

贝叶斯Lasso并行中介模型

张沥今

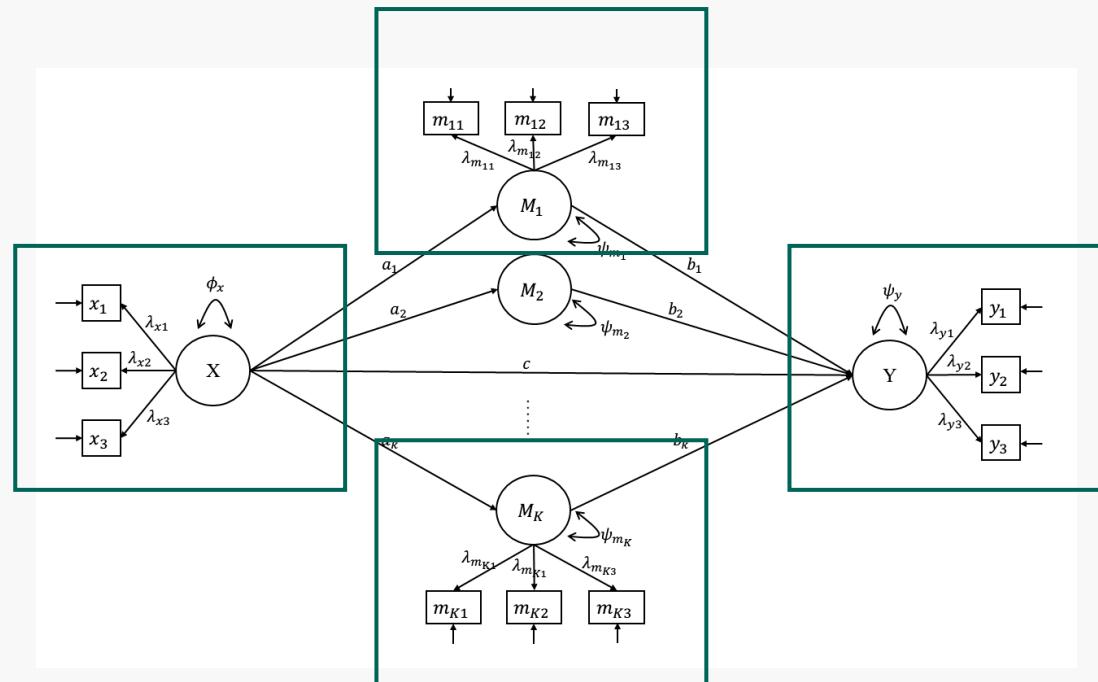
lijinzhang@stanford.edu

潘俊豪

panjunh@sysu.edu.cn

结构方程模型

- 可观测变量、潜变量
- 测量模型：估计潜变量水平的同时考虑了测量误差的存在
- $y_i = \mu + \Lambda\omega_i + \varepsilon_i$



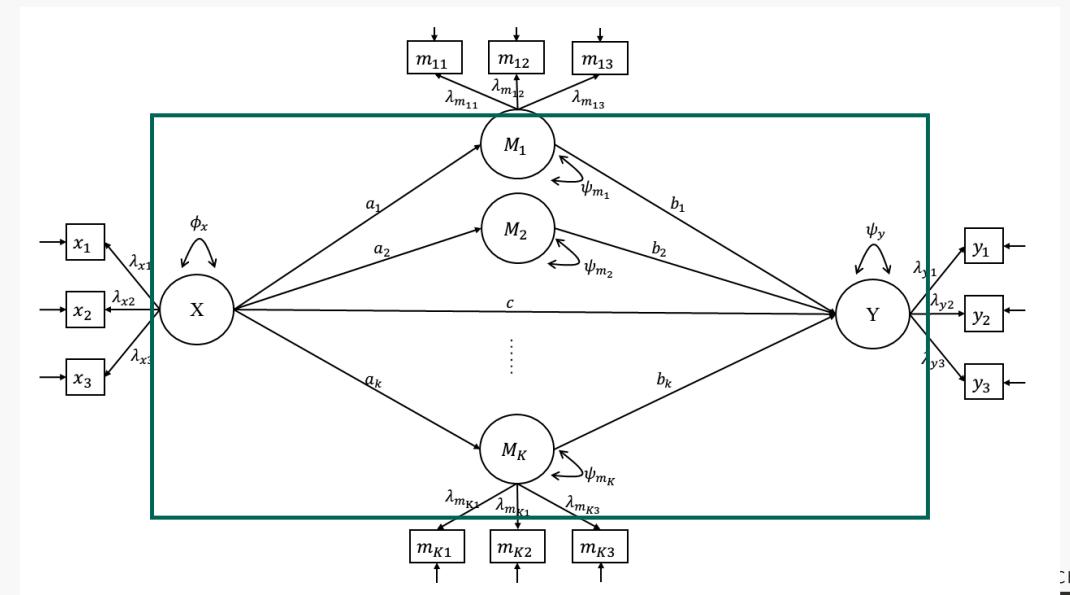
结构方程模型

- **结构模型**：检验研究中感兴趣的变量间的关系
- 并行中介模型 (Multiple Mediator Model)

$$Y_i = \mu_y + \sum_{k=1}^K b_k M_k + \epsilon_y$$

$$M_{ki} = \mu_{m_k} + a_k X_i + \epsilon_{m_k}$$

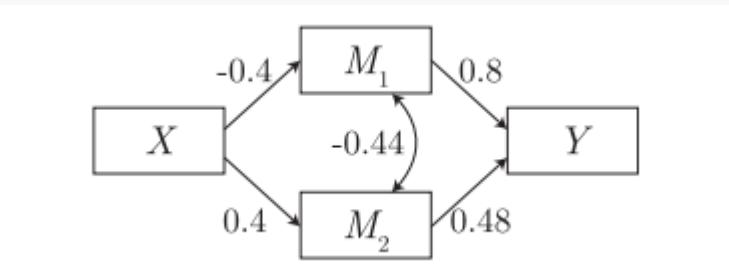
- 理解变量间的作用机制
- 对人们的行为做出有效的预测
- 探究在控制了其它中介变量的情况下，某一中介变量能够发挥的中介效应



過濾法

過濾法 (Filter Method, Preacher & Hayes, 2008)

- 遮掩中介效应



$$\text{cov}(M_2, Y) = 0.48 + 0.8 * \text{cov}(M_1, M_2) = 0$$

传统的中介变量选择方法

1

Sobel法 (Sobel, 1982)

$$\hat{a}_k * \hat{b}_k \text{ or } \hat{a}\hat{b}_k$$

假设 $\hat{a}\hat{b}_k$ 服从正态分布

2

$$[\theta_{2.5}, \theta_{97.5}]$$

3

偏差校正的Bootstrap法 (Efron & Tibshirani, 1993)

根据 $\hat{a}\hat{b}_k$ 与 θ_{50} 的距离调整区间边界

$$\widehat{z}_0 = \Phi^{-1}(q)$$

$$[\theta_{100*\Phi(2*\widehat{z}_0-z\frac{\alpha}{2})}, \theta_{100*\Phi(2*\widehat{z}_0+z\frac{\alpha}{2})}]$$

其中， q 指Bootstrap估计值 θ_i 小于 $\hat{a}\hat{b}_k$ 的比例

Lasso 并行中介模型

(Serang et al., 2017, 2020)

