arXiv:2603.07055v1 [stat.ME] 7 Mar 2026

随机化实验中的异质信息整合：统一校准框架*

Wei Ma† Zeqi Wu‡ Zheng Zhang§中国人民大学统计与大数据研究院

摘要

在现代随机化实验中，大规模数据收集日益产生丰富的基线协变量和来自多个来源的辅助信息。这类信息为更精确地估计治疗效应提供了机会，但也带来了如何在不损害有效性的前提下，以协调一致的方式整合异质信息这一挑战。协变量自适应随机化（CAR）广泛用于在设计阶段改善协变量平衡，但通常只平衡用于构建层的少量协变量，因此，在分析阶段进行协变量调整对于更高效地估计治疗效应至关重要。除了标准的协变量调整之外，人们往往还希望纳入辅助信息，包括跨层信息、各种机器学习模型的预测，以及来自历史试验或真实世界数据来源的外部数据。尽管这类辅助信息十分丰富，现有针对CAR的协变量调整方法主要利用层内协变量，无法提供一致的整合机制。我们提出一种统一校准框架，通过信息代理向量以及由凸优化问题定义的校准权重来整合这类信息。由此得到的估计量将许多近期的协变量调整程序涵盖为特例，同时在单一框架下为内部和外部信息借用提供系统化机制。我们确立了大样本有效性和无损效率保证，表明纳入额外信息来源不会增加渐近方差；并将理论扩展至层数和信息来源数量均随样本量增长的情形。模拟研究以及对乌干达和马拉维储蓄行为现场实验的实证分析，展示了该方法出色的有限样本表现和实用价值。

关键词：校准权重；协变量自适应随机化；协变量调整；信息整合；机器学习

*作者按姓氏字母顺序排列。模拟和实证应用的完整源代码托管在 GitHub，网址为 https://github.com/zeqiwu1202/Replication-for-CAR-calibration。

†电子邮箱：mawei@ruc.edu.cn。
‡通讯作者。电子邮箱：wuzeqi@ruc.edu.cn。§电子邮箱：zhengzhang@ruc.edu.cn。

1 引言

在现代随机化实验中，确保处理组之间的基线协变量平衡，对于减少偏倚、提高试验结果的可信度至关重要。协变量自适应随机化（CAR）方法，例如分层偏倚硬币随机化（Efron1971）、分层区组随机化（Zelen，1974年）和最小化法（Pocock & Simon1975，Taves 1974年），广泛用于在设计阶段实现这种平衡。然而在实践中，CAR通常只使用少量协变量构建层，因此并不会直接确保许多其他基线协变量的平衡。此外，一些重要的治疗前协变量可能只有在随机化后才能观测到，例如与结局变量一同收集时（Bai et al. 2024）。因此，在统计分析阶段进行协变量调整发挥着至关重要的补充作用。将基线协变量纳入统计分析后，协变量调整方法可以纠正残余不平衡，从而提高治疗效应估计的精确度和效率，并增强试验的有效性。

统计分析阶段对基线协变量的调整已有长期研究（参见，例如，Lin 2013年、Tsiatis 等人 2008年、Zhang 等人 2008年及其中的参考文献）。在 CAR 设计下，Gu 等人（2023）、Ma 等人（2022）、Ye 等人（2022）使用线性回归对额外协变量进行调整，而 Liu 等人（2023）应用了 Lasso（Tibshirani，1996年）。这些方法得到的平均治疗效应（ATE）估计量比朴素均值差估计量以及 Bugni 等人（2018）和 Bugni 等人（2019）提出的普通最小二乘估计量（将结局对层指标进行回归）更为有效。即使缺乏协变量与结局之间存在线性关系的有力证据，这一结论仍然成立。为进一步提高效率，Rafi（2023）、Tu 等人（2024）和 Bannick 等人（2025年）的近期工作提出了非线性协变量调整的统一框架。这些框架依托增广逆概率加权（增广逆概率加权，Robins 等人 1994 年），纳入随机森林和深度神经网络等机器学习技术来调整协变量。这些机器

学习方法能够有效捕捉协变量与结局之间复杂的非线性关系。

然而，现有协变量调整文献中的很大一部分可被理解为聚焦于内部信息借用的一种特定形式，即利用当前试验的基线协变量（通常在每一层内）来提高效率。这种聚焦为其他具有实际重要性的信息借用形式留下的空间相对有限，并限制了标准调整方法以系统方式整合异质信息的程度。在内部，当结局-协变量关系稳定时，通过跨层借用信息，以及在没有任何单一学习器始终可靠时组合多种机器学习方法的预测，往往可以提高效率。在外部，人们越来越有兴趣利用历史试验和真实世界数据来支持同期试验的分析，尤其是在样本量受到成本、伦理考量或招募挑战制约时（美国食品药品监督管理局 2019, Gu等人（2024））。现有基于增广逆概率加权的非线性调整框架（Bannick 等人, 2025, Rafi（2023）, Tu 等人，2024）并非旨在整合这些异质信息来源，因为它们通常依赖单一的干扰参数估计，并且缺乏将多个内部预测变量或外部信息源加以组合的系统机制。为填补这一空白，我们提出一个用于整合异质信息的统一校准框架，该框架同时容纳内部借用与外部借用，从而使估计量能够联合利用协变量、跨层信息和辅助数据源。

我们方法的一个核心方法学特征，是在受限随机化设计下使用校准权重。尽管校准已在调查抽样（Deville & Särndal1992、Kwon et al. 2025年）、缺失数据问题（Qin & Zhang 2007年、Tan 2014）以及观察性研究（Chan et al. 2016）中得到研究，但其在受限随机化设计下的分析引发了独特的理论问题。不同于上述样本通常独立同分布（i.i.d.）的情形，受限随机化设计会在层内的处理分配之间诱导出复杂的依赖结构。我们通过条件渐近

论证来处理这种依赖。具体而言，我们以实现的层指标和处理分配为条件，将其视为固定量，并利用条件大数定律和条件中心极限定理建立大样本结果。这一方法为受限随机化下的推断提供了一个易于处理的框架，并产生了可推广至层数不断增加以及信息来源数量不断增加情形的证明技巧，而这些技巧本身也可能具有独立的研究价值。我们将主要贡献概括如下。

- **统一校准框架**。我们提出一种基于校准的框架，用于受限随机化下的估计与推断。该框架在三个方面实现了统一。首先，它提供了一种通用表述，可将许多近期的协变量调整程序纳为特例（例如，Bannick 等人，2025、Bugni 等人，2019、Cohen 和 Fogarty，2024 、 Liu 等人，2023 、 Ma 等人，2022 、 Tu 等人，2024 、 Ye 等人，2023、2022）。其次，它将内部和外部信息借用纳入同一架构，从而为整合异质信息来源提供系统化方法。第三，该框架广泛适用于满足假设2的受限随机化方案，因此所得推断程序不依赖于特定的随机化方法。

- **灵活且稳健的信息借用**。我们提出了信息代理向量的实用构造方法，可适用于多种内部和外部信息来源。在内部，该框架能够跨层借用信息，并汇总异质机器学习预测。在外部，该框架可整合来自历史试验和真实世界数据的信息。重要的是，我们的框架对信息来源具有模型无关性。我们将这一特性称为稳健性，意即即使所用信息存在偏倚或由不准确的模型生成，我们的统计推断仍然有效。

- **受限随机化（CAR）下的一般推断理论**。 我们提供了严谨的理论基础，证明我们的估计量渐近正态，且其方差可一致估计。我们建立了无损 效率保证，确保利用额外的

信息来源能够提高估计效率，或至少使其保持不变。此外，我们开发了针对受限随机化（CAR）所致依赖关系的证明技巧，这些技巧不同于现有论证（例如，Bugni 等人，2018， 2019， Liu 等人，2023， Ma 等人，2022）。这些技巧适用于层数不断增加、信息来源数量日益增多的情形，也为相关问题提供了可能具有独立价值的工具。

本文组织如下。图1 提供了概览。第2 节介绍研究设定和我们的统一校准框架。第3 节讨论构建信息代理向量 $\boldsymbol{\xi}_n$的实用策略，包括从内部和外部来源借用辅助信息的方法。第4 节建立所提出的估计量的大样本性质，包括渐近正态性、一致方差估计和效率比较。第5 节介绍理论扩展，包括层数发散和维度不断增长的设定 $\boldsymbol{\xi}_n$，以及一般差异度量。最后，第6 节和第7节通过模拟评估有限样本表现，并使用来自Dupas 等人（2018）的实验数据演示该方法。所有证明均见附录（补充材料，Ma 等（2026））。

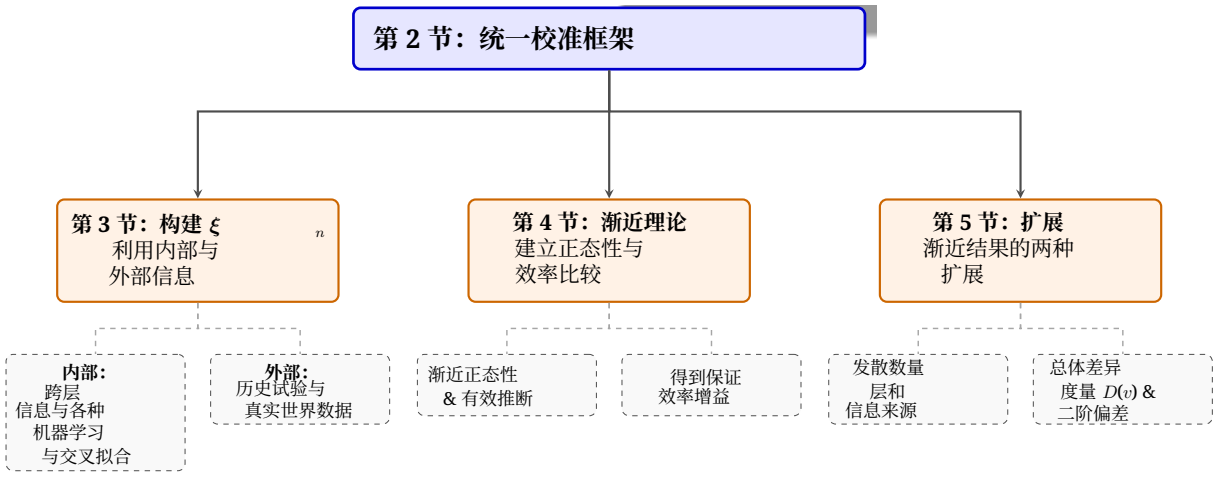


图1：所提出的统一校准框架及理论分析路线图。

符号约定。 本文始终采用以下符号约定。对于任意列向量 $\boldsymbol{x} = (x_1, x_2, \ldots\ldots, x_d)^\top \in \mathbb{R}^d$，其中 $\mathbb{R}^d$是 d维欧几里得空间，$\|\boldsymbol{x}\| = \left(\boldsymbol{x}^\top \boldsymbol{x}\right)^{1/2}$ 表示其欧几里得范数。对于任意矩阵 $A = (a_{ij})_{n\times m}$， $\|A\|$ 表示其

