

异方差模型平均的稀疏化赋权

刘芳琳, 邹晨晨*

青岛大学数学与统计学院, 山东 青岛

收稿日期: 2023年7月26日; 录用日期: 2023年8月18日; 发布日期: 2023年8月24日

摘要

本文针对具有异方差误差的线性回归模型, 在一种可行的Mallows' Cp准则提出的异方差稳健Cp(HRCp)模型平均方法基础上, 考虑HRCp模型平均权重的稀疏性, 使用坐标下降算法对异方差模型平均进行稀疏化赋权。数值模拟表明, 用坐标下降算法改进的HRCp模型平均方法在候选模型较少或者中等、样本量不大时, 相对于其他的模型平均方法有较小的损失风险, 能更好的拟合模型。

关键词

模型平均, 异方差, 稀疏化赋权, 坐标下降算法

Sparsified Weighting of Heteroscedasticity Model Averaging

Fanglin Liu, Chenchen Zou*

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: Jul. 26th, 2023; accepted: Aug. 18th, 2023; published: Aug. 24th, 2023

Abstract

Based on the averaging method of heteroscedasticity robust Cp (HRCp) model proposed by a feasible Mallows' Cp criterion, this paper considers the sparsity of the average weight of the HRCp model, and uses the coordinate-wise descent algorithm to sparse the average weight of the heteroscedasticity model. Numerical simulation shows that the improved HRCp model averaging method with coordinate-wise descent algorithm has less loss risk compared with other model averaging methods when the candidate models are few or medium and the sample size is not large, and can fit the model better.

*通讯作者。

Keywords

Model Averaging, Heteroscedasticity, Sparsified Weighting, Coordinate-Wise Descent Algorithm

Copyright © 2023 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

模型平均是当代统计学和计量经济学界备受瞩目的国际性话题, 它是用一定的权重对来自不同模型的估计或预测进行有效地组合, 从而获得较为稳健的估计, 在经济、金融、生物、医学等领域有广泛的应用前景。模型平均没有忽视模型选择过程产生的不确定性, 估计一般比模型选择更加稳健, 可用减少有用信息的遗失。新近提出的模型平均方法多直接以降低估计或预测的风险作为目标, 因而目标更加明确, 比如渐近最优的模型平均方法。总之, 随着计算机技术的提升, 模型平均作为一种更加复杂的数据挖掘方法正在越来越多被使用。

对于异方差模型, Andrews (1991) [1]提出了异方差模型的模型选择方法, 但没有提供该准则的可行形式。Hansen 和 Racine (2012) [2]提出了 Jackknife model averaging (JMA), 其使用交叉验证标准来选择权重。Liu 和 Ryo Okui (2013) [3]针对具有异方差误差的线性回归模型, 提出了异方差稳健 C_p (HRCp)模型平均方法。Liu (2015) [4]在局部渐近框架中提出了模型平均估计量, 并基于渐近均方误差表达式推导了插入式平均估计量的渐近分布。Ando 和 Li (2014) [5]最近研究了高维回归的模型平均, 其主要特点是模型权重可以在 0 和 1 之间自由变化, 而不受求和为 1 的标准限制, 并建立了一个定理, 用留一交叉验证方法证明了权重选择的渐近最优性。Charkhi, Claeskens 和 Hansen (2016) [6]考虑了一般情况下由极大似然估计得到的估计量, 用其插入估计量替换未知的位置参数, 并建议选择最小化 MSE 估计值的权重。Zhang 等(2016) [7]研究了广义线性模型和广义线性混合效应模型的最优模型平均, 提出了基于 Kullback-Leibler (KL 距离)损失惩罚项的权重选择标准。Zhu 等(2018) [8]针对半参数变系数部分线性模型, 发展了一个分配模型权重的 Mallows 型准则, 并证明了其渐近最优性。Zhao 等(2020) [9]考虑了一个具有乘性异方差的回归模型, 通过结合模型均值和方差函数中未知参数的最大似然估计进行模型平均, 其权重选择标准是基于模型平均估计值平方预测风险的插入式估计值的最小化。Zhang 和 Liu (2022) [10]考虑准似然框架中的模型平均预测, 通过最小化 K 折交叉验证来选择模型平均的最优权重。本文针对异方差模型平均问题, 选取了常用的 HRCp 异方差模型平均方法, 探究其目标函数与 Lasso 的关系, 并对其稀疏解展开具体研究。

