系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
42(1) (2022, 1), 109–132

大规模数据下子抽样模型平均估计理论*

宗先鹏[1] 王彤彤[2]

(1. 北京工业大学理学部, 北京 100124; 2. 首都师范大学数学科学学院, 北京 100048)

摘要 随着信息时代的来临, 如何从海量数据中快速、有效地挖掘有用信息是目前面临的新挑战. 子抽样方法作为大规模数据分析的有效工具, 已经受到国内外学者的广泛关注. 不过, 传统的子抽样方法通常没有考虑到模型的不确定性. 当模型假设不正确时, 后面的统计推断将会出现偏差, 甚至导致错误的结论. 为了解决该问题, 文章利用频率模型平均的方法构建了子抽样模型平均估计 (简称 SSMA 估计). 理论上, 文章证明了 SSMA 估计是全部数据下模型平均估计的一个渐近无偏且相合的估计. 另外, 我们基于 Hansen (2007) 的 Mallows 模型平均方法提出了 SSMA 估计的权重选择准则, 并证明了方差已知和未知时权重估计的渐近最优性. 在这些理论性质的研究中, 文章同时考虑了模型和抽样设计带来的双重随机性. 最后, 数值分析进一步说明了所提出方法的有效性.

关键词 大数据分析, 子抽样方法, 模型平均, Mallows 准则, 渐近最优性.

MR(2000) 主题分类号 62F12, 62H12

Sub-Sampling Model Averaging Theory for Large Scale Data

ZONG Xianpeng[1] WANG Tongtong[2]

(1. *Faculty of Science, Beijing University of Technology, Beijing* 100124; 2. *School of Mathematical Sciences, Capital Normal University, Beijing* 100048)

Abstract With the development of information age, how to mine useful information from massive data quickly and effectively is a new challenge. As an effective tool for large scale data analysis, sub-sampling method has attracted extensive attention of scholars at home and abroad. However, the traditional sub-sampling method usually does not take into account the uncertainty of the model. When the assumed model is incorrect, the conclusions may be wrong. In order to solve this problem, a sub-sampling model averaging estimator (SSMA estimator) is constructed by the sampled data. Theoretically, we prove that the SSMA estimator is an asymptotically unbiased and consistent estimator of the model averaging estimator based on full data. In

* 北京市自然科学基金重点研究专项 (Z210003), 国家自然科学基金 (11971323, 12031016, 71973116) 和首都师范大学交叉科学研究院和生物统计交叉学科研究项目资助课题.

收稿日期: 2021-09-14, 收到修改稿日期: 2021-10-20.

编委: 邹国华.

addition, we propose a weight choice criterion for the SSMA estimator, which is based on the Mallows' criterion proposed by Hansen (2007), and derive the asymptotic optimality of the weight estimator. It is worth mentioning that, in the proofs of these theoretical properties, we consider the double randomness brought by the model and sampling design. Finally, numerical analysis further shows the effectiveness of the proposed method.

Keywords Big data analysis, sub-sampling method, model averaging, Mallows' criterion, asymptotic optimality.

1 引 言

随着科学技术的发展, 人们存储数据的工具越来越先进, 数据的存储量也越来越大. 如何对大数据快速、有效的分析成为统计学和数据挖掘领域的研究热点问题. 由于数据量庞大、维数较高, 传统的计算机工具 (如办公电脑) 会因内存或显卡的限制不能快速进行统计分析. 目前, 对大数据进行快速分析的方法主要分为三种: 并行运算、在线更新和子抽样方法. 顾名思义, 并行运算 (divide and conquer) 采用的是 "分而治之" 的策略, 其首先将整个数据集进行分块, 然后对分块的数据进行分析, 最后综合起来推断总体. 常见的文献有 Lin 和 Xi[1], Chen 和 Xie[2] 和 Song 和 Liang[3]. 在线更新方法 (online updating approach) 是针对大数据流提出的一种实时分析的方法, 其研究文献包括 Schifano 等 [4] 和 Wang 等 [5]. 而子抽样方法 (sub-sampling method) 则采用抽样调查的方法, 从所有数据 (可看作是一个有限总体) 中抽取部分数据进行推断. 子抽样方法不仅能够快速分析数据, 而且对分析工具的硬件要求不高, 其估计精度主要取决于抽中数据的样本量. 与传统的抽样调查不同, 大数据子抽样的有限总体是已知的, 只不过因为数据量庞大无法进行快速分析.

在大数据子抽样方法中, 等概率抽样是最常用的方法, 其操作简单, 可以从大数据中快速抽取子样本进行统计推断. 不过, 等概率抽中的样本是均匀的, 因此代表性不是很强. 在实际应用中, 我们可以通过多次抽样的方法提高估计精度. 另外, 重复抽样也可以进行其他的统计推断, 如方差估计、构造置信区间等. 不等概率抽样可以选择有代表性的样本, 但在大数据分析中, 所有数据的入样概率往往是未知的, 因此需要事先进行额外估计. 常见不等概率子抽样方法有: 杠杆值抽样和最优子抽样等. 另外, 从实施抽样的角度来讲, 等概率抽样和有放回的不等概率抽样耗时较短, 而不放回的不等概率抽样操作复杂. 因此, 重复利用等概率抽样或有放回的不等概率抽样是大数据子抽样中的常用方法. Kleiner 等 [6] 提出从所有数据中反复抽取子样本, 然后基于多个子样本的估计进行统计推断; Ma 等 [7] 分析了利用杠杆值抽取样本推断总体时的统计性质; Wang 等 [8] 研究了逻辑回归中大数据子抽样方法, 提出了最优子抽样方法; Deldossi 和 Tommasi[9] 提出了大数据最优设计子抽样方法. 其他关于子抽样方法的文献还有 Wang 等 [10], Liang 等 [11], Wang[12] 和 Ai 等 [13].

目前, 子抽样方法已经成为大规模数据分析的有效工具, 并受到众多学者的广泛关注. 不过, 已有的这些文献通常没有考虑模型的不确定性. 当模型假设不正确时, 后面的统计推断将会出现偏差, 甚至导致错误的结论. 特别是对最优子抽样方法, 当模型假设错误时, 入样概率的估计也会受到影响. 因此, 本文将结合频率模型平均方法的优势开发子抽样模型平均

估计. 近年来, 频率模型平均方法发展十分迅速, 它考虑了各个子模型的不确定性, 将每个子模型组合起来进行统计推断. 这种平均的思想可以避免选择一个很差的模型, 从而有望降低估计的风险. Buckland 等 [14] 提出了 Smoothed AIC 和 Smoothed BIC 方法. 该方法计算简单且便于操作, 是一种常用的权重选择方法. Yang[15] 提出了自适应模型平均方法; Hjort 和 Claeskens[16] 探究了频率模型平均估计的渐近分布; Hansen[17] 提出了 Mallows 权重选择准则, 并在嵌套模型下证明了权重估计的渐近最优性; Wan 等 [18] 进一步将嵌套的 Mallows 模型平均方法推广到了非嵌套模型中, 并给出了渐近最优性的证明; Hansen 和 Racine[19] 提出了 Jackknife 模型平均估计; Ando 和 Li 等 [20] 探究了高维模型平均方法. 这些基础文献的研究大大推动了频率模型平均理论的发展. 其他文献可参考 Zhang 和 Liang[21], Zhang 等 [22], Hansen[23], Zhang 等 [24], Zhang 等 [25] 和 Liao 等 [26].

结合频率模型平均方法的优势, 本文开发了等概率子抽样的模型平均估计方法. 具体地, 我们基于子抽样数据在每个候选模型下构建最小二乘估计, 然后将所有的估计加权起来, 提出了子抽样模型平均估计 (简称 SSMA 估计). 理论上, 我们证明了 SSMA 估计是全部数据下模型平均估计的一个渐近无偏且相合的估计. 另外, 我们基于 Hansen[17] 的 Mallows 模型平均方法提出了 SSMA 估计的权重选择准则, 并证明了方差已知和未知时权重估计的渐近最优性. 值得一提的是, 在这些理论性质的研究中, 我们同时考虑了模型和抽样设计带来的双重随机性. 这虽然使理论的证明变得非常复杂, 但与实际情况更加相符, 理论价值也较高. 最后, 数值分析进一步说明了本文提出方法的有效性.

本文后面的结构安排如下: 第 2 节根据抽样数据构建了子抽样模型平均估计, 并给出了权重选择准则; 第 3 节研究了 SSMA 估计的理论性质, 并证明了方差已知和未知时权重估计的渐近最优性; 在第 4 节中, 我们通过数值分析说明了提出方法的有效性; 第 5 节是总结; 所有引理和定理的证明在附录中.

2 子抽样模型平均估计

记所有数据 $D_N = \{(y_i, \boldsymbol{x}_i), i = 1, 2, \cdots, N\}$, 这里 y_i 是第 i 次观测的响应, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{ip})'$ 表示第 i 次观测的 p 维协变量. 维数 p 可以随着 N 的增加而增加. 为简单计, 将 D_N 简记为 $D_N = \{1, 2, \cdots, N\}$. 我们假设数据 D_N 来自于以下模型

$$y_i = \mu_i + e_i, \quad i = 1, 2, \cdots, N, \tag{2.1}$$

其中 $\mu_i = \sum_{j=1}^{\infty} x_{ij}\beta_j$, e_i 表示第 i 次观测的随机误差, 且满足 $E_m(e_i) = 0$ 和 $V_m(e_i) = \sigma^2$. 注意到, 模型 (2.1) 中的协变量个数是无限维的, 因此该模型是误设定的. 另外, 这里的 E_m 和 V_m 分别表示对模型的随机性求期望和方差.

考虑模型选择的不确定性, 我们可以利用 Hansen[17] 提出的 Mallows 模型平均方法提高估计的精度和稳健性. 但是随着数据量的增加, 普通计算机会因内存或显卡的限制无法对全部数据进行模型平均估计. 为了解决该问题, 我们采用简单随机抽样从全部数据中抽取部分样本进行分析, 提出子抽样模型平均估计.