最大奇异值，即算子范数；$\|A\|_F = \sqrt{\text{tr}(AA^\top)}$ 表示其弗罗贝尼乌斯范数，而 A^+ 表示其摩尔-彭若斯伪逆。对于两个正的非随机序列a_n，b_n 以及随机向量序列 X_n，$X_n = o_P(a_n)$表示 $P(\|X_n\| > a_n\epsilon) \to 0$ 对任意 $\epsilon > 0$ 和 $X_n = O_P(a_n)$成立，即对任意 $\epsilon > 0$，存在常数 $M > 0$ 使得 $\limsup_{n\to\infty} P(\|X_n\| \geq a_n M) < \epsilon$。符号 $\mathbf{1}(\cdot)$ 表示指示函数：括号内的条件成立时取值为 1，否则取值为 0。

2 统一校准框架

2.1 预备知识

令 $A_i \in \{0, 1\}$ $(i = 1, \ldots, n)$ 表示处理分配指示变量，其中 $A_i = 1$表示第 i 个单位被分配到处理组，$A_i = 0$ 则表示并非如此。我们假设分配 $\{A_i\}_{i=1}^n$ 由满足假设2的受限随机化设计生成。因此，变量 A_i 通常并非独立同分布。采用潜在结果框架（Imbens & Rubin 2015），我们将 $Y_i(1)$ 和 $Y_i(0)$ 分别定义为处理和对照条件下的潜在结果。观测结果 Y_i 由$Y_i = A_iY_i(1) + (1 - A_i)Y_i(0)$ 决定。

该实验设计将单位分为 K 层，其中 $B_i \in \{1, \ldots, K\}$ 表示单位 i 所属的层。为方便记号，令 $[k] = \{i : B_i = k\}$ 表示属于层 k 的单位索引集合。为确保各层均非空，我们假设分配概率为正：对于所有 $k \in \{1, \ldots, K\}$ 和 $i \in \{1, \ldots, n\}$，均有 $p_{[k]} = P(B_i = k) > 0$。第 k 层的目标处理分配比例为 $\pi_{[k]} = P(A_i = 1 \mid B_i = k) \in (0, 1)$。每个单位都有一个 p 维基线协变量向量 $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^\top \in \mathbf{X} \subset \mathbb{R}^p$，其维度可以是低维或高维。我们假设协变量向量 $\boldsymbol{X}_i$包含层指示变量 B_i；但为突出层指示变量 B_i，有时使用记号 $(\boldsymbol{X}_i, B_i)$。

下标 1 和 0 分别表示处理组和对照组。处理

组包含 $n_1 = \sum_{i=1}^{n} A_i$ 个单位，对照组包含 $n_0 = \sum_{i=1}^{n} (1 - A_i)$ 个单位。对于层特定的量，我们使用下标 $[k]$: 令 $n_{[k]} = \sum_{i \in [k]} 1$ 表示层大小，其中 $n_{1[k]} = \sum_{i \in [k]} A_i$ 和 $n_{0[k]} = \sum_{i \in [k]} (1 - A_i)$ 分别表示第 k 层中的处理组单位和对照组单位。层比例和处理分配比例分别定义为 $p_{n[k]} = n_{[k]}/n$ 和 $\pi_{n[k]} = n_{1[k]}/n_{[k]}$。

我们关注的参数是总体平均处理效应（ATE）：

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0)].$$

在受限随机化（CAR）下，ATE 参数 τ 可以通过汇总各层的治疗效应估计而得到一致估计。对此，一个简单的估计量是分层均值差估计量（Bugni 等人，2019，Ma 等人，2022）：

$$\widehat{\tau}_{\text{sdim}} := \sum_{k=1}^{K} p_{n[k]} \left(\overline{Y}_{1[k]} - \overline{Y}_{0[k]} \right),$$

其中，对于 $a \in \{0, 1\}$ 和 $k = 1, \ldots, K$，有 $\overline{Y}_{a[k]} := \frac{1}{n_{a[k]}} \sum_{i \in [k]} \mathbb{1}(A_i = a) Y_i$。为提高估计效率，近期文献提出了非线性协变量调整方法（Bannick 等人，2025；Rafi（2023）；Tu 等人，2024）。这些方法通常采用增广逆概率加权（AIPW，Robins 等人，1994）估计量，并依赖于对条件均值函数 $h^*_{a[k]}(\boldsymbol{X}_i) = \mathbb{E}[Y_i(a) \mid \boldsymbol{X}_i, B_i = k]$ 的估计。然而，基于 AIPW 的调整方法有其局限，无法结合 $h^*_{a[k]}(\boldsymbol{X}_i)$ 的不同估计值。例如，这些估计值可能利用跨层借用和异质机器学习预测来使用内部信息，也可能纳入来自历史试验、真实世界数据以及领域专家建议的函数形式的外部信息（详细讨论见第3节）。下文将介绍一种统一调整框架，用于结合 $h^*_{a[k]}(\boldsymbol{X}_i)$ 的这些不同估计值，从而得到更高效、更稳健的平均处理效应估计量。

2.2 校准估计量的构造

设 $D(v): \mathbb{R} \to \mathbb{R}$ 是一个二次连续可微且严格凸的函数，用于衡量与 1 的差异，例如 $D(v) = (v-1)^2/2$ 和 $D(v) = v - \log v$。假设 $\boldsymbol{\xi}_n : \mathcal{X} \to \mathbb{R}^d$ $(n \geq 1)$ 是一列取值于 $\mathbb{R}^d$ 的（可能的）$\boldsymbol{X}_i$ 随机函数，我们称之为信息代理向量。通常，$\boldsymbol{\xi}_n$ 的各个元素可能是条件均值函数 $h^*_{a[k]}(\boldsymbol{X}_i)$ 的不同估计值。基于 $\boldsymbol{\xi}_n$，我们提出如下校准估计量:

$$\widehat{\tau}_{\text{cal}} := \widehat{\tau}_{\text{sdim}} + \frac{1}{n}\sum_{i=1}^{n} \widehat{w}_i r_i, \tag{2.1}$$

其中 $r_i := \sum_{k=1}^{K} \left\{ \frac{A_i}{\pi_{n[k]}}(Y_i - Y_{1[k]}) - \frac{1-A_i}{1-\pi_{n[k]}}(Y_i - Y_{0[k]}) \right\} \mathbb{1}(B_i = k)$，校准权重$\widehat{w}_i$'s $(i = 1, \ldots\ldots, n)$ 满足以下校准问题:

$$\begin{cases} (\widehat{w}_1, \ldots, \widehat{w}_n) = \min_{w_i, 1 \leq i \leq n} \sum_{i=1}^{n} D(w_i) \text{ subject to} \\ \frac{1}{n}\sum_{i=1}^{n} w_i \left\{ A_i - \pi_{n[k]} \right\} \mathbb{1}(B_i = k) \left\{ \boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]} \right\} = 0, \quad \forall k = 1, \ldots, K. \end{cases} \tag{2.2}$$

在（2.2）中，$\overline{\boldsymbol{\xi}}_{n[k]} := \frac{1}{n_{[k]}} \sum_{i \in [k]} \boldsymbol{\xi}_n(\boldsymbol{X}_i)$被定义为$\boldsymbol{\xi}_n(\boldsymbol{X}_i)$的分层特定样本均值。校准问题（2.2）是一个涉及 dK 线性约束的凸优化问题。该问题可通过其对偶形式，使用标准凸优化软件高效求解（参见Boyd & Vandenberghe 2004年）。校准估计量（2.1）由两部分组成：（i）分层均值差估计量 $\widehat{\tau}_{\text{sdim}}$，以及（ii）由加权残差构造的校正项，其中权重由校准问题（2.2）确定。与线性回归分析类似，残差$r_i = \sum_{k=1}^{K} \left\{ \frac{A_i}{\pi_{n[k]}}(Y_i - Y_{1[k]}) - \frac{1-A_i}{1-\pi_{n[k]}}(Y_i - Y_{0[k]}) \right\} \mathbb{1}(B_i = k)$，$i = 1, \ldots\ldots, n$表示 $\sum_{k=1}^{K} \left\{ \frac{A_i}{\pi_{n[k]}} Y_i - \frac{1-A_i}{1-\pi_{n[k]}} Y_i \right\} \mathbb{1}(B_i = k)$ 中未被层均值解释的部分。从几何上看，残差 r_i 来自将（$\sum_{k=1}^{K} \left\{ \frac{A_i}{\pi_{n[k]}} Y_i - \frac{1-A_i}{1-\pi_{n[k]}} Y_i \right\} \mathbb{1}(B_i = k)$：$1 \leq i \leq n$）投影到由（1, ..., 1）张成的空间的正交补空间上，这本身意味着 $\sum_{i=1}^{n} r_i = 0$。当观测样本按以下方式加权时，校准问题（2.2）强制 $\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]}$在层内不同处理组之间保持平衡

$\widehat{w}_i$， $i = 1, \ldots, n$，从而自然地将 $\boldsymbol{\xi}_n$ 中包含的信息纳入校准权重 $\widehat{w}_i$。这种信息纳入使校准权重$\widehat{w}_i$能够进一步解释 $\sum_{k=1}^{K}\left\{\frac{A_i}{\pi_{n[k]}} Y_i - \frac{1-A_i}{1-\pi_{n[k]}} Y_i\right\}\mathbb{1}(B_i = k)$中的变异。因此，残差的加权平均 $\frac{1}{n}\sum_{i=1}^{n} \widehat{w}_i r_i$作为校正项，表示由校准权重所解释的部分。如果对所有 $i = 1, \ldots, n$ 都有 $\widehat{w}_i = 1$，例如当 $\boldsymbol{\xi}_n(\boldsymbol{X}) - \boldsymbol{\xi}_{n[k]}$ 已由随机化程序实现平衡时，则 $\widehat{\tau}_{\text{cal}}$ 退化为 $\widehat{\tau}_{\text{sdim}}$。