Lasso (Least absolute shrinkage and selection operator)方法用模型系数的绝对值函数作为惩罚来压缩模型系数, 使绝对值较小的系数自动压缩为 0, 从而同时实现显著性变量的选择和对对应参数的估计。相较于子集选择和岭回归等传统的 OLS 估计改进方法, Lasso 改进普通最小二乘(Ordinary Least Square, OLS)估计的优势更多一些。近些年来, Lasso 惩罚适用于多种模型, 新开发的计算算法允许将这些模型应用于大型数据集, 利用稀疏性获得统计信息并计算收益, 故有关 Lasso 的工作正在包括统计、工程、数学和计算机科学等许多领域如火如荼地展开。

本文将在 Liu 和 Ryo Okui (2012) [3]针对具有异方差误差的线性回归模型, 采用了一种可行的 Mallows' C_p 准则提出的异方差稳健 C_p (HRCp)模型平均方法基础上, 考虑 HRCp 模型平均权重的稀疏性,

使用坐标下降算法(Coordinate-wise Descent Algorithm, 简称 CD 算法)对异方差模型平均进行稀疏化赋权, 这也是本文的主要创新点。

2. HRCp-Lasso

考虑 n 个独立随机样本 $\{(x_i, y_i), i=1, \dots, n\}$, 其数据产生过程为:

$$y_i = \sum_{j=1}^{\infty} \theta_j x_{ij} + e_i = \mu_i + e_i, i=1, \dots, n \quad (1)$$

其中, y_i 是响应变量, $x_i = (x_{i1}, x_{i2}, \dots)$ 是解释变量, 其具有可数无限个元素, e_i 是随机误差项。令 $Y = (y_1, \dots, y_n)'$, $X = (x'_1, \dots, x'_n)'$, $\mu = (\mu_1, \dots, \mu_n)'$, $\varepsilon = (e_1, \dots, e_n)'$ 。假设 $E(e_i | x_i) = 0$, 故 Y 的条件期望为 $E(Y | X) = \mu$ 。假设 $\text{Var}(e_i | x_i) = \sigma_i^2$, 即 e_i 的条件方差与 x_i 有关, 此时随机误差项的方差是异方差。

考虑总共 M 个近似模型, 其中第 m 个近似模型为:

$$y_i = \sum_{j=1}^{k_m} \theta_{j(m)} x_{ij(m)} + e_i = \mu_i + e_i, i=1, \dots, n \quad (2)$$

即, 使用了 k_m 个 x_{ij} 作为解释变量, 用矩阵可表示为:

$$Y = X_{(m)} \Theta_{(m)} + b_{(m)} + e, m=1, \dots, M \quad (3)$$

其中, $X_{(m)}$ 是一个 $n \times k_m$ 的矩阵, 其第 (i, j) 个元素为 $x_{ij(m)}$, 它是由 X 的任意 k_m 列组成的, $\Theta_{(m)}$ 为相应的系数向量。此时, 第 m 个模型 μ 的最小二乘估计为:

$$\hat{\mu}_m = X_{(m)} \Theta_{(m)} = X_{(m)} \left(X'_{(m)} X_{(m)} \right)^{-1} X'_{(m)} y = P_m y \quad (4)$$

其中, $P_m = X_{(m)} \left(X'_{(m)} X_{(m)} \right)^{-1} X'_{(m)}$ 为投影矩阵。

记模型的权重向量 $w = (w_1, \dots, w_M)'$, 并限制在如下权重集合中:

$$\mathcal{H}_M = \left\{ w \in [0, 1]^M : \sum_{m=1}^M w_m = 1 \right\} \quad (5)$$

则 μ 的模型平均估计为:

$$\hat{\mu}(w) = \sum_{m=1}^M w_m \hat{\mu}_m = \sum_{m=1}^M w_m P_m y = P(w) y \quad (6)$$