假设有 M 个候选模型, M 可以随着 N 趋于无穷. 第 m 个候选模型使用前 k_m 个协变

量, 记为

$$y_i = \sum_{j=1}^{k_m} x_{ij}\beta_j + b_{i(m)} + e_i, \quad i = 1, 2, \cdots, N, \tag{2.2}$$

这里 $b_{i(m)} = \mu_i - \sum_{j=1}^{k_m} x_{ij}\beta_j$ 为第 m 个候选模型的近似误差, 最大模型协变量的个数为 k_M ($k_M \le p$). 由于大数据的维数可能很高, 因此本文采用了嵌套的模型平均估计, 以减少计算量. 在实际应用中, 我们可以先将所有协变量分组排序, 然后进行嵌套的模型平均估计. 模型 (2.2) 写成矩阵形式为

$$\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{X}}_{(m)}\boldsymbol{\beta}_{(m)} + \widetilde{\boldsymbol{b}}_{(m)} + \widetilde{\boldsymbol{e}},$$

其中 $\widetilde{\boldsymbol{Y}} = (y_1, y_2, \cdots, y_N)' \triangleq [y_i]_{i\in D_N}$, $\boldsymbol{\beta}_{(m)} = (\beta_1, \beta_2, \cdots, \beta_{k_m})'$, $\widetilde{\boldsymbol{b}}_{(m)} = [b_{i(m)}]_{i\in D_N}$, $\widetilde{\boldsymbol{e}} = [e_i]_{i\in D_N}$, 且

$$\widetilde{\boldsymbol{X}}_{(m)} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k_m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk_m} \end{bmatrix} \triangleq \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{ik_m} \end{bmatrix}_{i\in D_N}.$$

因此, 基于所有数据 D_N, 第 m 个候选模型下 $\boldsymbol{\beta}_{(m)}$ 的最小二乘估计为

$$\begin{aligned} \widetilde{\boldsymbol{\beta}}_{(m)} &= \left(\sum_{i\in D_N} \boldsymbol{x}_{i(m)}\boldsymbol{x}'_{i(m)}\right)^{-1} \left(\sum_{i\in D_N} \boldsymbol{x}_{i(m)} y_i\right) \\ &= \left(\widetilde{\boldsymbol{X}}'_{(m)}\widetilde{\boldsymbol{X}}_{(m)}\right)^{-1}\widetilde{\boldsymbol{X}}'_{(m)}\widetilde{\boldsymbol{Y}}, \end{aligned} \tag{2.3}$$

这里 $\boldsymbol{x}_{i(m)} = (x_{i1}, x_{i2}, \cdots, x_{ik_m})'$. 记 $\boldsymbol{\omega} = (\omega_1, \omega_2, \cdots, \omega_M)'$ 是一个 M 维的权重向量, 且来自于以下空间

$$\mathscr{H} = \left\{ \boldsymbol{\omega} \in \mathbb{R}^M : \omega_m \ge 0, \quad \sum_{m=1}^{M} \omega_m = 1 \right\},$$

则我们基于全部数据得到 $\boldsymbol{\beta}_{(M)}$ 的模型平均估计

$$\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega}) = \sum_{m=1}^{M} \omega_m \begin{pmatrix} \widetilde{\boldsymbol{\beta}}_{(m)} \\ 0 \end{pmatrix}, \tag{2.4}$$

这里 $\boldsymbol{\omega} = (\omega_1, \omega_2, \cdots, \omega_M)' \in \mathscr{H}$.

当数据量庞大时, (2.3) 式中的 $\left(\widetilde{\boldsymbol{X}}'_{(m)}\widetilde{\boldsymbol{X}}_{(m)}\right)^{-1}$ 在普通计算机上很难快速计算出来, 进而无法求出全部数据下的模型平均估计 $\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})$. 现采用简单随机抽样从所有数据 D_N 中抽取一个大小为 n ($n > k_M$) 的样本 s. 为简单计, 将 s 简记为 $s = \{1, 2, \cdots, n\}$. 利用抽中的数据 s, 第 m 个候选模型下 $\boldsymbol{\beta}_{(m)}$ 的一个基于设计估计为

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{(m)} &= \left(\sum_{i\in s} \boldsymbol{x}_{i(m)}\boldsymbol{x}'_{i(m)}\right)^{-1} \left(\sum_{i\in s} \boldsymbol{x}_{i(m)} y_i\right) \\ &= \left(\boldsymbol{X}'_{(m)}\boldsymbol{X}_{(m)}\right)^{-1}\boldsymbol{X}'_{(m)}\boldsymbol{Y}, \end{aligned}$$

这里 $\boldsymbol{X}_{(m)}=[x_{i1},x_{i2},\cdots,x_{ik_m}]_{i\in s}$, $\boldsymbol{Y}=[y_i]_{i\in s}$. 进一步, 将所有候选模型下的估计加权起来, 得到 $\boldsymbol{\beta}_{(M)}$ 的模型平均估计为

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})=\sum_{m=1}^{M}\omega_m\begin{pmatrix}\widehat{\boldsymbol{\beta}}_{(m)}\\0\end{pmatrix}. \tag{2.5}$$

这里 $\boldsymbol{\omega}=(\omega_1,\omega_2,\cdots,\omega_M)'\in\mathscr{H}$. 以上估计我们称为子抽样模型平均估计, 简称 SSMA 估计. 可以看出, SSMA 估计只依赖于被抽中的样本, 因此在普通计算机上可以快速进行计算. 如果权重向量已知, 给定任意的协变量 $\boldsymbol{x}_0=(x_{01},x_{02},\cdots,x_{0k_M})'$, 我们可以作出预测 $\widehat{\mu}_0=\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})$. 另外, SSMA 估计既依赖于抽样设计也与模型有关, 因此在探究理论性质时要同时考虑抽样和模型带来的双重随机性. 在本文中, 我们用 E_d 表示关于设计求期望, E_m 表示关于模型求期望.

在实际应用中, 如何选取模型平均估计的权重十分关键. Hansen[17] 基于全部数据提出了 Mallows 权重选择方法, 并证明了权重选择方法的渐近最优性. 类似地, 我们基于全部数据 D_N, 定义损失函数

$$L_N(\boldsymbol{\omega})=\left(\widetilde{\boldsymbol{\mu}}-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right)'\left(\widetilde{\boldsymbol{\mu}}-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right)$$

和风险函数

$$R_N(\boldsymbol{\omega})=E(L_N(\boldsymbol{\omega})|\widetilde{\boldsymbol{X}}),$$

这里 $E(\cdot)=E_mE_d(\cdot)$, $\widetilde{\boldsymbol{\mu}}=[\mu_i]_{i\in D_N}$, $\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})=\widetilde{\boldsymbol{X}}_{(M)}\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})$, $\widetilde{\boldsymbol{X}}=[x_{i1},x_{i2},\cdots]_{i\in D_N}$. 直观上, 使得全部数据的损失函数或者风险函数达到最小的权重是最优的. 不过, 由于 $\widetilde{\boldsymbol{\mu}}$ 在实际中未知, 因此损失函数和风险函数无法进行计算. 一种自然的想法就是来寻找它们的无偏估计或者渐近无偏估计. Hansen[17] 提出的 Mallows 权重选择准则就是风险函数的一个无偏估计 (见 Hansen[17] 中的引理 3).

在本文中, 我们利用抽中数据 s 提出以下权重选择准则

$$C_s(\boldsymbol{\omega})=\left(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right)'\left(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right)+2f\sigma^2k(\boldsymbol{\omega}), \tag{2.6}$$

这里 $\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})=\boldsymbol{X}_{(M)}\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})$, $f=\frac{n}{N}$ 为抽样比, $k(\boldsymbol{\omega})=\sum_{m=1}^{M}\omega_mk_m$. 进一步, 子抽样模型平均估计的权重确定方法为

$$\widehat{\boldsymbol{\omega}}=(\widehat{\omega}_1,\widehat{\omega}_2,\cdots,\widehat{\omega}_M)'=\arg\min_{\boldsymbol{\omega}\in\mathscr{H}}\ C_s(\boldsymbol{\omega}).$$

可以看出, 权重准则 (2.6) 式完全依赖于抽样数据, 因此可以实现快速计算. 特别地, 如果所有数据被抽中 (即 $n=N$), 本文提出的权重选择准则退化为 Mallows 权重选择准则. 在下一节中, 我们将研究该权重选择方法的理论性质.

3 SSMA 估计的理论性质

为了研究本文提出方法的理论性质, 我们作出以下正则假设.

条件 1 $E_m\left(e_i^4\right)=O(1)$, 且全部数据的协变量矩阵 $\widetilde{\boldsymbol{X}}_{(M)}$ 中每个元素都是有界的, 即

$$\max_{i\in D_N,\ j=1,2,\cdots,k_M}|x_{ij}|=O(1).$$

定义

$$\boldsymbol{\beta}^*_{(m)} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{R}^{k_m}} E_m\left(y_i - \boldsymbol{x}'_{i(m)}\boldsymbol{\theta}\right)^2,$$

则

$$E_m\left\{e^*_{i(m)}\boldsymbol{x}'_{i(m)}\right\} = \mathbf{0}', \tag{3.7}$$

这里 $e^*_{i(m)} = y_i - \boldsymbol{x}'_{i(m)}\boldsymbol{\beta}^*_{(m)}$.

条件 2 $\sup_{1\le m\le M}\left\{E_m\left(e^{*2q}_{i(m)}\right)\right\} = O(1)$, 这里 q 是一些正整数.

条件 3 对任意的 $1\le m\le M$, 有

$$NC_1 < \lambda_{\min}\left(\widetilde{\boldsymbol{X}}'_{(m)}\widetilde{\boldsymbol{X}}_{(m)}\right) \le \lambda_{\max}\left(\widetilde{\boldsymbol{X}}'_{(m)}\widetilde{\boldsymbol{X}}_{(m)}\right) < Nk_mC_2,$$

这里 C_1 和 C_2 为常数.

条件 4 对任意的 $1\le m\le M$, 有

$$nC_3 < \lambda_{\min}\left(\boldsymbol{X}'_{(m)}\boldsymbol{X}_{(m)}\right) \le \lambda_{\max}\left(\boldsymbol{X}'_{(m)}\boldsymbol{X}_{(m)}\right) < nk_mC_4,$$

这里 C_3 和 C_4 为常数.

条件 5 $\frac{Mk_M^2}{f\xi_N}\to 0$, 这里 $\xi_N = \inf_{\boldsymbol{\omega}\in\mathscr{H}} R_N(\boldsymbol{\omega})$.

条件 6 $\frac{nM^2k_M^2}{f^2\xi_N^2}\to 0$.

记 $\widetilde{L}_N(\boldsymbol{\omega}) = \left(\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right)'\left(\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right)$ 和 $\widetilde{R}_N(\boldsymbol{\omega}) = E(\widetilde{L}_N(\boldsymbol{\omega})|\widetilde{\boldsymbol{X}})$, 这里 $\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega}) = \widetilde{\boldsymbol{X}}_{(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})$.

条件 7 $M\widetilde{\xi}_N^{-2}\sum_{m=1}^M \widetilde{R}_N(\boldsymbol{\omega}_m^0)\to 0$, 这里 $\widetilde{\xi}_N = \inf_{\boldsymbol{\omega}\in\mathscr{H}}\widetilde{R}_N(\boldsymbol{\omega})$, $\boldsymbol{\omega}_m^0$ 表示第 m 个元素为 1, 其余为 0 的 M 维向量.

条件 8 $\widetilde{\boldsymbol{\mu}}'\widetilde{\boldsymbol{\mu}} = O(N)$ 且 $k_M^2 = O(N)$.

