

# 高维线性模型中参数组的显著性检验

史延美

青岛大学, 山东 青岛

收稿日期: 2022年9月10日; 录用日期: 2022年10月2日; 发布日期: 2022年10月10日

## 摘要

高维线性模型中参数的显著性检验是统计学研究的热点。现有的统计方法往往局限于参数的稀疏性假设, 但在实际中, 该假设是很难检验且极易被违反的。本文通过重构回归对参数组进行显著性检验, 首先利用原假设的结构得到重构回归模型, 并建立设计矩阵间的线性相关模型, 进而计算两模型误差的相关性, 以此将待测原假设转化为可检验的矩条件, 并建立检验统计量, 得到检验统计量的渐近分布。最后, 通过模拟试验, 观察该检验方法在不同的误差和设计矩阵设置下的第I类错误概率和势函数曲线。试验结果表明: 对于参数组检验, 矩方法能保证第I类错误概率控制在显著性水平附近, 并且可以证明检验的势特征。

## 关键词

高维线性模型, 重构回归模型, 参数组, 矩条件

# Significance Tests for Parameter Group in High-Dimensional Linear Models

Yanmei Shi

Qingdao University, Qingdao Shandong

Received: Sep. 10<sup>th</sup>, 2022; accepted: Oct. 2<sup>nd</sup>, 2022; published: Oct. 10<sup>th</sup>, 2022

## Abstract

The significance test of parameters in high-dimensional linear models is a hot topic in statistical research. Existing statistical methods are often limited to the assumption of parameter sparsity, but in practice, this assumption is difficult to test and easily violated. In this paper, the significance test of the parameter group is carried out through reconstruction regression. First, the reconstruction regression model is obtained by using the structure of the null hypothesis, and the linear correlation model between the design matrices is established, and then the correlation between

the errors of the two models is calculated. The hypothesis is transformed into a testable moment condition, establish the test statistic, and obtain the asymptotic distribution of the test statistic. Finally, through simulation experiments, the probability of Type I error of the test method and power function curve under different error and design matrix settings are observed. The experimental results show that for the parameter group test, the moment method can ensure that the probability of Type I error is close to the significance level, and can prove the power property of the test.

## Keywords

High-Dimensional Linear Models, Reconstruction Regression Model, Parameter Group, Moment Condition

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

信息技术的快速发展造成了数据收集和数据分析的重大变革。现如今，不同领域的科学研究越来越依赖高维观测数据，即数据维度  $p$  远大于样本容量  $n$ 。大量高维数据的建模、推断和预测逐渐成为统计学研究的热点。另外，由于数据维度和科技水平的限制，现有的方法通常假设参数的稀疏性，即参数向量中非零元素的个数  $s$  满足：当  $n \rightarrow \infty$  时， $s/n \rightarrow 0$  [1]。常见的惩罚方法包括 Lasso [2]、Dantzig Selector [3]、square root Lasso [4] 和岭估计 [5]。这些方法已经被广泛研究，而且惩罚方法已经应用于诸多领域，例如生物成像、疾病跟踪、政策制定及营销策略等。

假设检验在统计理论和应用中扮演着重要的作用。假设检验的现有方法往往局限在参数的超稀疏领域，即  $s = o(\sqrt{n}/\log p)$ 。[6] 中研究了具有随机设计的高维线性回归模型的置信区间，并分析了置信区间的期望长度及其在不同稀疏水平下的自适应性；[7] 中提出了一种偏差校正方法，证明了低维投影是一种构造置信区间和假设检验的有效方法；[8] 在稀疏水平  $s = \frac{n}{\log^2 p}$  下，提出了一种纠偏 Lasso 估计，并证明

了其渐近正态性；[9] 中将参数向量的单分量或低维分量估计推广到广义线性模型；[10] 在稀疏模型下提供了有效控制第 I 类错误概率的方法；[11] 在高维线性模型的框架下构造了去偏估计量，并建立其渐近正态性；[12] 考虑了高维稀疏逻辑回归模型的预测，通过方差 - 偏差均衡思想，提出了一种新的偏差校正估计方法；[13] 考虑了单样本及双样本回归条件下回归系数的全局检验和多重检验，利用偏差校正的广义低维投影构造了检验全局参数的检验统计量，并导出其渐近原分布；[14] 定义了一种新的去偏阈值岭回归方法，并证明了一致性结果和高斯近似定理；[15] 基于  $L^2$  范数建立检验统计量，用于检验高维平均向量是否等于 0，并表明在正则性条件下，其渐近原分布与  $\chi^2$  混合的极限分布相同。以上推理方法都假设了模型的稀疏性，但是现有的推理方法对稀疏性假设很敏感，一旦违背该假设，可能会失去对第 I 类错误概率的控制。为了摆脱稀疏性假设，[16] 提出了近似稀疏的 CorrT 检验，通过重构回归将原假设转化为可检验的矩条件；[17] 将矩方法推广到全高斯设计下回归参数线性泛函的假设检验。