无法提供区间估计

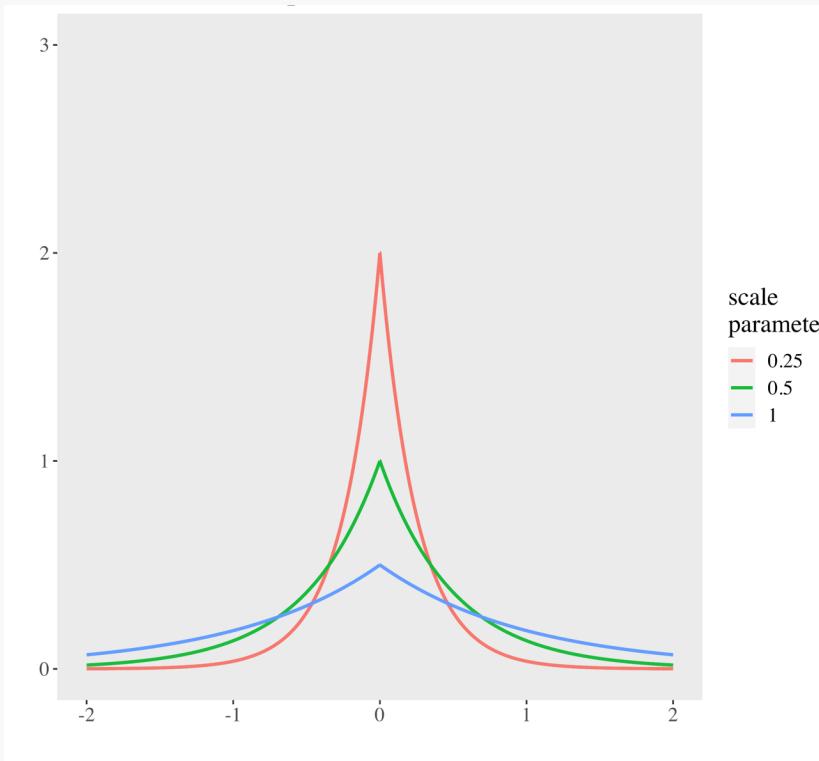
尚未扩展到潜变量模型

百分位Bootstrap法

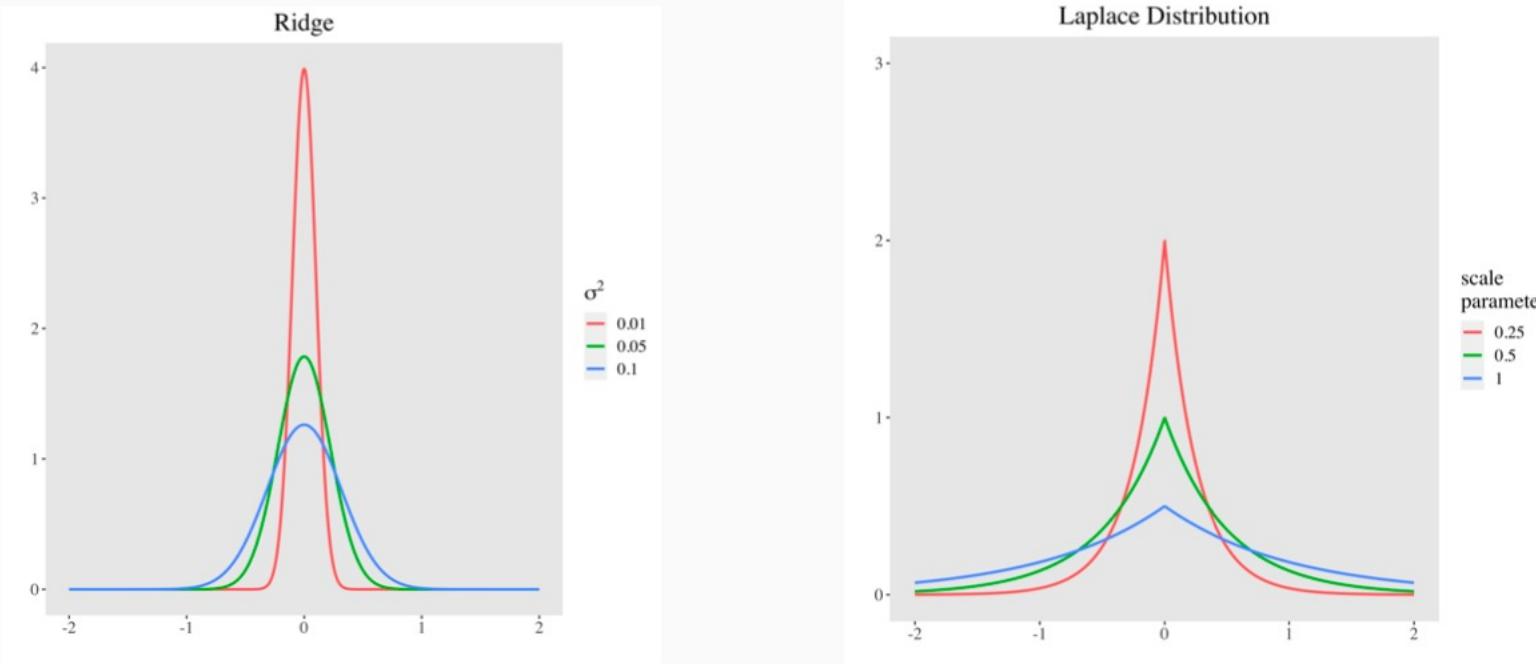
正则化方法

		频率学派	贝叶斯
标准误估计	难以获得，难以计算p值		
区间估计	Bootstrap法	基于MCMC (Markov Chain Monte Carlo) 算法	
估计惩罚参数 λ	模型比较 交叉验证	易于直接获得	
小样本	无法识别参数数目大于样本数 目的模型	表现优良	

贝叶斯Lasso

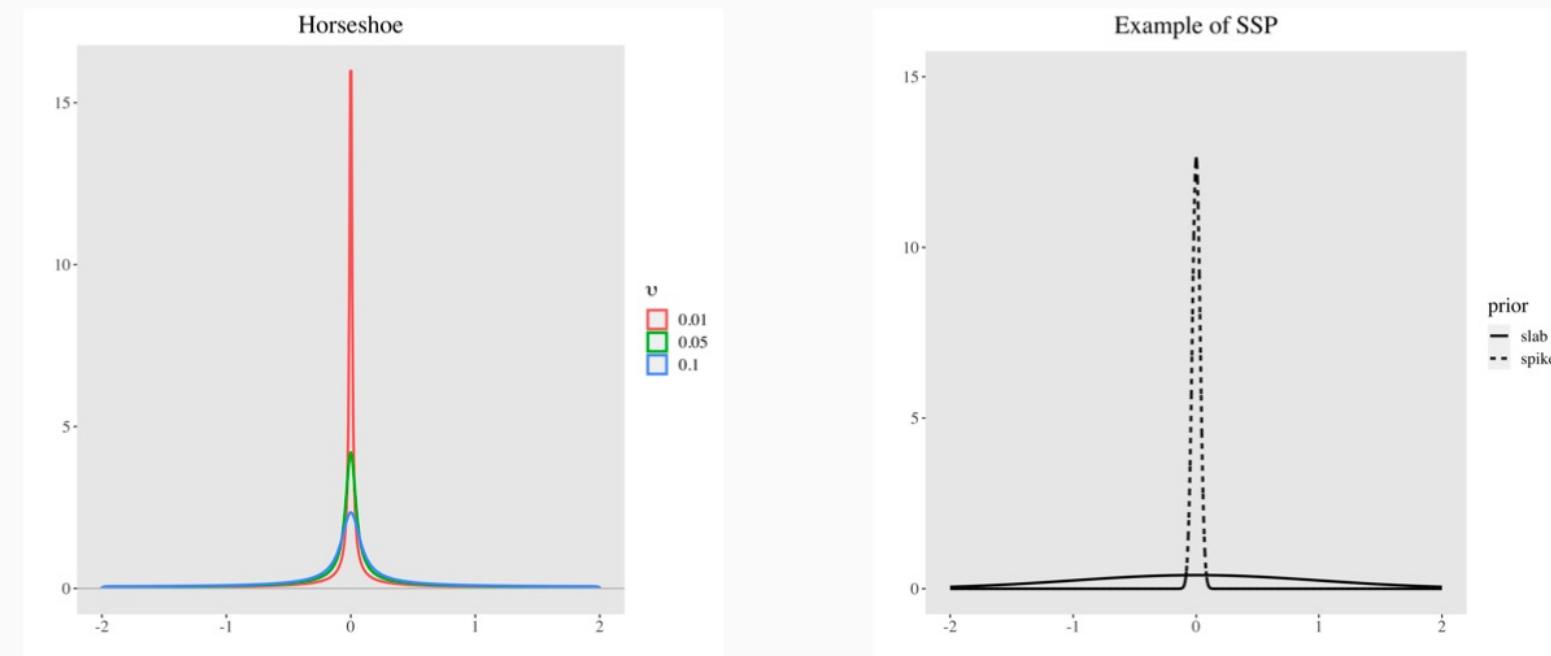


其他正则化方法



- Ridge: global shrinkage
- Lasso (least absolute shrinkage and selection operator; Park & Casella, 2008; Tibshirani, 1996): global shrinkage with Laplace prior
- Adaptive lasso (lasso; Zou, 2006): local shrinkage

其他正则化方法

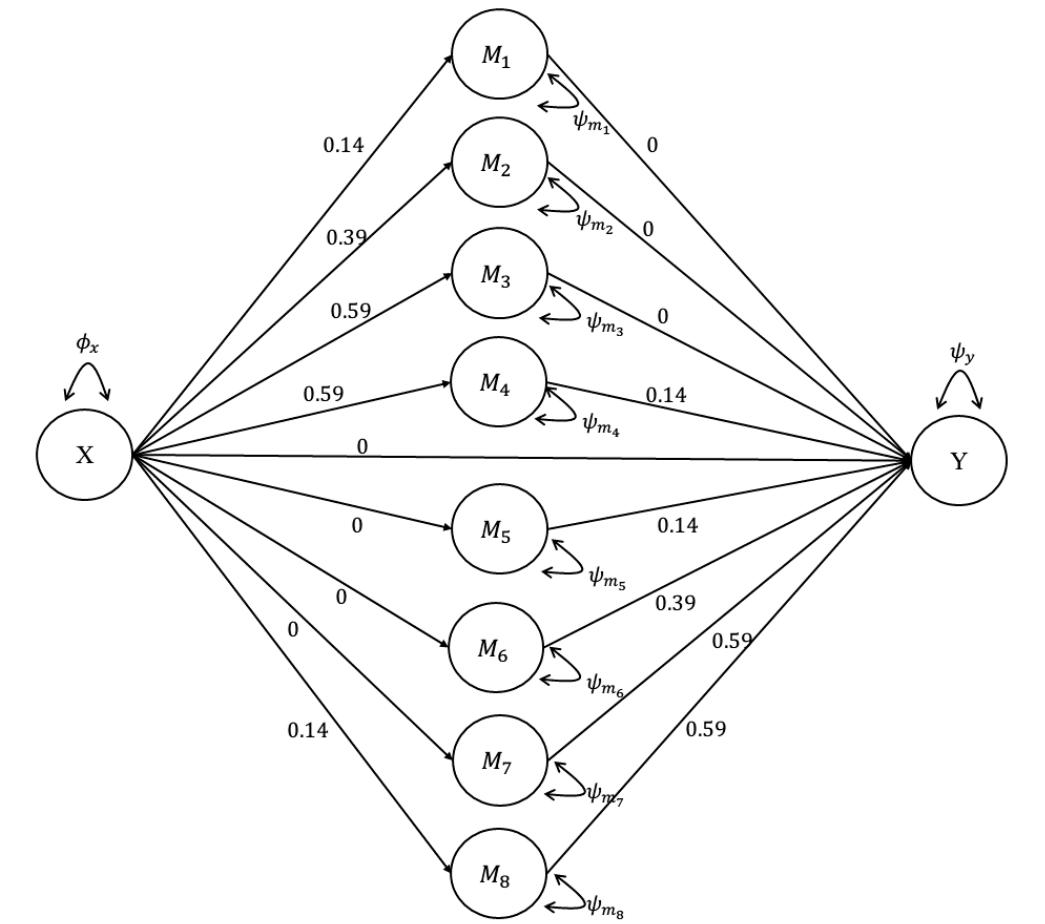
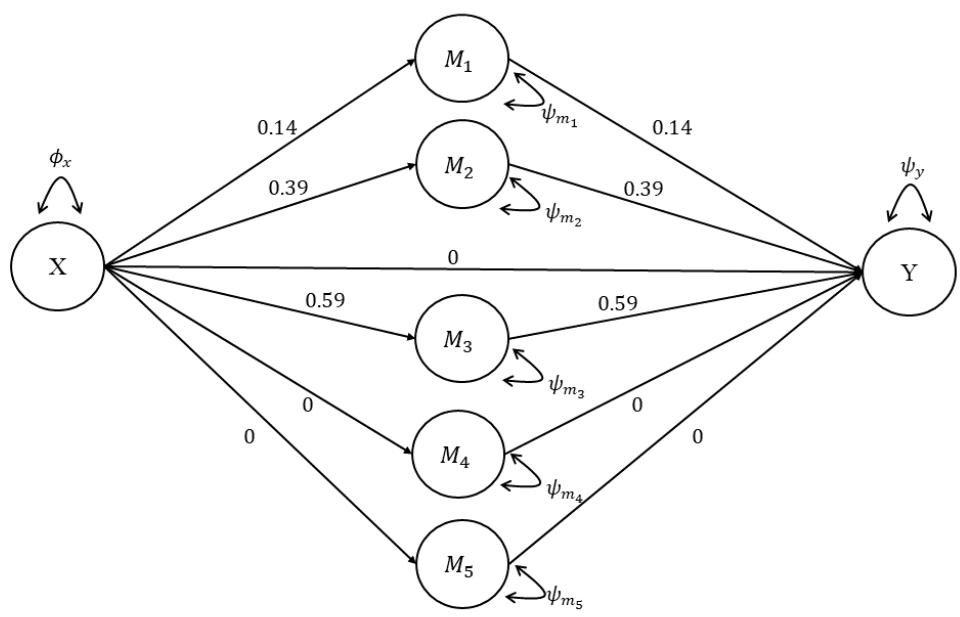


