弹性约束估计的显著性检验及其渐近分布

杨玥含^{1*}, 吴岚²

1. 中央财经大学统计与数学学院, 北京 100871;

2. 北京大学数学科学学院, 北京 100081

E-mail: yyh@cufe.edu.cn, lwu@pku.edu.cn

收稿日期: 2016-06-29; 接受日期: 2018-11-11; 网络出版日期: 2019-05-15; * 通信作者
国家自然科学基金 (批准号: 11671059) 资助项目

摘要 本文基于高维稀疏线性模型, 研究弹性约束估计 (elastic net, EN) 的相关显著性检验问题, 在弹性约束估计的解路径上建立 Cov-EN 检验. 为了获取该检验的理论结果, 本文回顾 KKT (Karush-Kuhn-Tucker) 条件, 通过 Lars 算法计算得到弹性约束估计的解路径上每个节点的解析表达式, 证明该检验在一般数据下渐近收敛于参数为 1 的指数分布. 本文的数值模拟和实证分析进一步阐述 Cov-EN 检验的特点与作用, 并与 Lasso 的协方差检验进行比较.

关键词 显著性检验 模型选择 协方差检验

MSC (2010) 主题分类 62J05, 62J07

1 引言

给定线性回归模型如下:

$$y = X\beta + \epsilon, \quad (1.1)$$

其中 y 为 $n \times 1$ 维的响应变量, X 表示 $n \times p$ 维设计矩阵 (或称作预测矩阵), $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ 是独立同分布的随机变量, 服从均值为 0、方差为 σ^2 的正态分布, β 表示 $p \times 1$ 的系数向量. 考虑回归模型是 q 维稀疏模型, 即在样本数量为 n 、变量数量为 p 的模型中, 真实系数 β 至多有 q 个非零项.

模型选择是高维数据线性回归分析里的关键问题, 统计学家研究并提出了多种罚项正则化方法, 如 Lasso、自适应 Lasso 和弹性约束估计等 (参见文献 [1–7]). 这一类方法易于计算, 且具有极佳的理论性质, 得到了广泛应用. 罚项正则化方法在模型选择理论、算法方面已积累了大量的研究和进展, 但对该类方法的统计推断领域的研究仍有缺失, 例如, p 值和置信区间等有效的检验手段在罚项正则化方法中讨论不足, 目前已有的研究大多需要对数据进行重抽样或分割.

英文引用格式: Yang Y H, Wu L. A significance test for the elastic net and its asymptotic distribution with general predictors (in Chinese). Sci Sin Math, 2019, 49: 1119–1138, doi: 10.1360/SCM-2016-0427

部分文献总结了模型选择过程的统计特征, 例如, 文献 [8–11] 在重抽样和数据分割的基础上构造了相关方法; 文献 [12] 针对 Lasso 构造了统计显著性检验, 该统计量可以粗略看作是 Lasso 拟合值的协方差, 因此被称为协方差检验. 弹性约束估计是应用广泛的模型选择方法, 并因它处理相关性数据的表现优于 Lasso 而广受欢迎, 然而目前还没有对该方法的显著性检验的研究, 因此, 本文在已有工作的基础上, 研究模型选择过程的显著性检验问题, 并针对罚项正则化方法弹性约束估计设计了有效的检验.

首先, 基于线性回归模型 (1.1) 假设, 给出弹性约束估计的估计量如下:

$$\hat{\beta}(\lambda_1, \lambda_2) = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2} \|y - X\beta\|_2^2 + \frac{1}{2} \lambda_2 \|\beta\|_2^2 + \lambda_1 \|\beta\|_1 \right\}, \quad (1.2)$$

其中 $\lambda_1, \lambda_2 > 0$ 分别是一范罚项和二范罚项的调整参数, 本文称之为正则参数. 本文所构造的弹性约束估计在模型选择时的显著性检验, 描述了弹性约束估计解路径上的各节点的拟合值与真实观测值之间的协方差的变化, 因此被称为 Cov-EN 检验. 该想法的灵感来自于文献 [12], 我们的检验针对弹性约束估计设计, 由于弹性约束估计用于处理高维相关性数据, 目标函数包含 l_2 罚项及其正则参数, 相关检验及理论论证更加复杂. 本文通过提出 Cov-EN 统计量检验了变量被选入模型时的显著性, 为了获得 Cov-EN 检验的理论结果, 使用 Lars 算法^[13] 计算弹性约束估计的解路径上每个节点的解析表达式, 论证该统计量渐近收敛于参数为 1 的指数分布, 并在进一步模拟算例和实证分析中将弹性约束估计与 Lasso 的检验效果进行对比.

2 弹性约束估计的解及节点表达式

首先, 在 KKT 条件下, $\hat{\beta}$ 是弹性约束估计的解当且仅当其满足以下等式:

$$X^T(y - X\hat{\beta}) = \lambda_2 \hat{\beta} + \lambda_1 \gamma, \quad (2.1)$$

其中 $\gamma = (\gamma_1, \dots, \gamma_p)^T$ 是函数 $f(x) = \|x\|_1$ 的次梯度 (subgradient) 参数, 且满足

$$\gamma_i \in \begin{cases} \{\text{sign}(\hat{\beta}_i)\}, & \text{若 } \hat{\beta}_i \neq 0, \\ [-1, 1], & \text{若 } \hat{\beta}_i = 0. \end{cases}$$

由 (1.2) 可知, 弹性约束估计的目标函数是严格凸的, 因此, 总是存在唯一解. 与之相比, Lasso 的目标函数是凸函数, 如果 Lasso 想要获得唯一解, 需要对预测矩阵有一定的约束假设, 如 “常规设定” (general position), 详情参见文献 [14, 15]. 因此, 定义等相关集合 $A = \{i \in \{1, \dots, p\} : |X^T(y - X\hat{\beta}) - \lambda_2 \hat{\beta}| = \lambda_1\}$, 弹性约束估计有以下解析解.

引理 2.1 通过回归模型 (1.1) 给定 y 和 X , 对任意 $\lambda_1, \lambda_2 > 0$, 弹性约束估计 (1.2) 的解为

$$\hat{\beta}_{A^c} = 0, \quad \hat{\beta}_A = (X_A^T X_A + \lambda_2 I)^{-1} (X_A^T y - \lambda_1 s_A),$$

其中 $s_A = \text{sign}(\hat{\beta}_A)$, $\text{sign}(\cdot)$ 表示对象的符号.

Lars 算法可用于求解带有 l_1 罚项的正则化方法, 该算法由 Efron 等^[13] 提出, 为罚项正则化方法的研究和推广作出了卓越的贡献. 弹性约束估计的解同样可由改良后的 Lars 算法计算获得, 本文省略文献已有的算法细节, 进一步推导弹性约束估计的解路径上各个节点的解析表达式, 更具体地, 对任意正整数 k , 研究弹性约束估计的解路径在第 k 个节点时的表达式, 以及该节点正则参数 λ_1 的解析解.

给定 λ_2 , 弹性约束估计在 Lars 算法中的解以 $\lambda_1 = \infty$ 为起始点, 此时所有变量的系数估计值被压缩至 0, 包含有效变量的集合为空集. 此后, λ_1 逐步降低, 变量逐个被选入有效变量集合, 或在某节点从该集合剔除. 令 $\lambda_{1,1}, \dots, \lambda_{1,k}, \dots$ 表示每个节点上的正则参数值, 并设 A 为第 k 个节点之前的有效变量集合, 则有 $\beta_{A^c}(\lambda_{1,k}, \lambda_2) = 0$. 设 s_A 表示选入变量的符号, 通过引理 2.1 可得第 k 个节点上的估计值 $\hat{\beta}_A$ 如下:

$$\hat{\beta}_A(\lambda_{1,k}, \lambda_2) = (X_A^T X_A + \lambda_2 I)^{-1} (X_A^T y - \lambda_1 s_A) = F_A y - \lambda_{1,k} C_A^{-1} s_A, \quad (2.2)$$

其中 $C_A^{-1} = (X_A^T X_A + \lambda_2 I)^{-1}$, $F_A = (X_A^T X_A + \lambda_2 I)^{-1} X_A^T$. 将 (2.2) 看作关于 $\lambda_{1,k}$ 的线性函数, 当 λ_1 降低到下一个节点 $\lambda_{1,(k+1)}$ 时, 会出现两种情形: (1) 某一个 A 以外的变量被选入集合 A ; (2) A 中的某个变量被剔除送还 A^c . 第一种情形下, 称该节点为“加入节点”, 记为 $\lambda_{1,(k+1)}^{\text{join}}$, 第二种情形则称之为“离开节点”, 记为 $\lambda_{1,(k+1)}^{\text{leave}}$. 因此, 第 $k+1$ 个节点的 λ_1 取值为

$$\lambda_{1,(k+1)} = \max\{\lambda_{1,(k+1)}^{\text{join}}, \lambda_{1,(k+1)}^{\text{leave}}\}.$$

考虑第一种情形, 对任意的 A^c 中的变量进入有效集合 A , 由 (2.1) 可得以下等式:

$$X_j^T (y - X_A \hat{\beta}_A) = \pm \lambda_{1,(k+1)}, \quad \text{其中 } j \notin A,$$

当 $j \notin A$ 时, $\lambda_2 \hat{\beta}_j = 0$, 因此, 上述等式成立. 通过计算, “加入节点”的 $\lambda_{1,(k+1)}$ 的值为

$$\lambda_{1,(k+1)}^{\text{join}} = \max_{j \in A^c, s \in \{-1, 1\}} \frac{X_j^T (I - R_A) y}{s - X_j^T F_A^T s_A} \cdot 1 \left\{ \frac{X_j^T (I - R_A) y}{s - X_j^T F_A^T s_A} < \lambda_{1,k} \right\}, \quad (2.3)$$

其中 $R_A = X_A (X_A^T X_A + \lambda_2 I)^{-1} X_A^T$. 类似地, 可计算得到“离开节点”的 $\lambda_{1,(k+1)}$ 的取值如下:

$$\lambda_{1,(k+1)}^{\text{leave}} = \max_{j \in A} \frac{[F_A y]_j}{[C_A^{-1} s_A]_j} \cdot 1 \left\{ \frac{[F_A y]_j}{[C_A^{-1} s_A]_j} < \lambda_{1,k} \right\}.$$

如果 $\lambda_{1,(k+1)}^{\text{join}} > \lambda_{1,(k+1)}^{\text{leave}}$, 将下标为 j^{join} 的变量加入有效集合 A , 并将该变量对应的系数估计值符号 s^{join} 加入有效符号集 s . 否则, 从集合 A 中删除下标为 j^{leave} 的变量, 并从符号集 s 中删去该变量对应的系数估计值符号 s^{leave} .

