Advances in Applied Mathematics 应用数学进展, 2025, 14(2), 329-340
Published Online February 2025 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2025.142074

Hans汉斯

基于秩回归的多阈值变平面模型研究

方贤珍

西南大学数学与统计学院，重庆

收稿日期：2025年1月24日；录用日期：2025年2月17日；发布日期：2025年2月25日

摘　要

为了更好地应用个性化医疗技术，识别出导致治疗效果出现异质性的亚组人群，进而为这些特定的患者群体提供更加精准有效的治疗方法，文章介绍了一种多阈值的变平面模型，该模型将研究对象划分为具有不同协变量效应的亚组。针对此模型，文章提出了一种新的基于秩回归的两阶段估计方法来确定亚组数量、阈值的位置以及其他的回归参数。在第一阶段，采用分组选择原则，一致地识别亚组的数量；在第二阶段，采用惩罚诱导平滑技术来精确估计变点位置和模型的其他参数。通过将该方法应用到艾滋病临床试验组研究175所得数据中，并对比其他的模型方法，可以发现基于秩回归的估计方法结果更好，同时该方法具有很好的鲁棒性。

关键词

亚组识别，诱导平滑，惩罚函数，稀疏解

Research on Multi-Threshold Variable Plane Model Based on Rank Regression

Xianzhen Fang

School of Mathematics and Statistics, Southwest University, Chongqing

Received: Jan. 24th, 2025; accepted: Feb. 17th, 2025; published: Feb. 25th, 2025

Abstract

In order to better apply personalized medical technology, identify subgroups that cause heterogeneity in treatment outcomes, and provide more accurate and effective treatment methods for these specific patient groups, this paper introduces a multi-threshold variable plane model, which divides the research subjects into subgroups with different covariate effects. This article proposes a new two-stage estimation method based on rank regression to determine the number of subgroups, the position of thresholds, and other regression parameters for this model. In the first stage, the principle

文章引用: 方贤珍. 基于秩回归的多阈值变平面模型研究[J]. 应用数学进展, 2025, 14(2): 329-340.
DOI: 10.12677/aam.2025.142074

of grouping selection is adopted to consistently identify the number of subgroups. In the second stage, the penalty-induced smoothing technique is used to accurately estimate the position of the inflection point and other parameters of the model. By applying the method proposed in this paper to the data obtained from the AIDS clinical trial group study 175, compared with other model methods, we can find that the estimation method based on rank regression has better results, and this method has good robustness.

Keywords

Subgroup Identification, Inducing Smoothness, Punishment Function, Sparse Release

1. 引言

随着大数据时代的到来，对每一个患者的医疗健康记录也越来越详细。通常情况下，不同患者对不同疾病的治疗方案会产生不同的反应，患者之间治疗效果的异质性使医疗决策复杂化，从而促使医疗研究者们从传统的“一刀切”方法过渡到量身定制疗法，即个性化医疗。

个性化医疗的一个新兴研究方向是使用适当的算法技术对患者进行分类，得到不同的亚组，并分析这些亚组的治疗效果。在现有的文献中，已经开发出了多种基于数据驱动的亚组识别方法。Zhao 等人[1]引入了结果加权学习(O-Learning)框架，从分类的角度直接寻找最佳的二元处理方法。Loh [2]提出了广义无偏交互检测与估计方法，旨在识别那些在接受治疗后效果得到显著提升的患者亚组。此外，Foster 等人[3]设计了一种虚拟双胞胎方法，用于发现治疗效果增强的特定亚组。同时，Cai 等人[4]和 Zhao 等人[5]则采用参数评分体系来估算个体患者的治疗差异，并据此筛选出从新治疗中可能获得更大益处的人群。

事实上，上述大多数亚组识别工作使用了与变点检测相似的技术[6]，并且可以使用传统的变点理论来严格证明。因此，可以考虑基于变点模型提出亚组识别方法，变点模型也称为阈值回归或分段回归模型，它的一个自然扩展是变平面模型。变平面模型相对于变点模型的优点是，它允许潜在的分组变量是协变量的线性组合，而不是局限于单个协变量，可以处理的变量更多。

到目前为止，由于难以处理由变平面定义的多维断点和不连续的亚组指标，对多阈值变平面模型只进行了少量的研究。Li 和 Jin [7]提出了一种基于惩罚的加速失效时间回归模型框架。在 Jin 等人[8]的研究中，他们将阈值问题定义为群体模型选择问题，并应用了快速计算工具。Wei 和 Kosorok [9]使用切片逆回归技术提出了协变量空间中具有变平面的不规则 Cox 模型，Fan 等人[10]使用双鲁棒性评分统计量考虑了使用变平面方法来检验亚组是否存在，但是该方法只允许一个阈值(即只有两个子组)，并且在一个单位球上搜索平方分数测试统计数据的最大值相当具有挑战性，尤其是在针对多个组时。为了解决上述问题，本文考虑了一个具有未知阈值的变平面模型，该模型允许存在多个变平面，自动生成具有不同协变量效应的亚组，从而促进个性化医疗在医学领域的应用[11]。

本文提出的方法其优势主要体现在以下三个方面。首先，本文研究的变平面的概念不只是使用预先指定的索引变量，而且允许协变量的线性组合，可以得到更有意义的亚组定义，进而为精准医疗提供更灵活的工具。其次，本文对普通的基于最小二乘的多阈值变平面模型估计进行了重要改进，考虑基于秩回归的多阈值变平面模型估计，将阈值识别问题转化为模型选择问题，所得结果在鲁棒性方面具有更显

著的优势。此外，在实践中，亚组可能只是在少数选定的协变量效应上有所不同，在其他协变量上具有相同的效应。因此，本文允许某些效应为零，从而获得稀疏解，这是通过惩罚诱导平滑估计方法实现的。

2. 多阈值变平面

2.1. 多阈值变平面模型基本理论

假设 $\left\{\left(Y_i, X_i^{\mathrm{T}}, Z_i^{\mathrm{T}}\right)^{\mathrm{T}}, i=1,\cdots,n\right\}$ 遵循以下具有 s 个阈值的变平面模型：

$$Y_i = X_i^{\mathrm{T}}\left[\beta + \sum_{j=1}^{s}\delta_j I\left(Z_i^{\mathrm{T}}\theta > a_j\right)\right] + \varepsilon_i,\ i=1,\cdots,n \tag{1}$$

