100		n = 400		n = 100		n = 400	
	h = 0.1	h = 0.2	h = 0.1	h = 0.2	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.1$	$\tau = 0.2$
0	0.024	0.027	0.030	0.040	0.039	0.042	0.045	0.051
0.2	0.017	0.027	0.034	0.040	0.044	0.047	0.047	0.050
0.4	0.017	0.034	0.035	0.051	0.043	0.047	0.045	0.046

Table 2. Number of modified Shao statistics at change points $m = 1$ and empirical power at the 5% level

表 2. 修正的 Shao 统计量在变点个数 $m = 1$, 经验水平为 5% 时的经验势

h	Δ	k^*	n = 100			n = 400		
			d = 0	d = 0.2	d = 0.4	d = 0	d = 0.2	d = 0.4
0.1	1	0.25	0.092	0.037	0.025	0.560	0.137	0.051
		0.50	0.094	0.039	0.025	0.577	0.132	0.053
		0.80	0.091	0.038	0.026	0.598	0.15	0.058
	2	0.25	0.455	0.149	0.058	0.987	0.547	0.140
		0.50	0.477	0.176	0.068	0.995	0.568	0.140
		0.80	0.461	0.158	0.064	0.988	0.597	0.152
0.2	1	0.25	0.381	0.119	0.061	0.961	0.340	0.099
		0.50	0.412	0.145	0.064	0.958	0.340	0.092
		0.80	0.276	0.091	0.052	0.9	0.263	0.091
	2	0.25	0.928	0.485	0.182	1	0.886	0.284
		0.50	0.937	0.509	0.204	1	0.869	0.278
		0.80	0.867	0.999	0.407	0.805	0.145	0.248

Table 3. Number of Shao statistics at change points $m=1$ and empirical power at the 5% level
表 3. Shao 统计量在变点个数 $m=1$, 经验水平为 5% 时的经验势

τ	Δ	k^*	$n=100$			$n=400$		
			$d=0$	$d=0.2$	$d=0.4$	$d=0$	$d=0.2$	$d=0.4$
0.1	1	0.25	0.617	0.241	0.104	0.981	0.49	0.102
		0.50	0.680	0.283	0.092	0.996	0.576	0.127
		0.80	0.612	0.204	0.078	0.98	0.475	0.094
	2	0.25	0.986	0.691	0.215	1	0.932	0.324
		0.50	0.997	0.8	0.318	1	0.983	0.402
		0.80	0.971	0.643	0.221	1	0.88	0.267
0.2	1	0.25	0.77	0.292	0.098	0.997	0.579	0.2
		0.50	0.863	0.4	0.129	1	0.707	0.3
		0.80	0.673	0.234	0.089	0.991	0.486	0.184
	2	0.25	0.996	0.759	0.278	1	0.96	0.485
		0.50	1	0.87	0.411	1	0.997	0.634
		0.80	0.987	1	0.656	0.904	0.232	0.399

检验势。从中可以看出, 随着样本容量的增大和跳跃度的增大, 两个统计量的检验势都逐渐增大, 这验证了两个统计量都是一致检验的定理结论。此外, 随着长记忆参数 d 的增大, 检验势逐渐降低, 变点位置越靠近中间时检验势也越高。由于本文检验统计量只用了部分样本, 所以在数据中只有 1 个变点时, 本文的修正检验统计量的检验势低于原统计量, 但随着样本容量的增大和窗宽参数的增大, 这种差距逐渐减小。

表 4~7 分别给出了两个统计量在变点个数 m 取值为 2 和 3 时的检验势, 此时可以发现本文提出的滑动比率方法的经验势普遍高于原统计量, 尤其在 m 取 3 时更为显著。在大部分情况下, 原统计量均很难检测到变点, 这说明本文基于部分样本的滑动比率方法在多变点的检验问题中是可行的。由于在实际问题中无法提前知道数据中有几个变点, 所以可以两种方法同时使用, 如果两种方法都不拒绝无变点的原假设, 则可认为数据中不存在变点; 若至少有一个统计量拒绝原假设则可认为数据中存在变点。特别地, 若统计量(1)没有拒绝原假设, 而统计量(2)拒绝了原假设。则可怀疑数据中存在多个变点, 需进一步仔细分析。