- Horseshoe (Carvalho et al., 2010): global-local shrinkage
- Spike-and-slab prior (SSP; Mitchell & Beauchamp, 1988): assigns a discrete mixture of normal distributions on parameters

模拟研究和实证演示



研究设计



研究设计

Software: R, JAGS (Plummer, 2003)

Hyperparameters of priors for path coefficients:

- Diffuse prior: $N(0, 1000)$
- Ridge: $N(0, 0.01)$ (Muthén & Asparouhov, 2012)
- Lasso and Alasso: $\alpha_l = 1, \beta_l = 0.01$ (Chen et al., 2021)
- SSP: $\sigma_p^2 = 0.001, c_p^2 \sim IG(0.5, 0.5)$ (van Erp et al., 2019)
- Horseshoe: $\rho_p \sim C^+(0, 1), v \sim C^+(0, 1)$
- Priors for other parameters (e.g., loadings): diffuse priors.
- Model convergence criteria: The estimated potential scale reduction (EPSR; Gelman et al., 1996) value should be less than 1.05 within 5,000 - 20,000 burn-in iterations.

路径系数

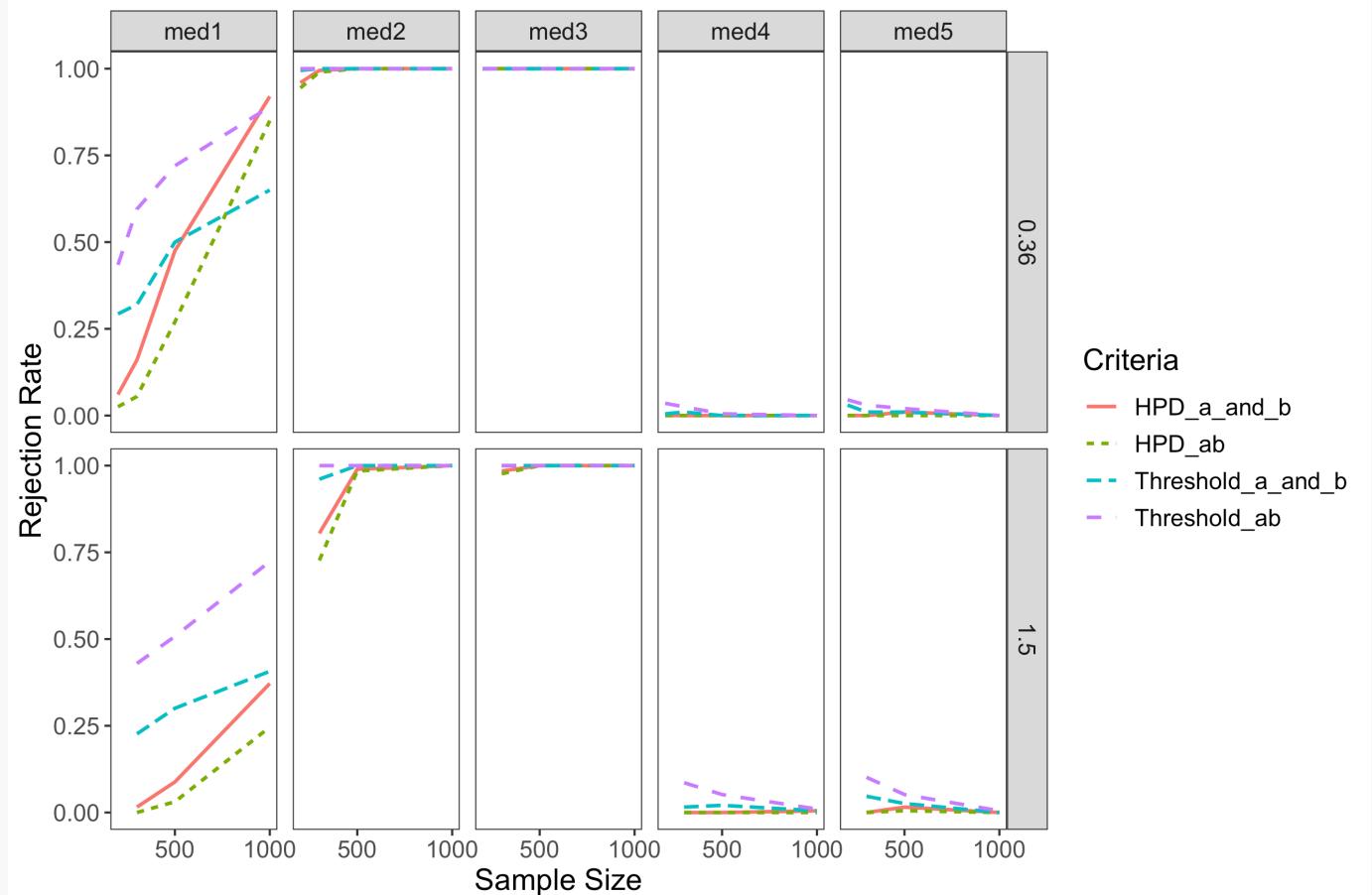
样本量：100 , 200 , 300 , 500 , 1000

残差水平：0.36 , 1.5 (Feng et al., 2020)

每个数据条件下生成200个数据集用于模拟研究

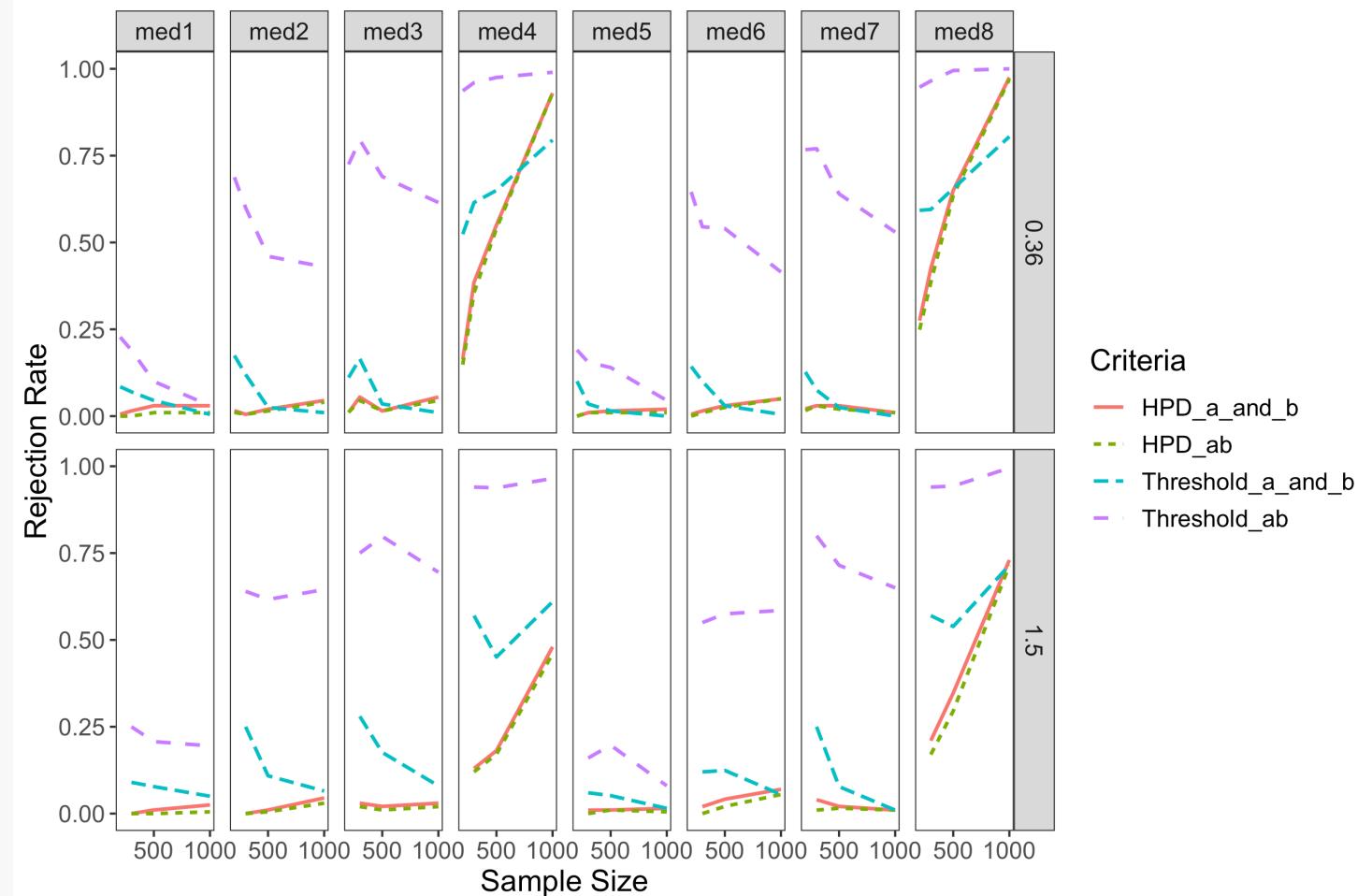
变量选择标准

- 1) a_k 和 b_k 的 HPD 区间均不包括 0。
- 2) $a_k * b_k$ 的 95%HPD 区间不包括 0；
- 3) $|a_k|$ 和 $|b_k|$ 均大于 0.1；
- 4) $|a_k * b_k|$ 大于 0.01 (即 $0.1 * 0.1$)；