上述文献考虑的是单参数分量或全局参数，但据我们所知，目前对非稀疏参数组的假设检验研究甚少。在全基因组基因表达谱研究领域， $\beta \in \mathfrak{R}^p$  代表整个基因组，但是我们感兴趣的往往是与某一性状有关的基因组，因此参数组的假设检验问题具有非常重要的应用价值。设  $\beta_G \subseteq \beta$  是我们感兴趣的参数组，

则本文要研究以下假设检验问题:

$$H_0 : \beta_G = \beta_G^0 \quad \text{v.s.} \quad H_1 : \beta_G \neq \beta_G^0 .$$

结合上述文献中的方法和应用场景, 根据原假设结构, 首先重构原始线性模型, 然后建立两个设计矩阵的相关模型, 进而计算两模型误差间的相关关系, 以此将原假设转化为可检验的矩条件, 估计矩条件中的未知参数, 并构造检验统计量, 推导其渐近分布。此过程在不施加任何稀疏性假设的情况下, 保证了检验的第 I 类和第 II 类错误概率。

本文的剩余部分组织如下: 在第 2 节中, 我们利用矩方法对参数组进行显著性检验, 将原假设转换为可检验的矩条件, 并提出检验统计量及渐近分布; 在第 3 节中, 我们通过模拟试验验证了所提检验方法在不同的设计矩阵和误差设置下的第 I 类和第 II 类错误概率; 在第 4 节中, 进行了总结和讨论。

## 2. 参数组的显著性检验

在本节中, 我们提出了高维线性模型中参数组的显著性检验方法。首先, 利用原假设的结构, 重构高维线性回归模型, 并建立该模型中的两组设计矩阵的线性相关模型, 计算重构回归模型及线性相关模型中的两误差的矩条件, 以此将原假设转化为可检验的矩条件; 然后, 利用 Lasso 方法估计矩条件中的未知参数; 最后, 建立检验统计量, 并推导其渐近分布, 并利用模拟的方法得到检验的临界值。

### 2.1. 符号说明

我们首先定义下面的符号。对于一个向量  $v \in \mathfrak{R}^k$ ,  $\|v\|_\infty = \max_{1 \leq i \leq k} |v_i|$ ,  $\|v\|_0 = \sum_{i=1}^k I(v_i \neq 0)$ , 其中  $I(\cdot)$  表示示性函数。 $v_G$  表示由  $v_j, j \in G$  组成的子向量,  $v_{-G}$  表示由  $v_j, j \in G^c$  组成的子向量。对于矩阵  $A$ ,  $a_i$  表示该矩阵的第  $i$  行,  $A_j$  表示其第  $j$  列。 $A_G$  表示由向量  $A_j, j \in G$  构成的子矩阵,  $A_{-G}$  表示由向量  $A_j, j \in G^c$  构成的子矩阵,  $\|A\|_\infty = \max |A_{i,j}|$ 。对于两个序列  $a_n, b_n > 0$ ,  $a_n \approx b_n$  表示存在常数  $C_1, C_2 > 0$ , 使得对于任意  $n$ ,  $a_n \leq C_1 b_n$  且  $b_n \leq C_2 a_n$ 。

### 2.2. 重构回归模型

对于高维线性模型:

$$y_i = x_i^T \beta + \varepsilon_i, i = 1, 2, \dots, n, \tag{1}$$

其中,  $y_i$  是响应变量,  $x_i \in \mathfrak{R}^p$  是解释向量, 其满足  $x_i \sim N(0, \Sigma)$ , 其中  $\Sigma$  是一个未知协方差矩阵。误差向量  $\varepsilon \in \mathfrak{R}^n$  和设计矩阵  $X = (x_1^T, x_2^T, \dots, x_n^T)^T$  是不相关的, 且满足  $E(\varepsilon_i) = 0$ ,  $\text{Var}(\varepsilon_i) = \sigma_\varepsilon^2$ ,  $0 < \sigma_\varepsilon < \infty$ 。参数向量  $\beta \in \mathfrak{R}^p$  是未知的, 在上述模型中, 我们允许  $p \gg n$ 。

本文的目的是对参数组进行假设检验, 即:

$$H_0 : \beta_G = \beta_G^0 \quad \text{v.s.} \quad H_1 : \beta_G \neq \beta_G^0, \tag{2}$$

其中, 子向量  $G \subseteq \{1, 2, \dots, p\}$ , 且  $\beta_G \in \mathfrak{R}^k$  是我们感兴趣的参数组,  $\beta_G^0$  是给定参数组。根据待测原假设, 我们将参数向量分解为  $\beta = (\beta_G^T, \beta_{-G}^T)^T \in \mathfrak{R}^p$ , 此时可将模型(1)重写为:

$$y = z^T \beta_G + w^T \beta_{-G} + \varepsilon, \tag{3}$$

其中,  $z = x_G \in \mathfrak{R}^k$ ,  $w = x_{-G} \in \mathfrak{R}^{p-k}$ 。

为了检验原假设, (2)式两端同时减  $z^T \beta_G^0$ , 得到

$$y - z^T \beta_G^0 = z^T (\beta_G - \beta_G^0) + w^T \beta_{-G} + \varepsilon .$$