5. 实例分析

本节用两组实际数据来验证本文所述的假设检验问题。第一组数据选自从 1871 年到 1970 年尼罗河阿斯旺水坝年度流量数据(见图 1), 第二组数据选自从 1854 年到 1989 年北半球气温数据(见图 2), Shao [5] 基于统计量(1)检验发现这两组数据中存在一个均值变点。下面利用本文提出的新统计量重新检验这两组数据, 第一组数据中发现当 $h=0.1$ 时, 检验统计量(2)在第 28 个样本处达到最大值 52.614, 其对应的临界值为 34.73, 当 $h=0.2$ 时, 检验统计量(2)在第 28 个样本处达到最大值 37.65, 其对应的临界值为 22.784, 由于统计量的值大于对应的临界值, 说明该数据集在第 28 个样本处发生了变化。经计算发现前 28 个样本的均值为 1097.75, 后 72 个样本的均值为 849.972, 前后差异较大, 说明该组数据在第 28 个样本处存在变点, 该结果与 Shao 的检验结果相一致。同样的, 对于第二组数据发现当 $h=0.1$ 和 $h=0.2$ 时, 检验统计量(2)在第 826 个样本处达到最大值 75.28, 用同样的检验方法发现该组数据在第 826 个样本处存在均值变点, 该结果也与 Shao 的检验结果相一致。

Table 4. Number of Shao statistics at change points $m = 2$ and empirical power at the 5% level
表 4. 修正的 Shao 统计量在变点个数 $m = 2$, 经验水平为 5% 时的经验势

h	$\Delta 1$	$\Delta 2$	k_1^*	k_2^*	$n = 100$			$n = 400$		
					$d = 0$	$d = 0.2$	$d = 0.4$	$d = 0$	$d = 0.2$	$d = 0.4$
0.1	-1	0.5	0.25	0.50	0.305	0.1	0.053	0.958	0.408	0.095
			0.50	0.80	0.293	0.091	0.044	0.96	0.411	0.11
			0.25	0.80	0.306	0.094	0.05	0.959	0.407	0.104
		1	0.25	0.50	0.520	0.192	0.074	0.999	0.617	0.160
			0.50	0.80	0.498	0.173	0.064	0.995	0.648	0.176
			0.25	0.80	0.506	0.177	0.071	0.994	0.637	0.17
	2	1	0.25	0.50	0.478	0.159	0.06	0.991	0.594	0.158
			0.50	0.80	0.515	0.191	0.073	1	0.615	0.153
			0.25	0.80	0.472	0.154	0.062	0.996	0.592	0.148
		0.5	0.25	0.50	0.552	0.188	0.068	1	0.684	0.183
			0.50	0.80	0.593	0.222	0.078	0.996	0.695	0.18
			0.25	0.80	0.552	0.186	0.067	0.998	0.677	0.174
0.2	-1	0.5	0.25	0.50	0.792	0.455	0.129	1	0.695	0.186
			0.50	0.80	0.700	0.277	0.108	1	0.690	0.185
			0.25	0.80	0.769	0.297	0.105	1	0.698	0.190
		1	0.25	0.50	0.941	0.719	0.211	1	0.884	0.288
			0.50	0.80	0.882	0.447	0.164	1	0.851	0.275
			0.25	0.80	0.913	0.464	0.168	1	0.858	0.289
	2	1	0.25	0.50	0.931	0.708	0.197	1	0.914	0.298
			0.50	0.80	0.947	0.527	0.218	1	0.889	0.282
			0.25	0.80	0.939	0.507	0.202	1	0.910	0.297
		0.5	0.25	0.50	0.961	0.755	0.235	1	0.942	0.348
			0.50	0.80	0.962	0.572	0.244	1	0.928	0.322
			0.25	0.80	0.967	0.569	0.230	1	0.940	0.335

