

# 基于似然比扫描方法的INGARCH模型变点探测

王萌, 卢飞龙\*

辽宁科技大学理学院, 辽宁 鞍山

收稿日期: 2024年2月21日; 录用日期: 2024年3月22日; 发布日期: 2024年3月29日

## 摘要

整数值时间序列数据在许多领域中非常常见, 可用整数值广义自回归条件异方差(INGARCH)模型来拟合。本文研究了INGARCH模型中的变点探测问题, 基于似然比扫描方法(LRSM), 讨论了分段平稳的INGARCH模型中变点的数量和位置。然后通过大量数值模拟, 验证LRSM的有效性, 最后并将其应用于实际数据的分析中。

## 关键词

整数值时间序列, INGARCH模型, 变点探测, 似然比扫描方法

## Change Point Detection of INGARCH Models Based on Likelihood Ratio Scanning Method

Meng Wang, Feilong Lu\*

College of Science, University of Science and Technology Liaoning, Anshan Liaoning

Received: Feb. 21<sup>st</sup>, 2024; accepted: Mar. 22<sup>nd</sup>, 2024; published: Mar. 29<sup>th</sup>, 2024

## Abstract

Integer-valued time series data are very common in many fields and can be fitted by integer-valued generalized autoregressive conditional heteroscedasticity (INGARCH) models. In this paper, the problem of change point detection in INGARCH model is studied. Based on the likelihood ratio scanning method (LRSM), the number and location of change points in piecewise stationary INGARCH process are discussed. Then the validity of LRSM is verified by a large number of numerical simulations, and finally it is applied to the analysis of real data.

\*通讯作者。

## Keywords

**Integer-Valued Time Series, INGARCH Model, Change Point Detection, Likelihood Ratio Scanning Method**

Copyright © 2024 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

近几十年来, 变点探测问题一直是统计学中的一个热门话题。如今它不仅被广泛应用于工业质量控制[1], 还被广泛应用于经济[2]、金融[3]、医学[4]等领域。许多学者对变点探测问题进行了研究。例如, Picard [5]研究了自回归 AR 模型变点的最大似然估计。Bai [6]对最小二乘估计线性回归模型的多变点问题进行了深入探索。Kokoszka 和 Leipus [7]研究了自回归条件异方差(ARCH)模型中单变点检测的相关问题。Liu [8]提出了一个新的经验似然比统计量来检验无变点的零假设。Foygel 和 Drton [9]基于群 LASSO 法, 解决了数据稀疏条件下线性回归模型的变点问题。Hudecová, Hušková 和 Meintanis [10]研究了 INARCH 模型中的变点探测问题。Chen 和 Lee [11] [12]提出了一种贝叶斯方法来检测广义泊松 INGARCH 模型中的变点。Pein 和 Sieling [13]提出了一种 H-SMUCE 的统计量, 用于检测异质高斯回归模型中信号的多个变点问题。Lee [14]利用支持向量回归(SVR)-自回归移动平均(ARMA)模型获得的残差, 考虑了基于位置和尺度的 CUSUM 检验的时间序列变点检验问题。Lee 和 Kim [15]重点介绍了整数自回归(INAR)和 INGARCH 模型的 CUSUM 检验和整数时间序列模型变点检验的最新进展。

在时间维度上, 计数过程体现在生产生活各方面, 比如一个城市在数月内道路交通事故发生次数、一家医院某种传染疾病每天的感染人数、商场中销售的某个商品数量、某只股票每天价格变动的次数、某一地区的犯罪数量和某个国家新增的失业人数等等。取值非负整数和有一定的自相关性是这种数据的两个明显的特点。而对于这类数据, 如果用广义自回归条件异方差模型(GARCH)来拟合, 效果不好。虽然这个模型可用来拟合有自相关关系的数据, 但这个模型是在实数域基础上而非整数域, 故对于这种数据, 用整数时间序列模型进行拟合要更好一些。近三十年来, 许多整数时间序列模型应运而生, 其应用前景也越来越广阔。比如为描述癫痫患者的发作次数, Franke 和 Seligmann [16]提出自激励门限自回归模型、Gauthier 和 Latour [17]提出的广义整数自回归模型等等。本文研究的整数时间序列模型分别是由 Ferland [18]提出的泊松 INGARCH 模型, Fokianos 和 Tjøstheim [19]提出对数线性泊松 INGARCH 模型。

变点探测常用的方法是通过优化特定的目标函数来寻找变点, 如最小二乘法[20]和贝叶斯方法[21]。为了避免优化问题, BS 方法[22]是一种流行的解决方案。此外, 还有 Davis 等人[23]提出的遗传算法、Killic 等人[24]提出的 Pruned Exact Linear Time (PELT)方法和 Yau 等人[25]提出的 LRSM。相比较于 BS 和 PELT, LRSM 不仅将计算难度从  $n^2$  降低到  $\log(n)$ , 且提高了变点估计的准确性。目前, 据我们所知, LRSM 还没有运用到整数时间序列的变点探测中。

本文的梗概如下。第 2 节中, 我们将详细介绍基于 LRSM 在分段平稳过程中估计多个变点的三个步骤。第 3 节讨论了窗口半径的选取问题。第 4 节对 LRSM 进行了大量的模拟研究。通过对不同模型的重复试验, 深入研究了 LRSM 的变点探测性能。第 5 节研究了 LRSM 在实际生活中的应用。第 6 节为对全文内容做出总结, 概述本论文所做的工作。

## 2. INGARCH 模型以及似然比扫描方法

### 2.1. 泊松 INGARCH 模型

拟合整数时间序列, 其中最为经典的就是用 Ferland 等(2006) [18]提出的泊松 INGARCH 模型来拟合, 其定义如下: 一个整数值过程  $\{X_t\}_{t \in \mathbb{Z}}$ , 满足

$$\begin{cases} X_t | \mathcal{F}_{t-1} : \mathcal{P}(\lambda_t) & \forall t \in \mathbb{Z}, \\ \lambda_t = \gamma_0 + \sum_{i=1}^q \gamma_i X_{t-i} + \sum_{j=1}^p \delta_j \lambda_{t-j}, \end{cases} \quad (1)$$

其中,  $\mathbb{Z}$  是正整数域,  $\gamma_0 > 0$ ,  $\gamma_i \geq 0$ ,  $i=1, \dots, q$ ,  $\delta_j \geq 0$ ,  $j=1, \dots, p$ ,  $\mathcal{F}_{t-1}$  表示由  $\{X_1, \dots, X_{t-1}\}$  生成的  $\sigma$  域,  $\mathcal{P}(\lambda_t)$  表示均值为  $\lambda_t$  的泊松分布。

为了简单起见, 本节主要考虑泊松 INGARCH (1, 1) 模型的变点探测。

泊松 INGARCH (1, 1) 模型为

$$\begin{cases} X_t | \mathcal{F}_{t-1} : \mathcal{P}(\lambda_t) & \forall t \in \mathbb{Z}, \\ \lambda_t = \gamma_0 + \gamma_1 X_{t-1} + \delta_1 \lambda_{t-1}, \end{cases} \quad (2)$$

