系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
42(5) (2022, 5), 1314–1329

当前状态数据的可加风险模型变量选择方法*

赵 慧 董庆凯

(中南财经政法大学统计与数学学院, 武汉 430073)

摘要 当协变量数目较多时, 变量选择对模型构建至关重要. 近年来, 以 LASSO 为代表的各种惩罚变量选择方法备受关注, 但生存分析领域的惩罚变量选择方法研究大多基于 Cox 比例风险模型, 且研究对象多为右删失数据. 文章对当前状态数据 (也称 I 型区间删失数据) 在可加风险模型下的变量选择方法进行研究. 在失效时间服从可加风险模型及观测时间与协变量相关的假定下, 从计数过程的角度来构造风险函数, 并给出一种基于重复迭代加权的 BAR (Broken Adaptive Ridge) 惩罚似然变量选择方法, 证明了 Oracle 性质. 通过模拟实验来比较 BAR 与其他常用惩罚似然方法在变量选择方面的效果, 最后利用文章提出的方法分析一项阿尔茨海默病的研究数据. 模拟实验和实证分析都表明了 BAR 方法在变量选择方面表现良好.

关键词 可加风险模型, BAR 估计, 当前状态数据, 变量选择.

MR(2000) 主题分类号 62N01, 62P10

DOI 10.12341/jssms21468

A Variable Selection Method for the Additive Hazards Model with Current Status Data

ZHAO Hui DONG Qingkai

(*School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan* 430073)

Abstract Variable selection is vital to statistical modelling when the number of covariates is large. In recent years, penalty function-based methods represented by LASSO have attracted much attention. But most researches in survival analysis are based on the Cox proportional hazards model and right-censored data. In this paper, we consider current status data (also called type I interval-censored data) with the additive hazards model which is less studied. Under the assumption that the failure

* 国家自然科学基金面上项目 (12171483), 中南财经政法大学研究生实践创新项目 (202251311) 资助课题.
收稿日期: 2021-09-01, 收到修改稿日期: 2021-12-08.
编委: 林华珍.

time follows the additive hazards model and the censoring time is dependent on covariates, the hazard function is constructed from the perspective of counting process, and then a simple likelihood function is derived. A BAR (Broken Adaptive Ridge) variable selection method is proposed, which is based on iteratively reweighted penalization and enjoys Oracle property. We compare BAR with some popular penalized methods through simulation and apply it to the current status data arising from the Alzheimer's disease study. Both simulation and application show that BAR performs better compared with popular penalized methods.

Keywords Additive hazards model, broken adaptive ridge estimate, current status data, variable selection.

1 引 言

生存分析中的一个主要任务就是探究对感兴趣事件发生时间有影响的协变量及影响程度. 但在许多实际问题中, 潜在的协变量个数较多, 这时就需要从中筛选出重要的协变量.

一些传统的变量选择方法通过剔除系数不显著的变量来完成变量选择, 如向前选择法, 向后选择法和最优子集法. 另外, 基于 AIC[1] 和 BIC[2] 等信息准则的变量选择方法也已相当成熟. 在这些成果的基础上, 许多学者提出惩罚函数变量选择方法, 其中 Tibshirani[3] 提出的 LASSO 方法是早期最具代表性的惩罚函数变量选择方法. 之后 Fan 和 Li[4] 改进了 LASSO 方法对绝对值大的系数过度压缩的缺陷, 提出了 SCAD 方法. Zou[5] 在 LASSO 的基础上改进了系数的权重项, 提出 ALASSO 方法. Lü 和 Fan[6] 提出了将 L_0 和 L_1 惩罚函数相结合的 SICA 方法. Dicker[7] 提出 SELO 方法, 构造了连续的惩罚函数来近似 L_0 惩罚函数. Liu 和 Li[8] 提出了迭代更新惩罚函数权重项的 BAR 方法, 其思想是用重加权的 L_2 惩罚函数来近似 L_0 惩罚函数.

目前上述变量选择方法已被广泛应用于针对删失数据的生存分析模型, 特别是 Cox 比例风险模型 [9–11] 和可加风险模型 [12, 13], 但这些研究针对的多是右删失数据. 与右删失数据相比, 当前状态数据蕴含的信息量更少, 估计更困难. 据我们了解, 目前对当前状态数据的变量选择问题的研究较少. Tian 等 [14] 对当前状态数据的广义线性回归模型进行了惩罚函数变量选择研究, Zhao 等 [15, 16] 在 Cox 比例风险模型下研究了一般区间删失数据的变量选择问题, 并在其提出的模型下对多种惩罚函数变量选择方法进行比较, 评估各种方法的估计效果.

一般来讲, Cox 比例风险模型关注风险函数的相对比值, 可加风险模型则主要关注风险函数的绝对差值. 作为对 Cox 模型的一种重要补充, 可加风险模型在生存分析中也占据重要地位, 但是在可加模型下对当前状态数据的变量选择问题的研究却比较少见. 另外, 在删失数据的现有研究中, 出于简单考虑, 大多假设观测时间与协变量无关, 但在很多实际问题中, 观测时间依赖于协变量, 例如临床治疗中对患者进行观测的时间常与其身体状况有关. 这种情况在 Lin 等 [17] 和 Wang 等 [18] 的文章中有所提及. 在多元协变量的情形下, 这种假设将使模型中的待定参数增加, 给参数估计和变量选择带来更多困难.

阿尔茨海默病 (AD) 是一种不可逆的神经退行性疾病, 会对人的记忆、思维和行为造成极大损害. 轻度认知障碍期 (MCI) 是 AD 发病之前的阶段. 调查显示, 截至 2014 年, 我国已

有约 2386 万名 65 岁以上老年人处于 MCI 期 [19]. 结合我国正处在人口老龄化上升期的时代背景, 探究影响 MCI 向 AD 转化的因素具有重要的医学和经济价值. 本文的实证部分将探讨针对这种疾病的研究所产生的当前状态数据的变量选择问题.

综合研究现状, 本文将在可加风险模型及观测时间与协变量相关的假定下, 探究当前状态数据下的变量选择问题. 具体内容安排为: 第 1 节介绍研究背景及国内外学者的研究情况; 第 2 节介绍本文所用到的模型假定及惩罚函数, 提出一种基于 BAR 惩罚函数的可加风险模型下当前状态数据变量选择方法, 并在适当的条件下给出其 Oracle 性质及证明; 第 3 节探究在不同模拟实验设定下各惩罚方法的表现; 第 4 节利用本文提出的方法分析阿尔茨海默病真实数据; 第 5 节总结全文并提出展望.

2 模型介绍

2.1 数据结构及模型假定

2.1.1 当前状态数据

考虑 n 个独立观测个体, 记个体 i 的失效时间为 T_i, 观测时间为 C_i $(i=1,2,\cdots,n)$. 若我们只知道 T_i 是发生在 C_i 时刻之前或之后, 而无法得知 T_i 的准确数值, 这样的数据就被称为当前状态数据或 I 型区间删失数据. 当前状态数据在生物医学和经济学等科学领域中较为常见. 例如对于某种癌症的发病时间, 不知道患者癌症发病的准确时刻 T, 而只能得知患者在观测时刻 C 是否已经发病.

对于真实失效时间 T 的观测, 当前状态数据能提供的信息量比右删失数据更少. 在右删失情形, 我们能观测到 $\min(T,C)$ 和示性函数 $I(T\le C)$, 对于个体 i, $I(T_i\le C_i)=1$ 意味着 $\min(T_i,C_i)=T_i$, 即 T_i 的准确值被观测到. 而在当前状态数据中, 我们只能得到 C 和示性函数 $I(T\ge C)$, 对个体 i, $I(T_i\ge C_i)=1$ 表示在观测时个体 i 未失效, 说明该个体右删失; 而 $I(T_i\ge C_i)=0$ 则表示在观测时个体 i 已失效, 该个体左删失. 可见当前状态数据可看作左删失和右删失的混合, 在这种数据产生机制下, 无论如何都观测不到 T_i 的准确值.

2.1.2 模型假定

设个体 i 的失效时间受协变量 $\boldsymbol{Z}_i=(Z_{i1},Z_{i2},\cdots,Z_{ip})'$ 影响, $\boldsymbol{Z}_i$ 与时间 t 无关. 可加风险模型假定在已知协变量的条件下, 失效时间 T_i 在 t 时刻的风险函数为

$$\lambda_i(t|\boldsymbol{Z}_i)=\lambda_0(t)+\boldsymbol{\beta}'\boldsymbol{Z}_i, \tag{2.1}$$

其中 $\lambda_0(\cdot)$ 是未知的非负基准风险函数, $\boldsymbol{\beta}$ 是未知回归系数, 其分量表示各协变量对风险函数的影响程度. Cox 比例风险模型只能从相对比值的角度来比较不同协变量取值对风险的影响程度, 而可加风险模型可以衡量风险的绝对差异, 在很多应用场景下更易解释.

