

基于MMAe的模型平均置信区间研究

李馨, 邹晨晨*

青岛大学数学与统计学院, 山东 青岛

收稿日期: 2024年8月12日; 录用日期: 2024年9月5日; 发布日期: 2024年9月14日

摘要

本文研究并比较了基于MMAe赋权的模型平均下, Wald置信区间, MATA置信区间, 及Bootstrap置信区间的覆盖率、区间长度、左右错误率等各方面的表现性能。不同信噪比水平下的模拟显示, Wald置信区间在低信噪比时有更好的覆盖率, 高信噪比下, 三者覆盖率相近, MATA置信区间相对长度更短。在与MMA等其他模型平均方法的横向比较中, MMAe的赋权在Wald、MATA及Bootstrap三种置信区间的构建下, 均比其他赋权方式在更低的样本量下更早达到名义覆盖率。最后在实例中考察MMAe在不同置信区间下的表现, 与模拟表现一致。

关键词

模型平均, 置信区间, MMAe, MMA, MATA

Research on Model-Averaged Confidence Interval Based on MMAe

Xin Li, Chenchen Zou*

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: Aug. 12th, 2024; accepted: Sep. 5th, 2024; published: Sep. 14th, 2024

Abstract

This article studies and compares the performance of models based on MMAe weighting in terms of coverage, interval length, left and right error rates under average, Wald confidence interval, MATA confidence interval, and Bootstrap confidence interval. Simulations at different levels of signal show that the Wald confidence interval has better coverage at low signal, while at high signal, the three's confidence interval has similar coverage, and the MATA confidence interval is relatively shorter in length. In the horizontal comparison with other model averaging methods such as MMA, the weighting of MMAe reached nominal coverage earlier than other weighting methods at lower sample sizes under the construction of Wald, MATA, and Bootstrap confidence intervals. Finally, the performance of MMAe at different confidence intervals was examined in the example, which was

*通讯作者。

consistent with the simulation results.

Keywords

Model Average, Confidence Interval, MMAe, MMA, MATA

Copyright © 2024 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

随着频率学派模型平均理论的不断研究和进展, 多模型下的置信区间整合也日益体现出其研究价值与实用价值。事实上, 使用模型平均的主要原因之一是在计算这样一个区间时允许模型的不确定性。模型选择通常会导致一个比模型平均的区间覆盖率更低、稳定性更差的区间[1] [2]。

越来越多的方法被提出来构建频率模型平均的置信区间。最经典和简单的 Wald 区间形式为 $\hat{\theta} \pm z_{\alpha} \widehat{se}(\hat{\theta})$, 其中 z_{α} 是标准正态分布的 $100(1-\alpha)\%$ 分位数, Wald 区间取决于 $\hat{\theta}$ 的标准误差估计 $\widehat{se}(\hat{\theta})$ 的准确估计。但是, 估计 $\widehat{se}(\hat{\theta})$ 是相当困难的。Burnham 和 Anderson (2002) [3]提出了一些 $\widehat{se}(\hat{\theta})$ 的估计形式。Turek 和 Fletcher (2012) [4]提出将每个置信限定义为单模型 Wald 区间误差率的加权和等于期望误差率的方法, 因为涉及对单个模型估计的抽样分布尾部区域进行平均, 因此该方法被称为模型平均尾部区域 Wald 区间。针对偏斜数据的模型平均置信区间构建, Zeng 和 Fletcher 等人(2019) [5]提出了学生化 Bootstrap 模型平均尾部区域区间, 利用参数 Bootstrap 方法来估计每个模型的学生化估计的分布, 该方法只需要在模型是真实模型时, 每个模型的学生化估计是渐近枢轴的。

MMA 是一种以 Mallows Cp 准则为基础的经典模型平均方法 Hansen (2007) [6], 自提出以来被广泛的应用与拓展。Liu 和 Okui (2013) [7]将 MMA 的应用范围扩展到了异方差情况。Wan 等(2010) [8]将其渐近最优性推广至非嵌套模型。Feng 等(2020) [9]将 MMA 改进为一种类似于 Lasso (Zou, 2006) [10]的凸优化方法, 并提出了高效的稀疏权重解算法。Kun Cao 等(2023) [11]提出了 MMAe, 一种修正的 MMA 方法, 基于对 Mallows Cp 准则的惩罚修正, 以应对模型之间不可避免的共线性问题。