其中 $X=\left(X_{i1},\cdots,X_{ip}\right)^{\mathrm{T}}$ 是 p 维向量，$Y_i\in\mathbb{R}$ 是与第 i 个研究对象 X_i 对应的响应变量。β 为基线组的协变量效应向量，δ_j 为第 j 个亚组相对于基线组的增强效应向量，其子组由变平面 $I\left(Z_i^{\mathrm{T}}\theta > a_j\right)$ 定义。当 $\delta=0$ 时，参数 θ 不被识别。Z_i 是不包含截距项的 d 维分组变量，θ 为变平面参数，$a_1,\cdots,a_s$ 是阈值位置，且满足 $a_1<a_2<\cdots<a_s$。ε_i 是由 ε_{ij} 组成的独立的误差项，$i=1,2,\cdots,n$，$1\le j\le p$，且 ε_i 的概率密度函数均为 $f(\cdot)$。ε_{ij} 是连续的随机变量，且对于任何成对的差异值 $\varepsilon_{ik}-\varepsilon_{jl}$，$i,j=1,\cdots,n$，$j,l=1,\cdots,p$，其中位数均为 0。在这个模型中，可以通过将不同的治疗组编码为 X_j 中的虚拟协变量来合并多个治疗效果，相应的系数 β 和 δ_j 表示不同处理方案的效果。为了更精确地识别模型，本文考虑对 θ 做出假设：$\theta\in\Theta=\left\{\theta\in\mathbb{R}^d:\|\theta\|=1,\theta_r>0,1\le r\le d\right\}$，且固定的第 r 个元素为正。

2.2. 基于秩回归的多阈值变平面模型求解

Hettmansperger [12]指出基于秩回归的估计方法是不依赖于总体分布的、具有稳健性的、高度有效的。此外，相对于一般的最小二乘估计法，秩回归方法能够有效地处理异方差和异常值问题。

记 $\eta=\left(\beta^{\mathrm{T}},\delta^{\mathrm{T}},a^{\mathrm{T}},\theta^{\mathrm{T}}\right)^{\mathrm{T}}$。根据 Jaeckel [13]的思想，结合秩回归的理论知识可知，秩估计的目的是最小化残差的离散度。因此，当 a_j，$j=1,\cdots,s$ 已知时，未知参数 β,δ,θ 可以通过最小化，如下目标函数来估计：

$$L_n(\eta)=\frac{1}{n}\sum_{i<j}\left|e_i-e_j\right| \tag{2}$$

其中 $e_i=Y_i-X_i^{\mathrm{T}}\beta-X_i^{\mathrm{T}}\sum_{j=1}^{s}\delta_j I\left(Z_i^{\mathrm{T}}\theta>a_j\right)$，并约束 $\theta\in\Theta$。

然而，在一般情况下，变平面的数量和位置都是未知的，而且在回归模型中，需要在选择出显著协变量的同时及时更新参数估计值。显然，公式(2)并不能同时进行显著变量选择和参数估计，因此考虑使用惩罚估计。在实践中，本文提出了一种迭代的两阶段多阈值变平面估计方法。在第一阶段中，对于任意给定的一致估计量 $\hat{\theta}$，可通过带有惩罚的变点检测算法得到 s 的一致估计量 $\hat{s}$。得到 $\hat{s}$ 之后，在第二阶段使用诱导平滑方法来估计参数 (β,δ,θ,a)。具体过程如下。

2.2.1. 初步分割阶段

对于给定的 $\hat{\theta}$，记 $\hat{W}_i=Z_i^{\mathrm{T}}\hat{\theta}, i=1,\cdots,n$，生成秩映射 $\left\{l_{(i)}:1\le i\le n\right\}$，使得 $\hat{W}_{l_{(i)}}$ 是 $\left\{\hat{W}_i:1\le i\le n\right\}$ 中的第 i 个最小值，为了避免位置估计不一致，将其按升序排列，即 $\hat{W}_{l_{(1)}}\le\hat{W}_{l_{(2)}}\le\cdots\le\hat{W}_{l_{(n)}}$。

首先，本文根据 $\hat{W}_{l_{(i)}}$ 将数据序列分为 q_n+1 段，其中当 $n\to\infty$ 时，q_n 趋于无穷大。即拆分数据序列使得第一段 $\mathcal{F}_1=\left\{i:\hat{W}_i\le\hat{W}_{l_{(n-q_nm)}}\right\}$，包含第 $n-q_nm$ 个观测值，其他 q_n 段 $\mathcal{F}_j=\left\{i:\hat{W}_{l_{(n-(q_n-j+2)m)}}\le\hat{W}_i\le\hat{W}_{l_{(n-(q_n-j+1)m)}}\right\}$，

$j=2,\cdots,q_n+1$，分别包含 m 个观测值，其中 $m=\lceil n/q_n \rceil$。定义 $b_j=\left|\mathcal{F}_j\right|, j=1,\cdots,q_n+1$，表示集合 $\mathcal{F}_j$ 中所含元素个数，即集合 $\mathcal{F}_j$ 的大小。令 $Y_{(j)}=\left(Y_i, i\in\mathcal{F}_j\right)^{\mathrm{T}}$，$X_{(j)}=\left(X_i, i\in\mathcal{F}_j\right)^{\mathrm{T}}$，$\tilde{Y}=\left(Y_{(1)}^{\mathrm{T}},\cdots,Y_{(q_n+1)}^{\mathrm{T}}\right)^{\mathrm{T}}$，$\tilde{X}=\left(X^{(1)},\cdots,X^{(q_n+1)}\right)$，$X^{(1)}=\left(X_{(1)}^{\mathrm{T}},\cdots,X_{(q_n+1)}^{\mathrm{T}}\right)^{\mathrm{T}}$，$X^{(j)}=\left(0_{p\times\sum_{i=1}^{j-1}b_j},X_{(j)}^{\mathrm{T}},\cdots,X_{(q_n+1)}^{\mathrm{T}}\right)^{\mathrm{T}}$，$j=2,\cdots,q_n+1$，则估计值 $\tilde{\gamma}'=\left(\tilde{\beta}^{\mathrm{T}},\tilde{\delta}_1'^{\mathrm{T}},\cdots,\tilde{\delta}_{q_n}'^{\mathrm{T}}\right)^{\mathrm{T}}$ 可以通过最小化以下带有惩罚的目标函数得到：

$$L_n'\left(\tilde{\gamma}'\right)=\frac{1}{n}\sum_{i<j}\left|e_i'-e_j'\right|+\sum_{j=1}^{q_n}p_{\lambda_n}\left(\left\|\delta_j'\right\|\right) \tag{3}$$

其中 $e_i'=Y_{(i)}-X^{(i)}\gamma_i$，$p_{\lambda_n}(\cdot)$ 为惩罚函数。

本文选取 Fan 和 Li [14]提出的具有较好性质的非凹惩罚函数，即平滑截断的绝对偏差(SCAD)惩罚，其定义为其一阶导数，并且是关于原点对称的。即对于 $\theta>0$，

$$p_{\lambda_n}'(\theta)=\lambda\left\{I\left(\theta\le\lambda_n\right)+\frac{\left(a\lambda-\theta\right)_+}{(a-1)\theta}I\left(\theta>\lambda_n\right)\right\}$$

其中 $a>2$，且 λ_n 是一个非负惩罚参数，用来控制模型的变量选择或稀疏性。一般地，当 p 较大时，通常假设 γ 为稀疏结构。现考虑使用迭代局部二次近似算法来找到式(3)的最小值。Leng [15]指出式(2)中的目标函数可以视为 Jaeckel [13]提出的 Wilcoxon 分数的秩离散函数，所以式(3)中的第一项可以近似为：

$$\frac{1}{n}\sum_{i<j}\left|e_i'-e_j'\right|\approx\sum_{i=1}^{n}\omega_i\left(e_i'-\xi\right)^2$$

