

响应变量缺失下可伸缩的模型平均估计

马 跃, 黄 彬*, 刘淑洁

北京化工大学数理学院, 北京

收稿日期: 2024年4月29日; 录用日期: 2024年5月22日; 发布日期: 2024年5月31日

摘 要

当响应变量缺失时, 本文研究了线性模型的可伸缩的模型平均估计问题。基于逆概率加权方法和奇异值分解对原模型进行转换, 所构建的模型平均估计最多只需要选择 p 个候选模型的权重。利用Jackknife准则选择权重, 且权重之和不必限制为1。在一定的条件下证明了所提方法在实现最小二次损失意义下是渐近最优的, 并通过数值模拟进一步验证了所提方法在预测效率和计算成本方面的优良性。

关键词

可伸缩的, 模型平均, 缺失的响应变量, 奇异值分解, 渐近最优性

Scalable Model Averaging Estimation with Missing Responses

Yue Ma, Bin Huang*, Shujie Liu

School of Mathematics and Physics, Beijing University of Chemical Technology, Beijing

Received: Apr. 29th, 2024; accepted: May 22nd, 2024; published: May 31st, 2024

Abstract

In this paper, a scalable model averaging method is developed for the linear regression models with response missing. By using inverse probability weighted method and the singular value decomposition to transform the original models, this method enables us to find the optimal weights by considering at most p candidate models. The weights can be selected by Jackknife criterion, and the sum of weights is not necessarily restricted to one. Under some mild conditions, it is shown that the proposed method is asymptotically optimal in the sense of achieving the lowest possible squared error. Some simulation studies are conducted to illustrate the superiority of the proposed method in terms of both predictive efficiency and computational cost.

*通讯作者。

Keywords

Scalable, Model Averaging, Missing Responses, Singular Value Decomposition, Asymptotic Optimality

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

模型平均是当代统计分析中用于处理模型不确定性和提高预测精度的一种流行且有效的方法。其中，针对频率模型平均(Frequency model Averaging, FMA)的方法和理论研究获得了广泛的关注。Hansen [1]开创性地提出了 Mallows 模型平均(MMA)方法，该方法利用极小化 Mallows 准则来选择嵌套模型的权重，并证明了该方法的渐近最优性。Wan 等[2]将 MMA 方法拓展到连续权重和非嵌套模型的情形。当随机误差项为异方差时，Hansen 和 Racine [3]，Liu 和 Okui [4]分别提出了 JMA (Jackknife Model Averaging)和 HRCp (Heteroscedasticity Robust Cp)模型平均方法，Zhang 等[5]进一步将 JMA 方法拓展到相关数据的情形。基于这些重要工作，很多学者将这些方法扩展到不同的数据类型和不同的模型中。

其中，缺失数据是在很多实际问题研究中经常会遇到的数据类型，因此缺失数据下模型平均估计也得到很多的关注和研究。当响应变量缺失时，Wei 和 Wang [6]，Wei 等[7]提出了线性模型的模型平均估计，并分别利用 JMA 和 HRCp 准则选择权重。Zeng 等[8]扩展到变系数偏线性模型的情形。但是，当协变量的个数 p 较大时，上述方法可选的候选模型的个数可高达 2^p ，这给每个候选模型的估计和权重的选择都带来了极大的困难和挑战。

为了提高估计效率和计算速度，Zhu 等[9]提出了可伸缩的模型平均估计，该方法利用奇异值分解对模型进行转换，所构建的模型平均估计只需要估计 p 个权重，且其最小二次损失在渐近意义下不差于传统的模型平均估计。因此，受[9]的启发，本文考虑了线性模型在响应变量缺失下的可伸缩的模型平均估计。在响应变量随机缺失(missing at random, MAR)的前提下，我们基于逆概率加权方法和奇异值分解对模型进行转换，使得转换后的模型只需估计 p 个候选模型，并通过极小化 Jackknife 准则选择 p 个权重，从而有效地减轻了计算压力，提高了模型平均的预测效率。而且，在一定的条件下证明了所提模型平均估计方法的渐近最优性，并通过数值模拟实验进一步验证了所提方法的优良性。

2. 模型及估计

2.1. 模型假设和估计方法

考虑线性回归模型：

$$y_i = \mu_i + e_i = \mathbf{x}_i^T \boldsymbol{\theta} + e_i = \sum_{j=1}^p \theta_j x_{ij} + e_i, i=1, \dots, n, \quad (1)$$

其中 y_i 是响应变量， $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ 是预测变量， $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ 是回归系数， e_1, \dots, e_n 是相互独立且为异方差的随机误差，满足 $E(e_i | x_i) = 0$ ， $E(e_i^2 | x_i) = \sigma_i^2$ 。我们考虑 $p < n$ 的情形。这里假定预测变量 x_i 的值可以被完全观测，响应变量 y_i 是随机缺失的，令 δ_i 为缺失示性变量，即当 y_i 缺失时 $\delta_i = 0$ ，否则 $\delta_i = 1$ ，且缺失机制满足 MAR 条件，即：

$$P(\delta_i = 1 | y_i, \mathbf{x}_i) = P(\delta_i = 1 | \mathbf{x}_i) = \pi(\mathbf{x}_i)。$$

假设 $\{(y_i, \mathbf{x}_i, \delta_i), i = 1, \dots, n\}$ 是独立同分布的。当选择概率函数 $\pi(\mathbf{x}_i)$ 已知时, 记 $z_{\pi,i} = \delta_i y_i / \pi(\mathbf{x}_i)$, 在 MAR 假设下, 易得 $E(z_{\pi,i} | \mathbf{x}_i) = \mu_i = \mathbf{x}_i^T \boldsymbol{\theta}$, 因此有

$$z_{\pi,i} = \mathbf{x}_i^T \boldsymbol{\theta} + e_{\pi,i} = \sum_{j=1}^p \theta_j x_{ij} + e_{\pi,i}, i = 1, \dots, n, \tag{2}$$

其中 $e_{\pi,i} = \delta_i (\mu_i + e_i) / \pi(\mathbf{x}_i) - \mu_i$, $E(e_{\pi,i} | \mathbf{x}_i) = 0$, $E(e_{\pi,i}^2 | \mathbf{x}_i) = \sigma_{\pi,i}^2 = (\sigma_i^2 + \mu_i^2) / \pi(\mathbf{x}_i) - \mu_i^2$ 。模型(2)可表示为矩阵形式:

$$\mathbf{z}_{\pi} = \boldsymbol{\mu} + \mathbf{e}_{\pi} = \mathbf{X} \boldsymbol{\theta} + \mathbf{e}_{\pi}, \tag{3}$$

其中 $\mathbf{z}_{\pi} = (z_{\pi,1}, \dots, z_{\pi,n})^T$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$, $\mathbf{e}_{\pi} = (e_{\pi,1}, \dots, e_{\pi,n})^T$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ 。