本文在 MMAe 赋权方法下, 构造了 Wald, MATA, Bootstrap 模型平均置信区间, 通过模拟计算来分析说明。第二部分介绍如何得到权重表达式, Wald 区间的构造方法以及相应的标准差的估计、MATA-Wald 模型平均置信区间的方法及 Bootstrap 模型平均置信区间的构造方法。第三部分进行模拟研究, 从平均区间长度、覆盖率、左右错误率方面分别比较了 Wald 区间、MATA 区间和 Bootstrap 区间在不同信噪比下的表现情况; 并且在中信噪比水平下, 对比了四种不同赋权方法在不同置信区间下覆盖率与错误率的表现情况。第四部分进行了实证分析, 展示了 MMAe 加权方法在不同置信区间下的表现。第五部分对模拟结果进行总结, 得出结论。

2. 置信区间构建

本文在使用 MMAe 赋权方法得到的权重下, 比较 Wald, MATA, Bootstrap 模型平均置信区间的表现情况。区间的频率性质使用覆盖率以及左右错误率这两个指标。

2.1. MMAe 方法

MMAe 方法是一种基于惩罚 Mallows Cp 的修正 MMA, 为了处理模型间不可避免的共线性, 添加

了一个加权弹性网惩罚项, 文中通过 CD 算法得到了 MMAe 权重[11], 在 Kun Cao [11]的基础上, w 的表达式通过以下过程计算:

$$\min_{w^T \mathbf{1}_M=1, w \geq 0} \frac{1}{2} w^T \hat{e}^T \hat{e} w + (1-\alpha) \hat{\sigma}^2 w^T K + \alpha \hat{\sigma}^2 w^T \text{diag}(K) w$$

其中, $w = (w_1, w_2, \dots, w_m)^T$, $\hat{e}_m = y - \hat{\mu}_m$, $K = (k_1, k_2, \dots, k_m)^T$, $\text{diag}(K) = \begin{pmatrix} k_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & k_m \end{pmatrix}$, $\alpha \in [0, 1]$, α 决

定了 L_1 和 L_2 惩罚的比例, $\hat{\sigma}^2$ 是完全模型或最大模型对 σ^2 的估计。

2.2. Wald 置信区间

假设 $T_{\hat{\theta}} = \frac{\hat{\theta} - \theta}{\widehat{se}}$ 服从标准正态分布, 其中 \widehat{se} 是 $\hat{\theta}$ 标准差的估计, 模型平均的 Wald 置信区间为[12]:

$$\hat{\theta} \pm z \cdot \widehat{se}$$

其中 z 是标准正态分布的 $100(1-\alpha)\%$ 分位数。本文中的 $\widehat{se} = \sqrt{\sum_{m=1}^M w_m \left[\left(\frac{\widehat{se}(\hat{\theta}_m) t_{v_m}}{z} \right)^2 + (\hat{\theta}_m - \hat{\theta})^2 \right]}$, 其中 t_{v_m} 为自由度为 v_m 的 t 分布的分位数, v_m 是与模型 M_i 相关的残差的自由度。

2.3. MATA 置信区间

MATA 区间[4]通过将模型平均的上下错误率设置为等于所需要的错误率来选择置信限, 其置信限的定义是: 每个候选模型的错误率的加权平均值等于所需的总体错误率。于是, $100(1-2\alpha)\%$ MATA 置信区间的下限 θ_L 满足:

$$\sum_{i=1}^R w_i \alpha_i(\theta_L) = \alpha$$

其中, 当模型 M_i 为真时, $\alpha_i(\theta_L)$ 是与使用 θ_L 作为 θ 的 $100(1-2\alpha)\%$ 置信下限相关的错误率的估计, 置信上限同理。

2.4. Bootstrap 置信区间

Bootstrap 置信区间[13]假设恰有一个候选模型是真实的, 记 c_i 为未知的指示变量, 如果模型 M_i 为真, 则 $c_i = 1$ 否则 $c_i = 0$, 假设我们知道 $\hat{\theta}_i$ 的真实分布记为 $G_i(\cdot)$, 参数 θ 的置信区间 $[\theta_L, \theta_U]$ 满足下列方程:

$$\begin{cases} \sum_{i=1}^R c_i G_i(\theta_L) = \alpha \\ \sum_{i=1}^R c_i G_i(\theta_U) = 1 - \alpha \end{cases}$$

由于 c_i 是未知的, 我们用与模型相对应的基于某种信息准则的模型权重 w_i 来估计上述公式中的 c_i , 则参数 θ 的置信区间 $[\theta_L, \theta_U]$ 满足下列方程:

$$\begin{cases} \sum_{i=1}^R w_i G_i(\theta_L) = \alpha \\ \sum_{i=1}^R w_i G_i(\theta_U) = 1 - \alpha \end{cases}$$

如果 $\hat{\theta}_i$ 的真实分布情况未知, 假设每个模型的学生化的参数估计 $\frac{\hat{\theta}_i - \theta}{se(\hat{\theta}_i)}$ 服从标准正态分布或 t 分布,