我们的校准估计量（2.1）提供了一个统一的信息整合框架。通过适当选择信息代理向量 $\boldsymbol{\xi}_n$，它可以还原许多现有的协变量调整方法，这些方法可视为内部信息借用的特例（例如，Bannick 等人，2025、Bugni 等人，2019、Cohen 和 Fogarty，2024、Gu等人（2024）、Liu 等人，2023、Ma 等人，2022、Tu 等人，2024、Ye 等人，2023以及2022）。例如，当 $\boldsymbol{\xi}_n$ 被指定为基于Lasso的估计时，（2.1）与Liu 等人，(2023）提出的Lasso调整估计量类似。除了内部调整之外，（2.1）也支持使用外部信息代理，与Gu 等人(2024）等外部借用方法相似，因此能够构造同时利用内部和外部信息的新估计量。

我们的校准估计量 (2.1) 的一个关键性质是，它在对 $\boldsymbol{\xi}_n(\boldsymbol{X})$ 进行仿射变换后保持不变。具体而言，在优化问题 (2.2) 中用 $\mathrm{Q}\boldsymbol{\xi}_n(\boldsymbol{X}) + \boldsymbol{q}$ 替换 $\boldsymbol{\xi}_n(\boldsymbol{X})$，得到的校准估计量 $\widehat{\tau}_{\text{cal}}$ 不变，其中 $\mathrm{Q} \in \mathbb{R}^{d\times d}$ 是任意可逆矩阵，$\boldsymbol{q} \in \mathbb{R}^d$ 是任意 d 维向量。如果 $\boldsymbol{\xi}_n(\boldsymbol{X})$ 用作条件均值函数 $h^*_{a[k]}(\boldsymbol{X}_i)$ 的估计，这一性质允许其误设程度达到仿射变换，为标准的基于 AIPW 的协变量调整方法所不具备的灵活性。

3 构造信息代理 $\boldsymbol{\xi}_n$ 的策略

所提出的统一框架的有效性和灵活性取决于 $\boldsymbol{\xi}_n$ 的构建，它可作为辅助信息的代理变量。因此，我们将相关策略分为

用于确定 $\boldsymbol{\xi}_n$ 的策略，分为内部信息借用和外部信息借用。第 3.1–3.3 节重点介绍内部策略，包括跨层信息借用、整合不同的机器学习预测以及交叉拟合技术；第 3.4节讨论如何纳入历史试验和真实世界数据等外部信息。

3.1 跨层信息借用

假设 $h_{a[k]}(\cdot)$ 是 $h^*_{a[k]}(\cdot)$ 的估计量，其中 $a \in \{0,1\}$ 且 $1 \le k \le K$。根据定理1（将在第4节中证明），如果 $h_{a[k]}(\cdot)$ 一致估计 $h^*_{a[k]}(\cdot)$，那么在 (2.1) 中取 $\boldsymbol{\xi}_n(\boldsymbol{X}) = \left(\sum_{k=1}^K \widehat{h}_{1[k]}(\boldsymbol{X})1(B=k)\ \sum_{k=1}^K \widehat{h}_{0[k]}(\boldsymbol{X})1(B=k)\right)^\top$，将得到 τ 的半参数有效估计量。不过，需要注意的是，这种方法确保每一层仅利用该层内的信息进行协变量调整。然而，跨层借用信息通常有利，尤其是在协变量 $\boldsymbol{X}$ 与潜在结果（$Y(1)$ $Y(0)$）之间的关系在不同层之间保持稳定时。我们的框架通过在 (2.1) 中取 $\boldsymbol{\xi}_n(\boldsymbol{X}) = \left(\widehat{h}_{1[k]}(\boldsymbol{X})\ \widehat{h}_{0[k]}(\boldsymbol{X}) : 1 \le k \le K\right)^\top$，自然地实现了这一点。这种设定使每一层都能利用所有层的信息，从而得到更有效的校准估计量。

3.2 整合 异质机器学习预测的整合

在实践中，可以使用多种机器学习方法来估计条件均值函数 $h^*_{a[k]}(\boldsymbol{X})$，例如深度神经网络（Farrell 等人，2021，Jiao 等人，2023，LeCun等，2015年）、随机森林（Breiman，2001年，Wager 和 Athey，2018）、Lasso（Tibshirani，1996年）等。我们的框架提供了一种自然的方法来组合这些机器学习估计。假设 $\widehat{h}^{\mathrm{rf}}_{a[k]}(\boldsymbol{X})$和 $\widehat{h}^{\mathrm{nn}}_{a[k]}(\boldsymbol{X})$分别是基于随机森林和深度神经网络对$h^*_{a[k]}(\boldsymbol{X})$的估计。然后，我们可以令 $\boldsymbol{\xi}_n(\boldsymbol{X}) = \left(\widehat{h}^{\mathrm{rf}}_{a[k]}(\boldsymbol{X})\ \widehat{h}^{\mathrm{nn}}_{a[k]}(\boldsymbol{X}) :\ a \in \{0,1\}\ 1 \le k \le K\right)^\top$ 在(2.1)中，从而推导出 τ的校准估计量。根据定理2（将在第4节中建立），该估计量

当单一机器学习方法未能充分刻画条件均值函数 $h^*_{a[k]}(\boldsymbol{X})$时，其效率将高于仅基于随机森林或神经网络的估计量。这一改进提高了协变量调整方法的效率。

3.3 通过交叉拟合和样本分割进行实现

使用机器学习方法估计 $\boldsymbol{\xi}_n$ 时，通常希望依赖独立样本，以缓解过拟合，并使假设3 更为合理。通过样本分割可以实现这种独立性。然而，样本分割通常会降低效率。为恢复效率，我们采用交叉拟合技术（参见Bannick 等人，2025、Chernozhukov 等人，2018、Rafi（2023）、Tu 等人，2024）。具体而言，我们遵循Rafi （2023，第4节）所述的样本分割程序，确保每一折都包含来自每个层和治疗组的数据。假设单位$\{1,\ldots,n\}$被划分为两个大小相等的折， I_0 和 I_1。交叉拟合校准估计量定义为 $\widehat{\tau}_{\text{CFcal}} := \frac{1}{2}\widehat{\tau}_{(0)\,\text{cal}} + \frac{1}{2}\widehat{\tau}^{(1)}_{\text{cal}}$，其中

$$\widehat{\tau}^{(\iota)}_{\text{cal}} = \sum_{k=1}^{K} p^{(\iota)}_{n[k]} \left(\overline{Y}^{(\iota)}_{1[k]} - \overline{Y}^{(\iota)}_{0[k]} \right) + \frac{1}{|I_\iota|} \sum_{i \in I_\iota} \widehat{w}_i \sum_{k=1}^{K} \left\{ \frac{A_i}{\pi^{(\iota)}_{n[k]}} \left(Y_i - \overline{Y}^{(\iota)}_{1[k]} \right) - \frac{1 - A_i}{1 - \pi^{(\iota)}_{n[k]}} \left(Y_i - \overline{Y}^{(\iota)}_{0[k]} \right) \right\} \mathbb{1}(B_i = k)$$

其中 $|I_\iota| = \sum_{i \in I_\iota} 1$, $p^{(\iota)}_{n[k]} = \sum_{i \in I_\iota} \mathbb{1}(B_i = k)/|I_\iota|$, $Y_{(\iota)\,a[k]} = \sum_{i \in I_\iota} \mathbb{1}(A_i = a, B_i = k) Y_i / \sum_{i \in I_\iota} \mathbb{1}(A_i = a, B_i = k)$, $\pi^{(\iota)}_{n[k]} = \sum_{i \in I_\iota} \mathbb{1}(A_i = a, B_i = k) / \sum_{i \in I_\iota} \mathbb{1}(B_i = k)$ 对于$\iota \in \{0,1\}$。权重 $\widehat{w}_i$'s $(i \in I_\iota)$ 通过求解以下校准问题获得：

$$\begin{cases} (\widehat{w}_i)_{i \in I_\iota} = \min_{w_i : i \in I_\iota} \sum_{i \in I_\iota} D(w_i) \text{ subject to} \\ \sum_{i \in I_\iota} w_i \left\{ A_i - \pi^{(\iota)}_{n[k]} \right\} \mathbb{1}(B_i = k) \left\{ \boldsymbol{\xi}^{(1-\iota)}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}^{(1-\iota)}_{n[k]} \right\} = 0, \quad \forall k = 1, \ldots, K, \end{cases}$$

其中，$\boldsymbol{\xi}_n^{(1-\iota)}$ 关于 $\{(Y_i, \boldsymbol{X}_i, A_i) : i \in I_\iota\}$ 可测，且 $\boldsymbol{\xi}_{n[k]}^{(1-\iota)} := \sum_{i \in I_\iota} \boldsymbol{\xi}_n(\boldsymbol{X}_i) \mathbb{1}(B_i = k) / \sum_{i \in I_\iota} \mathbb{1}(B_i = k)$ 表示 $\boldsymbol{\xi}_n^{(1-\iota)}(\boldsymbol{X})$ 的分层特定样本均值。结合定理1的证明与 Rafi（2023，定理 4.1）证明中的论证，可以建立 $\hat{\tau}_{\text{cal}}^{\text{CF}}$ 的渐近性质；此处省略细节。我们将在模拟研究中评估这一交叉拟合估计量 $\hat{\tau}_{\text{cal}}^{\text{CF}}$ 的有限样本表现。

3.4 利用历史数据与真实世界数据

除内部策略之外，利用外部信息构建 $\boldsymbol{\xi}_n$ 为提升效率提供了一种强有力的途径。在本小节中，我们聚焦于两个主要来源：历史临床试验和真实世界数据。尽管这些来源提供了有价值的信息，但它们在分布上往往与当前试验不同。本框架的一个关键特点在于，它能够利用此类异质外部数据，而无需施加限制性的相似性假设。

首先，历史临床试验通常可提供与当前研究背景相似、且易于获取的数据。由于各试验之间的治疗方案往往不同，现有文献主要侧重于利用历史对照数据来提高当前试验中估计的精确度（Callegaro et al. 2023，Pocock & Simon 1975）。这些方法通常依赖于历史对照组与当前对照组具有可比性这一假设。相比之下，我们统一框架的一个显著优势在于其弱假设性质：我们对所借用信息的有效性或直接可迁移性不作任何假设。因此，我们并不局限于仅从历史对照组中借用信息；我们可以灵活地纳入来自历史研究的信息，即使这些研究涉及的治疗可能与当前试验中的治疗不同。

其次，除了历史临床试验，真实世界数据也是一个庞大的信息来源。真实世界数据包括医疗服务常规提供过程中及患者监测时生成的数据，例如电子健康记录（EHRs）、疾病登记系统，

以及越来越多来自可穿戴设备的数据。与受到严格控制的临床试验不同，基于真实世界数据的观察性研究通常涉及规模大得多、更多元化的患者群体，这可以显著增强分析的统计功效。