3 弹性约束估计的 Cov-EN 检验

3.1 定义与假设

参考上节设定, 令 $\lambda_{1,k}$ 和 $\lambda_{1,(k+1)}$ 表示 λ_1 在第 k 和 $k+1$ 个节点上的取值, A 表示第 k 个节点之前的有效变量集合, $\hat{\beta}(\lambda_{1,(k+1)}, \lambda_2)$ 表示弹性约束估计在其解路径的第 $k+1$ 个节点上的解, 该节点的有效集合是 $A \cup \{j\}$ 、 $j \notin A$ 或 $A/\{i\}$, $i \in A$. 回顾上节结论与符号, 在第 k 个节点上的 $\hat{\beta}_A$ 为

$$\hat{\beta}_A(\lambda_{1,k}, \lambda_2) = F_A y - \lambda_{1,k} C_A^{-1} s_A.$$

同理, 易得在第 $k+1$ 个节点上的 $\hat{\beta}(\lambda_{1,(k+1)}, \lambda_2)$ 为

$$\hat{\beta}_{A \cup \{j\}}(\lambda_{1,(k+1)}, \lambda_2) = F_{A \cup \{j\}} y - \lambda_{1,(k+1)} C_{A \cup \{j\}}^{-1} s_{A \cup \{j\}}.$$

另一方面, 设 $\tilde{\beta}_A(\lambda_{1,(k+1)}, \lambda_2)$ 为只使用集合 A 中的变量得到的弹性约束估计在第 $k+1$ 个节点上的解, 则有

$$\tilde{\beta}_A(\lambda_{1,(k+1)}, \lambda_2) = F_A y - \lambda_{1,(k+1)} C_A^{-1} s_A.$$

由此, 本文定义 Cov-EN 检验统计量如下:

$$T_k^{\text{EN}} = \frac{y^T X \hat{\beta}(\lambda_{1,(k+1)}, \lambda_2) - y^T X \tilde{\beta}_A(\lambda_{1,(k+1)}, \lambda_2)}{\sigma^2}. \quad (3.1)$$

该统计量考虑 $X \hat{\beta}(\lambda_{1,(k+1)}, \lambda_2)$ 和 $X_A \tilde{\beta}_A(\lambda_{1,(k+1)}, \lambda_2)$ 二者与 y 内积之差, 与其协方差表达式相似, 因此取名为 Cov-EN 统计量. 同时, 假设第 k 个节点为“加入节点”, 即没有变量被删除, 可获得如下等式:

$$\text{sign}(\hat{\beta}_A(\lambda_{1,k}, \lambda_2)) = \text{sign}(\tilde{\beta}_A(\lambda_{1,(k+1)}, \lambda_2)). \quad (3.2)$$

当 $k=1$ 时, 该假设显然成立. 当 X 满足正共轭条件 (包含 X 正交) 时, 在弹性约束估计的解路径中, 任意变量被选入有效集合后, 都不会被删除, 因此, 该假设也成立. 在其余情形, 该假设意味着 Cov-EN 检验只研究“加入节点”, 不考虑“离开节点”. 因此, 进一步获得 Cov-EN 检验的解析表达式.

定理 3.1 假设满足条件 (3.2), 弹性约束估计的解的 Cov-EN 检验统计量 (3.1) 可表示为

$$T_k^{\text{EN}} = \frac{\lambda_{1,k}(\lambda_{1,k} - \lambda_{1,(k+1)})}{\sigma^2} C(A, s_A, j, s) + \lambda_2 \frac{(\lambda_{1,k} - \lambda_{1,(k+1)})}{\sigma^2 \lambda_{1,k}} D(A, j, s, y),$$

其中

$$\begin{aligned} C(A, s_A, j, s) &= \|F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A\|_2^2, \\ D(A, j, s, y) &= \|F_{A \cup \{j\}} y - F_A y\|_2^2. \end{aligned}$$

为了得到该检验统计量在一般数据下的理论性质, 本文需要如下假设: 令 $A^* = \text{supp}(\beta)$ 表示真实的 β 中非零项的集合, 对任意给定的正整数 k_0 , A_0 表示该节点上的有效集合, 假设其包含真实变量集 A^* . 该假设可通过弹性约束估计的模型选择一致性来获得, 即在满足以下条件且样本量增加时, 弹性约束估计将大概率地选择正确的模型.

条件 (a): 存在常数 $0 < \eta < 1$, 对所有的 $j \notin A_0$, 下列不等式成立:

$$\|F_{A_0 \cup \{j\}} X_j\|_1 \leq 1 - \eta. \quad (3.3)$$

条件 (b): 对任意给定的 $j \notin A_0$, 存在集合 $S \subseteq \{1, \dots, p\} \setminus (A_0 \cup \{j\})$, 该集合 S 中元素个数 d_p 大于 $\log p$, 即

$$|S| \geq d_p, \quad \text{其中 } \frac{d_p}{\log p} \rightarrow \infty, \quad \text{当 } p \rightarrow \infty. \quad (3.4)$$

条件 (c): 定义矩阵 $G_{ij} = X_i^T (I - P_{A_0}) X_j$, $i, j \notin A_0$, 设该矩阵 G 的对角元素同阶, 存在常数 δ , 对所有的 $i \in S$, 有

$$\frac{|G_{ij}|}{G_{jj}} < \frac{\eta}{2 - \eta}, \quad (3.5)$$

$$\|F_{A_0 \cup \{j\}} X_j\|_1 < 1, \quad (3.6)$$

$$1 - G_{i, S \setminus \{i\}} (G_{S \setminus \{i\}, S \setminus \{i\}} + \lambda_2 I)^{-1} G_{S \setminus \{i\}, i} \geq \delta^2, \quad \forall i \in S, \quad (3.7)$$

$$\lambda_2 [s_{A_0 \cup \{j\}}^T (X_{A_0 \cup \{j\}}^T X_{A_0 \cup \{j\}} + \lambda_2 I)^{-2} s_{A_0 \cup \{j\}} - s_{A_0}^T (X_{A_0}^T X_{A_0} + \lambda_2 I)^{-2} s_{A_0}] \geq 0. \quad (3.8)$$

条件 (a) 与弹性约束估计获取模型选择一致性所需的不可表示性条件类似, 该条件是弹性约束估计获得模型选择一致性的关键条件 (参见文献 [16]). 条件 (b) 使得子集 S 大于 $\log p$. 条件 (c) 的 (3.5) 和 (3.6) 用于使 $R_{ij} < 1$. (3.6) 可以看作弱化的不可表示性条件. 条件 (c) 的 (3.7) 可以近似看作 $X_j y$ 的方差的下界. 条件 (c) 的 (3.8) 用于证明附录 A 的引理 A.3. 以上条件均是罚项正则化方法的常规条件, 曾多次在文献中被使用, 如文献 [12, 16] 等. 在以上条件的帮助下, 本文在后文展开对 Cov-EN 检验的理论性质的研究.

3.2 第一个节点

本节讨论在第一个节点处 Cov-EN 检验的渐近分布, 为了表达方便, 本节假设 X 满足 $\|X_i\|_2 = 1$, $i = 1, \dots, p$. 该假设可通过对预测矩阵标准化获得. 首先, 通过定理 3.1 以及弹性约束估计在解路径上的解析表达式, 可以得到

$$D(A, j, s, y) = \frac{\lambda_{1,1}^2}{(1 + \lambda_2)^2},$$

以及

$$C(A, s_A, j, s) = \frac{1}{(1 + \lambda_2)^2}.$$

因此, 经过简单计算, Cov-EN 检验在第一个节点上的表达式为

$$T_1^{\text{EN}} = \frac{1}{1 + \lambda_2} \cdot \frac{\lambda_{1,1}(\lambda_{1,1} - \lambda_{1,2})}{\sigma^2}, \quad (3.9)$$

其中 $\lambda_{1,1}$ 和 $\lambda_{1,2}$ 分别表示正则参数 λ_1 在弹性约束估计解路径上的第一和二节点的取值, 这两个取值的具体表达式在以下引理中给出.