即 $G_i(\bullet)$ 为标准正态分布或者 t 分布的分布函数, 本文利用参数 Bootstrap 方法构造变量的分布, 即用参数估计的 Bootstrap 分布 $\widehat{G}_i(\bullet)$ 来近似上述公式中的分布 $G_i(\bullet)$, 从而计算参数的模型平均置信区间。获取的参数 θ 的 $100(1-2\alpha)\%$ 的置信区间即为 z 版本或者 t 版本的 MATA 区间。

3. 数据模拟

考虑正态线性模型, 包括 8 个候选模型, 第 m 个模型中包含预测变量 $x_1, x_2, x_3, \dots, x_{3+m}$, $m = 1, \dots, 8$, 每个模型 $\mu_i = \sum_{j=1}^k \theta_j x_{ji}$ 。假设数据的产生过程如下[11]:

$$y_i = \mu_i + \varepsilon_i = \sum_{j=1}^k \theta_j x_{ji} + \varepsilon_i, i = 1, \dots, n$$

其中, $x_{1i} = 1$, $(x_{2i}, \dots, x_{ki}) \sim N_{k-1}(0, Q)$, $i = 1, \dots, n$, $\varepsilon_i \sim N(0, 1)$ 。 Q 的对角线元素为 1, 非对角线元素为 ρ 。 $\theta = c \left(\frac{1}{a}, \frac{1}{a}, \frac{1}{a}, \frac{1}{\sqrt{n}} \left(1, \frac{q-1}{q}, \dots, \frac{1}{q} \right) \right)$, 其中 q 是协变量的数目。在这里令 $a = 12$, 并且 c, ρ 满足 $R^2 = \frac{\tilde{\theta}' Q \tilde{\theta}}{1 + \tilde{\theta}' Q \tilde{\theta}}$, $\tilde{\theta} = (\theta_2, \dots, \theta_k)'$, $q = 8$ 。 x 是在每轮模拟中随机产生的。样本量分别取 35, 45, \dots , 95 这 7 种情形。对于感兴趣的参数选取为预测空间 x 各个点的 μ 值, 这里考虑预测空间 x 在不同信噪比下不同值处的 μ 值。这里模拟次数为 10^3 次, 在计算 Bootstrap 模型平均置信区间时 Bootstrap 样本量取 $B = 1000$ 。

3.1. MMAe 权重下不同置信区间比较

由图 3、图 4 和图 5 可以看出响应变量均值在不同置信区间的覆盖率和左、右错误率随样本容量 n 的变化趋势, 图 3 右图、图 4 右图和图 5 右图的下方和上方分别为不同置信区间在不同样本量下的左、右错误率。为了表示方便, 左错误率取负值后展示在图 3 右图、图 4 右图和图 5 右图的图下方, 图 1 和图 2 展示了不同置信区间在 $R^2 = 0.4$ 、 $R^2 = 0.6$ 和 $R^2 = 0.8$ 时不同样本量下的平均区间长度。

通过图 1、图 2、图 3、图 4 和图 5 的数据可以得到以下结论: 1) 右错误率方面, Wald 区整体要好于其它两种区间, Bootstrap 模型平均置信区间在样本量较小时具有较低的右错误率水平, 在 $R^2 = 0.4$

