

分段平稳自回归过程的变点估计与模型选择

——基于改进的自适应 LASSO 方法

刘杰, 陈啸远, 吴遵

(中国科学技术大学管理学院国际金融研究院, 安徽合肥 230601)

摘要: 针对非平稳时间序列中一类分段平稳自回归(PSAR)过程的变点估计和模型选择问题,在已有的将变点估计问题转化成变量选择问题方法的基础上,提出一种基于两阶段 LASSO(TS-LASSO)算法同时进行变点估计和模型选择.具体地,在第一阶段中,通过 LASSO 算法对序列中的变点和模型进行初步的估计和选择,然后在第二阶段中结合改进的自适应 LASSO 算法对过估计的 LASSO 结果进行筛选,最终实现变点的一致估计和模型的准确选择.并对变点估计结果的大样本性质进行了分析.此外,对于特殊情形下的均值变化序列和无变点序列,TS-LASSO 算法也能实现有效的估计和识别.最后,结合不同类型序列的模拟检验以及地震波数据的实例分析,证明 TS-LASSO 算法是有效的,并具有一定的实用意义.

关键词: 分段平稳自回归; 变点估计; 模型选择; 自适应 LASSO

中图分类号: O213 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2020.06.005

2010 Mathematics Subject Classification: 60G10;60C05;62J05

引用格式: 刘杰,陈啸远,吴遵. 分段平稳自回归过程的变点估计与模型选择——基于改进的自适应 LASSO 方法[J]. 中国科学技术大学学报,2020,50(6):744-751,776.

LIU Jie, CHEN Xiaoyuan, WU Zun. Change-points estimation and model selection for piecewise stationary autoregressive processes based on modified adaptive LASSO method [J]. Journal of University of Science and Technology of China, 2020,50(6):744-751,776.

Change-points estimation and model selection for piecewise stationary autoregressive processes based on modified adaptive LASSO method

LIU Jie, CHEN Xiaoyuan, WU Zun

(International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230601, China)

Abstract: Considering the problems of change-points estimation and model selection for nonstationary time series such as piecewise stationary autoregressive (PSAR) processes, a method which can simultaneously conduct change-point estimation and model selection with a two-stage LASSO (TS-LASSO) algorithm based on the existing method to transform the problem of change-points estimation into the problem of variable selection was proposed. Specifically, in the first stage, the preliminary estimation of change-points and the selection of the models can be derived by LASSO algorithm. Then, in the second stage, a modified adaptive LASSO algorithm was used to screen the overestimated results, so the consistent estimation could be obtained and the accurate model could be selected. The large sample properties of the results for the variable-point estimation. In addition, the TS-LASSO algorithm can also achieve the estimation and recognition for the mean change-points sequences and no change-points sequences in special cases effectively. Finally, combined with the test of different type of simulative sequences and the case study of a seismic wave data, it was shown that TS-LASSO algorithm is effective and has certain practicability.

Key words: piecewise stationary autoregressive; change-points estimation; model selection; adaptive LASSO

0 引言

近几十年来,有关变点问题的研究一直受到众

多学者的关注.起初,这些研究主要集中于序列中的单变点检测.早在1954年,Page^[1]通过构造的CUSUM检验统计量,有效估计了独立同分布序列

收稿日期:2019-12-06;修回日期:2020-04-20

基金项目:国家自然科学基金(71771201,71874171,71731010,71631006,71991464)资助.

作者简介:刘杰(通讯作者),男,1981年生,博士/副教授.研究方向:随机网络和时间序列. E-mail:jiel@ustc.edu.cn

中水平均值的变化位置. 具体到模型参数变点的研究中, Picard^[2] 最先得到了分段自回归模型中参数变点的极大似然估计. 类似的, Takanami 等^[3] 基于局部平稳自回归模型的假设, 结合 AIC 准则, 有效地识别出地震波的初至时刻. 在实际情况中, 由于多变点问题更加普遍, 随后便出现了许多有关多变点估计问题的研究. 一类称之为“二分法”的多变点估计方法, 其原理是对分割到的子序列重复使用单变点估计方法. 这是一种简单有效的多变点估计方法, 但得到的变点估计往往不是全局最优的. 对于全局最优的变点估计方法, 它们更多地依赖于复杂的算法, 如 Davis 等^[4] 基于最小描述长度 (minimum description length, MDL) 准则, 采用遗传算法比较不同模型对数据的拟合效果, 从而选出最佳变点数量、变点位置和自回归阶数对应的分段平稳自回归模型. Yau 等^[5] 提出的似然比扫描方法, 同样解决了非平稳时间序列中的多个自回归参数变点的估计. 吴楠等^[6] 则采用一种两阶段估计方法, 先通过似然比扫描统计量初步确定变点的大体位置, 然后基于 MDL 准则得到变点的一致估计.

复杂算法自然不利于变点的快速有效估计, 为解决非平稳时间序列中的多变点估计问题, 一种将变点估计问题转化成变量选择问题的方法得到众多学者的关注. 该转换方法的原理是通过特定的设计矩阵, 将相邻平稳序列之间的参数以增量的形式表示, 从而转换成一般的回归变量选择问题, 进而借鉴变量选择的方法以实现多变点同时估计的目标. 对应的非零回归系数意味着相邻平稳序列的参数发生了改变, 即变点产生. 具体地, Harchaoui 等^[7] 提出一种 fused LASSO 多均值变点同时估计的方法, 有效解决了独立序列中水平均值变化对应的变点估计问题. 对于独立观测到的样本 (包括响应变量和解释变量), Jin 等^[8] 提出了一种基于自适应 LASSO、SCAD 和 MCP 惩罚的两阶段变点估计方法, 实现了分段线性回归模型中的多变点估计目标. 然而在时间序列中, 序列前后之间有时存在着某种依赖关系, 属于非独立序列. 具体到分段平稳自回归序列中, 采用同样的问题转换方法, Chan 等^[9] 基于 group LASSO 惩罚的变量选择方法, 实现同时多变点估计的目标.

在非平稳时间序列中, 对各平稳子序列的模型选择, 一般是在变点估计后进行的. 而本文试图在变点估计的同时, 考虑各平稳自回归模型的变量选择, 以实现同时变点估计和模型选择的目标. 考虑到各平稳序列间相同自回归变量的假设, 文献^[9] 采用的 group LASSO 方法对组内的变量无法起到稀疏的作用, 因而它需要在变点估计后再进行各模型的选择. 本文基于单变量惩罚提出的两阶段 LASSO (two-stage LASSO, TS-LASSO) 方法可同时进行变点估计和模型选择. 此外, Jin 等^[8] 研究的分段线性回归模型中的变点估计, 其样本之间的假设是独立的. 而本文研究的分段自回归序列属于非独立序列, 变量选择方法对应的传统渐近理论在这里失效, 需要建立新的渐近理论.