引理 3.1 令 $U_j = X_j^T y$, $j = 1, \dots, p$, $C = X^T X$. 并设 j_1 和 s_1 表示第一个被选中变量的下标及对应系数估计的符号, 则有

$$\begin{aligned} \lambda_{1,1} &= \max_{j=1, \dots, p, s \in \{-1, 1\}} s U_j, \\ \lambda_{1,2} &= \max_{j \neq j_1, s \in \{-1, 1\}} \frac{s U_j - \frac{1}{1 + \lambda_2} s C_{j, j_1} U_{j_1}}{1 - \frac{1}{1 + \lambda_2} s s_1 C_{j, j_1}}. \end{aligned}$$

由于对任意的 $j \neq j_1$, 有 $|C_{j, j_1}| < 1$, 且 $1 - s s_1 C_{j, j_1} > 0$, 其中 $s \in \{-1, 1\}$, (2.3) 中的示性函数可以省略.

上述结果可以通过 (2.3) 直接计算获得. 在第一个节点之前, 有效变量集合为空集, 即当 $\lambda_1 = \infty$ 时, $A = \emptyset$. 因此, 在第一个节点处, 下标为 j_1 的变量被选中后, 无法立刻离开有效集合, 满足该检验对“加入节点”的假设.

给定零假设 $H_0 : y \sim N(0, \sigma^2 I)$. 针对 Cov-EN 检验在第一个节点处的表达式 (3.9), 仅需要条件 (b) 和部分条件 (c) 就可得到该检验的渐近分布, 如下所示.

定理 3.2 假设条件 (b) 和 (c) 的 (3.7) 成立, 则在零假设下, 有

$$P(T_1^{\text{EN}} > t) \rightarrow e^{-(1+\lambda_2)t}, \quad \text{当 } p \rightarrow \infty, \quad \forall t \geq 0.$$

3.2.1 第 k 个节点 ($k \geq 1$)

考虑 Cov-EN 检验的一般情形, 对任意给定的 $k_0 \geq 0$, 记解路径的第 k_0 个节点处的有效变量集合为 A_0 . 令 $\lambda_{1,(k_0+1)}$ 和 $\lambda_{1,(k_0+2)}$ 分别表示第 k_0+1 和 k_0+2 个节点上的正则参数的值, 我们需要下列重要定义.

定义 3.1 (k_0 节点假设 ($B_{k_0}^{\text{EN}}$)) 对任意整数 $k_0 \geq 0$, 设 $s_{A_0} = \text{sign}(F_{A_0}y)$, 则 $\lambda_{1,(k_0+1)}$ 和 $\lambda_{1,(k_0+2)}$ 的值分别为

$$\lambda_{1,(k_0+1)} = \max_{j \notin A_0 \cup \{j_{k_0}\}, s \in \{-1, 1\}} \frac{X_j^T(I - R_{A_0})y}{s - X_j^T F_{A_0}^T s_{A_0}}, \quad (3.10)$$

$$\lambda_{1,(k_0+2)} = \lambda_{1,(k_0+1)}^{\text{join}}. \quad (3.11)$$

(3.10) 意味着第 k_0+1 个节点上为“加入节点”并省略了示性函数的约束, (3.11) 表示 k_0+2 同样是“加入节点”. 由此, 得到以下结果.

定理 3.3 假设 $X \in \mathbb{R}^{n \times p}$ 和 $y \in \mathbb{R}^n$ 由线性模型 (1.1) 给定. 如果满足条件 (a)–(c), 且 k_0 节点假设依概率趋近于 1 ($P(B_{k_0}^{\text{EN}}) \rightarrow 1$), 令 $k = k_0 + 1$, 则有

$$\lim_{p \rightarrow \infty} P(T_k^{\text{EN}} > t) \leq e^{-(1+\lambda_2)t}, \quad \forall t \geq 0.$$

在定理 3.3 中, (3.3) 和 (3.6) 可分别看作不可表示条件和弱不可表示性条件, 常见于研究弹性约束估计的模型选择理论性质的文献, 在该条件下, 选择合适的参数, 弹性约束估计将大概率选出正确模型. (3.5) 和 (3.8) 用于证明附录 A 的引理 A.3. k_0 节点假设 ($B_{k_0}^{\text{EN}}$) 用于保证该节点是“加入节点”.

3.3 正交预测矩阵 X

本节讨论特殊情形正交预测矩阵 X , 即 $X^T X = I$. 对比一般矩阵的检验统计量. 正交阵下的 Cov-EN 检验简单易得, 可以不加证明地给出以下结论: 对任意的正交阵 X , 通过定理 3.1, 弹性约束估计解路径上第 k 个节点的 Cov-EN 检验统计量为

$$T_k^{\text{EN}} = \frac{1}{1 + \lambda_2} \cdot \frac{\lambda_{1,k}(\lambda_{1,k} - \lambda_{1,(k+1)})}{\sigma^2}.$$

设 $U_j = X_j^T y$, 已知 $U_j \sim N(\beta_j, \sigma^2)$ 且互相独立, $j = 1, \dots, p$, 设 A^* 表示真实非零系数的下标集合, 假设真实系数满足

$$\min_{j \in A^*} |\beta_j^*| - \sigma \sqrt{2 \log p} \rightarrow \infty, \quad \text{当 } p \rightarrow \infty.$$

由罚项正则化方法的估计一致性理论研究^[17]和模型选择一致性理论研究^[16]可得, 在正交阵下, 选择合适的参数, 弹性约束估计将大概率选中正确模型. 由于预测矩阵正交, 选入有效集合的变量将不会被剔除, 假设真实系数有 k_0 个非零项, 该假设意味着弹性约束估计将大概率地在解路径第 k_0 个节点选出正确模型. 设定以下事件:

$$B = \left\{ \min_{j \in A^*} |U_j| > \max_{j \notin A^*} |U_j| \right\},$$

因此, 当 $p \rightarrow \infty$ 时, $P(B) \rightarrow 1$. 基于定理 3.3, Cov-EN 检验在第 $k_0 + 1$ 个节点的渐近性质可不加证明地由以下引理给出.

引理 3.2 假设 $X \in \mathbb{R}^{n \times p}$ 是正交的, 真实系数 β 有 k_0 个非零项, 设集合 B 依概率趋近于 1, 则对任意 $k_0 \geq 0$, 有

$$(1 + \lambda_2)T_{k_0+1}^{\text{EN}} \xrightarrow{d} \text{Exp}(1), \quad \text{当 } p \rightarrow \infty.$$

4 模拟算例

4.1 正交预测矩阵与一般预测矩阵

本节研究正交预测矩阵数据和一般预测矩阵数据, 分别给出零假设下弹性约束估计的 Cov-EN 检验效果. 假设预测矩阵来自多元正态分布 $N(0, \Sigma)$, Σ 有两种不同设定: (1) 独立同分布, $\Sigma = I$ (正交模型); (2) 相关系数为 $\rho = 0.5$ (一般模型). 对其他参数的设定如下: 样本数量为 $n = 100$, 预测变量个数为 $p = 10$, 真实系数为 $\beta = (3, 0, 0, 0, 0, 0, 0, 0, 0, 0)$. λ_2 设定 5 个不同的取值 $\{0, 0.1, 1, 9, 49\}$. 但当 $\lambda_2 = 0$ 时, 弹性约束估计退化为 Lasso.

首先计算正交预测矩阵下的 Cov-EN 检验统计量和 $\text{Exp}(1)$ 分布的分位数图, 图 1 给出第二个预

图 1 该图展示模型为 $n = 100$ 和 $p = 10$ 的正交预测矩阵, 真实参数仅有一个非零项, 重复模拟 400 次后的分位数图, 检验四个不同的 λ_2 值下的弹性约束估计的解路径的第二个节点的 Cov-EN 检验效果

测变量进入有效集合的检验效果, 该模拟重复了 400 次. 图 1(a) 表示 $\lambda = 0.1$ 时的 Cov-EN 检验效果; 图 1(b) 表示 $\lambda = 1$ 时的 Cov-EN 检验效果; 图 1(c) 表示 $\lambda = 9$ 时的 Cov-EN 检验效果; 图 1(d) 表示 $\lambda = 49$ 时的 Cov-EN 检验效果. 图 2 展示一般预测矩阵下的 Cov-EN 检验统计量的分布情形, 与图 1 类似, 给出不同 $\lambda_2 = \{0.1, 1, 9, 49\}$ 取值下, 第二个预测变量进入有效变量集合的 Cov-EN 检验效果.

4.2 弹性约束估计与 Lasso 的检验效果对比

本节通过数值模拟比较弹性约束估计与 Lasso 的检验效果差异, 后者采用文献 [12] 中的协方差检验方法. 图 3 和 4 分别展示在正交预测矩阵和一般矩阵下弹性约束估计与 Lasso 的检验结果, 设定与前两例类似, 假设 $n = 100$, $p = 10$, 真实系数仅有一个非零项, 即第一项等于 3, 余项为 0. 图 3(a) 表示正交阵下, $\lambda = 9$ 时弹性约束估计的检验效果; 图 3(b) 表示 $\lambda_2 = 0$ 即 Lasso 的检验结果.