另外, 一般的生存分析模型通常假设观测时间 C 的分布与协变量无关. 但在很多实际问题中, 观测时间 C 可能受到某些协变量的影响. 这里用如下 Cox 比例风险模型来描述观测时间与协变量的关系

$$\mathrm{d}\Lambda_c(t|\boldsymbol{X}_i)=\mathrm{e}^{\boldsymbol{\gamma}'\boldsymbol{X}_i}\mathrm{d}\Lambda_{c,0}(t), \tag{2.2}$$

其中 $\Lambda_{c,0}(\cdot)$ 是未知的基准累积风险函数, $\boldsymbol{\gamma}$ 是未知回归参数. $\boldsymbol{X}_i=(X_{i1},X_{i2},\cdots,X_{id})'$ 通常是 $\boldsymbol{Z}_i$ 的某个子集. 本文中假设 d 较小, 且 $\boldsymbol{X}_i$ 与时间 t 无关. 另外假设在给定 $\boldsymbol{Z}_i$ 的前提下,

C 和 T 是独立的.

2.1.3 风险函数转换

Lin 等 [17] 从计数过程的角度构造风险函数, 用比例风险模型下风险函数的形式来表示可加风险模型下的风险函数. 定义如下的计数过程

$$N_i(t) = \delta_i I(C_i \le t), \tag{2.3}$$

其中, $\delta_i = I(C_i \le T_i)$ 是当前状态数据的示性函数. 设 $\mathrm{d}H_i(t)$ 表示个体 i 的计数过程发生跳跃 ($N_i(t)$ 从 0 变为 1, 即 $\mathrm{d}N_i(t) = 1$) 的概率. 根据 $N_i(t)$ 的定义, 计数过程发生跳跃的条件是: 个体 i 在 t 时刻被观测且未失效, 即 $C_i = t$ 且 $T_i \ge t$. 用 C_i 在 t 时刻的风险函数 $\mathrm{d}\Lambda_c(t|\boldsymbol{X}_i) = \mathrm{e}^{\boldsymbol{\gamma}'\boldsymbol{X}_i}\mathrm{d}\Lambda_{c,0}(t)$ 表示前者的概率, 用 $\Pr(T_i \ge t|\boldsymbol{Z}_i) = \mathrm{e}^{-\Lambda_0(t)-\boldsymbol{\beta}'\boldsymbol{Z}_i^*(t)}$ 表示后者的概率, 其中 $\Lambda_0(t) = \int_0^t \lambda_0(s)\,\mathrm{d}s$, $\boldsymbol{Z}_i^*(t) = \int_0^t \boldsymbol{Z}_i \mathrm{d}s$. 二者相乘得

$$\mathrm{d}H_i(t) = \mathrm{d}H_0(t)\,\mathrm{e}^{-\boldsymbol{\beta}'\boldsymbol{Z}_i^*(t)+\boldsymbol{\gamma}'\boldsymbol{X}_i}. \tag{2.4}$$

可见, $N_i(t)$ 发生跳跃的概率类似于 Cox 模型的风险函数, 其中 $\mathrm{d}H_0(t) = \mathrm{e}^{-\Lambda_0(t)}\mathrm{d}\Lambda_{c,0}(t)$ 被视为基准风险函数. 这样就可以按照 Cox 模型下构造偏似然的方法, 基于 $\mathrm{d}H_i(t)$ 构造如下偏似然函数

$$L(\boldsymbol{\beta},\boldsymbol{\gamma}) = \prod\nolimits_{i=1}^{n}\left[\frac{\mathrm{e}^{-\boldsymbol{\beta}'\boldsymbol{Z}_i^*(C_i)+\boldsymbol{\gamma}'\boldsymbol{X}_i}}{\sum_{j=1}^{n} Y_j(C_i)\,\mathrm{e}^{-\boldsymbol{\beta}'\boldsymbol{Z}_j^*(C_i)+\boldsymbol{\gamma}'\boldsymbol{X}_j}}\right]^{\delta_i}, \tag{2.5}$$

其中, $\boldsymbol{Z}_i^*(C_i) = \int_0^{C_i}\boldsymbol{Z}_i\mathrm{d}s = C_i\boldsymbol{Z}_i$, $Y_j(C_i) = I(C_j \ge C_i)$. 对任意 C_i, $\{j|Y_j(C_i) = 1\}$ 构成了 C_i 时刻的风险集. 向 (2.5) 式中代入数据 $\{C_i, \delta_i, \boldsymbol{Z}_i, \boldsymbol{Z}_i^*(\cdot), \boldsymbol{X}_i\}\,(i = 1, 2, \cdots, n)$ 并极大化, 就能估计出未知参数 $\boldsymbol{\beta}$ 和 $\boldsymbol{\gamma}$. 可见 (2.5) 式与未知的基准风险函数无关, 从而避免了对未知基准风险函数进行近似估计所带来的误差. 并且, 符合 (2.2) 式的观测时间 C 准确值已被全部观测到, 所以可基于完全数据 $\{C_i, \boldsymbol{X}_i\}$ $(i = 1, 2, \cdots, n)$, 通过极大化 Cox 比例风险模型的偏似然函数 $L_{\boldsymbol{\gamma}}(\boldsymbol{\gamma})$ 来得到 $\widehat{\boldsymbol{\gamma}}$

$$L_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) = \prod_{i=1}^{n}\left(\frac{\mathrm{e}^{\boldsymbol{\gamma}'\boldsymbol{X}_i}}{\sum_{j=1}^{n} Y_j(C_i)\,\mathrm{e}^{\boldsymbol{\gamma}'\boldsymbol{X}_j}}\right). \tag{2.6}$$

注意到在给定 $\widehat{\boldsymbol{\gamma}}$ 后, (2.5) 式仅比一般的 Cox 比例风险模型偏似然函数多了一个与 $\boldsymbol{\beta}$ 无关的常数项, 因此 $L(\boldsymbol{\beta},\widehat{\boldsymbol{\gamma}})$ 的理论性质与 Cox 比例风险模型偏似然函数的理论性质很接近, 后文的定理也将印证这一点. 另外不难发现, 当 $\boldsymbol{\gamma} = \mathbf{0}$ 时, C 与协变量无关, 此时 (2.5) 式变为 Cox 比例风险模型的偏似然函数形式, 对此式求极值所得到的估计有着与 Cox 比例风险模型参数估计相同的性质.

2.2 惩罚函数变量选择方法

2.2.1 常见惩罚方法

当协变量的个数较多时, 容易出现多重共线性, 且对数据搜集和模型解释造成困难. 此时可以采用惩罚函数变量选择方法来精简模型, 剔除对因变量影响不显著的协变量. 在代入 $\widehat{\boldsymbol{\gamma}}$ 后, (2.5) 式中的未知参数只剩 $\boldsymbol{\beta}$, 可将 (2.5) 式记为 $L(\boldsymbol{\beta})$. 基于 $L(\boldsymbol{\beta})$ 构造惩罚偏似然函数 $L_p(\boldsymbol{\beta})$

$$L_p(\boldsymbol{\beta}) = -2\log L(\boldsymbol{\beta}) + \sum_{j=1}^{p} p(\boldsymbol{\beta}_j, \lambda_n), \tag{2.7}$$

其中, $p(\cdot)$ 是惩罚函数, λ_n 是待定调节参数.

LASSO 方法的惩罚函数为 $p_{LASSO}(\boldsymbol{\beta}_j,\lambda_n)=\lambda_n|\boldsymbol{\beta}_j|$. 以 LASSO 为基础, ALASSO 方法的惩罚函数为 $p_{ALASSO}(\boldsymbol{\beta}_j,\lambda_n)=\lambda_n\omega_j|\boldsymbol{\beta}_j|$, 其中 ω_j 是 $\boldsymbol{\beta}_j$ 的权重, Zou 等 [5] 建议取 $\omega_j=1/\left|\widetilde{\boldsymbol{\beta}}_j\right|$, 其中 $\widetilde{\boldsymbol{\beta}}_j$ 由极大化 $L(\boldsymbol{\beta})$ 而得. SCAD 方法的惩罚函数为

$$p_{SCAD}(\boldsymbol{\beta}_j,\lambda_n)=\begin{cases}\lambda_n|\boldsymbol{\beta}_j|, & |\boldsymbol{\beta}_j|\le\lambda_n,\\ -\dfrac{\boldsymbol{\beta}_j^2-2\alpha\lambda_n|\boldsymbol{\beta}_j|+\lambda_n^2}{|2(\alpha-1)|}, & \lambda_n<|\boldsymbol{\beta}_j|\le\alpha\lambda_n,\\ \dfrac{(\alpha+1)\lambda_n^2}{2}, & \alpha\lambda_n<|\boldsymbol{\beta}_j|.\end{cases}\tag{2.8}$$

按照 Fan 和 Li[4] 的结论, 取 $\alpha=3.7$. 易知, 当 $|\boldsymbol{\beta}_j|$ 较小时, 对 $\boldsymbol{\beta}_j$ 的惩罚随 $|\boldsymbol{\beta}_j|$ 的增大而增大. 当 $|\boldsymbol{\beta}_j|$ 较大时, 对 $\boldsymbol{\beta}_j$ 的惩罚仅由 α 和 λ_n 控制.