6. 总结

本文研究了长记忆时间序列均值多变点的检验问题, 基于已有的一种比率型统计量, 通过引进一个窗宽参数, 提出了一种滑动比率型检验统计量。在原假设下推导出了新统计量的极限分布, 在备择假设下证明了统计量的一致性, 并通过蒙特卡罗模拟说明了本文所选方法的有效性, 最后通过一组实际数据说明了本文方法的可行性。由于本问提出的滑动比率方法相对更保守, 且在数据中只有一个变点时, 检验势低于原检验方法, 所以建议在实际问题中两种方法同时使用, 如果两种方法都不拒绝无变点的原假设, 则可认为数据中不存在变点; 若至少有一个统计量拒绝原假设则可认为数据中存在变点。特别地, 若原没有拒绝原假设, 而本文提出的新统计量拒绝了原假设, 则可怀疑数据中存在多个变点, 需进一步仔细分析。

Table 5. Number of Shao statistics at change points $m = 2$ and empirical power at the 5% level
表 5. Shao 统计量在变点个数 $m = 2$, 经验水平为 5% 时的经验势

τ	$\Delta 1$	$\Delta 2$	k_1^*	k_2^*	$n = 100$			$n = 400$		
					$d = 0$	$d = 0.2$	$d = 0.4$	$d = 0$	$d = 0.2$	$d = 0.4$
0.1	-1	0.5	0.25	0.50	0.053	0.055	0.051	0.011	0.107	0.063
			0.50	0.80	0.001	0.017	0.039	0	0.007	0.049
			0.25	0.80	0.009	0.045	0.064	0	0.016	0.065
		1	0.25	0.50	0.304	0.179	0.107	0.394	0.392	0.135
			0.50	0.80	0.028	0.074	0.068	0	0.051	0.094
			0.25	0.80	0.07	0.103	0.088	0.001	0.064	0.123
	2	1	0.25	0.50	0.085	0.109	0.083	0.014	0.153	0.092
			0.50	0.80	0.402	0.243	0.127	0.569	0.315	0.179
			0.25	0.80	0.172	0.174	0.127	0.051	0.155	0.15
		0.5	0.25	0.50	0	0.003	0.024	0	0	0.028
			0.50	0.80	0.014	0.045	0.062	0.001	0.01	0.06
			0.25	0.80	0.003	0.025	0.055	0	0.007	0.056
0.2	-1	0.5	0.25	0.50	0.146	0.094	0.064	0.079	0.092	0.086
			0.50	0.80	0.002	0.009	0.032	0	0.002	0.041
			0.25	0.80	0.007	0.038	0.051	0	0.009	0.062
		1	0.25	0.50	0.568	0.262	0.135	0.831	0.358	0.167
			0.50	0.80	0.006	0.038	0.047	0	0.02	0.076
			0.25	0.80	0.058	0.094	0.083	0.001	0.056	0.111
	2	1	0.25	0.50	0.133	0.135	0.091	0.038	0.108	0.096
			0.50	0.80	0.666	0.344	0.166	0.92	0.426	0.215
			0.25	0.80	0.351	0.225	0.144	0.311	0.229	0.167
		0.5	0.25	0.50	0	0.004	0.022	0	0	0.028
			0.50	0.80	0.085	0.071	0.077	0.007	0.035	0.083
			0.25	0.80	0.016	0.038	0.062	0	0.009	0.064

Table 6. The empirical power of the modified Shao statistic when the number of change points is $m = 3$, the position of change point is 0.25, 0.5, 0.8, and the empirical level is 5%

