广义线性模型中条件主效应的自适应双层变量选择

Kexin Xie 和 Xinwei Deng[1]

美国弗吉尼亚理工学院统计学系

arXiv:2603.06871v1 [stat.ME] 6 Mar 2026

摘要

理解变量之间的交互效应，对于各种应用中的回归建模至关重要。传统方法通常将变量的乘积作为交互作用的量化方式，但这种方法往往缺乏清晰的可解释性，尤其是在复杂系统中。条件主效应（CME）这一概念提供了更直观、可解释的框架：通过量化一个变量在另一个变量不同水平下的效应来捕捉交互效应。近期提出的一种名为 $cmenet$ 的方法，利用条件主效应自然的分组结构（如同胞组和表亲组），通过惩罚实现条件主效应的双层选择。然而，cmenet方法仍存在若干局限，包括组内条件主效应惩罚项的耦合能力不足、组间惩罚缺乏自适应性，以及仅适用于连续响应的线性模型。为克服这些局限，我们提出了一种适用于广义线性模型（GLM）框架下条件主效应选择的自适应 cmenet 方法。所提方法采用带自适应权重的惩罚似然方法，以实现有效的双层变量选择，并提升组间和组内选择的效果。此外，我们还通过采用迭代加权最小二乘过程，开发了一种高效的参数估计算法。我们通过模拟研究以及基因关联分析中的真实数据研究，评估了所提方法的性能。

关键词： 条件效应；自适应权重；坐标下降；基因关联；交互分析；高维数据

[1]通信地址：Xinwei Deng，教授，弗吉尼亚理工学院统计学系，美国弗吉尼亚州布莱克斯堡，邮编24061（电子邮箱：xdeng@vt.edu）。

1 引言

在统计分析中，变量之间的交互作用对于建模和揭示变量间的复杂关系具有重要意义。传统上，在回归模型中，两个变量 x_1 和 x_2 之间的交互作用通过其乘积项 x_1x_2 来建模。然而，这种表示方式往往缺乏直观的可解释性，尤其对非专业人士而言，因此难以在实践中应用。Wu（2015）提出的条件主效应（CMEs）提供了一种新颖的替代方案，其关注的是一个变量在另一个变量处于固定水平时的效应，即 $x_1|x_2$。Su 和 Wu（2017）进一步拓展了这一思想，提出了一种系统策略，用于识别适合分析分部因子设计的一组条件主效应。条件主效应的概念提供了一个可解释的框架，有助于理解特定因素在特定条件下如何影响结果变量。例如，在基因关联研究中，识别与多个单核苷酸多态性相关的上位性（基因—基因交互作用）效应，对于这些寡基因/多基因疾病至关重要（Shibata 等，2024）。条件主效应 $A|B+$ 不再量化基因 A 与基因 B之间的交互作用 AB ，而是提供一个变量，用于量化仅当基因 B 存在时基因 A的重要性，从而增强生物学可解释性。

然而，当存在 p 个变量时，即使每个变量只有两个水平，CME 的总数也为 $2p$（$p-1$），即使 p 处于中等水平，这一数量也可能很大。因此，有效的变量选择至关重要。此外，CME 自然呈现出结构化的分组。根据 Mak 和 Wu（2019），一般有两种组结构：兄弟 组和表亲组。具体而言，主效应 A 的兄弟组由共享同一主效应的条件主效应组成，例如 $\{A|B+, A|B-, A|C+, A|C-, \dots\}$。相比之下， A 的表亲组由共享同一条件效应的条件主效应组成，包括 $\{B|A+, B|A-, C|A+, C|A-, \dots\}$。考虑到这类组结构，更适合采用双层变量选择，以识别活跃的条件主效应组以及这些组中的显著效应。
Mak 和 Wu（2019）的一项开创性研究提出了一种名为$cmenet$ 的方法，在条件主效应的线性模型背景下，采用“指数-MC+”层级惩罚进行双层变量选择。该方法能够

通过在惩罚回归中采用复合惩罚（Breheny 和 Huang，2009），分别控制组间 和组内变量选择，从而实现双层变量选择。

尽管$cmenet$ 方法在选择条件主效应方面表现良好，但仍存在若干局限。其一，耦合程度是固定的，无法根据组内信号强度进行调整：当某个系数与其他重要预测变量分组时，施加于该系数的惩罚会降低，即使引入耦合参数 τ，这一点依然如此。此外，调节参数 λ_S （针对兄弟组）和 λ_C（针对表亲组）在所有组中均为固定值，这限制了模型捕捉组特定变化的灵活性。我们还注意到，$cmenet$ 方法主要适用于具有连续响应的线性模型，因此不适用于响应为二元、计数型或遵循广义线性模型（GLM）结构的情形。

为解决这些问题，我们提出一种自适应 *cmenet* 方法，用于广义线性模型框架下的双层CME选择。通过对cmenet所用惩罚项的系统研究，我们提出一种自适应惩罚策略，以解决变量选择中的静态耦合局限，并将CME变量选择扩展到广义线性模型。本文的主要贡献如下：首先，受自适应 Lasso（Zou 2006）启发，所提方法同时纳入组自适应权重和个体自适应权重，以实现更准确的条件主效应变量选择。组自适应权重通过对同一组内可能发生耦合的效应赋予更高权重，缓解静态耦合问题。个体自适应权重可以调节收缩率，从而实现更准确的系数估计。我们给出了自适应权重的理论解释，以论证权重选择的合理性。此外，我们将CME变量选择扩展到广义线性模型，并开发了一种高效的迭代加权最小二乘（IRLS）算法（Green 1984）用于参数估计。所提迭代重加权最小二乘算法结合了坐标下降方法（Fu 1998；Friedman 等人 2007；Mazumder 等人 2011）和主化-最小化技术（Hunter and Lange 2000），使每次迭代中的优化具有闭式解。所提方法通过R 软件包 `glmcmenet`实现，并将公开提供，以供实际应用使用。

本文其余部分组织如下。我们在第 2.1 节简要描述条件主效应的概念，并回顾 cmenet方法。我们在第 3 节详细介绍面向广义线性模型的所提自适应 cmenet方法，并在第 4 节给出参数估计算法。第 5 节通过若干数值模拟展示自适应 cmenet的性能。第 6 节分析其在基因关联研究中的应用。第 7 节对本文进行总结，并讨论未来工作。

2 背景与动机

2.1 条件主效应概述

条件主效应（CME）的定义最早由 Wu（2015）提出。遵循 Mak and Wu（2019）的研究，我们采用基于主效应（ME）推导出的 CME 定义。假设模型中有 n 个观测值，考虑 p 个二水平因子，用二元协变量向量 $\mathbf{x}_j = (x_{1,j}, x_{2,j}, \ldots, x_{n,j})^T \in \{-1, 1\}^n$, $j = 1, \ldots, p$ 表示。简而言之，对于协变量 $x_{i,j}$ ，在对象 i 的 $+1$ 水平上以 $x_{i,k}$ 为条件时，CME 协变量定义为 $x_{i,j|k+} = x_{i,j}$ （当 $x_{i,k} = +1$ 时），并在 $x_{i,k} = -1$ 时定义为 $x_{i,j|k+} = 0$ 。$x_{i,j|k-}$ 的 CME 协变量定义方式类似（详情请参阅 Mak and Wu（2019））。为方便起见，我们采用记号 $A, B, C, D, \ldots$ 表示分别量化协变量 $\mathbf{x}_1, \mathbf{x}2, \ldots, \mathbf{x}_p$ 效应的主效应。同样，量化协变量 $\mathbf{x}_{j|k+}$ 和 $\mathbf{x}_{j|k-}$ 效应的条件主效应分别记为 $J|K+$ 和 $J|K-$。在条件主效应 $J|K+$ 和 $J|K-$ 中，效应 J 和 K 分别称为父效应和子效应。本文其余部分将始终采用这一记号。

需要注意的是，传统的交互项往往不易解释。考虑一个涉及两个协变量 A 和 B的简单例子。传统上，它们的交互作用用乘积项 AB表示。然而，仅凭这一项 AB 无法明确说明 A 和 B各自的取值或水平，因此在实际应用中很难厘清它们之间的交互作用。例如， $AB = 1$ 既可能由 $A = +1$, $B = +1$ 导致，也可能由 $A = -1$, $B = -1$导致。CME 可消除这种歧义

表 1：主效应 A 和 B 的模型矩阵，以及其交互效应和四个条件主效应。

A	B	AB	$A\|B+$	$A\|B-$	$B\|A+$	$B\|A-$
$+1$	$+1$	$+1$	$+1$	0	$+1$	0
$+1$	-1	-1	0	$+1$	-1	0
-1	$+1$	-1	-1	0	0	$+1$
-1	-1	$+1$	0	-1	0	-1

通过更明确地刻画条件关系。表 1 列出了由这些主效应导出的四种可能的条件主效应。

显然，主效应协变量与条件主效应协变量之间存在复杂的层级结构，它们可以是父项，也可以是子项。按照 Mak 和 Wu（2019）中的分组结构，可分为三大类：(1)同胞条件主效应，共享相同的父效应，例如 $\{A|B+, A|B-, A|C+, A|C-, \cdots\}$；(2) 堂表亲条件主效应，共享相同的子效应，例如 $\{B|A+, B|A-, C|A+, C|A-, \cdots\}$；以及 (3) 父效应-子效应 对，例如 $\{A, A|B+\}$, $\{A, A|B-\}$, ...。Mak 和 Wu（2019）考察了这些组内的共线性和两两相关性，并发现相关性大小形成了效应组之间的自然层级，导致使用标准变量选择方法时选择结果不一致。为此，他们提出了一种针对条件主效应的双层变量选择方法，并在 cmenet 框架中实现。这一思路建立在早期双层变量选择研究的基础上，例如 Huang 等人（2009）提出的组桥惩罚，该方法对预定义组的 ℓ_1-范数施加桥惩罚，以促进组级稀疏性。随后，Breheny 和 Huang（2009）提出了复合惩罚的一般框架，该框架采用外层 惩罚来约束一系列内层 惩罚。该框架通过控制组间和组内的变量选择，实现了双层变量选择。

令（$\mathbf{y}$，$\mathbf{X}$）表示观测值的完整集合，其中 $\mathbf{y} \in \mathbb{R}^{n\times 1}$ 是与完整模型矩阵 $\mathbf{X}$ =（$\mathbf{x}_1$, $\mathbf{x}_2$，.....，$\mathbf{x}_p$，.....，$\mathbf{x}_{p'}$） $\in \mathbb{R}^{n\times p'}$相对应的响应变量。此处，$\{\mathbf{x}_j\}_{j=1}^{p}$ 和 $\{\mathbf{x}_j\}_{j=p+1}^{p'}$ 分别是 p主效应和 $4\binom{p}{2}$ 条件主效应的标准化向量，

分别定义 $\mathcal{S}(j) = \{J, J|A+, J|A-, J|B+, J|B-, \cdots\}$为父效应 j的兄弟组，定义 $\mathcal{C}(j) = \{J, A|J+, A|J-, B|J+, B|J-, \cdots\}$为子效应 j、 $j = 1$、.....、 p的表亲组。随后进行参数估计，具体如下:

$$\begin{aligned} \min_{\boldsymbol{\beta}} Q(\boldsymbol{\beta}) &= \min_{\boldsymbol{\beta}} \left\{ \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + P_S(\boldsymbol{\beta}) + P_C(\boldsymbol{\beta}) \right\}, \\ P_{\mathcal{S}}(\boldsymbol{\beta}) = \sum_{j=1}^{p} f_{out,\mathcal{S}} &\left\{ \sum_{k \in \mathcal{S}(j)} f_{in,\mathcal{S}}(\beta_k) \right\}, P_{\mathcal{C}}(\boldsymbol{\beta}) = \sum_{j=1}^{p} f_{out,\mathcal{C}} \left\{ \sum_{k \in \mathcal{C}(j)} f_{in,\mathcal{C}}(\beta_k) \right\}, \end{aligned} \tag{1}$$

其中，$f_{out,\mathcal{S}}$和 $f_{in,\mathcal{S}}$ （$f_{out,\mathcal{C}}$和 $f_{in,\mathcal{C}}$）分别表示外层和内层惩罚，分别用于同胞组（或表亲组）的组间和组内选择。通过采用不同的外层和内层惩罚组合，Mak 和 Wu（2019）针对外层惩罚采用了指数惩罚 $\eta_{\lambda_s,\tau}$（θ）（或 $\eta_{\lambda_c,\tau}$（θ））（Breheny 2015），并针对内层惩罚采用了缩放的极小极大凹加（MC+）惩罚 $g_{\lambda_s,\gamma}$（θ）（或 $g_{\lambda_c,\gamma}$（θ））（Zhang 2010）。这些惩罚定义如下:

$$\begin{aligned} \eta_{\lambda,\tau}(\theta) &= \frac{\lambda^2}{\tau} \left\{ 1 - \exp(-\frac{\tau\theta}{\lambda}) \right\}, \\ \text{and} \quad g_{\lambda,\gamma}(\beta) &= \int_0^{|\beta|} \left(1 - \frac{x}{\lambda\gamma} \right)_+ dx, \end{aligned} \tag{2}$$

其中， λ、 γ和 τ 是调节参数，分别用于控制正则化程度、非凸性和耦合程度。具体而言，两个惩罚参数 $\lambda_s > 0$ 和$\lambda_c > 0$ 分别用于同胞组和表亲组，并且在所有组中保持不变。MC+惩罚包含参数 $\gamma > 1$，该参数用于在 ℓ_0-范数惩罚（近似为 $\gamma \to 1$）和 LASSO 中的 ℓ_1-范数惩罚（近似为 $\gamma \to \infty$）之间进行转换。指数惩罚中的耦合参数 τ 控制同一组内变量之间关系的强度。这些惩罚函数能够有效控制组间耦合和稀疏性，从而应对 CME 双层变量选择中的挑战。

2.2 现实世界动机

生物学家经常会观察到上位性，即基因（或遗传变异）对表型的影响取决于是否存在另一种变异，或某条通路是否处于活跃状态。

主效应筛查可能会遗漏此类模式，而对所有交互作用进行全面搜索则可能规模庞大且难以解读。条件主效应（CMEs）可以弥补这一不足。诸如“SNP J 在SNP $K = 1$条件下”的表述，能够直接体现生物学背景，并以科学家惯用的方式将相关效应归组。在该生物学背景下，待评估其效应的SNP（J）是焦点位点，而作为条件的SNP（K）是修饰位点。这种表示方式明确了某个位点在何种遗传背景下会影响表型，同时保持模型简洁且易于解读。为说明CMEs如何在实践中捕捉这类依赖背景的效应，我们考察了拟南芥的一项全基因组关联研究。

我们研究 Atwell 等人（2010年）的全基因组关联研究中的可见黄化表型（详情见第 6 节）。拟南芥是一种主要通过自花受精繁殖的模式植物，在许多性状上都存在显著的遗传变异。条件主效应有助于将单核苷酸多态性与特定生物学通路联系起来。例如，某个条件主效应可能表明，只有当铁相关基因中也存在某个变异时，叶绿素基因附近的单核苷酸多态性才会产生影响。另一个条件主效应可能表明，当抗氧化伙伴基因缺失时，某个氧化还原酶单核苷酸多态性的影响更强。这类“条件性”表述有助于与现有功能注释进行比较，也能引出可实验检验的假设。同时，$cmenet$ 方法在这种数据情境下也存在局限。首先，黄化表型是二元结局，而基于线性高斯模型构建的$cmenet$ 方法无法在适当的尺度上对二元响应进行建模。其次，遗传标记往往彼此相关。当$cmenet$ 发现一个强效的条件主效应时，往往也会选出一组与之相关且位置相近的条件主效应。这会导致结果冗余，输出中充斥着指向同一遗传来源的长串条件主效应，从而降低可解释性。

我们的框架旨在应对这些挑战。我们保留 CME 表示，并采用逻辑回归模型分析二元性状。随后，我们在组水平和个体水平上分别进行惩罚。其主要优势在于，两个 CME 组的相对影响会随证据而变化，而非预先固定。该框架能够得到一组可解释的情境特异性关联，从而揭示值得后续研究的候选基因—基因相互作用模式。具体内容将在

所提方法将在下文第3.1节中介绍。

3 所提方法：自适应cmene t

本节首先探讨 cmenet 方法中的耦合问题，并在第3.1节指出其局限性。随后，我们在第3.2节提出一种新的惩罚方法，以解决这些局限性。最后，我们在第3.3节将所提出的方法与 cmenet 方法进行比较，以考察其有效性。