SELO 方法的惩罚函数为 $p_{SELO}(\boldsymbol{\beta}_j,\lambda_n)=\frac{\lambda_n}{\log(2)}\log\left(\frac{|\boldsymbol{\beta}_j|}{|\boldsymbol{\beta}_j|+\tau}+1\right)$, 其中 λ_n 和 τ 是待定调节参数. 当 τ 取值较小时 $p_{SELO}(\boldsymbol{\beta}_j,\lambda_n)\approx\lambda_n I(\boldsymbol{\beta}_j\neq0)$. Dicker 等 [7] 建议取 $\tau=0.01$. SICA 方法的惩罚函数为 $p_{SICA}(\boldsymbol{\beta}_j,\lambda_n)=\lambda_n\frac{|\boldsymbol{\beta}_j|(\tau+1)}{|\boldsymbol{\beta}_j|+\tau}$. Lü和 Fan[6] 建议取 $\tau=0.01$. 这两种惩罚函数是对 L_0 惩罚函数的连续近似, 在一定程度上克服了 L_0 惩罚函数的不连续性所带来的计算复杂性和不稳定性.

2.2.2 BAR 方法及其性质

BAR 方法是一种迭代重加权方法. 首先寻找一个迭代初值 $\widehat{\boldsymbol{\beta}}^{(0)}$, 此初值可以由极大化 $\log L(\boldsymbol{\beta})$ 来获得, 也可以用岭估计. 迭代过程中, 第 k 步迭代所求的 $\widehat{\boldsymbol{\beta}}^{(k)}$ 由重加权的 L_2 惩罚函数来更新

$$\widehat{\boldsymbol{\beta}}^{(k)}=\arg\min_{\boldsymbol{\beta}}\left\{-2\log L(\boldsymbol{\beta})+\lambda_n\sum_{j=1}^{p}\frac{\boldsymbol{\beta}_j^2}{\left(\widehat{\boldsymbol{\beta}}_j^{(k-1)}\right)^2}\right\},\quad k\ge1,\tag{2.9}$$

其中 λ_n 是调节参数, 惩罚函数为 $p_{BAR}(\boldsymbol{\beta}_j,\lambda_n)=\lambda_n\frac{\boldsymbol{\beta}_j^2}{(\widehat{\boldsymbol{\beta}}_j^{(k-1)})^2}$, BAR 估计量被定义为 $\widehat{\boldsymbol{\beta}}=\lim_{k\to\infty}\widehat{\boldsymbol{\beta}}^{(k)}$. 可以证明, 在一定条件下, $p_{BAR}(\boldsymbol{\beta}_j,\lambda_n)$ 将收敛到 $\lambda_n I(\boldsymbol{\beta}_j\neq0)$. 因此这种方法也可看作对 L_0 惩罚函数的 L_2 近似, 且无需 λ_n 以外的待定调节参数. 已经有学者采用 BAR 方法研究线性模型 [8], Cox 比例风险模型 [16] 等模型, 并在多种模型下证明了 BAR 估计量在变量选择上的一致性, Oracle 性质及变量自动分组特性.

据我们所知, 尚未有将 BAR 方法引入当前状态数据的可加风险模型研究的文章, 下面在 $p<n$ 的情形下, 给出 BAR 估计 $\widehat{\boldsymbol{\beta}}$ 在本文模型下的 Oracle 性质, 这些性质很容易推广到 p 可随 n 的增大而增大的情形. $p>n$ 的讨论可参见 [20].

首先引入一些计数过程中的记号及证明所需条件. 记

$$M_i(t)=N_i(t)-\int_0^1 Y_i(s)\mathrm{e}^{-\boldsymbol{\beta}'\boldsymbol{Z}_i^*(s)+\boldsymbol{\gamma}'\boldsymbol{X}_i}\mathrm{d}H_0(s),\quad i=1,2,\cdots,n\tag{2.10}$$

为定义在 σ 域 $\mathscr{F}_t:=\sigma\{N_i(s),Y_i(s),\boldsymbol{Z}_i(s):s\leqslant t,i=1,2,\cdots,n\}$ 上的鞅, 不失一般性, 设最晚

的观测时间为 1. 记

$$S^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},t)=\sum_{j=1}^{n}\left(\boldsymbol{Z}_j^*(t)\right)^{\otimes k}Y_j(t)\mathrm{e}^{-\boldsymbol{\beta}'\boldsymbol{Z}_j^*(t)+\boldsymbol{\gamma}'\boldsymbol{X}_j},\quad k=0,1,2, \tag{2.11}$$

其中 $k=0,1,2$ 时 $\boldsymbol{x}^{\otimes k}=1,\boldsymbol{x},\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$. 由 (2.5) 式, 在给定 $\widehat{\boldsymbol{\gamma}}$ 后, 记关于 $\boldsymbol{\beta}$ 的对数偏似然函数为

$$\begin{aligned}l(\boldsymbol{\beta})=&\log L(\boldsymbol{\beta},\widehat{\boldsymbol{\gamma}})\\=&\sum_{i=1}^{n}\int_0^1\left(-\boldsymbol{\beta}'\boldsymbol{Z}_i^*(s)+\widehat{\boldsymbol{\gamma}}'\boldsymbol{X}_i\right)\mathrm{d}N_i(s)\\&-\int_0^1\log\left(\sum_{j=1}^{n}Y_j(s)\mathrm{e}^{-\boldsymbol{\beta}'\boldsymbol{Z}_j^*(s)+\widehat{\boldsymbol{\gamma}}'\boldsymbol{X}_j}\right)\mathrm{d}\overline{N}(s),\end{aligned} \tag{2.12}$$

其中 $\overline{N}=\sum_{i=1}^{n}N_i$. 记 $U_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\gamma})=\frac{\partial\log L(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\boldsymbol{\beta}}$, $U_{\boldsymbol{\gamma}}(\boldsymbol{\gamma})=\frac{\partial\log L_{\boldsymbol{\gamma}}(\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}$, $\widehat{\Omega}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\gamma})=-n^{-1}\frac{\partial U_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\boldsymbol{\beta}'}$, $\widehat{\Omega}_{\boldsymbol{\beta}\boldsymbol{\gamma}}(\boldsymbol{\beta},\boldsymbol{\gamma})=-n^{-1}\frac{\partial U_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}'}$, $\widehat{D}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma})=-n^{-1}\frac{\partial U_{\boldsymbol{\gamma}}(\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}'}$.

设 $\boldsymbol{\beta}_0$ 和 $\boldsymbol{\gamma}_0$ 为 $\boldsymbol{\beta}$ 和 $\boldsymbol{\gamma}$ 的真值. 记 $\Omega_{\boldsymbol{\beta}}$, $\Omega_{\boldsymbol{\beta}\boldsymbol{\gamma}}$ 和 $D_{\boldsymbol{\gamma}}$ 分别为 $\widehat{\Omega}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\gamma})$, $\widehat{\Omega}_{\boldsymbol{\beta}\boldsymbol{\gamma}}(\boldsymbol{\beta},\boldsymbol{\gamma})$ 和 $\widehat{D}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma})$ 在 $\boldsymbol{\beta}=\boldsymbol{\beta}_0$ 和 $\boldsymbol{\gamma}=\boldsymbol{\gamma}_0$ 时的极限, 并假设 $\Omega_{\boldsymbol{\beta}}$ 和 $D_{\boldsymbol{\gamma}}$ 非奇异.

令 $\widetilde{\boldsymbol{\beta}}$ 表示对 $l(\boldsymbol{\beta})$ 求极大值而得的估计. 基于 (2.10)–(2.12), 文献 [17] 已证明 $\frac{1}{\sqrt{n}}U_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\gamma}}\right)$ 以及 $\sqrt{n}\left(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)$ 都收敛到 0 均值多元正态分布, 且协方差阵分别为 $M\left(\boldsymbol{\beta}_0\right)=\Omega_{\boldsymbol{\beta}}-\Omega_{\boldsymbol{\beta}\boldsymbol{\gamma}}D_{\boldsymbol{\gamma}}^{-1}\Omega_{\boldsymbol{\beta}\boldsymbol{\gamma}}'$ 及 $V\left(\boldsymbol{\beta}_0\right)=\Omega_{\boldsymbol{\beta}}^{-1}-\Omega_{\boldsymbol{\beta}}^{-1}\Omega_{\boldsymbol{\beta}\boldsymbol{\gamma}}D_{\boldsymbol{\gamma}}^{-1}\Omega_{\boldsymbol{\beta}\boldsymbol{\gamma}}'\Omega_{\boldsymbol{\beta}}^{-1}$. 不失一般性, 可以将 $\boldsymbol{\beta}_0$ 记为 $(\boldsymbol{\beta}_{01}',\boldsymbol{\beta}_{02}')'$, $\boldsymbol{\beta}_{01}$ 表示值非零的分量, 维数为 q, $\boldsymbol{\beta}_{02}$ 表示值为零的分量, 维数为 $p-q$. 对应地, $\widetilde{\boldsymbol{\beta}}$ 表示为 $\left(\widetilde{\boldsymbol{\beta}}_1',\widetilde{\boldsymbol{\beta}}_2'\right)'$, $\widehat{\boldsymbol{\beta}}$ 表示为 $\left(\widehat{\boldsymbol{\beta}}_1',\widehat{\boldsymbol{\beta}}_2'\right)'$, $M_1\left(\boldsymbol{\beta}_{01}\right)$ 是 $M\left(\boldsymbol{\beta}_0\right)$ 的前 $q\times q$ 子矩阵, $V_1\left(\boldsymbol{\beta}_{01}\right)$ 是 $V\left(\boldsymbol{\beta}_0\right)$ 的前 $q\times q$ 子矩阵, 下同.