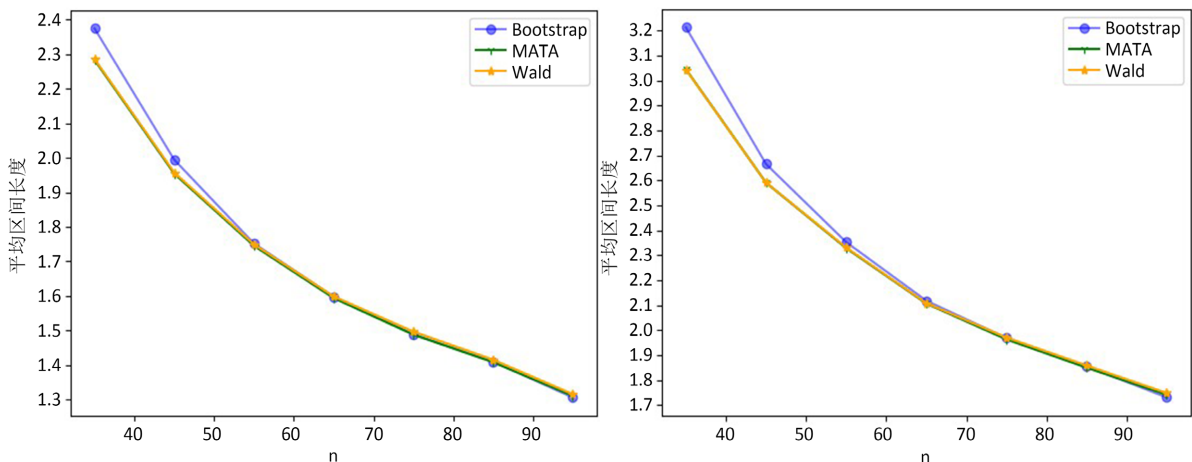


Figure 1. The average interval length corresponding to different confidence intervals when $R^2 = 0.4$ (left) and $R^2 = 0.6$ (right) 图 1. $R^2 = 0.4$ (左)与 $R^2 = 0.6$ (右)时不同置信区间对应的平均区间长度

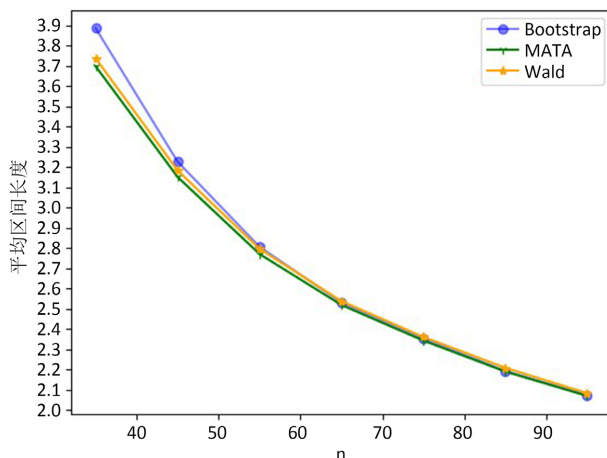


Figure 2. The average interval length corresponding to different confidence intervals when $R^2 = 0.8$

图 2. $R^2 = 0.8$ 时不同置信区间对应的平均区间长度

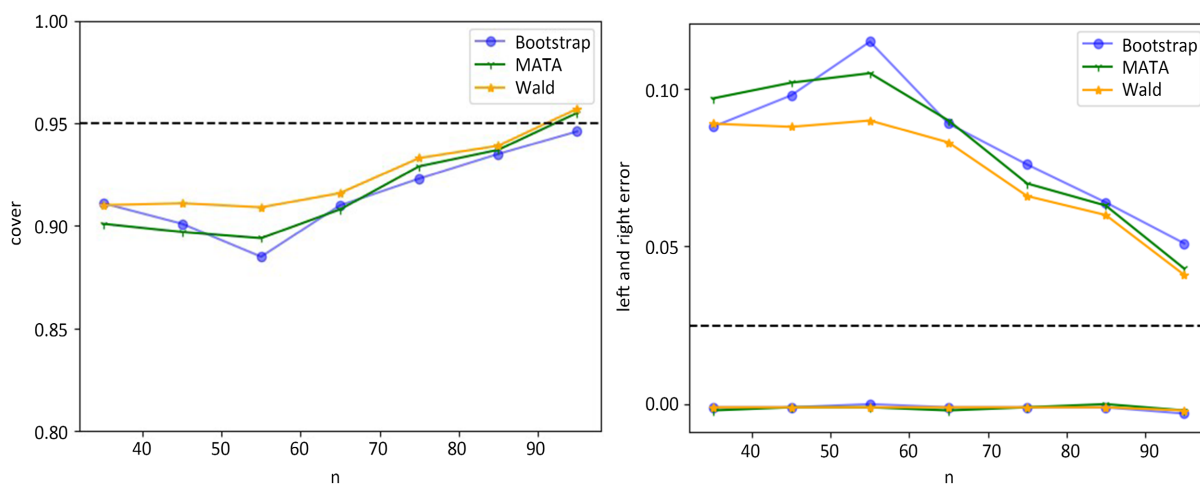


Figure 3. The coverage rate (left) and error rate (right) corresponding to different confidence intervals when $R^2 = 0.4$

图 3. $R^2 = 0.4$ 时不同置信区间对应的覆盖率(左), 左右错误率(右)