图 2 该图展示模型为 $n = 100$ 和 $p = 10$ 的一般预测矩阵, 真实参数仅有一个非零项, 重复模拟 400 次后的分位数图, 检验四个不同的 λ_2 值下的弹性约束估计解路径上的第二个节点的 Cov-EN 检验效果

图 5 考虑如下模型, 在正交矩阵下, 设定 $n = 100, p = 400$, 真实系数 β^* 包含 3 个非零项, 前三项均等于 3, 其余为 0. 图 5 展示了第 5、6、7 个变量分别进入弹性约束估计解路径、Lasso 解路径时的检验效果. 该图通过分位数图展示 Cov-EN 检验和协方差检验在这三个节点时被选中变量的显著性, 其中,

图 3 该图展示模型为 $n = 100$ 和 $p = 10$ 的正交预测矩阵, 真实系数仅有一个非零项, 对比弹性约束估计与 Lasso 在相同设定下的检验效果差异. (a) 弹性约束估计的 Cov-EN 检验; (b) Lasso 的协方差检验

图 4 该图展示模型为 $n = 100$ 和 $p = 10$ 的一般预测矩阵, 真实系数仅有一个非零项, 对比弹性约束估计与 Lasso 在相同设定下的检验效果差异. (a) 弹性约束估计的 Cov-EN 检验; (b) Lasso 的协方差检验

图 5 模型为 $n = 100$ 和 $p = 10$ 的正交预测矩阵, 真实系数包含 3 个非零项, 三条虚线的斜率分别为 $1/2$ 、 $1/3$ 和 $1/4$, (a)–(c) Elastic, 展示弹性约束估计的 Cov-EN 检验; (d)–(f) Lasso, 展示 Lasso 的协方差检验

三条虚线的斜率分别为 $1/2$ 、 $1/3$ 和 $1/4$, 如图所示, 基于弹性约束估计解路径的 Cov-EN 检验比基于 Lasso 的协方差检验更贴合虚线.

5 实际应用

5.1 红酒数据分析

本节采用文献 [12] 使用的红酒评分案例, 该数据来自 UCI (University of California Irvine) 数据库, 总计 1,599 个样本, 11 个预测变量, 分别是酒精、挥发性酸、硫酸盐、氯化物、pH 值、总二氧化硫、糖度、酸度、密度、酸度和游离二氧化硫, 响应变量是红酒质量评价, 打分在 0 到 10 之间. 将所有样本随机地等分为测试集和训练集, 并将这种等分重复 500 次, 通过交叉验证从 λ_2 的 5 个取值 $\{0.1, 0.5, 1, 5, 10, 100\}$ 中选取最优 λ_2 .

图 6 展示红酒数据的测试结果. 图 6(a) 展示通过训练集计算得到的 p 值, 实线表示弹性约束估计的 p 值, 前三项极小, 第四项的 p 值为 0.03, 第五项的 p 值接近 0.1. 与之相比, Lasso 的前三项的 p 值同样较小, 第四项的 p 值增大到 0.2. 图 6(b) 展示测试集的预测效果, 弹性约束估计与 Lasso 具有相似的预测误差, 均在第二和三个变量被选中时显著降低, 当第四个变量被选中时, 误差降低到最低. 从预测误差的表现上来看, Cov-EN 检验的 p 值更符合估计效果, 即在选中第四个变量后停止.

5.2 股票数据分析

本节研究金融问题的应用, 采用道琼斯指数的数据, 数据来自万德金融终端, 包含从 2016 年 1 月到 2017 年 12 月的指数日收盘价和对应 30 支成份股日收盘价. 为了测试 $p > n$ 的高维数据检验结果,

我们将数据集分为 20 个样本用于建模, 20 个样本用于预测的滚动时间窗, 其中 σ^2 的估计和 λ_2 均通过交叉验证获得, λ_2 的可选集为 $\{1, 5, 10, 100, 200\}$.

图 7 展示股票数据分析的结果. 图 7(a) 显示通过训练集计算获得的 p 值; 图 7(b) 显示由测试集计算获得的 p 值. 由图 7(a) 所示, 弹性约束估计通过 Cov-EN 检验获取的前两个变量选入时的 p 值极小, 第三个 p 值接近 0.1, 作为对比, Lasso 的协方差检验在第二个变量选中时 p 值便接近 0.15, 这意味着 Lasso 的协方差检验仅选中了一个变量, 而弹性约束估计的 Cov-EN 检验选中了两个变量, 其中第一个

图 6 红酒数据分析. (a) 展示两个方法各自在训练集下的检验 p 值; (b) 展示对应模型的预测误差

图 7 股票数据分析: 数据被划分为 49 份, 训练集和测试集均包含 20 个样本. (a) 显示两个方法各自在训练集下的检验 p 值; (b) 显示其对应测试集的 p 值

变量是美孚公司的股票, 第二个变量是 IBM 公司的股票. 根据图 7(b) 所示, 两个检验均认为美孚公司对道琼斯指数来说是最重要的. 对应实际情形分析, IBM 公司对于道琼斯指数来说是不可或缺的组成成分, 因此, Cov-EN 检验结果更合理.

6 讨论

本文基于高维线性模型, 提出了 Cov-EN 检验统计量, 用于检验弹性约束估计选中变量时的显著性. 我们研究了该统计量的渐近性质, 分别在正交预测矩阵和一般预测矩阵下, 论证该 Cov-EN 检验渐近收敛于参数为 1 的指数分布. 为了举例说明该结论, 本文进行了相关的数值模拟分析和实际数据应用分析, 并将该检验与 Lasso 的协方差检验进行比较.

致谢 感谢审稿人的仔细阅读和中肯意见.

参考文献

- 1 Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B Stat Methodol*, 1996, 58: 267–288
- 2 Wu L, Yang Y H. Nonnegative elastic net and application in index tracking. *Appl Math Comput*, 2014, 227: 541–552
- 3 Wu L, Yang Y H, Liu H. Nonnegative-Lasso and application in index tracking. *Comput Statist Data Anal*, 2014, 70: 116–126
- 4 Yang Y H, Wu L. Nonnegative adaptive Lasso for ultra-high dimensional regression models and a two-stage method applied in financial modeling. *J Statist Plann Inference*, 2016, 174: 52–67
- 5 Zou H. The adaptive Lasso and its oracle properties. *J Amer Statist Assoc*, 2006, 101: 1418–1429
- 6 Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol*, 2005, 67: 301–320
- 7 Zou H, Zhang H H. On the adaptive Elastic-Net with a diverging number of parameters. *Ann Statist*, 2009, 37: 1733–1751
- 8 Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Ser B Stat Methodol*, 2010, 72: 417–473
- 9 Meinshausen N, Meier L, Bühlmann P. p -values for high-dimensional regression. *J Amer Statist Assoc*, 2009, 104: 1671–1681
- 10 Minnier J, Tian L, Cai T. A perturbation method for inference on regularized regression estimates. *J Amer Statist Assoc*, 2011, 106: 1371–1382
- 11 Wasserman L, Roeder K. High-dimensional variable selection. *Ann Statist*, 2009, 37: 2178–2201
- 12 Lockhart R, Taylor J, Tibshirani R J, et al. A significance test for the Lasso. *Ann Statist*, 2014, 42: 413–468
- 13 Efron B, Hastie T, Johnstone L, et al. Least angle regression. *Ann Statist*, 2004, 32: 407–451
- 14 Tibshirani R J. The Lasso problem and uniqueness. *Electron J Stat*, 2013, 7: 1456–1490
- 15 Tibshirani R J, Taylor J. Degrees of freedom in Lasso problems. *Ann Statist*, 2012, 40: 1198–1232
- 16 Jia J, Yu B. On model selection consistency of elastic net when $p \gg n$. *Statist Sinica*, 2010, 20: 595–611
- 17 Negahban S, Ravikuma P, Wainwright M, et al. A unified framework for high-dimensional analysis of M -estimators with decomposable regularizers. *Statist Sci*, 2012, 27: 1348–1356

附录 A 定理的证明

定理 3.1 的证明 基于第 2 节的讨论, 给定正则参数 λ_1 和 λ_2 , $\hat{\beta}$ 是弹性约束估计 (1.2) 的最优解当且仅当下式成立:

$$\hat{\beta}_A(\lambda_1, \lambda_2) = (X_A^T X_A + \lambda_2 I)^{-1} (X_A^T y - \lambda_1 s_A).$$

因此, 令 $R_A = X_A(X_A^T X_A + \lambda_2 I)^{-1} X_A^T$, $F_A = (X_A^T X_A + \lambda_2 I)^{-1} X_A^T$, 得到

$$X_A \hat{\beta}_A(\lambda_1, \lambda_2) = R_A y - \lambda_1 F_A^T s_A.$$

设 $\lambda_{1,(k+1)}$ 为 λ_1 在第 $k+1$ 个节点上的正则参数取值, 则在该节点有如下等式:

$$X_{A \cup \{j\}} \hat{\beta}_{A \cup \{j\}}(\lambda_{1,(k+1)}, \lambda_2) = R_{A \cup \{j\}} y - \lambda_{1,(k+1)} F_{A \cup \{j\}}^T s_{A \cup \{j\}},$$