本文旨在从变量选择的角度出发, 针对非平稳时间序列中的分段平稳自回归模型, 提出一种能够同时解决变点估计和模型选择问题的方法. 具体地, 基于特殊设计矩阵得到的线性回归模型, 采用 LASSO 算法对高维回归变量进行稀疏, 以实现初步的变点估计和模型选择. 然后, 基于 LASSO 算法得到的初步估计, 采用自适应 LASSO 算法, 剔除多余的变点, 从而实现变点的一致估计. 此外, 在选择自适应惩罚权重时, 采用了一种改进的自适应惩罚函数, 以适应时间序列可能存在的对时间先后顺序的敏感性^[10].

1 模型假设及问题确定

本节将具体介绍分段平稳自回归 (piecewise stationary autoregressive, PSAR) 模型以及相关的符号说明, 然后确定本文所要研究的具体问题.

1.1 分段平稳自回归过程

假设一组长度为 $n + p$ 的观测序列 $y_{-(p-1):n} = \{y_{-(p-1)}, \dots, y_{-1}, y_0, y_1, \dots, y_n\}$ 满足如式 (1) 所示的 PSAR 过程.

$$y_t = \begin{cases} \beta_0^{(1)} + \beta_1^{(1)} y_{t-1} + \dots + \beta_{p_1}^{(1)} y_{t-p_1} + \varepsilon_t, & \tau_0 \leq t < \tau_1; \\ \beta_0^{(2)} + \beta_1^{(2)} y_{t-1} + \dots + \beta_{p_2}^{(2)} y_{t-p_2} + \varepsilon_t, & \tau_1 \leq t < \tau_2; \\ \vdots \\ \beta_0^{(m+1)} + \beta_1^{(m+1)} y_{t-1} + \dots + \beta_{p_{m+1}}^{(m+1)} y_{t-p_{m+1}} + \varepsilon_t, & \tau_m \leq t < \tau_{m+1} \end{cases} \quad (1)$$

式中, $m, \Omega = \{\tau_1, \dots, \tau_m\}$ 分别表示序列 $y_{-(p-1):n}$ 中的变点数量和变点位置, 且满足 $1 = \tau_0 < \tau_1 < \dots < \tau_m < \tau_{m+1} = n + 1$. 通常, 假定 PSAR 模型中的变点数量 m 、变点位置 Ω 以及各自回归模型 (具体的模型变量和阶数) 都是未知的. 整个序列被 m 个变点分割成 $m + 1$ 段平稳序列, 其中第 j 段平稳序列的自回归系数是 $(\beta_0^{(j)}, \beta_1^{(j)}, \dots, \beta_{p_j}^{(j)})$, p_j 是模型的阶数, $j = 1, \dots, m + 1$. $\{\varepsilon_{-(p-1)}, \dots, \varepsilon_n\}$ 表示零均值单位方差的随机误差向量.

可以注意到, 当模型 (1) 中所有自回归模型的阶数都为零时, 各段序列的差异只体现在常数项上, 此时 PSAR 模型退化成一般的均值变化模型.

1.2 同时变点估计和模型选择

从上述的模型假设可以看出, 序列 $y_{-(p-1):n}$ 中的变点数量、各变点的位置以及各模型的回归变量和阶数是确定具体 PSAR 模型的关键. 本文的主要目标是寻找能够拟合时间序列 $y_{-(p-1):n}$ 的最佳 PSAR 模型, 也就是解决该模型中的变点估计和模型选择问题. 已有的大部分研究主要关注模型中的变点估计问题, 至于各自回归模型的选择是在变点估计后进行的. 本文关注 PSAR 模型中的关键因素, 尝试同时解决序列中的变点估计和模型选择两方面问题. 接下来, 我们将具体介绍一种两阶段估计方法, 以实现同时变点估计和模型选择的目标.

2 方法介绍

在 PSAR 模型中,各自回归模型的阶数可能是不一样的,我们不妨都先假设成 p 阶自回归模型,且 p 是已知的.在实际操作中,可取一个较大的 p 值,以满足 $p \geq \max\{p_1, \dots, p_{m+1}\}$.通过构造恰当的设计矩阵,我们将 PSAR 模型转化成一般的线性回归模型.回归系数对应着相邻自回归模型参数的增量,当相邻自回归模型参数发生变化时,增量非零,意味着变点出现.反之,当相邻自回归模型参数相同时,对应的增量为零,即没有变点出现.于是,回归模型参数中非零元素的位置便对应着 PSAR

$$X = \begin{pmatrix} 1 & y_0 & \cdots & y_{-(p-1)} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 1 & y_1 & \cdots & y_{-(p-2)} & 1 & y_1 & \cdots & y_{-(p-2)} & \cdots & 0 & 0 & \cdots & 0 \\ & & \cdots & & & & \cdots & & \cdots & & & \cdots & \\ 1 & y_{n-2} & \cdots & y_{n-(p+1)} & 1 & y_{n-2} & \cdots & y_{n-(p+1)} & \cdots & 0 & 0 & \cdots & 0 \\ 1 & y_{n-1} & \cdots & y_{n-p} & 1 & y_{n-1} & \cdots & y_{n-p} & \cdots & 1 & y_{n-1} & \cdots & y_{n-p} \end{pmatrix} \quad (2)$$

记 n 维响应向量 $Y^T = (y_1, \dots, y_n)$, 自回归阶数 $p \geq \max\{p_1, \dots, p_{m+1}\}$, n 维随机误差向量 $\varepsilon^T = (\varepsilon_1, \dots, \varepsilon_n)$. $n(p+1)$ 维回归系数 $\theta^{*T} = (\theta_{1,0}^*, \theta_{1,1}^*, \dots, \theta_{n,p}^*)$, 并满足如下表达式:

$$\left. \begin{aligned} & \left\{ \begin{array}{l} \theta_{1,0}^* = \beta_0^{(1)} \\ \cdots \\ \theta_{1,p}^* = \beta_p^{(1)} \end{array} \right\}; i = \tau_j \left\{ \begin{array}{l} \theta_{i,0}^* = \beta_0^{(j+1)} - \beta_0^{(j)} \\ \cdots \\ \theta_{i,p}^* = \beta_p^{(j+1)} - \beta_p^{(j)} \end{array} \right\}; \\ & i \neq \tau_j \left\{ \begin{array}{l} \theta_{i,0}^* = 0 \\ \cdots \\ \theta_{i,p}^* = 0 \end{array} \right\} \end{aligned} \right\} \quad (3)$$

式中, $i=2, \dots, n; j=1, \dots, m$. 此外,当第 j 段子序列的真实阶数小于 p 时,我们令 $\beta_j^{(j)} = 0$, $p_j < l \leq p$. 于是,向量 θ^* 对应着 PSAR 模型中相邻自回归参数间的增量.最后,基于上述假设,我们可以得到线性回归方程

$$Y = X\theta^* + \varepsilon \quad (4)$$

记集合 $\mathcal{A} = \{(i-1)(p+1) + j \mid \theta_{ij}^* \neq 0, 1 \leq i \leq n, 0 \leq j \leq p\}$, 它表示向量 θ^* 中非零元素的位置.从式(3)可以发现,式(4)中的回归参数向量 θ^* 是稀疏的.其原因有两方面,一方面,只有变点位置对应的回归系数不为零,其余非变点位置的回归系数均为零;另一方面,即使在变点位置,由于真实自回归模型的阶数可能小于所假设的阶数 p , 相邻自回归模型的差异可能还表现在某些回归变量上,从而导致变点位置的回归系数 $(\theta_{\tau_j,0}^*, \theta_{\tau_j,1}^*, \dots, \theta_{\tau_j,p}^*)$, $j \in \{1, \dots, m+1\}$, 中也存在零元素.