3.1 cmenet中的耦合能力

需要注意的是，“指数-MC+”惩罚框架的选择源于两项关键原则：（1）*CME*耦合，即若同一同胞或表亲组中的其他效应已被选入，则某个CME更有可能被纳入模型；以及（2）*CME*约简，即当其同胞或表亲组中的效应已进入模型时，主效应更容易被选中。为阐明这些原则，我们采用Breheny（2015）提出的“选择阈值”概念，它指协变量与结果变量之间为使该协变量被选入模型所必需的最低关联水平。令 z 表示 β_j的无惩罚估计，即 $z = \frac{1}{n}\mathbf{x}_j^T(\mathbf{y} - \mathbf{X}_{-j}\beta_{-j})$。此处，$\mathbf{X}_{-j} \in \mathbb{R}^{n\times(p'-1)}$ 表示排除第 j个协变量后的设计矩阵，$\boldsymbol{\beta}_{-j} \in \mathbb{R}^{(p'-1)}$ 为相应的系数向量。估计 z 对应于在保持所有其他ˆ系数固定的情况下使损失函数最小化的值。“选择阈值”正式定义为 $t_{\beta_j} := \inf\{|z| : \beta_j \neq 0\}$对于 $j = 1, \ldots, p'$。换言之，它表示使相应系数 β_j 非零的所有 $|z|$ 取值的下确界。由（1）推导，$\beta_{j|k+}$ 的选择阈值 $t_{\beta_{j|k+}}$ 可表示为：

$$t_{\beta_{j|k+}} = \lambda_s \exp\{-\frac{\tau\|\boldsymbol{\beta}_{\mathcal{S}(j)}\|_{\lambda_s,\gamma}}{\lambda_s}\} + \lambda_c \exp\{-\frac{\tau\|\boldsymbol{\beta}_{\mathcal{C}(k)}\|_{\lambda_c,\gamma}}{\lambda_c}\}. \tag{3}$$

这里，组 $\mathcal{G}$ 的系数向量记为 $\boldsymbol{\beta}_{\mathcal{G}} \in \mathbb{R}^{|\mathcal{G}|}$，惩罚项定义为$\|\boldsymbol{\beta}_{\mathcal{G}}\|_{\lambda_{\mathcal{G}},\gamma} := \sum_{l\in\mathcal{G}} g_{\lambda_{\mathcal{G}},\gamma}(\beta_l)$。系数 β_j 的选择阈值 t_{β_j}

以类似方式定义。为简便起见，我们引入记号 $\Delta_{\mathcal{S}(j)} := \lambda_s \exp\{-\frac{\tau\|\boldsymbol{\beta}_{\mathcal{S}(j)}\|_{\lambda_s,\gamma}}{\lambda_s}\}$ 和 $\Delta_{\mathcal{C}(k)} := \lambda_c \exp\{-\frac{\tau\|\boldsymbol{\beta}_{\mathcal{C}(k)}\|_{\lambda_c,\gamma}}{\lambda_c}\}$，这些记号将在本文余下部分中使用。

让我们考虑一个已选择 $A|B+$的模型。目前正在考虑两个候选效应： $A|C+$ 和 A。图1(a)说明了 $\beta_{A|B+}$ 的变化将如何分别影响 $\beta_{A|C+}$ 和 β_A的选择阈值。当模型中最初未包含任何变量时，无论系数大小如何，选择阈值都保持较高水平。然而，随着同胞变量（$\mathbf{x}_{A|\cdot}$）或表亲变量（对于$A|C+$为$\mathbf{x}_{\cdot|C\cdot}$，或对于 A为$\mathbf{x}_{\cdot|A\cdot}$）被纳入模型， $\beta_{A|C+}$和 β_A 的选择阈值都会降低。当同胞效应和表亲效应同时纳入时，这种降低尤为明显，从而有利于将同一组中的其他效应纳入模型。这一现象构成了第一项原则：*CME*耦合。为与同胞效应的行为进行对比，图1(b)考虑了一个包含 $B|A+$ 并改变其系数 $\beta_{B|A+}$的模型。β_A 的选择阈值（以虚线表示）随着 $\beta_{B|A+}$而下降，而 $\beta_{A|C+}$ 的阈值（以实线表示）基本保持不变，因为 C 的表亲变量是固定的。 β_A 的阈值较低，表明主效应 A比CME $A|C+$更有可能被纳入模型。也就是说，当涉及 A的效应已经从 $\mathcal{S}$（A）或从 $\mathcal{C}$（A）中被选入后，主效应 A更容易进入模型。当许多此类效应被选中时，条件效应集合往往会收缩至其潜在的主效应。这一行为体现了第二项原则：*CME*约简。

一个重要的观测是，随着系数幅度增大，选择阈值会稳定在某个固定水平。也就是说，耦合程度本质上由组大小和耦合参数 τ 控制。这一现象源于式 (2) 中内部惩罚函数 $g_{\lambda,\gamma}(\beta)$ 的特性：该函数单调递增，并随着 $|\beta|$ 增大而逐渐减轻惩罚，直至临界点 $|\beta| = \gamma\lambda$。超过这一临界值后，$g_{\lambda,\gamma}(\beta)$ 会达到其最大值 $\gamma\lambda/2$ 并保持不变。因此，选择阈值会达到固定值 $\eta'_{\lambda,\tau}(\theta)|_{\theta=\gamma\lambda/2}$ 的平台，此后不再受系数幅度进一步增大的影响。该平台可防止过度耦

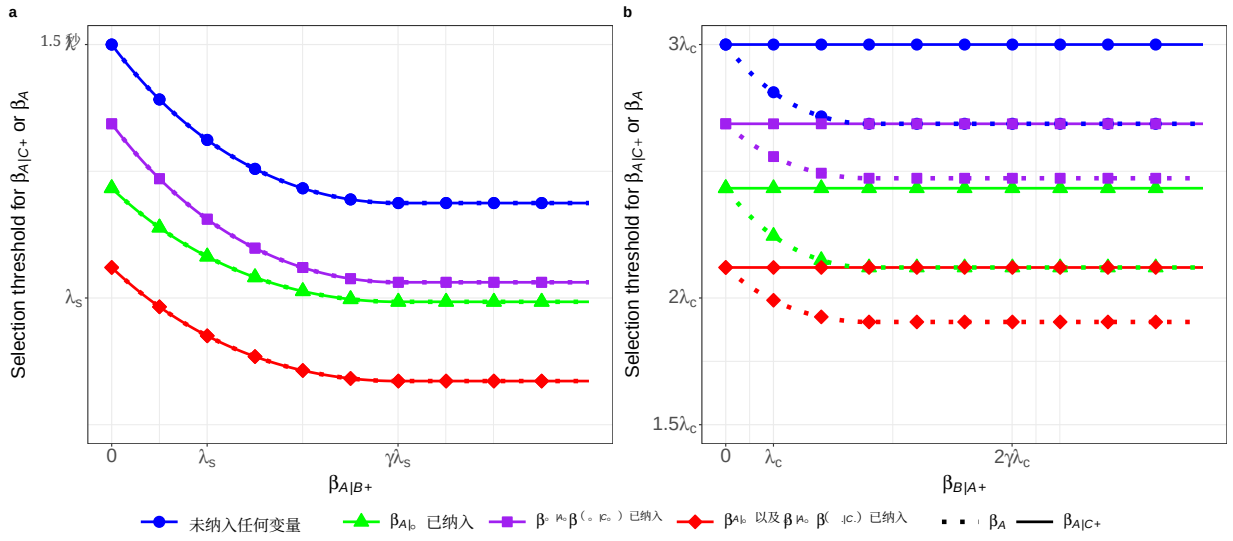


图1：CME系数变化对*cmenet*选择阈值的影响：(a) 变化$\beta_{A|B+}$；(b) 变化 $\beta_{B|A+}$。曲线显示了在四种模型情境下，$\beta_{A|C+}$（虚线）和 β_A（实线）的阈值：未纳入其他变量（蓝色）、纳入一个同胞（绿色）、纳入一个堂表亲（紫色），以及纳入一个同胞和一个堂表亲（红色）。基线 `cmenet` 设置：$(\lambda_s, \lambda_c, \gamma, \tau) = (1, 0.5, 3, 0.25)$。新增变量的系数与 $\beta = 2$相同。

合，因为 $\eta_{\lambda,\ \tau}(\theta)$ 中指数的每一项贡献都是有限的；一旦各项达到饱和，增大系数便无法使阈值无限下降。

尽管信号极强时出现这种平台效应是可取的，但该方法存在静态耦合局限，具体表现在两个方面。首先，组间相对影响由非自适应参数 λ_s 和 λ_c固定决定。这种平衡从一开始就被锁定，无法根据数据进行调整。例如，在图1中，设定 $\lambda_s = 2\lambda_c$ 会导致绿色曲线始终低于紫色曲线，无论信号强度有多大。这种固定设定锁定了组间的预设比例，使阈值无法根据证据实际强度的变化而调整。其次，也是更重要的一点，无论当前组的强度如何，在组内添加一个新变量所产生的边际效应始终不变。图1(a)清楚地显示了这一点：阈值曲线之间的间距（例如蓝线与绿线之间的间距）

（或紫线与红线之间的间距）会随着 $\beta_{A|B+}$ 增大而保持不变。这些曲线彼此平行。此外，由于内部惩罚具有一致有界性，添加第二个同胞所带来的益处与添加第一个完全相同。因此，即使一个组已经“饱和”，随着添加更多同胞，阈值仍会持续下降。缺乏边际收益递减可能导致对已有主导组过度强调，同时无法充分响应竞争组中的新兴信号。

这些考虑表明，让惩罚能够根据各组当前的信号水平进行自适应调整非常重要。首先，为了解决组间相对影响固定的问题，最好让主惩罚参数（λ_s和 λ_c）根据各自组内的总信号强度进行调整。这样，该过程便能动态地“重新平衡”同胞组与表亲组的影响。其次，为了解决组内边际效应恒定的问题，我们需要对每个系数的内部惩罚进行自适应调整。这意味着，施加于某个系数（例如 $\beta_{A|C+}$）的内部惩罚应随着相关系数（$\beta_{A|B+}$）的信号增强而减小。这样，当组内出现初始的微弱信号时，选择阈值会更早下降；而当组内已有较强信号时，也能减轻其边际影响。因此，我们提出了一个自适应惩罚框架，并将在下一节中详细介绍。

3.2 提出的自适应cmenet和 `glmcmenet`

如第3.1节所述，静态耦合局限主要源于内部惩罚的一致有界，以及应用于各组和个体的非自适应惩罚参数。受 Zou（2006）的自适应 Lasso 和 Fang 等人（2015）提出的自适应稀疏组 Lasso 启发，我们考虑为不同的组和个体效应引入自适应权重。因此，所提出的自适应 cmenet 惩罚定义如下：

$$\tilde{P}_{\mathcal{G}}(\boldsymbol{\beta}) = \sum_{j=1}^{p} f_{out,\mathcal{G}}^{(j)} \left\{ \sum_{k \in \mathcal{G}(j)} \omega_k f_{in,\mathcal{G}}^{(j)}(\beta_k^{(j)}) \right\}, \quad \mathcal{G} \in \{\mathcal{S}, \mathcal{C}\} \text{ and } k \in \mathcal{G}(j) \tag{4}$$

其中 $f_{out}^{(j)}$, $g(\theta) = \eta_{\lambda_{\mathcal{G}(j)},\ \tau}(\theta)$ 和 $f_{(j)in}$, $g(\beta) = g_{\lambda_{\mathcal{G}(j)},\ \gamma}(\beta)$。这里， $\lambda_{\mathcal{G}(j)}$是组 $\mathcal{G}(j)$ 的自适应权重，而 ω_k 是组 $\mathcal{G}(j)$ 内主效应或 CME 的个体权重。引入自适应权重的理由是：$\lambda_{\mathcal{G}(j)}$ 解决了各组之间惩罚参数非自适应的问题，而 ω_k 则放宽了内部惩罚的一致边界。为了说明 cmenet 惩罚与所提出的自适应 cmenet 惩罚之间的区别，读者可以参考示例 1 中所示的简化示例。因此，参数估计可表示为：

$$\min_{\boldsymbol{\beta}} Q(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} \Big\{ \frac{1}{2} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \tilde{P}_S(\boldsymbol{\beta}) + \tilde{P}_C(\boldsymbol{\beta}) \Big\}. \tag{5}$$

示例 1. 考虑一个仅包含两个主效应 A 和 B的简单模型，则相应的兄弟组为 $\mathcal{S}(A) = \{A, A|B+, A|B-\}$, $\mathcal{S}(B) = \{B, B|A-, B|A+\}$。对于式 (1) 中的 cmenet 惩罚，兄弟组的惩罚定义如下：

$$P_{\mathcal{S}}(\boldsymbol{\beta}) = \frac{\lambda^2}{\tau}\left\{1 - \exp\left[-\frac{\tau}{\lambda}\Big\{g_{\lambda,\gamma}(\beta_A) + g_{\lambda,\gamma}(\beta_{A|B+}) + g_{\lambda,\gamma}(\beta_{A|B-})\Big\}\right]\right\} + \frac{\lambda^2}{\tau}\left\{1 - \exp\left[-\frac{\tau}{\lambda}\Big\{g_{\lambda,\gamma}(\beta_B) + g_{\lambda,\gamma}(\beta_{B|A+}) + g_{\lambda,\gamma}(\beta_{B|A-})\Big\}\right]\right\}.$$

相比之下，对于式 (4) 中提出的自适应 cmenet 惩罚，引入了兄弟组权重 $\mathcal{S}(j)$ 和个体权重 ω_k：

$$\begin{aligned}\tilde{P}_{\mathcal{S}}(\boldsymbol{\beta}) =& \frac{\lambda_{\mathcal{S}(A)}^2}{\tau}\left\{1 - \exp\left[-\tau\Big\{\frac{\omega_A}{\lambda_{\mathcal{S}(A)}} \cdot g_{\lambda_{\mathcal{S}(A)},\gamma}(\beta_A) + \right.\right. \\ & \left.\left. \frac{\omega_{A|B+}}{\lambda_{\mathcal{S}(A)}} \cdot g_{\lambda_{\mathcal{S}(A)},\gamma}(\beta_{A|B+}) + \frac{\omega_{A|B-}}{\lambda_{\mathcal{S}(A)}} \cdot g_{\lambda_{\mathcal{S}(A)},\gamma}(\beta_{A|B-})\Big\}\right]\right\} + \\ & \frac{\lambda_{\mathcal{S}(B)}^2}{\tau}\left\{1 - \exp\left[-\tau\Big\{\frac{\omega_B}{\lambda_{\mathcal{S}(B)}} \cdot g_{\lambda_{\mathcal{S}(B)},\gamma}(\beta_B) + \right.\right. \\ & \left.\left. \frac{\omega_{B|A+}}{\lambda_{\mathcal{S}(B)}} \cdot g_{\lambda_{\mathcal{S}(B)},\gamma}(\beta_{B|A+}) + \frac{\omega_{B|A-}}{\lambda_{\mathcal{S}(B)}} \cdot g_{\lambda_{\mathcal{S}(B)},\gamma}(\beta_{B|A-})\Big\}\right]\right\}.\end{aligned}$$

表亲组惩罚也可以写出类似的公式。

对于每个组 $\mathcal{G}(j)$，我们将组权重设为 $\Omega_{\mathcal{G}(j)} = (\|\hat{\boldsymbol{\beta}}^{ridge}_{\mathcal{G}(j)}\|_1 + \frac{1}{n})^{-1}$，其中 $\|\hat{\boldsymbol{\beta}}^{ridge}_{\mathcal{G}(j)}\|_1$ 表示组 $\mathcal{G}(j)$ 内岭估计量的 l_1-范数。因此，

该组的自适应惩罚参数为 $\lambda_{\mathcal{G}(j)} = \lambda_{\mathcal{G}}\Omega^{\mathcal{G}(j)}$。此组内的个体权重定义为 $\omega_k = (|\hat{\beta}_k^{ridge}| + \frac{1}{n})^{-1}$, $k \in \mathcal{G}(j)$。使用岭估计量构造权重与 Zou（2006）的方法一致。鉴于 CME 组结构中存在隐式多重共线性，岭估计比 $\hat{\beta}^{MLE}$ 更可取，因为它具有更高的一致性和稳健性。为避免除以零，我们遵循 Fang 等人（2015）的做法，在分母中加上 $1/n$ 。自适应 cmenet 方法保留了 CME 耦合和 CME 约简的原则。通过引入自适应权重，该方法有效缓解了静态耦合问题，详见第 4 节。