其中 $s_{A \cup \{j\}} = \text{sign}(\hat{\beta}_{A \cup \{j\}}(\lambda_{1,(k+1)}))$. 采用正文的设定, $\tilde{\beta}_A(\lambda_{1,(k+1)}, \lambda_2)$ 是在该节点上仅使用集合 A 作为有效集合, 但使用正则参数 $\lambda_{1,(k+1)}$ 和 λ_2 计算得到的弹性约束估计的解, 即

$$\begin{aligned} X_A \tilde{\beta}_A(\lambda_{1,(k+1)}, \lambda_2) &= X_A (X_A^T X_A + \lambda_2 I)^{-1} (X_A^T y - \lambda_{1,(k+1)} s_A) \\ &= R_A y - \lambda_{1,(k+1)} F_A^T s_A. \end{aligned}$$

因此, Cov-EN 检验可以表示为

$$T_k^{\text{EN}} = \frac{y^T (R_{A \cup \{j\}} - R_A) y}{\sigma^2} - \frac{\lambda_{1,(k+1)} y^T (F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A)}{\sigma^2}. \quad (\text{A.1})$$

由弹性约束估计的解路径在各节点上的连续性可得

$$R_A y - \lambda_{1,k} F_A^T s_A = R_{A \cup \{j\}} y - \lambda_{1,k} F_{A \cup \{j\}}^T s_{A \cup \{j\}}, \quad (\text{A.2})$$

上式的等号左边是第 k 个节点的解, 右边是第 $k+1$ 个节点上的解. 由上式进一步计算得到

$$(R_{A \cup \{j\}} - R_A) y = \lambda_{1,k} (F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A).$$

已知

$$R_{A \cup \{j\}}^2 = R_{A \cup \{j\}} - \lambda_2 F_{A \cup \{j\}}^T F_{A \cup \{j\}},$$

$$R_A^2 = R_A - \lambda_2 F_A^T F_A,$$

$$R_{A \cup \{j\}} R_A = R_A + \lambda_2 F_{A \cup \{j\}}^T F_A,$$

因此有

$$(R_{A \cup \{j\}} - R_A)^2 = (R_{A \cup \{j\}} - R_A) - \lambda_2 (F_{A \cup \{j\}} - F_A)^T (F_{A \cup \{j\}} - F_A).$$

计算 (A.1) 等号右边的第一项得到

$$\begin{aligned} y^T (R_{A \cup \{j\}} - R_A) y &= y^T [(R_{A \cup \{j\}} - R_A)^2 + \lambda_2 \|F_{A \cup \{j\}} - F_A\|_2^2] y \\ &= \lambda_{1,k}^2 \|F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A\|_2^2 + \lambda_2 \|F_{A \cup \{j\}} y - F_A y\|_2^2. \end{aligned}$$

计算 (A.1) 等号右边的第二项得到

$$\begin{aligned} y^T (F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A) &= \frac{1}{\lambda_{1,k}} y^T (R_{A \cup \{j\}} - R_A) y \\ &= \lambda_{1,k} \|F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A\|_2^2 + \frac{\lambda_2}{\lambda_{1,k}} \|F_{A \cup \{j\}} y - F_A y\|_2^2. \end{aligned}$$

因此, Cov-EN 检验统计量计算如下:

$$T_k^{\text{EN}} = \frac{y^T (R_{A \cup \{j\}} - R_A) y}{\sigma^2} - \frac{\lambda_{1,(k+1)} y^T (F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A)}{\sigma^2}$$

$$\begin{aligned}
 &= \frac{\lambda_{1,k}^2 \|F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A\|_2^2 + \lambda_2 \|F_{A \cup \{j\}} y - F_A y\|_2^2}{\sigma^2} \\
 &\quad - \lambda_{1,(k+1)} \frac{[\lambda_{1,k} \|F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A\|_2^2 + \frac{\lambda_2}{\lambda_{1,k}} \|F_{A \cup \{j\}} y - F_A y\|_2^2]}{\sigma^2} \\
 &= \lambda_{1,k} (\lambda_{1,k} - \lambda_{1,(k+1)}) \frac{\|F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A\|_2^2}{\sigma^2} + \lambda_2 \left(1 - \frac{\lambda_{1,(k+1)}}{\lambda_{1,k}}\right) \frac{\|F_{A \cup \{j\}} y - F_A y\|_2^2}{\sigma^2}.
 \end{aligned}$$

证毕. $\square$

定义 A.1 在零假设下, 定义期望为 0 的 Gauss 过程 $g(j, s) = sU_j$, $j = 1, \dots, p$, $s \in \{-1, 1\}$, 其中 $U_j = X_j^T y$, $j = 1, \dots, p$, 其协方差函数为 $E[g(j, s)g(j_1, s_1)] = ss_1 C_{j,j_1} \sigma^2$, 则有 $\lambda_{1,1} = \max_{j,s} g(j, s)$, 该等式由引理 3.1 给出, 考虑如下等式:

$$\begin{aligned}
 h_{j_1, s_1}^{\text{EN}}(j, s) &= \frac{g(j, s) - \frac{1}{1+\lambda_2} s C_{j,j_1} g(j_1, s_1)}{1 - \frac{1}{1+\lambda_2} ss_1 C_{j,j_1}}, \quad j \neq j_1, \quad s \in \{-1, 1\}, \\
 M^{\text{EN}}(j_1, s_1) &= \max_{j \neq j_1, s \in \{-1, 1\}} h^{j_1, s_1}(j, s).
 \end{aligned}$$

若将 j_1 和 s_1 看作最大化 $g(j, s)$ 的取值, 则有 $\lambda_{1,2} = M^{\text{EN}}(j_1, s_1)$, 继而得到第一个节点的 Cov-EN 统计量表达如下:

$$T_1^{\text{EN}} = \frac{\lambda_{1,1}(\lambda_{1,1} - \lambda_{1,2})}{(1 + \lambda_2)\sigma^2} = \frac{g(j_1, s_1)(g(j_1, s_1) - M^{\text{EN}}(j_1, s_1))}{(1 + \lambda_2)\sigma^2},$$

定理 3.2 的证明 引用定义 A.1 的设定, 弹性约束估计解路径上第一个节点的 Cov-EN 检验的概率分布为

$$P(T_1^{\text{EN}} > t) = \sum_{j_1, s_1} P\left(\frac{1}{1 + \lambda_2} \cdot \frac{g(j_1, s_1)(g(j_1, s_1) - M^{\text{EN}}(j_1, s_1))}{\sigma^2} > t, g(j_1, s_1) > g(j, s), \forall j, s\right). \quad (\text{A.3})$$

将该概率分布的前半部分看作关于对象 g 的二次方程, 简化得到

$$\frac{1}{1 + \lambda_2} \cdot \frac{g(g - M^{\text{EN}})}{\sigma^2} = t.$$

通过直接计算, 下式可用于帮助求解该二次方程:

$$\{g(j_1, s_1) > g(j, s), \forall j, s\} = \{g(j_1, s_1) > M^{\text{EN}}(j_1, s_1)\}. \quad (\text{A.4})$$

将 (A.4) 代入 (A.3) 求解二次方程并设 u 为较大值根, 则有

$$u\left((1 + \lambda_2)t, \frac{M^{\text{EN}}}{\sigma}\right) = \frac{1}{2} \left(\frac{M^{\text{EN}}}{\sigma} + \sqrt{\frac{(M^{\text{EN}})^2}{\sigma^2} + \frac{4M^{\text{EN}}}{\sigma}} \right).$$

因此得到

$$P(T_1^{\text{EN}} > t) = \sum_{j_1, s_1} P\left(\frac{g(j_1, s_1)}{\sigma^2} > u\left((1 + \lambda_2)t, \frac{M^{\text{EN}}}{\sigma}\right)\right).$$

令 $\bar{\Phi}$ 表示标准正态剩余函数, $F_M^{\text{EN}}(j_1, s_1)$ 表示 $M^{\text{EN}}(j_1, s_1)$ 的分布, (A.3) 可以进一步地表示为

$$P(T_1^{\text{EN}} > t) = \sum_{j_1, s_1} \int_0^\infty \bar{\Phi}\left(u\left(t(1 + \lambda_2), \frac{m}{\sigma}\right)\right) F_{M(j_1, s_1)}^{\text{EN}}(dm). \quad (\text{A.5})$$

接下来考虑 (A.5) 与指数分布的差的渐近分布, 即

$$|P(T_1^{\text{EN}} > t) - e^{-(1+\lambda_2)t}| = \left| \sum_{j_1, s_1} \int_0^\infty \left(\frac{\bar{\Phi}((1+\lambda_2)t, m/\sigma)}{\bar{\Phi}(m/\sigma)} - e^{(1+\lambda_2)t} \right) \bar{\Phi}\left(\frac{m}{\sigma}\right) F_M(dm) \right|. \quad (\text{A.6})$$

(A.6) 可以被分割为以下两个部分:

$$\begin{aligned} & \left| \sum_{j_1, s_1} \int_0^\infty \left(\frac{\bar{\Phi}((1+\lambda_2)t, m/\sigma)}{\bar{\Phi}(m/\sigma)} - e^{(1+\lambda_2)t} \right) \bar{\Phi}\left(\frac{m}{\sigma}\right) F_M(dm) \right| \\ & \leq \sum_{j_1, s_1} \int_0^{m_0} \left| \frac{\bar{\Phi}((1+\lambda_2)t, m/\sigma)}{\bar{\Phi}(m/\sigma)} - e^{(1+\lambda_2)t} \right| \bar{\Phi}\left(\frac{m}{\sigma}\right) F_M(dm) \\ & \quad + \sum_{j_1, s_1} \int_{m_0}^\infty \left| \frac{\bar{\Phi}((1+\lambda_2)t, m/\sigma)}{\bar{\Phi}(m/\sigma)} - e^{(1+\lambda_2)t} \right| \bar{\Phi}\left(\frac{m}{\sigma}\right) F_M(dm). \end{aligned} \quad (\text{A.7})$$