在本文中, 我们假设协变量是非随机的. 条件 1, 3 和 4 是对协变量和模型误差的一些常用假设. 条件 2 是对候选模型下的预测误差的矩进行假设, 其中的 $\beta^*_{(m)}$ 可看作是第 m 个模型下的线性投影系数. 条件 5 和条件 6 是对样本量、抽样比、维数、模型个数与风险函数下界的关系进行假设. 类似的条件在模型平均的论文中比较常见, 如 Ando 和 Li[20], Liao 等 [26] 和 Gao 等 [27]. 当 $n^{-1}k_M^2 = O(1)$ 时, 条件 6 推出条件 5. 条件 7 和条件 8 是文献 [18] 中的假设. 需要注意的是, 本文考虑了子抽样模型平均估计, 其抽取的样本量 n 与数据量 N 有关. 当 N 趋于无穷时, n 也趋于无穷.

定理 3.1 如果条件 1–4 成立, 则有

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}} E\left\|\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega}) - \widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right\|^2 = O\left(n^{-1}Mk_M\right).$$

证 见附录.

从定理 3.1 可以看出, 当 $n^{-1}Mk_M$ 趋于 0 时, 部分数据下的模型平均估计是全部数据的模型平均估计相合且渐近无偏估计.

定理 3.2 如果条件 1–4 成立, 则有

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{R_N(\boldsymbol{\omega}) - \widetilde{R}_N(\boldsymbol{\omega})}{R_N(\boldsymbol{\omega})}\to 0.$$

证 见附录.

注意到, $\widetilde{R}_N(\boldsymbol{\omega})$ 表示用全部数据作模型平均估计时的风险函数. 从定理 3.2 可以看出, 在一定的正则条件下, 本文使用的风险函数与全部数据作模型平均估计时的风险函数近似. 事实上, 两个估计都是定义在全部数据下的风险函数, 只不过在进行模型平均估计时所用的数据量不同, 因此随着抽样数据的增多两者趋于相同.

定理 3.3 如果条件 1–6 成立, 则有

$$E(C_s(\boldsymbol{\omega})) = fR_N(\boldsymbol{\omega}) + fN\sigma^2 + o\left(fR_N(\boldsymbol{\omega})\right).$$

证 见附录.

从定理 3.3 中可以看出, 本文提出的权重准则是 $fR_N(\boldsymbol{\omega}) + fN\sigma^2$ (风险函数的 f 倍加上一个与 $\boldsymbol{\omega}$ 无关的量) 的近似无偏估计. 因此, 当从大数据中抽取部分数据进行模型平均估计时, 最小化权重选择准则 $C_s(\boldsymbol{\omega})$ 是合理的. 特别地, 随着抽取的样本量增加, 该定理逐渐退化成 Hansen[17] 中的引理 3.

定理 3.4 如果条件 1–7 成立, 则有

$$\frac{L_N(\widehat{\boldsymbol{\omega}})}{\inf\limits_{\boldsymbol{\omega}\in\mathscr{H}} L_N(\boldsymbol{\omega})} \xrightarrow{p} 1.$$

证 见附录.

定理 3.4 证明了 SSMA 估计的权重渐近最优性, 即随着 N 趋于无穷 (n 也趋于无穷), 估计权重的损失函数会趋于最小损失函数. 特别地, 随着抽取的样本量增加, 该定理逐渐退化成 Hansen[17] 中的定理 1.

在实际中, 如果方差 σ^2 是未知的, 我们将用最大模型的方差估计 $\widehat{\sigma}^2_{(M)}$ 来代替, 这里

$$\widehat{\sigma}^2_{(M)} = \frac{1}{f(N-k_M)} \left\| \boldsymbol{Y} - \boldsymbol{X}_{(M)}\widehat{\boldsymbol{\beta}}_{(M)} \right\|^2.$$

用 $\widehat{\boldsymbol{\omega}}^*$ 表示对应的权重估计.

定理 3.5 如果条件 1–8 成立, 则有

$$\frac{L_N(\widehat{\boldsymbol{\omega}}^*)}{\inf\limits_{\boldsymbol{\omega}\in\mathscr{H}} L_N(\boldsymbol{\omega})} \xrightarrow{p} 1.$$

证 见附录.

从定理 3.5 可以看出, 当 σ^2 未知时, 用最大模型的方差估计 $\widehat{\sigma}^2_{(M)}$ 代替后的权重选择方法仍然具有渐近最优性. 与 Hansen[17] 不同, 本文是对部分抽样数据进行模型平均, 因此 $\widehat{\sigma}^2_{(M)}$ 与抽样设计和模型都有关.

4 数值分析

在本节中, 我们通过数值分析说明了本文提出方法的有效性. 类似于 Hansen[17] 的设置, 数据的产生过程为

$$y_i = \sum_{j=1}^{M_0} \theta_j x_{ij} + e_i, \quad i = 1, 2, \cdots, N,$$

这里 $x_{i1} = 1$ 作为截距项, 其余 x_{ij} 和随机误差项 e_i 都由标准正态分布独立产生. 设置 $M_0 = 200$, $N = 10000$, 回归系数 $\theta_j = c\sqrt{2\alpha}j^{-\alpha-\frac{1}{2}}$. 相关系数 $R^2 = \frac{c^2}{1+c^2}$ 与参数 c 有关.

我们采用简单随机抽样从 N 个生成数据中抽取部分数据进行模拟, 样本量 $n = (200, 600, 1000, 1400)$. 最大候选模型个数 M 取值约为 $3n^{\frac{1}{3}}$ (对应四个样本量的最大模型个数分别为 17, 25, 30, 33). 选取相应的系数 c, 控制 R^2 取值在 0.3 至 0.9 之间. 我们主要考虑了以下 7 种估计方法

1) 全部数据下的 Mallows 模型平均 (MMA); 2) 本文的子抽样模型平均估计 (SSMA);
3) 抽样数据下的 Mallows 模型选择 (C_p); 4) 抽样数据下的 AIC 模型选择 (AIC);
5) 抽样数据下的 BIC 模型选择 (BIC); 6) 抽样数据下的 SAIC 模型平均 (SAIC);
7) 抽样数据下的 SBIC 模型平均 (SBIC).

在模拟过程中, 全部数据下的 Mallows 模型平均的权重选择由 Hansen[17] 中的 (12) 式估计得到. 基于抽样数据, 第 m 个模型的 AIC 和 BIC 分别为

$$\mathrm{AIC}_m = n\ln\left(\widehat{\sigma}_m^2\right) + 2m$$

和

$$\mathrm{BIC}_m = n\ln\left(\widehat{\sigma}_m^2\right) + \ln(n)m,$$

其中 $\widehat{\sigma}_m^2 = \frac{1}{n}\big(Y - X_{(m)}\widehat{\boldsymbol{\beta}}_{(m)}\big)'\big(Y - X_{(m)}\widehat{\boldsymbol{\beta}}_{(m)}\big)$. 模拟中, AIC 和 BIC 模型选择分别选取使 AIC 值和 BIC 值最小的模型进行估计. SAIC 和 SBIC 模型平均的第 m 个候选模型权重估计分别为

$$\omega_m^{\mathrm{AIC}} = \exp\left(-\frac{1}{2}\mathrm{AIC}_m\right) \Big/ \sum_{m=1}^{M}\exp\left(-\frac{1}{2}\mathrm{AIC}_m\right)$$

和

$$\omega_m^{\mathrm{BIC}} = \exp\left(-\frac{1}{2}\mathrm{BIC}_m\right) \Big/ \sum_{m=1}^{M}\exp\left(-\frac{1}{2}\mathrm{BIC}_m\right).$$

为了评价以上各种方法的好坏, 我们重复 $D = 200$ 次分别计算各种估计量的平均损失, 即

$$\overline{\mathrm{Los}} = \frac{1}{D}\sum_{d=1}^{D}\left\|\widetilde{\boldsymbol{\mu}}_{(d)} - \widehat{\boldsymbol{\mu}}_{(d)}\right\|^2,$$

这里 $\widetilde{\boldsymbol{\mu}}_{(d)}$ 表示第 d 次重复时全部数据对应的真实值, $\widehat{\boldsymbol{\mu}}_{(d)}$ 表示第 d 次重复时全部数据对应的预测值. 表 1–4 分别展示了不同 $\alpha = (0.2, 0.4, 0.6, 0.8)$ 下不同估计方法的平均损失, 其中抽样数据下平均损失的最小值用黑色粗体显示, 次小值右上角用星号标记.

可以发现, 在子抽样方法中, 本文提出的 SSMA 估计大部分情况下平均损失最小或次小, 有随 R^2 增大逐渐变好的趋势. 特别是在 α 取值较小时, SSMA 估计往往能取得很好的效果, 这可能是因为 α 较小时变量系数衰减较慢, 模型误设定比较强, 有利于进行模型平均估计. 另外, 由于全部数据下的模型平均估计所用的数据量最多, 因此它的平均损失在所有方法中最小. 随着样本量的增加, 所有子抽样方法的平均损失逐渐减少. 随着相关系数 R^2 的增加, 所有子抽样方法的平均损失与全数据下 MMA 方法的平均损失逐渐拉近.

表 1 $\alpha = 0.2$ 时不同估计方法的平均损失

(Table 1 Average losses of different methods, $\alpha = 0.2$)

	R^2	MMA	SSMA	Cp	AIC	BIC	SAIC	SBIC
	0.3	865.26	1867.97*	1923.76	1933.96	2290.32	**1754.58**	2097.97
	0.4	1336.49	2379.15*	2501.75	2520.53	3105.36	**2358.83**	2878.22
	0.5	1996.91	**3073.83**	3232.93	3275.22	4172.21	3135.94*	3880.74
n=200	0.6	2986.62	**4183.71**	4300.51	4343.94	5404.21	4246.65*	5079.09
	0.7	4640.22	**5989.53**	6088.47	6170.36	7445.54	6085.00*	7087.48
	0.8	7922.84	**9514.24**	9577.22*	9677.07	11144.91	9640.87	10704.14
	0.9	17859.59	**20479.46**	20508.04*	20625.95	21895.31	20630.60	21588.93
	0.3	683.80	**1106.71**	1165.90	1171.23	1592.51	1121.92*	1502.76
	0.4	1049.17	**1487.27**	1536.73	1546.61	2106.91	1513.67*	2008.05
	0.5	1561.54	**2041.23**	2084.50	2091.21	2733.32	2069.46*	2623.58
n=600	0.6	2326.41	**2833.39**	2858.44*	2874.49	3509.93	2860.38	3391.40
	0.7	3606.84	**4166.74**	4187.28*	4193.12	4705.43	4193.78	4595.57
	0.8	6163.22	**6833.58**	6840.21*	6847.29	7193.89	6855.09	7100.00
	0.9	13811.15	**14774.46**	14774.55*	14784.34	14918.38	14796.21	14927.21
	0.3	606.91	**882.75**	923.00	926.85	1335.31	901.45*	1272.45
	0.4	925.32	**1216.51**	1243.92	1251.64	1700.52	1235.91*	1630.23
	0.5	1374.02	**1684.82**	1710.73	1713.68	2197.63	1707.70*	2125.61
n=1000	0.6	2048.56	**2391.88**	2407.38*	2410.08	2814.13	2408.89	2744.84
	0.7	3166.81	**3525.70**	3534.20*	3536.91	3848.13	3538.50	3783.67
	0.8	5402.67	**5811.27**	5813.08*	5814.88	5932.48	5821.56	5904.36
	0.9	12114.19	**12742.38**	12742.61*	12744.33	12796.60	12749.66	12791.31
	0.3	569.89	**776.51**	805.21	808.20	1152.94	793.27*	1106.54
	0.4	866.08	**1089.09**	1109.88	1113.88	1491.06	1105.81*	1437.12
	0.5	1280.15	**1509.15**	1523.79	1525.98	1889.77	1523.05*	1845.67
n=1400	0.6	1912.40	**2161.39**	2166.92*	2169.26	2446.69	2170.95	2393.55
	0.7	2948.53	**3220.74**	3224.49*	3226.63	3367.65	3229.16	3343.72
	0.8	5042.99	**5353.97**	5355.60*	5356.33	5419.18	5360.32	5407.20
	0.9	11290.66	**11735.44**	11735.78*	11736.97	11752.07	11738.00	11754.10