至关重要的是，历史试验数据和真实世界数据都可能存在“协变量偏移”，即这些外部来源中的协变量 $\boldsymbol{X}$ 分布通常不同于当前试验中的分布。然而，一种常见现象是，尽管$\boldsymbol{X}$ 的边缘分布发生变化，给定协变量 $\boldsymbol{X}$ 时潜在结果（$Y(1)$ $Y(0)$）的条件分布仍具有可比性。这种稳定性十分重要，因为最优的 $\boldsymbol{\xi}_n(\boldsymbol{X})$仅取决于给定 $\boldsymbol{X}$ 时（$Y(1)$ $Y(0)$）的条件分布（见定理1），并能确保由此得到的估计量 $\hat{\tau}_{\text{cal}}$ 达到半参数效率。在实践中，我们可以利用这一点，从这些外部数据集中估计条件均值函数 $h^*_{1[k]}(\boldsymbol{X})$和 $h^*_{0[k]}(\boldsymbol{X})$（$1 \le k \le K$），并将 $\boldsymbol{\xi}_n(\boldsymbol{X})$设为这些估计值组成的向量。定理2保证，将这些外部数据集中的估计值纳入后，得到的校准估计量渐近方差不会增大。因此，我们的框架提供了一种稳健的策略，能够提高估计效率，同时避免“负迁移”的风险。与依赖贝叶斯框架（Hobbs等（2011），Ibrahim等（2015））或Trans-Lasso（Gu等人（2024））的现有方法不同——这些方法中信息迁移取决于先验分布和/或模型假设——我们的方法完全不依赖模型。值得注意的是，该方法不对外部数据与目标数据生成过程之间的相似性施加任何限制，因此更为灵活，也能更稳健地应对真实世界数据的复杂性。

4 渐近性质

我们作出以下假设。

假设1。 $\boldsymbol{W}_i = (Y_i(1)\ Y_i(0)\ \boldsymbol{X}_i^\top, B_i)^\top, i = 1, \ldots, n$ 是来自总体分布的独立同分布样本，其中总体分布为 $\boldsymbol{W} = (Y(1)\ Y(0)\ \boldsymbol{X}^\top, B)^\top$，记作 $\boldsymbol{W}^{(n)} = \{\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n\}$。

此外，对于 $a=0$，$\sup_{1\le k\le K}\left[|Y_i(a)|^{2+\epsilon}\mid B_i=k\right]<\infty$ 均成立，其中 $0<\epsilon\le 1$ 为常数。

假设2 (处理分配)。 令 $A^{(n)}=\{A_1,\ldots,A_n\}$ 和 $B^{(n)}=\{B_1,\ldots,B_n\}$。处理分配机制满足以下条件：

(1) 在层指标 $B^{(n)}$给定的条件下，处理分配 $A^{(n)}$ 与 $\boldsymbol{W}^{(n)}$相互独立，即 $\boldsymbol{W}^{(n)}\perp A^{(n)}\mid B^{(n)}$;

(2) 当 $n\to\infty$ 趋于无穷时，$\sup_{1\le k\le K}\left|\pi_{n[k]}-\pi_{[k]}\right|\xrightarrow{p}0$，其中 $\pi_{[k]}$、 $k=1$、..、 K 满足

$$0<\inf_{n\ge 1}\inf_{1\le k\le K}\pi_{[k]}\le\sup_{n\ge 1}\sup_{1\le k\le K}\pi_{[k]}<1.$$

此外，$\lim_{n\to\infty}P(\inf_{1\le k\le K}n_{[k]}\ge d+2)=1$，其中 d 表示 $\boldsymbol{\xi}_n$的维数。**假设3**。 d 和 K 为固定数。序列$\boldsymbol{\xi}_n$ （可能是随机函数）满足以下条件：

(1) 存在一个非随机函数序列 $\boldsymbol{\xi}_n^*:\boldsymbol{X}\to\mathbb{R}^d$ *($n\ge 1$)*，使得

(a) 上确界$_{n\ge 1}\mathbb{E}\left[\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\|^{2+\epsilon}\right]<\infty$，其中 ϵ 在假设 *1*中定义；

(b) $\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top\mid B_i=k\right]$ 的最小非零奇异值大于某个常数 $c>0$，且对 $n\ge 1$、 $k=1$、.....、 K均匀成立；

(c) 对每个 $k=1$、.....、 K，

$$\frac{1}{n_{1[k]}}\sum_{i\in[k]}A_i\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}-\frac{1}{n_{0[k]}}\sum_{i\in[k]}(1-A_i)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}=o_P(n^{-1/2})$$

$and\ \frac{1}{n_{[k]}}\sum_{i\in[k]}\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^2=o_P(1)$;

$\liminf_{n\to\infty}(\varsigma^2_{\widetilde{Y}}-\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*})>0$，其中 $\varsigma^2_{\widetilde{Y}}$和 $\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}$ 在附录 *B*中定义；

(2) 对每一个 $k=1$，.....， K，

$$\left\|\left\{\frac{1}{n}\sum_{i\in[k]}(A_i-\pi_{n[k]})^2\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}\right\}\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}\right\}^\top\right\}^+\right\|=O_P(1).$$

在本节中，我们关注层数 K 固定的情形，并将 K 发散的情形留至第 5.1 节。假设 1 和 2 在受限随机化下统计推断的文献中是标准的（例如，见 Bannick 等人，2025，Bugni 等人，2018，2019，Jiang 等人，2023，Ma 等人，2022，Tu 等人，2024 及其中的参考文献）。假设2 对许多随机化方法成立，例如分层区组随机化（Zelen，1974）以及 Pocock 和 Simon 的最小化法（Pocock & Simon 1975）。假设 3(1)(a) 是关于 $\boldsymbol{\xi}_n^*$（即 $\boldsymbol{\xi}_n$ 的概率极限）的一个温和矩假设。假设 3(1)(b) 要求 $\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right]$ 的最小非零奇异值有界且远离零。重要的是，该假设允许矩阵 $\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right]$ 为奇异的。这是合理的，因为在许多应用中，$\boldsymbol{\xi}_n$ 的各分量可能共享共同信息，从而导致矩阵 $\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right]$ 确实奇异的情形。假设 3(1)(c) 要求 $\boldsymbol{\xi}_n$ 收敛于其概率极限 $\boldsymbol{\xi}_n^*$。该假设类似于 Tu 等人（2024，假设 4）、Bannick 等人（2025，假设2）和 Jiang 等人（2023，假设3）。当 $\boldsymbol{\xi}_n$ 由线性回归得到时，它简化为 Liu 等人（2023，假设3）。当 $\boldsymbol{\xi}_n$ 由局部线性核回归得到时，该假设可由 Tu 等人（2024，定理 3）加以验证。若 $\boldsymbol{\xi}_n$ 是依赖于整个观测数据集的一般机器学习估计量，则验证假设 3(1)(c) 通常要求包含 $\boldsymbol{\xi}_n$ 的函数类满足 Donsker 条件（见 Bannick 等人，2025，假设3）；若 $\boldsymbol{\xi}_n$ 由复杂的机器学习算法得到，该条件可能被违反。为绕过 Donsker 条件，可以采用第 3.3 节所讨论的样本分割和/或交叉拟合技术（Chernozhukov 等人，2018，Rafi（2023），Tu 等人，2024）。通过使用交叉拟合，假设 3(1)(c) 可由均方误差收敛条件替代（见 Rafi（2023），假设 4.1 以及 Tu 等人，2024，假设6），该条件对许多机器学习方法成立（例如，见 Chi 等人，2022，Farrell 等人，2021，Jiao 等人，2023，Wager 和 Athey，2018）。假设 3(1)(d) 确保结果变量 Y 不能被 $\boldsymbol{\xi}(\boldsymbol{X})$ 完全解释。假设 3(2) 是一项技术性要求，用以确保所估计的摩尔–彭若斯广义逆保持良好性态且不会发散。该条件是

有必要处理摩尔–彭若斯广义逆运算的不连续性（例如参见Stewart 1977，定理 3.1），因为仅凭假设1–2 和3(1)(a)–(c) 并不能保证该逆的收敛。然而，在极限矩阵$\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid B_i=k\right]$非奇异的标准情形下，摩尔–彭若斯广义逆与标准矩阵逆一致。由于标准求逆是连续运算，假设 3(2) 便由前述假设自动蕴含。

回顾 $h^*_{a[k]}(\boldsymbol{X}) = \mathbb{E}[Y(a) \mid \boldsymbol{X},\ B=k]$ 并令 $\widetilde{h}^*_{a[k]}(\boldsymbol{X}) := h^*_{a[k]}(\boldsymbol{X}) - \mathbb{E}\left[h^*_{a[k]}(\boldsymbol{X}) \mid B=k\right]$与 $\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}) := \boldsymbol{\xi}_n^*(\boldsymbol{X}) - \mathbb{E}[\boldsymbol{\xi}_n^*(\boldsymbol{X}) \mid B=k]$。我们有如下定理。

定理1。 假设 *1–3* 成立，且 $D(v) = (v-1)^2/2$。则

$$\frac{\sqrt{n}\,(\widehat{\tau}_{\mathrm{cal}} - \tau)}{\sqrt{\varsigma_H^2 + \varsigma_{\widetilde{Y}}^2 - \varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}}} \xrightarrow{d} N(0,1) \ \textit{and}\ \widehat{\varsigma}_H^2 + \widehat{\varsigma}_{\widetilde{Y}}^2 - \widehat{\varsigma}^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*} = \varsigma_H^2 + \varsigma_{\widetilde{Y}}^2 - \varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*} + o_P(1),$$

渐近方差 ς_H^2、 $\varsigma_{\widetilde{Y}}^2$、 $\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}$及其估计量 $\widehat{\varsigma}_H^2$、 $\widehat{\varsigma}_{\widetilde{Y}}^2$和 $\widehat{\varsigma}^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}$ 的定义见附录 *B*。此外，若对于每个 $k=1,\ \ldots\ldots,\ K$，存在一个非随机向量 $\boldsymbol{\alpha}_k \in \mathbb{R}^d$ 使得

$$\left\{\sqrt{\frac{1-\pi_{[k]}}{\pi_{[k]}}}\widetilde{h}^*_{1[k]}(\boldsymbol{X}) + \sqrt{\frac{\pi_{[k]}}{1-\pi_{[k]}}}\widetilde{h}^*_{0[k]}(\boldsymbol{X})\right\}\mathbb{1}(B=k) = \boldsymbol{\alpha}_k^\top \widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X})\mathbb{1}(B=k),$$

then

$$\frac{\sqrt{n}\,(\widehat{\tau}_{\mathrm{cal}} - \tau)}{\sqrt{\varsigma_H^2 + \varsigma_{\widetilde{Y}}^2 - \varsigma^2_{\widetilde{Y}|\widetilde{h}^*}}} \xrightarrow{d} N(0,1),$$