假设满足条件 (3.7) 和 (3.4), 对任意常数 m_0 , 有

$$\sum_{j_1, s_1} P(M^{\text{EN}}(j_1, s_1) \leq m_0) \rightarrow 0, \quad \text{当 } p \rightarrow \infty. \quad (\text{A.8})$$

以上结论可用于论证 (A.7) 不等式右边的第一项. 该结果的证明细节与后文帮助证明定理 3.1 的引理 A.3 类似, 所以这里省略证明过程. 计算 (A.7) 不等式右边第二项, 由 L'Hospital 法则可得

$$\lim_{m \rightarrow \infty} \frac{\bar{\Phi}((1+\lambda_2)t, m/\sigma)}{\bar{\Phi}(m/\sigma)} = e^{-(1+\lambda_2)t}.$$

因此, 对任意给定的 $m > m_0$ 和 ϵ , 下式成立:

$$\left| \frac{\bar{\Phi}((1+\lambda_2)t, m/\sigma)}{\bar{\Phi}(m/\sigma)} - e^{-(1+\lambda_2)t} \right| \leq \epsilon. \quad (\text{A.9})$$

结合 (A.8) 和 (A.9) 得到

$$\begin{aligned} |P(T_1^{\text{EN}} > t) - e^{-(1+\lambda_2)t}| & \leq \sum_{j_1, s_1} \int_0^{m_0} \left| \frac{\bar{\Phi}((1+\lambda_2)t, m/\sigma)}{\bar{\Phi}(m/\sigma)} - e^{(1+\lambda_2)t} \right| \bar{\Phi}\left(\frac{m}{\sigma}\right) F_M(dm) \\ & \quad + \sum_{j_1, s_1} \int_{m_0}^\infty \left| \frac{\bar{\Phi}((1+\lambda_2)t, m/\sigma)}{\bar{\Phi}(m/\sigma)} - e^{(1+\lambda_2)t} \right| \bar{\Phi}\left(\frac{m}{\sigma}\right) F_M(dm) \\ & \leq \sum_{j_1, s_1} \int_{m_0}^\infty F_M(dm) + \epsilon \sum_{j_1, s_1} \int_{m_0}^\infty \bar{\Phi}\left(\frac{m}{\sigma}\right) F_M(dm) \\ & \leq \sum_{j_1, s_1} P(M(j_1, s_1) \leq m_0) + \epsilon \sum_{j_1, s_1} P(g(j_1, s_1) > M(j_1, s_1)) \\ & \rightarrow 0, \quad \text{当 } p \rightarrow \infty. \end{aligned}$$

证毕. □

定义 A.2 在零假设下, 定义离散时间的 Gauss 过程

$$g_{A, s_A}^{\text{EN}}(j, s) = \frac{X_j^T(I - R_A)y}{s - X_j^T F_A^T s_A}, \quad j \notin A, \quad s \in \{-1, 1\},$$

其中 A 和 s_A 分别是弹性约束估计解路径刚到 k 节点的有效变量集合及其对应系数估计值的符号集合. 与定义 A.1 类似, 可计算得到 $\lambda_{1,k} = \max_{j \notin A, s \in \{-1, 1\}} g_{A, s_A}^{\text{EN}}(j, s)$. 设 j_k 是第 k 个节点加入集合 A 的变量, 它的系数符号是 s_k , 则有

$$g_{A \cup \{j_k\}, s_{A \cup \{j_k\}}}^{\text{EN}}(j, s) = \frac{X_j^T(I - R_{A \cup \{j_k\}})y}{s - X_j^T F_{A \cup \{j_k\}}^T s_{A \cup \{j_k\}}}, \quad j \notin A \cup \{j_k\}, \quad s \in \{-1, 1\},$$

$$M_{A, s_A}^{\text{EN}}(j_k, s_k) = \max_{j_k \notin A \cup \{j_k\}, s \in \{-1, 1\}} g_{A \cup \{j_k\}, s_{A \cup \{j_k\}}}^{\text{EN}}(j, s) \cdot 1 \left\{ g_{A \cup \{j_k\}, s_{A \cup \{j_k\}}}^{\text{EN}}(j, s) < \max_{j \notin A, s \in \{-1, 1\}} g_{A, s_A}^{\text{EN}}(j, s) \right\}.$$

假设第 k 个节点满足 k_0 节点假设 ($B_{k_0}^{\text{EN}}$), 则有 $\lambda_{1, (k+1)} = M_{A, s_A}^{\text{EN}}(j_k, s_k)$, 因此, 该节点的 Cov-EN 统计量有以下表达式:

$$T_k^{\text{EN}} = \frac{\lambda_{1,k}(\lambda_{1,k} - \lambda_{1, (k+1)})}{\sigma^2} C(A, s_A, j, s) + \lambda_2 \frac{(\lambda_{1,k} - \lambda_{1, (k+1)})}{\sigma^2 \lambda_{1,k}} D(A, j, s, y)$$

$$= C(A, s_A, j, s) \frac{g_{A, s_A}^{\text{EN}}(j_k, s_k)(g_{A, s_A}^{\text{EN}}(j_k, s_k) - M_{A, s_A}^{\text{EN}}(j_k, s_k))}{\sigma^2}$$

$$+ \lambda_2 D(A, j, s, y) \frac{g_{A, s_A}^{\text{EN}}(j_k, s_k) - M_{A, s_A}^{\text{EN}}(j_k, s_k)}{\sigma^2 g_{A, s_A}^{\text{EN}}(j_k, s_k)},$$

上式在事件 $\{g_{A, s_A}^{\text{EN}}(j_k, s_k) > g_{A, s_A}^{\text{EN}}(j, s), \forall j, s\}$ 中成立.

由于事件 $\{g_{A, s_A}^{\text{EN}}(j_k, s_k) > g_{A, s_A}^{\text{EN}}(j, s), \forall j, s\}$ 包含于 k_0 节点假设中, 因此, 定义 A.2 的结论是直观的. 接下来需要考虑的问题是当 $k > 1$, $g_{A, s_A}^{\text{EN}}(j_k, s_k)$ 和 $M_{A, s_A}^{\text{EN}}(j_k, s_k)$ 在固定 j_k 和 s_k 时不再独立的处理方法. 为此, 在接下来的引理 A.1 中引入三组变量 M^+ 、 M^- 和 M^0 , 使其在固定 j_k 和 s_k 时与 g^{EN} 独立, 该结论无需证明.

引理 A.1 引用定义 A.2 的设定 $g^{\text{EN}}(j, s)$, 令 $\Sigma_{j, j'}$ 表示 $g^{\text{EN}}(j, s)$ 的协方差, 则有

$$\Sigma_{j, j'} = E[g_{A, s_A}^{\text{EN}}(j, s)g_{A, s_A}^{\text{EN}}(j', s')] = \frac{X_j^T(1 - R_A + \lambda_2 F_A^T F_A)X_{j'}\sigma^2}{[s - X_j^T F_A^T s_A][s' - X_{j'}^T F_A^T s_A]},$$

其中 $j \notin A, s \in \{-1, 1\}$. 定义以下三个变量:

$$M^+(j, s) = \max_{(j', s') \in S^+(j, s)} \frac{g(j', s') - (\Sigma_{j, j'}/\Sigma_{jj})g(j, s)}{1 - \Sigma_{j, j'}/\Sigma_{jj}},$$

$$M^-(j, s) = \min_{(j', s') \in S^-(j, s)} \frac{g(j', s') - (\Sigma_{j, j'}/\Sigma_{jj})g(j, s)}{1 - \Sigma_{j, j'}/\Sigma_{jj}},$$

$$M^0(j, s) = \max_{(j', s') \in S^0(j, s)} g(j', s') - \left(\frac{\Sigma_{j, j'}}{\Sigma_{jj}} \right) g(j, s),$$

其中 $S^+(j, s) = \{(j', s') : \frac{\Sigma_{j, j'}}{\Sigma_{jj}} < 1\}$, $S^-(j, s) = \{(j', s') : \frac{\Sigma_{j, j'}}{\Sigma_{jj}} > 1\}$, $S^0(j, s) = \{(j', s') : \frac{\Sigma_{j, j'}}{\Sigma_{jj}} = 1\}$. 由于

$$E \left[g^{\text{EN}}(j_k, s_k) \left(g^{\text{EN}}(j, s) - \left(\frac{\Sigma_{j_k, j}}{\Sigma_{j_k, j_k}} \right) g^{\text{EN}}(j_k, s_k) \right) \right] = 0,$$

因此, $M^+(j_k, s_k)$ 、 $M^-(j_k, s_k)$ 和 $M^0(j_k, s_k)$ 关于 $g^{\text{EN}}(j_k, s_k)$ 独立, 且事件 $\{g_{A, s_A}^{\text{EN}}(j_k, s_k) > g_{A, s_A}^{\text{EN}}(j, s), \forall j, s\}$ 可以被表示为

$$\{g_{A, s_A}^{\text{EN}}(j_k, s_k) > g_{A, s_A}^{\text{EN}}(j, s), \forall j, s\}$$

$$= \{g_{A, s_A}^{\text{EN}}(j_k, s_k) > M^+(j_k, s_k)\} \cap \{g_{A, s_A}^{\text{EN}}(j_k, s_k) < M^-(j_k, s_k)\} \cap \{0 > M^0(j_k, s_k)\},$$

其中 j_k 和 s_k 通过最大化 $g^{\text{EN}}(j, s)$ 获得.