对数据生成过程(1), 考虑 M 个候选模型, 相应地, 对(3)式, 也有 M 个候选模型:

$$\mathbf{z}_{\pi} = \mathbf{X}_m \boldsymbol{\theta}_m + \mathbf{e}_{\pi}, m = 1, \dots, M, \tag{4}$$

其中 \mathbf{X}_m 第 m 个模型对应的 $n \times k_m$ 阶预测矩阵, $\boldsymbol{\theta}_m$ 是 k_m 维回归系数, 记 S_m 为该模型包含的预测变量的指标集, 则 S_m 的维度为 k_m 。

由于 $\pi(\mathbf{x}_i)$ 一般都是未知的, 故这里对 $\pi(\mathbf{x}_i)$ 假设一个参数模型 $\bar{\pi}(\mathbf{x}_i; \boldsymbol{\alpha})$, 其中 $\bar{\pi}(\cdot; \boldsymbol{\alpha})$ 是已知的函数, $\boldsymbol{\alpha}$ 是未知的参数向量。记 $\hat{\boldsymbol{\alpha}}$ 为 $\boldsymbol{\alpha}$ 的一个估计量, 如极大似然估计量, 记 $\hat{\pi}(\mathbf{x}_i) = \bar{\pi}(\mathbf{x}_i; \hat{\boldsymbol{\alpha}})$ 。因此, 当需要估计 $\pi(\mathbf{x}_i)$ 时, 用 $\hat{\pi}(\mathbf{x}_i)$ 替代上述模型(2)~(4)中的 $\pi(\mathbf{x}_i)$, 相应地, 记号的下标 π 替换成 $\hat{\pi}$, 如: $z_{\hat{\pi},i} = \delta_i y_i / \hat{\pi}(\mathbf{x}_i)$, $\mathbf{z}_{\hat{\pi}} = (z_{\hat{\pi},1}, \dots, z_{\hat{\pi},n})^T$, $\mathbf{e}_{\hat{\pi}} = (e_{\hat{\pi},1}, \dots, e_{\hat{\pi},n})^T$ 等。

对第 m 个模型, 利用最小二乘(OLS)估计可得到 $\boldsymbol{\theta}_m, \boldsymbol{\mu}$ 的估计 $\tilde{\boldsymbol{\theta}}_{\hat{\pi},m} = (\mathbf{X}_m^T \mathbf{X}_m)^{-1} \mathbf{X}_m^T \mathbf{z}_{\hat{\pi}}$, $\tilde{\boldsymbol{\mu}}_{\hat{\pi},m} = \mathbf{P}_m \mathbf{z}_{\hat{\pi}}$, 其中 $\mathbf{P}_m = \mathbf{X}_m (\mathbf{X}_m^T \mathbf{X}_m)^{-1} \mathbf{X}_m^T$, $m = 1, \dots, M$ 。令 $\tilde{\mathbf{w}} = (\tilde{w}_1, \dots, \tilde{w}_M)^T$ 为集合 $\tilde{W}^M = \{\tilde{\mathbf{w}} \in [0, 1]^M : \sum_{m=1}^M \tilde{w}_m = 1\}$ 中的权重向量, 可得 $\boldsymbol{\mu}$ 的一个模型平均估计 $\tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\tilde{\mathbf{w}}) = \sum_{m=1}^M \tilde{w}_m \tilde{\boldsymbol{\mu}}_{\hat{\pi},m} = \mathbf{P}(\tilde{\mathbf{w}}) \mathbf{z}_{\hat{\pi}}$, 其中 $\mathbf{P}(\tilde{\mathbf{w}}) = \sum_{m=1}^M \tilde{w}_m \mathbf{P}_m$ 。进一步, 利用 Jackknife 准则[3]来选择权重, 令

$$\tilde{J}_{\hat{\pi}}(\tilde{\mathbf{w}}) = \|\mathbf{z}_{\hat{\pi}} - \tilde{\mathbf{P}}(\tilde{\mathbf{w}}) \mathbf{z}_{\hat{\pi}}\|^2,$$

其中 $\tilde{\mathbf{P}}(\tilde{\mathbf{w}}) = \sum_{m=1}^M \tilde{w}_m \tilde{\mathbf{P}}_m$, $\tilde{\mathbf{P}}_m = \mathbf{D}_m (\mathbf{P}_m - \mathbf{I}_n) + \mathbf{I}_n$, \mathbf{I}_n 为 n 阶单位矩阵, \mathbf{D}_m 为对角元素为 $(1 - p_{m,ii})^{-1}$ 的 n 阶对角阵, $p_{m,ii}$ 为 \mathbf{P}_m 的第 i 个对角线元素。通过极小化 Jackknife 准则得到权重向量 $\tilde{\mathbf{w}}_{\hat{\pi}} = \arg \min_{\tilde{\mathbf{w}} \in \tilde{W}^M} \tilde{J}_{\hat{\pi}}(\tilde{\mathbf{w}})$, 进一步得到 $\boldsymbol{\mu}$ 的响应变量缺失下 Jackknife Model Averaging (M-JMA)估计 $\tilde{\boldsymbol{\mu}}_{\hat{\pi}}(\tilde{\mathbf{w}}_{\hat{\pi}})$ 。由文献[3] [6], 在一定的条件下, 该 M-JMA 估计在实现最小二次损失意义下是渐近最优的, 因此, 该方法得到了广泛的应用, 且有很好的预测性能。

但是, 当协变量的个数 p 较大时, 可选的候选模型的个数 M 可高达 2^p , 这给每个模型的估计和权重的选择都带了极大的困难和挑战, 从而降低了估计效率和计算速度。因此, 我们需要尝试新的方法去改进上述模型平均估计。参考文献[9], 本文提出了响应变量缺失下可伸缩的模型平均估计, 基于逆概率加权和奇异值分解(SVD)将模型进行转换, 转换后所构建的模型平均估计只需估计 p 个一元回归模型和选择 p 个权重, 从而有效地减轻了计算压力, 提高了预测精度。

假设 $\text{rank}(\mathbf{X}) = p$, 由奇异值分解(SVD), $\mathbf{X} = \mathbf{U} \mathbf{D} \mathbf{V}^T$, 使得 $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}_p$, 其中 \mathbf{U} 是 $n \times p$ 阶的列正交矩阵, \mathbf{V} 是 p 阶正交矩阵, \mathbf{D} 是对角线为非负实数的 p 阶对角矩阵。相应地, \mathbf{X}_m 可分解为 $\mathbf{X}_m = \mathbf{U}_m \mathbf{D}_m \mathbf{V}_m^T$, 则候选模型可表示为

$$\mathbf{z}_{\hat{\pi}} = \mathbf{U}_m \boldsymbol{\beta}_m + \mathbf{e}_{\hat{\pi}}, m = 1, \dots, M, \tag{5}$$