对方程(4)中稀疏解 θ^* 的求解,常采用最小绝对收缩与选择算子^[11] (LASSO) 变量选择算法.具体地,LASSO 通过对回归变量的绝对值之和进行适当的约束,从而将某些变量的系数压缩至零,以实现变量选择的目的.这是一种可以同时进行变量选择和参数估计的有效方法.式(4)对应的估计为

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{n} \|Y - X\theta\|_2^2 + \lambda_n \sum_{i=1}^n \sum_{j=0}^p |\theta_{ij}| \quad (5)$$

模型中的变点位置.通常变点数量小于序列长度,意味着回归模型参数中存在零元素,需要确定其中的非零元素.这样,PSAR 模型中的变点估计和模型选择问题便转化成了回归模型中的变量选择问题.

关于稀疏解的估计方法,统计学中常采用变量选择算法.于是,本文将结合具体的 LASSO 类变量选择算法,提出一种 TS-LASSO 同时变点估计和模型选择方法,其原理是基于第一阶段估计结果的先验信息,使得第二阶段的估计结果更加精确.

2.1 第一阶段:基于 LASSO 算法的初步变点估计

为了将 PSAR 模型中的变点估计问题转换成变量选择问题^[9],首先构造设计矩阵 X :

式中, λ_n 是惩罚权重,其值越大则 $\hat{\theta}$ 越稀疏.对估计 $\hat{\theta}$ 中的非零元素,记集合 $\hat{\mathcal{A}}$ 为

$$\hat{\mathcal{A}} = \{(i-1)(p+1) + j \mid \hat{\theta}_{ij} \neq 0, 1 \leq i \leq n, 0 \leq j \leq p\}.$$

于是,所要估计的变点位置为 $\hat{\Omega} = \{\hat{\tau}_1, \dots, \hat{\tau}_{|\hat{\mathcal{A}}|}\}$, 变点数量为 $\hat{m} = \lceil |\hat{\mathcal{A}}| / (p+1) \rceil$.

2.2 第二阶段:基于改进自适应 LASSO 方法的一致变点估计

在实际估计过程中,第一阶段的结果往往存在过估计现象,即 $\hat{\Omega}$ 中包含有非变点估计.其原因是在某些情况下 LASSO 的变量选择结果是不一致的.Zhao 等^[12] 指出,LASSO 的变量选择结果满足一致性的前提条件是:在回归模型(4)中,无关变量 (θ^* 中为零的变量)对应的 X 列向量不可由真实变量 (θ^* 中不为零的变量)表示,即所谓的“不可表示性条件”.否则,LASSO 算法在变量选择的过程中便会将与真实变量相关性很强的变量纳入模型中,以弥补对真实变量的过分惩罚.在自回归序列中,由于数据前后间有着既定的线性关系,这使得设计矩阵 X 违背了“不可表示性条件”,从而导致估计结果中包含有非变点估计,即存在过估计现象.

此外,Zou^[13] 也指出 LASSO 的变量选择结果并不总是一致的,前提是要满足一定的必要条件,并指出该条件类似于文献^[12] 提出的“不可表示性条件”.为解决非一致的变量选择结果,Zou 提出了一种改进版的 LASSO 算法——自适应 LASSO 算法.其原理是对不同的变量采用不同的惩罚权重,并得到自适应 LASSO 估计具有的 Oracle 性质^[14]: (i)能确定出正确的变量子集;(ii)有最优的估计率.

于是,结合第一阶段的估计结果,我们接下来采用自适应 LASSO 算法对过估计的变点进行筛选,以剔除多余的变点.对应的自适应 LASSO 估计为

$$\hat{\theta}^* = \arg \min_{\theta} \frac{1}{n} \|Y - X\theta\|_2^2 + \lambda_n \sum_{i=1}^n \sum_{j=0}^p \hat{w}_{ij} |\theta_{ij}|.$$

式中, λ_n 为共同的惩罚权重, \hat{w}_{ij} 为各变量自适应的惩罚权重. 对于惩罚权重 \hat{w}_{ij} 的确定, 常依据最小二乘估计或 LASSO 估计进行确定.

此外, 在时间序列模型中, 通常滞后阶数越靠后的变量对未来的预测能力越弱, 这是时间序列所特有的时间敏感性. 王国长等^[10]在选择自回归模型时考虑了该规律, 从而提升对金融时间序列的预测精度. 由于本文的研究对象同样是时间序列, 于是在惩罚权重 \hat{w}_{ij} 中引入关于滞后阶数的惩罚, 采用文献^[10]中同样形式的自适应惩罚权重

$$\hat{w}_{ij} = l_j^{\gamma_1} / |\hat{\theta}_{ij}|^{\gamma_2}, \gamma_1 > 0, \gamma_2 > 0.$$

其中, γ_1 和 γ_2 为惩罚参数, $l_j (l_0 = 1; l_j = j, j = 1, \dots, p)$ 为滞后阶数. 将 \hat{w}_{ij} 带入式(6), 得到如下改进的自适应 LASSO 估计 $\hat{\theta}^*$:

$$\hat{\theta}^* = \arg \min_{\theta} \frac{1}{n} \|Y - X\theta\|_2^2 + \lambda_n \sum_{i=1}^n \sum_{j=0}^p \frac{l_j^{\gamma_1}}{|\hat{\theta}_{ij}|^{\gamma_2}} |\theta_{ij}| \quad (7)$$

式(7)仍属于自适应 LASSO 估计问题, 是一个凸优化问题, 存在全局最优解. 令向量 $\alpha = A\theta$, 其中 A 是满足如下形式的对角矩阵:

$$A = \text{diag}\left(\frac{1^{\gamma_1}}{|\hat{\theta}_{1,0}|^{\gamma_2}}, \frac{1^{\gamma_1}}{|\hat{\theta}_{1,1}|^{\gamma_2}}, \dots, \frac{p^{\gamma_1}}{|\hat{\theta}_{1,p}|^{\gamma_2}}, \dots, \frac{p^{\gamma_1}}{|\hat{\theta}_{n,p}|^{\gamma_2}}\right).$$

将 $\theta = A^{-1}\alpha$ 带入式(7), 并令新的设计矩阵 $X^* = XA^{-1}$. 于是, 通过简单的变量替换便可得到如下一般的 LASSO 估计为