定理 1 (Oracle 性质) 记 $Q_1(\boldsymbol{\theta}_1)=-2l_1(\boldsymbol{\theta}_1)+\lambda_n\boldsymbol{\theta}_1'D_1(\boldsymbol{\beta}_1)\boldsymbol{\theta}_1$, 其中 $\boldsymbol{\theta}_1$ 是 q 维向量, $D_1(\boldsymbol{\beta_1})=\mathrm{diag}\{\boldsymbol{\beta}_1^{-2},\boldsymbol{\beta}_2^{-2},\cdots,\boldsymbol{\beta}_q^{-2}\}$, $l_1(\boldsymbol{\theta}_1)$ 是 $l(\boldsymbol{\theta})$ 的前 q 个分量, 设 $f(\boldsymbol{\beta}_1)$ 是方程 $\dot{Q}_1(\boldsymbol{\theta}_1)=0$ 的解. 若附录中条件 (C1)–(C7) 成立, 那么对 BAR 估计 $\widehat{\boldsymbol{\beta}}=(\widehat{\boldsymbol{\beta}}_1',\widehat{\boldsymbol{\beta}}_2')'$, 依概率 1 有

(a) $\widehat{\boldsymbol{\beta}}_2=0$, $\widehat{\boldsymbol{\beta}}_1$ 存在且为 $f(\boldsymbol{\beta}_1)$ 的唯一不动点.

(b) $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_{01}\right)\xrightarrow{D}N_q(0,V_1(\boldsymbol{\beta}_{01}))$.

证 见附录.

3 模拟实验

3.1 实验设定

本节通过模拟实验来评估有限样本下, BAR 及其他常见惩罚方法在本文假定模型下的表现. 为此考虑以下几组实验: 前两组考虑协变量的来源分布复杂的情况; 后两组考虑稀疏性较强的情况.

第 1 组 设协变量个数 $p=8$, 其中 $\boldsymbol{Z}_1$ 和 $\boldsymbol{Z}_2$ 由标准正态分布随机生成, $\boldsymbol{Z}_3$, $\boldsymbol{Z}_5$ 和 $\boldsymbol{Z}_7$ 由参数为 1 的泊松分布随机生成, $\boldsymbol{Z}_4$, $\boldsymbol{Z}_6$ 和 $\boldsymbol{Z}_8$ 由参数为 1 的指数分布随机生成, 各协变量相互独立. 失效时间 T_i 的风险函数 (2.1) 式中 $\lambda_0\left(t\right)=2$, 参数真值 $\boldsymbol{\beta}_0=(1,0,1,0,1,0,1,0)$. 本组实验考虑观测时间与协变量无关的情况, 设 $C_i\sim U\left(0,1/3\right)$, 数据右删失率约为 50%. 利用

坐标下降法求出 (2.7) 式的极大值点 $\widehat{\boldsymbol{\beta}}$. 重复实验 100 次. 各惩罚函数中的待定调节参数 λ_n 均根据 BIC 准则确定.

第 2 组 在第一组的基础上, 考虑观测时间与协变量相关的情况. 由观测时间 C_i 的风险函数 (2.2) 式, 令 C_i 由参数为 $2\mathrm{e}^{\boldsymbol{\gamma}'\boldsymbol{X}_i}$ 的指数分布随机生成, 并假设 $\boldsymbol{X}_i=\boldsymbol{Z}_i, i=1,2,\cdots,n$. 参数真值设为 $\boldsymbol{\gamma}_0=(0.5,1,0,0,0,0,0.5,0.5)$, 这样能使对失效时间和观测时间有影响的协变量集合不完全相同. 其它设定与第一组相同.

第 3 组 设协变量个数 $p=20$, 各协变量由多元正态分布随机生成, 均值都为 0, 不同协变量 $\boldsymbol{Z}_i$ 和 $\boldsymbol{Z}_k$ 之间的协方差满足为 $\rho^{|i-j|}$, $\rho=0.5$, $i,k=1,2,\cdots,p$. Σ 是协方差阵. 失效时间 T_i 的风险函数 (2.1) 式中 $\lambda_0(t)=0.5$, 参数真值 $\boldsymbol{\beta}_0=(1,1,0,0,\cdots,0,0,1,1)$. 本组实验考虑观测时间与协变量无关的情况, 设 C_i 服从参数为 0.5 的指数分布, 数据右删失率约为 50%.

第 4 组 在第 3 组的基础上, 考虑观测时间与协变量相关的情况. 由观测时间 C_i 的风险函数 (2.2) 式, 令 C_i 由参数为 $0.5\mathrm{e}^{\boldsymbol{\gamma}'\boldsymbol{X}_i}$ 的指数分布随机生成, 参数真值 $\boldsymbol{\gamma}_0=(0.5,1,0,\cdots,0,0.5,0.5)$, 其它设定与第 3 组相同.

我们采用参数估计值的 MSE, TP 和 FP 等指标来评价本文模型下不同惩罚函数的参数估计表现和变量选择表现. 表中 MMSE 和 SD 分别表示重复实验得到的均方误差中位数和标准差, 均方误差 MSE 由 $\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)'\Sigma\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)$ 得到. TP 代表被正确选入模型 (即对应的参数真值非零, 估计的结果也非零) 的协变量数目的平均值. FP 代表被错误选入模型 (即对应的参数真值为零, 估计的结果非零) 的协变量数目的平均值.

3.2 结果分析

在给定的实验设定下, 各组的实验结果见表 1–4.

表 1 第一组实验结果

(Table 1 Simulation results for Case 1)

样本量	方法	MMSE / SD	TP	FP
$n=800$	LASSO	1.1392 / 0.8139	3.74	2.14
	ALASSO	0.7468 / 0.7331	3.72	1.00
	SCAD	0.8964 / 0.8240	3.82	1.60
	SICA	0.8448 / 0.6644	3.88	1.73
	SELO	0.8175 / 0.9218	3.67	1.57
	BAR	0.6539 / 0.6255	3.76	0.59
	Oracle	0.3924 / 0.3628	4	0
$n=1500$	LASSO	0.9040 / 0.6962	3.82	1.80
	ALASSO	0.3845 / 0.6807	3.84	0.30
	SCAD	0.3785 / 0.2385	3.98	1.16
	SICA	0.3839 / 0.4518	3.96	1.11
	SELO	0.4377 / 0.4622	3.94	1.15
	BAR	0.3551 / 0.3408	3.91	0.24
	Oracle	0.1795 / 0.1899	4	0

表 2 第二组实验结果

(Table 2 Simulation results for Case 2)

样本量	方法	MMSE / SD	TP	FP
	LASSO	1.0198 / 0.5464	3.66	1.48
	ALASSO	0.7969 / 0.9933	3.64	0.87
	SCAD	1.0646 / 0.8042	3.80	1.56
$n = 800$	SICA	1.0283 / 0.8939	3.74	1.46
	SELO	0.8743 / 0.9473	3.76	1.68
	BAR	0.7360 / 0.8537	3.79	0.64
	Oracle	0.7100 / 0.6271	4	0
	LASSO	0.5393 / 0.3823	3.78	1.06
	ALASSO	0.3805 / 0.8348	3.78	0.31
	SCAD	0.3680 / 0.5898	3.90	0.60
$n = 1500$	SICA	0.3524 / 0.6321	3.90	0.98
	SELO	0.3228 / 0.6728	3.88	0.75
	BAR	0.3537 / 0.3816	3.95	0.27
	Oracle	0.3587 / 0.4332	4	0

表 3 第三组实验结果

(Table 3 Simulation results for Case 3)

样本量	方法	MMSE / SD	TP	FP
	LASSO	0.7244 / 1.1272	3.78	5.62
	ALASSO	0.2670 / 0.8791	3.84	2.00
	SCAD	0.2208 / 1.0656	3.95	3.03
$n = 400$	SICA	0.1991 / 1.2143	3.90	2.80
	SELO	0.2572 / 0.7710	3.98	2.86
	BAR	0.2247 / 0.8291	4	1.09
	Oracle	0.1648 / 0.3667	4	0
	LASSO	0.4053 / 0.5761	4	2.15
	ALASSO	0.1045 / 0.8544	3.92	0.39
	SCAD	0.0669 / 0.1429	4	0.54
$n = 800$	SICA	0.0801 / 0.6421	3.96	0.45
	SELO	0.0882 / 0.2119	4	0.57
	BAR	0.0781 / 0.2074	4	0.07
	Oracle	0.0959 / 0.1635	4	0

表 4 第四组实验结果

(Table 4 Simulation results for Case 4)