且渐近方差 $\varsigma_H^2 + \varsigma_{\widetilde{Y}}^2 - \varsigma^2_{\widetilde{Y}|\widetilde{h}^*}$ 与 *Rafi (2023，定理 3.1)* 中提出的半参数效率界相匹配。

在定理1中，我们着重讨论 $D(v) = (v-1)^2/2$这一情形；一般的 $D(v)$ 将在第5.2节处理。定理1 表明，我们的校准估计量渐近正态分布，且其渐近方差可以得到一致估计。此外，只要存在 $\boldsymbol{\xi}_n^*(\boldsymbol{X})\mathbb{1}(B=k)$ 的一个线性组合等于 $\left\{\sqrt{\frac{1-\pi_{[k]}}{\pi_{[k]}}}h^*_{1[k]}(\boldsymbol{X}) +\right.$

$\sqrt{\frac{\pi_{[k]}}{1-\pi_{[k]}}}h^*_{0[k]}(\boldsymbol{X})\Big\}\mathbb{1}(B=k)$，我们的校准估计量具有半参数效率。据我们所知，这是文献中实现效率界的一项新条件。现有研究（例如 Tu 等人，2024）要求对 $h^*_{1[k]}(\boldsymbol{X})$ 和 $h^*_{0[k]}(\boldsymbol{X})$ 均进行一致估计。相比之下，我们只要求对这些函数的线性组合进行一致估计，条件更弱。这一条件与 Bai（2022，注2）的论述一致：只要知道函数$\Big\{\sqrt{\frac{1-\pi_{[k]}}{\pi_{[k]}}}h^*_{1[k]}(\boldsymbol{X}) + \sqrt{\frac{\pi_{[k]}}{1-\pi_{[k]}}}h^*_{0[k]}(\boldsymbol{X})\Big\}$，就能实现最优分层。在此语境下，最优分层指使均值差估计量的均方误差最小的分层方法。由于 $\boldsymbol{\xi}_n(\boldsymbol{X})$ 的各个分量可以选为来自不同预先指定模型的 $\Big\{\sqrt{\frac{1-\pi_{[k]}}{\pi_{[k]}}}h^*_{1[k]}(\boldsymbol{X}) + \sqrt{\frac{\pi_{[k]}}{1-\pi_{[k]}}}h^*_{0[k]}(\boldsymbol{X})\Big\}$ 的估计值，校准估计量具有多重效率：只要这些模型中至少有一个设定正确，它就仍具有半参数效率。相比之下，基于 AIPW 的协变量调整方法不具备这种多重效率。

定理1 表明，校准估计量（2.1）具有两项理论优势。首先，假设2 对随机化方案的选择没有任何限制，可以采用简单随机化、分层区组随机化或最小化法。由于校准估计量的渐近分布不随随机化方法而变化，因此它具有文献中所称的“普遍适用性”（Bannick 等人，2025，Ye 等人，2023）。其次，由于 $\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}^*_n} \geq 0$，定理1结合Ma 等人（2022，命题 3）表明，$\widehat{\tau}_{\text{cal}}$ 的效率始终不低于分层均值差估计量 $\widehat{\tau}_{\text{sdim}}$；后者对应于（$\widehat{w}_i = 1$ 对所有 $i = 1, \ldots\ldots, n$）中的特殊情形2.1）。因此，校准估计量能够保证效率增益。更一般地，我们有如下效率比较结果。

定理2。将估计量 $\widehat{\tau}_{\text{cal}}$ 在（*2.1*）中改写为 $\widehat{\tau}_{\text{cal}}(\boldsymbol{\xi}_n)$。设 $\boldsymbol{\Lambda}$ 是一个 $d \times d$ 非随机矩阵，其中 $d \geq 1$ 是固定的。假设 *1–3* 对 $\boldsymbol{\xi}_n$ 和 $\boldsymbol{\Lambda}\boldsymbol{\xi}_n$ 均成立。则

$\widehat{\tau}_{\text{cal}}(\boldsymbol{\xi}_n)$ 和 $\widehat{\tau}_{\text{cal}}(\boldsymbol{\Lambda}\boldsymbol{\xi}_n)$ 满足

$$\frac{\sqrt{n}\left(\widehat{\tau}_{\text{cal}}(\boldsymbol{\xi}_n)-\tau\right)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}} \xrightarrow{d} N(0,1) \ and \ \frac{\sqrt{n}\left(\widehat{\tau}_{\text{cal}}(\boldsymbol{\Lambda}\boldsymbol{\xi}_n)-\tau\right)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\boldsymbol{\Lambda}\widetilde{\boldsymbol{\xi}}_n^*}^2}} \xrightarrow{d} N(0,1),$$

分别。此外，渐近方差满足 $\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2 \leq \varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\boldsymbol{\Lambda}\widetilde{\boldsymbol{\xi}}_n^*}^2$.定理2 表明，在 $\boldsymbol{\xi}_n$ 中纳入更多元素具有无害性质；也就是说，这样做必定能提高或至少维持效率。因此，从理论角度来看，在$\boldsymbol{\xi}_n$中尽可能多地纳入元素似乎是最优的。然而，随着 $\boldsymbol{\xi}_n$ 的维数变大，不再适合将维数 d 视为固定值。我们将在第 $\boldsymbol{\xi}_n$ 节讨论5.1中维数发散的情形。

5 扩展

5.1 $\boldsymbol{\xi}_n(\cdot)$ 的维数和层数均趋于无穷

本节允许 $\boldsymbol{\xi}_n(\cdot)$ 的维数随 n 增大，即当 $n \to \infty$ 时， $d = d_n \to \infty$。此外，我们也允许层数随 n 增加，即当 $n \to \infty$ 时，$K = K_n \to \infty$；这在许多应用中都很常见。我们首先给出一个假设。

假设 4。 随机函数序列 $\boldsymbol{\xi}_n(\cdot)$ 满足以下条件。

（1）存在一列非随机函数 $\boldsymbol{\xi}_n^*(\cdot)$: $\boldsymbol{X} \to \mathbb{R}^d$ $(n \geq 1)$，使得

（a） $K^2 r_n^2/n \to 0$， $K^2\zeta_n^2 r_n/n \to 0$ 以及 $\max\left\{\zeta_n^2 \log(2Kd),\ r_n \log^2(2Kd)\right\}/(n \inf_{1\leq k\leq K} p_{[k]}) \to 0$ 当 $n \to \infty$时，其中 $r_n := \max_{1\leq k\leq K}$秩 $\left\{\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right]\right\}$ 以及$\zeta_n := \sup_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{\xi}_n^*(\boldsymbol{x})\|$;

(b) $\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right]$ 的最大奇异值在 $n \geq 1$、 $k = 1$、.....、 K 上一致地小于某个有限常数 $C > 0$； 最小非

零奇异值 $\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid B_i=k\right]$ 在 $n\geq 1$、$k=1$、.....、K 上一致地大于某个常数 $c>0$；

(c) 满足 (d) $\liminf_{n\to\infty}\left(\varsigma^2_{\widetilde{Y}}-\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}\right)>0$。(2) $\sup_{1\leq k\leq K}\left\|\left\{\frac{1}{n_{[k]}}\sum_{i\in[k]}(A_i-\pi_{n[k]})^2\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_{n[k]}\right\}\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_{n[k]}\right\}^\top\right\}^+\right\|=O_P(1)$。假设 4 与假设3非常相似，不同之处在于我们允许 $\boldsymbol{\xi}_n$ 的维度和层数都趋于无穷。假设4(1)(a)限制了 d 和 K的增长率。假设矩阵 $\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i=k\right]$ 是非奇异的，且 $\boldsymbol{\xi}_n^*(\boldsymbol{x})$中的元素均匀有界，则有 $r_n=d$ 和 $\zeta_n=\sqrt{d}$。考虑各层大小相等的情形，即 $p_{[k]}=1/K$。此时，假设4(1)(a)的一个充分条件是 $Kd=o(\sqrt{n})$，这意味着我们允许 $\boldsymbol{\xi}_n$ 的维度与层数的乘积增长，但增长速度慢于 $\sqrt{n}$。当 K 固定时，这一条件与 Lei & Ding (2021) 在完全随机化和有限总体渐近理论下针对回归调整的平均处理效应估计量提出的类似结果一致。为进一步放宽对 d增长率的限制，Lei & Ding (2021)、Lu 等（2025）和 Gu 等（2025）提出了去偏回归调整估计量，允许 d 以快于 $\sqrt{n}$的速度增长。相比之下，Jiang 等（2025）通过假设潜在结果与协变量之间存在线性关系且模型设定正确，实现了相同目标。也可以对校准估计量 (2.1) 进行去偏处理，从而放宽假设4(1)(a)。然而，这超出了本文的研究范围，将留待未来研究。假设4(1)(b)要求最大和最小非零奇异值

$$\sup_{1\leq k\leq K}\left\|\frac{1}{n_{1[k]}}\sum_{i\in[k]}A_i\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}-\frac{1}{n_{0[k]}}\sum_{i\in[k]}(1-A_i)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}\right\|$$

$$\text{and}\ \sup_{1\leq k\leq K}\frac{1}{n_{[k]}}\sum_{i\in[k]}\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|=O_P(n^{-1/2}),$$

$\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\,\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid B_i=k\right]$ 有界且远离零和无穷大，同时仍允许该矩阵为奇异矩阵。假设4(1)(c)强化了假设3(1)(c)。在交叉拟合下，假设4(1)(c)要求 $\boldsymbol{\xi}_n$ 以 $\boldsymbol{\xi}_n^*$ 的速率收敛至其概率极限， $n^{-1/4}$ 且在 $k=1,\ldots,K$ 上均匀成立。许多现代机器学习方法都能达到这样的速率（例如，参见Farrell 等人，2021、Jiao 等人，2023）。假设4(2)与假设3(2)类似，不同之处在于它要求在 $k=1,\ldots,K$ 上均匀成立。现在我们可以陈述以下定理。

定理 3。 假设 *1–2* 和 *4* 成立，且 $D(v)=(v-1)^2/2$。则

$$\frac{\sqrt{n}\,(\widehat{\tau}_{\text{cal}}-\tau)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\xrightarrow{d}N(0,1)\ \textit{and}\ \widehat{\varsigma}_H^2+\widehat{\varsigma}_{\widetilde{Y}}^2-\widehat{\varsigma}_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2=\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2+o_P(1).$$