关于该模型的统计性质有

- 1) 当  $\gamma_1 + \delta_1 < 1$  时, 该模型是严平稳的。
- 2) 该过程的期望值为  $E(X_t) = \frac{\gamma_0}{1 - (\gamma_1 + \delta_1)}$ 。
- 3) 该过程的方差为  $Var(X_t) = \frac{\mu(1 - (\gamma_1 + \delta_1)^2 + \gamma_1^2)}{1 - (\gamma_1 + \delta_1)^2}$ 。

### 2.2. 似然比扫描方法

LRSM 相较于其它变点探测方法, 能够快速有效的探测出时间序列中的变点。本节详细介绍此方法的三个步骤。具体 LRSM 变点探测步骤如下:

步骤 1: 扫描观测序列  $\{X_t\}_{t=1, \dots, n}$ , 获得初始变点估计  $\hat{\mathcal{J}}^{(1)}$ ;

步骤 2: 结合最小描述长度准则(MDL), 得到一致变点估计  $\hat{\mathcal{J}}^{(2)}$ ;

步骤 3: 获得最终的变点估计  $\hat{\mathcal{J}}^{(3)}$ 。

具体变点探测问题流程图如图 1 所示。

本节详细介绍分段平稳的泊松 INGARCH 模型中的变点探测和估计问题。假设观测值  $\{X_t\}_{t=1, \dots, n}$  来自于  $m+1$  段平稳的泊松 INGARCH 过程, 则第  $j$  段泊松 INGARCH 的表达式为:

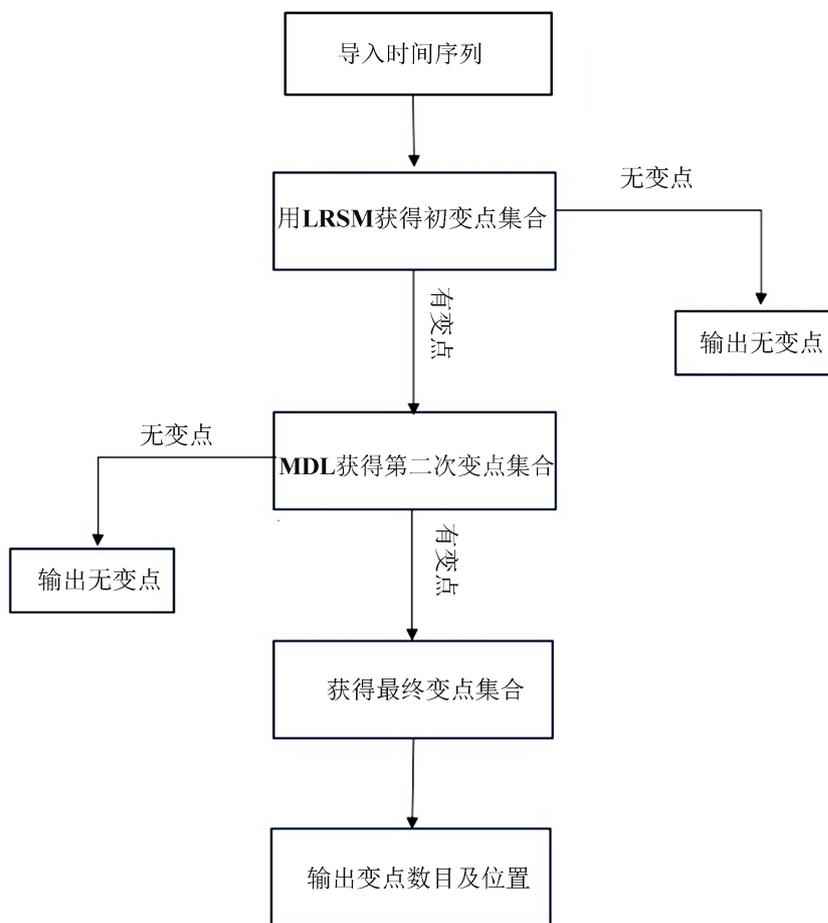
$$Y_{t,j} = X_t, \quad \tau_{j-1} < t \leq \tau_j, \quad (3)$$

其中  $j=1, \dots, m$ ,  $\tau_j$  是泊松 INGARCH (1, 1) 过程的第  $j$  个变点, 令  $\tau_0 = 1$ ,  $\tau_{m+1} = n$ ,  $\{Y_{t,j}\}$  是平稳的泊松 INGARCH (1, 1) 过程, 满足

$$\begin{cases} Y_{t,j} | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_{t,j}) & \forall t \in \mathbb{Z}, \\ \lambda_{t,j} = \gamma_{j,0} + \gamma_{j,1} Y_{t-1,j} + \delta_{j,1} \lambda_{t-1,j}, \end{cases} \quad (4)$$

其中  $\gamma_{j,0} > 0$ ,  $\gamma_{j,1} \geq 0$ ,  $\delta_{j,1} \geq 0$ ,  $\gamma_{j,1} + \delta_{j,1} < 1$ 。记  $\mathcal{J} = (\tau_1, \dots, \tau_m)$  为变点集合, 定义  $\tau_j = \lceil \varpi_j n \rceil$ ,  $j=1, \dots, m+1$ ,  $\varpi_j \in [0, 1]$  是第  $j$  个变点的相对位置, 对于  $\varepsilon_\sigma > 0$ , 有  $\min_{j=0, \dots, m} (\varpi_{j+1} - \varpi_j) > \varepsilon_\sigma$ 。下面对基于 LRSM 探测

分段平稳的泊松 INGARCH 过程中变点的三个步骤进行详细介绍。



**Figure 1.** The flowchart depicts the procedure for detecting change points in a piecewise stationary autoregressive process

**图 1.** 分段平稳自回归过程变点探测流程图

第一步：扫描来自于分段平稳泊松 INGARCH 过程的观测序列  $\{X_t\}_{t=1,\dots,n}$ ，得到最初估计的变点。定义  $t$  点的扫描窗口为

$$W_t(h) = \{t-h+1, \dots, t+h\},$$

和相应的观察值为

$$X_{W_t(h)} = (X_{t-h+1}, \dots, X_{t+h}),$$

其中  $t = h, \dots, n-h$ ， $h$  为窗口半径，关于  $h$  的选择详细可见第 3 节。

为了获得扫描窗口中初始估计的变点，可以选择似然比统计量。对于一个来自泊松 INGARCH 模型的样本  $d = (d_1, \dots, d_n)$  的条件似然函数为

$$l(\theta) = \ln \sum_{t=1}^n \left[ p(d_t | d_{t-1}, \dots, d_{t-p}) \right], \quad (5)$$

其中  $p(d_t | d_{t-1}, \dots, d_{t-p}) = \frac{e^{-\lambda_t} \lambda_t^{d_t}}{d_t!}$  是先前给定观测值  $d_s$  的条件密度函数，且当  $s \leq 0$  时  $d_s = 0$ 。

定义扫描窗口  $W_t(h)$  的似然比扫描统计量为

$$S_h(t) = \frac{1}{h} l_h(t, \hat{\theta}_1) + \frac{1}{h} l_{2h}(t, \hat{\theta}_2) - \frac{1}{h} l_h(t, \hat{\theta}), \quad (6)$$