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{n} \|Y - X^*\alpha\|_2^2 + \lambda_n \sum_{i=1}^n \sum_{j=0}^p |\alpha_{ij}| \quad (8)$$

进而有 $\hat{\theta}^* = A^{-1}\hat{\alpha}$. 同样, 记集合 $\hat{\mathcal{A}}^*$ 为

$$\hat{\mathcal{A}}^* = \{(i-1)(p+1) + j \mid \hat{\theta}_{ij}^* \neq 0, 1 \leq i \leq n, 0 \leq j \leq p\}.$$

它表示估计向量中非零元素的位置. 于是便可得到最终的变点估计为

$$\hat{\Omega}^* = \{\hat{\tau}_1^*, \dots, \hat{\tau}_{|\hat{\mathcal{A}}^*|}^*\} \quad (9)$$

对应的变点数量为 $\hat{m}^* = \lceil |\hat{\mathcal{A}}^*| / (p+1) \rceil$. 结合式(3), 可以得到各自回归模型对应的参数估计为

$$\left. \begin{cases} \hat{\beta}_0^{(1)} = \hat{\theta}_{1,0}^* \\ \dots \\ \hat{\beta}_p^{(1)} = \hat{\theta}_{1,p}^* \\ \dots \\ \hat{\beta}_0^{(j)} = \sum_{i=1}^{\hat{\tau}_{j-1}^*} \hat{\theta}_{i,0}^* \\ \dots \\ \hat{\beta}_p^{(j)} = \sum_{i=1}^{\hat{\tau}_{j-1}^*} \hat{\theta}_{i,p}^* \end{cases} \right\} \quad (10)$$

为方便表示最终各平稳自回归模型的选择结果, 构造如下形式的模型选择矩阵: 记 $p+1$ 维向量 v_j 只含 0 或 1 元素, 它代表 TS-LASSO 方法对 $y_{-(p-1):n}$ 中第 j 段平稳序列的模型选择结果. 在 v_j 中, 元素 1 表示该位置对应的自回归变量被选中, 元素 0 则表示该位置对应的自回归变量未被选中. 于是, TS-LASSO 方法对所有 $\hat{m}^* + 1$ 段平稳序列的模型选择结果可表示为 $(\hat{m}^* + 1) \times p$ 维矩阵 \hat{V} ,

$$\hat{V} = (v_1, \dots, v_{\hat{m}^*+1})^T \quad (11)$$

特别地, 当矩阵 \hat{V} 中第一列元素都为 1, 其余列元素都为零时, TSAR 模型退化成一般的均值变化模型, 此时 TS-LASSO 方法与 Harchaoui 等^[7]提出的均值变点估计方法是一样的.

2.3 渐近性质

为研究 TS-LASSO 方法估计结果的渐近性质, 本文提出以下几条合理的假设:

假设 2.1 $\{\epsilon_t : -(p-1) \leq t \leq n\}$ 为零均值单位方差的白噪声序列; 对任意 $0 < \eta < 1$ 有 $E\epsilon_t^{2+\eta} < \infty$; 对任意的 $i < t$, 有 $\text{cov}(\epsilon_t, y_i) = 0$.

假设 2.2 当 $n \rightarrow \infty$ 时, 有 $\lambda_n \rightarrow 0$ 和 $\sqrt{n}\lambda_n \rightarrow \infty$.

假设 2.3 (平稳性) 对任意实数 $|h| < 1$, 有 $1 - \sum_{i=1}^p \beta_i^j h^i \neq 0, j = 1, \dots, m+1$.

假设 2.4 $\min\{\tau_i/n - \tau_{i-1}/n : i = 1, \dots, m+1\} > \delta > 0$, 其中 δ 为常数.

假设 2.5 $X_{\mathcal{A}^*}^T X_{\mathcal{A}^*} / n \rightarrow_p \Gamma_{\mathcal{A}^*} > 0$.

其中, 假设 2.1 对序列中的噪声做出了常规的约束, 并假设当前时刻的噪声 ϵ_t 与过去样本之间是独立的, 即 $\text{cov}(\epsilon_t, y_i) = 0$. 假设 2.2 则对(自适应) LASSO 估计中的正则参数 λ_n 做出了规范, 此外由 $\lambda_n \rightarrow 0$ 和 $\sqrt{n}\lambda_n \rightarrow \infty$, 可以直接得到 $\lambda_n / \sqrt{n} \rightarrow 0$ (用于后续定理 2.1 的证明). 假设 2.3 保证了各自回归序列的平稳性. 假设 2.4 和假设 2.5 则分别对变点间距和设计矩阵做了约束, 其中 $\Gamma_{\mathcal{A}^*}$ 是一个正定矩阵.

接下来, 先提出以下 2 个引理:

引理 2.1 定义 $n(p+1)$ 维向量 $\check{\theta}$, 使其满足 $\check{\theta}_{\mathcal{A}^*} = (X_{\mathcal{A}^*}^T X_{\mathcal{A}^*})^{-1} X_{\mathcal{A}^*}^T Y$ 以及 $\check{\theta}_{\mathcal{A}^*c} = 0$, 在假设 2.1、2.2 的前提下, 当 $n \rightarrow \infty$ 时, 有 $\hat{\theta} - \check{\theta} \rightarrow_p 0$.

证明 若 $\check{\theta}$ 是式(5)的一个解, 由 Karush-Kuhn-Tucker(KKT)最优条件可知, 对任意的 $1 \leq i \leq n, 0 \leq j \leq p$, 有

$$|X_{ij}^T (Y - X\check{\theta})| \leq \frac{1}{2} n \lambda_n \quad (12)$$

式中, X_{ij} 为矩阵 X 的第 $(i-1)(p+1) + j$ 列. 于是, 接下来只需要证式(12)依概率成立即可.

记 $P_n = X_{\mathcal{A}^*} (X_{\mathcal{A}^*}^T X_{\mathcal{A}^*})^{-1} X_{\mathcal{A}^*}^T$, 于是

$$|X_{ij}^T (Y - X\check{\theta})| =$$

$$|X_{ij}^T (Y - X_{\mathcal{A}^*} \check{\theta}_{\mathcal{A}^*})| = |X_{ij}^T (I_n - P_n) \epsilon|.$$

因为 $\text{tr}(P_n) \leq (m+1)(p+1)$ 为一常数, 所以只需

要证

$$P\left\{\max_{1 \leq i \leq n, 0 \leq j \leq p} |X_{ij}^T \epsilon| \leq \frac{1}{2} n \lambda_n\right\} \rightarrow 1 \quad (13)$$

成立即可.