其中 ξ 是 $\left\{e_i'\right\}_{i=1}^n$ 的中位数，且

$$\omega_i=\begin{cases}\dfrac{R\left(e_i'\right)}{n+1}-\dfrac{1}{2}, & e_i'\ne\xi\\ e_i'-\xi & \\ 0, & \text{其他}\end{cases}$$

使用简单的泰勒展开，给定目标函数(3)的初始估计 γ'（等效于给定了 β 和 δ_j'），可以相应地获得权重 ω_i 和残差的中位数 ξ。对于式(3)的第二项，本文考虑近似正则项为：

$$p_{\lambda_n}\left(\left\|\delta_j'\right\|\right)\approx p_{\lambda_n}\left(\left\|\tilde{\delta}_j'\right\|\right)+\frac{1}{2}\frac{p_{\lambda_n}'\left(\left\|\tilde{\delta}_j'\right\|\right)}{\left\|\tilde{\delta}_j'\right\|}\left\{\left\|\delta_j'\right\|^2-\left\|\tilde{\delta}_j'\right\|^2\right\}$$

因此，式(3)可以近似为：

$$L_n''\left(\tilde{\gamma}'\right)\approx\left(\tilde{Y}-\tilde{X}\tilde{\gamma}'\right)^{\mathrm{T}}\tilde{W}\left(\tilde{Y}-\tilde{X}\tilde{\gamma}'\right)+\frac{n^2}{2}\tilde{\gamma}'^{\mathrm{T}}\Omega\tilde{\gamma}' \tag{4}$$

其中 $\tilde{Y}=\left(Y_{(1)}^{\mathrm{T}},\cdots,Y_{(q_n+1)}^{\mathrm{T}}\right)^{\mathrm{T}}$，且

$$\Omega=diag\left(\frac{p_{\lambda_n}'\left(\left\|\tilde{\delta}_1'\right\|\right)}{\left\|\tilde{\delta}_1'\right\|},\cdots,\frac{p_{\lambda_n}'\left(\left\|\tilde{\delta}_{q_n}'\right\|\right)}{\left\|\tilde{\delta}_{q_n}'\right\|}\right),\quad \tilde{W}=diag\left(\tilde{\omega}_1,\cdots,\tilde{\omega}_n\right)$$

为方便起见，记估计量 $\tilde{\gamma}'=\left(\tilde{\gamma}_1'^{\mathrm{T}},\cdots,\tilde{\gamma}_{q_n+1}'^{\mathrm{T}}\right)$，$\hat{\mathcal{A}}=\left\{j:\tilde{\gamma}_j'\ne0, j=1,\cdots,q_n+1\right\}$，

$$\hat{\mathcal{A}}^*=\left\{j:j\in\hat{\mathcal{A}}, j-1\notin\hat{\mathcal{A}}, j=2,\cdots,q_n+1\right\}=\left\{\hat{\kappa}_1,\cdots,\hat{\kappa}_{\hat{s}}\right\},\ \hat{\kappa}_1<\cdots<\hat{\kappa}_{\hat{s}} \tag{5}$$

显然，由上述定义可知 $\hat{\mathcal{A}}^*$ 是 $\hat{\mathcal{A}}$ 的子集。若 $j-1\notin\hat{\mathcal{A}}, j\in\hat{\mathcal{A}}$ 和 $j+1\in\hat{\mathcal{A}}$，则 $j\in\hat{\mathcal{A}}^*$ 且 $j+1\notin\hat{\mathcal{A}}^*$。因此，当给定每个估计量 $\hat{\theta}$ 的值，就可以获得变平面的阈值数量估计值 $\hat{s}=\left|\hat{\mathcal{A}}^*\right|$。如果 $\hat{s}=0$，则表明没有子组。如果 $\hat{s}>0$，则真阈值 a_j 极有可能位于区间 $\left(\hat{W}_{l_{\left(n-\left(q_n-\hat{\kappa}_j+3\right)m\right)}},\hat{W}_{l_{\left(n-\left(q_n-\hat{\kappa}_j+1\right)m\right)}}\right), j=1,\cdots,\hat{s}$。若给定的估计量 $\hat{\theta}$ 是一致的，则初步分割阶段得到的估计值 $\hat{s}$ 也将以很高的概率收敛。

2.2.2. 平滑优化阶段

根据初步分割阶段得到的变平面阈值数量的估计值 $\hat{s}$，可通过最小化以下平滑目标函数来估计模型中的参数 a,θ 和 $\gamma=\left(\beta^{\mathrm{T}},\delta_1^{\mathrm{T}},\cdots,\delta_{\hat{s}}^{\mathrm{T}}\right)^{\mathrm{T}}$：

$$\frac{1}{n}\sum_{i<j}\left|e_i-e_j\right| \tag{6}$$

其中 $e_i=Y_i-X_i^{\mathrm{T}}\beta-X_i^{\mathrm{T}}\sum_{k=1}^{\hat{s}}\delta_k I\left(\frac{Z_i^{\mathrm{T}}\theta-a_k}{h}\right)$。

针对以上目标函数，本文考虑使用一个迭代估计过程，它可以产生相对稳定的解。对于给定的 θ，目标函数(6)可以简单地视为分段函数，可以通过秩回归方法来估计基线系数 β 和增强效应 δ。然而，对于给定的 γ，目标函数(6)不是连续的，找到它的最小值比较困难。因此，本文考虑使用光滑的分布函数 $\Phi(\cdot/h)$ 作为不连续的示性函数的平滑逼近，其中带宽 h 随着样本量 n 的增加逐渐趋于 0。如果 $Z_i^{\mathrm{T}}\theta>a_k$，则当 $h\to 0$ 时，$\Phi\left(\left(Z_i^{\mathrm{T}}\theta-a_k\right)/h\right)\to 1$。因此，目标函数(6)可以用如下函数来近似：

$$Q_n^*(\eta)=\frac{1}{n}\sum_{i<j}\left|e_i^*-e_j^*\right| \tag{7}$$

其中 $e_i^*=Y_i-X_i^{\mathrm{T}}\beta-X_i^{\mathrm{T}}\sum_{k=1}^{\hat{s}}\delta_k\Phi\left(\frac{Z_i^{\mathrm{T}}\theta-a_k}{h}\right)$。

记 $\tilde{\eta}^*=\arg\min_{\theta\in\Theta}\left\{Q_n^*(\eta)\right\}$。对于非稀疏问题，可以直接通过 Newton-Type 算法来最小化目标函数(7)。对于稀疏问题，考虑在式(7)中增加惩罚函数，即最小化以下惩罚目标函数来估计 η：

$$Q_n^{**}(\eta)=\frac{1}{n}\sum_{i<j}\left|e_i^*-e_j^*\right|+\sum_{j=1}^{q_n+1}p_{\lambda_n}\left(\left\|\gamma_j\right\|\right) \tag{8}$$

记 $\tilde{\eta}^{**}=\arg\min_{\theta\in\Theta}\left\{Q_n^{**}(\eta)\right\}$，可通过迭代诱导平滑方法来处理上述带有惩罚的目标函数，进而得到参数的估计值。具体的计算类似初步分割阶段的处理方法，对式(8)的第一项和第二项进行近似即可。