上式中, 包含原假设  $H_0 : \beta_G - \beta_G^0 = 0$ 。令  $v = y - z^T \beta_G^0$ ,  $e = z^T (\beta_G - \beta_G^0) + \varepsilon$ , 得到如下重构回归模型:

$$v = w^T \beta_{-G} + e. \tag{4}$$

在此模型中，当原假设成立时， $E(ew) = E(\varepsilon w) = E(\varepsilon)E(w) = 0$ 。否则，

$$E(ew) = E\left[\left(z^T(\beta_G - \beta_G^0) + \varepsilon\right)w\right] = E\left[z^T(\beta_G - \beta_G^0)w\right] = (\beta_G - \beta_G^0)^T E(zw^T).$$

由于  $z$  和  $w$  之间可能存在线性相关性，故  $E(zw^T)$  不一定为 0。下面我们建立  $z$  和  $w$  之间线性相关模型：

$$z = \Pi^T w + u, \tag{5}$$

其中， $\Pi \in \mathfrak{R}^{(p-k) \times k}$  是相关系数矩阵， $u \in \mathfrak{R}^k$  是误差项，且本文假设  $u$  和  $w$  是独立的，且  $u$  服从高斯分布，满足  $E(u) = 0$ 。在本文中，我们合理假设  $\Pi$  的列稀疏性，实际上， $\Pi$  的列稀疏性假设表明了精度矩阵  $\Omega_X = \Sigma_X^{-1}$  的稀疏性。

考虑(4)中的伪误差  $e$  和(5)中的误差  $u$  之间的相关性，建立以下矩条件：

$$E(ue) = E\left[u\left(z^T(\beta_G - \beta_G^0) + \varepsilon\right)\right] = E(uu^T)(\beta_G - \beta_G^0), \tag{6}$$

在(2)中的原假设下， $E(ue) = 0$ ；否则， $E(ue) \neq 0$ 。因此，(2)中的检验问题等价于检验：

$$H_0 : E\left[(z - \Pi^T w)(v - w^T \beta_{-G})\right] = 0 \quad \text{v.s.} \quad H_1 : E\left[(z - \Pi^T w)(v - w^T \beta_{-G})\right] \neq 0, \tag{7}$$

其中， $\Pi$  和  $\beta_{-G}$  是未知的待估参数。根据  $\Pi$  的列稀疏性，容易获得其一一致估计  $\hat{\Pi}$ ，此时，对于  $\beta_{-G}$  的任意估计  $\tilde{\beta}_{-G}$ ，我们有：

$$E\left[(z - \hat{\Pi}^T w)(v - w^T \tilde{\beta}_{-G})\right] \rightarrow E\left[(z - \Pi^T w)(v - w^T \tilde{\beta}_{-G})\right] = E(ue).$$

因此，上述内积结构不依赖对  $\beta_{-G}$  的良好估计的需要。

### 2.3. 参数估计

(7)中含有未知参数  $\Pi$  和  $\beta_{-G}$ ，我们建立它们的 Lasso 估计。对于设计矩阵  $X = (x_1^T, x_2^T, \dots, x_n^T)^T \in \mathfrak{R}^{n \times p}$ ，我们定义新的响应向量  $V = Y - Z\beta_G^0 \in \mathfrak{R}^n$ ，新的设计矩阵  $Z = X_G \in \mathfrak{R}^k$ ， $W = X_{-G} \in \mathfrak{R}^{p-k}$ ，且  $e = Z(\beta_G - \beta_G^0) + \varepsilon = (e_1, e_2, \dots, e_n)^T \in \mathfrak{R}^n$ 。 $\beta_{-G}$  的 Lasso 估计为：

$$\hat{\beta}_{-G} = \arg \min_{\beta_{-G} \in \mathfrak{R}^{p-k}} \left\{ \|V - W\beta_{-G}\|_2^2 + \lambda_1 \|\beta_{-G}\|_1 \right\}, \tag{8}$$

其中，参数  $\lambda_1$  满足  $\lambda_1 \approx \sqrt{n^{-1} \log p}$ 。 $\Pi_j$  的 Lasso 估计为：

$$\hat{\Pi}_j = \arg \min_{\Pi_j \in \mathfrak{R}^{p-5}} \left\{ \|Z_j - W\Pi_j\|_2^2 + \lambda_2 \|\Pi_j\|_1 \right\}, j \in G, \tag{9}$$

其中， $\lambda_2 \approx \sqrt{n^{-1} \log p}$ 。 $\Pi$  的 Lasso 估计为：

$$\hat{\Pi} = (\hat{\Pi}_j, j \in G) \in \mathfrak{R}^{(p-k) \times k}. \tag{10}$$

### 2.4. 检验统计量

在本小节中，对于(7)中的假设问题，我们用  $\hat{\beta}_{-G}$  和  $\hat{\Pi}$  构造检验统计量，并用模拟的方法获得检验的临界值。