表 2 $\alpha = 0.4$ 时不同估计方法的平均损失

(Table 2 Average losses of different methods, $\alpha = 0.4$)

	R^2	MMA	SSMA	Cp	AIC	BIC	SAIC	SBIC
	0.30	389.97	1313.02	1303.74*	1307.07	1588.26	**1146.47**	1427.59
n=200	0.40	595.59	1549.31*	1626.68	1624.21	2058.18	**1482.75**	1841.74
	0.50	885.78	1813.39*	1944.64	1954.99	2559.38	**1813.00**	2351.23
	0.60	1318.58	**2378.63**	2490.88	2518.44	3258.09	2400.30*	2993.82

续表 2 $\alpha = 0.4$ 时不同估计方法的平均损失

(Table 2 Average losses of different methods, $\alpha = 0.4$ (Continued))

	R^2	MMA	SSMA	Cp	AIC	BIC	SAIC	SBIC
n=200	0.30	389.97	1313.02	1303.74*	1307.07	1588.26	**1146.47**	1427.59
	0.70	2039.64	**3093.40**	3200.13	3235.31	4043.74	3162.97*	3797.05
	0.80	3494.96	**4720.73**	4814.10*	4869.21	5794.40	4820.97	5522.31
	0.90	7806.64	**9474.76**	9509.68*	9581.20	10432.69	9583.80	10209.50
n=600	0.30	284.59	686.23*	720.17	723.18	971.94	**659.51**	904.96
	0.40	429.33	**851.85**	911.04	914.06	1259.83	860.81*	1183.63
	0.50	628.86	**1057.82**	1109.73	1112.65	1529.90	1073.95*	1448.00
	0.60	929.76	**1379.06**	1426.45	1430.39	1916.73	1405.84*	1819.49
	0.70	1435.33	**1885.09**	1926.37	1931.34	2457.55	1916.70*	2340.46
	0.80	2434.03	**2948.95**	2966.49	2975.13	3329.22	2974.84*	3242.28
	0.90	5466.45	**6123.14**	6126.55*	6132.29	6303.80	6142.91	6272.38
n=1000	0.30	246.02	517.13*	552.61	554.10	782.93	**516.19**	745.43
	0.40	366.21	**643.29**	685.65	685.93	1001.48	655.14*	950.95
	0.50	534.58	**830.47**	864.49	868.56	1236.80	849.29*	1182.40
	0.60	786.50	**1086.96**	1116.95	1121.48	1497.04	1108.53*	1441.48
	0.70	1205.97	**1517.35**	1537.68	1541.03	1911.17	1535.67*	1848.57
	0.80	2045.09	**2379.25**	2386.49*	2388.53	2592.77	2390.32	2547.92
	0.90	4561.58	**4969.49**	4971.44*	4972.83	5036.20	4977.08	5030.81
n=1400	0.30	230.15	**426.15**	457.00	457.74	657.69	432.91*	623.60
	0.40	339.07	**544.80**	575.84	575.62	848.40	558.74*	807.67
	0.50	491.43	**703.40**	726.22	728.56	1039.31	717.40*	999.85
	0.60	719.29	**939.93**	958.13	959.64	1271.58	954.20*	1229.49
	0.70	1099.46	**1329.78**	1341.04*	1344.23	1584.58	1342.75	1536.77
	0.80	1861.81	**2107.20**	2111.11*	2112.28	2254.35	2114.49	2220.33
	0.90	4144.44	**4443.53**	4444.77	4445.18	4472.17	4447.15*	4469.37

表 3 $\alpha = 0.6$ 时不同估计方法的平均损失

(Table 3 Average losses of different methods, $\alpha = 0.6$)

	R^2	MMA	SSMA	Cp	AIC	BIC	SAIC	SBIC
n=200	0.3	149.27	1043.14	905.31*	916.15	1054.08	**771.00**	917.73
	0.4	220.70	1167.10	1129.12	1124.91*	1346.91	**988.30**	1182.40
	0.5	321.11	1258.92*	1279.59	1270.37	1591.39	**1129.82**	1449.15
	0.6	473.49	1407.25*	1489.67	1495.13	1910.58	**1353.25**	1724.10
	0.7	729.55	**1686.02**	1799.11	1805.30	2354.29	1691.03*	2156.81
	0.8	1234.40	**2264.92**	2406.67	2419.06	3127.84	2323.90*	2854.42
	0.9	2755.16	**3917.34**	3985.84*	4019.87	4792.16	4000.43	4541.62

续表 3 $\alpha = 0.6$ 时不同估计方法的平均损失

(Table 3 Average losses of different methods, $\alpha = 0.6$ (Continued))

	R^2	MMA	SSMA	Cp	AIC	BIC	SAIC	SBIC
	0.3	105.71	486.17	468.44*	470.45	620.10	**411.35**	565.36
	0.4	150.69	545.70*	558.54	560.31	746.64	**503.89**	687.42
	0.5	212.72	609.15*	646.33	645.94	907.46	**594.53**	835.67
n=600	0.6	306.96	**719.57**	771.22	773.08	1089.85	725.73*	1011.70
	0.7	462.53	**882.24**	933.44	939.57	1325.61	900.80*	1241.86
	0.8	776.85	**1213.61**	1258.08	1262.28	1721.03	1243.55*	1620.16
	0.9	1714.71	**2198.93**	2215.09	2217.90	2544.45	2221.51*	2462.36
	0.3	93.53	351.71*	345.92	345.78	469.01	**308.26**	435.37
	0.4	129.43	391.63*	407.51	407.44	584.11	**373.42**	541.42
	0.5	177.92	443.42*	473.06	473.29	680.39	**442.61**	633.60
n=1000	0.6	252.57	**533.71**	568.87	567.93	829.65	539.03*	786.88
	0.7	376.81	**668.01**	703.66	705.20	993.03	684.25*	944.77
	0.8	621.57	**908.33**	935.73	935.75	1284.21	927.83*	1228.31
	0.9	1361.30	**1672.77**	1681.27*	1682.62	1894.70	1684.92	1847.13
	0.3	89.37	275.40*	282.47	281.72	396.15	**253.48**	369.86
	0.4	120.30	310.15*	332.23	332.96	482.09	**304.04**	449.10
	0.5	164.40	**355.06**	381.69	381.33	582.46	357.74*	547.47
n=1400	0.6	228.53	**425.61**	454.69	454.28	687.11	435.84*	648.42
	0.7	337.24	**545.79**	571.21	571.15	830.78	559.60*	788.58
	0.8	553.81	**766.48**	781.78*	783.04	1055.75	778.53	1014.56
	0.9	1203.66	**1429.64**	1435.36*	1435.55	1579.10	1437.93	1549.25

表 4 $\alpha = 0.8$ 时不同估计方法的平均损失

(Table 4 Average losses of different methods, $\alpha = 0.8$)

	R^2	MMA	SSMA	Cp	AIC	BIC	SAIC	SBIC
	0.3	61.06	969.97	716.28	727.38	772.30	**606.97**	664.10*
	0.4	85.03	1007.30	810.17	808.92	913.30	**682.01**	788.96*
	0.5	118.71	1063.00	933.99*	944.37	1100.28	**810.96**	963.47
n=200	0.6	168.56	1090.79	1049.84	1048.93*	1298.73	**923.32**	1133.97
	0.7	250.89	1103.27*	1153.66	1154.95	1481.76	**1021.04**	1317.93
	0.8	418.11	1363.73*	1453.45	1466.54	1899.22	**1339.00**	1717.43
	0.9	920.23	**1915.95**	2021.42	2020.75	2591.84	1952.75*	2375.56

续表 4 $\alpha = 0.8$ 时不同估计方法的平均损失

(Table 4 Average losses of different methods, $\alpha = 0.8$ (Continued))

	R^2	MMA	SSMA	Cp	AIC	BIC	SAIC	SBIC
	0.3	47.72	441.31	337.46*	339.47	414.70	**290.13**	368.29
	0.4	62.05	438.99	373.69*	378.56	479.54	**322.41**	434.82
	0.5	79.97	464.72	434.15*	434.67	585.30	**379.75**	530.85
n=600	0.6	107.55	503.19*	503.20	505.17	686.38	**449.72**	625.01
	0.7	151.88	537.63*	562.33	564.42	794.79	**512.75**	729.07
	0.8	240.86	**647.39**	693.75	693.76	978.59	649.82*	908.81
	0.9	510.39	**934.48**	977.89	979.15	1337.59	959.79*	1265.77
	0.3	44.95	297.68	233.37*	234.76	318.39	**206.31**	289.28
	0.4	55.83	301.72	270.56*	271.58	367.51	**235.26**	333.66
	0.5	70.47	318.93	309.86*	310.58	435.09	**273.58**	400.62
n=1000	0.6	91.54	356.29*	363.21	363.16	497.83	**328.81**	459.23
	0.7	124.32	391.60*	416.53	416.51	610.91	**380.12**	563.32
	0.8	189.73	**465.03**	495.59	494.89	715.33	469.38*	674.13
	0.9	389.35	**663.58**	694.43	693.64	1008.10	681.02*	941.12
	0.3	43.77	226.54	190.91*	192.95	256.36	**167.83**	235.28
	0.4	53.92	235.95	219.45*	220.20	307.39	**197.01**	283.63
	0.5	66.93	256.29*	256.61	256.85	348.02	**231.49**	326.10
n=1400	0.6	84.45	274.00*	288.25	288.74	408.43	**260.77**	382.24
	0.7	114.69	311.12*	333.42	333.25	492.90	**308.33**	463.44
	0.8	170.11	**368.18**	399.36	400.46	589.22	377.20*	558.83
	0.9	340.67	**551.89**	568.93	568.82	812.76	562.85*	769.18

5 总 结

为了实现对大规模数据的快速分析, 本文结合频率模型平均方法构建了 SSMA 估计. 具体地, 我们首先基于子抽样数据在每个候选模型下得到最小二乘估计, 然后将所有的估计加权起来进行参数推断. 理论上, 我们证明了 SSMA 估计是全部数据下模型平均估计的一个渐近无偏且相合的估计. 另外, 我们基于 Hansen[17] 的 Mallows 模型平均方法提出了 SSMA 估计的权重选择准则, 并证明了方差已知和未知时权重估计的渐近最优性. 值得一提的是, 在这些理论性质的研究中, 我们同时考虑了模型和抽样设计带来的双重随机性. 这虽然使理论的证明变得非常复杂, 但与实际情况更加相符, 理论价值也较高. 本文的理论研究可以类似推广到杠杆值子抽样、最优子抽样方法中. 另外, 重复子抽样不但可以提高估计精度, 而且能够进行方差估计或构造置信区间, 因此重复子抽样模型平均估计值得进一步研究.