重复进行上述初步分割阶段和平滑优化阶段，直到满足收敛条件。尤其当变平面的估计数量保持不变时，考虑终止迭代。现将本文提出的求解基于秩回归的多阈值变平面模型的算法记为 TSMCPLD，其具体算法过程如下。

算法. TSMCPLD

步骤 0：给定 θ 的初始估计，记为 $\tilde{\theta}_{int}^*$，并设置 $\tilde{\theta}^*=\tilde{\theta}_{int}^*/\left\|\tilde{\theta}_{int}^*\right\|$；

步骤 1：实现初步分割阶段。最小化式(4)得到估计值 $\tilde{\gamma}^*$，然后计算式(5)中定义的索引集 $\hat{\mathcal{A}}^*$，通过 $\hat{s}=\left|\hat{\mathcal{A}}^*\right|$ 获得阈值数量；

步骤 2：对于给定的 $\hat{s}$，在平滑优化阶段中，通过最小化惩罚目标函数(8)来更新参数 $\left(\hat{\theta}^*,\hat{a}^*,\hat{\gamma}^*\right)$；

步骤 3：迭代步骤 1 和 2 直至收敛。

在上述算法中，初步分割阶段的性能依赖于片段长度 m，并且最佳 m 的选择可以遵循[7]中的建议，将数据序列进行 L 次初步分割阶段，事件(不包括第一段)的公共长度为 $m_\ell, \ell=1,\cdots,L$，设 $m_\ell=\lfloor \kappa_\ell \sqrt{n} \rfloor, \ell=1,\cdots,L$，其中 κ_ℓ 取区间内 ℓ 个网格点的值，贝叶斯信息准则(BIC)可用于选择最佳索引 ℓ。

在平滑优化阶段的具体计算中，可以使用 R 包 BB 中的 BBoptim 函数来优化高维非线性目标函数。此外，参数的数量可能相当大，尤其是有大量的异构子组时，包含惩罚函数会导致稀疏解。在中等或高维环境下，调谐参数 λ_n 也可以通过贝叶斯信息准则(BIC)来选择。具体的计算公式为：

$$\mathrm{BIC}(\lambda_n)=n\log(\mathrm{RSS}/n)+\log(n)(p+1)(\hat{s}+1)$$

其中 RSS 为残差平方和。

3. 渐进理论性质

为了更好地建立渐近理论性质，本文考虑增加以下必要条件：

条件 1：a) $E(X_iX_i^{\mathrm{T}})=\Sigma_0$ 是有限且正定的。$E(Z_iZ_i^{\mathrm{T}})$ 是正定的。Z_i 和 ε_i 是独立的，$i=1,\cdots,n$。$E(\varepsilon_i \mid X_i)=0$ 几乎必然成立。

b) 对于某些 $\xi>1$，设 $0<E\left\|(X_i^{\mathrm{T}},Z_i^{\mathrm{T}})^{\mathrm{T}}(X_i^{\mathrm{T}},Z_i^{\mathrm{T}})\right\|^{\xi}<\infty$，且 $E\left\|(X_i^{\mathrm{T}},Z_i^{\mathrm{T}})^{\mathrm{T}}\varepsilon_i\right\|^{\xi}<\infty$，$E(X_iX_i^{\mathrm{T}} \mid Z_i)>0$。

条件 2：参数 η 的参数空间是紧空间，其中 $\min_{1\le l\le k\le s}\{|a_l-a_k|\}$ 且 $\min\left\{\left\|\beta/\sqrt{p}\right\|,\left\|\delta_1/\sqrt{p}\right\|,\cdots,\left\|\delta_s/\sqrt{p}\right\|\right\}$ 以零为界。设 $\rho(t)=\lambda_n^{-1}p_{\lambda_n}(t)$ 且 $\overline{\rho}(t)=\rho'(|t|)\mathrm{sgn}(t)$。本文假设罚函数 $p_{\lambda_n}(\cdot)$ 满足以下条件：

条件 3：$p_{\lambda_n}(\cdot)$ 是一个对称函数，它在 $[0,\infty)$ 上是不递减的凹函数。存在一个常数 $v>0$，使得 $\rho(t)$ 对于所有 $|t|\ge v\lambda_n$ 都是常数，且 $\rho(0)=0$。除了有限的 t，$\rho'(t)$ 存在且连续，而且 $\rho'(0+)=1$。本文设 $a_0=-\infty$ 且 $a_{s+1}=\infty$。记 $\mathcal{F}_j^0=\{i:a_{j-1}<Z_i^{\mathrm{T}}\theta<a_j\}, j=1,\cdots,s$，其包含真实向量阈值位置 a，变平面参数 θ。与本文中 $\tilde{X}$ 和 $\tilde{Y}$ 的定义类似，可以通过将 $\mathcal{F}_j^0$ 替换为 $\mathcal{F}_j$ 来定义 $\tilde{X}_a$ 和 $\tilde{Y}_a$。根据条件 1 可知，$\tilde{X}_a^{\mathrm{T}}\tilde{X}_a/n\xrightarrow{a.s.}\Upsilon$，其中 Υ 是一个正定矩阵。为了获得式(8)中 $\hat{\gamma}^*$ 的渐近性质，考虑以下假设条件：

条件 4：$\max_{u\ge 0}\{p''_{\lambda_n}(u)\}+\Lambda_{(s+1)p}(\Upsilon)>0$，其中 $\Lambda_{(s+1)p}(\Upsilon)$ 是 Υ 的最小特征值。

条件 5：设置 $W_i=Z_i^{\mathrm{T}}\theta$ 且 $V_i=(X_i,Z_i)$，$f_{W|V}(\cdot)$ 表示给定 $V_i=V$ 时 W_i 的条件密度，$f_W(\cdot)$ 表示 W_i 的密度，其中 $f_{W|V}(\cdot)$ 有紧支撑且二阶导数有界。$P(W_i\le a_j)=\tau_j, 0<\tau_1<\cdots<\tau_s<1$。此外，对于某些 $M<\infty$，$E(\varepsilon_i^4 \mid V_i)<M$ 几乎必然成立。

条件 6：$h\to 0$ 和 $nh^2\to 0$，当 $n\to\infty$ 时。

设计矩阵的条件 1 允许特定制度的异方差性。误差假设可以被放宽为 $\varepsilon_i=\sigma(X_i^{\mathrm{T}}\beta)e_i$，其中 e_i 独立于 X_i，且 $e_1,\cdots,e_n$ 是 i.i.d.的，均值为零和方差为 σ^2。条件 2 是关于参数空间的，它通过要求 $a_{j-1}<a_j, j=1,\cdots,s$ 来排除小于 $s+1$ 个亚组的简化模型的可能性。条件 3 和 4 通常用于高维数据环境的收缩回归中。SCAD 这类凹形惩罚满足条件 3，对于 SCAD 惩罚，条件 4 等价于 $\Lambda_{(s+1)p}(\Upsilon)>1/(v-1)$，这确保了目标函数(8)是全局凸的。条件 5 是一个标准的平滑条件，其表明存在 s 个不同的跳跃，否则模型无法识别。条件 6 可以来确定 h 的速率。