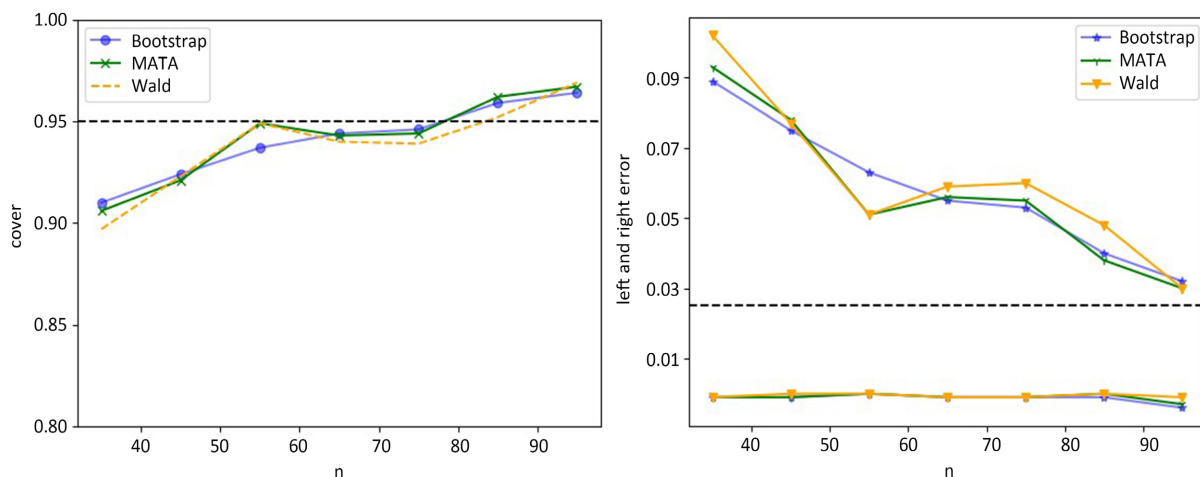


Figure 4. The coverage rate (left) and error rate (right) corresponding to different confidence intervals when $R^2 = 0.6$

图 4. $R^2 = 0.6$ 时不同置信区间对应的覆盖率(左), 左右错误率(右)

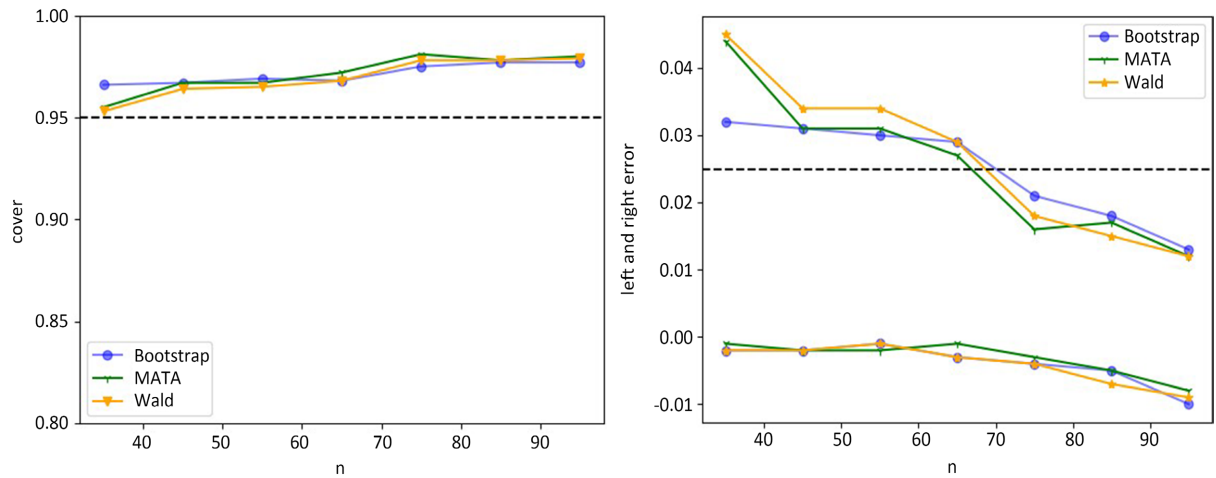


Figure 5. The coverage rate (left) and error rate (right) corresponding to different confidence intervals when $R^2 = 0.8$
图 5. $R^2 = 0.8$ 时不同置信区间对应的覆盖率(左), 左右错误率(右)