首先,对任意常数 $b > 0$ 以及 $1 \leq k \leq n, 0 \leq j \leq p$, 由 Markov 不等式我们可以得到

$$\begin{aligned} P\{|y_k^j \epsilon_k| > x\} &\leq \\ P\{\exp\{b |y_k^j \epsilon_k|\} > \exp\{bx\}\} &\leq \\ \frac{E \exp\{b |y_k^j \epsilon_k|\}}{\exp\{bx\}} &\leq a \exp\{-bx\} \end{aligned} \quad (14)$$

式中, $a = \max_{1 \leq k \leq n, 0 \leq j \leq p} E \exp\{b |y_k^j \epsilon_k|\}$, $j=0$ 时 $y_k^j = 1$, $j=1, \dots, p$ 时 $y_k^j = y_{k-j}$. 于是,在假设 2.1~2.3 的前提下,当 $n \rightarrow \infty$ 时有

$$\begin{aligned} P\left\{\max_{1 \leq i \leq n, 0 \leq j \leq p} |X_{ij}^T \epsilon| \leq \frac{1}{2} n \lambda_n\right\} &= \\ P\left\{\max_{1 \leq i \leq n, 0 \leq j \leq p} \left|\sum_{k=i}^n y_k^j \epsilon_k\right| \leq \frac{1}{2} n \lambda_n\right\} &\leq \\ \sum_{i=1}^n \sum_{j=0}^p \sum_{k=i}^n \{P\{|y_k^j \epsilon_k| > \frac{1}{2} n \lambda_n\}\} &\leq \\ a(p+1)n^2 \exp\{-\frac{b}{2} n \lambda_n\} &\rightarrow 0 \end{aligned} \quad (15)$$

所以,当 $n \rightarrow \infty$ 时式(13)成立,证毕.

引理 2.2 在假设 2.1、2.2 和 2.5 的前提下,当 $n \rightarrow \infty$ 时,有 $\tilde{\theta} - \theta^* \rightarrow_p 0$.

证明 因为

$$\begin{aligned} \tilde{\theta}_{\mathcal{A}} - \theta_{\mathcal{A}}^* &= (X_{\mathcal{A}}^T X_{\mathcal{A}}^*)^{-1} X_{\mathcal{A}}^T (X_{\mathcal{A}}^* \theta_{\mathcal{A}}^* + \epsilon) - \theta_{\mathcal{A}}^* = \\ &= \frac{X_{\mathcal{A}}^T X_{\mathcal{A}}^*}{n} (X_{\mathcal{A}}^T \epsilon) \rightarrow_p 0, \end{aligned}$$

即 $\tilde{\theta} - \theta^* \rightarrow_p 0$.

由引理 2.1 和引理 2.2,我们可以得到 $\hat{\theta} - \theta^* \rightarrow_p 0$, 即 LASSO 估计 $\hat{\theta}$ 依概率收敛于真实回归参数 θ^* , 但该结论无法保证 $\hat{\theta}$ 选择的非零元素也收敛于 θ^* 中的非零元素. 为此,基于上述结论需要进一步建立如下结论.

定理 2.1 在假设 2.1、2.2、2.4 和 2.5 的前提下,当 $n \rightarrow \infty$ 时,有 $P(\hat{\mathcal{A}}^* = \mathcal{A}^*) = 1$.

证明 欲证 $P(\hat{\mathcal{A}}^* = \mathcal{A}^*) \rightarrow 1$, 只要分别证 $\forall s \in \mathcal{A}^* \text{ s.t. } P(s \in \hat{\mathcal{A}}^*) \rightarrow 1$ 和 $\forall s' \notin \mathcal{A}^* \text{ s.t. } P(s' \in \hat{\mathcal{A}}^*) \rightarrow 0$ 成立即可.

首先,令 $\theta = \theta^* + u/\sqrt{n}$, 定义

$$\begin{aligned} \Psi(u) &= \frac{1}{n} \|Y - X(\theta^* + u/\sqrt{n})\|_2^2 + \\ &= \lambda_n \sum_{i=1}^n \sum_{j=0}^p \hat{w}_{ij} |\theta_{ij}^* + u_{ij}/\sqrt{n}|, \end{aligned}$$

于是,令 $V(u) = \Psi(u) - \Psi(0)$, 接下来有

$$V(u) = \frac{u^T}{\sqrt{n}} \left(\frac{X^T X}{n}\right) \frac{u}{\sqrt{n}} - \frac{u^T}{\sqrt{n}} \frac{2X^T \epsilon}{\sqrt{n}} +$$

$$\lambda_n \sum_{i=1}^n \sum_{j=0}^p \hat{w}_{ij} (|\theta_{ij}^* + u_{ij}/\sqrt{n}| - |\theta_{ij}^*|).$$

进而有 $\hat{u} = \arg \min_u \Psi(u) = \arg \min_u V(u)$, $\hat{\theta}^* = \theta^* +$

\hat{u}/\sqrt{n} .

当 $\theta_{ij}^* \neq 0$ 时,则 $\hat{w}_{ij} \rightarrow_p \lambda_n^{|\gamma_1|} / |\theta_{ij}^*|^{\gamma_2}$,

$\sqrt{n}(|\theta_{ij}^* + u_{ij}/\sqrt{n}| - |\theta_{ij}^*|) \rightarrow_p u_{ij} \text{sign}(\theta_{ij}^*)$, 进而,由 Slutsky 定理,可以得到

$$\frac{\lambda_n}{\sqrt{n}} (|\theta_{ij}^* + u_{ij}/\sqrt{n}| - |\theta_{ij}^*|) \rightarrow_p 0.$$

于是,当 $\theta_{ij}^* \neq 0$ 时,

$$V(u) \rightarrow_p \frac{u_{\mathcal{A}^*}^T}{\sqrt{n}} \left(\frac{X_{\mathcal{A}^*}^T X_{\mathcal{A}^*}}{n}\right) \frac{u_{\mathcal{A}^*}}{\sqrt{n}} - \frac{u_{\mathcal{A}^*}^T}{\sqrt{n}} W_{\mathcal{A}^*} \quad (16)$$

式中, $W_{\mathcal{A}^*} = \frac{2X_{\mathcal{A}^*}^T \epsilon}{\sqrt{n}}$. 由式(16)可以得到 $\frac{\hat{u}_{\mathcal{A}^*}}{\sqrt{n}} \rightarrow_p$

0, 即 $\hat{\theta}_{\mathcal{A}^*} - \theta_{\mathcal{A}^*}^* \rightarrow_p 0$. 于是有 $\forall s \in \mathcal{A}^* \text{ s.t. } P(s \in \hat{\mathcal{A}}^*) \rightarrow 1$ 成立. 下面,证明 $\forall s' \notin \mathcal{A}^* \text{ s.t. } P(s' \in \hat{\mathcal{A}}^*) \rightarrow 0$ 成立.