通过插入(8)和(10)中的估计值，我们构造如下检验统计量：

$$T_n = n^{-1/2} \hat{\sigma}_e^{-1} \left\| (Z - W\hat{\Pi})^T (V - W\hat{\beta}_{-G}) \right\|_{\infty}, \tag{11}$$

其中,  $\hat{\sigma}_e = \frac{\|V - W\hat{\beta}_{-G}\|_2}{\sqrt{n}}$ 。当原假设成立时,  $T_n$  的值接近为 0; 反之,  $T_n$  的值较大。因此, 当  $T_n$  的值大于某临界值时, 应该拒绝原假设。

在(7)中的原假设下,

$$n^{-1/2}\hat{\sigma}_e^{-1}(Z - W\hat{\Pi})^T(V - W\hat{\beta}_{-G}) = n^{-1/2}\hat{\sigma}_e^{-1}(\Pi - \hat{\Pi})^T W^T \hat{e} + n^{-1/2}\hat{\sigma}_e^{-1}U^T \hat{e}。$$

注意, 我们构造的  $U$  独立于  $(V, W)$ , 但是  $\hat{\beta}_{-G}$  只与  $(V, W)$  有关, 因此  $\hat{e} = V - W\hat{\beta}_{-G}$  也完全依赖于  $(V, W)$ , 故  $U$  和  $\hat{e}$  是不相关的。设  $\delta_n$  是一个由  $(V, W)$  生成的  $\sigma$ -代数, 则

$$E\left(n^{-1/2}\hat{\sigma}_e^{-1}(\Pi - \hat{\Pi})^T W^T \hat{e} + n^{-1/2}\hat{\sigma}_e^{-1}U^T \hat{e} \mid \delta_n\right) \rightarrow 0,$$

$$\text{Var}\left(n^{-1/2}\hat{\sigma}_e^{-1}U^T \hat{e}\right) = n^{-1}\hat{\sigma}_e^{-2}\text{Var}\left(\sum_{i=1}^n \hat{U}_{i,G}^T \hat{e}_i\right) = Q,$$

其中,  $Q$  是一个分块对角矩阵, 其中每一个对角块为  $q = n^{-1} \sum_{i=1}^n \hat{\sigma}_e^{-2} \hat{e}_i^2 E(\hat{u}_i \hat{u}_i^T) = E(\hat{u}_1 \hat{u}_1^T) = n^{-1} \sum_{i=1}^n \hat{u}_i \hat{u}_i^T$ , 其中  $\hat{u}_i^T$  是  $\hat{U}$  的第  $i$  行。因此, 可以得到在  $\delta_n$  下,  $\left\|n^{-1} \sum_{i=1}^n u_i \hat{e}_i \hat{\sigma}_e^{-1}\right\|_\infty$  与  $\|\xi\|_\infty$  的分布渐近相同, 其中  $\xi \sim N(0, q)$ 。我们引入函数  $\Gamma(x, A) := P(\|\xi\|_\infty \leq x)$ , 其中  $\xi \sim N(0, A)$ , 则  $\Gamma(x, q)$  与  $T_n$  的分布渐近相等, 而且  $\Gamma(x, q)$  可以通过模拟实现。

### 3. 仿真试验

我们将通过仿真试验展示所提方法的有限样本属性。不失一般性, 我们检验  $\beta$  的前 5 个分量:

$$H_0: \beta_G = \beta_G^0,$$

其中  $G = \{1, 2, 3, 4, 5\}$ 。

在所有的模拟试验中, 我们设置  $n = 200$ ,  $p = 500$ , 检验的名义水平为  $\alpha = 0.05$ 。拒绝概率是基于 100 次重复试验得到的, 并设置参数  $\lambda_1 = \lambda_2 = 0.5\sqrt{n^{-1} \log p}$ , 这是一个常用的选择, 我们在以下的模拟中也证明了该选择可提供良好的结果。

#### 3.1. 试验设置

我们在以下 3 种不同高斯设计下证明试验结果:

在这种情况下, 我们考虑标准的 Toeplitz 设计矩阵, 即  $X$  各行独立同分布地取自  $N(0, \Sigma_1)$ , 其中  $(\Sigma_1)_{i,j} = 0.4^{|i-j|}$ 。

在此情况下, 我们考虑不相关设计矩阵, 其中  $X$  各行独立同分布地取自  $N(0, \Sigma_2)$ , 且

$$\Sigma_2 = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}。$$

在此情况下, 我们考虑等相关设计矩阵, 即  $X$  各行独立同分布地取自  $N(0, \Sigma_3)$ , 且

$$\Sigma_3 = \begin{cases} 1, & i = j \\ 0.4, & i \neq j \end{cases}。$$

除此之外, 我们考虑两种不同的误差分布:

轻尾误差分布, 在此情况下, 误差  $\varepsilon$  取自标准正态分布;