变量选择标准

- 1) a_k 和 b_k 的 HPD 区间均不包括 0;
- 2) $a_k * b_k$ 的 95%HPD 区间不包括 0;
- 3) $|a_k|$ 和 $|b_k|$ 均大于 0.1;
- 4) $|a_k * b_k|$ 大于 0.01 (即 $0.1 * 0.1$);



方法比较

3

极大似然估计

Sobel法

百分位Bootstrap法

偏差校正的Bootstrap法

VS.

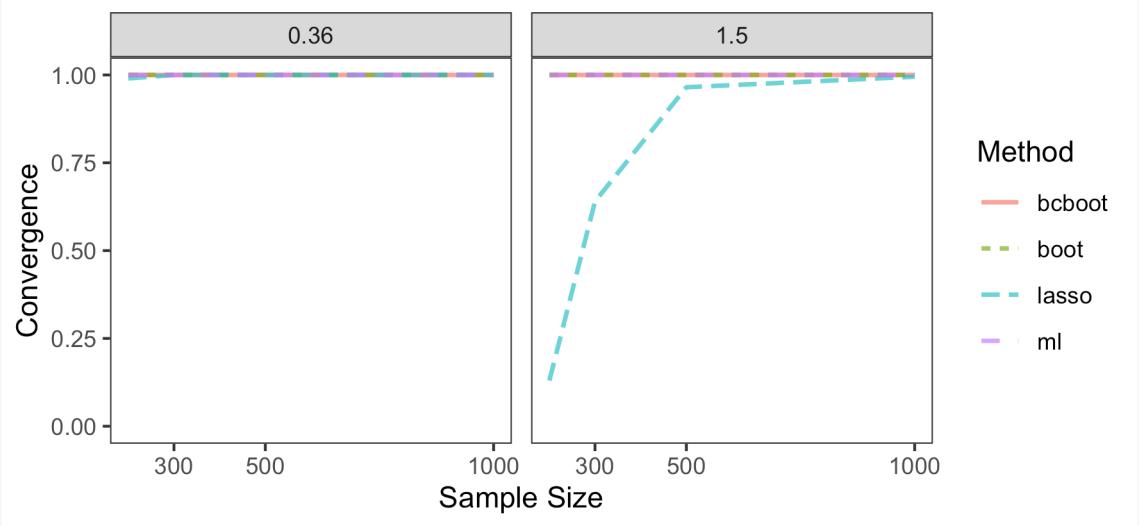
1

贝叶斯估计

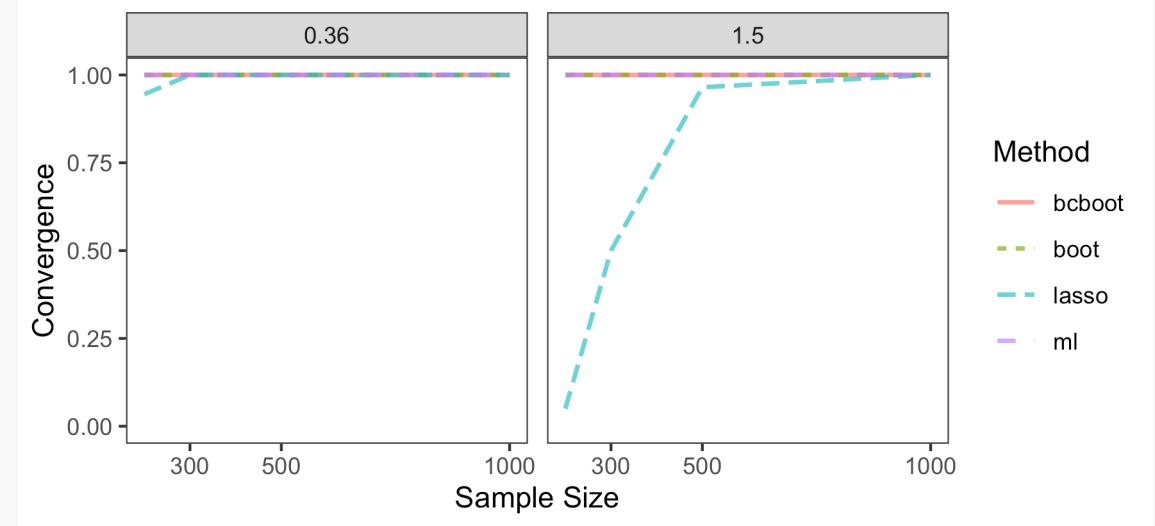
Lasso

收敛率

相比于频率学方法，贝叶斯Lasso方法在低信度(即残差水平较高)且样本量较小(200, 300)时收敛率较低。

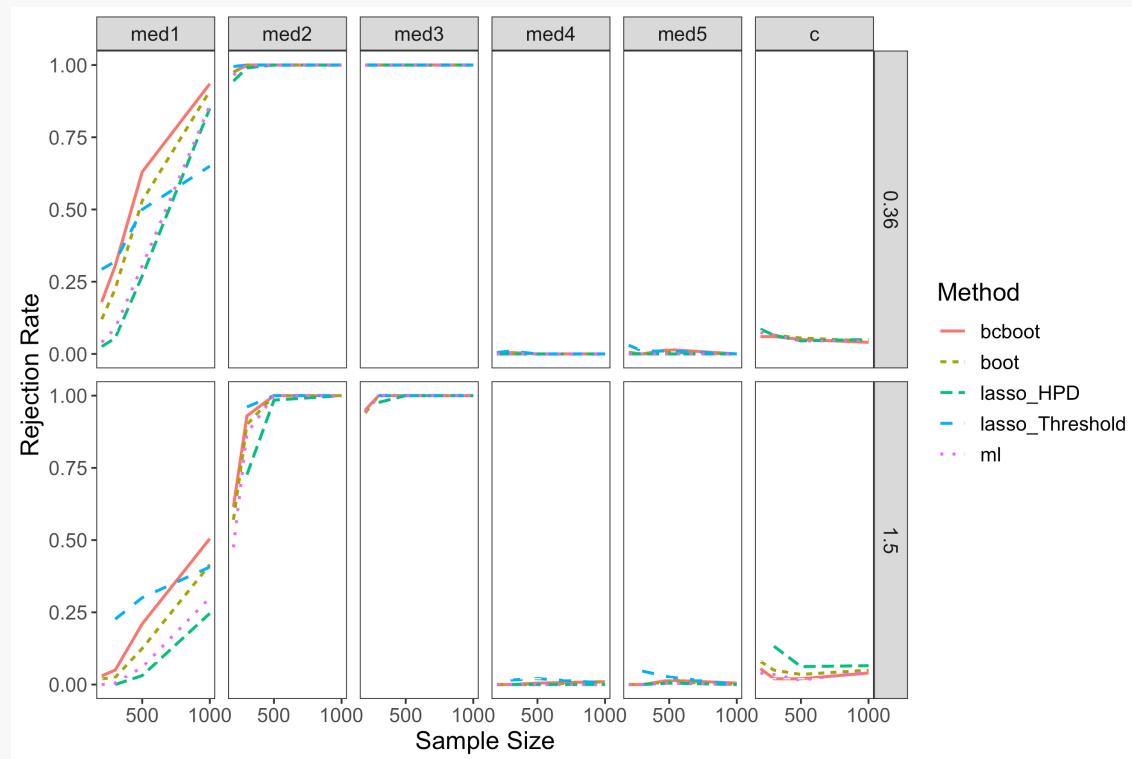


a) 对称模型

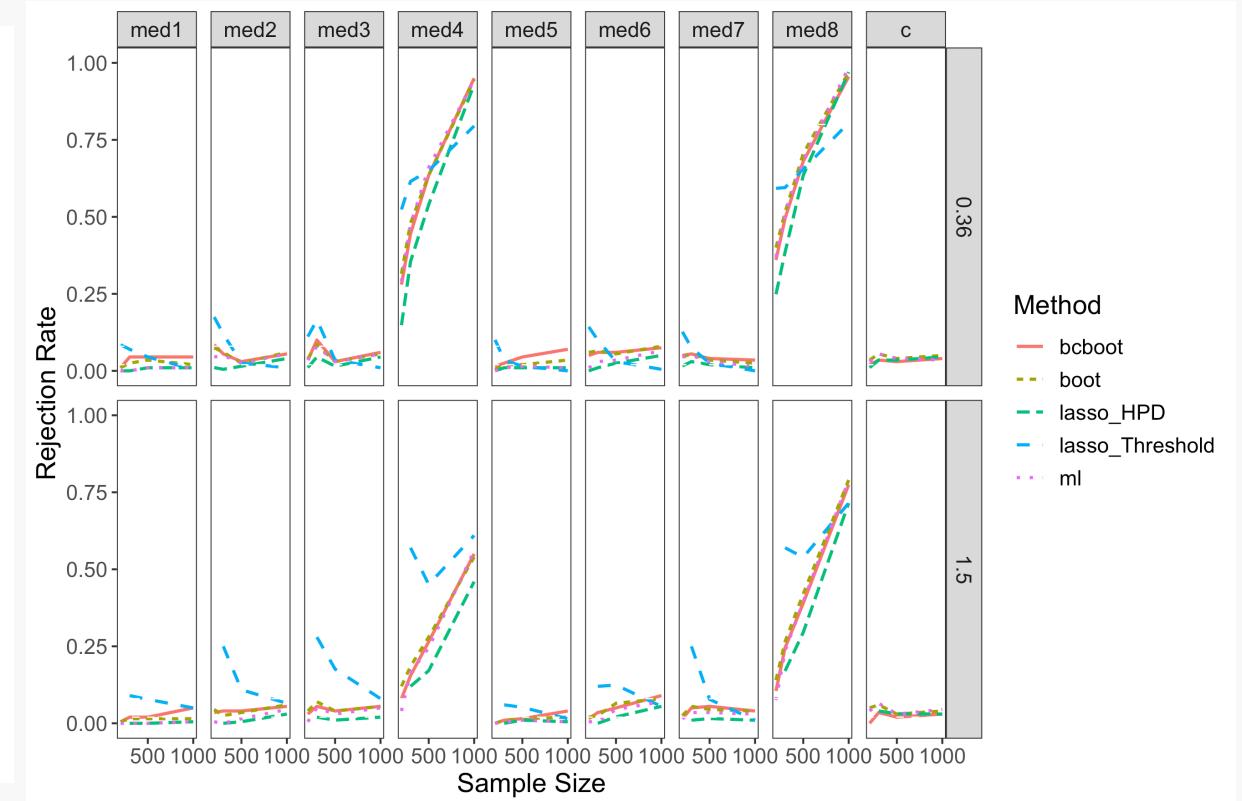