请注意，cmenet 方法主要是为线性模型设计的。为拓展条件主效应的适用范围，我们将自适应 cmenet 框架扩展到广义线性模型（GLM），下文称为 `glmcmenet`。对于 GLM，损失函数通常是由指数族导出的负对数似然。广义线性模型通过连接函数将响应均值与预测变量关联起来，从而扩展线性模型以适应各种类型的响应变量。令 $\eta_i = \mathbf{x}_i^T\boldsymbol{\beta}$ 表示线性预测子，其中 $\mathbf{x}_i$ 是第 i 个观测的预测变量向量。响应均值 μ_i 与线性预测子 η_i 之间的关系由典范连接函数 $g(\mu_i) = \eta_i$指定。广义线性模型的损失函数（即负对数似然函数）为 $-\frac{1}{n}\sum_{i=1}^{n}[y_i \cdot \eta_i - b(\eta_i) + c(y_i)]$，其中 $b(\eta_i)$ 和 $c(y_i)$ 是由响应变量的指数族分布定义的函数。所提方法的优化问题可写为

$$\min_{\boldsymbol{\beta}} Q(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}}\Big\{-\frac{1}{n}\sum_{i=1}^{n}[y_i \cdot \eta_i - b(\eta_i) + c(y_i)] + \tilde{P}_S(\boldsymbol{\beta}) + \tilde{P}_C(\boldsymbol{\beta})\Big\}, \tag{6}$$

其中，$\tilde{P}_S(\boldsymbol{\beta})$ 和 $\tilde{P}_C(\boldsymbol{\beta})$ 定义见式 (1)。参数估计的详细内容将在第 4 节中介绍。

3.3 自适应 cmenet 的耦合能力

本节进一步探讨第3.1节所述的耦合能力，重点关注引入自适应权重后，$J|K+$ 的选择阈值 $\tilde{t}_{\beta_{j|k+}}$ ，该权重由

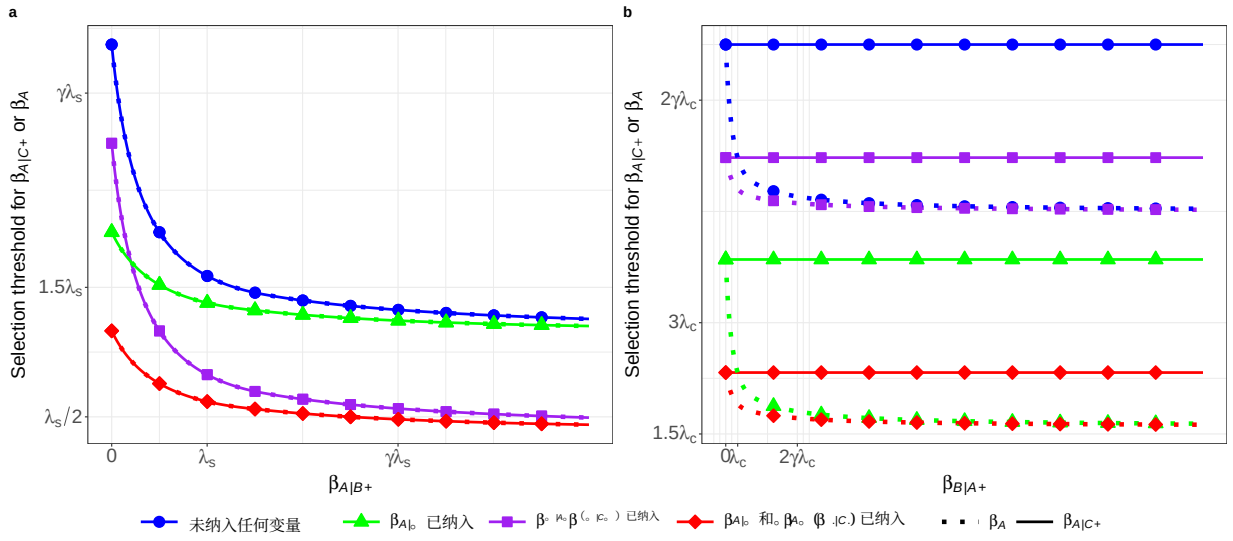


图2：CME系数变化对自适应 *cmenet*选择阈值的影响：(a) 变化 $\beta_{A|B+}$；(b) 变化 $\beta_{B|A+}$。曲线显示了在四种情境下 $\beta_{A|C+}$ (虚线) 和β_A (实线)的阈值：模型中未纳入其他变量时（蓝色）、纳入一个同胞时（绿色）、纳入一个堂表亲时（紫色），以及纳入一个同胞和一个堂表亲时（红色）。自适应 cmenet 的基线设置：$(\lambda_s, \lambda_c, \gamma, \tau) = (1, 0.5, 3, 0.25)$。添加的变量与 $\beta = 0.5$ 的系数相同。

式（5）得出：

$$\tilde{t}_{\beta_{j|k+}} = \lambda_{\mathcal{S}(j)} \exp\{-\frac{\tau\|\boldsymbol{\beta}_{\mathcal{S}(j)}\|_{\lambda_{\mathcal{S}(j)},\omega,\gamma}}{\lambda_{\mathcal{S}(j)}}\} + \lambda_{\mathcal{C}(k)} \exp\{-\frac{\tau\|\boldsymbol{\beta}_{\mathcal{C}(k)}\|_{\lambda_{\mathcal{C}(k)},\omega,\gamma}}{\lambda_{\mathcal{C}(k)}}\}. \tag{7}$$

与图1中的固定平台不同，图2显示，自适应权重会产生更低且依赖于信号的平台。这种模式源于自适应权重： $\mathcal{S}(A)$中的小系数会获得较大的权重，因此信号刚出现时，阈值会发生显著变化。随着 $\mathcal{S}(A)$中的系数增大，其权重会减小，惩罚项也会逐渐趋近其平台。这使得阈值在开始时迅速下降，随后更早趋于平稳。此外，图2更清楚地展示了 *CME*耦合原理：一旦选中某个同胞效应或表亲效应，阈值便会下移。图2(b)中 A的虚线也呈现出类似的下降趋势，而 $A|C+$ 的实线则保持平坦，不随 $\beta_{B|A+}$ 增大而变化，这与 *CME*约简 原理一致。所提方法解决了先前的两项局限：首先，它能动态地

“重新平衡”组间影响，这一点可从图2(a)中绿色线与紫色线的交叉得到证明。起初，当同胞信号较弱（接近 $\beta_{A|B+}=0$）时，紫色线较低。然而，随着同胞信号增强，绿色线下降得更陡，并与紫色线交叉，这证明相对组影响不再固定。其次，所提方法成功减弱了组内边际效应。在图2(a)中，随着 $\beta_{A|B+}$ 增大，绿色线在较高的平台处趋近蓝色线，而红色线在较低的平台处趋近紫色线。这表明，一旦兄弟组变强，额外兄弟的边际效应便微不足道，因为其贡献受到降权，且内部惩罚接近饱和。来自表亲侧的信号仍然重要，因为对表亲效应的调整并不取决于 $\mathcal{S}(A)$中的活动量。图2(b)中的虚线也呈现出类似的现象：随着表亲效应 $\beta_{B|A+}$ 增大，增加一个新表亲的边际效应会减弱，而增加一个新同胞仍然会产生影响。

4 参数估计

在广义线性模型的参数估计中，通常使用迭代加权最小二乘（IRLS）算法。IRLS 利用二阶泰勒展开，在当前估计值 $\boldsymbol{\beta}^{(m)}$ 附近对损失函数进行二次近似。随后，系数按照加权最小二乘的形式迭代更新为 $\frac{1}{2n}(\tilde{\mathbf{y}}-\mathbf{X}\boldsymbol{\beta})^T\mathbf{W}(\tilde{\mathbf{y}}-\mathbf{X}\boldsymbol{\beta})$，其中 $\tilde{\mathbf{y}}=\mathbf{X}\boldsymbol{\beta}^{(m)}+\mathbf{W}^{-1}(\mathbf{y}-\mathbf{v})$ 表示工作响应，$\mathbf{v}$ 和 $\mathbf{W}$ 分别是 $L(\boldsymbol{\eta})$ 关于 $\boldsymbol{\eta}$ 的一阶和二阶导数，在 $\boldsymbol{\eta}^{(m)}=\mathbf{X}\boldsymbol{\beta}^{(m)}$ 处取值。值得注意的是，$\mathbf{W}$ 是一个对角矩阵，其中每个对角元素对应最小二乘损失下第 i 个观测的权重。在每次迭代中，采用自适应 cmenet 惩罚的目标函数可表示为：

$$Q(\boldsymbol{\beta})=\frac{1}{2n}(\tilde{\mathbf{y}}-\mathbf{X}\boldsymbol{\beta})^T\mathbf{W}(\tilde{\mathbf{y}}-\mathbf{X}\boldsymbol{\beta})+\tilde{P}_S(\boldsymbol{\beta})+\tilde{P}_C(\boldsymbol{\beta}). \tag{8}$$

在本文余下部分中，$Q(\cdot)$ 将专指公式 (8) 中的形式。

为最小化式（8）中的目标函数，我们采用局部坐标下降（LCD）算法（Fu 1998；Friedman 等人2007；Mazumder 等人 2011），以高效地

推导出逐坐标目标函数 $Q_j(\cdot)$ 关于系数 β_j 的闭式阈值算子，同时保持其余系数不变:

$$\begin{aligned} Q_j(\boldsymbol{\beta}) = \frac{1}{2n}(\tilde{\mathbf{y}} - \mathbf{X}_{-j}\boldsymbol{\beta}_{-j} - \mathbf{x}_j^T\boldsymbol{\beta}_j)^T\mathbf{W}(\tilde{\mathbf{y}} - \mathbf{X}_{-j}\boldsymbol{\beta}_{-j} - \mathbf{x}_j^T\boldsymbol{\beta}_j) \\ + \eta_{\lambda_{\mathcal{S}(j)},\gamma}\{\|\boldsymbol{\beta}_{\mathcal{S}(j)}\|_{\lambda_{\mathcal{S}(j)},\omega,\gamma}\} + \eta_{\lambda_{\mathcal{C}(j)},\gamma}\{\|\boldsymbol{\beta}_{\mathcal{C}(j)}\|_{\lambda_{\mathcal{C}(j)},\omega,\gamma}\}. \end{aligned} \tag{9}$$

关于系数 $\beta_{j|k+}$ 的逐坐标目标函数 $Q_{j|k+}(\cdot)$定义如下:

$$\begin{aligned} Q_{j|k+}(\boldsymbol{\beta}) = \frac{1}{2n}(\tilde{\mathbf{y}} - \mathbf{X}_{-j|k+}\boldsymbol{\beta}_{-j|k+} - \mathbf{x}_{j|k+}^T\boldsymbol{\beta}_{j|k+})^T\mathbf{W}(\tilde{\mathbf{y}} - \mathbf{X}_{-j|k+}\boldsymbol{\beta}_{-j|k+} - \mathbf{x}_{j|k+}^T\boldsymbol{\beta}_{j|k+}) \\ + \eta_{\lambda_{\mathcal{S}(j)},\gamma}\{\|\boldsymbol{\beta}_{\mathcal{S}(j)}\|_{\lambda_{\mathcal{S}(j)},\omega,\gamma}\} + \eta_{\lambda_{\mathcal{C}(k)},\gamma}\{\|\boldsymbol{\beta}_{\mathcal{C}(k)}\|_{\lambda_{\mathcal{C}(k)},\omega,\gamma}\}. \end{aligned} \tag{10}$$

LCD算法每次固定其余坐标，并迭代优化一个坐标，因此具有较高的计算效率和稳定性，尤其适用于高维问题。然而，由于存在惩罚项，目标函数没有闭式一维解。为解决这一局限，我们采用 majorization-minimization（MM）方法（Hunter and Lange 2000）构造 $Q_j(\cdot)$ 或 $Q_{j|k+}(\cdot)$ 的替代函数，从而推导出闭式阈值算子。

4.1 凸性与稳定性

对于参数估计，考察优化中目标函数的凸性很有帮助。为简洁清晰起见，本文余下部分均假设矩阵 $\mathbf{x}j$ 的每一列均已归一化，使得 $\mathbf{x}_j^T\mathbf{I}_n = 0$ 且 $\frac{1}{n}\|\mathbf{x}_j\|_2^2 = 1$ 对所有 $j = 1, \ldots, p'$ 均成立。除非另有说明，本文将式（9）和（10）中的所有逐坐标目标函数统称为 $Q_j(\cdot)$, 其中 $j = 1, \ldots, p'$，不区分主效应和条件主效应。

命题 1. 定义 $\omega_{\max} = \text{最大}_{1\le j\le p'}\,\omega_j$ 为最大个体权重。当(8) 中的目标函数 $Q(\boldsymbol{\beta})$ 满足 $\tau + \frac{1}{\gamma\omega_{\max}} < \frac{\zeta_{\min}(\mathbf{X^TWX})}{2n\omega_{\max}^2}$ 时，该函数严格凸，其中$\mathbf{W}$ 在 $\boldsymbol{\beta}$ 处求值，而 $\zeta_{\min}$ 为最小特征值。

命题 1 表明，$\tau + \frac{1}{\gamma\omega_{\max}}$ 较小可确保目标函数具有凸性，从而得到全局最小值。然而，这一条件相当严格，通常仅适用于满足 $n \le p'$ 的低维情形。在高维

情形下，坐标下降算法仍然有效，因为它只要求每个坐标方向都具有局部凸性。对于 $\mathbf{W}$ $=\mathbf{I}$ 的线性回归，当 $\tau + \frac{1}{\gamma\omega_{\max}} \leq \frac{1}{2\omega_{\max}^2}$ 时，逐坐标目标函数 $Q_j(\beta_j)$ 的局部凸性得以保证。

对于广义线性模型，由于引入了非单位权重矩阵$\mathbf{W}$，其中 $\frac{1}{n}\mathbf{X^T W X} \neq \mathbf{I}$，情况变得更加复杂。尽管如此，仍可通过对$\mathbf{W}$施加上界来确保局部凸性；该矩阵对应于损失函数关于线性预测子 $\boldsymbol{\eta}$ 的海森矩阵。令$\bar{w} = \max_i \sup_\eta \{\nabla^2 L_i(\eta)\}$，其中 $L_i(\eta)$ 是第 i 个观测的损失函数。对于逻辑回归，这个上界很容易推导为 $\bar{w} = 1/4$，因为 $w_i = \pi_i(1-\pi_i)$。以下推论为如何在逻辑回归的高维设定下保持局部凸性并确保坐标下降算法有效提供了一些见解。

推论1.1. 对于逻辑回归的逐坐标目标函数 $Q_j(\beta_j)$ ，$Q_j(\beta_j)$ 在任意 $j = 1, \ldots, p'$ 下均为严格凸，只要 $\tau + \frac{1}{\gamma\omega_{\max}} \leq \frac{1}{8\omega_{\max}^2}$。

备注 1。 需要注意的是，推论1.1未必能推广到其他广义线性模型，因为由不同指数族分布导出的损失函数的海森矩阵可能没有有限上界。然而，这一限制并非不可克服，因为仍可找到合适的 τ 和 γ 值，以确保目标函数的凸性。或者，可以通过在每次迭代开始时初始化 $\bar{w} \leftarrow \max_i \{\nabla^2 L_i(\eta_i^*)\}$ 来引入伪上界。

4.2 阈值算子

为推导闭式阈值算子，我们首先根据引理1，通过对自适应惩罚进行一阶泰勒展开，定义$Q_j(\beta_j)$和 $Q_{j|k+}(\beta_{j|k+})$的主化函数。

引理1。 采用 (4) 中定义的自适应 *cmenet* 惩罚，令 $\tilde{\eta}_{\lambda_\mathcal{G},\tau}(\boldsymbol{\beta}_\mathcal{G}|\boldsymbol{\beta}_\mathcal{G}^*)$ 为 $\eta_{\lambda_\mathcal{G},\tau}(\boldsymbol{\beta}_\mathcal{G})$ 在固定 $\boldsymbol{\beta}_\mathcal{G}^* \in \mathbb{R}^{|\mathcal{G}|}$ 处的一阶泰勒展开，其中 $\mathcal{G}$ 表示一个预先指定效应的同胞

或表亲组。于是，

$$\tilde{\eta}_{\lambda_{\mathcal{G}},\tau}(\boldsymbol{\beta}_{\mathcal{G}}|\boldsymbol{\beta}^*_{\mathcal{G}}) = \eta_{\lambda_{\mathcal{G}},\gamma}\{\|\boldsymbol{\beta}^*_{\mathcal{G}}\|_{\lambda_{\mathcal{G}},\omega,\gamma}\} + \Delta^*_{\mathcal{G}} \times \{\|\boldsymbol{\beta}_{\mathcal{G}}\|_{\lambda_{\mathcal{G}},\omega,\gamma} - \|\boldsymbol{\beta}^*_{\mathcal{G}}\|_{\lambda_{\mathcal{G}},\omega,\gamma}\} \tag{11}$$