根据大数定律和条件 5，可以得到 $\sum_{i=1}^{n}I(a_{j-1}<W_i\le a_j)/n\xrightarrow{p}\tau_j-\tau_{j-1}>0$，且当 $m\to\infty$ 和 $m/n\to 0$ 时，有 $\sum_{i=1}^{n}I\left(\hat{W}_{l_{(n-(q_n-j+2)m)}}<\hat{W}_i\le\hat{W}_{l_{(n-(q_n-j+1)m)}}\right)/n=m/n\to 0$。因此，当 θ 是已知的或是一致估计值，即

$\hat{W}_i = W_i + o_p(1)$ 时，对于足够大的 n，在每个分段 $\mathcal{F}_j = \left\{ \hat{W}_{l_{(n-(q_n-j+2)m)}} < \hat{W}_i \le \hat{W}_{l_{(n-(q_n-j+1)m)}} \right\}$ 中，最多有一个阈值，其中，$\hat{W}_{l_{(n-(q_n-j+2)m)}}$ 和 $\hat{W}_{l_{(n-(q_n-j+1)m)}}$，$j = 1, \cdots, q_n + 1$ 的定义在第 2.2.1 节中。由以下定理 1 即可保证初步分割阶段得到变平面数量估计值 $\hat{s}$ 是 s 的一致估计。

定理 1. 假设 $m \to \infty$ 且 $m = O(n^r)$，其中 $0 < r \le 1/2$ 为常数。当 $n \to \infty$ 时，$\lambda_n \to 0$ 和 $\lambda_n \sqrt{n} / \log n \to \infty$。如果条件 1~5 成立，那么有 $\lim_{n\to\infty} P(\hat{s} = s) = 1$。

设 $\gamma = \left(\gamma_1, \cdots, \gamma_{(s+1)p}\right)^{\mathrm{T}} = \left(\beta^{\mathrm{T}}, \delta_1^{\mathrm{T}}, \cdots, \delta_s^{\mathrm{T}}\right)^{\mathrm{T}}$ 是(1)中的回归参数，$\mathcal{G} = \left\{ j : \gamma_j \ne 0, j = 1, \cdots, (s+1)p \right\}$ 是模型中的重要变量集。对于给定的一致估计量 $\hat{s}$，由定理 1 可以获得最小化非正则目标函数(7)的平滑秩回归估计量 $\tilde{\eta}^*$ 的一致性，其中 $s = 1$。本文考虑通过最小化惩罚平滑目标函数(8)来获得估计量 $\hat{\eta}^*$，以下定理保证了我们估计量的一致性。

定理 2. 在条件 1~6 下，当 $n \to \infty$ 时，$\hat{s} = s$ 且 $\lambda_n \to 0$，$L_n^*(\eta)$ 有局部最小值 $\hat{\eta}^*$，使得 $\left\| \hat{\gamma}^* - \gamma \right\| = O_p\left(\sqrt{1/n}\right)$，$\left\| \hat{a}^* - a \right\| = O_p\left(\sqrt{h/n}\right)$，且 $\left\| \hat{\theta}^* - \theta \right\| = O_p\left(\sqrt{h/n}\right)$，其中 $\left\| \hat{\theta}^* \right\| = \left\| \theta \right\| = 1$。

现重新记 $\mathcal{G} = \{g_1, \cdots, g_{s+1}\}$，其中 $g_{j+1} = \left\{ j_1, \cdots, j_{p_j} \right\}$ 是第 j 个子群中 p_j 个非零协变量集的索引集，$j = 0, 1, \cdots, s$。不失一般性，考虑将 γ 记为 $\gamma_p = \left(\gamma_{(1)}^{\mathrm{T}}, \gamma_{(2)}^{\mathrm{T}}\right)^{\mathrm{T}}$，其中 $\gamma_{(1)} = \left(\gamma_{g_1}^{\mathrm{T}}, \cdots, \gamma_{g_{s+1}}^{\mathrm{T}}\right)$，$\gamma_{g_{j+1}} = \left(\gamma_{j_1}, \cdots, \gamma_{j_{p_j}}\right)^{\mathrm{T}}$ 且 $\gamma_{(2)} = 0$。记 $X_{i,g_{j+1}} = \left(X_{i,j_1}, \cdots, X_{i,j_{p_j}}\right)^{\mathrm{T}}$，$j = 0, 1, \cdots, s$。记 $\Sigma_1 = \left(\sigma_{1,jk}\right)_{0 \le j,k \le s}$ 表示 $\sum_{j=0}^{s} p_j \times \sum_{j=0}^{s} p_j$ 个分块矩阵，其中每一块矩阵分别为 $\sigma_{1,jk} = 4\sigma^2 E X_{i,g_{j+1}} X_{i,g_{k+1}}^{\mathrm{T}} I\left(Z_i^{\mathrm{T}}\theta > a_j \vee a_k\right)$，$\Sigma_2 = \frac{4}{nh} diag\left\{ \frac{\sigma^2}{2\sqrt{\pi}} A_j + \Pi \cdot B_j, j = 2, \cdots, s \right\}$ 是 $s \times s$ 维对角矩阵，其中 $A_j = E\left\{ \left(\delta_j^{\mathrm{T}} X_i\right)^2 \mid Z_i^{\mathrm{T}}\theta = a_j \right\} f_W\left(a_j\right)$，$B_j = E\left\{ \left(\delta_j^{\mathrm{T}} X_i\right)^4 \mid Z_i^{\mathrm{T}}\theta = a_j \right\} f_W\left(a_j\right)$，且 $\Pi = \int_{-\infty}^{\infty} \phi(s)^2 \left(I(s > 0) - \Phi(s)\right)^2 \mathrm{d}s$，而且 $\Sigma_3 = \frac{4}{nh} \sum_{j=1}^{s} \left\{ \frac{\sigma^2}{2\sqrt{\pi}} G_j + \Pi \cdot H_j \right\}$ 是 $d \times d$ 维矩阵，其中 $G_j = E\left\{ Z_i Z_i^{\mathrm{T}} \left(\delta_j^{\mathrm{T}} X_i\right)^2 \mid Z_i^{\mathrm{T}}\theta = a_j \right\} f_W\left(a_j\right)$，$H_j = E\left\{ Z_i Z_i^{\mathrm{T}} \left(\delta_j^{\mathrm{T}} X_i\right)^4 \mid Z_i^{\mathrm{T}}\theta = a_j \right\} f_W\left(a_j\right)$。