b) 非对称模型

拒绝率

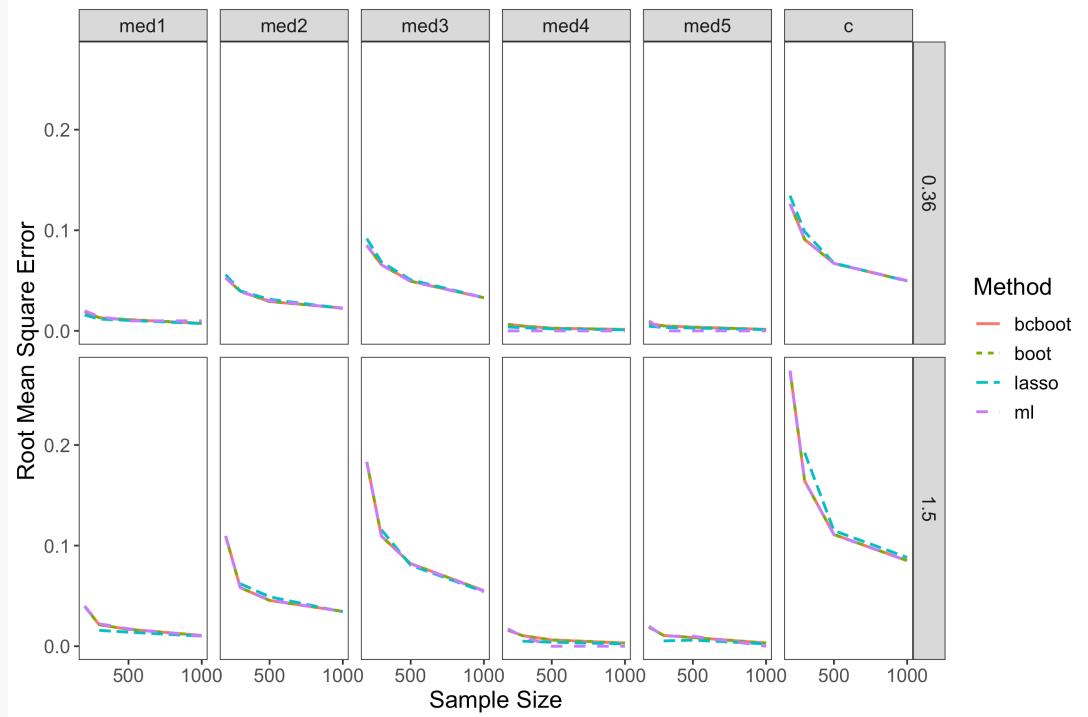


a) 对称模型

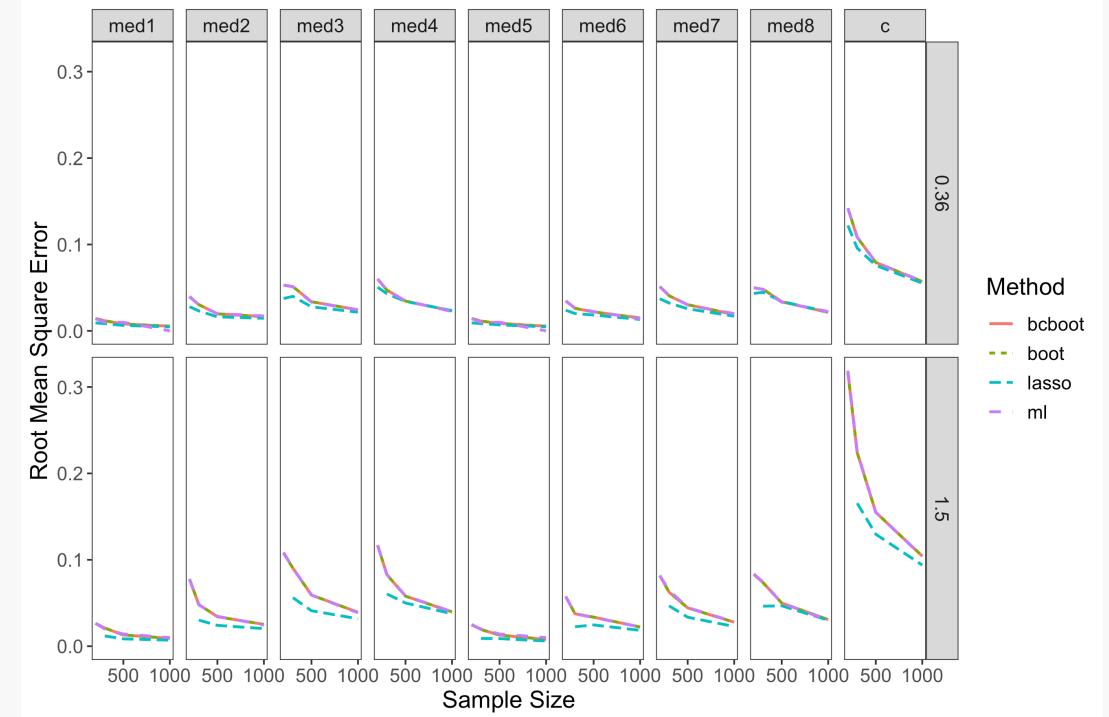


b) 非对称模型

均方根误差



a) 对称模型



b) 非对称模型

结果

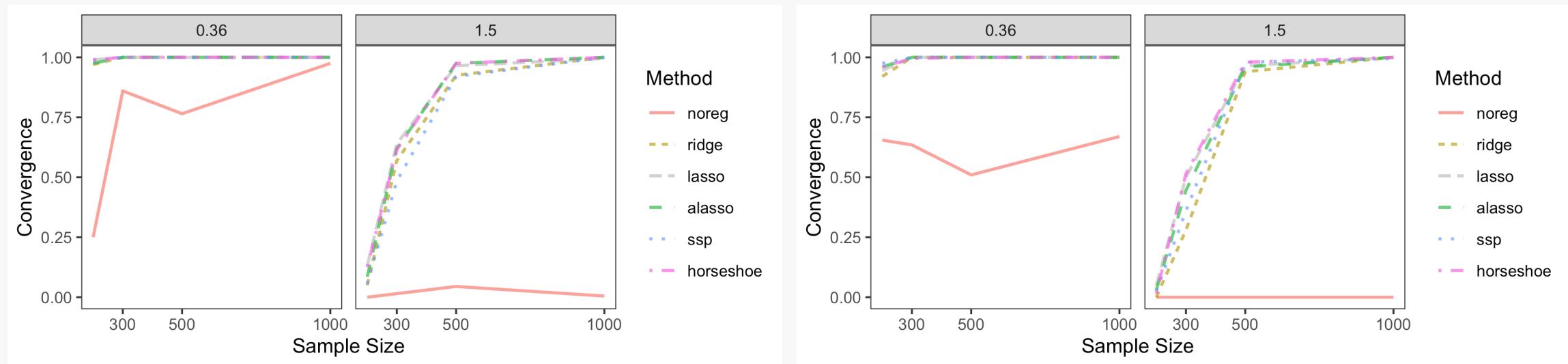
变量选择标准

- 基于中介效应整体效应量的阈值方法假阳性错误率较高，不推荐使用
- 基于两条路径单独效应量的阈值方法表现较好，对样本量不敏感
- 两种HPD区间选择方法表现相似

对比不同方法

- 偏差校正的Bootstrap法检验力优于百分位Bootstrap法，Sobel法最差
- 基于95%HPD区间的Lasso方法表现与极大似然估计类似，而基于阈值的Lasso方法能够更好地处理小样本情况或效应急量较低的情况
- 几种方法的检验力都会随着信度的降低而降低，而在低信度的情况下，基于阈值的Lasso方法优于其他方法