其中 $\beta_m = D_m V_m^T \theta_m$ ，且 β_m 的 OLS 估计为

$$\hat{\beta}_{\hat{\pi},m} = (U_m^T U_m)^{-1} U_m^T z_{\hat{\pi}} = U_m^T z_{\hat{\pi}} = (\hat{\beta}_{\hat{\pi},j_1}, \dots, \hat{\beta}_{\hat{\pi},j_{k_m}})^T,$$

其中 j_1, \dots, j_{k_m} 是 S_m 中的元素， $\hat{\beta}_{\hat{\pi},j} = (u_{(j)}^T u_{(j)})^{-1} u_{(j)}^T z_{\hat{\pi}} = u_{(j)}^T z_{\hat{\pi}}$ 是一元回归模型 $z_{\hat{\pi}} = u_{(j)} \beta_j + e_{\hat{\pi}}$ 的 OLS 估计， $u_{(j)}$ 是 U 的第 j 列向量， $1 \leq j \leq p$ ，即 β_m 可以通过 $z_{\hat{\pi}}$ 关于 U_m 的每一列做一元线性回归进行估计。从而， μ 可估计为 $\hat{\mu}_{\hat{\pi},m} = U_m \hat{\beta}_{\hat{\pi},m} = \sum_{j \in S_m} u_{(j)} \hat{\beta}_{\hat{\pi},j}$ ，故 μ 的模型平均估计为

$$\hat{\mu}_{\hat{\pi}}(\tilde{w}) = \sum_{m=1}^M \tilde{w}_m \hat{\mu}_{\hat{\pi},m} = \sum_{m=1}^M \tilde{w}_m \left(\sum_{j \in S_m} u_{(j)} \hat{\beta}_{\hat{\pi},j} \right) = \sum_{j=1}^p \left(\sum_{m=1}^M \tilde{w}_m I_{S_m}(j) \right) u_{(j)} \hat{\beta}_{\hat{\pi},j}, \quad (6)$$

其中 $I_{S_m}(j)$ 为示性函数，即当 $j \in S_m$ 时 $I_{S_m}(j) = 1$ ，否则为 0。若定义 $w_j = \sum_{m=1}^M \tilde{w}_m I_{S_m}(j) \in [0,1]$ 为权重，则式(6)可看成为 p 个一元回归模型的预测平均。

令 $w = (w_1, \dots, w_p)^T$ 为集合 $W = \{w_j \in [0,1]: 1 \leq j \leq p\}$ 中的权重向量，构造 μ 的模型平均估计量

$$\hat{\mu}_{\hat{\pi}}(w) = \sum_{j=1}^p w_j \hat{\mu}_{(j)} = \sum_{j=1}^p w_j u_{(j)} \hat{\beta}_{\hat{\pi},j} = H(w) z_{\hat{\pi}},$$

其中 $w \in W$ ， $\hat{\mu}_{(j)} = u_{(j)} \hat{\beta}_{\hat{\pi},j}$ ， $H(w) = \sum_{j=1}^p w_j H_j$ ， $H_j = u_{(j)} u_{(j)}^T$ ，注意权重之和 $\sum_{j=1}^p w_j$ 无须等于 1。进一步，利用 Jackknife 准则[3]来选择权重 w ，令

$$J_{\hat{\pi}}(w) = \left\| z_{\hat{\pi}} - \hat{H}(w) z_{\hat{\pi}} \right\|^2,$$

其中 $\hat{H}(w) = \sum_{j=1}^p w_j \hat{H}_j$ ， $\hat{H}_j = A_j (H_j - I_n) + I_n$ ， $h_{j,ii}$ 为 H_j 的第 i 个对角线元素， A_j 为对角元素为 $(1 - h_{j,ii})^{-1}$ 的 n 阶对角阵。通过极小化 Jackknife 准则得到权重向量 $\hat{w}_{\hat{\pi}} = \arg \min_{w \in W} J_{\hat{\pi}}(w)$ ，进一步得到 μ 的响应变量缺失下 Scalable Jackknife Model Averaging (M-SJMA)估计 $\hat{\mu}_{\hat{\pi}}(\hat{w}_{\hat{\pi}})$ 。该估计量一个直接的好处是，所选权重个数完成了从 2^p 到 p 的大幅缩小，极大地降低计算成本，减少计算中的数值误差，从而提高模型平均的预测效率，这在第 3 节的数值模拟结果中得到了很好的验证。

2.2. 渐近性质

本小节将讨论所提出的 M-SJMA 估计的渐近性质。为了说明的方便，先引入一些记号和定义。记 Θ_{α} 为 α 的参数空间， $l(\alpha) = E[\delta \log \bar{\pi}(x; \alpha) + (1 - \delta) \log(1 - \bar{\pi}(x; \alpha))]$ 。定义二次损失函数 $\tilde{L}_{\pi}(\tilde{w}) = \|\hat{\mu}_{\pi}(\tilde{w}) - \mu\|^2$ ， $\tilde{w} \in \tilde{W}^M$ ， $L_{\pi}(w) = \|\hat{\mu}_{\pi}(w) - \mu\|^2$ ， $w \in W$ ，相应的风险函数 $R_{\pi}(w) = E[L_{\pi}(w) | X]$ ，其中 $\hat{\mu}_{\pi}(w) = H(w) z_{\pi}$ 。记 $\xi_{\pi} = \inf_{w \in W} R_{\pi}(w)$ 。

接下来，给出证明渐近性质所需要的条件。需要指出的是，所有的大样本条件和结论均是在 $n \rightarrow \infty$ 的意义上的。

(C1) Θ_{α} 是有界闭集， $l(\alpha)$ 在 Θ_{α} 中关于 α 有唯一的极大值点 α_0 ，且 α_0 为 Θ_{α} 的内点。 $\bar{\pi}(x; \alpha)$ 关于 α 二阶连续可微，且 $E\left[\left. \frac{\partial \bar{\pi}(x; \alpha)}{\partial \alpha} \cdot \frac{\partial \bar{\pi}(x; \alpha)}{\partial \alpha} \right|_{\alpha = \alpha_0}\right]$ 为正定矩阵。 $\inf_{\alpha \in \Theta_{\alpha}} \bar{\pi}(x; \alpha) > 0$ 。

(C2) 存在常数 c_1, c_2 ，使得 $\max_{1 \leq i \leq n} E(e_i^4) \leq c_1$ 和 $\max_{1 \leq i \leq n} |\mu_i| \leq c_2$ 几乎处处成立。