重尾分布, 在此情况下,  $\varepsilon$  取自自由度为 3 的 t-分布。

设  $s = \|\beta\|_0$  表示模型稀疏性, 即  $\beta$  中非零元素的个数, 并设置从  $s = 3$  (超稀疏) 到  $s = p$  (稠密)。对于稀疏度  $s$ , 我们设置模型参数为:

$$\beta_j = \begin{cases} \frac{3}{\sqrt{s}}, & 1 \leq j \leq s \\ 0, & j > s \end{cases}.$$

为了说明本文所示的矩方法的鲁棒性, 我们还将我们的矩方法与[9]中提出的用纠偏 Lasso 估计构造 Wald 检验的方法(WL)进行比较。矩方法的第 I 类错误概率计算如下:

$$P = \frac{\text{num}\left(I\left(T_n > \Gamma^{-1}(1-\alpha, q) \mid H_0\right)\right)}{\text{Num}},$$

即在所有的试验中, 原假设成立时拒绝原假设的试验所占比例, 其中  $\text{Num} = 100$ 。WL 的具体实现过程为: 首先计算参数  $\beta$  的纠偏估计  $\hat{\beta}_{debias}$  以及精度矩阵的 nodewise Lasso 估计  $\hat{\Theta}_{Lasso}$ , 当且仅当

$$\max_{j \in G} \frac{\sqrt{n} \left| \hat{\beta}_{debias, j} - \beta_j^0 \right|}{\sigma_\varepsilon \sqrt{\hat{\Omega}_{j, j}}} > \Psi^{-1}(1-\alpha),$$

拒绝原假设, 其中,  $\Psi = \max_{j \in G} \frac{|W_j|}{\sigma_\varepsilon \sqrt{\hat{\Omega}_{j, j}}}$ , 且  $W = \hat{\Theta}_{Lasso} X^T \varepsilon / \sqrt{n} \sim N(0, \sigma_\varepsilon^2 \hat{\Omega})$ ,  $\hat{\Omega} = \hat{\Theta} \hat{\Sigma} \hat{\Theta}^T$ ,  $\hat{\Sigma} = X^T X / n$ 。

且 Wald 检验的第 I 类错误概率计算如下:

$$p = \text{num} \left( I \left( \max_{j \in G} \frac{\sqrt{n} \left| \hat{\beta}_{debias, j} - \beta_j^0 \right|}{\sigma_\varepsilon \sqrt{\hat{\Omega}_{j, j}}} > \Psi^{-1}(1-\alpha) \mid H_0 \right) \right) / \text{Num}.$$

### 3.2. 试验结果及分析

#### 3.2.1. 第 I 类错误概率

我们将轻尾误差下的第 I 类错误结果收集在表 1 中, 其中包括轻尾误差联合 Toeplitz 设计矩阵, 轻尾误差联合不相关设计矩阵, 该表的每一行代表  $\beta$  的不同稀疏性水平, 在以上设置下, 我们考虑矩检验和 Wald 检验的第 I 类错误。

**Table 1.** The Type I errors of moment method and Wald test under light-tailed error settings  
**表 1.** 轻尾误差设置下矩方法和 Wald 检验的第 I 类错误

方法	轻尾误差 + Toeplitz 设计		轻尾误差 + 不相关设计	
	矩方法	Wald	矩方法	Wald
$s = 3$	0.02	0.23	0.03	0.04
$s = 5$	0.04	0.40	0.02	0.06
$s = 10$	0.07	0.56	0.04	0.19
$s = 20$	0.04	0.72	0.04	0.58
$s = 50$	0.06	0.79	0.09	0.90
$s = 100$	0.03	0.92	0.07	0.86
$s = n$	0.08	0.89	0.06	0.87
$s = p$	0.04	0.83	0.07	0.75



从上表容易看出，在表 1 的两种设置下，矩方法在不同的稀疏水平下的第 I 类错误概率始终维持在名义水平 0.05 左右，即使  $s = p$ ，即稠密模型，矩方法的第 I 类错误概率仍然保持稳定，这说明矩方法对设计矩阵和模型稀疏性具有鲁棒性。相反，对于 Wald 检验，在超稀疏水平( $s = 3$  和  $s = 5$ )上，轻尾误差分布联合不相关矩阵设置下的第 I 类错误概率可以控制在 0.05 左右，但是轻尾误差分布联合 Toeplitz 设计下的第 I 类错误概率远高于名义水平 0.05，且随着稀疏水平不断增加，检验的第 I 类错误概率不断提高，甚至达到 0.9 左右，这说明 Wald 检验对于设计矩阵设置不具备鲁棒性，对于稀疏模型，无法控制第 I 类错误概率，而对于稠密模型( $s = p$ )，第 I 类错误概率高达 0.9 左右，检验完全失效。

我们将重尾误差下的第 I 类错误概率结果收集在表 2 中，其中包括重尾误差联合 Toeplitz 设计矩阵，重尾误差联合不相关设计矩阵及重尾误差联合等相关设计矩阵，该表的每一行代表  $\beta$  的不同稀疏性水平，在以上设置下，我们考虑矩检验和 Wald 检验的第 I 类错误概率。