是 $\eta_{\lambda_{\mathcal{G}},\,\tau}(\boldsymbol{\beta}_{\mathcal{G}})$ 的主化函数，其中 $\|\boldsymbol{\beta}_{\mathcal{G}}\|_{\lambda_{\mathcal{G}},\,\omega,\,\gamma} := \sum_{l\in\mathcal{G}} \omega_l g_{\lambda_{\mathcal{G}},\,\gamma}(\beta_l)$。此外，我们可以推导出，在固定 $\boldsymbol{\beta}^* \in \mathbb{R}^{p'}$ 时，$\tilde{Q}_{j|k+}(\beta_{j|k+}|\boldsymbol{\beta}^*)$ 是 $Q_{j|k+}(\beta_{j|k+})$ 的主化函数，

$$\begin{aligned}\tilde{Q}_{j|k+}(\beta_{j|k+}|\boldsymbol{\beta}^*) = {} & L(\beta_{j|k+}|\boldsymbol{\beta}^*) + \eta_{\lambda_{\mathcal{S}(j)},\tau}(\|\boldsymbol{\beta}^*_{\mathcal{S}(j)}\|_{\lambda_{\mathcal{S}(j)},\omega,\gamma}) + \eta_{\lambda_{\mathcal{C}(k)},\tau}(\|\boldsymbol{\beta}^*_{\mathcal{C}(k)}\|_{\lambda_{\mathcal{C}(k)},\omega,\gamma}) \\ & + \Delta^*_{\mathcal{S}(j)}\omega_{j|k+}\{g_{\lambda_{\mathcal{S}(j)},\gamma}(\beta_{j|k+}) - g_{\lambda_{\mathcal{S}(j)},\gamma}(\beta^*_{j|k+})\} \\ & + \Delta^*_{\mathcal{C}(k)}\omega_{j|k+}\{g_{\lambda_{\mathcal{C}(k)},\gamma}(\beta_{j|k+}) - g_{\lambda_{\mathcal{C}(k)},\gamma}(\beta^*_{j|k+})\},\end{aligned} \tag{12}$$

其中，$*$ 表示该量是使用 $\boldsymbol{\beta}^*$ 的当前（即最近一次更新的）˜值计算得到的。$Q_j(\beta_j|\boldsymbol{\beta}^*)$ 是 $Q_j(\beta_j)$ 的主化函数，其定义方式类似。

对于给定的坐标，最小化目标函数 $Q_j(\beta_j)$ 等价于将偏残差 $\mathbf{y} - \mathbf{X}_{-j}\boldsymbol{\beta}^*_{-j}$ 对 $\mathbf{x}_j$ 进行回归。对 $L(\boldsymbol{\beta})$ 关于 β_j 求梯度，同时固定其他系数，可得 $\nabla L(\beta_j|\boldsymbol{\beta}^*) = -z_j + v_j\beta_j$，其中 $z_j = \frac{1}{n}\mathbf{x}_j^T\mathbf{W}(\tilde{\mathbf{y}} - \mathbf{X}_{-j}\boldsymbol{\beta}^*_{-j})$ 表示 β_j 的未惩罚解，而 $v_j = \frac{1}{n}\mathbf{x}_j^T\mathbf{W}\mathbf{x}_j$。因此，在采用自适应 cmenet 惩罚的广义线性回归中，β_j（或 $\beta_{j|k+}$）的逐坐标更新依赖于关于 z_j（或 $z_{j|k+}$）的闭式阈值函数，表达式如下：

$$\begin{aligned}\beta_j &\leftarrow S_{\lambda_{\mathcal{S}(j)},\lambda_{\mathcal{C}(j)}}(z_j; \Delta_{\mathcal{S}(j)}, \Delta_{\mathcal{C}(j)}), \\ \text{or} \quad \beta_{j|k+} &\leftarrow S_{\lambda_{\mathcal{S}(j)},\lambda_{\mathcal{C}(k)}}(z_{j|k+}; \Delta_{\mathcal{S}(j)}, \Delta_{\mathcal{C}(k)}).\end{aligned}$$

其中，$S(\cdot; \Delta_1, \Delta_2)$ 定义见下文的定理 1。

定理 1. 假设 $\tau + \frac{1}{\gamma\omega_{\max}} \leq \frac{1}{8\omega^2_{\max}}$，那么用于更新 a 的阈值函数

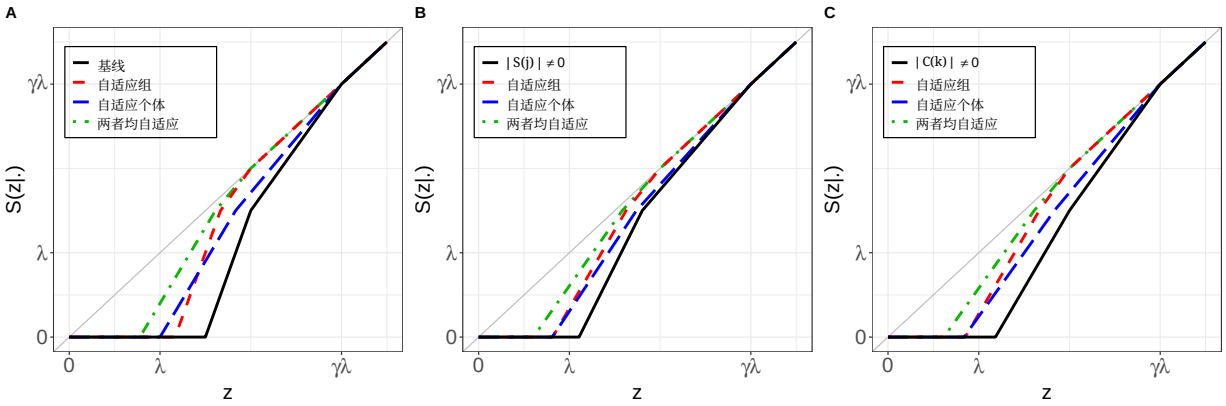


图3：在三种情形下比较 cmenet 与自适应 cmenet 的阈值函数 $S(\cdot;\Delta_1,\Delta_2)$：面板 (A) 未纳入变量，面板 (B) 纳入一个系数为 $\beta=2$ 的同胞，面板 (C) 纳入一个系数为 $\beta=2$ 的堂表亲。每幅图展示四种阈值函数：(实线) 无权重惩罚，即设置为 $(\lambda_s,\lambda_c,\gamma,\tau)$ =(1, 0.5, 3, 0.25) 的基线 cmenet；(虚线) 采用自适应权重的组加权惩罚 $(\Omega^{\mathcal{S}(j)},\Omega^{\mathcal{C}(j)},\omega_j)$ =(2/3, 1, 1)；(长虚线) 采用自适应权重的个体加权惩罚 $(\Omega^{\mathcal{S}(j)},\Omega^{\mathcal{C}(j)},\omega_j)$ =(1, 1, 1.5)；以及 (点划线) 采用自适应权重的组与个体组合加权惩罚 $(\Omega^{\mathcal{S}(j)},\Omega^{\mathcal{C}(j)},\omega_j)$ =(2/3, 1, 1.5)。

(8)的回归系数定义如下：

$$S_{\lambda_1,\lambda_2}(z_j;\Delta_1,\Delta_2)=$$

$$\begin{cases}\dfrac{z_j}{v_j}, & \text{if} \quad |z_j|\in[v_j\gamma\lambda_{(1),\infty}),\\ \dfrac{\mathrm{sgn}(z_j)(|z_j|-\Delta_{(1)}\omega_j)}{v_j-\frac{\Delta_{(1)}\omega_j}{\lambda_{(1)}\gamma}}, & \text{if} \quad |z_j|\in[v_j\lambda_{(2)}\gamma+\Delta_{(1)}\omega_j(1-\dfrac{\lambda_{(2)}}{\lambda_{(1)}}),v_j\lambda_{(1)}\gamma),\\ \dfrac{\mathrm{sgn}(z_j)(|z_j|-\Delta_{(1)}\omega_j-\Delta_{(2)}\omega_j)}{v_j-\frac{\Delta_{(1)}\omega_j}{\lambda_{(1)}\gamma}-\frac{\Delta_{(2)}\omega_j}{\lambda_{(2)}\gamma}}, & \text{if} \quad |z_j|\in[\Delta_{(1)}\omega_j+\Delta_{(2)}\omega_j,v_j\lambda_{(2)}\gamma+\Delta_{(1)}\omega_j(1-\dfrac{\lambda_{(2)}}{\lambda_{(1)}})),\\ 0, & \text{if} \quad |z_j|\in[0,\Delta_{(1)}\omega_j+\Delta_{(2)}\omega_j).\end{cases} \tag{13}$$

其中 $\lambda_{(1)}=\max(\lambda_1,\lambda_2)$ 和 $\lambda_{(2)}=\min(\lambda_1,\lambda_2)$，$\Delta_{(1)}$ 和 $\Delta_{(2)}$ 为其对应的斜率。

理解组权重与个体权重将如何影响阈值

函数 $S(\cdot;\Delta_1,\ \Delta_2)$ 至关重要。如图3所示，阈值函数展示了惩罚权重的变化如何影响曲线的转折点和斜率。在惩罚项中引入组权重（由三个图中贯穿的虚线表示），会使阈值函数的转折点向左移动。这一变化表明，模型在较低的选择阈值下更有可能纳入效应。此外，加入个体权重（由长虚线表示）会改变阈值函数的斜率。这种调整降低了系数的收缩率，使惩罚过程更加平缓。同时应用组权重与个体权重时（如点划线所示），得到的阈值函数所受惩罚最弱。这种配置更接近对角线，体现出一种更具一般性、限制更少的惩罚方法。这些结果凸显了自适应 cmenet 惩罚结构的灵活性。具体而言，组权重主要调整效应纳入的选择阈值，而个体权重则控制收缩率。通过结合这两种加权方案，所提出的框架为变量选择提供了一种更灵活且自适应的方法。

4.3 glmcmenet算法与参数调优

glmcmenet 使用阈值函数的详细算法可在补充材料中找到。参数调优时，我们使用 K 折交叉验证来寻找一组最优参数（$\lambda_s,\ \lambda_c,\ \gamma,\ \tau$）。为了高效探索调优参数空间，我们采用与 Mak 和 Wu（2019）相同的策略。具体而言，我们首先对结构参数（$\gamma,\ \tau$）扫描一个对数间隔网格，并通过交叉验证（CV）选出最佳参数对；在此选择的基础上，对惩罚强度（$\gamma,\ \tau$）参数对搜索一个对数间隔网格，然后对惩罚强度（$\lambda_s,\ \lambda_c$）搜索一个对数间隔网格。惩罚项 λ_s 和λ_c 是全局的，而自适应性则通过同级组和表亲组权重 $\Omega^{\mathcal{S}(j)}$、$\Omega^{\mathcal{C}(j)}$，以及效应 J的个体权重 w_j 来引入。为避免平凡解并缩小搜索空间，我们采用加权起始规则：令标准化后的 $\mathbf{X}$ 和中心化后的 $\mathbf{y}$，定义

$$\lambda_{\max}^{\omega}(\rho)=\max_j \frac{|\mathbf{x}_j^T\mathbf{y}|/n}{\rho\Omega_{\mathcal{S}(j)wj}+(1-\rho)\Omega_{\mathcal{C}(j)w_j}},$$

其中 $\rho = \lambda_s/(\lambda_s + \lambda_c) \in (0, 1)$。那么，只要满足 $\lambda_s + \lambda_c \geq \lambda_{\max}^{\omega}$，全零解就满足 KKT 条件（即为一个平稳解）。因此，对于每个固定的 ρ，我们将网格限制为 $\lambda_s + \lambda_c \geq \lambda_{\max}^{\omega}(\rho)$，并令（$\lambda_s$，$\lambda_c$）$=$（$\rho$（$\lambda_s + \lambda_c$），（$1-\rho$）（$\lambda_s + \lambda_c$））。

对于具有典范链接 $g(\mu) = \eta$ 的广义线性模型，我们采用类似的加权起始规则 $\lambda_{\max}^{\omega} = \max_j \frac{|\mathbf{x}_j^T \boldsymbol{u}|/n}{\rho\, \mathcal{S}(j)w_j + 1-\rho}\, \Omega(\,)\Omega^{\mathcal{C}(j)w_j}$，其中 $\boldsymbol{u} = \mathbf{y} - \boldsymbol{\mu}\, \mathrm{Var}(\mathbf{Y}|\boldsymbol{\eta}) \cdot \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}$。此外，如果 $Q(\boldsymbol{\beta})$ 严格凸，则根据命题 1，它是唯一的极小值。因此，我们还施加稳定性条件，以确保逐坐标子问题严格凸。为提高交叉验证期间的计算效率，我们采用与 cmenet 相同的技术，包括热启动和 Mak and Wu（2019）提出的活动集优化，因此同样具有稳定性保证和实际表现。

所提出算法的一项显著优势是，得益于 MM 框架并结合 LCD 的循环过程，它具有保证的下降性质。这些技术确保目标函数在每次迭代中单调递减。此外，利用局部凸性，我们可以证明算法收敛至一个驻点，从而提升优化过程的稳定性。

命题 2. 令 $\{\boldsymbol{\beta}^{(m)}\}$ 表示所提出的带自适应 *cmenet* 惩罚的逻辑回归算法在每次迭代中生成的系数估计序列。对于所有迭代 $m = 0$，1、2、……，下列不等式成立：

$$Q(\boldsymbol{\beta}^{(m+1)}) \leq Q(\boldsymbol{\beta}^{(m)}).$$

此外，当 $\tau + \frac{1}{\gamma\omega_{\max}} \leq \frac{1}{8\omega_{\max}^2}$ 时，该算法保证收敛到 $Q(\boldsymbol{\beta})$ 的一个驻点。

5 模拟研究

本节考虑两种情景，以考察所提方法的不同方面。在第一种情景（第 5.1 节）中，我们关注连续响应，并采用与 cmenet 论文（Mak and Wu 2019）完全相同的模拟设置。在这里，活跃效应完全由纯兄弟项、纯表亲项或纯主效应构成。本情景旨在评估加入自适应权重的有效性

与原始 cmenet方法相比的效果。本情景作为一项受控实验，用于展示自适应权重带来的改进。在第二种情景（第 5.2 节）中，我们考虑混合效应，其中活跃效应包含主效应及其相关兄弟或表亲效应的组合，这更贴近实际应用。本情景评估所提方法在连续响应和二元响应下的表现，从而更全面地考察其在不同数据类型下的有效性。为确保结果的稳健性，两种情景中的每个模拟案例均重复 100 次。

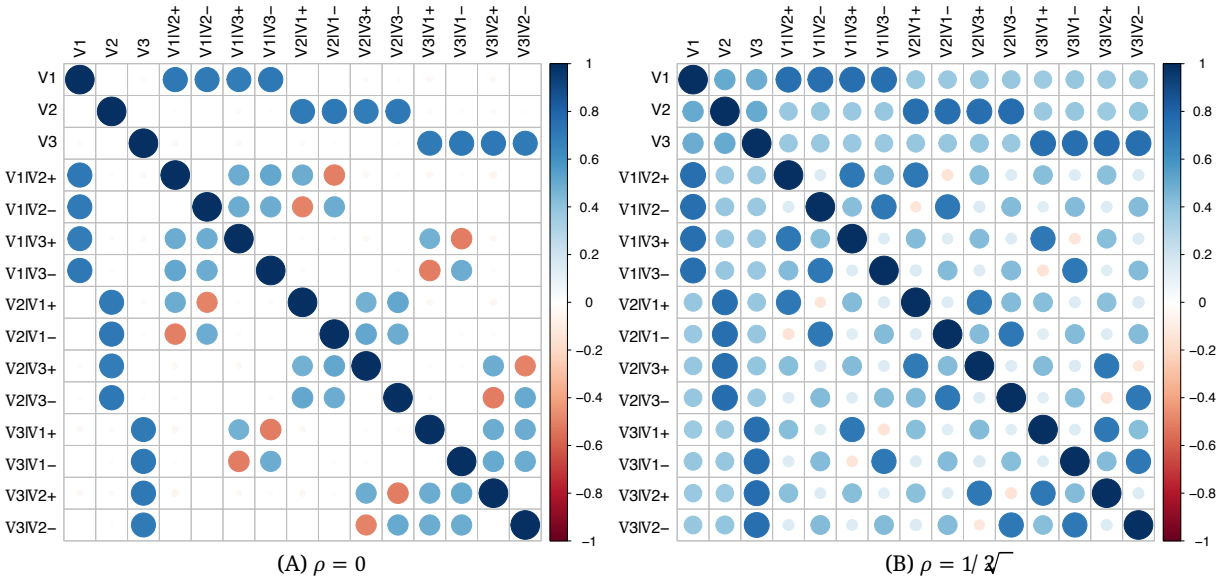
模型矩阵$\mathbf{X}$采用 Mak 和 Wu（2019）所述的等相关潜变量模型生成。该方法基于潜在矩阵$\mathbf{Z} = (z_{i,j})_{i=1\,j=1}^{n\quad p} \in \mathbb{R}^{n\times p}$构造主效应 $\{\mathbf{x}_j\}_{j=1}^{p} \in \{-1,+1\}^n$，其中每一行都服从多元正态分布$\mathbf{z}_i \sim \mathcal{N}\{\mathbf{0}, \rho\mathbf{J}_p + (1-\rho)\mathbf{I}_p\}$ ($\mathbf{J}_p$是全 1 矩阵，$\mathbf{I}_p$是单位矩阵，$\rho \in [0,1]$ 是相关参数）。随后根据这一潜在结构导出二元主效应：

$$x_{i,j} = -1\{z_{i,j} \le 0\} + 1\{z_{ij} > 0\}, \quad i = 1, \ldots .n, j = 1, \ldots, p. \tag{14}$$