其中  $l_{1h}(t, \hat{\theta}_1)$ ,  $l_{2h}(t, \hat{\theta}_2)$  和  $l_h(t, \hat{\theta})$  分别是  $\{X_s\}_{t-h+1, \dots, t}$ ,  $\{X_s\}_{t+1, \dots, t+h}$  和  $\{X_s\}_{W_t(h)}$  的条件似然函数,  $\hat{\theta}_1$ ,  $\hat{\theta}_2$ ,  $\hat{\theta}$  分别是基于序列  $\{X_s\}_{t-h+1, \dots, t}$ ,  $\{X_s\}_{t+1, \dots, t+h}$  和  $\{X_s\}_{W_t(h)}$  中参数  $\theta$  的估计量。

利用  $S_h(t)$  扫描观测序列, 可以得到一系列似然比扫描统计量  $(S_h(h), S_h(h+1), \dots, S_h(n-h))$ 。假如变点为  $t$ ,  $S_h(t)$  一般会较大。若选择的  $h$ , 满足  $2h < n\varepsilon_\sigma$ , 那么每个扫描窗口内最多存在一个变点。因此, 通过局部变点位置的估计, 可以得到潜在变点集合并定义为

$$\hat{\mathcal{J}}^{(1)} = \left\{ m \in \{h, h+1, \dots, n-h\} : S_h(m) = \max_{t \in [m-h, m+h]} S_h(t) \right\}. \quad (7)$$

如果  $S_h(m)$  是以点  $m$  为中心的窗口  $[m-h+1, m+h]$  上的最大值, 则  $m$  为一个局部变点。则有一系列的初始变点  $\hat{\mathcal{J}}^{(1)} = (\hat{\tau}_1^{(1)}, \hat{\tau}_2^{(1)}, \dots, \hat{\tau}_{\hat{m}^{(1)}}^{(1)})$ , 其中  $\hat{m}^{(1)} = |\hat{\mathcal{J}}^{(1)}|$ 。

**定理 1** 记录  $\mathcal{J}_0 = (\tau_1^0, \dots, \tau_{m_0}^0)$  为实际变点集, 假设  $2h < n\varepsilon_\sigma$  且  $\varepsilon_\sigma > cn^{-q}$ , 则有  $q \in [0, 1)$ ,  $c > 0$ , 满足  $\mathbf{P} \left( \max_{\tau_i^0 \in \mathcal{J}_0, k=1, \dots, \hat{m}^{(1)}} |\tau_i^0 - \hat{\tau}_k^{(1)}| < h \right) \rightarrow 1$ 。

定理 1 表明全部真实变点在初始估计变点集合  $\hat{\mathcal{J}}^{(1)}$  的  $h$ -邻域内, 但此时并不能保证  $\hat{m}^{(1)}$  等于真实变点数量  $m_0$ 。也就是说在第一步得到的变点估计往往会有过估计问题, 把一些不是变点的值当作了变点, 导致  $\hat{\mathcal{J}}^{(1)}$  内不但有真实变点集合, 还有一些非变点。下一步将利用合适的信息准则来解决以上问题, 获得一致的变点估计。定理 1 的证明过程可参考文献[25]。

第二步: 结合 MDL, 得到一致变点估计。

为了从  $\hat{\mathcal{J}}^{(1)}$  中找出准确的变点, 可以使用 MDL 准则进行 INGARCH 模型的选择过程。定义 MDL 如下

$$\text{MDL}(m, \mathcal{J}) = \log(m) + (m+1)\log(n) + \sum_{j=1}^{m+1} 2\log(n_j) - \sum_{j=1}^{m+1} l_j(\hat{\theta}_j), \quad (8)$$

其中  $\mathcal{J} = (\tau_1, \dots, \tau_m)$  是第一步中得到的初始估计变点的集合  $\hat{\mathcal{J}}^{(1)}$ ,  $(n_1, \dots, n_{m+1})$  是各段泊松 INGARCH 过程的长度,  $l_j(\hat{\theta}_j)$  为第  $j$  段的似然函数类似于表达式(5)中的  $l(\theta)$ 。给定局部变点估计  $\hat{\mathcal{J}}^{(1)}$ , 可通过(9)进行变点估计

$$(\hat{m}^{(2)}, \hat{\mathcal{J}}^{(2)}) = \arg \min_{m=|\mathcal{J}|, \mathcal{J} \subseteq \hat{\mathcal{J}}^{(1)}} \text{MDL}(m, \mathcal{J}), \quad (9)$$

其中  $\hat{\mathcal{J}}^{(2)} = (\hat{\tau}_1^{(2)}, \hat{\tau}_2^{(2)}, \dots, \hat{\tau}_{\hat{m}^{(2)}}^{(2)})$ ,  $\hat{m}^{(2)} = |\hat{\mathcal{J}}^{(2)}|$ 。

因  $\hat{\mathcal{J}}^{(1)}$  里面包括的因素已很大程度上小于样本容量, 故  $\hat{\mathcal{J}}^{(1)}$  上使用 MDL 能使计算复杂度大幅度降低, 很大程度的提高了计算效率。由于  $h \geq d \log(n)$ , 因此有  $\max_{j=1, \dots, m_0} |\hat{\tau}_j^{(2)}| = O_p(h)$ , 与  $O_p(1)$  相比其并不理想, 而 MDL 方法保证了  $\tau_j^0$  落在区间  $[\hat{\tau}_j - h + 1, \hat{\tau}_j + h]$  的概率接近于 1。那么在第三步中本文将利用这个性质定义变点的最终估计。

第三步: 获得最终估计  $\hat{\mathcal{J}}^{(3)}$ 。

定义第  $j$  个估计变点  $\hat{\tau}_j^{(2)}$  的扫描窗口  $E_j(h)$  和相应的相应观测值  $X_{E_j(h)}$  如下

$$E_j(h) = \{\hat{\tau}_j^{(2)} - 2h + 1, \dots, \hat{\tau}_j^{(2)} + 2h\},$$

$$X_{E_j(h)} = \left( X_{\hat{\tau}_j^{(2)}-2h+1}, \dots, X_{\hat{\tau}_j^{(2)}+2h} \right),$$

令

$$l_j(\tau, \theta_1, \theta_2) = \sum_{t=\hat{\tau}_j^{(2)}-2h+1}^{\tau} R_t(\theta_1) + \sum_{t=\hat{\tau}_j^{(2)}+2h}^{\tau} R_t(\theta_2), \quad (10)$$

其中  $R_t(\theta_1)$  是基于序列  $\{X_{\hat{\tau}_j^{(2)}-2h+1}, X_{\tau}\}$  的条件似然函数, 同理定义  $R_t(\theta_2)$ ,  $j=1, \dots, \hat{m}^{(2)}$ , 定义最终变点估计如下

$$\hat{\tau}_j^{(3)} = \arg \max_{\tau \in [\hat{\tau}_j^{(2)}-h, \hat{\tau}_j^{(2)}+h]} l_j(\tau, \hat{\theta}_j, \hat{\theta}_{j+1}), \quad (11)$$