此外，当 K 为固定数值时，若对于每个 $k=1$、……、K，都存在一个非随机向量 $\boldsymbol{\alpha}_k\in\mathbb{R}^d$，使得

$$\left\{\sqrt{\frac{1-\pi_{[k]}}{\pi_{[k]}}}\widetilde{h}_{1[k]}^*(\boldsymbol{X})+\sqrt{\frac{\pi_{[k]}}{1-\pi_{[k]}}}\widetilde{h}_{0[k]}^*(\boldsymbol{X})\right\}\mathbb{1}(B=k)=\boldsymbol{\alpha}_k^\top\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X})\mathbb{1}(B=k),$$

then

$$\frac{\sqrt{n}\,(\widehat{\tau}_{\text{cal}}-\tau)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{h}^*}^2}}\xrightarrow{d}N(0,1),$$

且渐近方差 $\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{h}^*}^2$ 与 *Rafi (2023*，定理 *3.1*）中推导出的半参数效率界相匹配。

定理3 表明，即使 $\boldsymbol{\xi}_n$ 的维度和层数都趋于无穷，校准估计量（2.1）仍然渐近正态。需要注意的是，由于我们允许层数随样本量增加，假设2 变得并非平凡。Xin 等人（2024）研究了层数不断增加时受限随机化下的推断，而我们的假设2 与他们的假设（B1）–（B3）直接对应。此外，Xin 等人（2024）针对简单随机化、分层自适应偏倚硬币随机化 2 和分层区组随机化的情形，推导出了假设

成立所需的更基本条件。关于假设2的进一步讨论，请参阅Xin 等人（2024，第3节）。

5.2 一般差异度量 $D(v)$

在定理 1中，我们的分析仅限于二次差异度量 $D(v) = (v-1)^2/2$。本节将分析扩展到一类一般差异度量$D(v)$，理论上可以证明，这类度量具有较小的二阶偏差。为此，我们对 $D(v)$施加以下较弱的正则性条件。

假设5。令 D' 为 D 的导数，$(D')^{-1}$ 为 D'的反函数。定义 $\rho(v) := D\{(D')^{-1}(-v)\} + v\cdot(D')^{-1}(-v)$。我们假设 $\rho(v)$是凹函数且三次连续可微，$\rho'(0) = 1$，$-\infty < \rho''(0) < 0$，并且存在常数 $\delta > 0$ 和$0 \le C_\rho < \infty$，使得对所有 $|v| \le \delta$，$|\rho'''(v) - \rho'''(0)| \le C_\rho |v|$。

假设5 适用于多种常用的差异度量，表1列举了几个常见的例子。当 $D(v) = v\log v - v$ 时，对应于Kitamura 和 Stutzer (1997年)研究的指数倾斜估计量；而当 $D(v) = v - \log v$时，对应于Qin 和 Lawless (1994年)研究的经验似然估计量。

表1： $D(v)$的不同取值及其对应的 $\rho(v)$。

$D(v)$	$\rho(v)$	$\rho'(v)$	$\rho''(0)$	$\rho'''(0)$
$(v-1)^2/2$	$-v^2/2 + v$	$-v+1$	-1	0
v 对数 $v - v$	$-e^{-v}$	e^{-v}	-1	1
$v -$ 对数 v	$1 +$ 对数$(1+v)$	$\frac{1}{1+v}$	-1	2

由于我们的目标是刻画校准估计量的二阶偏差，因此我们需要假设 3 的加强版本。

假设6。 *(1)* 随机函数序列 $\boldsymbol{\xi}_n(\cdot)$ 满足 $\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\|^4 = O_P(1)$;

(2) 非随机函数 $\boldsymbol{\xi}_n^*(\cdot)$: $\mathcal{X} \to \mathbb{R}^d$ $(n \ge 1)$ 在假设3 中满足

以下条件：$\sup_{n\geq 1}\mathbb{E}\left[\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\|^4\right]<\infty$，且对每个 $k=1,\ldots,K$，

$$\left\|\frac{1}{n_{1[k]}}\sum_{i\in[k]}A_i\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}-\frac{1}{n_{0[k]}}\sum_{i\in[k]}(1-A_i)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}\right\|=o_P(n^{-1}),$$

以及 $\frac{1}{n_{[k]}}\sum_{i\in[k]}\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\|^2=O_P(\Delta_n^2)$，其中 $\Delta_n\to 0$ 而 $n\to\infty$ 是一个实数序列。

我们有如下定理。

定理 4。 假设 *1–3* 和 *5* 成立。若 $\sup^{n\geq 1}\sup_{1\leq i\leq n}\mathbb{E}\left[\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\|^{2+\epsilon}\right]<\infty$ 对于某个 $\epsilon>0$ 且 $\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top\mid B_i=k\right]$ 对每个 $k=1,\ldots,K$, 都是非奇异的，则定理 *1* 的结论成立。

If, 此外，假设 *6* 成立，并且 $\mathbb{E}\left[Y_i(a)^4\right]<\infty$ 对所有 $a=0$，1，则

$$\widehat{\tau}_{\text{cal}}-\tau=\frac{1}{n}\sum_{i=1}^{n}\psi_{1,i}+\frac{1}{n}\psi_2+o_P(n^{-1})+O_P(\Delta_n n^{-1/2}),$$

其中 $\psi_{1,\ i}=\sum_{k=1}^{K}\left\{\left(\frac{A_i}{\pi_{n[k]}}-\frac{1-A_i}{1-\pi_{n[k]}}\right)\mathbb{1}(B_i=k).Y_i-\tau-\boldsymbol{\beta}_{[k],\ \mathcal{C}_n}^{\top}\boldsymbol{\Xi}_{i,\ [k]}^*\right\}$ 其中 $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\psi_{1,\ i}\right]=0$ 且 $\mathbb{E}\left[\psi_2\right]=\left\{\frac{\rho'''}{2\rho''}(0)\ (0)^2-1\right\}\sum_{k=1}^{K}\operatorname{tr}\left\{\mathbb{E}\left[\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}^{\mathcal{C}_n}\right]\right\}$。 、$\mathcal{C}_n\boldsymbol{\Xi}_{i,\ [k]}^*$和 $\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}^{\mathcal{C}_n}$ 的定义见附录 *B*。

定理 4 的第一个结论表明，只要假设5成立， $D(v)$ 的不同选择会得到具有相同渐近分布的校准估计量。我们还注意到，该结论要求 $\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top\mid B_i=k\right]$ 的非奇异性，以确保 (2.2) 在渐近意义下存在唯一解。在 $D(v)=(v-1)^2/2$ 这一特殊情形下，该条件并非必要。这是因为，在二次差异下，权重 $\widehat{w}_i, i=1,\ldots,n$ 具有闭式表达式。即使 (2.2) 存在无穷多个解，也始终可以选择经由摩尔-彭若斯伪逆得到的解（见附录 B.4 中的定理 4 的证明中的 (B.7)）。相比之下，当 $D(v)$ 为一般差异度量时，权重 $\widehat{w}_i$ 不再具有显式形式。在这种情况下，如果 (2.2) 存在无穷多个解，则不清楚哪一个解

优化算法会收敛到。在实践中，可以引入正则化项以强制解的唯一性，尽管这超出了本文的范围。此外，如果 $\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right]$ 是奇异的，则可以应用奇异值分解（SVD）来消除 $\boldsymbol{\xi}_n$ 的冗余分量，并使用 $\boldsymbol{\xi}_n$ 的约化版本来计算校准估计量（2.1）。我们将这一策略的严格理论发展留待未来研究。

定理 4 的第二个结论刻画了受限随机化设计下的二阶偏差；据我们所知，这是对现有文献的一项新贡献。与关于二阶偏差的现有研究（Newey & Smith 2004、Tan 2014）不同，受限随机化设计下的观测样本并非独立同分布。该结果的证明依赖于条件论证：我们首先考察在处理分配和层指标条件下校准估计量的渐近展开，然后应用迭代期望定律。由定理 4可知，当 $\Delta_n = o(n^{-1/2})$时，取 $D(v) = v - \log v$可得到二阶偏差为零的校准估计量 $\mathbb{E}[\psi_2]$。这与独立同分布情形下的发现一致：基于经验似然的估计量具有较小的二阶偏差（Newey & Smith 2004、Tan 2014）。

6 模拟研究

我们使用蒙特卡洛模拟评估校准估计量的性能。对于 $a \in \{0, 1\}$ 以及 $1 \le i \le n$，潜在结果按如下方式生成：

$$Y_i(a) = g_a(\boldsymbol{X}_i) + \epsilon_{a,i} \quad i = 1, \ldots, n, \ a \in \{0, 1\},$$

其中，$\boldsymbol{X}_i$、$\epsilon_{a,\,i}$ 和 $g_a(\cdot)$ 将在每个模型中具体指定，而三元组 $(\boldsymbol{X}_i, \epsilon_{0,\,i}, \epsilon_{1,\,i})$ 对于$1 \le i \le n$ 而言是独立同分布的。