使用等相关潜变量模型，可以根据第 2.1 节给出的定义轻松构造 $4\binom{p}{2}$ CME。因此，每次模拟的完整模型矩阵表示为$\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n} \in \mathbb{R}^{n\times p'}$，其中每个行向量定义为$\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p}, x_{i,p+1}, \ldots, x_{i,p'})' \in \mathbb{R}^{p'\times 1}$。当 $p = 3$ 时，该设计矩阵的相关结构如图4所示（更多细节请参见 Mak and Wu（2019）第 2.2 节）。

我们考虑两种不同类型的响应变量：连续和二元。这里我们以逻辑回归为代表性案例，但该方法同样适用于其他广义线性模型。本研究中的连续响应变量 y_i 变量由 $y_i = \beta_0 + \sum_{j=1}^{p'} \beta_j x_{i,j} + \epsilon_i$ 生成，其中 $\epsilon_i \sim N(0,1)$。二元响应基于逻辑回归模型生成。线性预测子 η_i 计算为 $\eta_i = \beta_0 + \sum_{j=1}^{p'} \beta_j x_{i,j}$。然后，二元响应变量 y_i 由伯努利分布模拟生成$y_i \sim$ 伯努利 $\left(\frac{1}{1+\exp(-\eta_i)}\right)$。

将 `cmenet` (R 软件包 `cmenet` V0.1.2；Mak 2022) 与所提出的方



(A) $\rho = 0$ (B) $\rho = 1/\sqrt{2}$

图4：由等相关潜在模型生成的模拟的完整模型矩阵（$n = 100$， $p = 3$）的相关矩阵，其中（A） $\rho = 0$ 和（B） $\rho = 1/\sqrt{2}$。

法进行比较，以评估自适应权重的有效性。我们还考虑了若干变量选择基准方法，包括 Lasso（R 软件包 `glmnet` V4.1.8；Fried- man et al. 2010；Tay et al. 2023）、自适应 Lasso （ R 软件包 `glmnet` V4.1.8 ； Zou 2006 ）、非凹惩罚方法 MCP （ R 软件包 `ncvreg`V3.14.3；Breheny and Huang 2011；Zhang 2010）、双层选择方法 GEL （R 软件包 `grpreg` V3.5.0；Breheny 2015）， 以及广受欢迎的双向交互选择方法 `hierNet`(R 软件包 `hierNet`V1.9；Bien and Tibshirani 2020)。我们使用三个指标衡量变量选择的准确性： F_1 分数、精确率和真正率（TPR）， 或召回率。分别用 TP、FP、TN 和 FN 表示真正例、假正例、真负例和假负例的数量。精确率可计算为 Precision= TP/(TP+FP)，召回率可计算为 Re- call=TP/(TP+FN)，而 F_1 分数可计算为 $F_1 =$ (2Precision $\times$ Recall)/(Precision +Recall)。除了评估变量选择性能，我们还衡量预测性能。预测评估指标取决于响应的类型。对于连续响应，我们使用均方预测误差（ MSPE）评估性能。对于二元响应，我们根据误分类率（MCR）评估性能。此处，

均方预测误差 $= \frac{1}{n}\sum_{i=1}^{n}(y_{i,\ new} - \hat{}y_{i,\ new})^2$, 和 MCR $= \frac{1}{n}\sum_{i=1}^{n} I(y_{i,\ new} \neq \hat{}y_{i,\ new})$。由于 `hierNet`无法编码交互作用中的条件关系，因此逐元素选择指标无法直接比较。因此，对于 `hierNet` ，我们报告模型规模（选定效应的数量）和样本外预测误差，作为简约性和预测性能的指标。根据简约原则（Stoica 和 S¨oderstr¨om 1982；Wu 和 Hamada 2011），当两个模型的预测性能相当时，参数较少的模型优于参数较多的模型。

5.1 自适应权重评估

本节在与 Mak and Wu（2019）用于 cmenet 的相同模拟设置下，评估自适应 cmenet 的性能，并仅关注连续响应。遵循他们的框架，我们考虑真实模型中三种不同类型的活跃效应：（1）所有活跃效应均为主效应，例如 $\{A, B, C, D\}$；（2）所有活跃效应均为同胞效应，例如 $\{A|B+, A|D+, B|C-, B|E+\}$；（3）所有活跃效应均为表亲效应，例如 $\{C|A+, B|A-, D|B+, E|B+\}$。对于涉及同胞效应和表亲效应的情形，我们确保活跃效应中不包含其他类型的组结构。本研究中，样本量为 $n = 100$，主效应数量为 $p = 40$，因此效应总数为 $p' = p + 4\binom{p}{2} = 3160$ （包括主效应和条件主效应）。活跃效应的系数设为 $\beta_j = 5$，截距为 $\beta_0 = 12$。为评估主效应之间相关性的影响，本研究考虑两种相关性设置：不相关（$\rho = 0$）和中度相关（$\rho = 1/\sqrt{2}$）。活跃组的数量有所变化（$\{4, 6, 8, 10, 12\}$），每组包含两个活跃效应。

–

图 5 展示了情景 1 的模拟结果。在非相关情形的几乎所有模拟设置中，所提方法的表现都显著优于 cmenet 及其他成熟方法。通过引入自适应权重，该方法取得了最高的 F_1 得分，在精确率和真阳性率之间实现了更好的平衡。自适应 cmenet 总是能通过筛选出更少的无关变量来获得显著更高的精确率，同时保持具有竞争力的真阳性率。与其他方法相比，它的均方预测误差也更低或相当，凸显了其提供出色预测

性能并提高变量选择准确性的能力。

将自适应 cmenet 与 Lasso 和自适应 Lasso 进行比较，可以看出其中存在一种权衡。有时，基于 Lasso 的方法表现出比自适应 cmenet 更高的真阳性率，这意味着它们识别出的真实活跃效应更多。然而，自适应 cmenet 的精确率更高，凸显了其筛选出较少无关变量的能力。这一规律在表亲效应、主效应和兄弟效应中始终存在，彰显了自适应 cmenet 在最大限度地实现真阳性检测、同时构建简约模型方面的相对优势。

比较三类活跃效应的相对表现，可以发现明显差异。主效应最容易识别，其选择指标值始终较高，许多接近或达到 1。这表明该方法几乎能够完美地选出所有活跃效应，且不会引入假正例。相比之下，表亲效应更难检测，选择指标值较低，尤其是随着活跃组数量增加时。兄弟效应则介于两者之间，在选择准确性方面表现中等。与表亲效应和兄弟效应相比，主效应的识别效果更优，这凸显了 CME约简原理在惩罚框架中的有效性。

现在考虑主效应之间存在相关性（$\rho = 1/\sqrt{2}$）的情形，如图 5(B) 所示。与其他方法相比，所提出的自适应 cmenet 在选择准确性方面具有明显优势，尤其是在活跃组数量较少时。在相关设置下，性能提升比非相关情形下更为显著。例如，在非相关情形下，某些方法（如 MCP）在堂表亲效应和同胞效应上的 F_1 得分与自适应 cmenet 相当。然而，与不相关设定相比，在相关情形下，自适应 cmenet 与 MCP 之间的 F_1 得分差距更为明显。只有当活跃组数量较大时，这种优势才会减弱。相关情形下的结果凸显了 CME 耦合原理在捕捉条件主效应组结构方面的有效性，因为随着相关性增强，CME 组结构会变得更加明显（图4）。

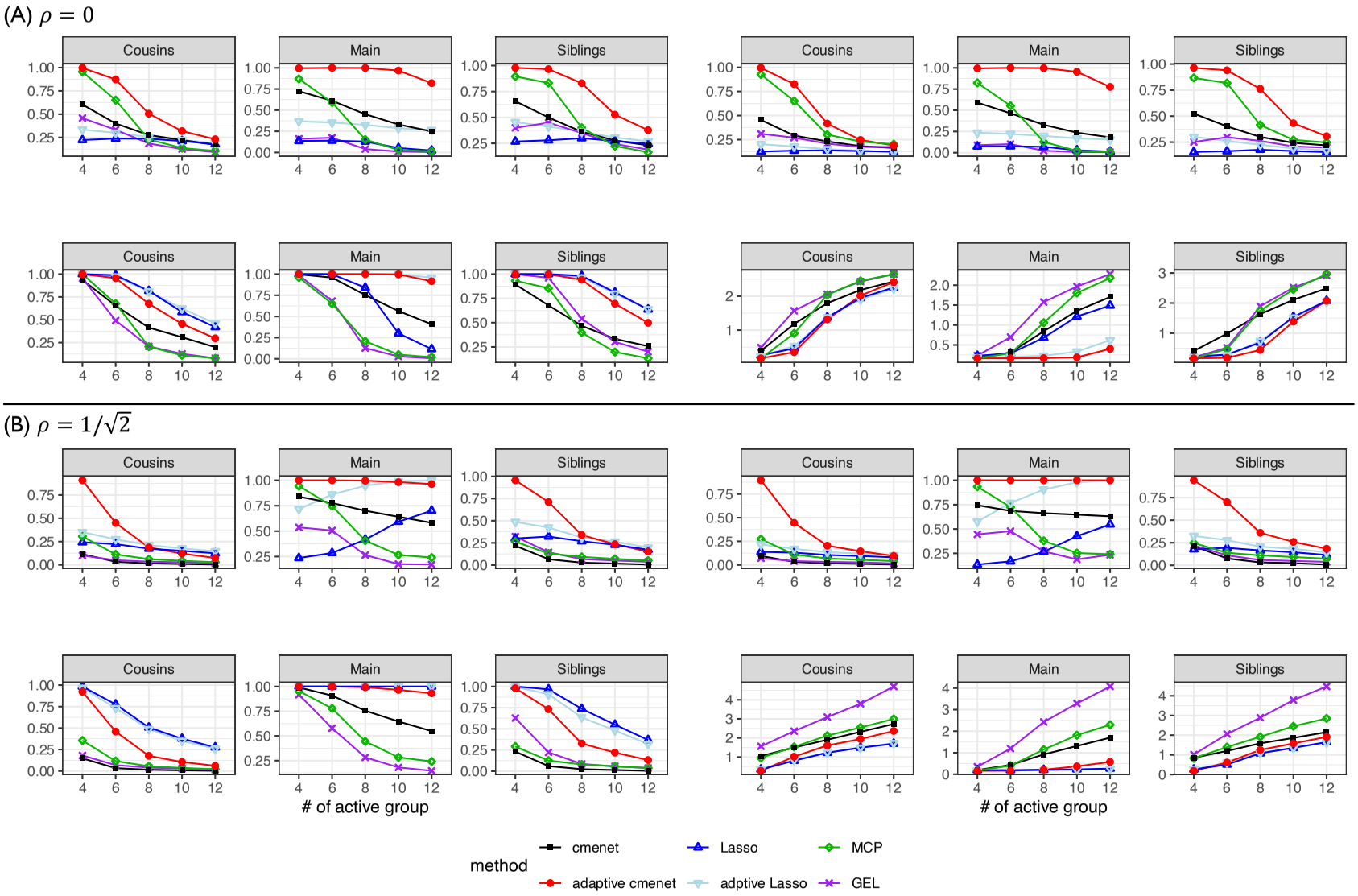


Figure 5: Example 1. Simulation results for continuous responses with $n = 100$, $p = 40$ for (A)uncorrelated MEs $\rho = 0$ and (B) correlated MEs $\rho = 1/\sqrt{2}$. The x-axis shows the number of active groups, while the y-axis displays the F_1 score, Precision, True Positive Rate (TPR), and MSPE. Each column corresponds to a type of active effect: cousins (left), main effects (middle), and siblings (right).

5.2 复杂效应结构的评估

在第二个模拟场景中，对于连续响应和二元响应，我们在底层真实模型中考虑两种不同类型的活跃效应：（1）活跃效应由主效应及其同胞组成，例如 $\{A, B, A|B+, A|D+, B|C-, B|E+\}$；（2）活跃效应由主效应及其表亲效应组成，例如 $\{A, B, C|A+, B|A-, D|B+, E|B+\}$。纳入主效应引入了重叠且含义模糊的组划分。不同于第一个示例，我们不确保活跃效应之间不存在其他类型的组结构。这种设置与现实世界的情形十分相似，为变量选择方法带来了额外挑战。在此，样本量为 $n = 50$，主效应数量为 $p = 20$，因此效应总数为 $p' = p + 4\binom{p}{2} = 780$。在本示例中，主效应的系数设为 $\beta_{\mathrm{ME}} = 5$，而条件主效应的系数为 $\beta_{\mathrm{CME}} = 1$。相关性设置和活跃组数量与情景 1 中的配置相同。作为敏感性分析，补充材料报告了另一项实验：固定总组数为四组且恰有两个活跃效应，其中效应量 β_{CME} 从 0.25 变化至 16（固定 β_{CME}=5），以评估信噪比和效应大小比 $\beta_{\mathrm{ME}}/\beta_{\mathrm{CME}}$ 对变量选择与预测的影响。

考虑图6(A)和6(B)中的连续响应，对于主组+表亲效应，自适应 cmenet 方法在不相关和相关设置下均优于其他方法。这一改进在相关情形中尤为明显，更强的组依赖关系增强了自适应 cmenet 平衡真阳性率与假阳性率的能力。对于主组+兄弟效应，自适应 cmenet 在非相关情形中的表现优于其他方法，这反映出当组结构定义较弱时检测兄弟效应的有效性。在相关情形中，与非相关情形相比，自适应 cmenet 表现出更好的性能，尤其是随着活跃组数量的增加。然而，自适应 cmenet 与自适应 Lasso 的 F_1 得分曲线在很大程度上重叠，表明总体性能相当，尽管自适应 cmenet 取得了更高的精确率和更低的真阳性率。一种可能的解释是，两种方法都依赖相同的岭估计量来生成初始权重，这影响了两种方法的惩罚结构。在所有情形中，自适应

cmenet 始终给出最低的均方预测误差，凸显了其将准确的变量选择与较强预测性能相结合的能力。

对于图6所示的二元响应结果，自适应 cmenet 在不相关和相关设定下均表现出竞争力，有效平衡了选择准确性与预测表现。对于主组+表亲效应，自适应 cmenet 在不相关设定下取得最高的 F_1 分数和真阳性率，凸显了其在保持较高精确率的同时捕捉真实正向关系的能力。对于主组+同胞效应，自适应 cmenet 在不相关和相关情形下的表现均可与自适应 Lasso 媲美，在精确率和真阳性率之间实现了良好的权衡。尽管在某些情形下，其 F_1 分数未能超过自适应 Lasso，但仍能保持较低的错误发现率。虽然 MCP 和 GEL 方法在 $\rho = 1/\sqrt{2}$ 时的精确率略高，但总体而言，自适应 cmenet 能更好地兼顾错误发现与真正例。

–

在图7中， `hierNet` 的预测精度相对高于自适应 cmenet，且在存在相关性（ $\rho = 1/\sqrt{2}$）以及活跃组数量增加时，这一优势会进一步扩大。不过，这种权衡十分明显： `hierNet` 选择的模型规模显著更大，表明它利用额外的成对交互作用来捕捉相依结构并提升预测表现。相比之下，自适应 cmenet 得到的模型更加简约，并具有可解释的 CME 结构，因此当可解释性和简约性是首要目标时，它更为合适。

6 基因关联案例研究

本节中，我们将所提方法应用于两项基因关联研究，以评估其在真实数据集中的表现：一项具有连续响应，另一项具有二元响应。在这两项研究中，主效应均为二元单核苷酸多态性（SNP）标记。每个SNP代表基因组中特定位置的一种遗传变异，并可呈现三种基因型：纯合参考型（AA）、杂合型（Aa）或纯合变异型（aa）。传统上，SNP采用加性 $\{0, 1, 2\}$ 编码方案，分别表示AA、Aa和aa（Lewis 2002）。