其中  $\hat{\theta}_j = \hat{\theta}_j(\tau) = \arg \max_{\theta_1} \sum_{t=\hat{\tau}_j^{(2)}-2h+1}^{\tau} R_t(\theta_1)$ ,  $\hat{\theta}_{j+1}$  定义类似。LRSM 中的每一步都需要对相应的 INGARCH 模型参数的估计值。将极大似然估计替换为矩估计, 具体原因如下: 极大似然估计法需要通过数值优化来求解, 故求解过程耗费时间长; 矩估计因其有显式表达式, 求解过程简单, 耗费时间短; 虽然极大似然估计值要优于矩估计值, 但通过模拟研究发现, 矩估计值与极大似然估计值近似, 采用矩估计值几乎不影响 LRSM 的性能。其参数估计详细过程见参考文献[18]。

### 2.3. 对数线性泊松 INGARCH 模型

尽管泊松 INGARCH 模型为计数相关数据的建模提供了一个适当的框架, 但至少有两个与它应用相关的缺点。第一, 当  $0 < \gamma_1 + \delta_1 < 1$ , 其自协方差函数大于 0, 故不能采用泊松 INGARCH 模型对负相关进行建模; 第二, 泊松 INGARCH 只包括导致正回归项的协变量, 否则泊松过程的均值将有可能变为负值。为解决这一问题, Fokianos 和 Tjøstheim [19] 提出的对数线性泊松 INGARCH 模型的自相关函数可正可负。在利用 LRSM 进行变点探测和估计的情况下, 对数线性泊松 INGARCH 模型与泊松 INGARCH 模型在步骤上相似, 但是其中的一些细节需要修正。

对数线性泊松 INGARCH (1, 1) 模型中的式(4)是

$$\begin{cases} Y_{t,j} | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_{t,j}) & t \geq 1, \\ v_{t,j} = \gamma_{j,0} + \gamma_{j,1} \log(Y_{t-1,j} + 1) + \delta_{j,1} v_{t-1,j}, \end{cases} \quad (12)$$

其中参数  $\gamma_{j,0}$ ,  $\gamma_{j,1}$ ,  $\delta_{j,1}$  属于  $\mathbb{R}$ ,  $v_{t,j} \equiv \log \lambda_{t,j}$ ,  $\theta_j = (\gamma_{j,0}, \gamma_{j,1}, \delta_{j,1})$  表示第  $j$  段对数线性泊松 INGARCH (1, 1) 过程的参数向量。

对数线性泊松 INGARCH (1, 1) 模型中(5)是

$$l(\theta) = \ln \sum_{t=1}^n \left[ p(d_t | d_{t-1}, \dots, d_{t-p}) \right], \quad (13)$$

其中  $p(d_t | d_{t-1}, \dots, d_{t-p}) = \frac{e^{-\lambda_t} \lambda_t^{d_t}}{d_t!}$ , 而  $\lambda_t = \exp(v_t)$ , 故有

$$l(\theta) = \sum_{t=1}^n (d_t v_t - \exp(v_t)) \quad (14)$$

对数线性泊松 INGARCH (1, 1) 模型中(8)是

$$\text{MDL}(m, \mathcal{J}) = \log(m) + (m+1) \log(n) + \sum_{j=1}^{m+1} 2 \log(n_j) - \sum_{j=1}^{m+1} l_j(\hat{\theta}_j), \quad (15)$$

其中  $l_j(\hat{\theta}_j)$  的定义类似于表达式(13)中的  $l(\theta)$ 。对数线性泊松 INGARCH (1, 1)模型的参数估计类似于泊松 INGARCH (1, 1)模型, 其详细过程见参考文献[19]。

### 3. 窗口半径的选取

在使用 LRSM 的初始扫描步骤中, 在给定窗口半径大小为  $h$  下, 那么扫描窗口  $W_j(h)$  的大小为  $2h$ ,  $O(h)$  是每一个时刻  $t$  的扫描统计量  $S_h(t)$  的计算复杂度。第二步利用 MDL 时,  $O\left(\left(\hat{m}^{(1)}\right)^2 n\right)$  是所需要的计算复杂度。在第三步中, 因在新的扫描窗口进行计算, 故此时计算复杂度为  $O\left(\hat{m}^{(2)} h^2\right)$ 。由以上所述, 因此最终计算复杂度的总和为  $O(nh + h^2)$ 。 $h$  较小的时候, 如  $h = O\{\log(n)\}$ , 那么完整的 LRSM 的计算复杂度为  $O\{n \log(n)\}$ , 显著低于  $O\{n^2\}$  的数量级。

为了说明窗口半径  $h$  的选择, 在本节中进行了仿真实验。LRSM 中涉及的调优参数包括窗口半径  $h$ 。为探讨  $h$  到底该如何取值, 我们使用下面第 4 节中模型 D 随机生成的数据进行了重复试验分析, 其中  $d$  分别取值(1, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7), 变点位置取为 1250。本章将三步程序应用于长度为  $n = 2024$  的模型 D 重复 100 次进行分析研究, 具有不同的  $h$  值, 其中  $h = d \log(n)^2$  使用不同的  $d$  值。结果汇总在表 1 中。

**Table 1.** Sensitivity analysis for the choice of  $h$

**表 1.** 窗口半径  $h$  的敏感性分析

$d$	$h = d \log(n)^2$	$\% : \hat{m}^{(2)} = m_0$	$\% :  \hat{\tau}_j - \tau_0  < h$
1	57	48	43
2	114	86	80
2.5	142	87	85
3	171	97	95
3.5	199	99	98
4	228	97	96
4.5	256	94	88
5	285	99	97
6	342	99	97
7	399	7	7

$\% : \hat{m}^{(2)} = m_0$  表示正确估计变点数量的百分比,  $\% : |\hat{\tau}_j - \tau_0| < h$  表示有效识别变点的百分比。

从表 1 的结果可以看出: 较小的  $h$  被选择时虽然会减少第一步的计算成本, 即  $O(nh)$ , 然而较小  $h$  的会产生非常大的  $\hat{m}^{(1)}$ 。也就是说, 它生成了一个用于选择变点的候选池, 在第二步中增加了  $O\left(\hat{m}^{(1)}\right)^2$  的计算量和计算时间。与之相反, 较大的  $h$  可能带来计算上的便利但可能违反  $h < n\varepsilon_m/2$ 。尽管如此, 从表 1 可以看出, 对于大多数不同的  $h$  值,  $m$  的估计值是一致的, 表 1 的结果可以说明的是当  $h = 342$  时, 得到的估计结果仍很不错。这种结果的产生可能与(Vostrikova, 1981) [26]的研究有着很大的关联, 也就是当序列存在变点时, LRSM 可探测出每个变点。故最终估计对  $h$  的选择并不敏感, 并不会因为  $h$  取值的稍微增大或缩小影响结果。因此对于泊松 INGARCH 模型这里选择  $d = 3$ 。

### 4. 模拟研究

本节进行了模拟研究, 以评估应用于 INGARCH 模型的 LRSM 的性能, 用统计软件 R 语言编程实现了仿真实验。将 LRSM 应用于泊松分布下的 INGARCH 模型和对数线性泊松下的 INGARCH 模型进行了探索研究, 并在仿真中设置了三种场景: 无变点、一个变点和两个变点。