(C3) $\xi_{\pi}^{-1} p \xrightarrow{a.s.} 0$ 。

(C4) 存在常数 c_3 ，使得 $\max_{i,j} u_{ij}^2 \leq c_3 n^{-1}$ ，其中 u_{ij} 为矩阵 U 中的元素。

这里，条件(C1)是极大似然估计量 $\hat{\alpha}$ 具有相合性和渐近正态性所需要的条件。条件(C2)~(C4)是模型平均估计中常用的条件，比如文献[5] [9]中都有类似的条件。

定理 1: 在条件(C1)~(C3)下, 当 $n \rightarrow \infty$ 时, 有 $\frac{\inf_{w \in W} L_{\hat{\pi}}(w)}{\inf_{\tilde{w} \in \tilde{W}^{2^p}} \tilde{L}_{\hat{\pi}}(\tilde{w})} \xrightarrow{P} 1$, 其中 \tilde{W}^{2^p} 是当 $M = 2^p$ 时的权重向量集合 \tilde{W}^M 。

定理 1 表明, 可伸缩的模型平均方法与传统的 2^p 个权重的模型平均方法相比, 它们的最小二次损失是渐近相等的。

定理 2: 在条件(C1)~(C4)下, 当 $n \rightarrow \infty$ 时, 有 $\frac{L_{\hat{\pi}}(\hat{w}_{\hat{\pi}})}{\inf_{w \in W} L_{\hat{\pi}}(w)} \xrightarrow{P} 1$ 。

定理 3: 在条件(C1)~(C4)下, 当 $n \rightarrow \infty$ 时, 有 $\frac{L_{\hat{\pi}}(\hat{w}_{\hat{\pi}})}{\inf_{\tilde{w} \in \tilde{W}^{2^p}} \tilde{L}_{\hat{\pi}}(\tilde{w})} \xrightarrow{P} 1$ 。

根据定理 1 和定理 2 的结论, 定理 3 显然成立。该定理表明被选出的 p 个权重 $\hat{w}_{\hat{\pi}}$ 所对应的二次损失渐近等价于不可获得的 2^p 个最优权重所对应的二次损失, 这说明所提出 M-SJMA 估计 $\hat{\mu}_{\hat{\pi}}(\hat{w}_{\hat{\pi}})$ 在实现最小二次损失意义下是渐近最优的。

3. 数值模拟

本节将通过数值模拟来检验所提 M-SJMA 方法的有限样本性质。为了更好地评估, 我们选取另外两种方法作为比较对象。第一种方法是基于完整数据(Complete Case, CC)分析的 SJMA 方法([9]), 记此方法为 CC-SJMA。第二种方法是响应变量缺失下的 M-JMA 方法([3] [6])。

考虑数据生成过程为

$$y_i = \mu_i + e_i = \sum_{j=1}^p \theta_j x_{ij} + e_i,$$

其中 $p=5$, $\theta = (0.3, 0.6, 0.3, 0.3, 0.3)^T$, $x_{i1} = 1$, 其余的 x_{ij} 是独立同分布的标准正态随机变量, $e_i \sim 0.5N(0, \sigma^2) + 0.5N(0, 3\sigma^2)$, $i=1, \dots, n$, $j=2, 3, 4, 5$ 。通过选取合适的 σ^2 将 R^2 控制在集合 $\{0.1, 0.2, \dots, 0.9\}$ 上变动, 其中:

$$R^2 = \frac{n^{-1} \sum_{i=1}^n (\mu_i - \bar{\mu})^2}{\sigma^2 + n^{-1} \sum_{i=1}^n (\mu_i - \bar{\mu})^2}, \bar{\mu} = n^{-1} \sum_{i=1}^n \mu_i。$$

选择概率函数 $\pi(x_i)$ 设定为

$$\pi(x_i) = \Phi(\alpha_1 + \alpha_2 x_{i2}),$$

其中 $\Phi(\cdot)$ 为标准正态分布的分布函数。选取两种不同的 α 的值: $\alpha = (0.6, 0.5)^T$ 和 $\alpha = (0, 0.5)^T$, 使得相应的平均缺失率分别大约为 0.3 和 0.5。对 M-JMA 方法, 考虑所有包含截距项的候选模型, 故共有 $M = 2^5$ 个候选模型。

为了评估上述三种方法的表现, 采用如下的均方误差(Mean Square Error, MSE)作为度量标准, 定义如下:

$$MSE = \sum_{d=1}^D \left\| \mu^{(d)}(w) - \mu^{(d)} \right\|^2 / D,$$

其中 D 是试验的重复次数, $\mu^{(d)}$ 表示 μ 在第 d 次试验中的值, $\mu^{(d)}(w)$ 是相应于 $\mu^{(d)}$ 的模型平均估计。重复次数 D 设定为 200, 考虑的样本量为 $n = 50, 150$ 。为了便于比较, 考虑基于完整数据的 SJMA 方法, 并将其视为“黄金标准”用于比较中, 记此方法为 CD-SJMA。我们分别计算了 CC-SJMA, M-JMA 和所提 M-SJMA 方法的 MSE 除以 CD-SJMA 方法的 MSE 的比值, 并报告最终的比值, 模拟结果如图 1 所示。