表 6. 修正的 Shao 统计量在变点个数 $m = 3$, 变点位置分别为 0.25、0.5、0.8, 经验水平为 5% 时的经验势

d	n	$h = 0.1$			$h = 0.2$		
		$\Delta 1 = -1$ $\Delta 2 = 1$ $\Delta 3 = 3$	$\Delta 1 = 3$ $\Delta 2 = 2$ $\Delta 3 = 1$	$\Delta 1 = 1$ $\Delta 2 = 3$ $\Delta 3 = -1$	$\Delta 1 = -1$ $\Delta 2 = 1$ $\Delta 3 = 3$	$\Delta 1 = 3$ $\Delta 2 = 2$ $\Delta 3 = 1$	$\Delta 1 = 1$ $\Delta 2 = 3$ $\Delta 3 = -1$
0	100	0.744	0.847	0.990	0.988	0.996	1
	400	1	1	1	1	1	1
0.2	100	0.315	0.441	0.765	0.708	0.847	0.961
	400	0.835	0.918	0.993	0.966	0.995	1
0.4	100	0.115	0.175	0.380	0.312	0.444	0.657
	400	0.258	0.358	0.628	0.447	0.600	0.789

Table 7. The empirical power of the Shao statistic when the number of change points is $m = 3$, the position of change point is 0.25, 0.5, 0.8, and the empirical level is 5%

表 7. Shao 统计量在变点个数 $m = 3$, 变点位置分别为 0.25、0.5、0.8 经验水平为 5% 时的经验势

d	n	$\tau = 0.1$			$\tau = 0.2$		
		$\Delta 1 = -1$ $\Delta 2 = 1$ $\Delta 3 = 3$	$\Delta 1 = 3$ $\Delta 2 = 2$ $\Delta 3 = 1$	$\Delta 1 = 1$ $\Delta 2 = 3$ $\Delta 3 = -1$	$\Delta 1 = -1$ $\Delta 2 = 1$ $\Delta 3 = 3$	$\Delta 1 = 3$ $\Delta 2 = 2$ $\Delta 3 = 1$	$\Delta 1 = 1$ $\Delta 2 = 3$ $\Delta 3 = -1$
0	100	0.089	0.003	0	0.567	0.008	0
	400	0.001	0	0	0.268	0	0
0.2	100	0.113	0.026	0	0.216	0.031	0
	400	0.022	0.002	0	0.100	0.005	0
0.4	100	0.101	0.073	0.007	0.137	0.080	0.005
	400	0.099	0.076	0.003	0.131	0.080	0

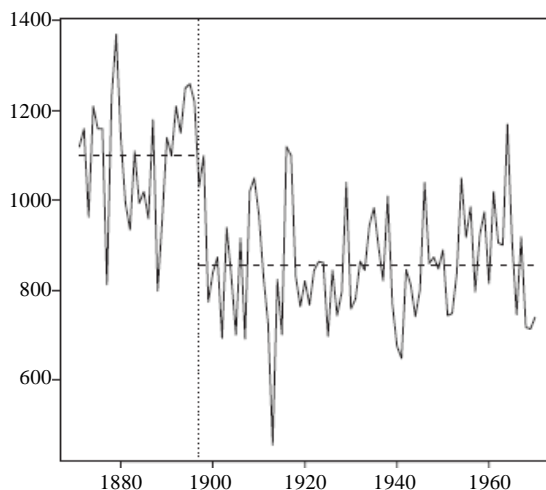


Figure 1. Annual flow time series of the Nile Aswan Dam
图 1. 尼罗河阿斯旺水坝年度流量时间序列图

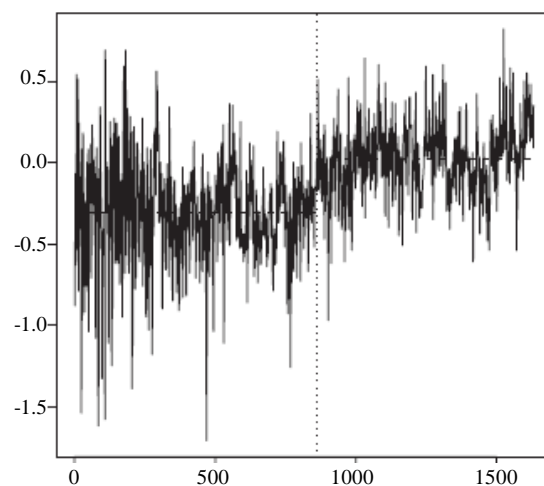


Figure 2. Time series of annual temperature data in the northern hemisphere
图 2. 北半球年度气温数据时间序列图