为了简单度量 LRSM 的估计精度, 作如下定义: 一个方法若能正确的探测出变点的数量, 当  $n = 1024$  或  $n = 2024$  时, 以及每对真实变点与估计变点之间的距离在 50 内, 则认为探测出来的变点是有效的。正

如文献中 Yau 和 Zhao [25]提到的, 本文也选择 50。

#### 4.1. 泊松 INGARCH (1, 1)模型

在本节中, 将对以下模型进行模拟研究, 并对结果进行分析。模型 A-F 的样本路径图见图 2, 模拟结果见表 2。

**Table 2.** Simulation results from Model A to Model F

**表 2.** 模型 A 到模型 F 的模拟结果

模型	$\% : \hat{m}^{(2)} = m_0$	$\% :  \hat{\tau}_j - \tau_0  < h$	变点均值
模型 A $\gamma_0 = 1, \gamma_1 = 0.2, \delta_1 = 0.4$	100	无	-
$\gamma_0 = 1, \gamma_1 = 0.3, \delta_1 = 0.5$	100	无	-
$\gamma_0 = 1, \gamma_1 = 0.4, \delta_1 = 0.1$	100	无	-
模型 B	94	92	514.2
模型 C	85	75	516.7
模型 D	97	95	1252.7
模型 E	96	96	676.7 1250.8
模型 F	81	78	680.3 1252.2

$\% : \hat{m}^{(2)} = m_0$  表示正确估计变点数量的百分比,  $\% : |\hat{\tau}_j - \tau_0| < h$  表示有效识别变点的百分比。

##### 模型 A: 没有变点的平稳 INGARCH (1, 1)模型

模型 A 场景用来测试在没有变点时的检测性能。采用多种参数组合的 INGARCH (1, 1)模型:  $\lambda_t = \gamma_0 + \gamma_1 X_{t-1} + \delta_1 \lambda_{t-1}$  生成观测值, 参数的选取见表 2, 取样本量  $n = 1024$ 。从表 2 可以看出, LRSM 是非常准确和敏感的, 无错误识别变点情况。

##### 模型 B: 分段平稳 INGARCH (1, 1)模型(1 变点)

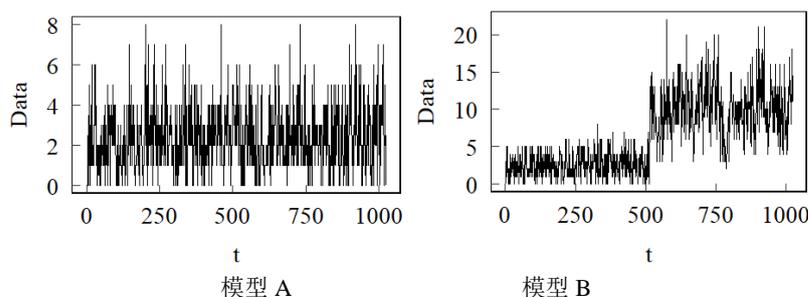
$$\lambda_t = \begin{cases} 1 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 1 \leq t \leq 512, \\ 2 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 513 \leq t \leq 1024. \end{cases}$$

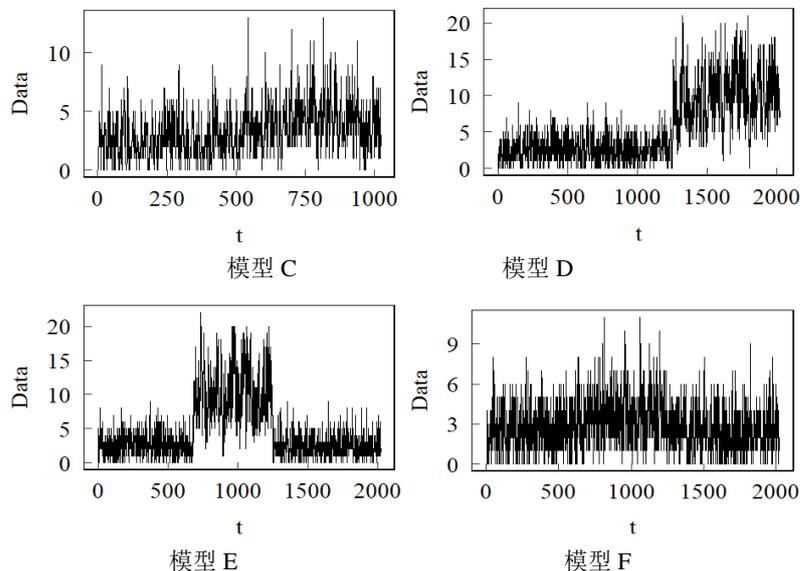
模型 B 将变点的位置设在  $t = 512$ , 取样本量  $n = 1024$ , 用来测试 LRSM 在变点靠近序列中端时的性能。从表 2 可以看出, LRSM 正确识别变点数量的频率高达 94 次, 且满足了估计变点与真实变点之间的距离均小于 50 的次数为 92 次。LRSM 在精确的性质上表现良好。

##### 模型 C: 分段平稳 INGARCH (1, 1)模型(1 变点)

$$\lambda_t = \begin{cases} 1 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 1 \leq t \leq 512, \\ 1.5 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 513 \leq t \leq 1024. \end{cases}$$

模型 C 将变点依然设在  $t = 512$ , 且样本量相同, 与模型 B 不同的就是参数选取的不同, 用来测试 LRSM 在变点前后参数变化不明显时的性能。由表 2 可知, LRSM 的性能也很好。





注: 模型 A 中参数取 1, 0.2, 0.4 进行展示。

Figure 2. Sample path from Model A to Model F

图 2. 模型 A 到模型 F 的样本路径图

**模型 D:** 分段平稳 INGARCH (1, 1)模型(1 变点)

$$\lambda_t = \begin{cases} 1 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 1 \leq t \leq 1250, \\ 2 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 1251 \leq t \leq 2024. \end{cases}$$

模型 D 变点位置在  $t = 1250$ , 样本量不同取  $n = 2024$ , 用来测试随着样本量增加时 LRSM 的性能。由表 2 可知, 随着样本量的增加, LRSM 的性能也很好。

**模型 E:** 分段平稳 INGARCH (1, 1)模型(2 变点)

$$\lambda_t = \begin{cases} 1 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 1 \leq t \leq 674, \\ 2 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 675 \leq t \leq 1250, \\ 1 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 1251 \leq t \leq 2024. \end{cases}$$

模型 E 分别在 674 和 1250 处设有两个变点。正确预测变点数量的频率为 96, 且均满足估计变点与实际变点之间的距离小于 50。因此 LRSM 在精确的性质上表现良好。

**模型 F:** 分段平稳 INGARCH (1, 1)模型(2 变点)

$$\lambda_t = \begin{cases} 1 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 1 \leq t \leq 674, \\ 1.5 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 675 \leq t \leq 1250, \\ 1 + 0.2X_{t-1} + 0.4\lambda_{t-1} & 1251 \leq t \leq 2024. \end{cases}$$