当 $\forall s' \notin \mathcal{A}^*$, 即 $\theta_{s'}^* = 0$; 有 $\sqrt{n} \lambda_n \hat{w}_{ij} \rightarrow_p \infty$; 根据 $V(u)$, 有 $\hat{u}_{s'} = \sqrt{n} (\hat{\theta}_{s'}^* - \theta_{s'}^*) \rightarrow_p 0$, 于是 $\frac{2X_{s'}^T (Y - X\hat{\theta}^*)}{\sqrt{n}} = \frac{2X_{s'}^T X \sqrt{n} (\theta^* - \hat{\theta}^*)}{n} + \frac{2X_{s'}^T \epsilon}{\sqrt{n}} \rightarrow_p 0$.

当 $\forall s' \in \hat{\mathcal{A}}^*$: 根据 KKT 最优条件,可以得到 $2X_{s'}^T (Y - X\hat{\theta}^*) = n \lambda_n \hat{w}_{s'}$. 于是,对于 $\forall s' \notin \mathcal{A}^*$, 有

$P(s' \in \hat{\mathcal{A}}^*) \leq P(2X_{s'}^T (Y - X\hat{\theta}^*) = n \lambda_n \hat{w}_{s'}) \rightarrow 0$. 证毕.

集合 $\hat{\mathcal{A}}^*$ 和 \mathcal{A}^* 分别代表自适应 LASSO 估计 $\hat{\theta}$ 和真实回归参数 θ^* 中非零元素的位置,于是定理 2.1 保证了 TS-LASSO 方法的变点估计收敛于真实变点.

2.4 惩罚参数的选择

从上述结论可以看出,TS-LASSO 方法的关键在于选取合适的惩罚参数. 通常可采用交叉验证^[18]算法进行自动选取. 接下来,本文将分别介绍上述两阶段中惩罚参数 $(\lambda_n, \gamma_1, \gamma_2)$ 的 5 折交叉验证选取步骤:

第一阶段惩罚参数的选取:

①在 λ_n 适当的取值范围内,将 λ_n 可能的取值按从小到大排列.

②对每一个 λ_n 值,将回归模型(4)中的样本 (X, Y) 随机分成 5 个相等的子集.

③将其中一份样本作为测试集 $(X_{\text{test}}, Y_{\text{test}})$, 剩余样本作为训练集 $(X_{\text{train}}, Y_{\text{train}})$.

④利用 $(X_{\text{train}}, Y_{\text{train}})$ 训练得到的线性模型(4)计算 $(X_{\text{test}}, Y_{\text{test}})$ 对应的预测值 Y_{pre} , 然后计算预测误差 $R_i = \sqrt{(Y_{\text{test}} - Y_{\text{pre}})^2}$.

⑤重复步骤③和④直到取遍②中所有的样本子集作为测试集.

⑥取得到的 5 个预测误差的平均值作为 λ_n 的交叉验证误差估计 $R_{cv}(\lambda_n) = \sum_{i=1}^5 R_i / 5$.

⑦重复步骤②至⑥,直到取遍①中所有的可能值.

⑧最后选出最小交叉验证误差估计对应的

λ_n 值.

第二阶段惩罚参数 $(\lambda_n, \gamma_1, \gamma_2)$ 的选取:

① 记适当的 γ_1, γ_2 取值范围分别为 B_1 和 B_2 .

② 对每一个 $(\gamma_1, \gamma_2) \in B_1 \times B_2$ (集合 B_1 和 B_2 的笛卡尔积), 按照第一阶段参数 λ_n 的选取步骤, 计算 λ_n 对应的最小交叉验证误差估计值.

③ 重复步骤②直到取遍 $B_1 \times B_2$ 中的所有组合 (γ_1, γ_2) .

④ 最后在所有最小交叉验证误差估计值中选取最小值所对应的 $(\lambda_n, \gamma_1, \gamma_2)$.

3 数据分析

为检验 TS-LASSO 方法的有效性, 本节先进行不同类型的模拟序列检验, 然后进行两组真实的地震波数据检验.

3.1 模拟数据分析

依次对如下四个序列模型进行变点估计和模型选择. 其中, 模型 1 为三阶平稳序列, 用以检验 TS-LASSO 方法能否准确识别无变点序列以及对模型选择的效果; 当自回归模型阶数为零时, PSAR 模型则退化成为独立样本下的均值变点模型, 于是在模型 2 中设置了均值变点序列; 模型 3 和模型 4 为一般形式的 PSAR 模型, 分别为多变点一阶模型和多变点混合阶数模型. 各序列模型对应的时序图如图 1 所示.

模型 1 平稳 AR(3) 过程

$$y_t = 1.13y_{t-1} - 0.5y_{t-3} + \epsilon_t, 1 \leq t \leq 1024.$$

模型 2 混合 $\{AR(0); AR(0)\}$ 过程

$$y_t = \begin{cases} 0.4 + \epsilon_t, & 1 \leq t \leq 512; \\ 1.23 + \epsilon_t, & 513 \leq t \leq 1024. \end{cases}$$

模型 3 混合 $\{AR(1); AR(1); AR(1)\}$ 过程

$$y_t = \begin{cases} 0.4y_{t-1} + \epsilon_t, & 1 \leq t \leq 512; \\ -0.6y_{t-1} + \epsilon_t, & 513 \leq t \leq 768; \\ 0.5y_{t-1} + \epsilon_t, & 769 \leq t \leq 1024. \end{cases}$$

模型 4 混合 $\{AR(1); AR(1); AR(3)\}$ 过程

$$y_t = \begin{cases} 0.5y_{t-1} + \epsilon_t, & 1 \leq t \leq 512; \\ 0.9y_{t-1} + \epsilon_t, & 513 \leq t \leq 768; \\ 0.6y_{t-1} - 0.51y_{t-3} + \epsilon_t, & 769 \leq t \leq 1024. \end{cases}$$

针对上述序列模型, 在构造设计矩阵 X 时, 本文取自回归模型阶数 $p = 5$. 参数 λ_n 的备选值取 100 个, 参数 γ_1, γ_2 的备选值设置在 $\{0, 0.1, 0.2, \dots, 2\}$ 范围内. 在表 1 分别给出各模拟序列的 TS-LASSO 算法估计结果以及选定的具体惩罚参数 $(\lambda_n, \gamma_1, \gamma_2)$.

从表 1 和图 1 中的结果可以看出: TS-LASSO 方法不仅能够准确地识别无变点序列 (模型 1), 还能对其模型进行正确地选择. 对于模型 2 中的均值变点, TS-LASSO 方法的处理结果也是正确的. 在一般的 PSAR 序列 (模型 3、4) 中, 虽然在变点估计上有一点偏差, 但偏差范围都较小 (10 个样本点范围内). 考虑到模拟序列的生成过程中带有随机误差, 上述结果仍可以表明 TS-LASSO 方法是有效的. 值得注意的是, TS-LASSO 方法的模型选择结果与真实模型都是一致的. 此外, 从各个模拟序列的惩罚参数可以看出, 关于滞后阶数的惩罚参数都是非零的, 这表明改进的自适应惩罚权重在变点估计过程中是有作用的.