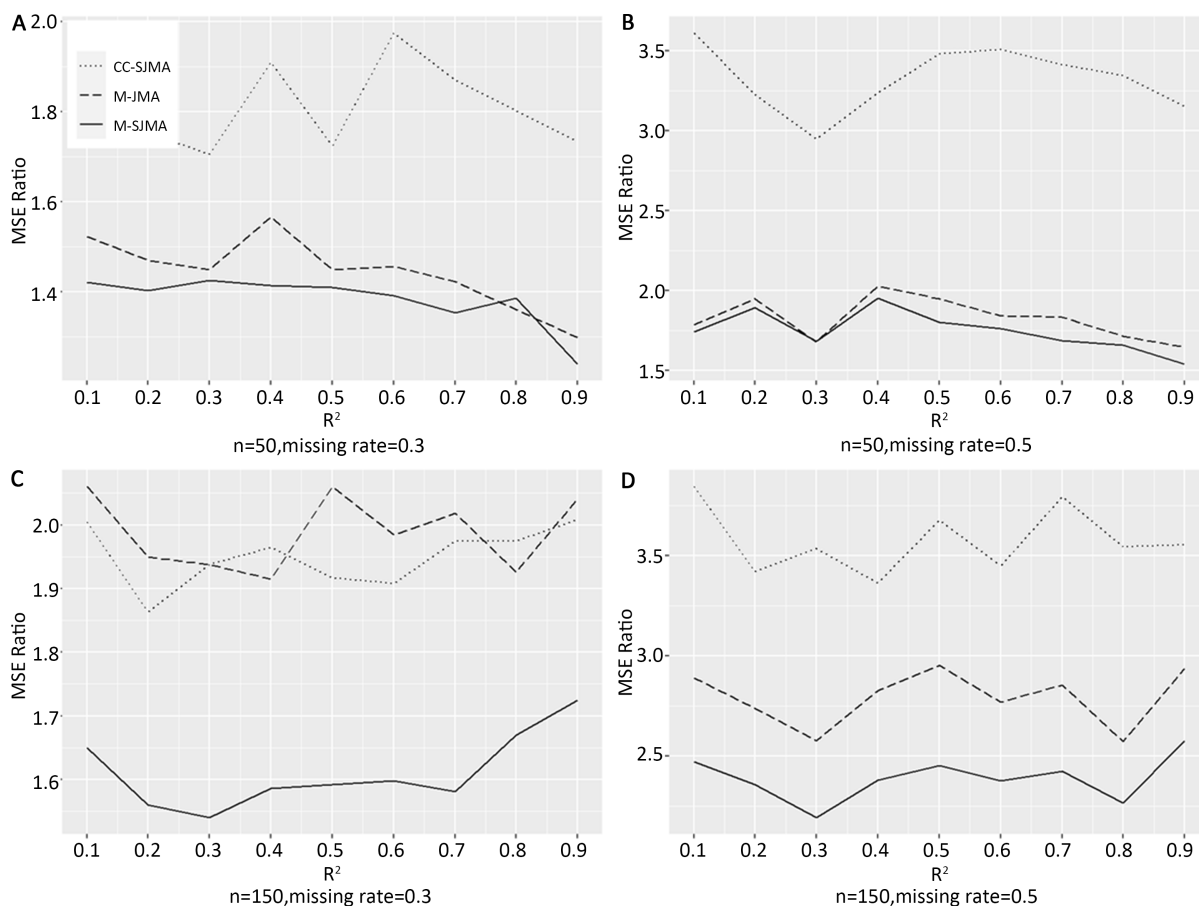


Figure 1. The MSE ratios of different model averaging estimators
图 1. 不同模型平均估计所对应的 MSE 比值

Table 1. The computational times (seconds) of different model averaging estimators with different feature dimension p
表 1. 不同特征维度 p 下模型平均估计的计算时间(秒)

| | p | 5 | 7 | 9 | 10 | 11 | 13 | 20 | 25 |
|-----------|---------|------|-------|-------|-------|--------|----------|--------|--------|
| $n = 50$ | M-JMA | 0.00 | 0.03 | 1.91 | 14.11 | 104.14 | 7899.01 | – | – |
| | M-SJMA | 0.11 | 0.19 | 0.26 | 0.31 | 0.51 | 0.68 | 1.81 | 2.69 |
| | CC-SJMA | 0.09 | 0.17 | 0.25 | 0.42 | 0.35 | 0.62 | 1.65 | 2.70 |
| $n = 300$ | M-JMA | 0.01 | 0.08 | 3.08 | 24.50 | 219.10 | 18127.52 | – | – |
| | M-SJMA | 7.78 | 19.54 | 34.29 | 38.03 | 47.61 | 67.86 | 203.68 | 273.28 |
| | CC-SJMA | 8.77 | 19.57 | 33.24 | 30.11 | 54.82 | 63.09 | 191.28 | 318.67 |

从图 1 中可以看出，本文所提出的 M-SJMA 方法的 MSE 在绝大多数情况下都是最小的。在考虑的所有情况下，M-SJMA 的表现明显优于 CC-SJMA 的表现。并且在大部分情况下，M-SJMA 都要比 M-JMA 表现更出色，特别当样本量较大时，如 $n = 150$ 。这表明随着样本量的增大，M-JMA 会因为计算成本增加产生更多的计算误差，从而导致预测精度降低。与 M-JMA 相比，CC-SJMA 的表现大多数都很差。但是当缺失率较低，样本量较大(如 $n = 150$ ，缺失率 = 0.3)时，M-JMA 可能比 CC-SJMA 表现更差。这也说明当计算量增大时，可伸缩的模型平均方法在预测精度上是优于传统的模型平均方法的。

另外, 针对不同的预测变量个数 p 和样本量 n , 我们比较了三种方法的计算时间, 结果见表 1。从表 1 中可以明显看出, 随着 p 的增大, M-JMA 的计算时间呈指数级增长, 特别当 p 大于 13 时, 由于权重向量维度过大, 计算机无法进行优化计算。但是, 对可伸缩的模型平均方法, 如 M-SJMA 和 CC-SJMA, 它们的计算时间都非常短。即使当 $n = 300$, $p = 25$ 时, M-SJMA 和 CC-SJMA 的计算时间都小于 273.3 和 318.7 秒。

总之, 模拟结果表明所提 M-SJMA 方法的有限样本性质要优于另外两种对比的方法。M-SJMA 方法在预测精度和计算成本方面均表现优越。

4. 结论

基于逆概率加权, 奇异值分解和 Jackknife 准则, 本文研究了响应变量缺失下线性模型的可伸缩的模型平均估计问题。在理论上证明了所提 M-SJMA 方法的渐近最优性, 并通过数值模拟进一步验证了所提方法的有限样本性质。未来可考虑将该方法推广到半参数模型的情况。

参考文献

- [1] Hansen, B.E. (2007) Least Squares Model Averaging. *Econometrica*, **75**, 1175-1189. <https://doi.org/10.1111/j.1468-0262.2007.00785.x>
- [2] Wan, A.T., 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>
- [3] 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>
- [4] 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>
- [5] Zhang, X., Wan, A.T. and Zou, G. (2013) Model Averaging by Jackknife Criterion in Models with Dependent Data. *Journal of Econometrics*, **174**, 82-94. <https://doi.org/10.1016/j.jeconom.2013.01.004>
- [6] Wei, Y. and Wang, Q. (2021) Cross-Validation-Based Model Averaging in Linear Models with Response Missing at Random. *Statistics & Probability Letters*, **171**, Article ID: 108990. <https://doi.org/10.1016/j.spl.2020.108990>
- [7] Wei, Y., Wang, Q. and Liu, W. (2021) Model Averaging for Linear Models with Responses Missing at Random. *Annals of the Institute of Statistical Mathematics*, **73**, 535-553. <https://doi.org/10.1007/s10463-020-00759-y>
- [8] Zeng, J., Cheng, W. and Hu, G. (2023) Optimal Model Averaging Estimation for the Varying-Coefficient Partially Linear Models with Missing Responses. *Mathematics*, **11**, Article 1883. <https://doi.org/10.3390/math11081883>
- [9] Zhu, R., Wang, H., Zhang, X. and Liang, H. (2023) A Scalable Frequentist Model Averaging Method. *Journal of Business & Economic Statistics*, **41**, 1228-1237. <https://doi.org/10.1080/07350015.2022.2116442>