设 $V_{11} = \left(v_{1k,l}\right)_{0 \le k,l \le s}$，其中 $v_{1k,l} = 2E X_{i,g_{k+1}} X_{i,g_{k+1}}^{\mathrm{T}} I\left(Z_i^{\mathrm{T}}\theta > a_k \vee a_l\right)$，$V_{22} = diag\left(\frac{A_k}{\sqrt{\pi}}, k = 1, \cdots, s \right)$，$V_{23} = \left(v_{23vk}\right)_{1 \le v \le d, 1 \le k \le s}$ 且 $v_{23vk} = -\frac{1}{\sqrt{\pi}} E\left[Z_{iv} \left(\delta_{g_{k+1}}^{\mathrm{T}} X_{i,g_{k+1}}\right)^2 \mid Z_i^{\mathrm{T}}\theta = a_k \right] f_W\left(a_k\right)$，$V_{33} = \frac{1}{\sqrt{\pi}} \sum_{j=1}^{s} G_j$。记：

$$\Gamma_{\lambda_n} = diag\left\{ p''_{\lambda_n}\left(\left|\gamma_{0_1}\right|\right), \cdots, p''_{\lambda_n}\left(\left|\gamma_{0_{p_0}}\right|\right), \cdots, p''_{\lambda_n}\left(\left|\gamma_{s_1}\right|\right), \cdots, p''_{\lambda_n}\left(\left|\gamma_{s_{p_s}}\right|\right) \right\}$$

以下定理中可以证明估计量的极限分布是成立的。

定理 3. 在条件 1~6 下，当 $n \to \infty$ 时，$\lambda_n \to 0$ 和 $\lambda_n \sqrt{n} / \log n \to \infty$，在概率趋近于 1 的情况下，定理 2 中的惩罚平滑估计量 $\hat{\eta}^* = \left(\hat{\gamma}_{(1)}^*, \hat{\gamma}_{(2)}^*, \hat{a}^*, \hat{\theta}^*\right)$ 满足：

a) 稀疏性：$\hat{\gamma}_{(2)}^* \to 0$ *a.s.*

b) 渐近正态性：

$$\sqrt{n}\left(\hat{\gamma}_{(1)}^* - \gamma_{(1)}\right) \xrightarrow{D} N\left(0, \left(V_{11} + \Gamma_{\lambda_N}\right)^{-1} \Sigma_1 \left(V_{11} + \Gamma_{\lambda_n}\right)^{-1}\right)$$

$$\sqrt{n/h}\tilde{V}\begin{pmatrix} \hat{a}^* - a \\ \hat{\theta}^* - \theta \end{pmatrix} \xrightarrow{D} N\left(0, \Omega\right)$$

其中 $\tilde{V}=\begin{pmatrix} V_{22} & V_{23} \\ P_\theta V_{23}^{\mathrm{T}} & P_\theta V_{33} \end{pmatrix}$，$\Omega=diag\left(\Sigma_2, P_\theta\Sigma_3\right)$ 且 $p_\theta=1-\theta\theta^{\mathrm{T}}$。此外，$\sqrt{n/h}\tilde{V}\begin{pmatrix} \hat{a}^*-a \\ \hat{\theta}^*-\theta \end{pmatrix}$ 和 $\sqrt{n}\left(\hat{\gamma}_{(1)}^*-\gamma_{(1)}\right)$ 都是渐进独立的。

定理 3 可以为许多比我们更简单但文献中尚未研究的模型提供推理工具。例如，考虑一维阈值变量(即 $d=1$)的情况很有趣，其中 $\theta=1$，$Z_i^{\mathrm{T}}\theta=Z_i$。然后，可以通过本文提出的估计方法估计 $\hat{\eta}^*$，并在以下推论中得到所得估计量的分布理论：

推论 1. 假设条件 1~6 成立，那么有 $\lim_{n\to\infty}P\left(\hat{s}=s\right)=1$，此外，$\sqrt{n}\left(\gamma^*-\gamma\right)$ 和 $\sqrt{n/h}\left(\hat{a}^*-a\right)$ 是渐近独立的，且有：

$$\sqrt{n}\left(\hat{\gamma}_{(1)}^*-\gamma_{(1)}\right)\xrightarrow{D}N\left(0,\left(V_{11}-\Gamma_{\lambda_N}\right)^{-1}\Sigma_1\left(V_{11}-\Gamma_{\lambda_n}\right)^{-1}\right)$$

$$\sqrt{n/h}\left(\hat{a}^*-a\right)\xrightarrow{D}N\left(0,V_{22}^{-1}\Sigma_2V_{22}^{-1}\right)$$

4. 实证分析

现考虑将本文所提出的方法应用于艾滋病临床试验组研究 175 所得数据。该数据是 R 软件的内置数据集(ACTG175)，可直接在 R 中调用。这项随机临床试验比较了不同药物治疗对感染人类免疫缺陷病毒 I 型成人的治疗效果[16]。本文的研究目标是对获取到的患者数据进行亚组分析，以便在(20 ± 5)周时得到更精确的每个组的 CD4 计数(细胞/mm^3)预测值。现结合实际病例与本文的研究目标，考虑选择以下 9 个协变量进行分析，具体见下表 1。

Table 1. Various covariates and their meanings
表 1. 各协变量及其含义

协变量	含义	单位
X_1	血友病(0 = 否，1 = 是)	—
X_2	性别(0 = 女性，1 = 男性)	—
X_3	基线 CD4 计数	—
X_4	直到首次出现以下情况的天数： 1) CD4T 细胞计数至少下降 50； 2) 表明进展为艾滋病的事件；3) 死亡	天
X_5	年龄	岁
X_6	重量	kg
X_7	Karnofsky 分数	—
X_8	基线 CD8 计数	—
X_9	之前接受抗逆转录病毒治疗的天数	天

本文首先将 $X_i=\left(1,X_{i1},\cdots,X_{i9}\right)^{\mathrm{T}}$ 作为自变量，$Y=20\pm5$ 周时的 CD4 计数作为因变量，结合相应数据拟合得到线性回归模型，即认为其没有亚组，并用 $\hat{\beta}^{ols}$ 表示对数据进行普通的最小二乘估计得到的各个参数的估计值。选择 $Z_i=\left(X_{i1},X_{i2},X_{i3},X_{i4},X_{i5}\right)^{\mathrm{T}}$ 作为阈值变量。

现考虑在艾滋病临床试验组研究所得数据中随机选取 500 名患者对应的变量数据，将其应用于本文所提出的基于秩回归估计的多阈值变平面模型中，可以得到系数估计值 $\hat{\beta}$ 和 $\hat{\delta}$，以及它们的标准误差(SE)，具体见下表 2。

由表 2 分别可以得到两个亚组的各种解释变量的推论。例如，第一组基线 CD4 计数变量 X_3 的系数

为 0.452，前两组间的系数差为 0.011，说明第二亚组的系数为 0.452 + 0.011 = 0.463，第二和第三亚组之间的系数差为 0.202，即说明第三亚组的系数为 0.463 + 0.202 = 0.665。然而，在没有进行的亚组分析的简单线性回归模型中，对于基线 CD4 计数变量，普通最小二乘法只返回一个恒定的系数 0.538，没有考虑到亚组内的基线 CD4 计数差异。其他系数也可以类似地解释。

Table 2. Estimated values of various coefficients and corresponding standard errors
表 2. 各系数估计值以及相应的标准误差