**Table 2.** The Type I errors of moment method and Wald test under heavy-tailed error settings

**表 2.** 重尾误差设置下矩方法和 Wald 检验的第 I 类错误

方法	重尾误差 + Toeplitz 设计		重尾误差 + 不相关设计		重尾误差 + 等相关设计	
	矩方法	Wald	矩方法	Wald	矩方法	Wald
$s = 3$	0.08	0.99	0.04	1	0.04	1
$s = 5$	0.08	1	0.07	1	0.04	1
$s = 10$	0.04	1	0.08	1	0.03	1
$s = 20$	0.02	1	0.06	1	0.03	1
$s = 50$	0.02	1	0.06	1	0.05	1
$s = 100$	0.07	1	0.02	1	0.04	1
$s = n$	0.09	1	0.03	1	0.06	1
$s = p$	0.03	1	0.03	1	0.09	1

从表 2 可以看出，在重尾误差联合不同的设计矩阵设置下，随着模型的稀疏水平不断增加，矩方法的第 I 类错误概率仍然维持在显著性水平 0.05 左右，这说明对于重尾误差分布，矩方法对设计矩阵和稀疏水平设置仍然具有鲁棒性，第 I 类错误概率始终保持稳定。对于 Wald 检验，即使在稀疏水平下，检验的第 I 类错误概率就高达 0.9 左右，这说明 Wald 检验对于重尾误差分布设置完全失效。

综上，对于矩方法，无论是轻尾误差分布，还是重尾误差分布，矩方法对于设计矩阵及稀疏水平设置都具有鲁棒性，这说明矩方法的第 I 类错误概率具有稳定性。与之相反，Wald 检验在轻尾误差分布下，随着稀疏水平的增加，第 I 类错误概率逐渐提高，并在稠密模型中，该概率高达 0.9；而在重尾误差分布下，即使对于稀疏模型，第 I 类错误概率已高达 0.9，这说明 Wald 检验误差分布不具有鲁棒性，难以控制第 I 类错误概率。

### 3.2.2. 势特征

为了观察检验的势特征，我们通过以下检验问题观察势特征：

$$H_0 : \beta_G = \beta_G^0 \quad \text{v.s.} \quad H_1 : \beta_G = \beta_G^0 + H,$$

其中， $G = \{1, 2, 3, 4, 5\}$ ，且我们设置  $H$  为  $H = (h, 0, 0, 0, 0)$ ， $h$  是一给定常数，其衡量了与原假设的距离。注意， $h = 0$  表示原假设成立。矩方法检验的势计算如下：

$$power = \frac{num\left(I\left(T_n > \Gamma^{-1}\left(1-\alpha, \hat{Q}\right) \mid H_1\right)\right)}{Num},$$

即在所有的试验中，原假设不成立时拒绝原假设的试验所占比例，其中  $Num = 100$ 。WL 的势计算如下：

$$power = num\left(I\left(\max_{j \in G} \frac{\sqrt{n}|\hat{\beta}_{debias,j} - \beta_j^0|}{\sigma_\varepsilon \sqrt{\hat{\Omega}_{j,j}}} > \Psi^{-1}(1-\alpha) \mid H_1\right)\right) / Num,$$

为了验证本文所用的矩方法在稠密模型下参数组检验中的优良性质，我们仅考虑  $s = p$  时的势函数。通过设置不同的  $h$  值，首先得到如下轻尾误差分布下的势函数曲线：