模型 F 分别在 674 和 1251 处设置两个变点, 与模型 E 参数设置不同, 表现出来变点前后样本变化较小, 其他相同正确预测变点数量的频率为 81, 满足估计变点与实际变点之间的距离小于 50 的次数为 78, 虽然比模型 E 稍微差点, 但也能说明在变点前后变化不大时 LRSM 的有效性。因此 LRSM 在精确的性质上表现良好。

## 4.2. 对数线性泊松 INGARCH (1, 1)模型

本节将对给出的以下几种对数线性泊松 INGARCH (1, 1)模型进行模拟研究, 并对结果进行分析。模型

G-L 的样本路径图见图 3, 模拟结果见表 3。

**模型 G:** 无变点的平稳对数线性泊松 INGARCH (1, 1)模型

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t), \lambda_t = e^{v_t},$$

$$v_t = 1 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} \quad 1 \leq t \leq 1024.$$

研究 LRSM 在对数线性泊松 INGARCH (1, 1)模型的变点识别性能。加入几组不同参数组合来测试无变点时性能, 不同参数组合见表 3。表 3 可以看出, LRSM 效果还是很不错的。

**模型 H:** 分段平稳对数线性泊松 INGARCH (1, 1)模型(1 变点)

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t), \lambda_t = e^{v_t},$$

$$v_t = \begin{cases} 1 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 1 \leq t \leq 400, \\ 0.6 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 401 \leq t \leq 1024. \end{cases}$$

模型 H 在 400 处设置一个变点,  $n = 1024$ 。从表 3 可以看出, LRSM 正确估计了变点的数量多达 97 次, 它们都满足估计的变点与实际变点之间的距离小于 50。所以 LRSM 在这个性质上表现的还是很不错的。

**模型 I:** 分段平稳对数线性泊松 INGARCH (1, 1)模型(1 变点)

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t), \lambda_t = e^{v_t},$$

$$v_t = \begin{cases} 1 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 1 \leq t \leq 400, \\ 0.9 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 401 \leq t \leq 1024. \end{cases}$$

模型 I 和模型 H 的样本量和变点设置相同, 但参数设置不同, 模型 H 设置的参数差距较大, 模型 I 设置的参数差距较小。表 3 可以看出, 当变点前后参数变化不是很明显时, LRSM 也能很好地检测到变点。

**模型 J:** 分段平稳对数线性泊松 INGARCH (1, 1)模型(2 变点)

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t), \lambda_t = e^{v_t},$$

$$v_t = \begin{cases} 1 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 1 \leq t \leq 312, \\ 0.6 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 313 \leq t \leq 612, \\ 1 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 613 \leq t \leq 1024. \end{cases}$$

模型 J 在 313 和 613 处设置了两个变点。表 3 可以看出, LRSM 对变点的正确估计次数高达 95 次, 它们都满足估计的变化与实际变点之间的距离小于 50。LRSM 效果还是很不错的。

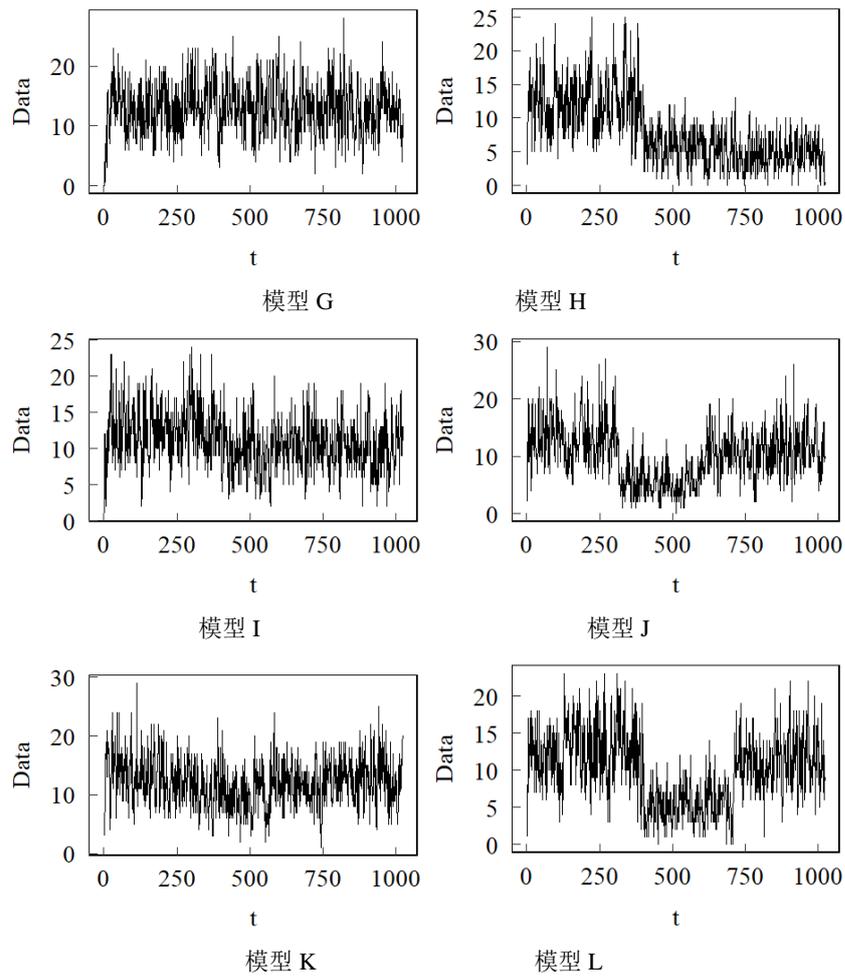
**模型 K:** 分段平稳对数线性泊松 INGARCH (1, 1)模型(2 变点)

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t), \lambda_t = e^{v_t},$$

$$v_t = \begin{cases} 1 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 1 \leq t \leq 312, \\ 0.9 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 313 \leq t \leq 612, \\ 1 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 613 \leq t \leq 1024. \end{cases}$$

模型 J 与模型 K 在样本量相同时设置了两个相同的变点, 但是参数设置不同, 模型 K 在参数上的差距设置很小, 用来测试 LRSM 在变点前后参数变化较小的性能。由表 3 看出, LRSM 效果还是很不错的。

**模型 L:** 分段平稳对数线性泊松 INGARCH (1, 1)模型(2 变点)



注: 模型 G 中参数取 1, 0.4, 0.2 进行展示。

Figure 3. Sample path from Model G to Model L  
图 3. 模型 G 到模型 L 的样本路径图

Table 3. Simulation results from Model G to Model L  
表 3. 模型 G 到模型 L 的模拟结果

模型	$\% : \hat{m}^{(2)} = m_0$	$\% :  \hat{\tau}_j - \tau_0  < h$	变点均值
模型 G $\gamma_0 = 1, \gamma_1 = 0.4, \delta_1 = 0.2$	100	无	-
$\gamma_0 = 1, \gamma_1 = 0.3, \delta_1 = 0.5$	100	无	-
$\gamma_0 = 1, \gamma_1 = 0.2, \delta_1 = 0.4$	100	无	-
模型 H	97	97	398.7
模型 I	90	88	401.6
模型 J	97	95	312.9 613.3
模型 K	80	77	315.4 617.3
模型 L	96	92	399.5 712.9