	$\hat{\beta}$		$\hat{\delta}_1$		$\hat{\delta}_2$		$\hat{\beta}^{ols}$	
	系数	标准误差	系数	标准误差	系数	标准误差	系数	标准误差
X_0	−0.073	−0.054	0.123	0.105	0.241	0.082	−0.003	0.126
X_1	−0.165	0.015	−0.002	0.025	−0.011	0.104	−0.086	0.024
X_2	0.001	0.020	0	—	0	—	0.023	0.066
X_3	0.452	0.052	0.011	0.035	0.202	0.032	0.538	0.035
X_4	0.222	0.048	0	—	0.053	0.024	0.244	0.043
X_5	0	—	−0.036	0.045	0	—	−0.025	0.014
X_6	−0.006	0.014	0	—	0	—	−0.006	0.031
X_7	0.075	0.031	0.018	0.021	0.163	0.012	0.032	0.019
X_8	0.079	0.019	−0.023	0.029	−0.105	0.084	−0.054	0.034
X_9	−0.175	0.042	−0.234	0.037	−0.047	0.033	−0.087	0.051

现考虑将选取的艾滋病临床试验数据以 7:3 的比例分为训练集和测试集两部分，选择前 70%作为训练集，后 30%作为测试集。将得到的训练集对应的数据带入基于秩回归的多阈值变平面模型中，通过使用本文提出的两阶段计算方法，可以得到各参数估计结果，再将各参数估计结果带入测试集中，结合测试集中自变量对应的数据，即可获得相应的因变量，即 CD4 计数预测值。类似地，可以得到基于普通最小二乘估计的多阈值变平面模型的各参数估计结果，将其带入测试集中也可以得到预测的 CD4 计数。

根据得到的两种不同方法的 CD4 计数的预测值，借助 R 软件，可以画出它们预测的 CD4 计数对比的散点图，具体见下图 1。

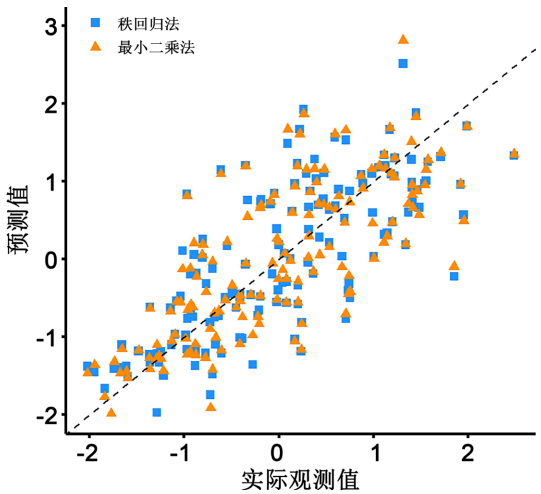
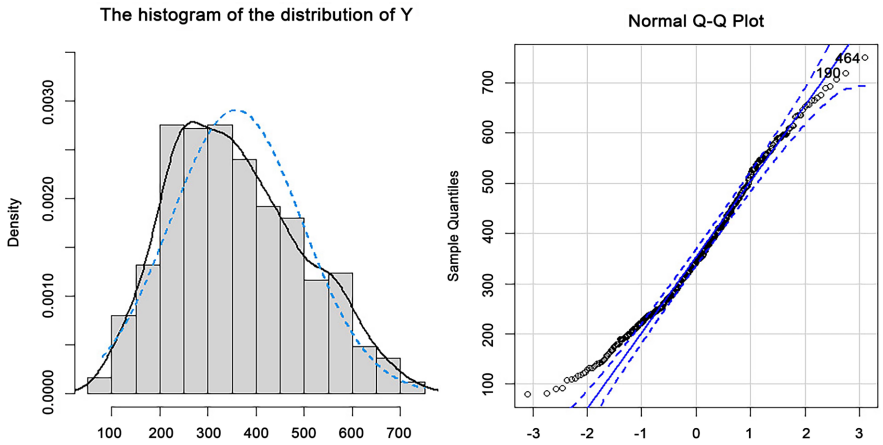


Figure 1. Comparison of prediction results by different methods scatter plot
图 1. 不同方法预测结果对比散点图

由上图 1 得到的不同方法预测结果的散点图对比，可以发现秩回归估计方法和最小二乘估计方法的预测值都与实际观测值有微小差距，因此考虑绘制实际观测的 CD4 计数的数据分布直方图和 Q-Q 图，进一步说明本文选取的秩回归估计方法的合理性，具体见下图 2。



(a) (b)

Figure 2. Histogram of CD4 count data distribution (left) and Q-Q chart (right)
图 2. CD4 计数的数据分布直方图(左)和 Q-Q 图(右)

根据图 2 展示的 CD4 计数数据分布直方图与 Q-Q 图可知，该数据并未完全遵循正态分布，特别是在分布的两端尾部，部分数据点显著偏离了正态分布的参照线，这表明数据集中存在异常值或离群数据，例如第 190 个与第 464 个数据点。因此，相较于一般的适用于独立同分布数据的最小二乘估计法，本文所提出的基于秩回归的多阈值变平面模型估计方法在此类数据上表现出更高的适用性。该方法展现出良好的鲁棒性，其估计效果更为优越。

此外，考虑将本文所提出的基于秩回归估计的多阈值变平面模型(MCPL)的预测性能与基于秩回归估计的单阈值变化平面(SCPL)模型、相关文献[7]提出的单阈值变点模型(MCPT)以及等加权多变量 $Z_i=\left(X_1+X_2+X_3+X_4+X_5\right)/5$ 的变平面模型(E-MCPL)的预测性能进行了比较。由于本例选取的数据中，X_1、X_2 为不连续变量，所以不能应用于单阈值变点模型。本文总结了上述提出的所有方法的预测误差，具体见下表 3。

Table 3. Prediction errors of model estimation results using different methods
表 3. 不同方法的模型估计结果的预测误差

方法	预测误差	$\hat{s}$	阈值 $\hat{a}$	分组结果
MCPL	0.565	2	(−0.436, 0.315)	390:66:44
SCPL	0.581	1	0.213	105:395
MCPT-X_3	0.578	1	−0.102	198:302
MCPT-X_4	0.585	0	—	—
MCPT-X_5	0.575	2	(−1.312, 1.513)	148:305:47
E-MCPL	0.587	0	—	—
OLS	0.597	0	—	—

注：MCPL 表示本文提出的基于秩回归的多阈值变平面模型估计方法；SCPL 表示基于秩回归的单阈值变平面模型估计方法；MCPT-X_3 表示阈值变量为 X_3 的单阈值变点模型估计方法；MCPT-X_4 表示阈值变量为 X_4 的单阈值变点模型估计方法；MCPT-X_5 表示阈值变量为 X_5 的单阈值变点模型估计方法；E-MCPL 表示等加权多变量 $Z_i=\left(X_1+X_2+X_3+X_4+X_5\right)/5$ 的变平面模型估计方法；OLS 表示普通的最小二乘模型估计方法。

由上表 3 结果可知，对比其他几种方法的估计结果，本文提出的基于秩回归的多阈值变平面模型估计方法的预测误差是最小的。即表明本文所提出的方法估计更精确、更有效。此外，为了研究亚组，本文总结了几种模型估计方法的分组结果，见上述表 3 的最后一列。现考虑通过所有协变量对应的不同亚组的平均值来展示其分组结果，具体见下图 3。