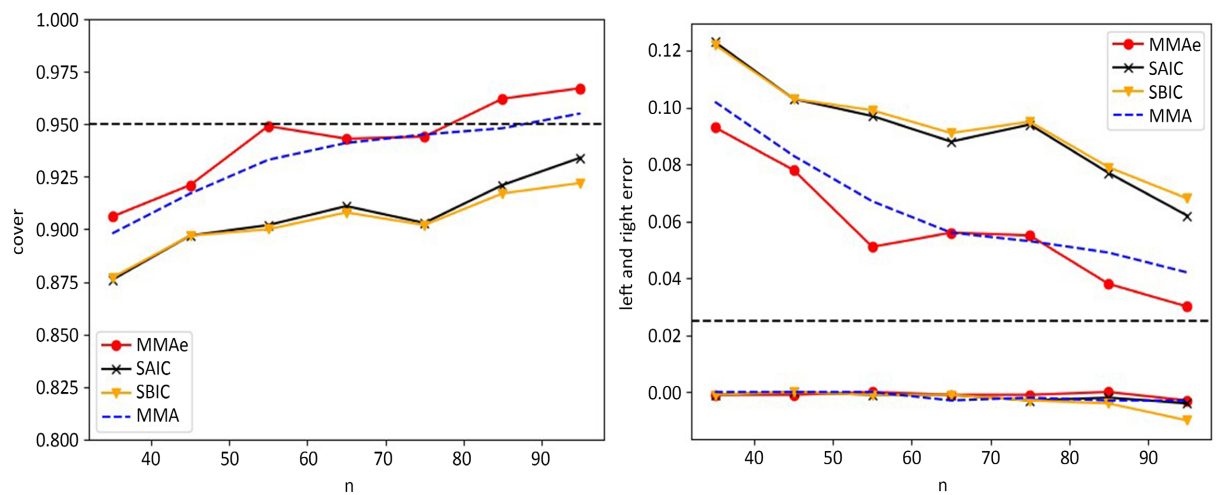


Figure 6. Coverage rate (left) and error rate (right) of MATA interval when $R^2 = 0.6$
图 6. $R^2 = 0.6$ 时 MATA 区间的覆盖率(左), 左右错误率(右)

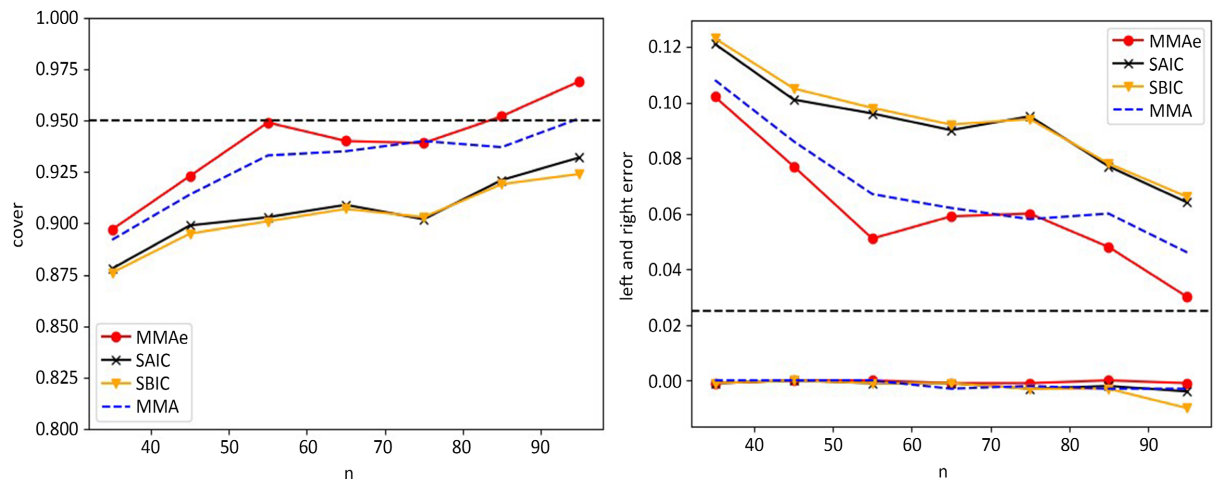


Figure 7. Coverage rate (left) and error rate (right) of Wald interval when $R^2 = 0.6$
图 7. $R^2 = 0.6$ 时 Wald 区间的覆盖率(左), 左右错误率(右)