样本量	方法	MMSE / SD	TP	FP
$n = 400$	LASSO	0.4269 / 0.6064	3.94	4.04
	ALASSO	0.1482 / 0.8382	3.80	1.32
	SCAD	0.2010 / 0.7159	4	1.98
	SICA	0.1539 / 0.4280	4	2.12
	SELO	0.2519 / 0.8381	3.95	1.67
	BAR	0.1472 / 0.4001	4	0.40
	Oracle	0.1535 / 0.4297	4	0
$n = 800$	LASSO	0.2634 / 0.5038	4	1.75
	ALASSO	0.0530 / 0.4992	3.84	0.06
	SCAD	0.0841 / 0.2010	4	0.21
	SICA	0.0587 / 0.3652	4	0.20
	SELO	0.0662 / 0.1856	4	0.22
	BAR	0.0775 / 0.2201	4	0.02
	Oracle	0.0695 / 0.1213	4	0

从表 1–4 中可得以下结论: 第一, 各惩罚方法都能较好地筛选出正确变量, 功效高, TP 接近真实值. 第二, 随着样本量的增大, 各方法的 MMSE 和 SD 明显减小, TP 和 FP 也更接近 Oracle 模型的表现, 说明参数估计和变量选择表现都随样本量增大而有明显提升. 第三, 从参数估计表现来看, 在多数情况下, LASSO 方法的 MMSE 都明显大于其他方法, 而其他方法之间的差距并不明显, BAR 方法在多数情况下的 MMSE 较小. 第四, 不论是在哪一种样本量下, 各方法的 TP 都比较接近, BAR 方法的 FP 都是最小的. 这说明在各组模拟实验设定下, BAR 方法剔除错误变量也就是控制假阳性方面表现最优. 其余方法虽在保留正确变量方面效果较好, 但假阳性率较高.

另外, 在模拟过程中我们发现了几个值得注意的现象: 第一, LASSO 方法得到的参数估计绝对值普遍偏小, MSE 值偏大, 这正是 LASSO 估计对真值非零的参数过度压缩的后果, 也是大样本性质较弱的体现, 其他如 SCAD 和 SELO 等方法的估计值普遍更加接近真值. 第二, SCAD, SICA 和 SELO 方法保留了相对多的错误变量, 可能的原因是这些方法都含有不止一个调节参数, 且受调节参数的影响较大, 导致其在本文模型设定下的变量选择表现不够好. 第三, 在后两组实验中, 各方法的 MMSE 普遍偏小, 但部分方法的 MSE 的标准差偏大, 这是因为其在重复实验中存在将所有参数全部估成零的极端情况, 但其在绝大多数实验中下都有较好的变量选择表现.

除上述指标外, 我们也将第一组和第二组的 100 次重复实验中各协变量的保留情况展示出来, 以分析不同类型的协变量是否对变量选择有影响.

从表 5 和表 6 可以看出, 随着样本量的增大, 各协变量被正确保留的次数几乎都会增大, 被错误保留的次数几乎都会减少. 在同一样本量下, 各协变量被保留的次数接近, 并没有因为生成协变量的分布不同而对变量选择有明显影响.

表 5 第一组实验的变量保留次数

(Table 5 The number of covariates retained in Case 1)

样本量	协变量	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
$n=800$	LASSO	96	60	96	42	92	62	90	50
	ALASSO	92	30	94	20	94	32	92	18
	SCAD	96	32	96	44	94	40	96	44
	SICA	99	32	95	47	97	51	97	43
	SELO	93	39	90	38	93	41	91	39
	BAR	95	13	94	20	93	12	94	14
$n=1500$	LASSO	96	50	94	46	96	52	96	32
	ALASSO	96	7	95	12	96	7	97	4
	SCAD	100	30	98	24	100	30	100	20
	SICA	99	23	99	30	99	34	99	24
	SELO	99	36	99	24	98	29	98	26
	BAR	99	3	97	9	98	6	97	6

表 6 第二组实验的变量保留次数

(Table 6 The number of covariates retained in Case 2)

样本量	协变量	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
$n=800$	LASSO	93	30	93	42	93	38	87	38
	ALASSO	94	17	93	22	93	24	84	24
	SCAD	98	30	94	36	92	56	96	34
	SICA	94	36	94	36	94	30	92	44
	SELO	94	42	96	34	96	50	96	42
	BAR	99	18	97	16	98	13	85	17
$n=1500$	LASSO	95	24	95	27	95	38	93	17
	ALASSO	95	11	95	7	95	6	93	7
	SCAD	98	14	98	18	98	18	96	10
	SICA	98	22	98	26	98	20	96	30
	SELO	97	17	97	22	97	13	97	22
	BAR	99	9	100	3	100	6	96	9

4 实证分析

4.1 数据介绍

前面提出的可加风险模型下的惩罚变量选择方法在模拟数据上表现良好, 下面利用这些方法进行实证分析. ADNI 是一项阿尔茨海默病神经影像研究 (the Alzheimer's Disease Neuroimaging Initiative), 通过对影像, 生化标志物和遗传信息的研究来对早期阿尔茨海默病患者进行检测和跟踪. 医学上通常把人的认知状况分为三个时期: CN (正常认知)、MCI (轻度认知障碍) 和 AD (阿尔茨海默病). 处于 MCI 的患者中有一部分发展为 AD, 另一部分并

不会继续恶化, 甚至恢复为 CN. 因此, 探究导致病情恶化的因素对该疾病的防治是至关重要的.

在这项研究中, 实验人员不定期地对参与实验的 MCI 患者进行观测, 记录下他们的认知状况及磁共振成像结果. 我们感兴趣的是从研究开始到 AD 发病的时间, 但发病时间无法被准确观测到. 因此, 若把发病时间看作失效时间 T, 把最后一次观测时间记为 C, 则可以认为这是一组当前状态数据. 数据包含 299 个有效样本, 其中 162 个样本的 $\delta_i = 1$, 右删失数据占比约 54%. 基于以往对 ADNI 数据集的研究 [21, 22], 我们提取了 24 个初始协变量, 其中一部分属于人口统计学与基因信息, 包括研究开始时参与者的年龄 (Age), 性别 (Male), 受教育年限 (PTEDUCAT), 婚姻状况 (Married) 和载脂蛋白基因型 (APOE4). 另一部分属于临床因素, 包括阿尔茨海默病评估量表的得分 (ADAS11 和 ADAS13), 延时词语回忆测试得分 (ADASQ4), 临床痴呆评分表的得分和 (CDRSB), 简单精神状态测试得分 (MMSE), Rey 听觉语言学习测试得分 (RAVLT_i, RAVLT_l, RAVLT_f, RAVLT_p), 数字符号替换测试得分 (DIGITSCOR), B 测试得分 (TRABSCOR) 和功能评估问卷得分 (FAQ). 最后一部分为参与者的磁共振成像体积数据, 包括脑室 (Ventricles), 海马 (Hippocampus), 全脑 (WholeBrain), 内嗅 (Entorhinal), 梭状回 (Fusiform), 颞中回 (MidTemp) 和脑内体积 (ICV).

与模拟实验的过程类似, 我们将连续协变量标准化后, 采用惩罚变量选择方法进行参数估计和变量选择, 结果如表 7 所示, 表中括号内的数值为采用 Bootstrap 方法进行 100 次重抽样计算出的参数标准差估计值.

表 7 ADNI 数据的参数估计及变量选择结果

(Table 7 Results of parameter estimation and variable selection for ADNI data)

协变量	LASSO	ALASSO	SCAD	SICA	SELO	BAR
Male	-	-	-	-	-	-
Married	-	-	-	-	-	-
Age	−0.0334(0.0365)	−0.0207(0.0497)	−0.0282(0.0454)	−0.0415(0.0590)	−0.0439(0.0463)	-
PTEDUCAT	-	-	-	-	-	-
APOE4	0.0438(0.0295)	0.0453(0.0400)	0.0410(0.0386)	-	-	0.0217(0.0276)
ADAS11	0.0544(0.0535)	-	0.0573(0.0640)	-	0.0383(0.0682)	-
ADAS13	-	0.0538(0.1196)	-	0.0800(0.0845)	0.0458(0.0845)	0.0427(0.0626)
ADASQ4	-	-	-	-	-	-
CDRSB	-	-	-	-	-	-
MMSE	-	-	-	-	-	-
RAVLT_*i*	−0.0573(0.0339)	−0.0210(0.0470)	−0.0497(0.0488)	−0.0610(0.0406)	−0.0639(0.0463)	-
RAVLT_l	0.0378(0.0314)	0.0397(0.0416)	0.0291(0.0421)	0.0502(0.0554)	0.0558(0.0464)	-
RAVLT_*f*	−0.0148(0.0408)	−0.0670(0.0818)	-	-	-	-
RAVLT_*p*	-	0.0800(0.1024)	-	-	-	0.0161(0.0755)
DIGITSCOR	-	-	-	-	-	-
TRABSCOR	-	-	-	-	-	-
FAQ	0.0512(0.0443)	0.0410(0.0540)	0.0430(0.0579)	0.0529(0.0566)	0.0587(0.0574)	0.02030 .0349)
Ventricles	0.0231(0.0287)	-	0.0233(0.0443)	-	-	-
Hippocampus	-	-	-	-	-	-
WholeBrain	0.0343(0.0317)	0.0451(0.0598)	0.0291(0.0792)	-	-	0.0106(0.0329)
Entorhinal	-	-	-	-	-	-
Fusiform	−0.0172(0.0298)	-	−0.0105(0.0401)	-	-	-
MidTemp	−0.0880(0.0446)	−0.1163(0.0860)	−0.0930(0.0792)	−0.1175(0.0988)	−0.1217(0.0863)	−0.0798(0.0609)
ICV	-	-	-	0.0585(0.0487)	0.0629(0.0488)	-

4.2 结果分析

从变量选择方面来说, Age, APOE4, ADAS13, RAVLT_i, RAVLT_l, FAQ, WholeBrain 和 MidTemp 这 8 个协变量被超过半数的方法保留, 而 Male 和 Married 等 10 个协变量被全部方法剔除. BAR 方法的模型最为精简, 仅保留 6 个协变量, 这也符合在 BAR 方法在模拟实验中的最优变量选择表现. SICA 和 SELO 方法的模型保留的变量个数适中, 分别保留了 7 个和 8 个. 其余 3 种方法保留了 10 个及以上的协变量. 从系数解释性的方面来说, 可以得出颞中回 (MidTemp) 磁共振成像体积更大的 MCI 患者恶化成 AD 的风险更小等结论. 类似的结论也被前人的研究成果 [21, 22] 所支持.