根据 Liu 和 Ryo Okui (2012) [10] 提出的异方差稳健 Cp (HRCp) 模型平均方法, HRCp 选择权重 w 是根据极小化下式得到:

$$\text{HRCp}(w) = \left\| Y - \sum_{m=1}^M w_m \hat{\mu}_m \right\|^2 + 2 \sum_{m=1}^M w_m \left(\sum_{i=1}^n \hat{e}_i^2 P_{ii}^{(m)} \right) \quad (7)$$

$$\text{s.t. } w \in \mathcal{H}_M = \left\{ w \in [0, 1]^M : \sum_{m=1}^M w_m = 1 \right\}$$

其中 $\sum_{m=1}^M w_m \hat{\mu}_m$ 是对 μ 的估计, $x_{i(m)}$ 表示 $X_{(m)}$ 的第 i 列, \hat{e}_i 表示残差。令 $Z = (\hat{\mu}_1, \dots, \hat{\mu}_M)_{n \times M}$, 其是一个 $n \times M$ 的矩阵, 它的列是来自 M 个候选模型的拟合响应向量, $P_{ii}^{(m)}$ 是第 m 个模型的帽子矩阵的第 i 个对角线元素。则 HRCp 优化问题可以写成最小化下列式子:

$$\text{HRCp}(w) = \frac{1}{2} \|y - Z w\|_2^2 + \sum_{m=1}^M w_m \left| \sum_{i=1}^n \hat{e}_i^2 P_{ii}^{(m)} \right| \quad (8)$$

$$\text{s.t. } w \in \mathcal{H}_M = \left\{ w \in [0, 1]^M : \sum_{m=1}^M w_m = 1 \right\}$$

而对于经典的线性模型 $y_{n \times 1} = Z_{n \times M} w_{M \times 1} + \varepsilon_{n \times 1}$, 我们可以利用 Lasso 方法找到一个 w 使得下式达到最

小:

$$\begin{aligned} \widetilde{\text{HRCp}}(w) &= \frac{1}{2} \|y - Zw\|_2^2 + \sum_{m=1}^M \lambda_m |w_m| \\ \text{s.t. } w &\in \mathcal{H}_M = \left\{ w \in [0, 1]^M : \sum_{m=1}^M w_m = 1 \right\} \end{aligned} \quad (9)$$

其中 $\lambda_i (i=1, \dots, M)$ 是不同 w 的惩罚向量。

对比上述两个式子, 可以看出 **HRCp** 的权重估计问题与 **Lasso** 之间的对应关系, 其中回归系数为权重 (w) , 惩罚权重 $\lambda_m = \hat{e}^2 p^{(m)} = \sum_{i=1}^n \hat{e}_i^2 p_{ii}^{(m)}$ 。因此, 我们可以将 **HRCp** 看作是带有约束条件的 **Lasso** 问题, 并可以使用 **Lasso** 问题的相关算法来求解模型平均问题中的权重。

3. 坐标下降算法

许多经典模型平均方法对于权重的求解采用的是二次规划算法, 例如 **MMA**、**JMA** 等。坐标下降法算法(**Coordinate-wise Descent Algorithm**, 简称 **CD** 算法)已经在处理高维数据的稀疏解问题上得到了广泛而成功的应用。如上文所述, 我们可以把异方差模型平均问题看作一个 **Lasso** 问题, 故我们可以利用 **CD** 算法来优化求解过程, 从而更好地对异方差模型平均进行稀疏化赋权。

我们参考 Feng 等(2020) [3]对于 **MMA** 的处理, 将关于 w 的等式限制条件 $\sum_{m=1}^M w_m = 1$ 转换为在目标函数中加上 $\frac{\gamma}{2} (1_M^\top w - 1)^2$, 其中 γ 是一个足够大的数。故我们可以将 **HRCp** 等价转换:

$$\begin{aligned} \min \left\{ f_n(w) = \frac{1}{2} \|y - Zw\|_2^2 + \hat{e}^2 \sum_{m=1}^M |w_m| p^{(m)} + \frac{\gamma}{2} (1_M^\top w - 1)^2 \right\} \\ \text{s.t. } w_m \geq 0, m = 1, 2, \dots, M \end{aligned} \quad (10)$$

其次梯度函数为:

$$\frac{\partial f_n(w)}{\partial w_m} = -\hat{\mu}'_m (y - Zw) + \hat{e}^2 p^{(m)} + \gamma (1'_m w - 1) \quad (11)$$