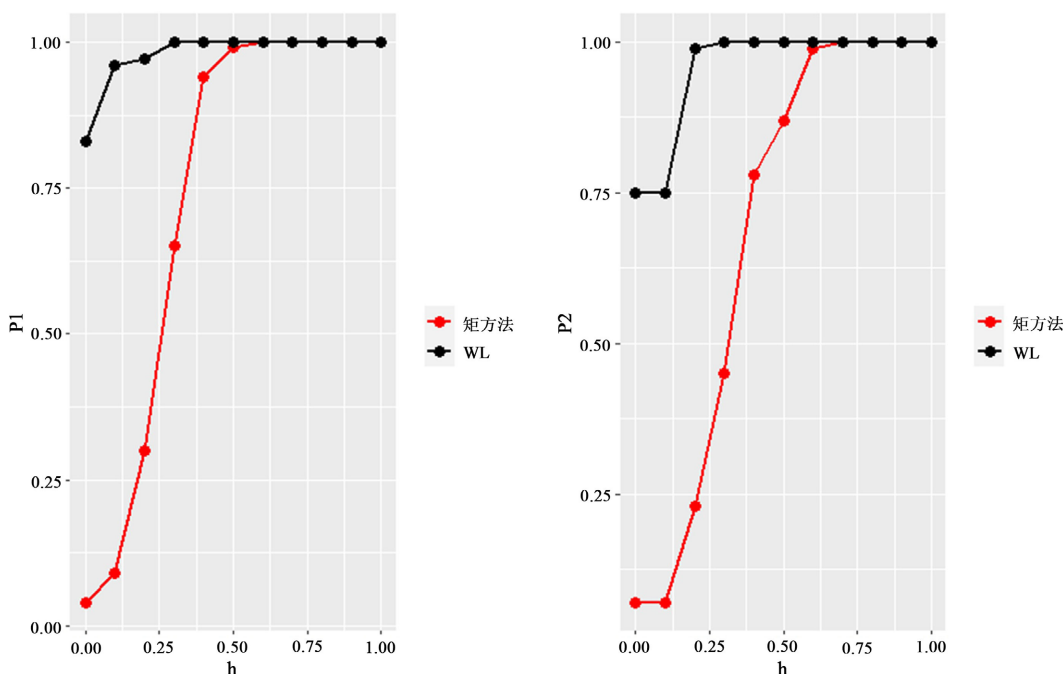


Figure 1. Power function curve under light-tailed error distribution

图 1. 轻尾误差分布下的势函数曲线

图 1 中左侧表示的是轻尾误差分布联合 Toeplitz 设计矩阵，右侧表示轻尾误差分布联合不相关设计矩阵，其中红色的线表示本文所利用的矩方法，黑色的线表示 Wald 检验，最左侧  $h = 0$  的点表示第 I 类错误概率，其余点表示  $h$  取不同值时检验的势。从上图可以看出，在轻尾误差分布设置中，矩方法的第 I 类错误概率均在名义水平 0.05 左右，且以很快的速度达到 1，这说明利用矩方法进行显著性检验时，该检验具有势属性，并且在原假设与备择假设距离较小时，就足以分离原假设和备择假设；反之，对于 Wald 检验，从图中可以看出，其势函数以更快的速度达到 1，但是当  $h = 0$ ，即原假设成立时，该检验的第 I 类错误概率显著地大于名义水平 0.05，且不相关设计下的第 I 类错误概率略大于 Toeplitz 设计矩阵下的第 I 类错误概率。我们将重尾误差分布下的势函数曲线绘制在如下图 2 中。

图 2 中最左侧图表示重尾误差分布联合 Toeplitz 设计矩阵，中间表示重尾误差分布联合不相关设计矩阵，右侧表示重尾误差分布联合等相关设计矩阵，其中红色的线表示矩方法，黑色的线表示 Wald 检验，最左侧  $h = 0$  的点表示第 I 类错误概率，其余点表示  $h$  取不同值时检验的势。从图中可以看出，在 3 种不



同的设计矩阵下,矩检验的第 I 类错误概率均在名义水平 0.05 左右,其势以一定的速度达到 1,其中 Toeplitz 设计矩阵下,矩检验的势达到 1 的速度最快,当  $h=10$  时,检验的势已达到 0.93; 不相关设计矩阵下的矩检验的势达到 1 的速度次之,当  $h=15$  时,检验的势已达到 0.98; 不相关设计矩阵下的矩检验的势达到 1 的速度最慢,当  $h=25$  时,检验的势已达到 0.95。而对于重尾误差分布下的 Wald 检验,其势一直维持在 1,但是当  $h=0$ ,即原假设成立时,该检验的第 I 类错误概率也几乎达到 1,这说明该设置下的 Wald 检验完全失效。

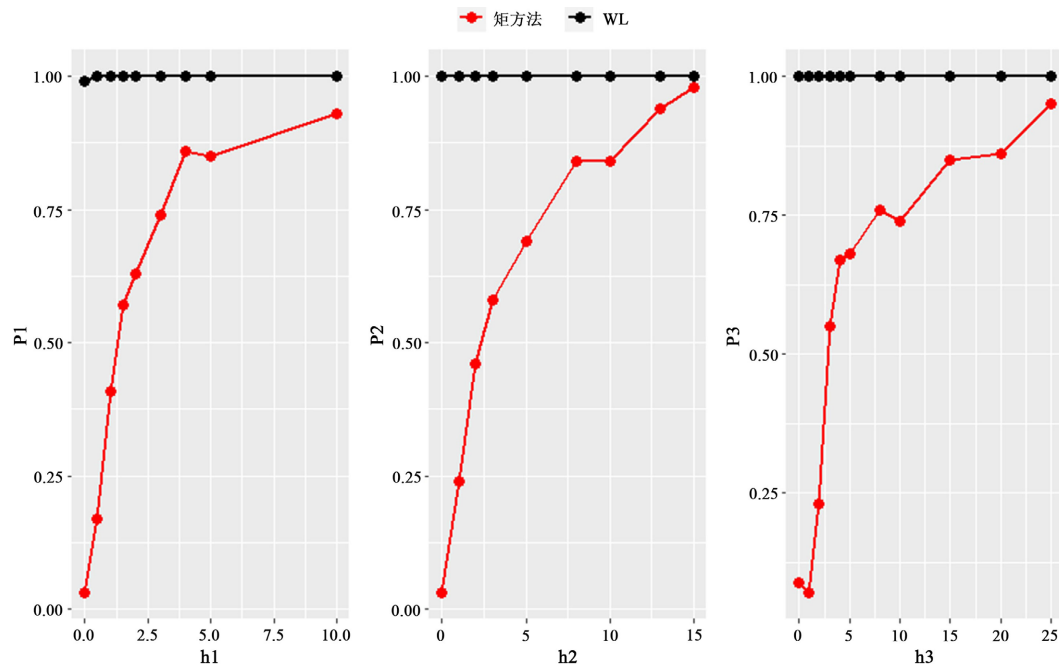


Figure 2. Power function curve under heavy-tailed error distribution  
图 2. 重尾误差分布下的势函数曲线