附录

定理 1 的证明: 令 $\mathbf{P} = \mathbf{U}\mathbf{U}^\top$, 通过使用与文献[9]中定理 1 类似的证明技巧, 我们有

$$\inf_{\tilde{\mathbf{w}} \in \tilde{W}^{2p}} \tilde{L}_{\hat{\pi}}(\tilde{\mathbf{w}}) \geq \|\mathbf{P}\boldsymbol{\mu} - \boldsymbol{\mu}\|^2,$$

$$\inf_{\mathbf{w} \in W} L_{\hat{\pi}}(\mathbf{w}) \leq \|\mathbf{P}\mathbf{z}_{\hat{\pi}} - \boldsymbol{\mu}\|^2 \leq \|\mathbf{P}\boldsymbol{\mu} - \boldsymbol{\mu}\|^2 + \mathbf{e}_{\pi}^\top \mathbf{P}\mathbf{e}_{\pi} + (\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi})^\top \mathbf{P}(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi}) + 2\|\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi}\| \|\mathbf{P}\mathbf{e}_{\pi}\|.$$

在条件(C1)、(C2)下, 存在常数 c , 使得 $\max_{1 \leq i \leq n} E(\mathbf{e}_{\pi,i}^4) \leq c$, 则

$$\mathbf{e}_{\pi}^\top \mathbf{P}\mathbf{e}_{\pi} = E(\mathbf{e}_{\pi}^\top \mathbf{P}\mathbf{e}_{\pi}) + O_p\left\{\sqrt{\text{Var}(\mathbf{e}_{\pi}^\top \mathbf{P}\mathbf{e}_{\pi})}\right\} = O_p(p). \quad (\text{A.1})$$

另外, 由文献[7]的引理 1, 有 $\|\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi}\|^2 = O_p(1)$, 可得

$$(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi})^\top \mathbf{P}(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi}) \leq \lambda_{\max}(\mathbf{P})\|\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi}\|^2 \leq \|\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi}\|^2 = O_p(1),$$

则有 $\inf_{\mathbf{w} \in W} L_{\hat{\pi}}(\mathbf{w}) \leq \|\mathbf{P}\boldsymbol{\mu} - \boldsymbol{\mu}\|^2 + O_p(p)$, 且 $\xi_{\hat{\pi}} = \inf_{\mathbf{w} \in W} E[L_{\hat{\pi}}(\mathbf{w})] \leq \|\mathbf{P}\boldsymbol{\mu} - \boldsymbol{\mu}\|^2 + O_p(p)$ 。

结合上述结果和条件(C3), 有

$$1 \leq \frac{\inf_{\mathbf{w} \in W} L_{\hat{\pi}}(\mathbf{w})}{\inf_{\tilde{\mathbf{w}} \in \tilde{W}^{2p}} \tilde{L}_{\hat{\pi}}(\tilde{\mathbf{w}})} \leq \frac{\|\mathbf{P}\boldsymbol{\mu} - \boldsymbol{\mu}\|^2 + O_p(p)}{\|\mathbf{P}\boldsymbol{\mu} - \boldsymbol{\mu}\|^2} \leq 1 + \frac{O_p(p)}{\|\mathbf{P}\boldsymbol{\mu} - \boldsymbol{\mu}\|^2} \leq 1 + \frac{O_p(p)}{\xi_{\hat{\pi}} - O_p(p)} = 1 + o_p(1),$$

故定理 1 得证。

定理 2 的证明: 记 $\mathbf{A}(\mathbf{w}) = \mathbf{I}_n - \mathbf{H}(\mathbf{w})$, $\boldsymbol{\Sigma}_{\pi} = E(\mathbf{e}_{\pi}\mathbf{e}_{\pi}^\top)$ 。经过简单的计算, 有

$$\mathbf{P}\mathbf{H}(\mathbf{w}) = \mathbf{H}(\mathbf{w}), \quad R_{\pi}(\mathbf{w}) = \|\mathbf{A}(\mathbf{w})\boldsymbol{\mu}\|^2 + \text{tr}(\mathbf{H}(\mathbf{w})\boldsymbol{\Sigma}_{\pi}\mathbf{H}(\mathbf{w})) \geq \|\mathbf{A}(\mathbf{w})\boldsymbol{\mu}\|^2,$$

$$L_{\pi}(\mathbf{w}) - R_{\pi}(\mathbf{w}) = \|\mathbf{H}(\mathbf{w})\mathbf{e}_{\pi}\|^2 - 2\mathbf{e}_{\pi}^\top \mathbf{H}(\mathbf{w})\mathbf{A}(\mathbf{w})\boldsymbol{\mu} - \text{tr}(\mathbf{H}(\mathbf{w})\boldsymbol{\Sigma}_{\pi}\mathbf{H}(\mathbf{w})). \quad (\text{A.2})$$

首先, $\mathbf{J}_{\hat{\pi}}(\mathbf{w}) = \|\mathbf{z}_{\hat{\pi}} - \hat{\mathbf{H}}(\mathbf{w})\mathbf{z}_{\hat{\pi}}\|^2 = \|(\mathbf{z}_{\hat{\pi}} - \boldsymbol{\mu}) + (\boldsymbol{\mu} - \mathbf{H}(\mathbf{w})\mathbf{z}_{\hat{\pi}}) - (\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}))\mathbf{z}_{\hat{\pi}}\|^2$ 可表示为

$$\mathbf{J}_{\hat{\pi}}(\mathbf{w}) = L_{\hat{\pi}}(\mathbf{w}) + t_{\hat{\pi}}(\mathbf{w}) + \|\mathbf{z}_{\hat{\pi}} - \boldsymbol{\mu}\|^2 + 2\mathbf{e}_{\pi}^\top \boldsymbol{\mu} - 2\mathbf{e}_{\pi}^\top \mathbf{P}\boldsymbol{\mu},$$

其中

$$\begin{aligned} t_{\hat{\pi}}(\mathbf{w}) = & \left\| (\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}))\mathbf{z}_{\hat{\pi}} \right\|^2 - 2(\boldsymbol{\mu} - \mathbf{H}(\mathbf{w})\mathbf{z}_{\hat{\pi}})^\top (\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}))\mathbf{z}_{\hat{\pi}} \\ & - 2(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi})^\top (\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}))\mathbf{z}_{\hat{\pi}} - 2\mathbf{e}_{\pi}^\top (\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}))\mathbf{z}_{\hat{\pi}} \\ & + 2(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi})^\top (\boldsymbol{\mu} - \mathbf{H}(\mathbf{w})\mathbf{z}_{\hat{\pi}}) + 2\mathbf{e}_{\pi}^\top \{ \mathbf{P}\mathbf{A}(\mathbf{w})\boldsymbol{\mu} - \mathbf{H}(\mathbf{w})(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi}) - \mathbf{H}(\mathbf{w})\mathbf{e}_{\pi} \}, \end{aligned}$$