故我们可以给定一个初始权重 $w^{(0)}$, 令次梯度函数 $\frac{\partial f_n(w)}{\partial w_m} = 0$, 得到 w_m 的权重更新公式:

$$\hat{w}_m = \frac{\left[\hat{\mu}'_m (y - Z_{-m} w_{-m}) - \gamma (1'_m w - 1) - \hat{e}^2 p^{(m)} \right]^+}{\gamma + \hat{\mu}'_m \hat{\mu}_m} \quad (12)$$

其中, Z_{-m} 表示 Z 去掉第 m 列, w_{-m} 表示没有第 m 个元素, $m = 1, 2, \dots, M$ 。坐标下降法具体算法见表 1:

Table 1. Coordinate-wise descent algorithm for HRCp

表 1. HRCp 的坐标下降算法

算法: HRCp 的坐标下降算法(HRCp_CD)

1. 初始化权重 $\hat{w}^{(0)} = \frac{1}{M}$, $m = 1, 2, \dots, M$
2. 令 $k = 1, \dots, K$, 根据上述权重更新公式(12)来更新权重, 迭代第 k 轮的权重记为 $\hat{w}^{(k)}$
3. 当 $\|\hat{w}^{(k)} - \hat{w}^{(k-1)}\| < \varepsilon$ 或 $k = K$ 时停止迭代, 获得此时的 $\hat{w}^{(k)}$

4. 数值模拟

我们将考虑两种情形：一种是嵌套模型；另一种是非嵌套模型。通过比较分析在不同的样本量和变量数的情况下，SAIC、SBIC、MMA、JMA、HRCp 以及 HRCp_CD 模型平均方法的估计风险，从而对本文改进的异模型平均方法进行评价。

估计风险由二次损失函数的均值衡量，即，

$$Risk = \frac{1}{D} \sum_{d=1}^D \left\| \hat{\mu}(\omega)_{(d)} - \mu_{(d)} \right\|^2 \tag{13}$$