参 考 文 献

[1] Lin N, Xi R. Aggregated estimating equation estimation. *Statistics and Its Interface*, 2011, **4**: 73–83.

[2] Chen X, Xie M. A split-and-conquer approach for analysis of extraordinarily large data. *Statistica Sinica*, 2014, **24**: 1655–1684.

[3] Song Q, Liang F. A split-and-merge Bayesian variable selection approach for ultrahigh dimensional regression. *Journal of the Royal Statistical Society, Series B*, 2014, **77**: 947–972.

[4] Schifano E, Wu J, Wang C, et al. Online updating of statistical inference in the big data setting. *Technometrics*, 2016, **58**: 393–403.

[5] Wang C, Chen M, Wu J, et al. Online updating method with new variables for big data streams. *The Canadian Journal of Statistics*, 2018, **46**: 123–146.

[6] Kleiner A, Talwalkar A, Sarkar P, et al. A scalable bootstrap for massive data. *Journal of the Royal Statistical Society, Series B*, 2014, **76**: 795–816.

[7] Ma P, Mahoney M W, Yu B. A statistical perspective on algorithmic leveraging. *Journal of Machine Learning Research*, 2015, **16**: 861–911.

[8] Wang H, Zhu R, Ma P. Optimal subsampling for large sample logistic regression. *Journal of the American Statistical Association*, 2018, **113**: 829–844.

[9] Deldossi L, Tommasi C. Optimal design subsampling from big datasets. *Journal of Quality Technology*, 2021, DOI: 10.1080/00224065.2021.1889418.

[10] Wang H, Yang M, Stufken J. Information-based optimal subdata selection for big data linear regression. *Journal of the American Statistical Association*, 2019, **114**: 393–405.

[11] Liang F, Cheng Y, Song Q, et al. A resampling-based stochastic approximation method for analysis of large geostatistical data. *Journal of the American Statistical Association*, 2013, **108**: 325–339.

[12] Wang H. More efficient estimation for logistic regression with optimal subsamples. *Journal of Machine Learning Research*, 2019, **20**: 1–59.

[13] Ai M, Yu J, Zhang H, et al. Optimal subsampling algorithms for big data regressions. *Statistica Sinica*, 2021, **31**: 749–772.

[14] Buckland S T, Burnham K P, Augustin N H. Model selection: An integral part of inference. *Biometrics*, 1997, **53**: 603–618.

[15] Yang Y. Adaptive regression by mixing. *Journal of the American Statistical Association*, 2001, **96**: 574–588.

[16] Hjort N L, Claeskens G. Frequentist model average estimators. *Journal of the American Statistical Association*, 2003, **98**: 879–899.

[17] Hansen B E. Least squares model averaging. *Econometrica*, 2007, **75**: 1175–1189.

[18] Wan A T K, Zhang X, Zou G. Least squares model averaging by Mallows criterion. *Journal of Econometrics*, 2010, **156**: 277–283.

[19] Hansen B E, Racine J S. Jackknife model averaging. *Journal of Econometrics*, 2012, **167**: 38–46.

[20] Ando T, Li K. A model-averaging approach for high-dimensional regression. *Journal of the American Statistical Association*, 2014, **109**: 254–265.

[21] Zhang X, Liang H. Focused information criterion and model averaging for generalized additive partial linear models. *Annals of Statistics*, 2011, **39**: 174–200.

[22] Zhang X, Wan A, Zou G. Model averaging by Jackknife criterion in models with dependent data. *Journal of Econometrics*, 2013, **174**: 82–94.

[23] Hansen B E. Model averaging, asymptotic risk, and regressor groups. *Quantitative Economics*, 2014, **5**: 495–530.

[24] Zhang X, Zou G, Liang H. Model averaging and weight choice in linear mixed effects models. *Biometrika*, 2014, **101**: 205–218.

[25] Zhang X, Yu D, Zou G, et al. Optimal model averaging estimation for generalized linear models and generalized linear mixed-effects models. *Journal of the American Statistical Association*, 2016, **111**: 1775–1790.

[26] Liao J, Zong X, Zhang X, et al. Model averaging based on leave-subject-out cross-validation for vector autoregressions. *Journal of Econometrics*, 2018, **209**: 35–60.

[27] Gao Y, Zhang X, Wang S, et al. Model averaging based on leave-subject-out cross-validation. *Journal of Econometrics*, 2016, **192**: 139–151.

[28] Li K C. Asymptotic optimality for Cp, CL, cross validation and generalized cross-validations: Discrete index set. *Annals of Statistics*, 1987, **15**: 958–975.

[29] Whittle P. Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability and Its Applications*, 1960, **5**: 302–305.

附 录

该附录包含所有引理和定理的证明. 如无特殊说明, 字母 C 在不同的地方表示不同的常数; 一个矩阵的范数指的是 2- 范数, 即该矩阵的最大奇异值.

A.1 定理的证明

定理 3.1 的证明 根据 c_r 不等式和三角不等式有

$$\begin{aligned}
E\left\|\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})-\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right\|^2 &= E\left\|\sum_{m=1}^{M}\omega_m\left(\widehat{\boldsymbol{\beta}}_{(m)}-\widetilde{\boldsymbol{\beta}}_{(m)}\right)\right\|^2 \\
&= E\left\|\sum_{m=1}^{M}\omega_m\left(\widehat{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}^*_{(m)}+\boldsymbol{\beta}^*_{(m)}-\widetilde{\boldsymbol{\beta}}_{(m)}\right)\right\|^2 \\
&\le 2E\left\|\sum_{m=1}^{M}\omega_m\left(\widehat{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}^*_{(m)}\right)\right\|^2+2E\left\|\sum_{m=1}^{M}\omega_m\left(\boldsymbol{\beta}^*_{(m)}-\widetilde{\boldsymbol{\beta}}_{(m)}\right)\right\|^2 \\
&\le CM\sum_{m=1}^{M}\omega_m^2E\left\|\widehat{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}^*_{(m)}\right\|^2+CM\sum_{m=1}^{M}\omega_m^2E\left\|\boldsymbol{\beta}^*_{(m)}-\widetilde{\boldsymbol{\beta}}_{(m)}\right\|^2,
\end{aligned}$$

再结合引理 2 和 4 得到

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}E\left\|\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})-\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right\|^2=O\left(n^{-1}Mk_M\right).$$

定理 3.1 证毕.

定理 3.2 的证明 注意到

$$\begin{aligned}
R_N(\boldsymbol{\omega}) &= E\left\|\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})+\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right\|^2 \\
&= E_m\left\|\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2+E\left\|\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right\|^2 \\
&\quad +2E\left(\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right)'\left(\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right).
\end{aligned}$$

又因为

$$E\left(\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right)'\left(\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right)\le\sqrt{E\left\|\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2E\left\|\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right\|^2}$$

$$\leq \sqrt{\left(2R_N(\boldsymbol{\omega})+2E\left\|\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right\|^2\right)E\left\|\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right\|^2}$$
$$\leq \sqrt{2R_N(\boldsymbol{\omega})E\left\|\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right\|^2+2\left(E\left\|\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right\|^2\right)^2},$$

再结合引理 6 和条件 5 得到

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{R_N(\boldsymbol{\omega})-\widetilde{R}_N(\boldsymbol{\omega})}{R_N(\boldsymbol{\omega})}\to 0.$$

定理 3.2 证毕.

定理 3.3 的证明 令 $\boldsymbol{e}=[e_i]_{i\in s}$, $\boldsymbol{\mu}=[\mu_i]_{i\in s}$. 则

$$\begin{aligned}C_s(\boldsymbol{\omega})&=\left\|\boldsymbol{Y}-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2+2f\sigma^2k(\boldsymbol{\omega})\\&=\left\|\boldsymbol{\mu}+\boldsymbol{e}-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2+2f\sigma^2k(\boldsymbol{\omega})\\&=\left\|\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2+2\boldsymbol{e}'\Big(\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)+\|\boldsymbol{e}\|^2+2f\sigma^2k(\boldsymbol{\omega}).\end{aligned}\tag{A.1}$$

注意到

$$\begin{aligned}2\boldsymbol{e}'\Big(\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)&=2\boldsymbol{e}'\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})+\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\\&=2\boldsymbol{e}'\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)+2\boldsymbol{e}'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\end{aligned}\tag{A.2}$$

且

$$\begin{aligned}\left\|\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2&=\left\|\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})+\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2\\&=\left\|\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\right\|^2+\left\|\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2\\&\quad+2\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\end{aligned}\tag{A.3}$$

这里 $\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})=\boldsymbol{X}_{(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})$. 令 $\widetilde{\boldsymbol{P}}_{(m)}=\widetilde{\boldsymbol{X}}_{(m)}\left(\widetilde{\boldsymbol{X}}'_{(m)}\widetilde{\boldsymbol{X}}_{(m)}\right)^{-1}\widetilde{\boldsymbol{X}}'_{(m)}$, $\widetilde{\boldsymbol{P}}(\boldsymbol{\omega})=\sum_{m=1}^{M}\omega_m\widetilde{\boldsymbol{P}}_{(m)}$ 且 $\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})=\widetilde{\boldsymbol{X}}_{(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})$, 则有

$$\begin{aligned}E\left\{2\boldsymbol{e}'\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)\right\}&=2E\Big(\boldsymbol{e}'\boldsymbol{\mu}-\boldsymbol{e}'\boldsymbol{X}_{(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\Big)\\&=2E_mE_d\left(\sum_{i\in s}e_i\mu_i\right)-2E_mE_d\left(\sum_{i\in s}\left[e_i\boldsymbol{x}'_{i(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right]\right)\\&=2E_m\left(\frac{n}{N}\sum_{i\in D_N}e_i\mu_i\right)-2E_m\left(\frac{n}{N}\sum_{i\in D_N}\left[e_i\boldsymbol{x}'_{i(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right]\right)\\&=2fE_m\Big(\widetilde{\boldsymbol{e}}'\widetilde{\boldsymbol{\mu}}\Big)-2fE_m\Big(\widetilde{\boldsymbol{e}}'\widetilde{\boldsymbol{X}}_{(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\Big)\\&=-2fE_m\Big(\widetilde{\boldsymbol{e}}'\widetilde{\boldsymbol{P}}(\boldsymbol{\omega})\widetilde{\boldsymbol{e}}\Big)\\&=-2f\sigma^2k(\boldsymbol{\omega})\end{aligned}\tag{A.4}$$