故可知 $\hat{\mathbf{w}}_{\hat{\pi}} = \arg \min_{\mathbf{w} \in W} L_{\hat{\pi}}(\mathbf{w}) + t_{\hat{\pi}}(\mathbf{w})$ 。据此, 要证定理 2 成立, 只用证:

$$\sup_{\mathbf{w} \in W} \left| \frac{L_{\hat{\pi}}(\mathbf{w})}{R_{\pi}(\mathbf{w})} - 1 \right| = o_p(1), \quad (\text{A.3})$$

$$\sup_{\mathbf{w} \in W} \left| \frac{t_{\hat{\pi}}(\mathbf{w})}{R_{\pi}(\mathbf{w})} \right| = o_p(1). \quad (\text{A.4})$$

对(A.3), 由于 $L_{\hat{\pi}}(\mathbf{w}) = \|\mathbf{H}(\mathbf{w})\mathbf{z}_{\hat{\pi}} - \boldsymbol{\mu}\|^2 = L_{\pi}(\mathbf{w}) + 2(\mathbf{H}(\mathbf{w})\mathbf{z}_{\pi} - \boldsymbol{\mu})^\top \mathbf{H}(\mathbf{w})(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi}) + \|\mathbf{H}(\mathbf{w})(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi})\|^2$, 则有

$$\left| \frac{L_{\hat{\pi}}(\mathbf{w})}{R_{\pi}(\mathbf{w})} - 1 \right| \leq \left| \frac{L_{\pi}(\mathbf{w})}{R_{\pi}(\mathbf{w})} - 1 \right| + \frac{2\{L_{\pi}(\mathbf{w})\}^{1/2} \|\mathbf{H}(\mathbf{w})(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi})\|}{R_{\pi}(\mathbf{w})} + \frac{\|\mathbf{H}(\mathbf{w})(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_{\pi})\|^2}{R_{\pi}(\mathbf{w})}, \quad (\text{A.5})$$

由(A.2), 若(i) $\sup_{\mathbf{w} \in W} \|\mathbf{H}(\mathbf{w})\mathbf{e}_{\pi}\|^2 / R_{\pi}(\mathbf{w}) = o_p(1)$; (ii) $\sup_{\mathbf{w} \in W} |\mathbf{e}_{\pi}^\top \mathbf{H}(\mathbf{w})\mathbf{A}(\mathbf{w})\boldsymbol{\mu}| / R_{\pi}(\mathbf{w}) = o_p(1)$, (iii)

$$\sup_{\mathbf{w} \in W} \text{tr}(\mathbf{H}(\mathbf{w})\boldsymbol{\Sigma}_\pi \mathbf{H}(\mathbf{w})) / R_\pi(\mathbf{w}) = o_p(1)$$

成立, 则有

$$\sup_{\mathbf{w} \in W} |L_\pi(\mathbf{w}) / R_\pi(\mathbf{w}) - 1| = o_p(1). \tag{A.6}$$

事实上, 由于 $\mathbf{H}(\mathbf{w}) = \sum_{j=1}^p w_j \mathbf{u}_{(j)} \mathbf{u}_{(j)}^\top \leq \mathbf{P}$, $\mathbf{H}^2(\mathbf{w}) = \left\{ \sum_{j=1}^p w_j \mathbf{u}_{(j)} \mathbf{u}_{(j)}^\top \right\}^2 = \sum_{j=1}^p w_j^2 \mathbf{u}_{(j)} \mathbf{u}_{(j)}^\top \leq \mathbf{P}$, 这里的不等号是 Loewner 序: 即矩阵 $\mathbf{A} \leq \mathbf{B}$ 表示当且仅当 $\mathbf{B} - \mathbf{A}$ 为半正定矩阵。据此, 类似于文献[9]中引理 1 中的证明, 由条件(C2), (C3)和结论(A.1), 可知(i)~(iii)成立。

另外, 注意到

$$\|\mathbf{H}(\mathbf{w})(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi)\|^2 = (\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi)^\top \mathbf{H}^2(\mathbf{w})(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi) \leq (\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi)^\top \mathbf{P}(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi) \leq \|\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi\|^2, \tag{A.7}$$

由条件(C2), (C3)和 $\|\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi\|^2 = O_p(1)$, 不难证明

$$\sup_{\mathbf{w} \in W} \|\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi\|^2 / R_\pi(\mathbf{w}) = o_p(1). \tag{A.8}$$

从而, 结合(A.5)~(A.8)的结果(A.3)得证。

对(A.4), 我们有

$$\begin{aligned} |t_{\hat{\pi}}(\mathbf{w})| \leq & \left\| \left(\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}) \right) \mathbf{z}_{\hat{\pi}} \right\|^2 + 2 \left\{ L_{\hat{\pi}}(\mathbf{w}) \right\}^{1/2} \left\| \left(\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}) \right) \mathbf{z}_{\hat{\pi}} \right\| \\ & + 2 \left\| \mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi \right\| \left\| \left(\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}) \right) \mathbf{z}_{\hat{\pi}} \right\| + 2 \left| \mathbf{e}_\pi^\top \left(\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}) \right) \mathbf{z}_{\hat{\pi}} \right| \\ & + 2 \left\| \left(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi \right) \right\|^\top \left\{ L_{\hat{\pi}}(\mathbf{w}) \right\}^{1/2} + 2 \left| \mathbf{e}_\pi^\top \mathbf{P} \mathbf{A}(\mathbf{w}) \boldsymbol{\mu} \right| + 2 \left| \mathbf{e}_\pi^\top \mathbf{H}(\mathbf{w}) \left(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi \right) \right| + 2 \mathbf{e}_\pi^\top \mathbf{H}(\mathbf{w}) \mathbf{e}_\pi, \end{aligned}$$

这一结果与(A.3)、(A.8)相结合, 不难看出, 要证(A.4)成立, 我们只用证

$$\sup_{\mathbf{w} \in W} \left\| \left(\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}) \right) \mathbf{z}_{\hat{\pi}} \right\|^2 / R_\pi(\mathbf{w}) = o_p(1), \tag{A.9}$$

$$\sup_{\mathbf{w} \in W} \left| \mathbf{e}_\pi^\top \left(\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}) \right) \mathbf{z}_{\hat{\pi}} \right| / R_\pi(\mathbf{w}) = o_p(1), \tag{A.10}$$

$$\sup_{\mathbf{w} \in W} \left| \mathbf{e}_\pi^\top \mathbf{P} \mathbf{A}(\mathbf{w}) \boldsymbol{\mu} \right| / R_\pi(\mathbf{w}) = o_p(1), \tag{A.11}$$

$$\sup_{\mathbf{w} \in W} \left| \mathbf{e}_\pi^\top \mathbf{H}(\mathbf{w}) \left(\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi \right) \right| / R_\pi(\mathbf{w}) = o_p(1), \tag{A.12}$$

$$\sup_{\mathbf{w} \in W} \mathbf{e}_\pi^\top \mathbf{H}(\mathbf{w}) \mathbf{e}_\pi / R_\pi(\mathbf{w}) = o_p(1). \tag{A.13}$$