$\% : \hat{m}^{(2)} = m_0$  表示正确估计变点数量的百分比,  $\% : |\hat{\tau}_j - \tau_0| < h$  表示有效识别变点的百分比。

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t), \lambda_t = e^{v_t},$$

$$v_t = \begin{cases} 1 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 1 \leq t \leq 400, \\ 0.6 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 401 \leq t \leq 712, \\ 1 + 0.4 \log(X_{t-1} + 1) + 0.2v_{t-1} & 713 \leq t \leq 1024. \end{cases}$$

模型 L 与模型 J 变点位置设置不同, 其他均相同, 用来测试变点位置设置不同时 LRSM 的性能。由表 3 看出, LRSM 还是有效的。

## 5. 实证分析

### 5.1. 癫痫发作次数数据

癫痫是一种神经疾病, 由于脑内的神经细胞异常放电, 突然引起脑功能障碍, 出现暂时性的异常行为、感觉和意识障碍、肌肉痉挛等症状。癫痫发作次数数据指的是患者在一段时间内, 通常指的是每天的发作次数。癫痫发作次数数据的收集和记录对于医生和研究人员来说很重要, 因为它可以用来评估患者病情的严重程度、治疗的有效性和疾病的倾向。通过分析癫痫发作次数数据, 医生可以制定个性化的治疗方案, 预测未来的发作风险, 评估药物的有效性和研究癫痫的病理生理。在统计分析中, 癫痫发作次数数据通常被视为时间序列数据, 时间序列分析方法可以应用于研究发作频率的趋势、季节的变化和预测未来的发作。更为重要的是, 变点检测方法也可以用来识别癫痫发作次数的变点, 帮助医生调整治疗方案。

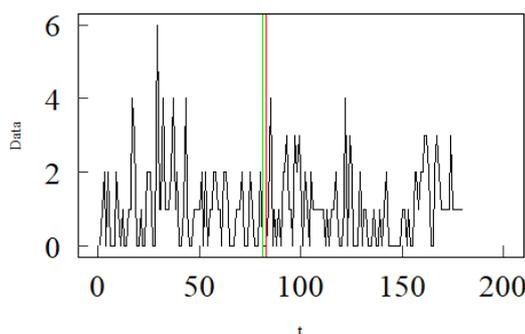


Figure 4. Epileptic seizure counts data  
图 4. 癫痫发作次数数据

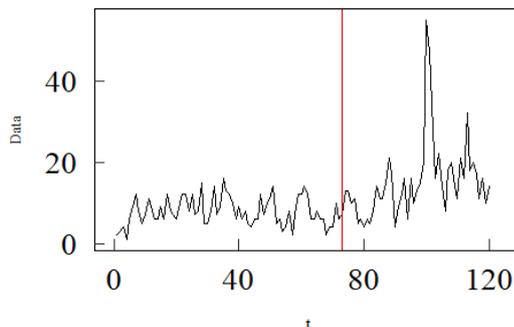
每日癫痫发作次数是评估各种治疗方法有效性的重要信息来源, 尤其是评估抗癫痫药物的使用情况。本节基于 LRSM 利用泊松 INGARCH (1, 1) 模型对癫痫发作次数数据进行变点探测。该数据来自于 Franke 和 Kirch [27], 其样本长度为 180, 样本路径图如图 4 所示。这些数据是一项大型临床研究的一部分, 该研究调查了一种药物对部分癫痫发作的影响。首先癫痫发作次数记录在常规药物下的一段时间, 然后从  $t=80$  时患者接受安慰剂。Franke 和 Kirch [27] 对该数据利用泊松 INGARCH (1, 1) 模型进行拟合并运用累计和方法探测到变点位置在  $t^*=81$ 。本节运用 LRSM 探测到变点位置在  $t^*=83$ 。为此, 对这两个可能的变点分别进行泊松 INGARCH (1, 1) 模型拟合, 并计算了均方误差。当计算得到变点为第 81 天时, 通过拟合的泊松 INGARCH (1, 1) 模型得到预测值的均方误差为 2.284; 变点为第 83 天时, 其均方误差为 2.045, 其均方误差的计算公式为  $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ , 其中  $n$  为样本量,  $y_i$  为实际观测值,  $\hat{y}_i$  为预测值。这明显地表明, 使用变点为第 83 天时的泊松 INGARCH 模型更适合该组数据。参数估计如下表 4。

**Table 4.** Parameter estimation of the Poisson INGARCH (1, 1) model**表 4.** 泊松 INGARCH (1, 1)模型参数估计

	$\gamma_0$	$\gamma_1$	$\delta_1$
[1, 83]	0.757	0.039	0.014
[84, 180]	0.366	0.423	0.012
[1, 180]	0.917	0.125	0.042

## 5.2. 弯曲杆菌数据

查阅相关资料得到, 弯曲杆菌病是由弯曲杆菌属细菌感染引起的急性细菌感染。主要攻击人体的消化系统, 然后引发疾病, 是导致胃肠炎最为常见的细菌。其中大多数患者在感染 2 到 5 天后就出现症状的现象比较常见, 但是也有可能连续 1 天到 1 周, 更有可能 10 天。弯曲杆菌感染一般来说是比较轻微的, 其中腹泻、腹痛、发烧、恶心或呕吐等现象是最为常见的临床症状, 一般持续 3 到 6 天。但是, 弯曲杆菌可能进入血液循环。因此, 对于非常年幼的儿童、老年人、免疫缺陷的人来说是致命的。

**Figure 5.** Campylobacter data**图 5.** 弯曲杆菌数据**Table 5.** Parameter estimation of the Log-linear Poisson INGARCH (1, 1) model**表 5.** 对数线性泊松 INGARCH (1, 1)模型参数估计

	$\gamma_0$	$\gamma_1$	$\delta_1$
[1, 73]	4.598	0.402	0.012
[74, 120]	5.444	0.585	0.057
[1, 120]	3.375	0.644	0.042

本节基于 LRSM 运用对数线性泊松 INGARCH 模型对弯曲杆菌数据[19]进行变点探测。这组数据记录了 1990 年 2 月至 2000 年 10 月期间加拿大魁北克省北部弯曲杆菌感染患者的数量。该地区每 28 天记录一次弯曲杆菌感染病例, 每年进行 13 次观察。选取前 120 个数据, 绘制时间序列图, 如图 5 所示。由图 5 可以看出, 在  $t = 70$  之前, 魁北克省北部弯曲杆菌感染人数保持相对稳定的波动, 但在  $t = 70$  之后, 弯曲杆菌感染人数“激增”并迅速达到最大值。对数线性泊松 INGARCH (1, 1)模型与该数据进行了拟合, 并使用 LRSM 检测变点, 在  $t = 73$  处检测到变点。这里使用赤池信息准则(AIC)和贝叶斯信息准则(BIC)在无变点模型和有变点模型中选择一个最优模型。当变点为 73 时的 AIC 值和 BIC 值分别为 702 和 717, 当无变点时的 AIC 值和 BIC 值分别为 763 和 774。最小信息量表明, 无论采用 AIC 准则还是 BIC 准则, 有变点的对数线性泊松 INGARCH (1, 1)是相对较优的模型。当变点为 73 时的均方误差为 2.282; 无变点