和

$$E\left\|\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\right\|^2 = E_m E_d\left\{\sum_{i\in s}\left(\mu_i-\boldsymbol{x}'_{i(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right)^2\right\}$$
$$= fE_m\left\{\sum_{i\in D_N}\left(\mu_i-\boldsymbol{x}'_{i(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right)^2\right\}$$
$$= fE_m\left\|\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2. \tag{A.5}$$

又因为

$$E\|\boldsymbol{e}\|^2 = E_m E_d\left(\sum_{i\in s}e_i^2\right) = fE_m\left(\sum_{i\in D_N}e_i^2\right) = Nf\sigma^2, \tag{A.6}$$

所以结合 (A.1)–(A.6) 式, 得到

$$E\Big(C_s(\boldsymbol{\omega})\Big) = fE_m\left\|\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2 + E\left\|\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2 + Nf\sigma^2$$
$$+ 2E\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big) + 2E\left\{\boldsymbol{e}'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\right\}. \tag{A.7}$$

因此, 根据定理 3.2, 为了完成整个定理的证明, 我们需验证以下结论

$$\frac{E\left\|\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2}{f\xi_N(\boldsymbol{\omega})}\to 0, \tag{A.8}$$

$$\frac{E\left|\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\right|}{f\xi_N(\boldsymbol{\omega})}\to 0, \tag{A.9}$$

和

$$\frac{E\left|\boldsymbol{e}'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\right|}{f\xi_N(\boldsymbol{\omega})}\to 0, \tag{A.10}$$

这里 $\xi_N=\inf_{\boldsymbol{\omega}\in\mathscr{H}}R_N(\boldsymbol{\omega})$.

根据引理 5 和条件 5 可知 (A.8) 式成立. 接下来证明 (A.9) 式. 利用柯西不等式和 (A.8) 式有

$$E\left|\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\right| \le \sqrt{E\left\|\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\right\|^2 E\left\|\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2}$$
$$= \sqrt{fE_m\left\|\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2 E\left\|\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2},$$

再结合条件 5, 引理 5 和定理 3.1 推出 (A.9) 式成立. 因为

$$E\left|\boldsymbol{e}'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\right| \le \sqrt{E\|\boldsymbol{e}\|^2\cdot E\left\|\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2},$$

所以根据条件 6, 引理 5 和 (A.6) 式推出 (A.10) 式成立.

定理 3.3 证毕.

定理 3.4 的证明 根据 (A.1)–(A.3) 式, 有

$$C_s(\boldsymbol{\omega}) = \left\|\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\right\|^2 + 2\boldsymbol{e}'\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big) + 2f\sigma^2k(\boldsymbol{\omega}) + \left\|\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2$$

$$+2\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)+2\boldsymbol{e}'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)+\|\boldsymbol{e}\|^2. \tag{A.11}$$

利用马尔科夫不等式和 (A.8)–(A.10) 式推出

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\Big\|\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big\|^2}{fR_N(\boldsymbol{\omega})}\xrightarrow{p}0,$$

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)}{fR_N(\boldsymbol{\omega})}\xrightarrow{p}0,$$

和

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\boldsymbol{e}'\Big(\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)}{fR_N(\boldsymbol{\omega})}\xrightarrow{p}0,$$

这里依概率收敛同时考虑了抽样设计和模型的双重随机性. 因此, 根据 Li[28] 和 Wan 等 [18] 可知, 为了完成整个定理的证明, 我们只需证明以下结论成立

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\boldsymbol{e}'\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)+f\sigma^2k(\boldsymbol{\omega})}{fR_N(\boldsymbol{\omega})}\xrightarrow{p}0, \tag{A.12}$$

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\Big\|\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big\|^2-fL_N(\boldsymbol{\omega})}{fR_N(\boldsymbol{\omega})}\xrightarrow{p}0, \tag{A.13}$$

且

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\left\{\frac{L_N(\boldsymbol{\omega})}{R_N(\boldsymbol{\omega})}-1\right\}\xrightarrow{p}0. \tag{A.14}$$

为了证明 (A.12) 式, 我们只需证明

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\boldsymbol{e}'\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)-f\widetilde{\boldsymbol{e}}'\Big(\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)}{fR_N(\boldsymbol{\omega})}\xrightarrow{p}0 \tag{A.15}$$

和

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\widetilde{\boldsymbol{e}}'\Big(\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)+\sigma^2k(\boldsymbol{\omega})}{R_N(\boldsymbol{\omega})}\xrightarrow{p}0. \tag{A.16}$$

注意到

$$\widetilde{\boldsymbol{e}}'\Big(\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)+\sigma^2k(\boldsymbol{\omega})=\widetilde{\boldsymbol{e}}'\Big(\boldsymbol{I}_N-\widetilde{\boldsymbol{P}}(\boldsymbol{\omega})\Big)\widetilde{\boldsymbol{\mu}}+\Big\{\sigma^2k(\boldsymbol{\omega})-\widetilde{\boldsymbol{e}}'\widetilde{\boldsymbol{P}}(\boldsymbol{\omega})\widetilde{\boldsymbol{e}}\Big\},$$

所以根据 Wan 等 [18] 中定理 1 证明的 (A.1) 和 (A.2) 式, 以及条件 7 推出

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\widetilde{\boldsymbol{e}}'\Big(\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)+\sigma^2k(\boldsymbol{\omega})}{\widetilde{R}_N(\boldsymbol{\omega})}\xrightarrow{p}0.$$

根据定理 3.2 知 (A.16) 式成立. 接下来证明 (A.15) 式. 用 I_i 表示第 i 个样本被抽中, $\widetilde{T}_{i(m)}=\mu_i-\boldsymbol{x}'_{i(m)}\widetilde{\boldsymbol{\beta}}_{(m)}$. 则对任意的 $\delta>0$, 根据 Bonferroni 不等式, 马尔科夫不等式和三角不等式有

$$P\left\{\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\boldsymbol{e}'\Big(\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\Big)-f\widetilde{\boldsymbol{e}}'\Big(\widetilde{\boldsymbol{\mu}}-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)}{fR_N(\boldsymbol{\omega})}>\delta\right\}$$

$$
\begin{aligned}
&\leq P\left\{\sup_{\boldsymbol{\omega}\in\mathscr{H}}\left|\sum_{i\in s}e_i\left(\mu_i-\boldsymbol{x}'_{i(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right)-f\sum_{i\in D_N}e_i\left(\mu_i-\boldsymbol{x}'_{i(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right)\right|>f\delta\xi_N\right\}\\
&=P\left\{\sup_{\boldsymbol{\omega}\in\mathscr{H}}\left|\sum_{i\in D_N}(I_i-f)\left(\mu_i-\boldsymbol{x}'_{i(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right)e_i\right|>f\delta\xi_N\right\}\\
&\leq P\left\{\sup_{\boldsymbol{\omega}\in\mathscr{H}}\sum_{m=1}^{M}\omega_m\left|\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}e_i\right|>f\delta\xi_N\right\}\\
&\leq\sum_{m=1}^{M}P\left\{\left|\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}e_i\right|>f\delta\xi_N\right\}\\
&\leq f^{-2}\delta^{-2}\xi_N^{-2}\sum_{m=1}^{M}E\left\{\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}e_i\right\}^2.
\end{aligned}
$$

再结合引理 7 推出 (A.15) 式成立. 因此, (A.12) 式成立.

为了证明 (A.13) 式, 我们需证明

$$
\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\left\|\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\right\|^2-f\widetilde{L}_N(\boldsymbol{\omega})}{fR_N(\boldsymbol{\omega})}\xrightarrow{p}0 \tag{A.17}
$$

和

$$
\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\widetilde{L}_N(\boldsymbol{\omega})-L_N(\boldsymbol{\omega})}{R_N(\boldsymbol{\omega})}\xrightarrow{p}0. \tag{A.18}
$$

由引理 8 可知 (A.18) 式成立. 下面证明 (A.17) 式. 注意到, 对任意的 $\delta>0$, 根据 Bonferroni 不等式, 马尔科夫不等式和三角不等式有

$$
\begin{aligned}
&P\left\{\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\left|\left\|\boldsymbol{\mu}-\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})\right\|^2-f\widetilde{L}_N(\boldsymbol{\omega})\right|}{fR_N(\boldsymbol{\omega})}>\delta\right\}\\
&\leq P\left\{\sup_{\boldsymbol{\omega}\in\mathscr{H}}\left|\sum_{i\in s}\left(\mu_i-\boldsymbol{x}'_{i(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right)^2-f\sum_{i\in D_N}\left(\mu_i-\boldsymbol{x}'_{i(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right)^2\right|>f\delta\xi_N\right\}\\
&=P\left\{\sup_{\boldsymbol{\omega}\in\mathscr{H}}\left|\sum_{i\in D_N}(I_i-f)\left(\mu_i-\boldsymbol{x}'_{i(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right)^2\right|>f\delta\xi_N\right\}\\
&\leq P\left\{\sup_{\boldsymbol{\omega}\in\mathscr{H}}\sum_{m=1}^{M}\sum_{t=1}^{M}\omega_m\omega_t\left|\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}\widetilde{T}_{i(t)}\right|>f\delta\xi_N\right\}\\
&\leq\sum_{m=1}^{M}\sum_{t=1}^{M}P\left\{\left|\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}\widetilde{T}_{i(t)}\right|>f\delta\xi_N\right\}\\
&\leq f^{-2}\delta^{-2}\xi_N^{-2}\sum_{m=1}^{M}\sum_{t=1}^{M}E\left\{\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}\widetilde{T}_{i(t)}\right\}^2.
\end{aligned}
$$

再结合引理 7 推出 (A.17) 式成立.

最后, 我们证明 (A.14) 式. 注意到,

$$L_N(\boldsymbol{\omega}) - R_N(\boldsymbol{\omega}) = L_N(\boldsymbol{\omega}) - \widetilde{L}_N(\boldsymbol{\omega}) + \widetilde{L}_N(\boldsymbol{\omega}) - \widetilde{R}_N(\boldsymbol{\omega}) + \widetilde{R}_N(\boldsymbol{\omega}) - R_N(\boldsymbol{\omega}).$$

因此, 为了证明 (A.14) 式成立, 根据引理 8 和定理 3.2, 我们只需证明

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}} \frac{\widetilde{L}_N(\boldsymbol{\omega}) - \widetilde{R}_N(\boldsymbol{\omega})}{R_N(\boldsymbol{\omega})} \xrightarrow{p} 0. \tag{A.19}$$

根据 Wan 等 [18] 中定理 1 证明的 (A.3) 式, 以及条件 7 推出

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}} \frac{\widetilde{L}_N(\boldsymbol{\omega}) - \widetilde{R}_N(\boldsymbol{\omega})}{\widetilde{R}_N(\boldsymbol{\omega})} \xrightarrow{p} 0,$$

进一步结合定理 3.2 可知 (A.19) 式成立.