记 \mathbf{Q}_j 为对角元素为 $h_{j,ii} (1 - h_{j,ii})^{-1}$ 的 n 阶对角阵, $\mathbf{A}_j = \mathbf{I}_n - \mathbf{H}_j$, 则 $\hat{\mathbf{H}}_j = \mathbf{H}_j - \mathbf{Q}_j \mathbf{A}_j$, $j = 1, \dots, p$ 。由条件(C4)可知 $h_{j,ii} = u_{ij}^2 \leq c_3 n^{-1}$, 且在条件(C1)和(C2)下, 有 $n^{-1} \|\mathbf{z}_{\hat{\pi}}\|^2 \leq n^{-1} (\|\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi\|^2 + \|\mathbf{z}_\pi\|^2) = O_p(1)$, $n^{-1} \|\mathbf{e}_\pi\|^2 = O_p(1)$ 。因此, 我们有

$$\begin{aligned} & \sup_{\mathbf{w} \in W} \left\| \left(\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w}) \right) \mathbf{z}_{\hat{\pi}} \right\|^2 / R_\pi(\mathbf{w}) \\ &= \sup_{\mathbf{w} \in W} \sum_{j=1}^p \sum_{k=1}^p w_j w_k \mathbf{z}_{\hat{\pi}}^\top \mathbf{A}_j \mathbf{Q}_j \mathbf{Q}_k \mathbf{A}_k \mathbf{z}_{\hat{\pi}} / R_\pi(\mathbf{w}) \\ &= \sup_{\mathbf{w} \in W} \sum_{j=1}^p \sum_{k=1}^p w_j w_k \mathbf{z}_{\hat{\pi}}^\top \left(\frac{\mathbf{A}_j \mathbf{Q}_j \mathbf{Q}_k \mathbf{A}_k + \mathbf{A}_k \mathbf{Q}_k \mathbf{Q}_j \mathbf{A}_j}{2} \right) \mathbf{z}_{\hat{\pi}} / R_\pi(\mathbf{w}) \\ &\leq \sup_{\mathbf{w} \in W} \sum_{j=1}^p \sum_{k=1}^p w_j w_k \lambda_{\max} \left(\mathbf{A}_j \mathbf{Q}_j \mathbf{Q}_k \mathbf{A}_k \right) \|\mathbf{z}_{\hat{\pi}}\|^2 / R_\pi(\mathbf{w}) \\ &\leq \xi_\pi^{-1} p^2 \left(\frac{c_3 n^{-1}}{1 - c_3 n^{-1}} \right)^2 \|\mathbf{z}_{\hat{\pi}}\|^2 = O_p(\xi_\pi^{-1} p) = o_p(1) \end{aligned}$$

$$\begin{aligned}
& \sup_{\mathbf{w} \in W} \left| \mathbf{e}_\pi^\top (\hat{\mathbf{H}}(\mathbf{w}) - \mathbf{H}(\mathbf{w})) \mathbf{z}_{\hat{\pi}} \right| / R_\pi(\mathbf{w}) \\
& \leq \sup_{\mathbf{w} \in W} \sum_{j=1}^p w_j \left| \mathbf{e}_\pi^\top \mathbf{Q}_j \mathbf{A}_j \mathbf{z}_{\hat{\pi}} \right| / R_\pi(\mathbf{w}) \\
& \leq \sup_{\mathbf{w} \in W} \sum_{j=1}^p w_j \left\| \mathbf{Q}_j \mathbf{e}_\pi \right\| \left\| \mathbf{A}_j \mathbf{z}_{\hat{\pi}} \right\| / R_\pi(\mathbf{w}) \\
& \leq \xi_\pi^{-1} p \frac{c_3 n^{-1}}{1 - c_3 n^{-1}} \left\| \mathbf{e}_\pi \right\| \left\| \mathbf{z}_{\hat{\pi}} \right\| = O_p(\xi_\pi^{-1} p) = o_p(1) \\
& \sup_{\mathbf{w} \in W} \left| \mathbf{e}_\pi^\top \mathbf{P} \mathbf{A}(\mathbf{w}) \boldsymbol{\mu} \right| / R_\pi(\mathbf{w}) \\
& \leq \sup_{\mathbf{w} \in W} \left\| \mathbf{P} \mathbf{e}_\pi \right\| \left\| \mathbf{A}(\mathbf{w}) \boldsymbol{\mu} \right\| / R_\pi(\mathbf{w}) \\
& \leq \sup_{\mathbf{w} \in W} \left\| \mathbf{P} \mathbf{e}_\pi \right\| / R_\pi^{1/2}(\mathbf{w}) = O_p(\xi_\pi^{-1/2} p^{1/2}) = o_p(1) \\
& \sup_{\mathbf{w} \in W} \left| \mathbf{e}_\pi^\top \mathbf{H}(\mathbf{w}) (\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi) \right| / R_\pi(\mathbf{w}) \\
& = \sup_{\mathbf{w} \in W} \left| \mathbf{e}_\pi^\top \mathbf{P} \mathbf{H}(\mathbf{w}) (\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi) \right| / R_\pi(\mathbf{w}) \\
& \leq \sup_{\mathbf{w} \in W} \left\| \mathbf{P} \mathbf{e}_\pi \right\| \left\| \mathbf{H}(\mathbf{w}) (\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi) \right\| / R_\pi(\mathbf{w}) \\
& \leq \sup_{\mathbf{w} \in W} \left(\mathbf{e}_\pi^\top \mathbf{P} \mathbf{e}_\pi \right)^{1/2} \left\{ (\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi)^\top \mathbf{P} (\mathbf{z}_{\hat{\pi}} - \mathbf{z}_\pi) \right\}^{1/2} / R_\pi(\mathbf{w}) \\
& = O_p(\xi_\pi^{-1} p^{1/2}) = o_p(1) \\
& \sup_{\mathbf{w} \in W} \mathbf{e}_\pi^\top \mathbf{H}(\mathbf{w}) \mathbf{e}_\pi / R_\pi(\mathbf{w}) \leq \sup_{\mathbf{w} \in W} \mathbf{e}_\pi^\top \mathbf{P} \mathbf{e}_\pi / R_\pi(\mathbf{w}) = O_p(\xi_\pi^{-1} p) = o_p(1)
\end{aligned}$$

故(A.9)~(A.13)成立，至此(A.4)证毕。进而定理 2 得证。