时的均方误差为 4.664, 这也表明了, 变点为第 73 天的对数线性泊松 INGARCH (1, 1) 模型拟合效果更好。其参数估计如上表 5。以上说明了 LRSM 的有效性。

## 6. 结论与展望

本文主要研究了分段平稳的 INGARCH 模型变点探测问题, 运用 LRSM 对变点数目及位置进行探测和估计。最后, 将 LRSM 应用于两个真实的数据集中。通过模拟研究和真实数据分析的结果表明: LRSM 能有效的探测出由 INGARCH 模型生成的分段平稳整数自回归过程中的变点。

## 基金项目

辽宁科技大学博士启动基金(601010391)。

## 参考文献

- [1] Eiauer, P. and Hackl, P. (1978) The Use of MOSUMS for Quality Control. *Technometrics*, **20**, 431-436. <https://doi.org/10.1080/00401706.1978.10489697>
- [2] Oh, K.J. and Han, I. (2000) An Intelligent Clustering Forecasting System Based on Change-Point Detection and Artificial Neural Networks: Application to Financial Economics. *IEEE Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, Maui, 3-6 January 2001, 8.
- [3] Pepelyshev, A. and Polunchenko, A.S. (2015) Real-Time Financial Surveillance via Quickest Change-Point Detection Methods. *Statistics and Its Interface*, **10**, 93-106. <https://doi.org/10.4310/SII.2017.v10.n1.a9>
- [4] Diop, M.L. and Kengne, W. (2022) Epidemic Change-Point Detection in General Causal Time Series. *Statistics & Probability Letters*, **184**, Article ID: 109416. <https://doi.org/10.1016/j.spl.2022.109416>
- [5] Picard, D. (1985) Testing and Estimating Change-Points in Time Series. *Advances in Applied Probability*, **17**, 841-867. <https://doi.org/10.2307/1427090>
- [6] Bai, J. and Perron, P. (1998) Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, **66**, 47-78. <https://doi.org/10.2307/2998540>
- [7] Kokozzka, P. and Leipus, R. (2000) Change-Point Estimation in ARCH Models. *Bernoulli*, **6**, 513-539. <https://doi.org/10.2307/3318673>
- [8] Liu, Z. and Qian, L. (2009) Changepoint Estimation in a Segmented Linear Regression via Empirical Likelihood. *Communications in Statistics-Simulation and Computation*, **39**, 85-100. <https://doi.org/10.1080/03610910903312193>
- [9] Foygel, R. and Drton, M. (2012) Exact Block-Wise Optimization in Group Lasso and Sparse Group Lasso for Linear Regression.
- [10] Hudecov, Á.Š., Huš Kov, Á.M. and Meintanis, S. (2014) Change Detection in INARCH Time Series of Counts. In: Cao, R., Manteiga, W.G. and Romo, J., Eds., *Nonparametric Statistics*, Springer, Berlin, 47-58. [https://doi.org/10.1007/978-3-319-41582-6\\_4](https://doi.org/10.1007/978-3-319-41582-6_4)
- [11] Chen, C.W.S. and Lee, S. (2016) Generalized Poisson Autoregressive Models for Time Series of Counts. *Computational Statistics & Data Analysis*, **99**, 51-67. <https://doi.org/10.1016/j.csda.2016.01.009>
- [12] Chen, C.W.S. and Lee, S. (2017) Bayesian Causality Test for Integer-Valued Time Series Models with Applications to Climate and Crime Data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **66**, 797-814. <https://doi.org/10.1111/rssc.12200>
- [13] Pein, F., Sieling, H. and Munk, A. (2017) Heterogeneous Change Point Inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **79**, 1207-1227. <https://doi.org/10.1111/rssb.12202>
- [14] Lee, S. and Moon, M. (2020) Hybrid Change Point Detection for Time Series via Support Vector Regression and CUSUM Method. *Applied Soft Computing*, **89**, Article ID: 106101. <https://doi.org/10.1016/j.asoc.2020.106101>
- [15] Lee, S. and Kim, B. (2021) Recent Progress in Parameter Change Test for Integer-Valued Time Series Models. *Journal of the Korean Statistical Society*, **50**, 730-755. <https://doi.org/10.1007/s42952-020-00102-4>
- [16] Franke, J. and Seligmann, T.H. (1993) Conditional Maximum Likelihood Estimates for INAR (1) Processes and Their Application to Modelling Epileptic Seizure Counts. In: *Developments in Time Series Analysis*, Springer, Berlin, 310-330.
- [17] Gauthier, G. and Latour, A. (1994) Convergence forte des estimateurs des paramètres d'un processus genar (P). *Annales des Sciences Mathématiques du Québec*, **18**, 49-71.

- 
- [18] Ferland, R., Latour, A. and Oraichi, D. (2006) Integer-Valued GARCH Process. *Journal of Time Series Analysis*, **27**, 923-942. <https://doi.org/10.1111/j.1467-9892.2006.00496.x>
- [19] Fokianos, K. and Tjøstheim, D. (2011) Log-Linear Poisson Autoregression. *Journal of Multivariate Analysis*, **102**, 563-578. <https://doi.org/10.1016/j.jmva.2010.11.002>
- [20] Bauwens, L., De Backer, B. and Dufays, A. (2014) A Bayesian Method of Change-Point Estimation with Recurrent Regimes: Application to GARCH Models. *Journal of Empirical Finance*, **29**, 207-229. <https://doi.org/10.1016/j.jempfin.2014.06.008>
- [21] Barry, D. and Hartigan, J.A. (1993) A Bayesian Analysis for Change Point Problems. *Journal of the American Statistical Association*, **88**, 309-319. <https://doi.org/10.1080/01621459.1993.10594323>
- [22] Bai, J. (1995) Estimating Multiple Breaks One at a Time. Massachusetts Institute of Technology (MIT), Cambridge.
- [23] Davis, R.A., Lee, T.C. and Rodriguez-Yam, G.A. (2005) Structural Breaks Estimation for Non-Stationary Time Series Signals. *IEEE/SP 13th Workshop on Statistical Signal Processing*, Bordeaux, 17-20 July 2005, 233-238. <https://doi.org/10.1109/SSP.2005.1628598>
- [24] Killick, R., Fearnhead, P. and Eckley, I.A. (2012) Optimal Detection of Changepoints with a Linear Computational Cost. *Journal of the American Statistical Association*, **107**, 1590-1598. <https://doi.org/10.1080/01621459.2012.737745>
- [25] Yau, C.Y. and Zhao, Z. (2016) Inference for Multiple Change Points in Time Series via Likelihood Ratio Scan Statistics. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **78**, 895-916. <https://doi.org/10.1111/rssb.12139>
- [26] Vostrikova, L.Y. (1981) Detecting "Disorder" in Multidimensional Random Processes. *Doklady Akademii Nauk. Russian Academy of Sciences*, **259**, 270-274.
- [27] Franke, J., Kirch, C. and Kamgaing, J.T. (2012) Changepoints in Times Series of Counts. *Journal of Time*, **33**, 757-770. <https://doi.org/10.1111/j.1467-9892.2011.00778.x>