定理 3.4 证毕.

定理 3.5 的证明 根据 Li[28] 和 (A.11) 式可知, 我们只需证明

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}} \frac{k(\boldsymbol{\omega}) \cdot \left|\widehat{\sigma}^2_{(M)} - \sigma^2\right|}{R_N(\boldsymbol{\omega})} \xrightarrow{p} 0. \tag{A.20}$$

注意到

$$\left|\widehat{\sigma}^2_{(M)} - \sigma^2\right| \le \left|\widehat{\sigma}^2_{(M)} - \widetilde{\sigma}^2_{(M)}\right| + \left|\widetilde{\sigma}^2_{(M)} - \sigma^2\right|,$$

这里

$$\widetilde{\sigma}^2_{(M)} = \frac{1}{N-k_M}\left\|\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{X}}_{(M)}\widetilde{\boldsymbol{\beta}}_{(M)}\right\|^2.$$

因此, 为了证明 (A.20) 式, 我们需证明

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}} \frac{k(\boldsymbol{\omega}) \cdot \left|\widehat{\sigma}^2_{(M)} - \widetilde{\sigma}^2_{(M)}\right|}{R_N(\boldsymbol{\omega})} \xrightarrow{p} 0 \tag{A.21}$$

和

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}} \frac{k(\boldsymbol{\omega}) \cdot \left|\widetilde{\sigma}^2_{(M)} - \sigma^2\right|}{R_N(\boldsymbol{\omega})} \xrightarrow{p} 0. \tag{A.22}$$

根据 Wan 等 [18] 的 (A.6) 式和条件 8 有

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}} \frac{k(\boldsymbol{\omega}) \cdot \left|\widetilde{\sigma}^2_{(M)} - \sigma^2\right|}{\widetilde{R}_N(\boldsymbol{\omega})} \xrightarrow{p} 0,$$

再根据定理 3.2 知 (A.22) 式成立. 下面证明 (A.21) 式. 注意到

$$\begin{aligned}
\sup_{\boldsymbol{\omega}\in\mathscr{H}} \frac{k(\boldsymbol{\omega}) \cdot \left|\widehat{\sigma}^2_{(M)} - \widetilde{\sigma}^2_{(M)}\right|}{R_N(\boldsymbol{\omega})} &\le \frac{k_M \left|\widehat{\sigma}^2_{(M)} - \widetilde{\sigma}^2_{(M)}\right|}{\xi_N} \\
&\le \frac{k_M \left|\left\|\boldsymbol{Y} - \boldsymbol{X}_{(M)}\widehat{\boldsymbol{\beta}}_{(M)}\right\|^2 - f\left\|\widetilde{\boldsymbol{Y}} - \widetilde{\boldsymbol{X}}_{(M)}\widetilde{\boldsymbol{\beta}}_{(M)}\right\|^2\right|}{(N-k_M)f\xi_N} \\
&\le \frac{k_M \left|\sum_{i\in s}\widetilde{T}^2_{i(M)} - f\sum_{i\in D_N}\widetilde{T}^2_{i(M)}\right|}{(N-k_M)f\xi_N}.
\end{aligned}$$

根据简单随机抽样方差公式, 条件 8 和 (A.25) 式有

$$\begin{aligned} E\left(\sum_{i\in s}\widetilde{T}^2_{i(M)}-f\sum_{i\in D_N}\widetilde{T}^2_{i(M)}\right)^2 &= n^2E\left(\frac{1}{n}\sum_{i\in s}\widetilde{T}^2_{i(M)}-\frac{1}{N}\sum_{i\in D_N}\widetilde{T}^2_{i(M)}\right)^2 \\ &\le \frac{n^2(1-f)}{n(N-1)}\sum_{i\in D_N}\left(\widetilde{T}^2_{i(M)}-\frac{1}{N}\sum_{i\in D_N}\widetilde{T}^2_{i(M)}\right)^2 \\ &= O(n), \end{aligned}$$

再结合条件 6 知 (A.21) 式成立.

定理 3.5 证毕.

A.2 引理的证明

引理 1 如果条件 1–2 成立, 则

$$E_m\left\|\frac{1}{N}\sum_{i\in D_N}\boldsymbol{x}'_{i(m)}e^*_{i(m)}\right\|^{2q}=O\left(\frac{k_m}{N}\right)^q.$$

证 利用 c_r 不等式, 有

$$\begin{aligned} E_m\left\|\frac{1}{N}\sum_{i\in D_N}\boldsymbol{x}'_{i(m)}e^*_{i(m)}\right\|^{2q} &= N^{-2q}\cdot E_m\left|\sum_{j=1}^{k_m}\left(\sum_{i\in D_N}x_{ij}e^*_{i(m)}\right)^2\right|^q \\ &\le N^{-2q}\cdot {k_m}^{q-1}\cdot\sum_{j=1}^{k_m}E_m\left|\sum_{i\in D_N}x_{ij}e^*_{i(m)}\right|^{2q}. \end{aligned}$$

根据 (3.7) 式, $E_m\left\{x_{ij}\left(y_i-\boldsymbol{x}'_{i(m)}\boldsymbol{\beta}^*_{(m)}\right)\right\}=0$. 再利用 Whittle[29] 定理 2 得到

$$E_m\left|\sum_{i\in D_N}x_{ij}e^*_{i(m)}\right|^{2q}\le 2^{2q}\cdot C(2q)\cdot\left(\sum_{i\in D_N}\left(E_m\left|x_{ij}e^*_{i(m)}\right|^{2q}\right)^{\frac{1}{q}}\right)^q. \tag{A.23}$$

进一步, 结合条件 1 和 2 有

$$E\left\|\frac{1}{N}\sum_{i\in D_N}\boldsymbol{x}'_{i(m)}e^*_{i(m)}\right\|^{2q}=O\left(\frac{k_m}{N}\right)^q.$$

引理 2 如果条件 1–3 成立, 则

$$E_m\left\|\widetilde{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}^*_{(m)}\right\|^{2q}=O\left(\frac{k_m}{N}\right)^q.$$

证 注意到,

$$\begin{aligned} E_m\left\|\widetilde{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}^*_{(m)}\right\|^{2q} &= E_m\left\|\left(\sum_{i\in D_N}\boldsymbol{x}_{i(m)}\boldsymbol{x}'_{i(m)}\right)^{-1}\sum_{i\in D_N}\boldsymbol{x}_{i(m)}y_i-\boldsymbol{\beta}^*_{(m)}\right\|^{2q} \\ &= E_m\left\|\left(\sum_{i\in D_N}\boldsymbol{x}_{i(m)}\boldsymbol{x}'_{i(m)}\right)^{-1}\sum_{i\in D_N}\boldsymbol{x}_{i(m)}e^*_{i(m)}\right\|^{2q} \end{aligned}$$

$$\leq \left\| \left(\frac{1}{N} \sum_{i \in D_N} \boldsymbol{x}_{i(m)} \boldsymbol{x}'_{i(m)} \right)^{-1} \right\|^{2q} \cdot E_m \left\| \frac{1}{N} \sum_{i \in D_N} \boldsymbol{x}_{i(m)} e^*_{i(m)} \right\|^{2q}$$

根据条件 3 和引理 1 即可得到结论.

引理 3 如果条件 1–2 成立, 则

$$E \left\| \frac{1}{n} \sum_{i \in s} \boldsymbol{x}'_{i(m)} e^*_{i(m)} \right\|^2 = O\left(\frac{k_m}{n} \right),$$

这里 $E(\cdot) = E_m E_d(\cdot)$.

证 用 I_i 表示第 i 个总体单元被抽到, 则

$$\begin{aligned}
& E \left(\frac{1}{n} \sum_{i \in s} x_{ij} e^*_{i(m)} - \frac{1}{N} \sum_{i \in D_N} x_{ij} e^*_{i(m)} \right)^2 \\
=& E_m E_d \left(\frac{1}{n} \sum_{i \in s} x_{ij} e^*_{i(m)} - \frac{1}{N} \sum_{i \in D_N} x_{ij} e^*_{i(m)} \right)^2 \\
=& \frac{1-f}{n(N-1)} \sum_{i \in D_N} E_m \left(x_{ij} e^*_{i(m)} - \frac{1}{N} \sum_{i \in D_N} x_{ij} e^*_{i(m)} \right)^2 \\
=& O(n^{-1}),
\end{aligned}$$

这里第二个等式由简单随机抽样的方差公式推出, 第三个等式是根据条件 1 和 2 得到. 因此, 结合引理 1 有

$$\begin{aligned}
E \left\| \frac{1}{n} \sum_{i \in s} \boldsymbol{x}_{i(m)} e^*_{i(m)} \right\|^2 &= \sum_{j=1}^{k_m} E \left(\frac{1}{n} \sum_{i \in s} x_{ij} e^*_{i(m)} \right)^2 \\
&\leq C \sum_{j=1}^{k_m} \left\{ E \left(\frac{1}{n} \sum_{i \in s} x_{ij} e^*_{i(m)} - \frac{1}{N} \sum_{i \in D_N} x_{ij} e^*_{i(m)} \right)^2 \right. \\
&\quad \left. + E_m \left(\frac{1}{N} \sum_{i \in D_N} x_{ij} e^*_{i(m)} \right)^2 \right\} \\
&= O\left(\frac{k_m}{n} \right). \qquad \text{(A.24)}
\end{aligned}$$

引理 4 如果条件 1, 2 和 4 成立, 则

$$E \left\| \widehat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}^*_{(m)} \right\|^2 = O\left(\frac{k_m}{n} \right).$$

证 注意到,

$$\begin{aligned}
E \left\| \widehat{\boldsymbol{\beta}}_{(m)} - \boldsymbol{\beta}^*_{(m)} \right\|^2 &= E \left\| \left(\sum_{i \in s} \boldsymbol{x}_{i(m)} \boldsymbol{x}'_{i(m)} \right)^{-1} \sum_{i \in s} \boldsymbol{x}_{i(m)} y_i - \boldsymbol{\beta}^*_{(m)} \right\|^2 \\
&= E \left\| \left(\sum_{i \in s} \boldsymbol{x}_{i(m)} \boldsymbol{x}'_{i(m)} \right)^{-1} \sum_{i \in s} \boldsymbol{x}_{i(m)} e^*_{i(m)} \right\|^2
\end{aligned}$$

$$\leq CE\left\|\frac{1}{n}\sum_{i\in s}\boldsymbol{x}_{i(m)}e^*_{i(m)}\right\|^2,$$

这里最后一行是由于条件 4 成立. 再结合引理 3 得到

$$E\left\|\widehat{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}^*_{(m)}\right\|^2=O\left(\frac{k_m}{n}\right).$$

引理 5 如果条件 1–4 成立, 则

$$E\left\|\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2=O(Mk_M^2).$$

证 利用 c_r 不等式和条件 4 得到

$$\begin{aligned}E\left\|\widetilde{\boldsymbol{\mu}}_s(\boldsymbol{\omega})-\widehat{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2&=E\left\|\boldsymbol{X}_{(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})-\boldsymbol{X}_{(M)}\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})\right\|^2\\&=E\left\|\sum_{m=1}^{M}\omega_m\cdot\left(\boldsymbol{X}_{(m)}\widetilde{\boldsymbol{\beta}}_{(m)}-\boldsymbol{X}_{(m)}\widehat{\boldsymbol{\beta}}_{(m)}\right)\right\|^2\\&\leq CM\sum_{m=1}^{M}\left\{\omega_m\cdot\lambda_{\max}\left(\boldsymbol{X}'_{(m)}\boldsymbol{X}_{(m)}\right)\cdot E\left\|\widetilde{\boldsymbol{\beta}}_{(m)}-\widehat{\boldsymbol{\beta}}_{(m)}\right\|^2\right\}\\&\leq CnM\sum_{m=1}^{M}\left\{\omega_m k_m\cdot E\left\|\widetilde{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}^*_{(m)}+\boldsymbol{\beta}^*_{(m)}-\widehat{\boldsymbol{\beta}}_{(m)}\right\|^2\right\}\\&=O(Mk_M^2),\end{aligned}$$

这里最后一行由引理 2 和 4 推出.