我们展示了三种随机化方法下各估计量的模拟结果：简单随机化、分层区组随机化和最小化法（Pocock &Simon1975）。

样本量 n 在 500、1000 和 2000 之间变化，协变量数量设为 $p = 30$。对于分层区组随机化，我们采用区组大小为 6。最小化法中，采用 0.75 的偏倚硬币概率和相等权重。如第 3.3 节所示，交叉拟合技术采用两折。除非另有说明，校准估计量通过设定 $D(v) = (v-1)^2/2$ 得到。我们比较九种估计量：(i) `cal_rf`：对于每个层 k，我们使用随机森林估计条件均值函数 $g_0(\cdot)$ 和 $g_1(\cdot)$，并将其估计值记为 $\widehat{g}^{\mathrm{rf}}_{0k}(\cdot)$ 和 $\widehat{g}^{\mathrm{rf}}_{1k}(\cdot)$。然后，在校准估计量 (2.1) 中取 $\boldsymbol{\xi}_n(\boldsymbol{X}_i) = (\sum_{k=1}^K \widehat{g}^{\mathrm{rf}}_{0k}(\boldsymbol{X}_i)1(B_i = k)\ \sum_{k=1}^K \widehat{g}^{\mathrm{rf}}_{1k}(\boldsymbol{X}_i)1(B_i = k))^\top$，得到 `cal_rf`；(ii) `cal_nn`：定义方式与 `cal_rf` 类似，但将随机森林估计值替换为神经网络估计值 $\widehat{g}^{\mathrm{rf}}_{0k}(\cdot)$ 和 $\widehat{g}^{\mathrm{rf}}_{1k}(\cdot)$；(iii)`cal_rfnn`：该估计量通过结合随机森林和神经网络估计值获得。在校准估计量 (2.1) 中取 $\boldsymbol{\xi}_n(\boldsymbol{X}_i) = (\sum_{k=1}^K \widehat{g}^{\sharp}_{0k}(\boldsymbol{X}_i)1(B_i = k)\ \sum_{k=1}^K \widehat{g}^{\sharp}_{1k}(\boldsymbol{X}_i)1(B_i = k) : \sharp \in \{\mathrm{rf}, \mathrm{nn}\})^\top$，即可得到该估计量；(iv) `cal_rflin`：定义方式与`cal_rfnn`类似，但通过结合随机森林和线性回归估计值获得。(v)`cal_rf_g`：取 $\boldsymbol{\xi}_n(\boldsymbol{X}_i) = (\widehat{g}^{\mathrm{rf}}_{0k}(\boldsymbol{X}_i)\ \widehat{g}^{\mathrm{rf}}_{1k}(\boldsymbol{X}_i) : 1 \le k \le K)^\top$，并将其用于校准估计量 (2.1)；(vi) `cal_nn_g`：定义方式与 `cal_rf_g` 类似，但采用神经网络估计值；(vii) `cal_lin_EL`：定义方式与 `cal_rf` 类似，但将随机森林估计值替换为线性回归估计值，并取 $D(v) = v - \log v$；(viii) `aipw_rf`：我们采用 AIPW-base 方法（Tu 等人，2024），使用随机森林估计值 $\widehat{g}^{\mathrm{rf}}_{0k}(\boldsymbol{X}_i)$ 和 $\widehat{g}^{\mathrm{rf}}_{1k}(\boldsymbol{X}_i)$ 调整协变量；(ix) `aipw_nn`：定义方式与 `aipw_rf` 类似，但将随机森林估计值替换为神经网络估计值 $\widehat{g}^{\mathrm{nn}}_{0k}(\cdot)$ 和 $\widehat{g}^{\mathrm{nn}}_{1k}(\cdot)$；(x)`aipw_lin`：定义方式与 `aipw_rf` 类似，但将随机森林估计值替换为线性回归估计值；(xi) `sdim`：分层均值差估计量 $\widehat{\tau}_{\mathrm{sdim}}$。对于每种估计量，我们报告基于 300 次重复模拟计算的以下指标：绝对偏差、经验标准差（SD）、平均估计标准误（SE），以及 95%置信区间的经验覆盖概率（CP）。M

odel 1. Model 1 对条件均值函数 $g_0(\boldsymbol{X})$ 和

$g_1(\boldsymbol{X})$ 设定线性模型。在该模型中，我们设定

$$g_0(\boldsymbol{X}_i) = \mu_0 + \sum_{j=1}^{4} \beta_{0j} X_{ij} \quad \text{and} \quad g_1(\boldsymbol{X}_i) = \mu_1 + \sum_{j=1}^{4} \beta_{1j} X_{ij},$$

其中 $\mu_0 = 1, \mu_1 = 4$，$(\beta_{01}, \ldots, \beta_{04}) = (75, 35, 125, 80)$，且 $(\beta_{11}, \ldots, \beta_{14}) = (100, 80, 60, 40)$。此外，变量定义如下：$\epsilon_{0,i} \sim N(0, 1)$，$\epsilon_{1,i} \sim N(0, 9)$，$X_{i1} \sim$贝塔分布$(3, 4)$，$X_{i2} \sim$均匀分布$(-2, 2)$，$X_{i3}$ 以相同概率取值于 $\{-1, 1\}$，X_{i4}以 0.6 和 0.4 的概率分别取值于 $\{3, 5\}$。这些变量彼此独立。其余变量 $X_{i5}, \ldots, X_{ip}$ 与 $X_{i1}, \ldots, X_{i4}$ 独立，且服从零均值多元正态分布，其协方差矩阵的所有非对角元素均为 0.2，对角元素均为 1。随机化变量是一个额外变量，以 0.2、0.3、0.3 和 0.2 的概率分别取值于 $\{1, 2, 3, 4\}$，并且与 $j = 1, \ldots, p$ 的 X_{ij} 独立。模型1的模拟结果见表 2。由于 $g_0(\boldsymbol{X}_i)$ 和 $g_1(\boldsymbol{X}_i)$ 都是协变量 X_{ij} 的线性函数，因此，正如预期，线性回归调整估计量 `aipw_lin` 在大样本（$n = 1000$，2000）中表现最佳。`cal_rflin` 的表现与 `aipw_lin` 非常接近，这与定理2的结论一致。然而，在较小样本（$n = 500$）中，`aipw_lin` 估计量的表现并非最佳，因为线性回归对异常值较为敏感。相比之下，`cal_rflin`的表现保持稳定，这表明将 $g_0(\boldsymbol{X}_i)$ 和 $g_1(\boldsymbol{X}_i)$ 的不同估计值纳入校准估计量，能够提高其稳健性。

模型2。模型2 对条件均值函数$g_0(\boldsymbol{X})$和 $g_1(\boldsymbol{X})$施加可加但非线性的模型。在此模型中，我们设定

$$g_0(\boldsymbol{X}_i) = \mu_0 + \beta_{01} \log(X_{i1} + 1) + \beta_{02} X_{i2}^2 + \beta_{03} \exp(X_{i3}) + \beta_{04}/(X_{i4} + 3)$$
$$g_1(\boldsymbol{X}_i) = \mu_1 + \beta_{11} \exp(X_{i1} + 2) + \beta_{12}/(X_{i1} + 1) + \beta_{13} X_{i2}^2,$$

其中 $\mu_0 = -3$, $\mu_1 = 0$、$(\beta_{01}, \ldots, \beta_{04}) = (10, 24, 15, 20)$，以及 $(\beta_{11}, \beta_{12}, \beta_{13}) = (20, 17, 10)$。

表 2：模型1下不同估计量性能的比较。

n	估计量	简单随机化			分层区组随机化				最小化法			
		Bias	SD	SE	覆盖概率偏倚	SD	SE	覆盖概率偏倚	SD	SE	CP	
500	cal_rf	0.12	6.85	7.39	0.973	0.12	6.74	7.31	0.960	0.36	6.98	7.33	0.953
	cal_nn	0.33	11.38	11.03	0.960	0.65	11.01	11.05	0.950	0.63	11.47	11.00	0.933
	cal_rfnn	0.47	6.90	7.33	0.960	0.18	6.90	7.24	0.960	0.12	7.03	7.27	0.950
	cal_rflin	0.08	4.36	5.23	0.983	0.35	4.78	5.19	0.967	0.37	4.60	5.19	0.973
	cal_rf_g	0.13	5.83	6.08	0.967	0.48	6.01	6.07	0.960	0.01	5.68	6.09	0.960
	cal_nn_g	0.48	9.23	9.10	0.960	0.79	9.41	9.15	0.943	0.31	9.31	9.10	0.950
	cal_lin_EL	0.10	4.25	5.26	0.980	0.32	4.49	5.21	0.983	0.14	4.39	5.22	0.987
	aipw_rf	0.31	9.03	8.91	0.957	0.07	8.74	8.88	0.940	0.65	9.13	8.88	0.943
	aipw_nn	0.29	14.17	13.55	0.923	0.74	14.45	13.54	0.910	1.22	13.98	13.52	0.950
	aipw_lin	3.22	50.59	14.23	0.953	0.69	25.12	9.04	0.967	1.78	19.45	7.30	0.953
	sdim	0.48	12.35	12.34	0.953	0.06	11.71	12.31	0.943	0.93	12.61	12.28	0.927
1000	cal_rf	0.21	4.00	4.13	0.957	0.06	3.70	4.12	0.960	0.09	3.85	4.12	0.963
	cal_nn	0.10	6.39	6.48	0.963	0.15	6.42	6.48	0.940	0.38	6.14	6.44	0.960
	cal_rfnn	0.05	3.93	4.06	0.963	0.06	3.78	4.05	0.960	0.17	3.90	4.05	0.973
	cal_rflin	0.17	2.94	3.28	0.973	0.24	2.84	3.27	0.973	0.05	3.14	3.27	0.957
	cal_rf_g	0.24	3.68	3.66	0.940	0.05	3.45	3.67	0.963	0.08	3.67	3.66	0.950
	cal_nn_g	0.40	4.48	4.73	0.967	0.18	4.83	4.76	0.950	0.55	4.74	4.74	0.950
	cal_lin_EL	0.12	2.87	3.27	0.973	0.29	2.76	3.27	0.977	0.05	3.06	3.27	0.970
	aipw_rf	0.07	5.11	5.24	0.967	0.04	5.25	5.23	0.947	0.16	5.27	5.24	0.950
	aipw_nn	0.17	7.96	7.51	0.943	0.26	7.52	7.53	0.953	0.36	7.38	7.44	0.973
	aipw_lin	0.08	2.81	2.90	0.943	0.32	2.74	2.91	0.960	0.05	3.03	2.91	0.940
	sdim	0.11	8.27	8.71	0.953	0.10	8.70	8.67	0.950	0.27	8.61	8.67	0.953
2000	cal_rf	0.10	2.49	2.53	0.960	0.02	2.32	2.53	0.960	0.11	2.50	2.53	0.967
	cal_nn	0.27	2.50	2.68	0.960	0.01	2.62	2.68	0.963	0.05	2.64	2.67	0.950
	cal_rfnn	0.17	2.23	2.33	0.960	0.04	2.17	2.33	0.963	0.04	2.32	2.33	0.953
	cal_rflin	0.12	2.04	2.19	0.967	0.04	2.06	2.19	0.973	0.02	2.11	2.19	0.970
	cal_rf_g	0.13	2.37	2.38	0.950	0.07	2.22	2.38	0.957	0.14	2.33	2.38	0.953
	cal_nn_g	0.28	2.09	2.23	0.957	0.15	2.19	2.24	0.957	0.05	2.18	2.23	0.967
	cal_lin_EL	0.16	2.01	2.19	0.970	0.00	2.03	2.19	0.970	0.02	2.08	2.19	0.973
	aipw_rf	0.12	3.04	3.05	0.957	0.15	3.00	3.05	0.953	0.20	3.10	3.05	0.947
	aipw_nn	0.29	2.77	2.91	0.967	0.04	2.93	2.89	0.927	0.06	2.81	2.88	0.953
	aipw_lin	0.16	1.99	2.06	0.967	0.00	2.02	2.06	0.957	0.03	2.06	2.05	0.963
	sdim	0.22	5.87	6.13	0.967	0.41	6.28	6.13	0.940	0.37	6.12	6.13	0.953

*Ab*缩写：随机化.，随机化；标准差，标准差，标准误：标准误；覆盖概率，覆盖概率 .