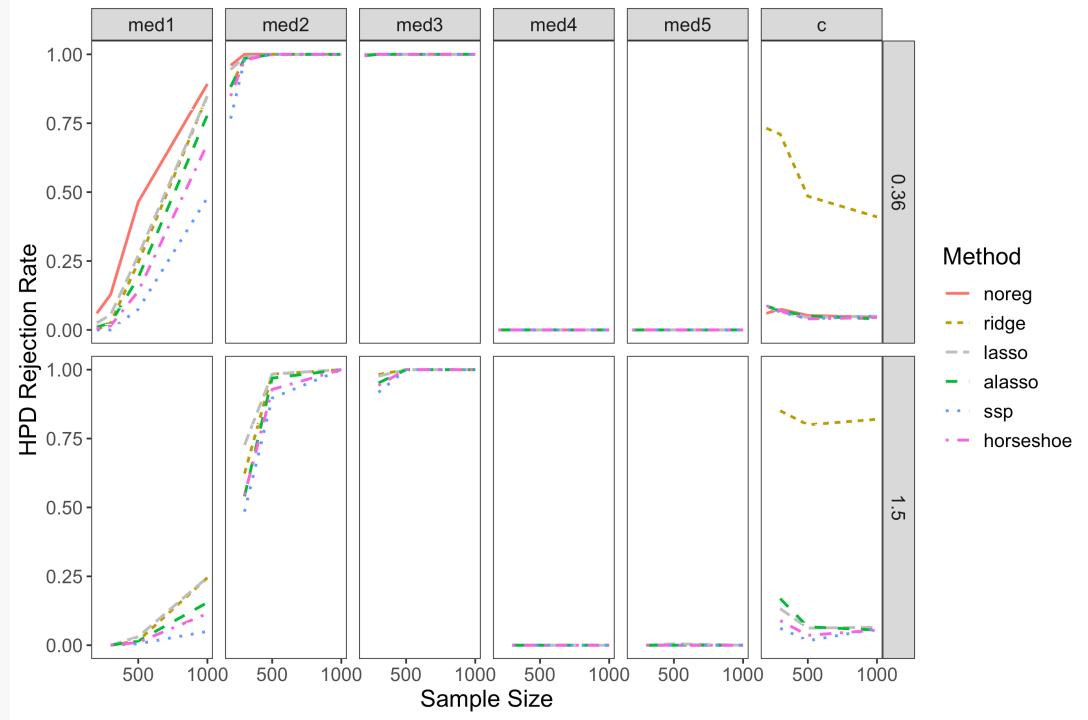
对比不同贝叶斯方法



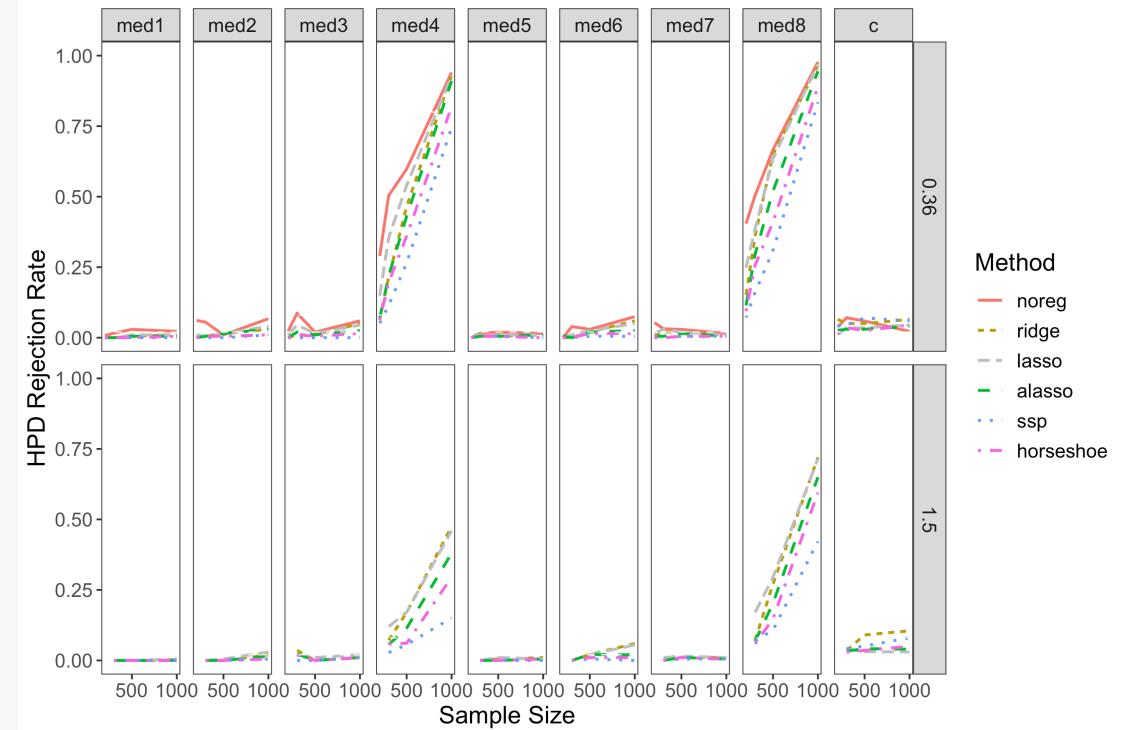
a) 对称模型

b) 非对称模型

HPD区间拒绝率

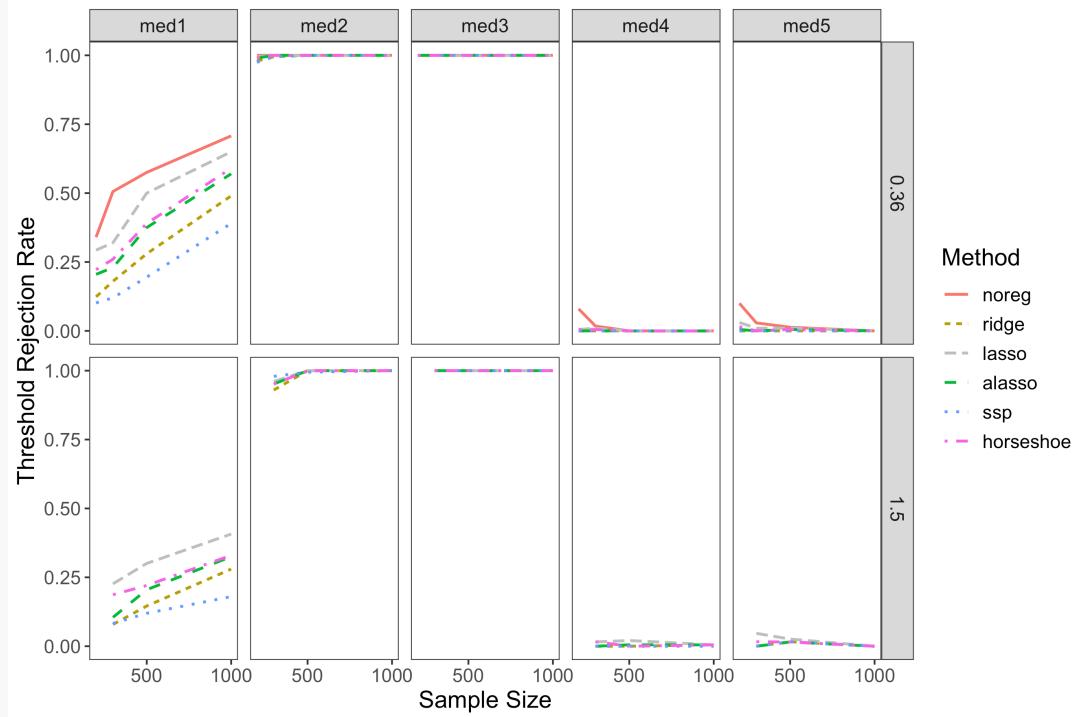


a) 对称模型

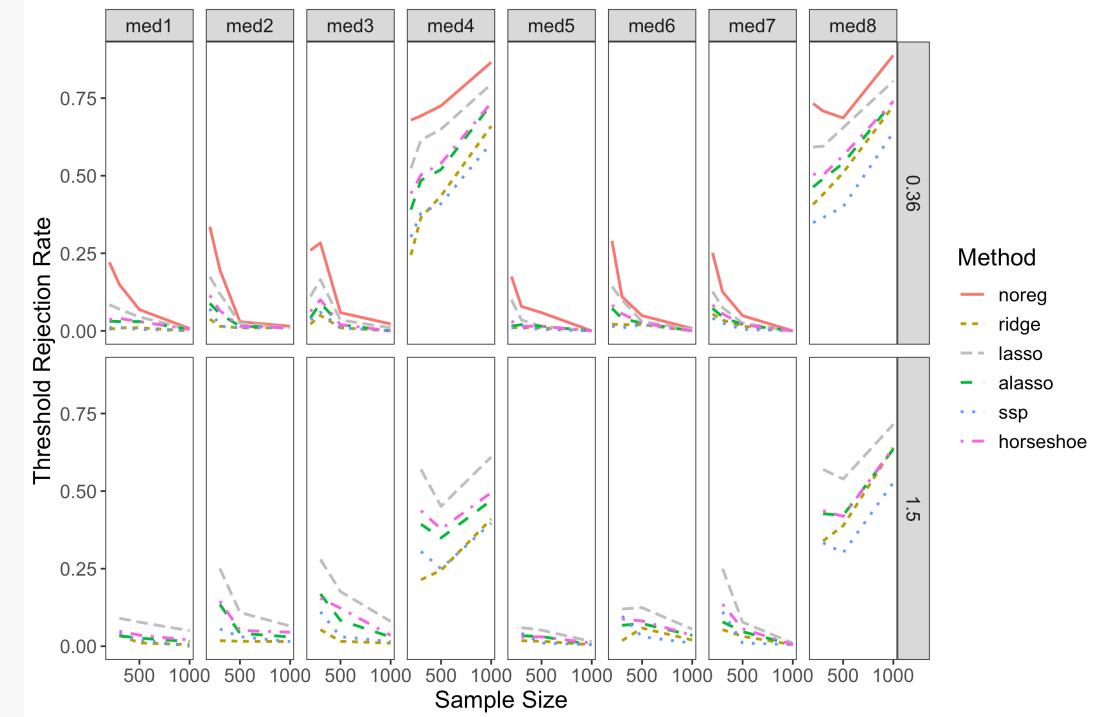


b) 非对称模型

阈值拒绝率

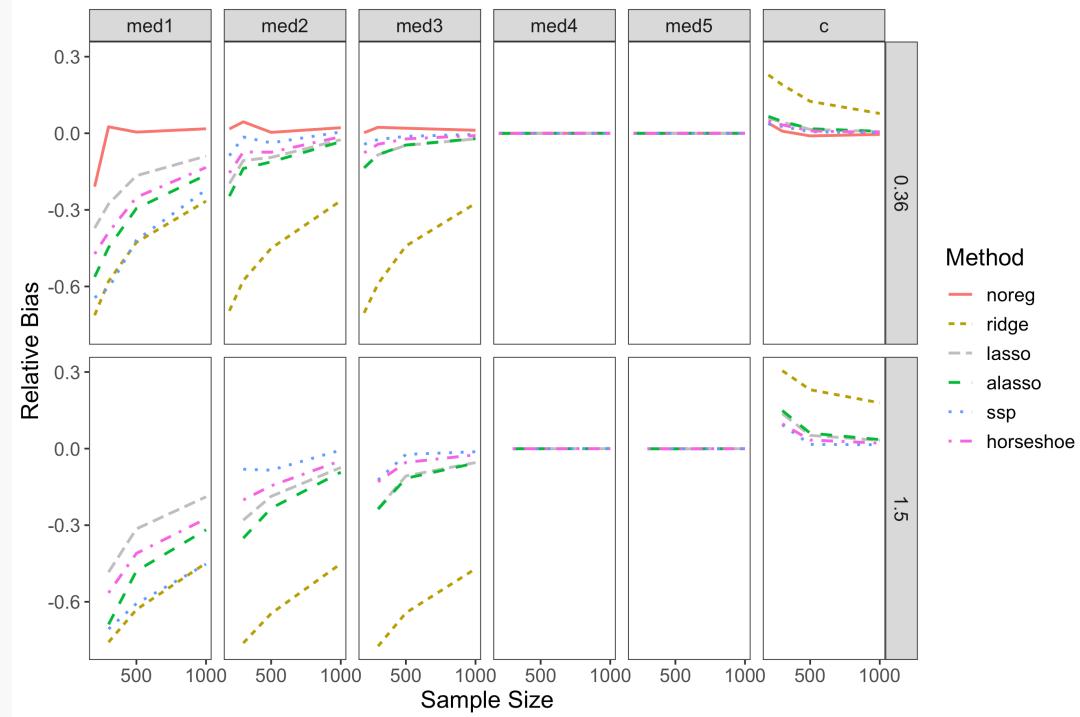


a) 对称模型