其中 d 表示第 d 次产生数据，在模型中，我们令 $D=1000$ 。为了便于比较，我们将每个方法的估计风险除以 HRCp_CD 的估计风险。

4.1. 嵌套模型

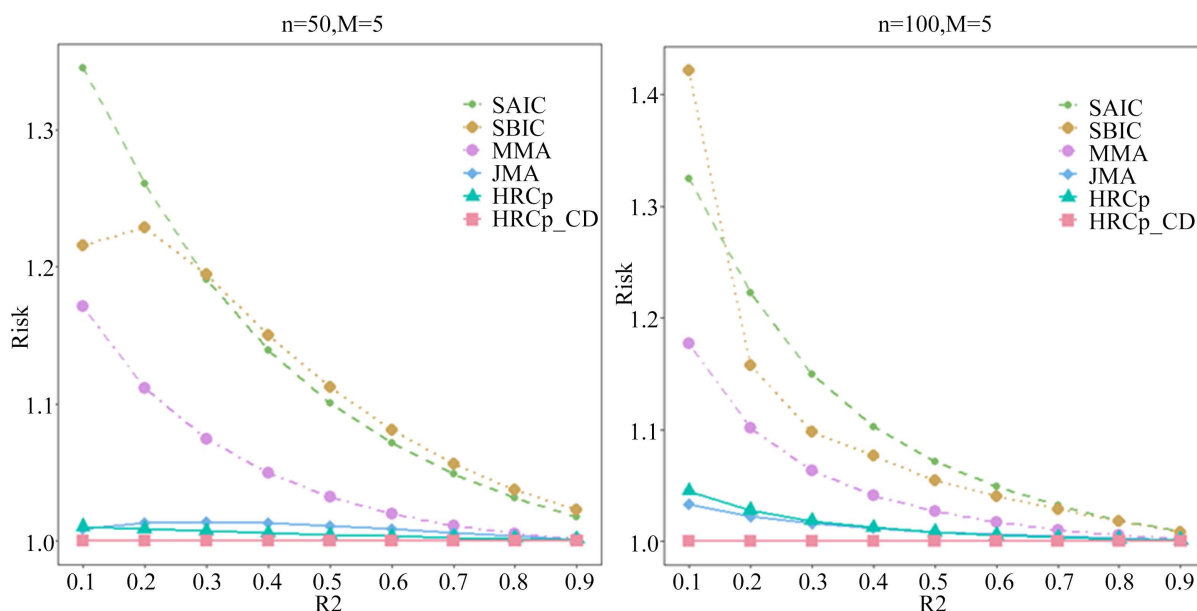
根据以下数据生成过程得到有限随机数据，即在 $k=1000$ 时截断：

$$y_i = \sum_{j=1}^k \theta_j x_{ji} + e_i, i=1, \dots, n \tag{14}$$

其中， $x_{1i} = 1, (x_{2i}, \dots, x_{ki}) \sim N_{k-1}(0, 1), e_i \sim N(0, \sigma_i^2), \sigma_i^2 = x_{2i}^4 + 0.01$ ，故我们保证了方差是严格正定的。令参数 $\theta_j = cj^{-1}, R^2 = (\text{var}(y_i) - \text{var}(e_i)) / \text{var}(y_i) = (\sum_{j=1}^{\infty} \theta_j^2) / (\sum_{j=1}^{\infty} \theta_j^2 + 3.01)$ ，我们可以通过改变 c 的值来控制 R^2 ，使之从 0.1 增加到 0.9。令 HRCp 的坐标下降算法收敛阈值 $\varepsilon = 1 \times 10^{-7}$ ，最大迭代步长 $K = 1 \times 10^5$ ，惩罚项参数 $\gamma = 100$ 。

嵌套模型，即第 $k+1$ 个模型中一定包含第 k 个模型中的所有解释变量。假设解释变量的个数为 p ，则候选模型的数量 $M = p$ ，第 m 个候选模型包括前 m 个解释变量。我们分别令 $n=50, M=5; n=100, M=5; n=100, M=10; n=400, M=15$ ，模拟结果见图 1：

根据图 1 可知，在嵌套模型设置情况下，SAIC 和 SBIC 总体来说表现不好，相对于其他模型平均方法 MMA、JMA、HRCp 以及 HRCp_CD 来说，有较大的损失风险。MMA 在误差项异方差的设置下仅优于 SAIC 和 SBIC，远不如 JMA、HRCp 和 HRCp_CD。这基本符合我们现有的研究认知，因为 JMA、HRCp 和 HRCp_CD 都是针对误差项异方差提出的模型平均方法，而 MMA 是针对误差项同方差提出的，故在