引理 A.2 任意给定 A 和 s_A , 对所有的 $j \notin A$, $s \in \{-1, 1\}$, 令 $s_{A \cup \{j\}}$ 表示 s_A 与 s 串联, 则有

$$\text{Var}(g^{\text{EN}}(j, s)) = \frac{X_j^T(1 - R_A - \lambda_2 F_A^T F_A)X_j \sigma^2}{[s - X_j^T F_A^T s_A]^2} \leq \frac{\sigma^2}{C(A, s_A, j, s) + \frac{\lambda_2}{\lambda_{1,k}^2} \cdot D(A, j, s, y)}. \quad (\text{A.10})$$

证明 首先,

$$\begin{aligned} C(A, s_A, j, s) &= \|F_{A \cup \{j\}}^T s_{A \cup \{j\}} - F_A^T s_A\|_2^2 \\ &= s_{A \cup \{j\}}^T (X_{A \cup \{j\}}^T X_{A \cup \{j\}} + \lambda_2 I)^{-1} s_{A \cup \{j\}} - s_A^T (X_A^T X_A + \lambda_2 I)^{-1} s_A \\ &\quad - \lambda_2 [s_{A \cup \{j\}}^T (X_{A \cup \{j\}}^T X_{A \cup \{j\}} + \lambda_2 I)^{-2} s_{A \cup \{j\}} - s_A^T (X_A^T X_A + \lambda_2 I)^{-2} s_A]. \end{aligned} \quad (\text{A.11})$$

考虑 (A.11) 等式右边的前两项, 需要证明如下等式:

$$\frac{[s - X_j^T F_A^T s_A]^2}{X_j^T(1 - R_A)X_j} = s_{A \cup \{j\}}^T (X_{A \cup \{j\}}^T X_{A \cup \{j\}} + \lambda_2 I)^{-1} s_{A \cup \{j\}} - s_A^T (X_A^T X_A + \lambda_2 I)^{-1} s_A.$$

令 $z^{\text{EN}} = (X_{A \cup \{j\}}^T X_{A \cup \{j\}} + \lambda_2 I)^{-1} s_{A \cup \{j\}}$, 则有

$$\begin{pmatrix} (X_A^T X_A + \lambda_2 I) & X_A^T X_j \\ X_j^T X_A & X_j^T X_j + \lambda_2 \end{pmatrix} \begin{pmatrix} z_1^{\text{EN}} \\ z_2^{\text{EN}} \end{pmatrix} = \begin{pmatrix} s_A \\ s \end{pmatrix}.$$

求解 z_1^{EN} 和 z_2^{EN} 得

$$\begin{aligned} z_1^{\text{EN}} &= (X_A^T X_A + \lambda_2 I)^{-1} s_A - F_A X_j z_2, \\ z_2^{\text{EN}} &= \frac{[s - s_A^T F_A X_j]}{X_j^T(1 - R_A)X_j}. \end{aligned}$$

将 z_1^{EN} 和 z_2^{EN} 的解代入 (A.10) 不等式右边, 得到

$$\begin{aligned} &s_{A \cup \{j\}}^T (X_{A \cup \{j\}}^T X_{A \cup \{j\}} + \lambda_2 I)^{-1} s_{A \cup \{j\}} - s_A^T (X_A^T X_A + \lambda_2 I)^{-1} s_A \\ &= s_A^T z_1 + s z_2 - s_A^T (X_A^T X_A + \lambda_2 I)^{-1} s_A \\ &= [s - s_A^T (X_A^T X_A + \lambda_2 I)^{-1} X_A^T X_j] z_2 \\ &= \frac{[s - s_A^T F_A X_j]^2}{X_j^T(1 - R_A)X_j}. \end{aligned}$$

结合条件 (c) 的 (3.8) 得到以下结果:

$$\begin{aligned} \text{Var}(g^{\text{EN}}(j, s)) &= \frac{X_j^T(1 - R_A - \lambda_2 F_A^T F_A)X_j \sigma^2}{[s - X_j^T F_A^T s_A]^2} \leq \frac{X_j^T(1 - R_A)X_j \sigma^2}{[s - X_j^T F_A^T s_A]^2} \\ &= [s_{A \cup \{j\}}^T (X_{A \cup \{j\}}^T X_{A \cup \{j\}} + \lambda_2 I)^{-1} s_{A \cup \{j\}} - s_A^T (X_A^T X_A + \lambda_2 I)^{-1} s_A]^{-1} \sigma^2 \\ &\leq \left[s_{A \cup \{j\}}^T (X_{A \cup \{j\}}^T X_{A \cup \{j\}} + \lambda_2 I)^{-1} s_{A \cup \{j\}} - s_A^T (X_A^T X_A + \lambda_2 I)^{-1} s_A \right. \\ &\quad \left. - \left(\lambda_2 \|C_{A \cup \{j\}}^{-1} s_{A \cup \{j\}} - C_A^{-1} s_A\|_2^2 - \frac{\lambda_2}{\lambda_{1,k}} \times D \right) \right]^{-1} \sigma^2 \\ &= \frac{\sigma^2}{C(A, s_A, j, s) + \frac{\lambda_2}{\lambda_{1,k}^2} \times D(A, j, s, y)}. \end{aligned}$$

证毕. □

引理 A.3 假设满足条件 (a)–(c), 则有

$$\sum_{j_k, s_k} P\left(M_{A, s_A}^{\text{EN}}(j_k, s_k) \leq \frac{m_0}{\sqrt{C(A, s_A, j, s)}}\right) \rightarrow 0, \quad \text{当 } p \rightarrow \infty.$$

证明 通过计算可知,

$$\frac{g_{A, s_A}^{\text{EN}}(j, s) - (\Sigma_{j_k, j} / \Sigma_{j_k, j_k}) g_{A, s_A}^{\text{EN}}(j_k, s_k)}{1 - \Sigma_{j_k, j} / \Sigma_{j_k, j_k}} \leq g_{A \cup j_k, s_A \cup j_k}^{\text{EN}}(j, s).$$

在事件 $\{g_{A, s_A}^{\text{EN}}(j_k, s_k) > g_{A, s_A}^{\text{EN}}(j, s), \forall j, s\}$ 成立时, 有

$$\begin{aligned} M^+(j_k, s_k) &= \max_{(j, s) \in S^+(j_k, s_k)} \frac{g_{A, s_A}^{\text{EN}}(j, s) - (\Sigma_{j_k, j} / \Sigma_{j_k, j_k}) g_{A, s_A}^{\text{EN}}(j_k, s_k)}{1 - \Sigma_{j_k, j} / \Sigma_{j_k, j_k}} \\ &\leq \max_{(j, s) \in S^+(j_k, s_k)} g_{A \cup \{j_k\}, s_A \cup \{j_k\}}^{\text{EN}}(j, s) \cdot 1\{g_{A \cup \{j_k\}, s_A \cup \{j_k\}}^{\text{EN}}(j, s) < g_{A, s_A}^{\text{EN}}(j_k, s_k)\} \\ &\leq \max_{j_k \notin A \cup \{j_k\}, s \in \{-1, 1\}} g_{A \cup \{j_k\}, s_A \cup \{j_k\}}^{\text{EN}}(j, s) \cdot 1\{g_{A \cup \{j_k\}, s_A \cup \{j_k\}}^{\text{EN}}(j, s) < g_{A, s_A}^{\text{EN}}(j_k, s_k)\} \\ &= M_{A, s_A}^{\text{EN}}(j_k, s_k). \end{aligned}$$

因此,

$$P\left(M_{A, s_A}^{\text{EN}}(j_k, s_k) \leq \frac{m_0}{\sqrt{C(A, s_A, j, s)}}\right) \leq P\left(M^+(j_k, s_k) \leq \frac{m_0}{\sqrt{C(A, s_A, j, s)}}\right).$$

令 $A = A_0, s_A = s_{A_0}$, 对任意固定的 m_0, j_k 和 s_k , 下式成立:

$$\begin{aligned} M^+(j_k, s_k) &\geq \max_{j \in S, s \in \{-1, 1\}} \frac{g(j, s) - (\Sigma_{j_k, j} / \Sigma_{j_k, j_k}) g(j_k, s_k)}{1 - \Sigma_{j_k, j} / \Sigma_{j_k, j_k}} \\ &= \max_{j \in S, s \in \{-1, 1\}} \frac{U_j^* - \theta_{j_k, j} U_{j_k}^*}{w_j s - X_j^T (X_{A \cup \{j_k\}}^T)^+ s_{A \cup \{j_k\}}} \\ &\geq \max_{j \in S} \frac{|U_j^* - \theta_{j_k, j} U_{j_k}^*|}{2}, \end{aligned} \tag{A.12}$$

其中 $U_j^* = X_j^T (I - P_A) y$, $\theta_{j_k, j} = X_{j_k}^T (I - P_A) X_j / (X_{j_k}^T (I - P_A) X_{j_k})$, $j \in S$. (A.12) 的第一个不等式由条件 (c) 的 (3.5) 达成. 条件 (a) 使得 $S^+(j_k, s_k) = \{(j, s) : \frac{\Sigma_{j_k, j}}{\Sigma_{j_k, j_k}} < 1\}$. (A.12) 的最后一个不等式通过条件 (c) 的 (3.6) 成立.