综上各模拟序列的估计结果, 可以表明 TS-LASSO 方法对解决 PSAR 模型中的变点估计和模型选择问题是有效的. 接下来, 在两组地震波数据上进一步检验 TS-LASSO 方法有效性和实用意义.

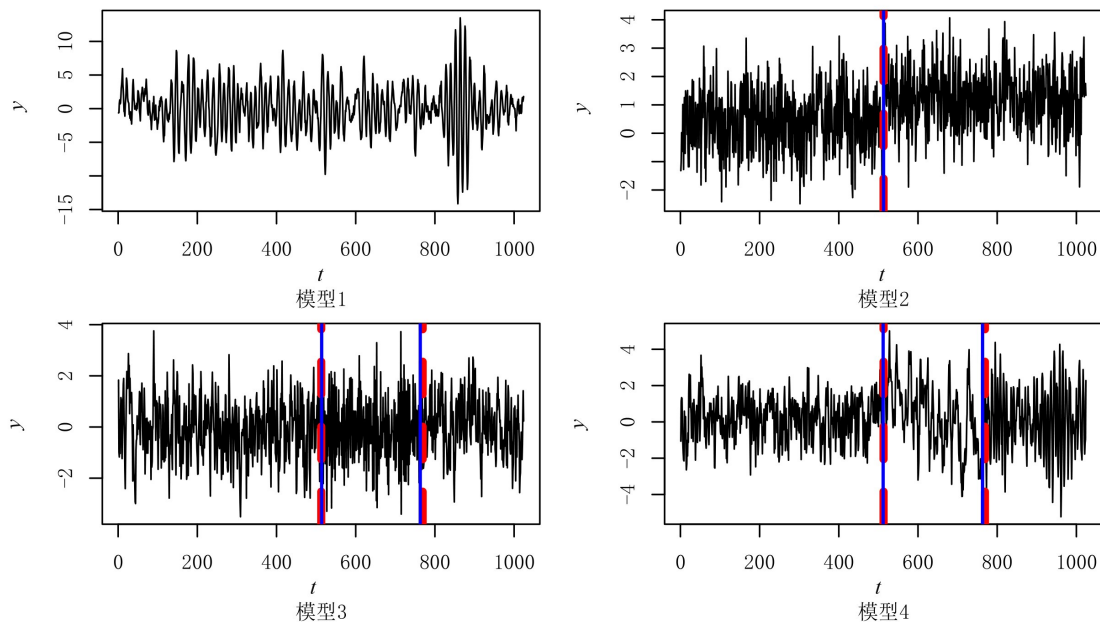


图 1 序列中的真实变点 (垂直红虚线) 和估计变点 (垂直蓝实线)

Fig. 1 Real change-points (vertical red dotted line) and estimated change-points (vertical blue solid line) in sequences

表 1 模拟序列的 TS-LASSO 选择结果
Tab. 1 Results by TS-LASSO for simulative sequences

模型	真实变点位置	估计变点位置	真实模型	选择模型	惩罚参数
1	无	无	(0,1,0,1,0,0)	(0,1,0,1,0,0)	(0.1147,0.7,0.6)
2	513	513	(1,0,0,0,0,0; 1,0,0,0,0,0)	(1,0,0,0,0,0; 1,0,0,0,0,0)	(0.0146,1.7,0.6)
3	513,769	514,763	(0,1,0,0,0,0; 0,1,0,0,0,0; 0,1,0,0,0,0)	(0,1,0,0,0,0; 0,1,0,0,0,0; 0,1,0,0,0,0)	(0.0042,0.9,0.4)
4	513,769	512,769	(0,1,0,0,0,0; 0,1,0,0,0,0; 0,1,0,1,0,0)	(0,1,0,0,0,0; 0,1,0,0,0,0; 0,1,0,1,0,0)	(0.0025,0.8,0.1)