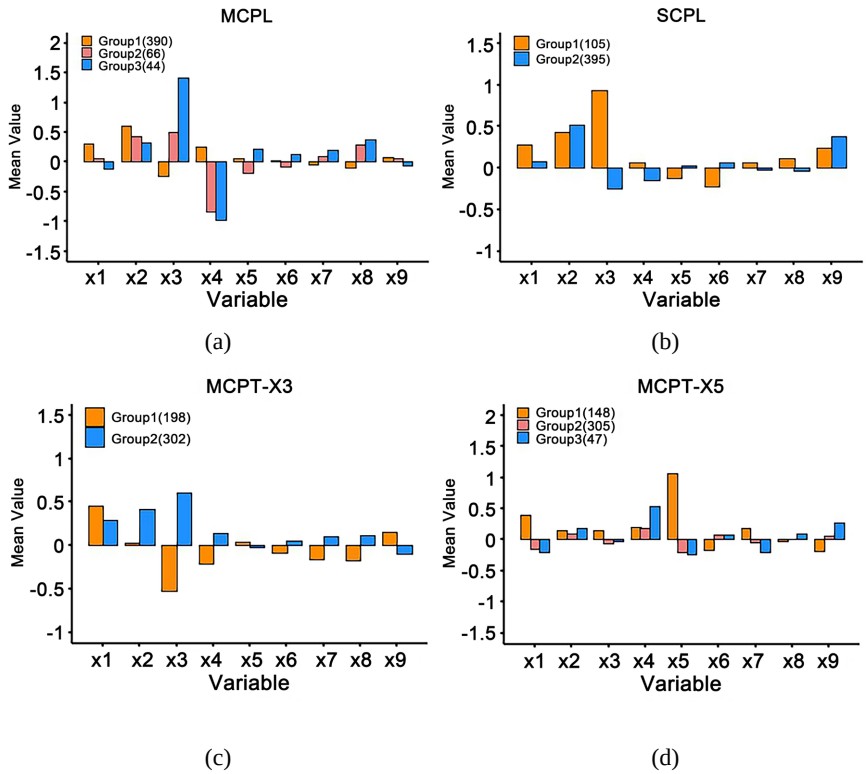


Figure 3. Average values of covariates corresponding to subgroups for different model estimation methods
图 3. 不同模型估计方法的协变量对应亚组的平均值

由上述图 3 中不同模型估计方法得到的每个协变量分组后组内平均值的条形分布图可知，本文提出的基于秩回归的多阈值变平面模型的分组效果更显著，能够更好地根据病人的不同特征(即自变量)将其分到相应的组别，采取适合该病人的治疗方法，即更好地体现了个性化医疗。

5. 结论

本文的创新点是，相对于普通的最小二乘估计法，本文所提出的基于秩回归的多阈值变平面模型估计方法具有很好的鲁棒性，而且此方法还添加了惩罚函数，可以扩展到高维的数据分析。此外，当特征空间的维数超高且为样本大小的指数阶时，可以通过基于秩回归的多阈值变平面模型估计方法来估计单个或多个标记的影响，进而考虑单个或多个阈值的变化，以便得到更精确的结果。

本文得出的结论可以应用于医学中的个性化医疗技术，以便为更多病情复杂的患者找到适合他们的治疗方案，从而更好、更有效地对其进行精准治疗，帮助患者早日康复，不再忍受病痛的折磨。

参考文献

[1] Zhao, Y., Zeng, D., Rush, A.J. and Kosorok, M.R. (2012) Estimating Individualized Treatment Rules Using Outcome

Weighted Learning. *Journal of the American Statistical Association*, **107**, 1106-1118. https://doi.org/10.1080/01621459.2012.695674

[2] Loh, W.Y. (2002) Regression Trees with Unbiased Variable Selection and Interaction Detection. *Statistica Sinica*, **12**, 361-386. http://www.jstor.org/stable/24306967

[3] Foster, J.C., Taylor, J.M.G. and Ruberg, S.J. (2011) Subgroup Identification from Randomized Clinical Trial Data. *Statistics in Medicine*, **30**, 2867-2880. https://doi.org/10.1002/sim.4322

[4] Cai, T., Tian, L., Wong, P.H. and Wei, L.J. (2010) Analysis of Randomized Comparative Clinical Trial Data for Personalized Treatment Selections. *Biostatistics*, **12**, 270-282. https://doi.org/10.1093/biostatistics/kxq060

[5] Zhao, L., Tian, L., Cai, T., Claggett, B. and Wei, L.J. (2013) Effectively Selecting a Target Population for a Future Comparative Study. *Journal of the American Statistical Association*, **108**, 527-539. https://doi.org/10.1080/01621459.2013.770705

[6] Bai, J. (1997) Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics*, **79**, 551-563. https://doi.org/10.1162/003465397557132

[7] Li, J. and Jin, B. (2018) Multi-threshold Accelerated Failure Time Model. *The Annals of Statistics*, **46**, 2657-2682. https://doi.org/10.1214/17-aos1632

[8] Jin, B., Shi, X. and Wu, Y. (2011) A Novel and Fast Methodology for Simultaneous Multiple Structural Break Estimation and Variable Selection for Nonstationary Time Series Models. *Statistics and Computing*, **23**, 221-231. https://doi.org/10.1007/s11222-011-9304-6

[9] Wei, S. and Kosorok, M.R. (2018) The Change-Plane Cox Model. *Biometrika*, **105**, 891-903. https://doi.org/10.1093/biomet/asy050

[10] Fan, A., Song, R. and Lu, W. (2017) Change-plane Analysis for Subgroup Detection and Sample Size Calculation. *Journal of the American Statistical Association*, **112**, 769-778. https://doi.org/10.1080/01621459.2016.1166115

[11] Li, J., Li, Y., Jin, B. and Kosorok, M.R. (2021) Multithreshold Change Plane Model: Estimation Theory and Applications in Subgroup Identification. *Statistics in Medicine*, **40**, 3440-3459. https://doi.org/10.1002/sim.8976

[12] Hettmansperger, T.P. and McKean, J.W. (1978) Statistical Inference Based on Ranks. *Psychometrika*, **43**, 69-79. https://doi.org/10.1007/bf02294090

[13] Jaeckel, L.A. (1972) Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals. *The Annals of Mathematical Statistics*, **43**, 1449-1458. https://doi.org/10.1214/aoms/1177692377

[14] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. https://doi.org/10.1198/016214501753382273

[15] Leng, C. (2010) Variable Selection and Coefficient Estimation via Regularized Rank Regression. *Statistica Sinica*, **20**,167-181. http://scholarbank.nus.edu.sg/handle/10635/105457

[16] Tsiatis, A.A., Davidian, M., Zhang, M. and Lu, X. (2008) Covariate Adjustment for Two-sample Treatment Comparisons in Randomized Clinical Trials: A Principled Yet Flexible Approach. *Statistics in Medicine*, **27**, 4658-4677. https://doi.org/10.1002/sim.3113