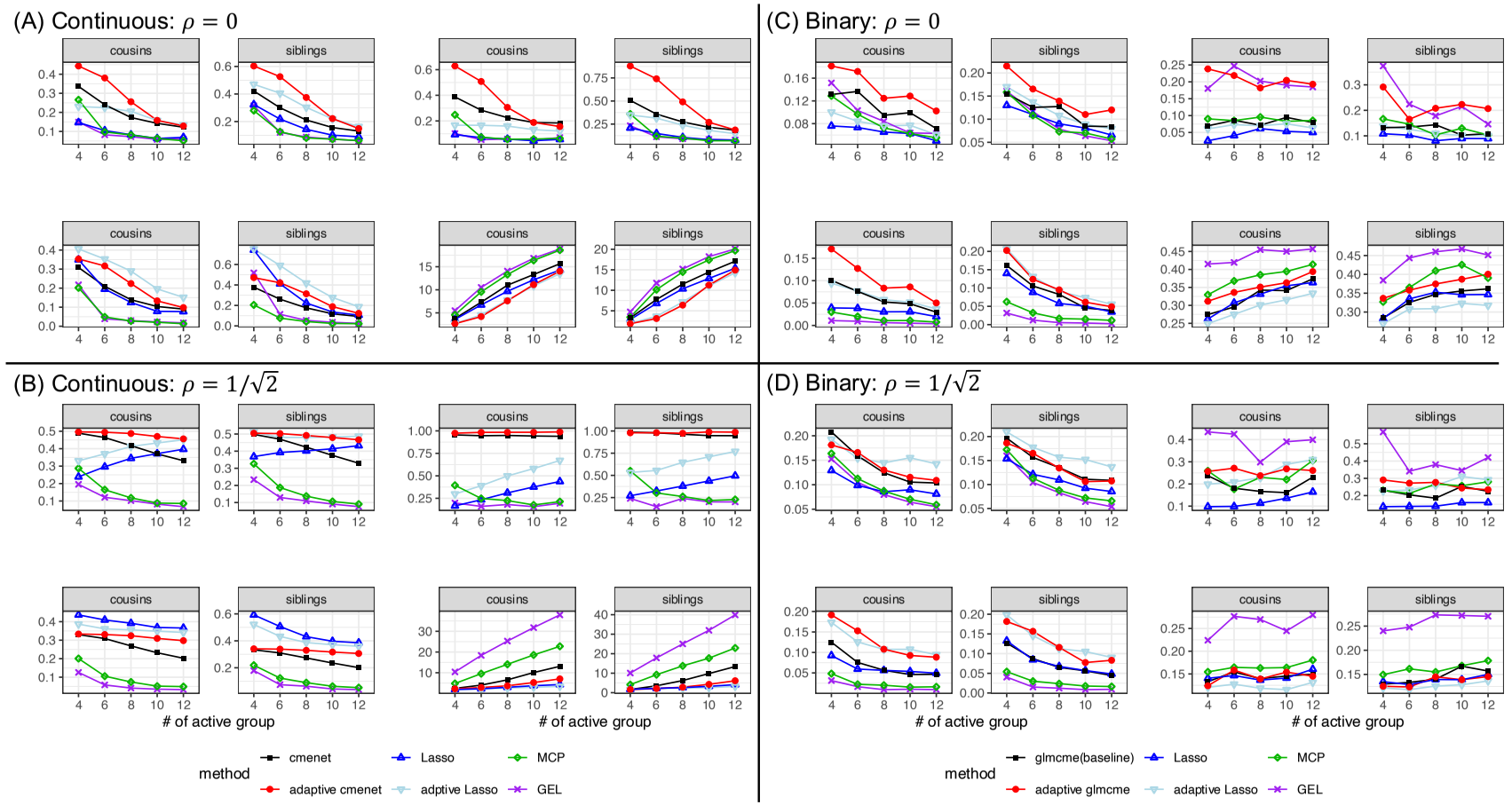


Figure 6: Example 2. Simulation results with $n = 50$, $p = 20$ for (A) continuous response with uncorrelated MEs $\rho = 0$, (B) continuous response with correlated MEs $\rho = 1/\sqrt{2}$, (C) binary response with uncorrelated MEs $\rho = 0$, and (D) binary response with correlated MEs $\rho = 1/\sqrt{2}$. The x-axis shows the number of active groups, while the y-axis displays the F_1 score, Precision, True Positive Rate (TPR), and MSPE (for continuous) or Misclassification Rate (for binary). Each column corresponds to a type of active effect: main+cousins (left), and main+siblings (right).

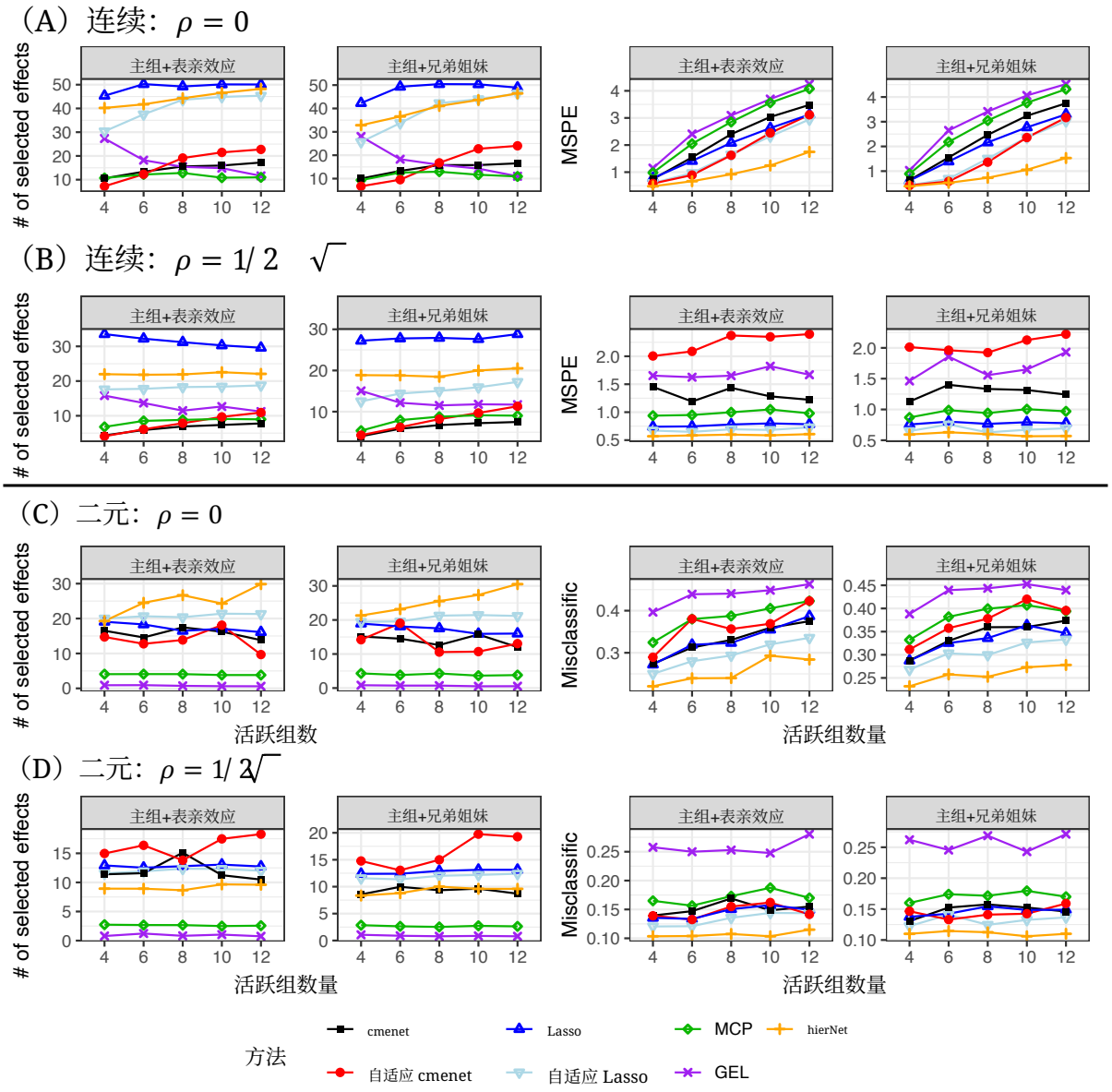


图7：例2。与hierNet的比较，其中 $n = 50$、 $p = 20$分别用于(A) 具有不相关主效应的连续响应 $\rho = 0$、(B) 具有相关主效应的连续响应$\rho = 1/\sqrt{2}$、(C) 具有不相关主效应的二元响应 $\rho = 0$，以及(D) 具有相关主效应的二元响应 $\rho = 1/\sqrt{2}$。横轴显示活跃组数量，纵轴显示选定效应的总数，以及均方预测误差（连续）或误分类率（二元）。每一列对应一种活跃效应类型：主组+表亲效应（左）和主组+兄弟姐妹（右）。

表 2：基因关联研究中六种方法的变量选择性能和预测性能。

	自适应 cmenet	cmenet Lasso		MCP	GEL hierNet	
（A）玉米开花时间						
选定效应的数量	19	24	43	12	21	66
显著入选效应（$p < 0.05$）	14	5	11	11	10	16
MSPE	3.83	5.47	4.95	6.38	5.81	4.68
（B）拟南芥						
选定效应的数量	9	-	13	11	5	93
显著入选效应（$p < 0.05$）	7	-	4	1	5	29
误分类率	0.1	-	0.1	0.13	0.3	0.01

在两项研究中，我们分析了来自近交系的数据集；近交过程确保所有单核苷酸多态性（单核苷酸多态性）均为纯合型（AA 或 aa），从而排除杂合基因型。因此，这些数据集中的单核苷酸多态性天然采用二元编码，不包含杂合类别。为了与我们所提方法的框架保持一致，我们将 $\{0, 1\}$ 重新标记为 $\{-1, 1\}$。所有调参和性能报告均通过5折交叉验证，在样本外进行。这些分析旨在识别显著的基因条件主效应，用以量化一个遗传变异在另一个遗传变异存在或不存在的条件下所产生的加性效应或显性效应。

在这两项研究中，我们将所提出的方法与 cmenet、Lasso、MCP、GEL 和 hierNet（R 软件包 `hierNet`V1.9；Bien 和 Tibshirani，2020）进行比较。hierNet方法（Bien 等，2013）是一种广泛用于在高维数据中选择双向交互作用（二因子交互）的方法。通过将自适应 cmenet 与 hierNet 进行比较，我们旨在评估其在识别基因间交互关系方面的实际优势。请注意，在这些真实世界的研究中，真实的活跃效应通常是未知的。这里，我们着重评估所估计的模型能否在保持较高预测精度的同时，选择数量更少的效应（Stoica 和 S¨oderstr¨om，1982）。这一标准对于基因组研究至关重要，因为效应过多、过于复杂的模型可能会降低可解释性并导致过拟合。一个所选效应更少而预测精度相当的模型

表 3：玉米开花时间研究中线性回归各方法所选部分效应的 P值。

方法	$g1\|g17-$	$g1\|g39-$	$g9\|g10-$	$g11\|g38+$	$g12\|g38-$	$g14\|g38+$	$g15\|g35+$	$g18\|g33-$	$g27\|g26-$
自适应 cmenet	-	0.0004	0.0183	0.0000	0.0351	0.0042	0.0003	0.0062	0.0143
cmenet	-	0.1214	0.1281	0.0091	0.0124	0.0966	0.0007	0.1629	0.0180
Lasso	0.0392	0.0953	0.1951	0.0043	0.1050	0.0790	0.0372	0.0343	0.0200
MCP	0.2573	0.0544	0.0196	0.0333	0.0646	0.8667	0.0022	0.0133	0.0242
GEL	0.0085	0.0483	0.0049	0.0029	0.1048	0.0001	0.0001	0.0008	0.0081
hierNet	0.0259	0.0510	0.0002	0.0057	0.0000	0.0259	0.0002	0.0001	0.0001

或预测精度更高的模型，更有可能捕捉到有意义的生物学关系，同时避免虚假发现。此外，我们使用常规（广义）线性回归报告选定效应的 P 值，以验证这些选定效应的相关性。这一步可以进一步证明该方法识别具有生物学意义关联的能力。

第一项研究聚焦于玉米开花时间，这是一项用于测量开花天数的连续变量。我们分析了Buckler等人（2009）数据集的一个子集，其中包含 $n = 150$ 个观测值，以及由二元SNP标记表示的 $p = 40$ 个主效应。与几种成熟方法相比，所提方法展现出更优的性能。如表2(a)所示，自适应 cmenet在选定的19个效应中识别出14个显著效应，同时保持了低于hierNet的均方预测误差。尽管hierNet选出了更多显著效应，但它也包含了更多不显著效应，因此均方预测误差更高。另一方面，MCP方法选出的总效应数较少（其中大多数为显著效应），但其均方预测误差仍高于自适应 cmenet。这些结果表明，所提方法能够在降低预测误差的同时，高效捕捉关键的基因互作。此外，表3中对选定效应的p值分析显示，自适应 cmenet识别出的显著效应多于其他方法；自适应 cmenet与hierNet选出的显著效应之间也存在较大重叠。

第二个 st　　本研究利用拟南芥的黄化表型进行分析　　g

表 4：拟南芥研究中逻辑回归的各方法所选部分效应的 P值。

	自适应 cmenet	Lasso	MCP	GEL	hierNet	单变量回归
2： 5804076|2: 8750047+	0.0069	0.0076	-	0.0000	0.0214	0.0007
2： 8750047|5: 1931167+	0.0337	0.0960	-	0.0004	0.0197	0.0005
2： 18752840|3: 4818602+	0.0122	0.0298	-	0.0008	0.0077	0.0066
2： 18752840|4: 11580143−	0.0096	0.0133	-	0.0011	0.0276	0.0006
3： 7392098|3: 21632608+	0.0028	0.0491	-	-	0.3579	0.0034
4： 17038400|2: 9256095+	0.5980	0.0502	-	-	0.0020	0.0012
4： 17038400|5: 1931167−	0.0314	0.2270	-	-	0.0197	0.0013
5： 20409250|2: 9256095−	0.0063	0.5506	-	-	0.0020	0.0153

数据来自 Atwell 等人（2010年）开展的一项全基因组关联研究（GWAS）。拟南芥是一种分布广泛、主要通过自花受精繁殖的模式植物，在许多具有适应性意义的性状上表现出显著的遗传变异（Atwell 等人，2010年）。该数据集包含199个品系的250,000个单核苷酸多态性（SNP）基因型信息，以及107种性状的表型测量数据，数据可从 `https://cynin.gmi.oeaw.ac.at/home/resources/atpolydb` 公开获取。就可见黄化表型而言，在生长五周后于16°C条件下进行评估，84个品系表现出黄化，而92个品系未出现黄化。本分析选取了该数据集的一个子集，其中包含100个品系（55个出现可见黄化，45个未出现），$p = 63$ 主效应对应于二元SNP标记。响应变量为二元变量，表示是否出现可见黄化。参见表2(B)，自适应 cmenet 的表现优于其他方法：在保持具有竞争力的误分类率的同时，它识别出了更多显著效应。相比之下，hierNet方法选出的显著效应数量最多，但也纳入了更多不显著效应，导致模型过拟合。MCP方法选出的效应数量与自适应 cmenet 相近；然而，其中只有一个

具有统计显著性。值得注意的是，该方法识别出的效应与其他方法选出的效应没有任何重叠。尽管 GEL 方法识别出的效应较少，其中有五个具有显著性，但其误分类率较高。自适应 cmenet 通过选出更多显著的显性效应，同时尽量减少纳入不显著效应，实现了更好的平衡，有效兼顾了选择准确性与预测可靠性。表 4 中对选定效应的 p 值分析表明，自适应 cmenet 方法和 hierNet 方法均能持续识别出相互重叠的显著显性效应。值得注意的是，自适应 cmenet 识别出的所有条件主效应（CMEs）均在单变量回归模型中得到验证，具有统计显著性，这进一步凸显了所提方法卓越的选择性能。

为探究所选条件主效应（CMEs）的生物学含义，我们将每个相关单核苷酸多态性（SNP）映射到距离最近的 TAIR10 基因（范围为 ± 10 kb），并使用基因本体论、KEGG 和 AraCyc 对这些基因进行功能注释。值得注意的是，排名靠前的条件主效应通常将两个位点配对，而其注释功能表明它们之间存在与黄化相关的条件关系。例如，我们检测到 APX5（AT4G35970）与 2-Cys过氧化物还原酶B（AT5G06290）之间存在一个条件主效应（$4:17038400|5:1931167-$）。这两种酶都是光合组织中活性氧解毒和氧化还原缓冲的关键组成部分。该条件主效应的负向极性表明，缺少 2-Cys Prx B 会削弱其保护或调控作用，使与 APX5 相关的遗传变异成为决定黄化表型的更关键因素。这一发现与抗坏血酸过氧化物酶和过氧化物还原酶氧化还原通路之间的补偿机制一致：2-Cys Prx 活性丧失可由其他叶绿体抗氧化系统补偿，尤其是与超氧化物歧化酶协同作用的抗坏血酸-谷胱甘肽循环，因此依赖性会转向与 APX 相关的途径（Puerto-Gal´an 等，2013）。这一例子阐明了条件主效应超越传统交互作用的贡献：它们能够捕捉依赖情境的效应，即一个位点的影响只有在另一个具有生物学关联的位点存在（或缺失）时才会显现。