令 $V_j = \frac{(U_j^* - \theta_{j_k, j} U_{j_k}^*)}{2}$, 由引理 A.2 的结论, 设定 $\sigma_k = \text{Var}(g^{\text{EN}}(j_k, s_k))$, 得到

$$\begin{aligned} P\left(M^+(j_k, s_k) \leq \frac{m_0}{\sqrt{C(A, s_A, j, s)}}\right) &= P(M^+(j_k, s_k) \leq m_0 \sigma_k) \\ &\leq P(|V_1| \leq m_0 \sigma_k, \dots, |V_q| \leq m_0 \sigma_k), \end{aligned} \tag{A.13}$$

其中 $q = |S|$ ($q \ll p$), 第二个不等式由 (A.12) 达成. 接下来将归纳法应用到 (A.13) 中, 已知条件 (c) 的 (3.7) 满足, 则有

$$\frac{\text{Var}(U_i^* | U_l^*, l \in S \setminus \{i\})}{\text{Var}(U_i^*)} = \frac{G_{ii} - G_{i, S \setminus \{i\}} (R_{S \setminus \{i\}, S \setminus \{i\}})^{-1} G_{S \setminus \{i\}, i}}{G_{ii}}.$$

设 S' 是满足 $S' \in S \setminus \{i\}$ ($i \in S$) 的任意集合, 则有

$$\text{Var}(U_i^* | U_l^*, l \in S') \geq \text{Var}(U_i^* | U_l^*, l \in S \setminus \{i\}) \geq \delta^2 \sigma^2 G_{ii}.$$

由于 $U_1^*, \dots$ 是联合正态的, 因此, 上式第一个不等式成立. 又因 $V_1, \dots$ 关于 $U_{j_k}^*$ 独立, 所以, $E[V_j U_{j_k}^*] = 0$, $j = 1, \dots$, $\text{Var}(V_1) = \text{Var}(V_1 | U_{j_k}^*) = \text{Var}(U_1^*)/4 \geq \delta^2 \sigma^2 R_{11}/4$, 即意味着 $V_1 \sim N(0, \text{Var}(V_1))$,

$$P(|V_1| \leq m_0 \sigma_k) \leq P\left(|\Phi| \leq \frac{2m_0 \sigma_k}{\delta \sigma \sqrt{R_{11}}}\right) \leq P\left(|\Phi| \leq \frac{2m_0 \sqrt{C}}{\delta \eta}\right) = c.$$

同样地, 采用归纳法, 假设 $P(|V_1| \leq m_0 \sigma_k, \dots, |V_q| \leq m_0 \sigma_k) \leq c^q$,

$$\begin{aligned} & P(|V_1| \leq m_0 \sigma_k, \dots, |V_q| \leq m_0 \sigma_k, |V_{q+1}| \leq m_0 \sigma_k) \\ &= P(|V_{q+1}| \leq m_0 \sigma_k | |V_1| \leq m_0 \sigma_k, \dots, |V_q| \leq m_0 \sigma_k) \times c^q \\ &\leq c^q \cdot P\left(|\Phi| \leq \frac{2m_0 \sqrt{C}}{\delta \eta}\right) = c^{q+1}. \end{aligned}$$

由条件 (b) 可得

$$\sum_{j_k, s_k} P\left(M^+(j_k, s_k) \leq \frac{m_0}{\sqrt{C(A, s_A, j, s)}}\right) \leq 2p \cdot c^{|S|} \leq 2p \cdot c^{d_p} \rightarrow 0, \quad \text{当 } p \rightarrow \infty.$$

证毕. □

定理 3.3 的证明 由以上各引理可以得到 T_k 的概率如下:

$$\begin{aligned} P(T_k^{\text{EN}} > t) &= \sum_{j_k, s_k} P\left(\left(C(A, s_A, j, s) + \frac{\lambda_2}{\lambda_{1,k}^2} D(A, j, s, y)\right) \cdot \frac{g_{A,s_A}^{\text{EN}}(g_{A,s_A}^{\text{EN}} - M_{A,s_A}^{\text{EN}})}{\sigma^2} > t\right. \\ &\quad \left. \cap \{g_{A,s_A}^{\text{EN}}(j_k, s_k) > g_{A,s_A}^{\text{EN}}(j, s), \forall j, s\}\right) + o(1) \\ &= \sum_{j_k, s_k} P\left(\left(C(A, s_A, j, s) + \frac{\lambda_2}{\lambda_{1,k}^2} D(A, j, s, y)\right) \cdot \frac{g_{A,s_A}^{\text{EN}}(g_{A,s_A}^{\text{EN}} - M_{A,s_A}^{\text{EN}})}{\sigma^2} > t,\right. \\ &\quad \left. g_{A,s_A}^{\text{EN}} > M^+, g_{A,s_A}^{\text{EN}} < M^-, 0 > M^0\right) + o(1) \\ &= \sum_{j_k, s_k} P\left(\frac{g_{A,s_A}^{\text{EN}}(g_{A,s_A}^{\text{EN}} - M^+)}{\sigma^2 / (C(A, s_A, j, s) + \frac{\lambda_2}{\lambda_{1,k}^2} D(A, j, s, y))} > t,\right. \\ &\quad \left. g_{A,s_A}^{\text{EN}} > M^+, g_{A,s_A}^{\text{EN}} < M^-, 0 > M^0\right) + o(1) \\ &\leq \sum_{j_k, s_k} P\left(\frac{g_{A,s_A}^{\text{EN}}(g_{A,s_A}^{\text{EN}} - M^+)}{\text{Var}(g_{A,s_A}^{\text{EN}})} > t, g_{A,s_A}^{\text{EN}} > M^+, g_{A,s_A}^{\text{EN}} < M^-, 0 > M^0\right) + o(1). \end{aligned}$$

为了书写简便, 设 $\sigma_k^2 = \text{Var}(g^{\text{EN}}(j_k, s_k))$, 并采用与证明定理 3.2 类似的技巧, 将 g_{A,s_A}^{EN} 看作二元一次方程并求解下列方程:

$$\frac{g_{A,s_A}^{\text{EN}}(g_{A,s_A}^{\text{EN}} - M^+)}{\sigma^2} = t. \quad (\text{A.14})$$

(A.14) 的较大值解表示如下:

$$u(a, b) = \frac{b + \sqrt{b^2 + 4a}}{2}.$$

则有

$$\begin{aligned} & \sum_{j_k, s_k} \mathbb{P} \left(\frac{g_{A, s_A}^{\text{EN}} (g_{A, s_A}^{\text{EN}} - M^+)}{\sigma_k^2} > t, g_{A, s_A}^{\text{EN}} > M^+, g_{A, s_A}^{\text{EN}} < M^-, 0 > M^0 \right) + o(1) \\ &= \sum_{j_k, s_k} \mathbb{P} \left(\frac{g_{A, s_A}^{\text{EN}}}{\sigma_k} > u \left(t, \frac{M^+}{\sigma_k} \right), g_{A, s_A}^{\text{EN}} < M^-, 0 > M^0 \right) + o(1). \end{aligned}$$

由二元一次方程的性质得到下列不等式:

$$g < M^- + u(t\sigma^2, M^+) - M^+.$$

因此, T_k 与指数分布的差可以表示为

$$\begin{aligned} \mathbb{P}(T_k^{\text{EN}} > t) - e^{-(1+\lambda_2)t} &\leq \epsilon \sum_{j_k, s_k} \int_{m^+/\sigma_k \geq m_0} \left[\Phi \left(\frac{m^-}{\sigma_k} \right) - \Phi \left(\frac{m^+}{\sigma_k} \right) \right] G_{j_k, s_k}(dm^+, dm^-, dm^0) \\ &\quad + \sum_{j_k, s_k} \int_{m^+/\sigma_k \leq m_0} G_{j_k, s_k}(dm^+, dm^-, dm^0), \end{aligned} \quad (\text{A.15})$$

其中

$$G_{j_k, s_k}(dm^+, dm^-, dm^0) = 1\{m^+ < m^-, m^0 < 0\} \cdot F_{M^+, M^-, M^0}(dm^+, dm^-, dm^0).$$

(A.15) 不等式右边的第二项被引理 A.3 约束, 第一项小于 ϵ . □

A significance test for the elastic net and its asymptotic distribution with general predictors

Yuehan Yang & Lan Wu

Abstract This paper considers the significance test for the elastic net in high dimensional sparse linear regression models. A test statistic called Cov-EN test statistic is proposed which is based on the knots of the elastic net solution path. It is shown to have an $\text{Exp}(1)$ asymptotic distribution with general predictors. To prepare for the test statistic, we review KKT conditions and compute each knot of Lars algorithm for the elastic net solution path. Simulations and real examples given in this paper illustrate the performance of the Cov-EN test statistic and compare the results with the significance test for the Lasso.

Keywords significance test, model selection, covariance test statistic

MSC(2010) 62J05, 62J07

doi: 10.1360/SCM-2016-0427