致 谢

在此, 我要感谢我的导师陈占寿老师, 感谢陈老师对我在生活上的关心和学业上的精心指导。他严谨求实、精益求精的治学态度和对专业前沿问题敏锐的洞察力深深地影响了我, 使我不仅学到了本专业的理论知识, 而且培养了我发现问题、分析问题、解决问题的能力, 使我终身受益。我还要感谢国家自然科学基金、青海省自然科学基金的资助, 以及所引用文献的作者对科学做出的贡献。

基金项目

国家自然科学基金(11301291, 11661067), 青海省自然科学基金(2015-ZJ-717)资助。

参考文献

- [1] Hidalgo, J. and Robinson, P.M. (1996) Testing for Structural Change in a Long-Memory Environment. *Journal of Econometrics*, **70**, 159-174. [https://doi.org/10.1016/0304-4076\(94\)01687-9](https://doi.org/10.1016/0304-4076(94)01687-9)
- [2] Wright, J.H. (2010) Testing for a Structural Break at Unknown Date with Long-Memory Disturbances. *Journal of Time Series Analysis*, **19**, 369-376. <https://doi.org/10.1111/1467-9892.00097>
- [3] Štěpána, L. (2005) Testing for Structural Change in Regression with Long Memory Processes. *Journal of Econometrics*, **129**, 329-372. <https://doi.org/10.1016/j.jeconom.2004.09.011>
- [4] Wang, L. (2008) Change-in-Mean Problem for Long Memory Time Series Models with Applications. *Journal of Statistical Computation & Simulation*, **78**, 653-668. <https://doi.org/10.1080/00949650701216604>
- [5] Shao, X. (2011) A Simple Test of Changes in Mean in the Possible Presence of Long-Range Dependence. *Journal of Time Series Analysis*, **32**, 598-606. <https://doi.org/10.1111/j.1467-9892.2010.00717.x>
- [6] 赵文芝, 夏志明, 贺飞跃. 长相依序列均值变点检测[J]. 数学的实践与认识, 2016, 46(3): 204-208.
- [7] 马健琦, 陈占寿, 吕娜. 基于 Sieve Bootstrap 方法的长记忆过程均值变点的检验[J]. 青海师范大学学报(自科版), 2017, 33(2): 34-38.
- [8] Kejriwal, M. and Zhou, J. (2013) Wald Tests for Detecting Multiple Structural Changes in Persistence. *Econometric Theory*, **29**, 289-323. <https://doi.org/10.1017/S0266466612000357>
- [9] Bai, J. (1999) Likelihood Ratio Tests for Multiple Structural Changes. *Journal of Econometrics*, **91**, 299-323. [https://doi.org/10.1016/S0304-4076\(98\)00079-7](https://doi.org/10.1016/S0304-4076(98)00079-7)
- [10] 秦瑞兵, 田铮, 陈占寿. 独立随机序列均值多变点的非参数检测[J]. 应用概率统计, 2013, 29(5): 449-457.
- [11] Chen, Z., Tian, Z., et al. (2012) Moving Ratio Test for Multiple Changes in Persistence. *Journal of Systems Science & Complexity*, **25**, 582-593. <https://doi.org/10.1007/s11424-012-9255-9>
- [12] Chen, Z., Jin, Z., Tian, Z., et al. (2012) Bootstrap Testing Multiple Changes in Persistence for a Heavy-Tailed Sequence. *Computational Statistics & Data Analysis*, **56**, 2303-2316. <https://doi.org/10.1016/j.csda.2012.01.011>

Hans 汉斯

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>
期刊邮箱: sa@hanspub.org