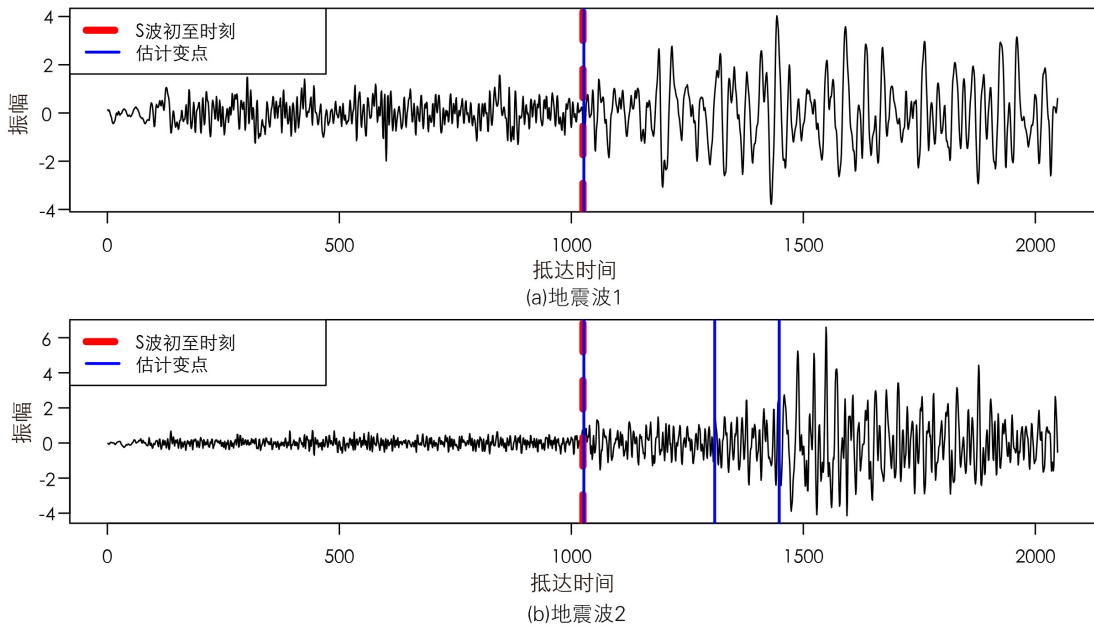


图 2 地震波形序列

Fig. 2 Seismic waveform sequences

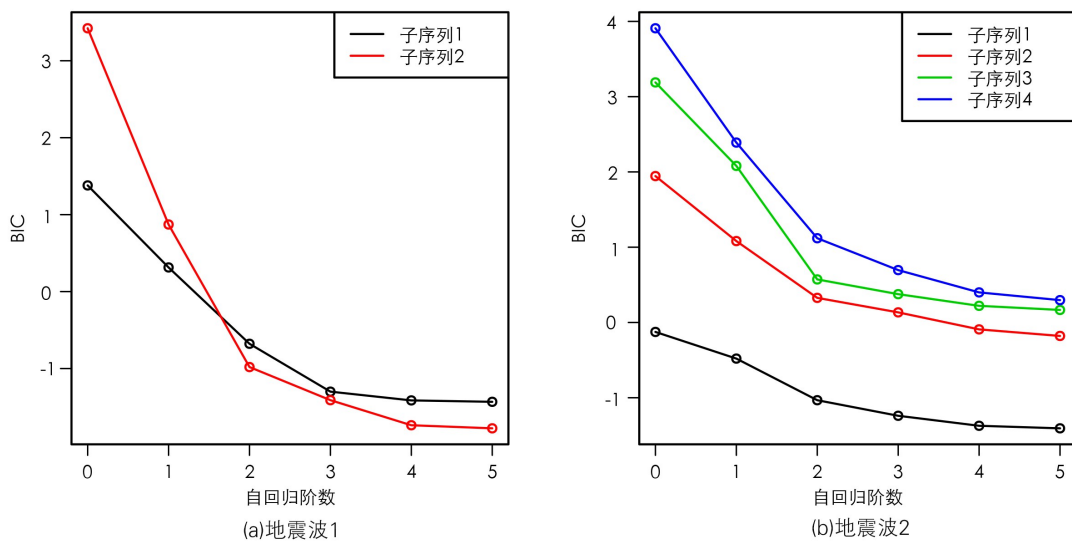


图 3 各子序列不同模型阶数对应的 BIC 值

Fig. 3 The BIC value for different model orders of each sub-sequence

3.2 地震波数据分析

在地震研究领域,地震 P 波和 S 波的初至时刻是计算震源位置的必要参数^[16]. 其中,P 波属于一种纵波,具有振幅小、传播速度快的特点,最先被仪器检测到;S 波属于一种横波,相比于 P 波,其传播速度慢,随后被检测到. 有关地震波初至时间的自动拾取方法有许多,如 Takanami 等^[16]基于同一地震波形是一条稳态自回归序列的假设,通过计算包含多个稳态地震波序列对应的最小 AIC 值,从而有效地识别出 P 波的初至时刻. 基于同样的假设,本文运用 TS-LASSO 方法对两组地震波数据进行震波初至时间的估计.

如图 2 中所示的两组采集自斯堪的纳维亚半岛的地震波数据^[17],它们都由 P 波和 S 波以及叠加在其中的随机噪声波组成,其中 1025 位置(图 2 中的垂直红虚线位置)对应 S 波的初至时刻.

TS-LASSO 方法对上述地震波序列 1 的估计变点位置为 1027,与 S 波初至时刻(1025)很接近;对地震波序列 2 的变点估计位置为 1027、1309 和 1448,其中的 1027 位置与 S 波初至时刻(1025)很接近. 此外,TS-LASSO 方法对两地震波序列各自回归模型的选择结果分别为 \hat{V}_1 和 \hat{V}_2 :

$$\hat{V}_1 = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{pmatrix},$$

$$\hat{V}_2 = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

即各子序列皆满足 2 阶自回归模型.

为检验 TS-LASSO 方法对地震波的模型选择是否准确,本文采用传统的 BIC 准则对分割后自回归模型进行阶数的选择. 如图 3 所示,从不同阶数对应的 BIC 值变化趋势可以看出,各平稳序列在阶数为 2 时对应的 BIC 值明显下降,且再提升阶数时,BIC 值下降幅度趋于平缓,这表明基于 BIC 准则得到的各地震波模型阶数的选择与 TS-LASSO 方法的结果是一致的. 对于图 2 中的地震波 2,TS-LASSO 方法还得到两个未知原因的变点估计(1309 和 1448). 为检验变点估计的合理性,我们分别对各子序列进行 2 阶自回归参数估计,结果如表 2 所示.

表 2 各地震波的 AR 参数估计

地震波	子波	滞后一阶系数	滞后二阶系数
1	1	1.4451	-0.7863
	2	1.8424	-0.9180
2	1	0.9098	-0.6610
	2	1.3337	-0.7344
	3	1.5593	-0.8800
	4	1.6380	-0.8497

从表 2 中可以看出,在两组地震波数据中,所有子序列均为平稳的,且相邻地震波的自回归参数的差距都比较大,这表明 TS-LASSO 方法的变点估计结果是合理的.

综上地震波数据的检验结果表明,TS-LASSO 方法在解决 PSAR 模型的变点估计和模型选择问题方面具有一定的实用意义.

4 结论

本文就具体的 PSAR 模型中的变点估计和模型选择问题,采用一种 TS-LASSO 算法,实现了对序列中的变点估计和模型选择问题的同时解决. 采用问题转换的方法,本文先利用恰当的设计矩阵,将变点估计问题转化成线性回归模型中的变量选择问题. 在处理变量选择问题上,为实现各段平稳序列的模型选择,采用单变量惩罚方法,先运用 LASSO 算法得到变点的初步估计,以及各模型的初步选择. 然后,基于初步的估计结果,进一步采用改进的自适应 LASSO 算法,以剔除初步估计中的多余变点,从而实现变点的一致估计和模型的准确选择. 为保证 TS-LASSO 方法的变点估计结果在大样本情形下的有效性,研究了估计结果的渐近性质. 此外,对于特殊情形下的平稳自回归序列(无变点序列)以及均值变化序列(阶数为零的 PSAR 序列),TS-LASSO 算法也能进行有效的识别和估计. 最后,进行模拟序列和地震波序列检验,结果表明 TS-LASSO 算法是有效的且具有一定的应用价值.

自回归模型是时间序列分析中最为常用的模型之一,因为其对数据有着较为直观的解释性. 本文中的 TS-LASSO 算法也是针对具体的 PSAR 模型提出的. 但在实际情况中,时间序列常表现出更为复杂的变化规律. 比如对金融股票价格收益率序列的刻画,往往采用的是一种包含方差变化规律的 ARCH 模型^[18]. 总之,对复杂序列中变点推断的研究会有更大的挑战,同时也具有更广的实用意义,这也是未来我们需要努力的方向.

参考文献(References)

- [1] PAGE E S. Continuous inspection schemes [J]. *Biometrika*, 1954, 41:100-115.
- [2] PICARD D. Testing and estimating change-points in time series [J]. *Advances in Applied Probability*, 1985, 17(4):841-867.
- [3] TAKANAMI T, KITAGAWA G. A new efficient procedure for the estimation of onset times of seismic waves [J]. *Journal of Physics of the Earth*, 1988, 36(6): 267-290.
- [4] DAVIS R A, LEE T C, RODRIGUEZYAM G A, et al. Structural break estimation for nonstationary time Series models[J]. *Journal of the American Statistical Association*, 2006, 101(473): 223-239.
- [5] YAU C Y, ZHAO Z. Inference for multiple change points in time series via likelihood ratio scan statistics [J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016, 78(4): 895-916.
- [6] 吴楠,胡尧,王丹. 分段平稳时间序列中的多变点检测 [J]. *理论数学*, 2018,8(2):136-148.
- [7] HARCHAOUI Z, LÉVY-LEDUC C. Multiple change point estimation with a total variation penalty [J]. *Journal of the American Statistical Association*, 2010, 105(492): 1480-1493.