Li 等 [22] 将比例风险模型和比例优势模型等半参数转换模型的惩罚变量选择方法应用到 ADNI 数据上. 将我们的结果与之对比, 可以看出在本文模型假设下所得到的最优方法与 Li 等 [22] 在其模型假设下得到的最优方法一致, 同为 BAR 方法.

5 总结与展望

本文研究了当前状态数据在可加风险模型下的变量选择问题, 基于计数过程理论, 用比例风险模型下风险函数的形式来表示可加风险模型下的风险函数, 给出一种重复迭代加权的 BAR 惩罚似然变量选择方法及其理论性质. 在多组模拟实验和实证分析中与其它常用惩罚似然方法进行比较, 验证了本文所提方法的良好效果.

在后续的研究工作中, 可以探究本文构造的各惩罚似然函数的其他大样本性质, 如变量分组特性. 本文转化风险函数的方法能避免估计基准风险函数所带来的误差及计算复杂度, 如果在可加风险模型框架下将此方法拓展到其他类型的区间删失数据, 可能需要重新构造风险函数. 除惩罚变量选择方法外, 常用的降维方法还包括主成分法等投影降维方法, 可以将惩罚变量选择方法与其他降维方法比较. 对高维或超高维数据的降维问题是近年来的学术热点, 在高维问题中, 本文所提模型及方法的表现仍有待研究. 此外, 已经有学者将模型平均与惩罚变量选择方法结合, 并在时间序列数据和经济数据上得到应用 [23]. 但相关方法在删失数据上的研究仍不多见, 因此这也是未来的一个拓展方向.

参 考 文 献

[1] Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 1974, **19**(6): 716–723.

[2] Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*, 1978, **6**(2): 461–464.

[3] Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 1996, **58**(1): 267–288.

[4] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 2001, **96**(456): 1348–1360.

[5] Zou H. The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, 2006, **101**(476): 1418–1429.

[6] Lü J, Fan Y. A unified approach to model selection and sparse recovery using regularized least squares. *The Annals of Statistics*, 2009, **37**(6A): 3498–3528.

[7] Dicker L, Huang B, Lin X. Variable selection and estimation with the seamless-L_0 penalty. *Statistica Sinica*, 2013, **23**(2): 929–962.

[8] Liu Z, Li G. Efficient regularized regression with penalty for variable selection and network construction. *Computational and Mathematical Methods in Medicine*, 2016, **2016**: 3456153.

[9] Tibshirani R. The Lasso method for variable selection in the Cox model. *Statistics in Medicine*, 1997, **16**(4): 385–395.

[10] Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. *Annals of Statistics*, 2002, **30**(1): 74–99.

[11] 曹永秀, 焦雨领, 石跃勇, 等. Cox 比例风险模型中基于 SELO 惩罚函数的变量选择方法. 中国科学: 数学, 2018, **48**(5): 643–660.
(Cao Y X, Jiao Y L, Shi Y Y, et al. Penalized variable selection procedure for Cox proportional hazards model via seamless-L_0 penalty. *Scientia Sinica Mathematica*, 2018, **48**(5): 643–660.)

[12] Leng C, Ma S. Path consistent model selection in additive risk model via Lasso. *Statistics in Medicine*, 2007, **26**(20): 3753–3770.

[13] Lin W, Lü J. High-dimensional sparse additive hazards regression. *Journal of the American Statistical Association*, 2013, **108**(501): 247–264.

[14] Tian GL, Wang M, Song L. Variable selection in the high-dimensional continuous generalized linear model with current status data. *Journal of Applied Statistics*, 2014, **41**(3): 467–483.

[15] Zhao H, Wu Q, Gilbert P B, et al. A regularized estimation approach for case-cohort periodic follow-up studies with an application to HIV vaccine trials. *Biometrical Journal*, 2020, **62**(5): 1176–1191.

[16] Zhao H, Wu Q, Li G, et al. Simultaneous estimation and variable selection for interval-censored data with broken adaptive ridge regression. *Journal of the American Statistical Association*, 2020, **115**(529): 204–216.

[17] Lin D, Oakes D, Ying Z. Additive hazards regression with current status data. *Biometrika*, 1998, **85**(2): 289–298.

[18] Wang L, Sun J, Tong X. Regression analysis of case II interval-censored failure time data with the additive hazards model. *Statistica Sinica*, 2010, **20**(4): 1709–1723.

[19] Jia J, Zhou A, Wei C, et al. The prevalence of mild cognitive impairment and its etiological subtypes in elderly Chinese. *Alzheimers's & Dement*, 2014, **10**(4): 439–447.

[20] Wu Q, Zhao H, Zhu L, et al. Variable selection for high-dimensional partly linear additive Cox model with application to Alzheimer's disease. *Statistics in Medicine*, 2020, **39**(23): 3120–3134.

[21] Li K, Chan W, Doody R S, et al. Prediction of conversion to Alzheimer's disease with longitudinal measures and time-to-event data. *Journal of Alzheimer's Disease*, 2017, **58**(2): 361–371.

[22] Li S, Wu Q, Sun J. Penalized estimation of semiparametric transformation models with interval-censored data and application to Alzheimer's disease. *Statistical Methods in Medical Research*, 2020, **29**(8): 2151–2166.

[23] 张新雨. 基于最小二乘近似的模型平均方法. 中国科学: 数学, 2021, **51**(3): 535–548.
(Zhang X Y. Model averaging by least squares approximation. *Scientia Sinica Mathematica*, 2021, **51**(3): 535–548.)

[24] 张怿瑾, 王成勇. 基于当前状态数据的加法风险模型的自适应 LASSO 回归选择 (英文). 数学杂志, 2021, **41**(3): 189–204.
(Zhang Y J, Wang C Y. Regression slection via the adaptive lasso for current status data under the additive hazards model. *Journal of Mathematics*, 2021, **41**(3): 189–204.)

[25] Kawaguchi E, Suchard M, Liu Z, et al. Scalable sparse Cox's regression for large-scale survival data via broken adaptive ridge. arXiv preprint arXiv: 1712.00561, 2017.

附 录

为证明定理 1, 需引入一些记号和两个引理. 定义 $Q(\boldsymbol{\theta};\boldsymbol{\beta})=-2l(\boldsymbol{\theta})+\lambda_n\boldsymbol{\theta}'D(\boldsymbol{\beta})\boldsymbol{\theta}$, 其中 $D(\boldsymbol{\beta})=\operatorname{diag}\left(\beta_1^{-2},\beta_2^{-2},\cdots,\beta_q^{-2},\beta_{q+1}^{-2},\cdots,\beta_p^{-2}\right)$. $Q(\boldsymbol{\theta};\boldsymbol{\beta})$ 类似于 BAR 方法中的惩罚偏似然函数, $\boldsymbol{\beta}$ 对应 BAR 方法迭代过程中的 $\widehat{\beta}_j^{(k-1)}$, 因此不妨将 $Q(\boldsymbol{\theta};\boldsymbol{\beta})$ 简记为 $Q(\boldsymbol{\theta})$. 记 $Q(\boldsymbol{\theta})$ 的前二阶导数分别为

$$\dot{Q}(\boldsymbol{\theta})=-2\dot{l}(\boldsymbol{\theta})+2\lambda_nD(\boldsymbol{\beta})\boldsymbol{\theta}, \tag{A.1}$$

$$\ddot{Q}(\boldsymbol{\theta})=-2\ddot{l}(\boldsymbol{\theta})+2\lambda_nD(\boldsymbol{\beta}). \tag{A.2}$$

下面给出定理及引理的条件

(C1) $\int_0^1\mathrm{d}H_0(t)<\infty$;

(C2) 存在 $\boldsymbol{\beta}_0$ 的紧邻域 $\mathcal{B}_0$ 和 $\boldsymbol{\gamma}_0$ 的紧邻域 $\mathcal{G}_0$, 使得对 $k=0,1,2$, 存在定义在 $\mathcal{B}_0\times\mathcal{G}_0\times[0,1]$ 上的有界且绝对连续的 $s^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},t)$ 满足 $\sup_{t\in[0,1],\boldsymbol{\beta}\in\mathcal{B}_0,\boldsymbol{\gamma}\in\mathcal{G}_0}\left\|S^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},t)-s^{(k)}(\boldsymbol{\beta},\boldsymbol{\gamma},t)\right\|\to\mathbf{0}$ 几乎必然成立. $\|\cdot\|$ 表示向量的欧几里得范数或矩阵的谱范数. 另外 $s^{(0)}(\boldsymbol{\beta},\boldsymbol{\gamma},t)$ 在 $\mathcal{B}_0\times\mathcal{G}_0\times[0,1]$ 上远离 0;