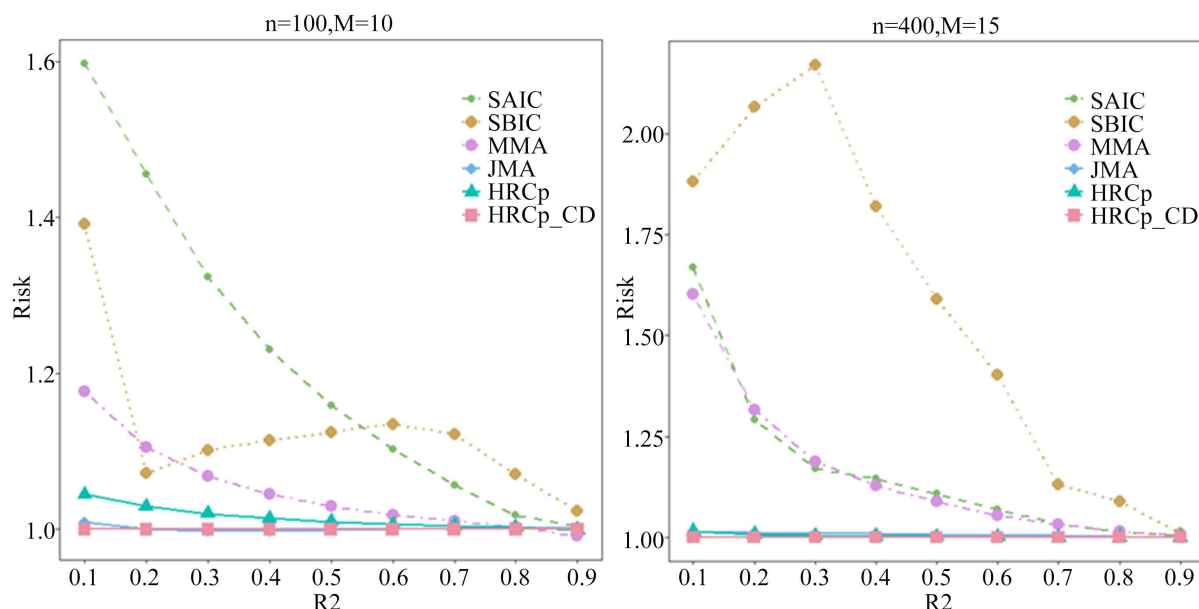


Figure 1. Comparison of various model-averaging methods in the nested setup

图 1. 嵌套模型各模型平均方 s 法的比较

处理异方差模型平均问题上效果比不上针对误差项异方差提出的模型平均方法。当 R^2 较小时, JMA 相对于 HRCp 来说有优势。从整体上看, 在损失风险方面, HRCp_CD 在有限样本中的性能优于其他方法, 且在候选模型数量较小或者中等时, 优势更加明显。此外我们注意到, 当 R^2 较小时 HRCp_CD 与其他模型平均方法的差距要比 R^2 较大时要大, 但整体来看, 本文基于坐标下降算法改进的异方差模型平均方法, 即 HRCp 还是很有优势的。

4.2. 非嵌套模型

非嵌套模型模拟采取线性回归模型, 并且回归变量个数有限。数据产生过程, 具体如下:

$$y_i = \mu_i + e_i = \sum_{j=1}^9 \theta_j x_{ji} + e_i, i=1, \dots, n \quad (15)$$

其中, $x_{1i} = 1$, $(x_{2i}, \dots, x_{ki}) \sim N_{k-1}(0, 1)$, $e_i \sim N(0, \sigma_i^2)$, $\sigma_i = (1 + 6x_{2i}^2)/11$, 令 x_{1i}, x_{2i} 和 x_{3i} 为关键核心变量, 即在每个候选模型中都必须含有, $x_{4i}, x_{5i}, x_{6i}, x_{7i}, x_{8i}$ 和 x_{9i} 为辅助变量, 即可能包含在候选模型中, 也可能不包含在候选模型中。给选定参数 $\theta = c \left(\frac{1}{a}, \frac{1}{a}, \frac{1}{a}, \frac{1}{\sqrt{n}} \left(1, \frac{q-1}{q}, \dots, \frac{1}{q} \right) \right)$, q 是辅助变量的个数。假设每个候选模型中只包含一个辅助变量, 则一共有 6 个候选模型, 每个模型中有 4 个解释变量(3 个核心变量和 1 个辅助变量)。定义为 $R^2 = \frac{\tilde{\theta}'\tilde{\theta}}{1 + \tilde{\theta}'\tilde{\theta}}$, 这里 $\tilde{\theta} = (\theta_2, \theta_3, \dots, \theta_9)'$, 参数 c 用来控制 R^2 , 使 R^2 限定在区间 $[0.1, 0.9]$ 。模拟结果见图 2:

根据图 2 可知, 在非嵌套模型设置情况下, 由于每个候选模型中的解释变量个数均相同, 故 SAIC 和 SBIC 相对于 HRCp_CD 的风险曲线大致重合, 总体来说表现不好, 比其他模型平均方法 MMA、JMA、HRCp 以及 HRCp_CD 有更大的损失风险。MMA 相对于 JMA、HRCp 和 HRCp_CD 来说表现欠佳, 不是处理异方差模型平均问题的首选。当 R^2 和 n 都较小时, JMA 相对于 HRCp 和 HRCp_CD 来说有一定的优势, 但随着 R^2 和 n 的增加, JMA 没有 HRCp_CD 有优势。从整体上看, 在损失风险方面, HRCp_CD 在有限样本中的性能优于其他方法, 且在样本数量相对较小时, 优势更加明显。