#### 4. 总结和讨论

在本文中,我们利用矩方法对参数组进行显著性检验。首先根据原假设的结构,重构回归模型,并建立该模型中两个设计矩阵的线性相关模型,通过建立以上两模型中误差的矩条件得到原假设的等价检验,该矩条件的建立放宽了对非稀疏参数良好估计的需求,并且不依赖任何非稀疏假设。仿真试验表明,在不同的误差分布和设计矩阵设置下,矩方法的第 I 类错误概率能稳定维持在显著性水平  $\alpha$  左右,势曲线很快达到 1; 与之相反,由于现有的参数估计方法大多依赖于稀疏性假设,因此对于稠密模型,往往难以控制第 I 类错误概率。除此之外通过建立参数估计的渐近分布的检验,对设计矩阵及误差分布不具有鲁棒性。

#### 参考文献

- [1] Bühlmann, P. and Geer, S. (2015) High-Dimensional Inference in Misspecified Linear Models. *Electronic Journal of Statistics*, **9**, 1449-1473. <https://doi.org/10.1214/15-EJS1041>
- [2] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [3] Candès, E. and Tao, T. (2007) Rejoinder: The Dantzig Selector: Statistical Estimation When  $p$  Is Much Larger than  $n$ . *The Annals of Statistics*, **35**, 2392-2404. <https://doi.org/10.1214/009053607000000532>

- 
- [4] Belloni, A. (2011) Square-Root Lasso: Pivotal Recovery of Sparse Signals via Conic Programming. *Biometrika*, **98**, 791-806. <https://doi.org/10.1093/biomet/asr043>
- [5] Nickl, R. and Geer, S. (2012) Confidence Sets in Sparse Regression. *The Annals of Statistics*, **41**, 2852-2876. <https://doi.org/10.1214/13-AOS1170>
- [6] Cai, T.T. and Guo, Z. (2015) Confidence Intervals for High-Dimensional Linear Regression: Minimax Rates and Adaptivity. *Statistics*, **45**, 615-646. <https://doi.org/10.1214/16-AOS1461>
- [7] Zhang, C.-H. and Zhang, S.S. (2014) Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models. *Journal of the Royal Statistical Society*, **76**, 217-242.
- [8] Javanmard, A. and Montanari, A. (2015) De-Biasing the Lasso: Optimal Sample Size for Gaussian Designs. *The Annals of Statistics*, **46**, 1997-2017. <https://doi.org/10.1214/17-AOS1630>
- [9] Sara, V., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014) On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models. *Annals of Statistics*, **42**, 1166-1202. <https://doi.org/10.1214/14-AOS1221>
- [10] Stadler, N. and Mukherjee, S. (2017) Two-Sample Testing in High Dimensions. *Journal of the Royal Statistical Society, Series B. Statistical Methodology*, **79**, 225-246. <https://doi.org/10.1111/rssb.12173>
- [11] Cai, T., Cai, T. and Guo, Z. (2019) Optimal Statistical Inference for Individualized Treatment Effects in High-Dimensional Models.
- [12] Guo, Z., Rakshit, P., Herman, D.S. and Chen, J. (2020) Inference for the Case Probability in High-Dimensional Logistic Regression.
- [13] Li, H.Z., et al. (2021) Global and Simultaneous Hypothesis Testing for High-Dimensional Logistic Regression Models. *Journal of the American Statistical Association*, **116**, 984-998. <https://doi.org/10.1080/01621459.2019.1699421>
- [14] Zhang, Y. and Politis, D.N. (2020) Ridge Regression Revisited: Debiasing, Thresholding and Bootstrap.
- [15] Zhang, J.T., Zhou, B. and Guo, J. (2022) Testing High-Dimensional Mean Vector with Applications. *Statistical Papers*, **63**, 1105-1137. <https://doi.org/10.1007/s00362-021-01270-z>
- [16] Zhu, Y. and Bradic, J. (2017) Significance Testing in Non-Sparse High-Dimensional Linear Models. *Electronic Journal of Statistics*, **12**, 3312-3364. <https://doi.org/10.1214/18-EJS1443>
- [17] Zhu, Y. and Bradic, J. (2016) Linear Hypothesis Testing in Dense High-Dimensional Linear Models. *Journal of the American Statistical Association*, **113**, 1583-1600. <https://doi.org/10.1080/01621459.2017.1356319>