7 讨论

本研究聚焦于广义线性模型中条件均值效应的自适应双层变量选择。所提方法采用的自适应惩罚结构能够实现更灵活、更具适应性的条件主效应选择。所提自适应 cmenet 增强了捕捉重要条件关系的能力，同时降低了耦合约束。

在此，我们想指出所提方法仍有若干可改进之处。首先，与迭代重加权最小二乘算法类似，第 4 节中的参数估计算法可能无法相对于实际目标函数收敛到全局最优解。进一步研究所提方法中参数估计的数值收敛性质将很有意义。尽管如此，所提方法在数值研究中表现出稳健的性能。其次，尽管自适应 cmenet 引入了额外惩罚以考虑表亲效应和兄弟姐妹等组结构，初始权重的选择可能会限制其充分利用这些结构的能力。因此，针对条件均值效应框架量身定制的替代加权方案可能进一步提升自适应 cmenet 的性能。第三，尽管自适应 cmenet 的类似理论框架可以扩展到典则连接的泊松回归和伽马回归，但计算细节因情形而异：方差函数（泊松 Var $=\mu$；伽马 Var $=\phi\mu^2$）和工作权重会改变曲率、步长和数值保护。交叉验证采用适当的偏差损失进行，热启动/活动集更新仍然有效，但需要进行这些针对具体情形的研究。第四，进一步提高算法的计算效率以实现可扩展性十分重要。最后，可以考虑将所提方法学扩展到处理多层或连续变量的条件主效应，从而拓宽其对多样化数据集和实验设置的适用性。此外，鉴于条件主效应起源于实验设计，可以探索 CME 在最优实验设计和序贯设计中的应用。

补充材料

Supplement.pdf 包含补充的方法学、证明、算法细节以及扩展结果。Si
mulation.zip 包含用于复现文稿中所报告模拟和图的 R 脚本。

致谢

作者感谢编辑、副编辑和匿名审稿人提出的深刻见解和建设性建议，这些意见极大地提升了本文的质量和清晰度。

利益声明

作者声明不存在利益冲突。

参考文献

Atwell, S.、Huang, Y. S.、Vilhj´almsson, B. J.、Willems, G.、Horton, M.、Li, Y.、Meng, D.、Platt, A.、Tarone, A. M.、Hu, T.T. 等（2010）。拟南芥近交系中107种表型的全基因组关联研究。《自然》，465(7298)：627–631。

Bien, J.、Taylor, J. 和 Tibshirani, R.（2013）。层次交互作用的 Lasso。《统计年刊》，41(3)：1111。

Bien, J. 和 Tibshirani, R.（2020）。*hierNet*：用于层次交互作用的 *Lasso*。R 软件包 1.9 版。
Breheny, P.（2015）。用于双层变量选择的组指数套索。《生物统计学》，71(3)：731–740。
Breheny, P. 和 Huang, J.（2009）。用于双层变量选择的惩罚方法。《统计学及其应用》，2(3)：369。

Breheny, P.and Huang, J. (2011). 非凸惩罚回归的坐标下降算法及其在生物特征选择中的应用。《应用统计年刊》, 5(1):232–253.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J. C., 等. (2009). 玉米开花时间的遗传结构。《科学》, 325(5941):714–718. Fang, K., Wang, X., Zhang, S., Zhu, J., and Ma, S. (2015). 基于自适应稀疏组 Lasso 的双层变量选择。《统计计算与模拟杂志》, 85(13):2750–2760. Friedman, J., Hastie, T., H¨ofling, H., and Tibshirani, R. (2007). 沿路径坐标优化。《应用统计年刊》, 1(2):302–332. Friedman, J., Tibshirani, R., and Hastie, T. (2010). 通过坐标下降求解广义线性模型的正则化路径。《统计软件杂志》, 33(1):1–22. Fu, W. J. (1998). 惩罚回归：桥惩罚与 Lasso 之比较。《计算与图形统计学杂志》, 7(3):397–416. Green, P. J. (1984). 用于最大似然估计的迭代加权最小二乘，以及一些稳健和抗差替代方法。《英国皇家统计学会杂志*B*辑：统计方法学》, 46(2):149–170. Huang, J., Ma, S., Xie, H., and Zhang, C.-H. (2009). 一种用于变量选择的组桥惩罚方法。《生物统计学》, 96(2):339–355. Hunter, D. R. and Lange, K. (2000). 基于 MM算法的分位数回归。《计算与图形统计学杂志》, 9(1):60–77. Lewis, C. M. (2002). 遗传关联研究：设计、分析与解释。*Briefings in Bioin-formatics*, 3(2):146–153. Mak, S. (2022). ***cmenet***：条件主效应的双层选择。R 软件包版本 0.1.2。Mak, S. and Wu, C. J. (2019). cmenet：一种用于条件主效应双层变量选择的新方法。《美国统计学会杂志》, 114(526):844–856. Mazumder, R., Friedman, J. H., and Hastie, T. (2011). Sparsenet：采用非凸惩罚的坐标下降。《美国统计学会杂志》, 106(495):1125–1138. Puerto-Gal´an, L., P´erez-Ruiz, J. M., Ferr´andez, J., Cano, B., Naranjo, B., N´ajera, V. A., Gonz´alez, M., Lindahl, A. M., and Cejudo, F. J. (2013). 叶绿体 2-Cys过氧化物还原蛋白的过度氧化：平衡过氧化氢的毒性与信号传导作用。《植物科学前沿》, 4:310.

Shibata, M., Terada, A., Kawaguchi, T., Kamatani, Y., Okada, D., Nagashima, K., Ohmura, K., Matsuda, F., Kawaguchi, S., Sese, J., 等（2024）。利用 LAMPLINK 和日本队列识别类风湿关节炎中的上位性 SNP 组合。人类遗传学杂志，69(10):541–547。Stoica, P. 和 S¨oderstr¨om, T.（1982）。论简约原则。国际控制杂志，36(3):409–418。Su, H. 和 Wu, C. J.（2017）。CME分析：一种揭示二水平部分因子试验中别名效应的新方法。质量技术杂志，49(1):1–10。Tay, J. K., Narasimhan, B. 和 Hastie, T.（2023）。适用于所有广义线性模型的弹性网络正则化路径。《统计软件杂志》，106(1):1–31。Wu, C. J.（2015）。后Fisher实验：从物理到虚拟。《美国统计学会杂志》，110(510):612–620。Wu, C. J. 和 Hamada, M. S.（2011）。实验：规划、分析与优化。John Wiley & Sons。Zhang, C.-H.（2010）。极小极大凹惩罚下近乎无偏的变量选择。统计学年刊，38(2):894–942。Zou, H.（2006）。自适应 Lasso 及其 Oracle 性质。《美国统计学会杂志，101(476):1418–1429。

《广义线性模型条件主效应的自适应双层变量选择》的补充材料

本补充材料包含：（i）命题 1、推论1.1、引理1和定理 1的技术性证明；（ii）第4.2节中自适应 cmenet 的算法；以及（iii）例2（第 5.2 节）中针对跨效应量 β_{CME} 的敏感性分析。

命题 1 的证明

$Q(\boldsymbol{\beta})$ 在 $\boldsymbol{\beta} = \mathbf{0_p}$ 处不可微，但一阶和二阶半可微。也就是说，对于所有点 $\boldsymbol{\beta}$ 以及所有方向$\mathbf{u} \in \mathbb{R}^p$ $(\|\mathbf{u}\| = 1)$（ $\boldsymbol{\beta}$），它都存在方向导数。我们采用与 Mak 和 Wu（2019）命题 1 类似的方法。令 ω_j为所有系数 $\beta_j, j = 1,\ldots,p'$对应的个体权重，并定义 $\omega_{\max} = \max_{1\le j\le p'} \omega_j$。 $Q(\boldsymbol{\beta})$沿$\mathbf{u}$ 方向的二阶导数为：

$$\nabla^2_{\mathbf{u}} Q(\boldsymbol{\beta}) = \frac{1}{n}\mathbf{u}^T\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{u} + \nabla^2_{\mathbf{u}}\tilde{P}_S(\boldsymbol{\beta}) + \nabla^2_{\mathbf{u}}\tilde{P}_C(\boldsymbol{\beta}), \tag{A.1}$$

其中，对于 $\mathcal{G} \in \{\mathcal{S},\mathcal{C}\}$， $\nabla^2_{\mathbf{u}}\tilde{P}_{\mathcal{G}}(\boldsymbol{\beta}) = \mathbf{u}^T\nabla^2_{\boldsymbol{\beta}}\tilde{P}_{\mathcal{G}}(\boldsymbol{\beta})\mathbf{u}$。不失一般性，海森矩阵 $H_{\boldsymbol{\beta}} \overset{\text{def}}{=} \nabla^2_{\boldsymbol{\beta}}\tilde{P}_{\mathcal{S}}(\boldsymbol{\beta})$ 的元素为：

$$\frac{\partial^2}{\partial\beta_{j|m+}\partial\beta_{h|n+}}\tilde{P}_{\mathcal{S}}(\boldsymbol{\beta}) = \begin{cases} f^{(j)\prime\prime}_{o,\mathcal{S}}\left\{\sum_{k\in\mathcal{S}(j)}\omega_k f^{(j)}_{i,\mathcal{S}}(\beta_k)\right\}(f^{(j)\prime}_{i,\mathcal{S}}(\beta_{j|m+}))^2\omega^2_{j|m+} + f^{(j)\prime}_{o,\mathcal{S}}\left\{\sum_{k\in\mathcal{S}(j)}\omega_k f^{(j)}_{i,\mathcal{S}}(\beta_k)\right\}f^{(j)\prime\prime}_{i,\mathcal{S}}(\beta_{j|m+})\omega_{j|m+}, & \text{if } j=h, m=n, \\ f^{(j)\prime\prime}_{o,\mathcal{S}}\left\{\sum_{k\in\mathcal{S}(j)}\omega_k f^{(j)}_{i,\mathcal{S}}(\beta_k)\right\}f^{(j)\prime}_{i,\mathcal{S}}(\beta_{j|m+})f^{(j)\prime}_{i,\mathcal{S}}(\beta_{j|n+})\omega_{j|m+}\omega_{j|n+}, & \text{if } j=h, m\neq n, \\ 0, & \text{if } j\neq h. \end{cases} \tag{A.2}$$

因此，很容易证明

$$\nabla^2_{\mathbf{u}}\tilde{P}_{\mathcal{G}}(\boldsymbol{\beta}) = \sum_{i=1}^{p'}\sum_{j=1}^{p'}\{H_{\boldsymbol{\beta}}\}_{i,j}u_iu_j \ge -\tau\exp(0)\omega^2_{\max} + \lambda_{\mathcal{G}(j)}\exp(0)(-\frac{1}{\lambda_{\mathcal{G}(j)}\gamma})\omega_{\max} = -\tau\omega^2_{\max} - \frac{\omega_{\max}}{\gamma},$$

对所有$\mathbf{u}$ 和 $\boldsymbol{\beta}$均成立。此外，令$\mathbf{u} = \sum_{i=1}^{p'} u_i\mathbf{e}_i$，其中$\mathbf{e}_i$ 表示$\mathbf{X}^T\mathbf{W}\mathbf{X}$的单位特征向量，满足特征值方程$\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{e}_i = \zeta_i\mathbf{e}_i$。于是，我们有$\mathbf{u}^T(\mathbf{X}^T\mathbf{W}\mathbf{X})\mathbf{u} = \sum_{i=1}^{p'} u_i^2\zeta_i \ge \zeta_{\min}$。因此，

$$\nabla^2_{\mathbf{u}}Q(\boldsymbol{\beta}) \ge \frac{\zeta_{\min}(\mathbf{X}^T\mathbf{W}\mathbf{X})}{n} - 2(\tau\omega^2_{\max} + \frac{\omega_{\max}}{\gamma}),$$

当 $\tau + \frac{1}{\gamma\omega_{\max}} < \frac{\zeta_{\min}(\mathbf{X^T W X})}{2n\omega_{\max}^2}$ 时，该值严格为正。

推论1.1的证明

由命题 1 可知，若以下条件成立 $\tau + \frac{1}{\gamma\omega_{\max}} < \frac{\zeta_{\min}(\mathbf{X}^T\mathbf{W}\mathbf{X})}{2n\omega_{\max}^2}$，则目标函数 $Q(\boldsymbol{\beta})$ 是严格凸的，其中 $\mathbf{W} = \mathrm{diag}(w_1, w_2, \ldots, w_n)$ 是权重的对角矩阵。在逻辑回归的情形下，权重项定义为 $w_i = \pi_i(1-\pi_i)$，其中 $\pi_i = \frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{1+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}$。由此可得 $w_i \le 1/4$，当 $\pi_i = 1/2$ 时等号成立。

此外，利用性质 $\|\mathbf{x}_j\|_2^2 = n$，我们证明：只要条件$\tau + \frac{1}{\gamma\omega_{\max}} < \frac{1}{8\omega_{\max}^2}$ 成立，逐坐标目标函数 $Q_j(\boldsymbol{\beta}_j)$ 对所有 $j = 1, \ldots, p'$ 仍保持严格凸。

引理1的证明

根据一阶泰勒展开，由于 $\eta_{\lambda,\tau}(\theta)$ 是凹可微函数，容易得出：

$$\eta_{\lambda_{\mathcal{G}},\tau}(\boldsymbol{\beta}_{\mathcal{G}}|\boldsymbol{\beta}_{\mathcal{G}}^*) \le \eta_{\lambda_{\mathcal{G}},\gamma}\{\|\boldsymbol{\beta}_{\mathcal{G}}^*\|_{\lambda_{\mathcal{G}},\omega,\gamma}\} + \Delta_{\mathcal{G}}^* \times \{\|\boldsymbol{\beta}_{\mathcal{G}}\|_{\lambda_{\mathcal{G}},\omega,\gamma} - \|\boldsymbol{\beta}_{\mathcal{G}}^*\|_{\lambda_{\mathcal{G}},\omega,\gamma}\},$$

其中，该不等式成立是因为 η 在 $[0,\infty)$ 上具有凹性。因此，直接代入后，(12) 中的 $Q_{j|k+}(\beta_{j|k+}|\boldsymbol{\beta}^*)$ 是 $Q_{j|k}(\beta_{j|k+})$ 的一个主化函数。

定理 1 的证明

为便于分析，对于任意效应系数 $\beta_j, j = 1$、.....、p'，令 λ_1 和 λ_2 分别表示与兄弟组和表亲组相关的调节参数。相应的惩罚斜率记为 Δ_1 和 Δ_2。考虑如下优化问题：

$$\hat{\beta}_j = \min_{\beta_j}\left[\frac{1}{2n}(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{W}(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}) + \Delta_1\omega_j g_{\lambda_1,\gamma}(\beta_j) + \Delta_2\omega_j g_{\lambda_2,\gamma}(\beta_j)\right] \tag{A.3}$$

这等价于

$$\begin{cases} \min_{\beta_j} L(\beta_j) = \frac{1}{2n}(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{W}(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}) \\ s.t., \omega_j g_{\lambda_1,\gamma}(\beta_j) \le 0, \omega_j g_{\lambda_2,\gamma}(\beta_j) \le 0 \end{cases} \tag{A.4}$$

T对目标函数关于 β_j 求导，得到 :

$$\begin{aligned}\frac{\partial L(\beta_j)}{\partial \beta_j} &= \frac{1}{2n}\left(-2\mathbf{x}_j^T\mathbf{W}\tilde{\mathbf{y}} + 2\mathbf{x}_j^T\mathbf{W}\mathbf{X}\boldsymbol{\beta}\right) \\ &= -\frac{1}{n}\mathbf{x}_j^T\mathbf{W}(\tilde{\mathbf{y}} - \mathbf{X}_{-j}\boldsymbol{\beta}_{-j}) + \frac{1}{n}\mathbf{x}_j^T\mathbf{W}\mathbf{x}_j\beta_j \\ &= -z_j + v_j\beta_j,\end{aligned}$$

其中 $z_j = \frac{1}{n}\mathbf{x}_j^T\mathbf{W}(\tilde{}\mathbf{y} - \mathbf{X}_{-j}\boldsymbol{\beta}_{-j})$，且 $v_j = \frac{1}{n}\mathbf{x}_j^T\mathbf{W}\mathbf{x}j$。对于约束函数 $g_{\lambda,\gamma}(\beta_j)$，关于 β_j 的导数为:

$$\partial g_{\lambda,\gamma}(\beta_j) = \begin{cases} 0, & \text{if } |\beta_j| > \lambda\gamma, \\ \operatorname{sgn}(\beta_j)\left(1 - \frac{|\beta_j|}{\lambda\gamma}\right), & \text{if } |\beta_j| \le \lambda\gamma, \\ [-1, 1], & \text{if } \beta_j = 0. \end{cases}$$