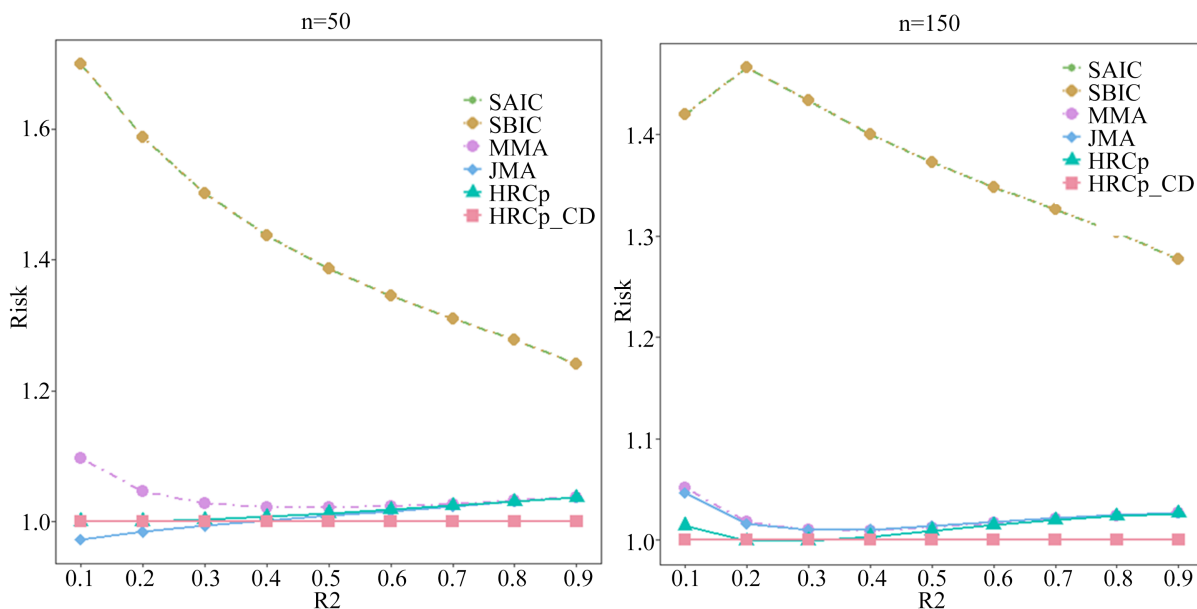


Figure 2. Comparison of various model-averaging methods in the singleton setup
图 2. 非嵌套模型各模型平均方法的比较

5. 实例分析

为了比较在数值模拟部分中提到的六种方法的表现, 我们根据 Zhang [11]将模型平均方法应用于我国的粮食产量预测, 选用了 1990~2021 年的相关数据, 数据来源于中国统计年鉴 (<http://www.stats.gov.cn/sj/ndsj/>)。我们将数据分为两部分, 分别记为训练集和测试集, 其中训练集的样本量为 $n_0 = 10, 20, 30$, 而测试集的样本量为 $n - n_0$, 其中 $n = 32$ 是整体的样本量。即我们分别用 1990~2004 年、1990 年~2009 年、1990 年~2014 年的粮食产量数据预测 2005 年~2021 年、2010~2021 年、2015~2021 年的粮食产量, 并计算样本均方预测误差(Mean Squared Prediction Error, MSPE), 即

$$MSPE = \frac{1}{n - n_0} \sum_{i=n_0+1}^n (y_i - \hat{\mu}_i)^2 \quad (16)$$

其中 $\hat{\mu}_i$ 是 μ_i 的预测值。我们计算了这六种方法的 MSPE, 同时为了方便比较, 我们将 HRCp_CD 的 MSPE 作为分母, 从而更直观地看出本文改进的异方差模型平均方法的有效性。结果见表 2:

Table 2. MSPE for each model averaging method
表 2. 各个模型平均方法的 MSPE

	MSPE1 ($n_0 = 15$)	MSPE2 ($n_0 = 20$)	MSPE3 ($n_0 = 25$)
HRCp_CD	1.000000	1.000000	1.000000
HRCp	1.179368	1.102594	1.071704
JMA	1.150299	1.050705	1.103771
MMA	1.183976	1.092502	1.071346
SBIC	1.281490	1.263368	1.297792
SAIC	1.278936	1.227782	1.241298

根据上述表格, 我们可以看出, 在不同的训练集样本量下, SAIC、SBIC 的预测效果是相对较差的,

HRCp_CD 的预测效果是最好的, MMA、JMA 和 HRCp 的预测效果比 SAIC、SBIC 要好, 但比 HRCp_CD 要差。

与此同时, 我们总结使用各个模型平均方法求出的权重中零权重的比例, 结果见表 3:

Table 3. Zero weights proportion for each model averaging method

表 3. 各个模型平均方法的零权重的比例

	$n_0 = 15$	$n_0 = 20$	$n_0 = 25$
HRCp_CD	0.59375	0.5625	0.53125
HRCp	0.5625	0.5	0.375
JMA	0.375	0.5	0.46875
MMA	0.5	0.40625	0.46875
SBIC	0	0	0
SAIC	0	0	0

根据表 3, 我们可以看出, HRCp_CD 方法得到的各个候选模型权重中零权重的比例最高, 大于 MMA、JMA 和 HRCp 得到的零权重的比例, 使用 SAIC、SBIC 方法一般并不能使模型平均某个候选模型的权重为 0。故说明本文利用坐标下降算法改进的 HRCp 模型平均方法能利用较少的候选模型得到较精确的预测, 即能有效地对异方差模型平均问题进行稀疏化赋权。

6. 总结

本文主要对异方差线性误差模型的平均估计进行了研究, 在现有的异方差模型平均方法 HRCp 的基础上, 将模型平均问题看作一个带约束条件的 Lasso 问题, 并对模型平均的权重进行稀疏化赋权。与此同时, 本文还提出了一种新的算法, 即利用坐标下降算法来对模型平均的权重进行稀疏化赋权。数值模拟和实例研究表明, 用坐标下降算法改进的 HRCp 模型平均方法在候选模型较少或者中等、样本量不大时, 相对于其他的模型平均方法有较小的损失风险, 能更好的拟合模型。此外, 在实际应用过程中非线性模型应用十分广泛, 所以把本文改进的方法推广到非线性模型是非常有意义的, 这需要进一步的研究。

参考文献

- [1] Andrews, D.W.K. (1991) Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, **59**, 817-858. <https://doi.org/10.2307/2938229>
- [2] Hansen, B.E. and Racine, J.S. (2012) Jackknife Model Averaging. *Journal of Econometrics*, **167**, 38-46. <https://doi.org/10.1016/j.jeconom.2011.06.019>
- [3] Liu, Q.F and Okui, R. (2013) Heteroscedasticity-Robust CP Model Averaging. *The Econometrics Journal*, **16**, 463-472. <https://doi.org/10.1111/ectj.12009>
- [4] Liu, C.A. (2015) Distribution Theory of the Least Squares Averaging Estimator. *Journal of Econometrics*, **186**, 142-159. <https://doi.org/10.1016/j.jeconom.2014.07.002>
- [5] Ando, T. and Li, K.C. (2014) A Model-Averaging Approach for Highdimensional Regression. *Journal of the American Statistical Association*, **109**, 254-265. <https://doi.org/10.1080/01621459.2013.838168>
- [6] Charkhi, A., Claeskens, G. and Hansen, B.E. (2016) Minimum Mean Squared Error Model Averaging in Likelihood Models. *Statistica Sinica*, **26**, 809-840. <https://doi.org/10.5705/ss.202014.0067>
- [7] Zhang, X.Y., Yu, D.L., Zou, G.H. and Liang, H. (2016) Optimal Model Averaging Estimation for Generalized Linear Models and Generalized Linear Mixed-Effects Models. *Journal of the American Statistical Association*, **111**, 1775-1790. <https://doi.org/10.1080/01621459.2015.1115762>
- [8] Zhu, R., Wan, T.K.A., Zhang, X.Y. and Zou, G.H. (2019) A Mallows-Type Model Averaging Estimator for the Varying-Coefficient Partially Linear Model. *Journal of the American Statistical Association*, **114**, 882-892.

- <https://doi.org/10.1080/01621459.2018.1456936>
- [9] Zhao, S.W, Ma, Y.Y., Wan, A.T.K., Zhang, X.Y. and Wang, S.Y. (2020) Model Averaging in a Multiplicative Heteroscedastic Model. *Econometric Reviews*, **39**, 1100-1124. <https://doi.org/10.1080/07474938.2020.1770995>
- [10] Zhang, X. and Liu, C.A. (2023) Model Averaging Prediction by K-Fold Cross-Validation. *Journal of Econometrics*, **235**, 280-301. <https://doi.org/10.1016/j.jeconom.2022.04.007>
- [11] Zhang, X.Y. and Zou, G.H. (2011) Model Averaging Method and Its Application in Forecast. *Statistical Research*, **28**, 98-103.