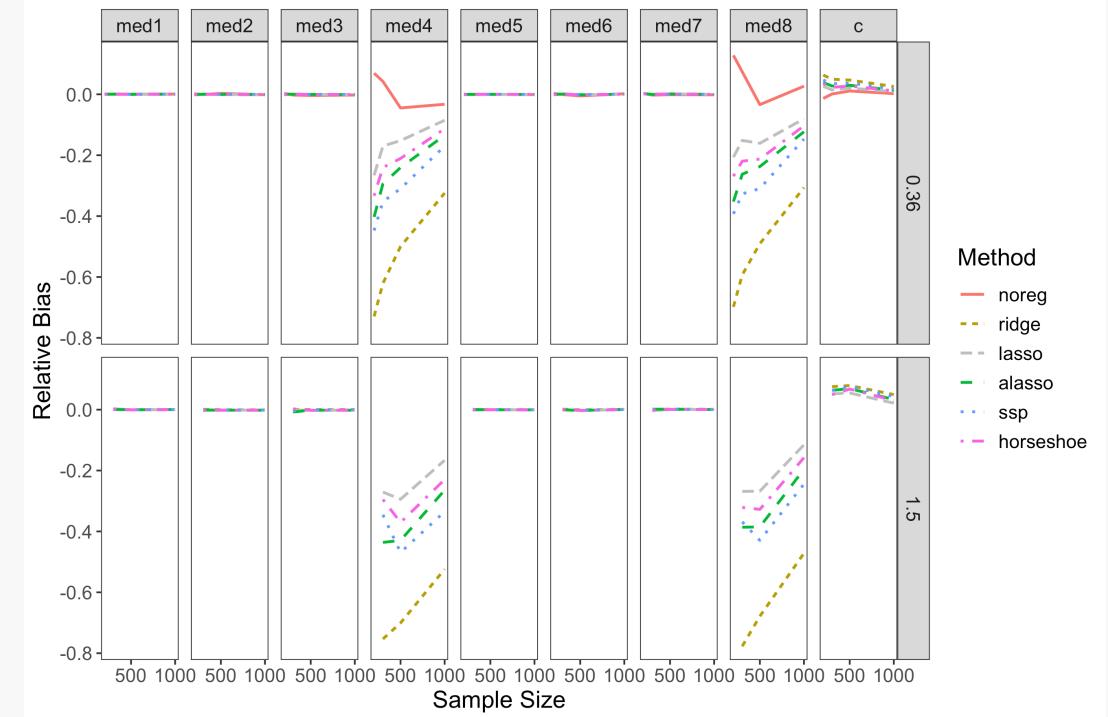


b) 非对称模型

相对偏差

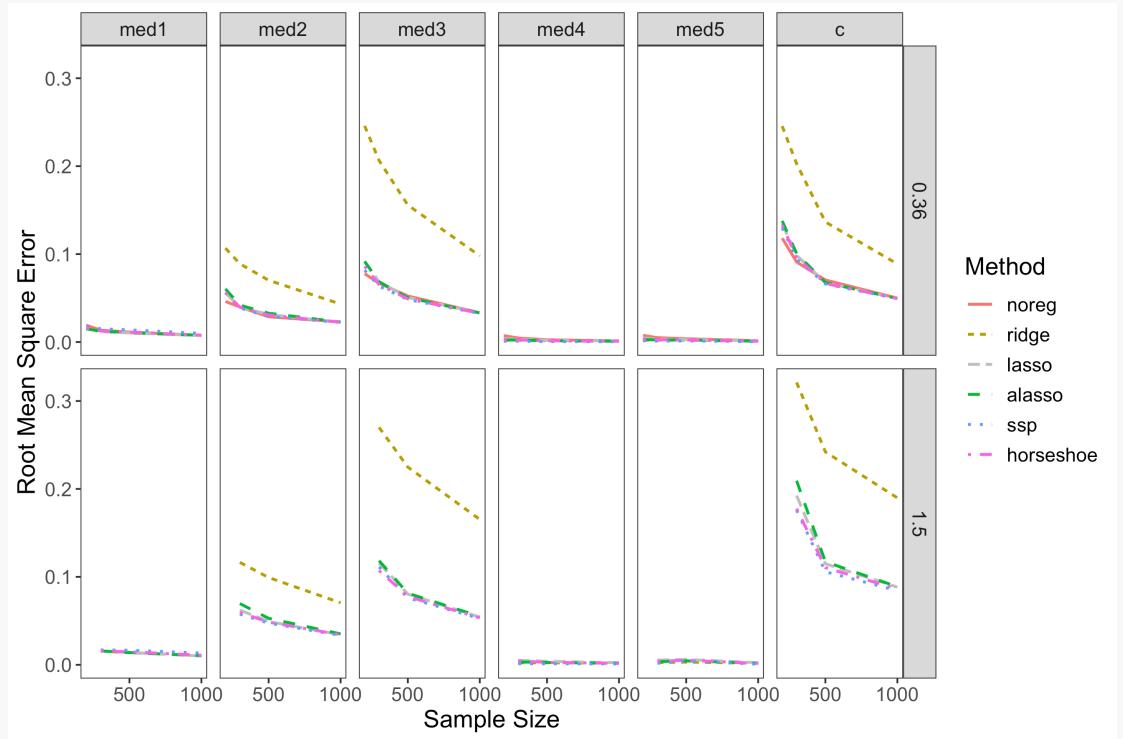


a) 对称模型

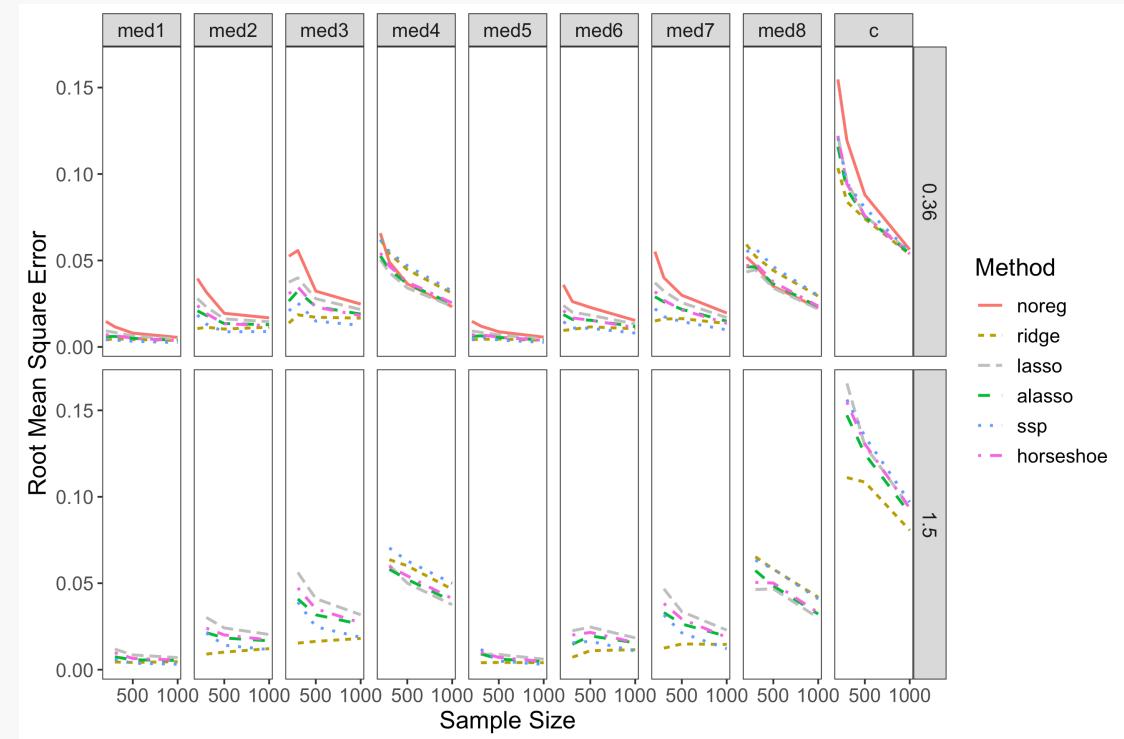


b) 非对称模型

均方根误差



a) 对称模型



b) 非对称模型

结果

对比无信息先验和正则化方法

- 阈值需要结合正则化方法使用，无信息先验下如果使用阈值容易带来较高的假阳性错误率。
- 正则化方法在小样本和低信度中具有优势，无信息先验在低信度条件下模型收敛率极低。