(C3) $\Omega_{\boldsymbol{\beta}}$ 是 $p\times p$ 的正定阵, 存在常数 $C>1$ 使 $C^{-1}<\lambda_{\min}\left(\Omega_{\boldsymbol{\beta}}\right)\leq\lambda_{\max}\left(\Omega_{\boldsymbol{\beta}}\right)<C$ 对充分大的 n 成立, 其中 $\lambda_{\min}(Q)$ 和 $\lambda_{\max}(Q)$ 表示矩阵 Q 的最小和最大特征值;

(C4) 设 $D_i=\int_0^1\left\{-\boldsymbol{Z}_i^*(s)-\mathrm{e}\left(\boldsymbol{\beta}_0;s\right)\right\}\mathrm{d}M_i(s)$, 对所有 $1\leq j,l\leq p$, 存在常数 K 使得 $\sup_{1\leq i\leq n}E\left(D_{ij}^2D_{il}^2\right)<K<\infty$, 其中 D_{ij} 是 D_i 的第 j 个元素;

(C5) 当 $n\to\infty$ 时, 有 $p^2q/\sqrt{n}\to0$, $\lambda_n/\sqrt{n}\to0$, $\lambda_n^2/\left(p\sqrt{n}\right)\to\infty$ 和 $\lambda_n\sqrt{q}/\sqrt{n}\to0$;

(C6) 存在常数 $0<a_0<a_1<\infty$ 使得 $\left|\boldsymbol{\beta}_{0j}\right|\in\left[a_0,a_1\right],j=1,2,\cdots,q$.

(C7) 迭代初值 $\widehat{\boldsymbol{\beta}}^{(0)}$ 满足 $\left\|\widehat{\boldsymbol{\beta}}^{(0)}-\boldsymbol{\beta}_0\right\|=O_p\left(\sqrt{p/n}\right)$.

注 (C5) 是证明定理 1 的充分条件但不是必要条件. 另外, 很容易找到满足 (C7) 的迭代初值 $\widehat{\boldsymbol{\beta}}^{(0)}$, 如偏似然估计 $\widetilde{\boldsymbol{\beta}}$, 文献 [24] 提出的估计或文献 [25] 采用的岭估计.

引理 1 设 $g(\boldsymbol{\beta})=\left(g_1(\boldsymbol{\beta})',g_2(\boldsymbol{\beta})'\right)'$ 是 $\dot{Q}(\boldsymbol{\theta})=\mathbf{0}$ 的一个解. 常数 $M_0>1$ 且满足 $\boldsymbol{\beta}_{01}\in[1/M_0,M_0]^q$, 给定定义域 $\mathcal{H}_n\equiv\left\{\boldsymbol{\beta}=\left(\boldsymbol{\beta}_1',\boldsymbol{\beta}_2'\right)':\left|\boldsymbol{\beta}_1\right|\in[1/M_0,M_0]^q,\left\|\boldsymbol{\beta}_2\right\|\leq\delta_n\sqrt{p}/\sqrt{n}\right\}$, δ_n 是正实数序列且 $\delta_n\to\infty$, $p\delta_n^2/\lambda_n\to0$, 那么在 (C1)–(C7) 的条件下, 依概率 1 有

(a) $g(\cdot)$ 是 $\mathcal{H}_n$ 到 $\mathcal{H}_n$ 的映射;

(b) 对某个常数 $C_0>1$, 有 $\sup_{\boldsymbol{\beta}\in\mathcal{H}_n}\frac{\|g_2(\boldsymbol{\beta})\|}{\|\boldsymbol{\beta}_2\|}<\frac{1}{C_0}$.

引理 2 记 $Q_1\left(\boldsymbol{\theta}_1\right)=-2l_1\left(\boldsymbol{\theta}_1\right)+\lambda_n\boldsymbol{\theta}_1'D_1\left(\boldsymbol{\beta}_1\right)\boldsymbol{\theta}_1$, 设 $f\left(\boldsymbol{\beta}_1\right)$ 是 $\dot{Q}_1\left(\boldsymbol{\theta}_1\right)=0$ 的一个解, 在 (C1)–(C7) 的条件下, 依概率 1 有: $f\left(\boldsymbol{\beta}_1\right)$ 是 $[1/M_0,M_0]^q$ 到 $[1/M_0,M_0]^q$ 的压缩映射且有唯一的不动点 $\widehat{\boldsymbol{\beta}}_1^{\circ}$.

引理 1 和引理 2 的证明参见文献 [25]. 接下来我们证明不动点 $\widehat{\boldsymbol{\beta}}_1^{\circ}$ 的渐近正态性, 并说明不动点 $\widehat{\boldsymbol{\beta}}_1^{\circ}$ 与 BAR 估计中 $\widehat{\boldsymbol{\beta}}_1$ 的关系, 就能得到定理 1 的结论. 将 $\dot{Q}_1(\boldsymbol{\beta_{01}})$ 在 $f(\boldsymbol{\beta}_1)$ 处进行泰勒一阶展开, 得

$$\dot{Q}_1\left(\boldsymbol{\beta}_{01}\right)=\dot{Q}_1(f(\boldsymbol{\beta}_1))+\ddot{Q}_1\left(\boldsymbol{\beta}_1^*\right)\left(\boldsymbol{\beta}_{01}-f(\boldsymbol{\beta}_1)\right), \tag{A.3}$$

其中 $\boldsymbol{\beta}_1^*$ 介于 $\boldsymbol{\beta}_{01}$ 和 $f(\boldsymbol{\beta}_1)$ 之间. 用 $H_1\left(\boldsymbol{\beta}_1\right)$ 表示 $\widehat{\Omega}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\widehat{\boldsymbol{\gamma}})$ 的前 $q\times q$ 子矩阵, 整理得

$$f\left(\boldsymbol{\beta}_1\right)=\left\{H_1\left(\boldsymbol{\beta}_1^*\right)+\frac{\lambda_n}{n}D_1\left(\boldsymbol{\beta}_1\right)\right\}^{-1}\left\{H_1\left(\boldsymbol{\beta}_1^*\right)\boldsymbol{\beta}_{01}+\frac{1}{n}\dot{l}_1\left(\boldsymbol{\beta}_{01}\right)\right\}. \tag{A.4}$$

由此可将 $\sqrt{n}H_1\left(\boldsymbol{\beta}_1^*\right)^{\frac{1}{2}}\left(\widehat{\boldsymbol{\beta}}_1^{\circ}-\boldsymbol{\beta}_{01}\right)$ 改写为

$$\begin{aligned}
&\sqrt{n}H_1\left(\boldsymbol{\beta}_1^*\right)^{\frac{1}{2}}\left(\widehat{\boldsymbol{\beta}}_1^{\circ}-\boldsymbol{\beta}_{01}\right)\\
=&\sqrt{n}H_1\left(\boldsymbol{\beta}_1^*\right)^{\frac{1}{2}}\left[\left\{H_1\left(\boldsymbol{\beta}_1^*\right)+\frac{\lambda_n}{n}D_1\left(\widehat{\boldsymbol{\beta}}_1^{\circ}\right)\right\}^{-1}H_1\left(\boldsymbol{\beta}_1^*\right)-I_q\right]\boldsymbol{\beta}_{01}\\
&+\sqrt{n}H_1\left(\boldsymbol{\beta}_1^*\right)^{\frac{1}{2}}\left[\left\{H_1\left(\boldsymbol{\beta}_1^*\right)+\frac{\lambda_n}{n}D_1\left(\widehat{\boldsymbol{\beta}}_1^{\circ}\right)\right\}^{-1}\frac{1}{n}\dot{l}_1\left(\boldsymbol{\beta}_{01}\right)\right]\\
=&I_1+I_2. \qquad (A.5)
\end{aligned}$$

先证明 $\|I_1\|\to 0$. 对 I_1 有

$$\begin{aligned}
I_1=&\sqrt{n}H_1\left(\boldsymbol{\beta}_1^*\right)^{\frac{1}{2}}\left[\left\{H_1\left(\boldsymbol{\beta}_1^*\right)+\frac{\lambda_n}{n}D_1\left(\widehat{\boldsymbol{\beta}}_1^{\circ}\right)\right\}^{-1}H_1\left(\boldsymbol{\beta}_1^*\right)-I_q\right]\boldsymbol{\beta}_{01}\\
=&-\frac{\lambda_n}{\sqrt{n}}H_1\left(\boldsymbol{\beta}_1^*\right)^{-\frac{1}{2}}D_1\left(\widehat{\boldsymbol{\beta}}_1^{\circ}\right)\left\{H_1\left(\boldsymbol{\beta}_1^*\right)+\frac{\lambda_n}{n}D_1\left(\widehat{\boldsymbol{\beta}}_1^{\circ}\right)\right\}^{-1}H_1\left(\boldsymbol{\beta}_1^*\right)\boldsymbol{\beta}_{01}, \qquad (A.6)
\end{aligned}$$