引理 6 如果条件 1–4 成立, 则

$$E\left\|\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2=O(f^{-1}Mk_M^2).$$

证 利用 c_r 不等式和条件 4 得到

$$\begin{aligned}E\left\|\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2&=E\left\|\widetilde{\boldsymbol{X}}_{(M)}\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})-\widetilde{\boldsymbol{X}}_{(M)}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right\|^2\\&=E\left\|\sum_{m=1}^{M}\omega_m\cdot\left(\widetilde{\boldsymbol{X}}_{(m)}\widehat{\boldsymbol{\beta}}_{(m)}-\widetilde{\boldsymbol{X}}_{(m)}\widetilde{\boldsymbol{\beta}}_{(m)}\right)\right\|^2\\&\leq CM\sum_{m=1}^{M}\left\{\omega_m\cdot\lambda_{\max}\left(\widetilde{\boldsymbol{X}}'_{(m)}\widetilde{\boldsymbol{X}}_{(m)}\right)\cdot E\left\|\widetilde{\boldsymbol{\beta}}_{(m)}-\widehat{\boldsymbol{\beta}}_{(m)}\right\|^2\right\}\\&\leq CNM\sum_{m=1}^{M}\left\{\omega_m k_m\cdot E\left\|\widetilde{\boldsymbol{\beta}}_{(m)}-\boldsymbol{\beta}^*_{(m)}+\boldsymbol{\beta}^*_{(m)}-\widehat{\boldsymbol{\beta}}_{(m)}\right\|^2\right\}\\&=O(f^{-1}Mk_M^2),\end{aligned}$$

这里最后一行由引理 2 和 4 推出.

引理 7 如果条件 1–6 成立, 则

$$f^{-2}\xi_N^{-2}\sum_{m=1}^{M}E\left\{\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}e_i\right\}^2\to 0$$

和

$$f^{-2}\xi_N^{-2}\sum_{m=1}^{M}\sum_{t=1}^{M}E\left\{\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}\widetilde{T}_{i(t)}\right\}^2\to 0,$$

这里 $\widetilde{T}_{i(m)}=\mu_i-\boldsymbol{x}'_{i(m)}\widetilde{\boldsymbol{\beta}}_{(m)}$.

证 利用引理 2, 条件 1 和 2 推出

$$\begin{aligned}
E_m\left(\widetilde{T}^4_{i(m)}\right)&=E_m\left(\mu_i-\boldsymbol{x}'_{i(m)}\widetilde{\boldsymbol{\beta}}_{(m)}\right)^4\\
&=E_m\left|y_i-\boldsymbol{x}'_{i(m)}\boldsymbol{\beta}^*_{(m)}+\boldsymbol{x}'_{i(m)}\boldsymbol{\beta}^*_{(m)}-\boldsymbol{x}'_{i(m)}\widetilde{\boldsymbol{\beta}}_{(m)}-e_i\right|^4\\
&\le CE_m\left|e^*_{i(m)}\right|^4+CE_m\left|\boldsymbol{x}'_{i(m)}\boldsymbol{\beta}^*_{(m)}-\boldsymbol{x}'_{i(m)}\widetilde{\boldsymbol{\beta}}_{(m)}\right|^4+CE_m\left|e_i\right|^4\\
&\le C\left\|\boldsymbol{x}_{i(m)}\right\|^4E_m\left\|\boldsymbol{\beta}^*_{(m)}-\widetilde{\boldsymbol{\beta}}_{(m)}\right\|^4+CE_m\left|e_i\right|^4+CE_m\left|e^*_{i(m)}\right|^4\\
&\le CN^{-2}k_m^4+C, \qquad (A.25)
\end{aligned}$$

再根据简单随机抽样的方差公式和条件 1 得到

$$\begin{aligned}
E\left\{\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}e_i\right\}^2&=E_mE_d\left\{\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}e_i\right\}^2\\
&=n^2E_m\left\{\frac{1-f}{n(N-1)}\sum_{i\in D_N}\left(\widetilde{T}_{i(m)}e_i-\frac{1}{N}\sum_{i\in D_N}\widetilde{T}_{i(m)}e_i\right)^2\right\}\\
&=\frac{n(1-f)}{(N-1)}\sum_{i\in D_N}E_m\left(\widetilde{T}_{i(m)}e_i-\frac{1}{N}\sum_{i\in D_N}\widetilde{T}_{i(m)}e_i\right)^2\\
&\le Cfk_M^2+nC
\end{aligned}$$

和

$$\begin{aligned}
E\left\{\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}\widetilde{T}_{i(t)}\right\}^2&=n^2E_m\left\{\frac{1-f}{n(N-1)}\sum_{i\in D_N}\left(\widetilde{T}_{i(m)}\widetilde{T}_{i(t)}-\frac{1}{N}\sum_{i\in D_N}\widetilde{T}_{i(m)}\widetilde{T}_{i(t)}\right)^2\right\}\\
&=\frac{n(1-f)}{(N-1)}\sum_{i\in D_N}E_m\left(\widetilde{T}_{i(m)}\widetilde{T}_{i(t)}-\frac{1}{N}\sum_{i\in D_N}\widetilde{T}_{i(m)}\widetilde{T}_{i(t)}\right)^2\\
&\le CfN^{-1}k_M^4+nC.
\end{aligned}$$

因此根据条件 5 和 6 有

$$f^{-2}\xi_N^{-2}\sum_{m=1}^{M}E\left\{\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}e_i\right\}^2\to 0$$

和

$$f^{-2}\xi_N^{-2}\sum_{m=1}^{M}\sum_{t=1}^{M}E\left\{\sum_{i\in D_N}(I_i-f)\widetilde{T}_{i(m)}\widetilde{T}_{i(t)}\right\}^2\to 0.$$

引理 8 如果条件 1–6 成立, 则

$$\sup_{\boldsymbol{\omega}\in\mathscr{H}}\frac{\widetilde{L}_N(\boldsymbol{\omega})-L_N(\boldsymbol{\omega})}{R_N(\boldsymbol{\omega})}\xrightarrow{p}0.$$

证 注意到

$$\begin{aligned}\widetilde{L}_N(\boldsymbol{\omega})&=\left\|\widetilde{\boldsymbol{\mu}}-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})+\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2\\&=L_N(\boldsymbol{\omega})+\left\|\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2+2\Big(\widetilde{\boldsymbol{\mu}}-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\Big)'\Big(\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\end{aligned}$$

因此, 为了完成引理的证明, 我们只需验证

$$\xi_N^{-1}\cdot E\left\|\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2=o(1)\tag{A.26}$$

和

$$\xi_N^{-1}\cdot E\left|\Big(\widetilde{\boldsymbol{\mu}}-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\Big)'\Big(\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\right|=o(1).\tag{A.27}$$

根据定理 3.1 和条件 3 有

$$\begin{aligned}E\left\|\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2&=E\left\|\widetilde{\boldsymbol{X}}_{(M)}\Big(\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})-\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\Big)\right\|^2\\&\le\left\|\widetilde{\boldsymbol{X}}_{(M)}\right\|^2\cdot E\left\|\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})-\widetilde{\boldsymbol{\beta}}(\boldsymbol{\omega})\right\|^2\\&=O(f^{-1}Mk_M^2),\end{aligned}\tag{A.28}$$

再结合条件 5 得到 (A.26) 式成立.

令 $\widehat{T}_{i(m)}=\mu_i-\boldsymbol{x}'_{i(m)}\widehat{\boldsymbol{\beta}}_{(m)}$. 利用引理 4, 条件 1 和 2 推出

$$\begin{aligned}E_m\left(\widehat{T}_{i(m)}^4\right)&=E_m\left(\mu_i-\boldsymbol{x}'_{i(m)}\widehat{\boldsymbol{\beta}}_{(m)}\right)^4\\&=E_m\left|y_i-\boldsymbol{x}'_{i(m)}\boldsymbol{\beta}^*_{(m)}+\boldsymbol{x}'_{i(m)}\boldsymbol{\beta}^*_{(m)}-\boldsymbol{x}'_{i(m)}\widehat{\boldsymbol{\beta}}_{(m)}-e_i\right|^4\\&\le CE_m\left|e^*_{i(m)}\right|^4+CE_m\left|\boldsymbol{x}'_{i(m)}\boldsymbol{\beta}^*_{(m)}-\boldsymbol{x}'_{i(m)}\widehat{\boldsymbol{\beta}}_{(m)}\right|^4+CE_m\left|e_i\right|^4\\&\le C\left\|\boldsymbol{x}_{i(m)}\right\|^4E_m\left\|\boldsymbol{\beta}^*_{(m)}-\widehat{\boldsymbol{\beta}}_{(m)}\right\|^4+CE_m\left|e_i\right|^4+CE_m\left|e^*_{i(m)}\right|^4\\&\le Cn^{-2}k_m^4+C,\end{aligned}$$

因此, 由 c_r 不等式得到

$$\begin{aligned}E\left\|\widetilde{\boldsymbol{\mu}}-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right\|^2&=\sum_{i\in D_N}E\left(\mu_i-\boldsymbol{x}'_{i(m)}\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})\right)^2\\&\le CM\sum_{i\in D_N}\sum_{m=1}^{M}\left\{\omega_m^2E\left(\widehat{T}_{i(m)}^2\right)\right\}\\&\le Cf^{-1}Mk_m^2+NMC,\end{aligned}$$

再结合 (A.28) 式和条件 6 有

$$\xi_N^{-1}E\left|\Big(\widetilde{\boldsymbol{\mu}}-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\Big)'\Big(\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\Big)\right|\le\xi_N^{-1}\sqrt{E\left\|\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\omega})\right\|^2E\left\|\widetilde{\boldsymbol{\mu}}-\widehat{\boldsymbol{\mu}}_f(\boldsymbol{\omega})\right\|^2}=o(1),$$

所以 (A.27) 式成立.