对比不同正则化方法

- SSP和Horseshoe方法的表现类似，对效应量较低的系数压缩力度更强，而对效应量较高的系数惩罚力度更小。
- 全局惩罚方法在变量选择中具有优势，尤其是对于效应量较小的系数，Lasso方法的检验力普遍高于其他方法。

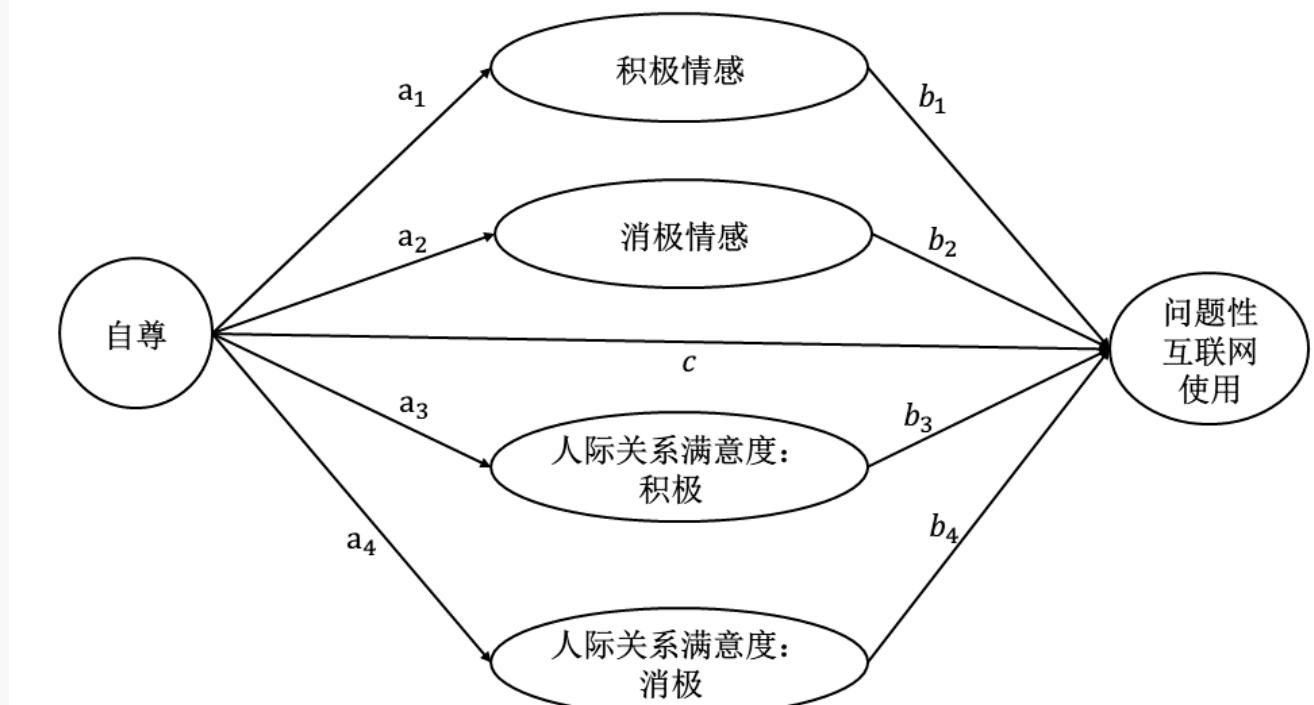
实证演示：研究背景

数据来源

507名本科生 (平均年龄20.41岁，男性占比14.5%)

问题性互联网使用

指过度的网络活动，可能会导致严重的功能损害



量表

问题性互联网使用 (Demetrovics, Szeredi, & Rozsa, 2008)

采用Likert五点计分，共18道题目。

包括三个子维度：痴迷(obsession)、忽视(Neglect)和控制障碍(Control Disorder)。

该量表在本研究中信度良好 (Cronbach's $\alpha = 0.862$)

积极消极情感量表 (Watson, Clark, & Tellegen, 1988)

Likert五点计分量表

包含积极和消极情感两个子维度，每个维度10道题目。

量表信度良好 (Cronbach's $\alpha = 0.846$)，积极和消极情感子维度问卷信度良好，Cronbach's α 分别为0.802和0.884。

人际关系满意度 (Mattson et al., 2013)

Likert八点计分量表 (0 = 完全不，7 = 总是)

包含积极和消极两个方面

量表信度良好，总量表和两个子维度的Cronbach's α 分别是0.764, 0.905和0.922

自尊 (Rosenberg Self-esteem; Rosenberg, Schooler, & Schoenbach, 1989)

采用Likert五点计分，共10道题目。

量表信度良好(Cronbach's $\alpha = 0.763$)

Sobel法、Bootstrap法

路径	β	p	95% CI ¹	95% CI ²
ab_1	-0.021	0.393	(-0.083, 0.035)	(-0.084, 0.034)
ab_2	-0.307	<0.001	(-0.442, -0.210)	(-0.443, -0.210)
ab_3	-0.045	0.282	(-0.141, 0.055)	(-0.152, 0.039)
ab_4	-0.122	0.005	(-0.236, -0.027)	(-0.224, -0.023)
c	0.031	0.753	-	-

	Ridge	Lasso	Alasso	SSP	Horseshoe
ab_1	-0.008 (-0.037, 0.020)	-0.014 (-0.067, 0.037)	-0.010 (-0.059, 0.033)	-0.005 (-0.040, 0.027)	-0.011 (-0.063, 0.036)
a_1	-0.029 (-0.121, 0.062)	0.029 (-0.134, 0.069)	-0.021 (-0.117, 0.067)	-0.011 (-0.074, 0.051)	-0.022 (-0.123, 0.070)
b_1	0.294 (0.186, 0.406)	0.493 (0.332, 0.656)	0.485 (0.325, 0.659)	0.513 (0.339, 0.682)	0.503 (0.341, 0.671)
ab_2	-0.178 (-0.244, -0.117)	-0.387 (-0.520, -0.262)	-0.378 (-0.506, -0.252)	0.399 (-0.535, -0.275)	-0.397 (-0.529, -0.267)
a_2	0.492 (0.397, 0.593)	0.675 (0.547, 0.806)	0.672 (0.545, 0.806)	0.683 (0.554, 0.813)	0.682 (0.550, 0.815)
b_2	-0.361 (-0.474, -0.253)	-0.574 (-0.741, -0.414)	-0.563 (-0.722, -0.392)	-0.585 (-0.757, -0.425)	-0.582 (-0.748, -0.426)
ab_3	-0.051 (-0.086, -0.019)	-0.073 (-0.130, -0.020)	-0.066 (-0.123, -0.014)	-0.069 (-0.129, 0.003)	-0.072 (-0.130, -0.018)
a_3	-0.262 (-0.352, -0.172)	-0.345 (-0.456, -0.241)	-0.337 (-0.444, -0.226)	-0.353 (-0.464, -0.248)	-0.347 (-0.457, -0.237)
b_3	0.196 (0.087, 0.308)	0.212 (0.071, 0.360)	0.196 (0.043, 0.340)	0.197 (-0.007, 0.351)	0.209 (0.059, 0.359)
ab_4	-0.078 (-0.124, -0.034)	-0.115 (-0.188, -0.040)	-0.106 (-0.183, -0.035)	-0.118 (-0.201, -0.018)	-0.113 (-0.189, -0.040)
a_4	0.380 (0.288, 0.472)	0.507 (0.398, 0.624)	0.501 (0.393, 0.618)	0.514 (0.397, 0.626)	0.512 (0.403, 0.630)
b_4	-0.205 (-0.312, -0.099)	-0.226 (-0.366, -0.092)	-0.213 (-0.356, -0.074)	-0.231 (-0.378, -0.032)	-0.220 (-0.358, -0.081)
c	-0.102 (-0.230, 0.035)	0.036 (-0.126, 0.206)	0.015 (-0.148, 0.178)	0.036 (-0.129, 0.200)	0.035 (-0.128, 0.202)

讨论：正则化方法对比传统方法



模型收敛

相比于无信息先验，正则化方法普遍能够获得更好的模型收敛率，尤其是在小样本或低信度情况下，但是中介模型中低信度下贝叶斯方法的收敛率不如极大似然估计



变量选择

基于阈值的正则化方法检验力普遍更高，在大部分情况下也能较好地控制假阳性错误率
基于HPD区间的正则化方法表现与传统估计方法类似



参数估计

正则化方法会低估参数，但是其RMSE普遍更低，能够获得更稳定的估计

讨论：全局正则与局部正则



变量选择和参数估计

全局惩罚方法被发现能够在变量选择中获得更高的检验力

含有局部惩罚参数的方法检验力更低，但是参数估计更准确，尤其是针对效应量较大的系数，估计更为准确



变量选择标准

- 阈值检验力更高，对数据条件变化不敏感，但是使用时变量需要经过标准化处理
- 取决于探索或验证的研究目的，采用多种标准进行充分考察

谢谢 !

Slide: https://lijinzhang.com/share/221125_r_conf.pdf

Improving lives through learning