这里用到了伍德伯里矩阵恒等式: 两个一致可逆矩阵满足 $(\Phi+\Psi)^{-1}=\Phi^{-1}-\Phi^{-1}\Psi(\Phi+\Psi)^{-1}$. 由 (C3), (C5) 和 (C6), 有

$$\|I_1\|\le\frac{M_0^2\lambda_n}{\sqrt{n}}\left\|H_1\left(\boldsymbol{\beta}_1^*\right)^{\frac{1}{2}}\right\|\|\boldsymbol{\beta}_{01}\|=O_p\left(\lambda_n\sqrt{q}/\sqrt{n}\right)\to 0. \qquad (A.7)$$

类似地,

$$\begin{aligned}
I_2=&\sqrt{n}H_1\left(\boldsymbol{\beta}_1^*\right)^{\frac{1}{2}}\left[\left\{H_1\left(\boldsymbol{\beta}_1^*\right)+\frac{\lambda_n}{n}D_1\left(\widehat{\boldsymbol{\beta}}_1^{\circ}\right)\right\}^{-1}\frac{1}{n}\dot{l}_1\left(\boldsymbol{\beta}_{01}\right)\right]\\
=&H_1\left(\boldsymbol{\beta}_1^*\right)^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\dot{l}_1\left(\boldsymbol{\beta}_{01}\right)-\frac{\lambda_n}{\sqrt{n}}H_1\left(\boldsymbol{\beta}_1^*\right)^{-\frac{1}{2}}D_1\left(\widehat{\boldsymbol{\beta}}_1^{\circ}\right)\left\{H_1^{-1}\left(\boldsymbol{\beta}_1^*\right)+D_1\left(\widehat{\boldsymbol{\beta}}_1^{\circ}\right)\right\}^{-1}\frac{1}{n}\dot{l}_1\left(\boldsymbol{\beta}_{01}\right)\\
=&H_1\left(\boldsymbol{\beta}_1^*\right)^{-\frac{1}{2}}\frac{1}{\sqrt{n}}\dot{l}_1\left(\boldsymbol{\beta}_{01}\right)+o_p(1). \qquad (A.8)
\end{aligned}$$

由文献 [17] 的结论,

$$\frac{1}{\sqrt{n}}\dot{l}_1\left(\boldsymbol{\beta}_{01}\right)\xrightarrow{D}N\left(\mathbf{0},M_1(\boldsymbol{\beta}_{01})\right) \qquad (A.9)$$

以及 (C3), 立即得到 $\sqrt{n}H_1\left(\boldsymbol{\beta}_1^*\right)\left(\widehat{\boldsymbol{\beta}}_1^{\circ}-\boldsymbol{\beta}_{01}\right)\xrightarrow{D}N\left(\mathbf{0},M_1(\boldsymbol{\beta}_{01})\right)$, 进而有 $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_1^{\circ}-\boldsymbol{\beta}_{01}\right)\xrightarrow{D}N\left(\mathbf{0},V_1(\boldsymbol{\beta}_{01})\right)$.

由引理 1 知 $\Pr\left(\widehat{\boldsymbol{\beta}}_2=\lim_{k\to\infty}g_2\left(\boldsymbol{\beta}^{(k)}\right)=\mathbf{0}\right)\to 1$. 下证 $\Pr(\lim_{k\to\infty}\left\|g_1\left(\boldsymbol{\beta}^{(k)}\right)-\widehat{\boldsymbol{\beta}}_1^{\circ}\right\|=0)\to 1$. 根据引理 1 中的定义, $g(\boldsymbol{\beta})$ 是

$$-\frac{1}{n}D(\boldsymbol{\beta})^{-1}\dot{l}_n(\boldsymbol{\theta})+\frac{1}{n}\lambda_n\boldsymbol{\theta}=\mathbf{0} \qquad (A.10)$$

的解, 显然当 $\boldsymbol{\beta}_{02}=\mathbf{0}$ 时 $g_2(\boldsymbol{\beta})=\mathbf{0}$. 因此可以将上式分解为两部分

$$\begin{aligned}
-\frac{1}{n}D_1^{-1}\left(\boldsymbol{\beta}_1\right)\dot{l}_{n1}\left(\boldsymbol{\theta}_1\right)+\frac{1}{n}\lambda_n\boldsymbol{\theta}_1=\mathbf{0},\\
-\frac{1}{n}D_2^{-1}\left(\boldsymbol{\beta}_2\right)\dot{l}_{n2}\left(\boldsymbol{\theta}_2\right)+\frac{1}{n}\lambda_n\boldsymbol{\theta}_2=\mathbf{0},
\end{aligned}$$

且 $g_1(\boldsymbol{\beta})$ 和 $g_2(\boldsymbol{\beta})$ 分别是这两个方程的解. 由引理 1 和引理 2 的结论可以得到

$$\lim_{\boldsymbol{\beta}_2\to 0} g_2\left(\boldsymbol{\beta}_2\right)=0,\quad \lim_{\boldsymbol{\beta}_2\to 0} g_1\left(\boldsymbol{\beta}_1\right)=f\left(\boldsymbol{\beta}_1\right),$$

可见 $g\left(\cdot\right)$ 在 $\boldsymbol{\beta}\in\mathcal{H}_n$ 上是连续的. 由 $g\left(\cdot\right)$ 的连续性, 在 $\boldsymbol{\beta}_2^{(k)}\to 0$ 时依概率 1 有

$$\omega_k\equiv\sup_{g_1(\boldsymbol{\beta})\in[1/M_0,M_0]^q}\left\|f\left(\boldsymbol{\beta}_1\right)-g_1\left(\boldsymbol{\beta}_1\right)\right\|\to 0. \tag{A.11}$$

由 $\widehat{\boldsymbol{\beta}}_1^{\circ}$ 是压缩映射 $f\left(\cdot\right)$ 的唯一不动点及压缩映射 $f\left(\cdot\right)$ 的性质 $\sup_{|\boldsymbol{\beta}_1|\in[1/M_0,M_0]^q}\left\|\dot{f}\left(\boldsymbol{\beta}_1\right)\right\|=o_p(1)$, 可得对某个 $C_1>1$, 有

$$\left\|f\left(\widehat{\boldsymbol{\beta}}_1^{(k)}\right)-\widehat{\boldsymbol{\beta}}_1^{\circ}\right\|=\left\|f\left(\widehat{\boldsymbol{\beta}}_1^{(k)}\right)-f\left(\widehat{\boldsymbol{\beta}}_1^{\circ}\right)\right\|\le\frac{1}{C_1}\left\|\widehat{\boldsymbol{\beta}}_1^{(k)}-\widehat{\boldsymbol{\beta}}_1^{\circ}\right\|, \tag{A.12}$$

进一步地, 有

$$\left\|\widehat{\boldsymbol{\beta}}_1^{(k+1)}-\widehat{\boldsymbol{\beta}}_1^{\circ}\right\|\le\left\|g_1\left(\widehat{\boldsymbol{\beta}}^{(k)}\right)-\widehat{\boldsymbol{\beta}}_1^{\circ}\right\|\le\left\|g_1\left(\widehat{\boldsymbol{\beta}}^{(k)}\right)-f\left(\widehat{\boldsymbol{\beta}}_1^{(k)}\right)\right\|+\left\|f\left(\widehat{\boldsymbol{\beta}}_1^{(k)}\right)-\widehat{\boldsymbol{\beta}}_1^{\circ}\right\|, \tag{A.13}$$

因此有

$$\left\|\widehat{\boldsymbol{\beta}}_1^{(k+1)}-\widehat{\boldsymbol{\beta}}_1^{\circ}\right\|\le\frac{1}{C_1}\left\|\widehat{\boldsymbol{\beta}}_1^{(k)}-\widehat{\boldsymbol{\beta}}_1^{\circ}\right\|+\omega_k, \tag{A.14}$$

经过递归计算可得

$$\Pr\left(\lim_{k\to\infty}\left\|\widehat{\boldsymbol{\beta}}_1^{(k)}-\widehat{\boldsymbol{\beta}}_1^{\circ}\right\|=\mathbf{0}\right)=1, \tag{A.15}$$

即 $\Pr\left(\widehat{\boldsymbol{\beta}}_1=\widehat{\boldsymbol{\beta}}_1^{\circ}\right)=1$. 至此已经得到 $\widehat{\boldsymbol{\beta}}=\left(\widehat{\boldsymbol{\beta}}_1',\widehat{\boldsymbol{\beta}}_2'\right)'\equiv\lim_{k\to\infty}\widehat{\boldsymbol{\beta}}^{(k)}=\left(g_1(\boldsymbol{\beta}^{(k)})',g_2(\boldsymbol{\beta}^{(k)})'\right)'=\left(\widehat{\boldsymbol{\beta}}_1^{\circ\prime},0'\right)'$, 定理 1 的 (a) 证毕, 定理 1 的 (b) 可由不动点 $\widehat{\boldsymbol{\beta}}_1^{\circ}$ 的渐近正态性得到. 定理 1 证毕.