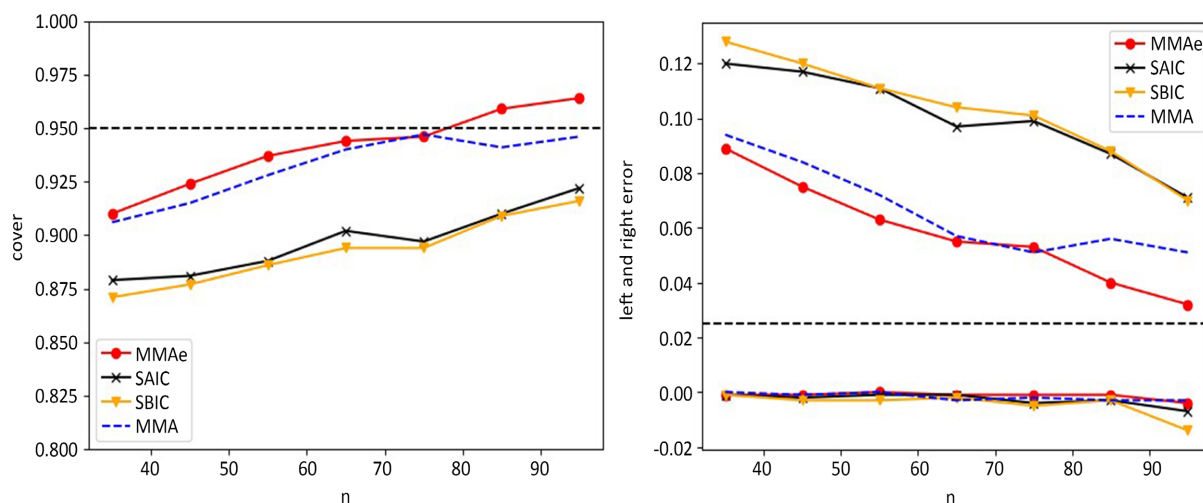


Figure 8. Coverage rate (left) and error rate (right) of Bootstrap interval when $R^2 = 0.6$

图 8. $R^2 = 0.6$ 时 Bootstrap 区间的覆盖率(左), 左右错误率(右)

和 $R^2 = 0.6$ 的情况下更接近名义错误率; 左错误率方面, 三种区间表现差距不大, 2) 覆盖率方面, Bootstrap 区间整体要逊于 MATA 区间和 Wald 区间, 在 $R^2 = 0.4$ 和 $R^2 = 0.6$ 情况下, Wald 置信区间整体表现较好。3) 平均区间长度方面, 随着样本量的增大, 三种置信区间平均长度的差距减小; 在较小样本条件下, Bootstrap 的平均区间长度长, 以达到较高的覆盖率。4) 随着样本量的增加, 三种区间估计方法都逐渐接近名义覆盖率。5) 随着 R^2 的增大, Bootstrap 区间, MATA 区间, Wald 区间得到的平均区间长度均增大; 且覆盖率在更低的样本量下到达名义覆盖率。

3.2. 不同赋权方法在不同置信区间下的比较

通过图 6、图 7、图 8 的数据可以得到以下结论: 1) 右错误率方面, MMAe 方法整体要好于其它三种方法, 且更接近名义错误率, 左错误率方面, 四种方法表现差异不大; 2) 覆盖率方面, MMAe 方法无论在何种区间下表现都好, 均在更低的样本量下到达名义覆盖率, MATA 置信区间和 Wald 置信区间相差不大, Bootstrap 置信区间在该信噪比下在样本量较大时到达名义覆盖率; 3) 随着样本量的增加, 三种区间估计方法都逐渐接近名义覆盖率。

4. 实证分析

本节应用提到的模型平均方法, 结合不同的模型平均区间方法, 利用财产估值数据来说明这些模型平均区间方法之间的差异。我们所考虑的数据来自 Subhash C. Narula & John F. Wellington [14] 的论文。数据包含 9 个预测变量, 分别为: x_1 (地方的, 学校的, 国家的) 税额/1000、 x_2 浴缸数、 x_3 占地面积(平方英尺 $\times 1000$)、 x_4 居住面积(平方英尺 $\times 1000$)、 x_5 车库数、 x_6 房间数、 x_7 卧室数、 x_8 房龄(年)、 x_9 壁炉数, 响应变量是房屋销售价格/1000。我们的目标是通过预测变量来预测房屋销售价格/1000 的值。我们考虑 7 个模型, 其中第 m 个模型中包含预测变量 $x_0, x_1, x_2, x_3, \dots, x_{2+m}$, $m = 1, \dots, 7$, x_0 为截距项, 每个模型 $\mu_i = \sum_{j=1}^k \theta_j x_{ji}$, 本实例给定置信区间的名义置信水平为 95%。

图 9 展示了关于财产估值数据预测均值在 MMAe 权重下不同置信区间(MATA, Bootstrap, Wald)的分析结果, Bootstrap 区间的区间长度大于 Wald 置信区间和 MATA 置信区间, Bootstrap 模型平均置信区间在样本量较小时具有较低的权利错误率水平, 且在样本量较小的情况下区间长度在三种方法中最长, 这与我们的模拟情况表现一致。