应用 Ka rush-Kuhn-Tucker（KKT）条件，我们推导出最优性条件:

1. 平稳性条件:

$$0 \in -z_j + v_j\beta_j + \Delta_1\omega_j\partial g_{\lambda_1,\gamma}(\beta_j) + \Delta_2\omega_j\partial g_{\lambda_2,\gamma}(\beta_j). \tag{A.5}$$

2. C互补松弛条件: $\Delta_1\omega_j g_{\lambda_1,\gamma}(\beta_j) = 0$, $\Delta_2\omega_j g_{\lambda_2,\gamma}(\beta_j) = 0$。

3. 原始可行性条件: $\omega_j g_{\lambda_1,\gamma}(\beta_j) \le 0$, $\omega_j g_{\lambda_2,\gamma}(\beta_j) \le 0$。

4. 对偶可行性条件: $\Delta_1 \ge 0, \Delta_2 \ge 0$。

不失一般性，假设 $z_j \ge 0$。令 $\lambda_{(1)} = \max(\lambda_1, \lambda_2)$，且 $\lambda_{(2)} = \min(\lambda_1, \lambda_2)$，其中 $\Delta_{(1)}$ 和 $\Delta_{(2)}$ 是其对应的斜率。考虑与式 (13) 中相同的四种情况: z_j

ˆ ˆ 1. 若 $\beta_j \ge \lambda_{(1)}\gamma$，则由平稳性条件 (A.5) 可得 $0 \in -z_j + v_j\beta_j$，这意味着: ˆ $\beta_j = \frac{z_j}{v_j}$。因此，_ $z_j \ge v_j\lambda_{(1)}\gamma$。由于 (A.3) 是严格凸的，ˆ $\beta_j = \frac{z_j}{v_j}$ 是 (A.3) 的唯一解。

ˆ ˆ 2. 若 $\lambda_{(2)}\gamma \le \beta_j < \lambda_{(1)}\gamma$，则根据平稳性条件 (A.5)， $0 \in -z_j + v_j\beta_j + \Delta_{(1)}\omega_j$ $(1 - \hat{}\frac{\beta_j}{\lambda_{(1)}\gamma})$，由此可得: ˆ $\beta_j = \frac{z_j - \Delta_{(1)}\omega_j}{v_j - \frac{\Delta_{(1)}\omega_j}{\lambda}}$ $_{(1)}\gamma$。因此， $v_j\lambda_{(2)}\gamma + \Delta_{(1)}\omega_j(1 - \frac{\lambda_{(2)}}{\lambda_{(1)}}) \le$ˆ $z_j < v_j\lambda_{(1)}\gamma$。由于 (A.3) 是严格凸的， β_j 是唯一解。

ˆ ˆ 3. 若 $0 \leq \beta_j < \lambda_{(2)}\gamma$，则根据平稳性条件 (A.5)，$0 \in \omega_j - z_j + v_j\beta_j + \Delta_{(1)}\omega_j(1 - \frac{\beta_j}{\lambda_{(1)}\gamma}) + \Delta_{(2)}\omega_j(1 - \frac{\beta_j}{\lambda_{(2)}\gamma})$，由此可得：$\hat{\beta}_j = \frac{v_j - \frac{\Delta_{(1)}\omega_j}{\lambda} - \frac{\Delta_{(2)}\omega_j}{\lambda}}{\gamma}$ $_{(1)(2)}$。因此，$\Delta_{(1)}\omega_j + \Delta_{(2)}\omega_j \leq z_j < v_j\lambda_{(2)}\gamma + \Delta_{(1)}\omega_j(1 - \frac{\lambda_{(2)}}{\lambda_{(1)}})$。由于 (A.3) 是严格凸的，$\hat{\beta}_j$ 是 (A.3) 的唯一解。

ˆ4. 若 $\beta_j = 0$，则根据平稳性条件 (A.5)，$0 \in -z_j + \Delta_{(1)}\omega_j[-1, 1] + \Delta_{(2)}\omega_j[-1, 1]$，由此可得：$0 \leq z_j < \Delta_{(1)}\omega_j + \Delta_{(2)}\omega_j$。由于 (A.3) 是严格凸的，$\hat{\beta}_j = 0$ 是 (A.3) 的唯一解。

命题 2 的证明

命题 2 可用与 Breheny（2015）命题3类似的方法证明 .

自适应 cmenet 算法

Algorithm 1 Adaptive cmenet: An algorithm for adaptive bi-level cmenet selection

function ADAPTIVE CMENET($\mathbf{X}, \mathbf{y}, \lambda_s, \lambda_c, \gamma, \tau, \boldsymbol{\beta} = \mathbf{0_{p'}}, \hat{\boldsymbol{\beta}}^{ridge}$)

  Initialize $\beta_0 = \log(\bar{y}/(1-\bar{y}))$, $\eta_i = \log(\bar{y}/(1-\bar{y}))$, $\Omega_{S(j)} = (\|\hat{\boldsymbol{\beta}}^{ridge}_{S(j)}\|_1 + \frac{1}{n})^{-1}$, $\Omega_{C(j)} = (\|\hat{\boldsymbol{\beta}}^{ridge}_{C(j)}\|_1 + \frac{1}{n})^{-1}$, $\Delta_{S(j)} = \lambda_s$, $\Delta_{C(j)} = \lambda_c$, $\lambda_{S(j)} = \lambda_s \Omega_{S(j)}$, $\lambda_{C(j)} = \lambda_s \Omega_{C(j)}$, for $j = 1, 2, ..., p$; $\boldsymbol{\omega} = \{(|\hat{\beta}^{ridge}_j| + \frac{1}{n})^{-1}\}_{j=1}^{p'}$

  repeat

    Calculate $\pi_i \leftarrow \frac{\exp(\eta_i)}{1+\exp(\eta_i)}$, $w_i \leftarrow \pi_i(1-\pi_i)$, $r_i \leftarrow \frac{y_i - \pi_i}{w_i}$, for $i = 1, 2, ..., n$

    Update $\beta_0^{(0)} \leftarrow \beta_0$, $\beta_0 \leftarrow \frac{\mathbf{1_n^T W r}}{\mathbf{1_n^T W 1_n}} + \beta_0^{(0)}$ ▷ For intercept

    $\mathbf{r} \leftarrow \mathbf{r} - \mathbf{1}_n(\beta_0 - \beta_0^{(0)})$, $\boldsymbol{\eta} \leftarrow \boldsymbol{\eta} + \mathbf{1}_n(\beta_0 - \beta_0^{(0)})$

    for $j = 1, \ldots, p$ **do** ▷ For all MEs

      Calculate $v_j \leftarrow \frac{1}{n}\mathbf{x}_j^T \mathbf{W} \mathbf{x}_j$, $z_j \leftarrow \frac{1}{n}\mathbf{x}_j^T \mathbf{W r} + v_j \beta_j^{(0)}$

      Update $\beta_j^{(0)} \leftarrow \beta_j$, $\beta_j \leftarrow S_{\lambda_{S(j)}, \lambda_{C(j)}}(z_j, \Delta_{S(j)}, \Delta_{C(j)})$

      $\mathbf{r} \leftarrow \mathbf{r} - \mathbf{x}_j(\beta_j - \beta_j^{(0)})$, $\boldsymbol{\eta} \leftarrow \boldsymbol{\eta} + \mathbf{x}_j(\beta_j - \beta_j^{(0)})$

      $\Delta_{S(j)} \leftarrow \Delta_{S(j)} \exp\{-\frac{\tau}{\lambda_{S(j)}} w_j [g_{\lambda_{S(j)}, \gamma}(\beta_j) - g_{\lambda_{S(j)}, \gamma}(\beta_j^{(0)})]\}$

      $\Delta_{C(j)} \leftarrow \Delta_{C(j)} \exp\{-\frac{\tau}{\lambda_{C(j)}} w_j [g_{\lambda_{C(j)}, \gamma}(\beta_j) - g_{\lambda_{C(j)}, \gamma}(\beta_j^{(0)})]\}$

    end for

    for $j = 1, \ldots, p$ and $k = 1, \ldots, p$ do **do** ▷ For all CMEs

      Calculate $v_{j|k+} \leftarrow \frac{1}{n}\mathbf{x}_{j|k+}^T \mathbf{W} \mathbf{x}_{j|k+}$, $z_j \leftarrow \frac{1}{n}\mathbf{x}_{j|k+}^T \mathbf{W r} + v_{j|k+} \beta_{j|k+}^{(0)}$

      Update $\beta_{j|k+}^{(0)} \leftarrow \beta_{j|k+}$, $\beta_{j|k+} \leftarrow S_{\lambda_{S(j)}, \lambda_{C(k)}}(z_{j|k+}, \Delta_{S(j)}, \Delta_{C(k)})$

      $\mathbf{r} \leftarrow \mathbf{r} - \mathbf{x}_{j|k+}(\beta_{j|k+} - \beta_{j|k+}^{(0)})$, $\boldsymbol{\eta} \leftarrow \boldsymbol{\eta} + \mathbf{x}_{j|k+}(\beta_{j|k+} - \beta_{j|k+}^{(0)})$

      $\Delta_{S(j)} \leftarrow \Delta_{S(j)} \exp\{-\frac{\tau}{\lambda_{S(j)}} w_{j|k+} [g_{\lambda_{S(j)}, \gamma}(\beta_{j|k+}) - g_{\lambda_{S(j)}, \gamma}(\beta_{j|k+}^{(0)})]\}$

      $\Delta_{C(k)} \leftarrow \Delta_{C(k)} \exp\{-\frac{\tau}{\lambda_{C(k)}} w_{j|k+} [g_{\lambda_{C(k)}, \gamma}(\beta_{j|k+}) - g_{\lambda_{C(k)}, \gamma}(\beta_{j|k+}^{(0)})]\}$

    end for

  until β converges

  return the converged coefficient vector $\boldsymbol{\beta}$

end function

信噪比的影响

图 A.1 总结了在连续和二元结果、非相关和相关设定，以及主组+表亲效应与主组+兄弟姐妹结构下，随着 CME 效应大小 β_{CME}增大（固定 $\beta_{\mathrm{ME}} = 5$），变量选择和预测性能如何变化；分析遵循第 5.2 节中的情景 2。在连续结果图中，随着 β_{CME} 增大（即比值 $\beta_{\mathrm{ME}}/\beta_{\mathrm{CME}}$减小），主组+表亲效应设计下的 F1 分数和真阳性率会下降，而主组+兄弟姐妹设计下则会提高；两种情况下，自适应 cmenet 的总体 F1 分数始终最高。在 $\rho = 0$ 条件下，精确率最高；在 $\rho = 1/\sqrt{2}$条件下，其精确率与非自适应 cmenet 相当。随着 β_{CME} 增大，均方预测误差也随之上升，因为响应尺度会随信号强度增大；值得注意的是，在各方法中，自适应 cmenet 的均方预测误差增长最慢。在二元结果图中,−随着 β_{CME} 增大，F1 分数或真阳性率也呈现相同的变化模式。在所有效应量下，无论相关性设置如何，其精确率都低于 GEL，但高于其他基线方法。在所有设定下，误分类率均低于或相当于其他方法，表明其样本外预测性能始终强劲。总体而言，较大的 $\beta_{\mathrm{ME}}/\beta_{\mathrm{CME}}$（较小的相对$\beta_{\mathrm{CME}}$）会拉大性能差距，而我们的方法优势最大；当$\beta_{\mathrm{ME}}/\beta_{\mathrm{CME}}$ 较小时，差距会缩小，但我们的方法仍优于其他方法。请注意，在连续且非相关的情形下，当主效应和条件主效应的效应量相近时（例如， $\beta_{\mathrm{ME}} = 5$、 $\beta_{\mathrm{CME}} = 4$），所有方法在所考察的信噪比中表现最差。当主效应和条件主效应的强度大致相同时，模型难以判断应保留哪些项。信号分散在主效应和条件主效应之间，因此每一类单独来看都只有中等强度。结果是，许多效应都接近选择阈值，更容易漏掉真实信号或纳入错误项。由于不存在相关结构来帮助“关联”相关项，模型无法借助特征间的信息来稳定变量选择。因此，尽管总体信号强度并不弱，F1 分数或真阳性率仍会降低，误差度量则会升高。

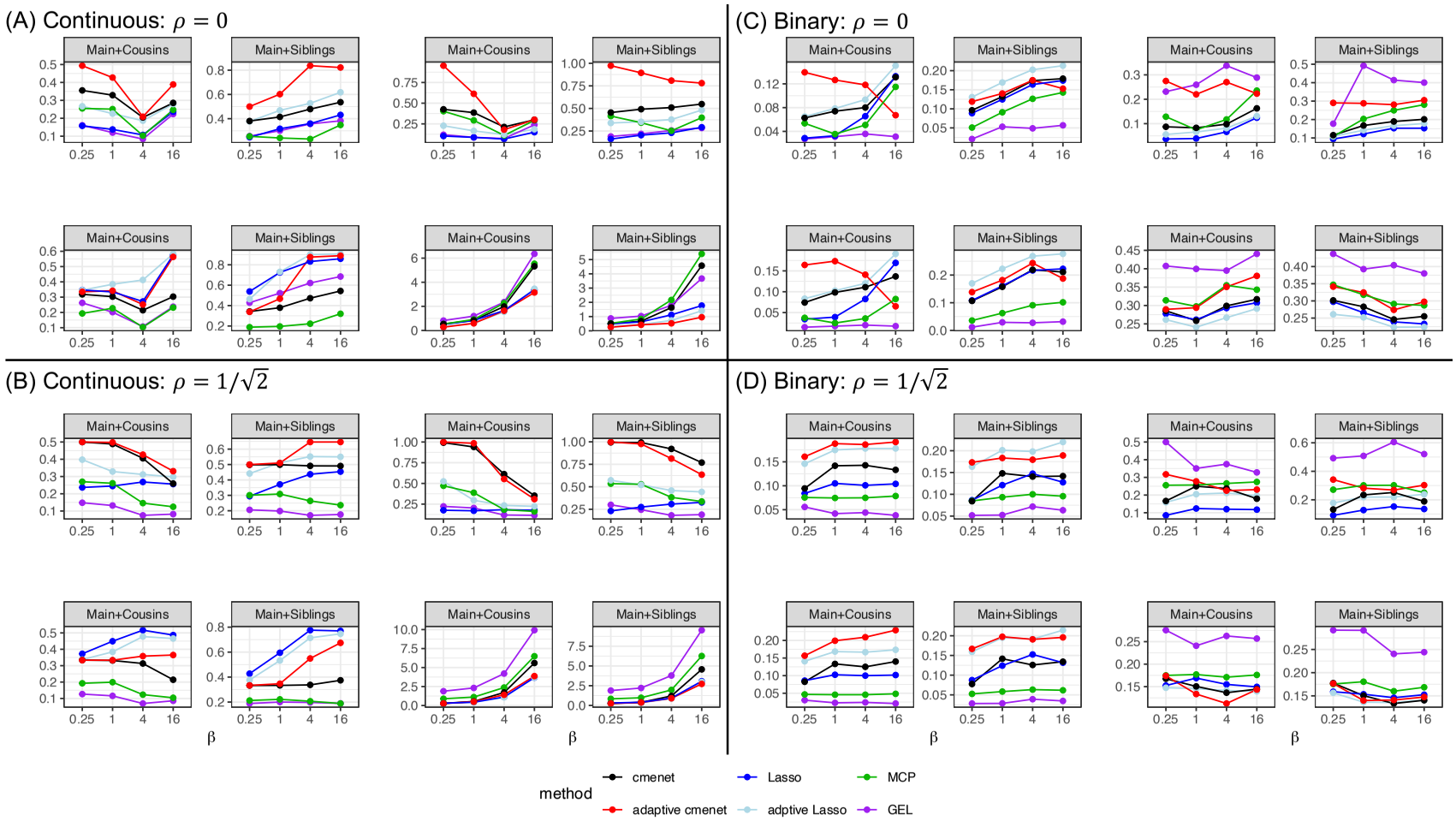


Figure A.1: Sensitivity analysis across effect size β_{CME} (Example 2; $n = 50$, $p = 20$; fixed $\beta_{\text{ME}} = 5$): 4 active groups with 2 active effects in each group. Panels (A)–(D) show: (A) continuous, uncorrelated ($\rho = 0$); (B) continuous, correlated ($\rho = 1/\sqrt{2}$); (C) binary, uncorrelated ($\rho = 0$); and (D) binary, correlated ($\rho = 1/\sqrt{2}$). The x-axis varies β_{CME}; the y-axis reports F_1, Precision, True Positive Rate (TPR), and MSPE (continuous) or Misclassification Rate (binary). Columns correspond to main+cousins (left) and main+siblings (right).