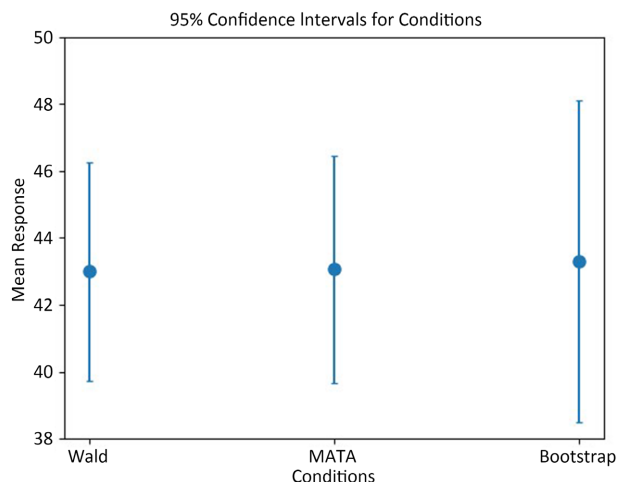


Figure 9. 95% confidence interval for models with different confidence intervals

图 9. 不同置信区间模型平均 95% 置信区间

5. 结论

通过上述模拟可知在一个数据集中, 当样本量小且使用 MMAe 方法赋值权重时, 使用 Bootstrap 置信区间会更好; 当有用的信号相对较少, 而噪音或随机变动相对较多时, 使用 Wald 置信区间会得到更好的覆盖率; 在噪音或随机变动相对较少时, 在 MMAe 权重下, 使用任何一种模型平均置信区间方法在覆盖率方面表现都较好, 但是 MATA 置信区间长度更短, 故在该情形下, MATA 置信区间更合适。

文中模拟了在假设情形下 MMAe 赋权方法的表现情况, 但是并未模拟模型平均置信区间的其它构造方法。随着模型平均的研究发展, 利用模型平均构造置信区间的方法也是层出不穷。除了本文中提到的模型平均置信区间的方法, 还有基于转换的模型平均尾部区域置信区间, 学生化 Bootstrap 模型平均尾部区域区间等, 因此其他的模型平均方法在 MMAe 赋权时还有待挖掘。

参考文献

- [1] Efron, B. (2014) Estimation and Accuracy after Model Selection. *Journal of the American Statistical Association*, **109**, 991-1007. <https://doi.org/10.1080/01621459.2013.823775>
- [2] Longford, N.T. (2008) An Alternative Analysis of Variance. *SORT-Statistics and Operations Research Transactions*, **32**, 77-92.
- [3] Burnham, K.P. and Anderson, D.R. (2002) Model Selection and Multimodel Inference: A Practical Information—Theoretic Approach. 2nd Edition, Springer. <https://doi.org/10.1007/b97636>
- [4] Turek, D. and Fletcher, D. (2012) Model-Averaged Wald Confidence Intervals. *Computational Statistics & Data Analysis*, **56**, 2809-2815. <https://doi.org/10.1016/j.csda.2012.03.002>
- [5] Zeng, J., Fletcher, D., Dillingham, P.W. and Cornwall, C.E. (2019) Studentized Bootstrap Model-Averaged Tail Area Intervals. *PLOS ONE*, **14**, e0213715. <https://doi.org/10.1371/journal.pone.0213715>
- [6] Hansen, B.E. (2007) Least Squares Model Averaging. *Econometrica*, **75**, 1175-1189. <https://doi.org/10.1111/j.1468-0262.2007.00785.x>
- [7] Liu, Q. and Okui, R. (2013) Heteroscedasticity-Robust C_p Model Averaging. *The Econometrics Journal*, **16**, 463-472. <https://doi.org/10.1111/ectj.12009>
- [8] Wan, A.T.K., Zhang, X. and Zou, G. (2010) Least Squares Model Averaging by Mallows Criterion. *Journal of Econometrics*, **156**, 277-283. <https://doi.org/10.1016/j.jeconom.2009.10.030>
- [9] Feng, Y., Liu, Q. and Okui, R. (2020) On the Sparsity of Mallows Model Averaging Estimator. *Economics Letters*, **187**, Article 108916. <https://doi.org/10.1016/j.econlet.2019.108916>

-
- [10] Zou, H. (2006) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, **101**, 1418-1429. <https://doi.org/10.1198/016214506000000735>
- [11] Cao, K., Li, X., Zhou, Y. and Zou, C. (2023) On Improvability of Model Averaging by Penalized Model Selection. *Stat*, **12**, e529. <https://doi.org/10.1002/sta4.529>
- [12] Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997) Model Selection: An Integral Part of Inference. *Biometrics*, **53**, 603-618. <https://doi.org/10.2307/2533961>
- [13] 郭庆光. 若干情形下的模型平均置信区间[D]: [硕士学位论文]. 青岛: 青岛大学, 2022.
- [14] Narula, S.C. and Wellington, J.F. (1977) Prediction, Linear Regression and the Minimum Sum of Relative Errors. *Technometrics*, **19**, 185-190. <https://doi.org/10.1080/00401706.1977.10489526>