此外，变量设定如下：$\epsilon_{0,\ i} \sim N(0,1)$，$\epsilon_{1,\ i} \sim N(0,9)$，$X_{i1}$ ~贝塔分布$(3,4)$，X_{i2} ~均匀分布$(-2,\ 2)$，这两个变量相互独立。额外协变量 X_{i3}、. . . 、 X_{ip} 首先按照模型1的方式生成。然后，我们从额外协变量中随机选取 $\lfloor p/3 \rfloor$ 个协变量，并以相同的概率将它们分别乘以 X_{i1} 或 X_{i2}，以形成最终的额外协变量。随机化变量是一个取值为{1、2、3、4}的额外变量，其对应概率分别为 0.2、0.3、0.3 和 0.2，并且与 X_{ij}（$j = 1$、. . . 、 p）相互独立。模型2的模拟结果为

见表3。结果表明，基于随机森林的校准估计量（cal_rf、 cal_rfnn、 cal_rflm、 cal_rf_g）在不同随机化方案和样本量下均持续优于其他方法。基于增广逆概率加权的估计量（aipw_rf和 aipw_nn）的表现往往不如校准估计量（cal_rf 和 cal_nn），尤其是在样本量较小（$n = 500$）时。经验覆盖概率表明，在大多数情况下，这些估计量得出的95%置信区间具有较高的可靠性。

表3：模型2下不同估计量的性能比较。

n	估计量	简单随机化			分层区组随机化				最小化法				
		Bias	SD	SE	覆盖概率偏倚		SD	SE	覆盖概率偏倚		SD	SE	CP
500	cal_rf	0.01	2.09	2.30	0.980	0.19	2.22	2.30	0.957	0.28	2.25	2.29	0.967
	cal_nn	0.16	3.09	3.14	0.957	0.14	3.19	3.14	0.940	0.10	3.08	3.13	0.960
	cal_rfnn	0.01	2.14	2.30	0.977	0.20	2.25	2.30	0.960	0.25	2.26	2.29	0.957
	cal_rflin	0.22	2.11	2.26	0.977	0.01	2.23	2.26	0.960	0.11	2.23	2.26	0.960
	cal_rf_g	0.10	2.00	2.16	0.957	0.19	2.25	2.16	0.937	0.15	2.18	2.16	0.950
	cal_nn_g	0.12	3.23	3.09	0.943	0.19	3.21	3.09	0.937	0.05	3.13	3.08	0.957
	cal_lin_EL	0.33	2.68	2.78	0.950	0.19	2.73	2.77	0.960	0.10	2.69	2.76	0.963
	aipw_rf	0.20	2.35	2.42	0.950	0.04	2.50	2.42	0.940	0.09	2.46	2.42	0.957
	aipw_nn	0.87	5.98	5.99	0.963	0.47	6.02	5.92	0.957	0.25	5.94	5.87	0.937
	aipw_lin	2.91	196.74	43.13	0.953	3.67	142.28	37.24	0.963	6.10	95.39	33.45	0.937
	sdim	0.22	3.07	3.12	0.950	0.14	3.19	3.12	0.940	0.08	3.06	3.12	0.970
1000	cal_rf	0.22	1.47	1.51	0.957	0.15	1.48	1.50	0.950	0.29	1.54	1.50	0.943
	cal_nn	0.04	2.24	2.15	0.960	0.09	2.16	2.13	0.963	0.27	2.15	2.14	0.950
	cal_rfnn	0.22	1.49	1.50	0.953	0.15	1.50	1.50	0.947	0.25	1.55	1.50	0.940
	cal_rflin	0.08	1.44	1.47	0.960	0.08	1.43	1.47	0.957	0.15	1.53	1.46	0.937
	cal_rf_g	0.17	1.46	1.45	0.940	0.12	1.46	1.45	0.947	0.25	1.50	1.45	0.937
	cal_nn_g	0.02	2.17	2.02	0.950	0.11	2.08	2.00	0.937	0.30	2.04	2.02	0.943
	cal_lin_EL	0.07	1.74	1.67	0.947	0.18	1.70	1.67	0.940	0.25	1.66	1.67	0.947
	aipw_rf	0.07	1.62	1.59	0.947	0.06	1.60	1.59	0.943	0.26	1.63	1.59	0.943
	aipw_nn	0.17	3.36	3.22	0.943	0.05	3.40	3.22	0.950	0.27	3.28	3.26	0.960
	aipw_lin	0.11	1.75	1.67	0.950	0.19	1.73	1.66	0.947	0.30	1.67	1.66	0.947
	sdim	0.01	2.25	2.20	0.950	0.07	2.27	2.20	0.933	0.31	2.20	2.20	0.960
2000	cal_rf	0.09	1.00	1.01	0.943	0.05	1.04	1.01	0.937	0.09	0.97	1.01	0.953
	cal_nn	0.03	1.26	1.27	0.950	0.13	1.31	1.27	0.950	0.05	1.25	1.27	0.950
	cal_rfnn	0.07	1.01	1.01	0.943	0.07	1.04	1.01	0.923	0.07	0.97	1.01	0.953
	cal_rflin	0.01	1.00	0.99	0.943	0.14	1.02	0.99	0.923	0.00	0.96	0.99	0.957
	cal_rf_g	0.05	1.00	0.99	0.950	0.08	1.04	0.99	0.927	0.07	0.97	0.99	0.950
	cal_nn_g	0.03	1.10	1.07	0.957	0.13	1.08	1.07	0.930	0.04	1.08	1.07	0.953
	cal_lin_EL	0.01	1.15	1.13	0.943	0.12	1.14	1.13	0.943	0.03	1.13	1.13	0.963
	aipw_rf	0.01	1.07	1.06	0.947	0.12	1.09	1.05	0.933	0.05	1.06	1.06	0.960
	aipw_nn	0.06	1.38	1.44	0.957	0.13	1.57	1.43	0.923	0.12	1.49	1.43	0.940
	aipw_lin	0.00	1.14	1.12	0.930	0.11	1.14	1.12	0.937	0.02	1.11	1.12	0.960
	sdim	0.03	1.58	1.55	0.943	0.15	1.58	1.55	0.930	0.12	1.64	1.55	0.933

缩写：随机化，随机化；标准差，标准差，标准误：标准误；覆盖概率，覆盖概率 y.

模型3。 模型3对条件均值采用非可加和非线性模型

函数 $g_0(\boldsymbol{X})$ 和 $g_1(\boldsymbol{X})$。在此模型中，我们设定

$$g_0(\boldsymbol{X}_i) = \mu_0 + \beta_{01} X_{i1} X_{i2}/(X_{i1} + X_{i2} + 2) + \beta_{02} X_{i1}^2 (X_{i2} + X_{i3})$$
$$g_1(\boldsymbol{X}_i) = \mu_1 + \beta_{11}(X_{i2} + X_{i4}) + \beta_{12} X_{i2}^2 / \exp(X_{i1} + 2),$$

其中 $\mu_0 = 5$, $\mu_1 = 2$, $(\beta_{01}, \beta_{02}) = (42, 83)$，以及 $(\beta_{11}, \beta_{12}) = (30, 75)$。此外，变量定义如下：$\epsilon_{0,i} \sim$ t(2)，$\epsilon_{1,i} \sim 3 \times$ t(2)，$X_{i1} \sim$ 贝塔分布(3, 4)，$X_{i2} \sim$均匀分布$(-2, 2)$，$X_{i3} \sim N(0, 1)$，$X_{i4} \sim$均匀分布(0, 2)，依次对应。这些变量彼此独立。其余变量 $X_{i5}, \ldots, X_{ip}$ 与 $X_{i1}, \ldots, X_{i4}$ 相互独立，且服从零均值的多元正态分布，其协方差矩阵为对称托普利茨矩阵，首行为首项为 1、公比为 0.5 的等比数列。随机化变量是一个额外变量，取值为 {1, 2} 的概率分别为 0.4 和 0.6，并且对于 $j = 1, \ldots, p$，它与 X_{ij} 独立。模型3的模拟结果见表4。结果表明，在不同的随机化方法和样本量下，基于随机森林的校准估计量（`cal_rf`, `cal_rfnn`, `cal_rflm`, `cal_rf_g`）表现最佳。随着样本量增加，校准估计量逐渐展现出更好的性能，标准差更低，同时保持正确的经验覆盖概率。然而，在标准差方面，`sdim` 估计量始终不如校准估计量，尤其是在样本量较小时。这一发现与定理1一致；定理1表明，校准估计量的效率始终高于 `sdim`估计量。

模型1-3假设不同层之间条件均值函数 $g_0(\boldsymbol{X})$ 和 $g_1(\boldsymbol{X})$具有同质性。在附录C中，我们还考察了一个能够考虑不同层之间条件均值函数 $g_0(\boldsymbol{X})$ 和 $g_1(\boldsymbol{X})$ 异质性的模型。我们的发现表明，其结果与模型1–3下得到的结果非常相似。

表4：模型3下不同估计量性能的比较。

n	估计量	简单随机化			分层区组随机化				最小化法				
		Bias	SD	SE	覆盖概率偏倚		SD	SE	覆盖概率 偏倚		SD	SE	CP
500	cal_rf	0.13	2.80	2.68	0.933	0.01	2.40	2.61	0.980	0.03	2.41	2.65	0.967
	cal_nn	0.17	3.62	3.43	0.943	0.08	3.20	3.35	0.960	0.08	3.29	3.38	0.960
	cal_rfnn	0.09	2.83	2.67	0.940	0.00	2.45	2.61	0.967	0.01	2.41	2.65	0.960
	cal_rflin	0.08	2.80	2.63	0.930	0.01	2.36	2.56	0.967	0.02	2.42	2.61	0.957
	cal_rf_g	0.16	2.74	2.58	0.927	0.01	2.27	2.52	0.980	0.11	2.35	2.56	0.957
	cal_nn_g	0.08	3.23	3.16	0.933	0.08	2.94	3.08	0.953	0.08	2.93	3.12	0.960
	cal_lin_EL	0.09	3.00	2.82	0.923	0.10	2.51	2.76	0.970	0.03	2.66	2.80	0.967
	aipw_rf	0.04	3.08	2.84	0.937	0.16	2.60	2.78	0.960	0.16	2.67	2.82	0.953
	aipw_nn	0.11	5.09	4.88	0.943	0.12	4.72	4.62	0.947	0.23	4.93	4.79	0.947
	aipw_lin	0.13	3.16	3.01	0.937	0.12	2.79	2.89	0.930	0.02	2.83	2.95	0.977
	sdim	0.19	4.31	4.00	0.930	0.23	3.77	3.93	0.950	0.26	3.87	3.97	0.960
1000	cal_rf	0.03	1.87	1.79	0.947	0.06	1.82	1.80	0.940	0.07	1.66	1.78	0.957
	cal_nn	0.09	2.31	2.18	0.937	0.02	2.15	2.20	0.947	0.02	2.20	2.18	0.960
	cal_rfnn	0.06	1.89	1.79	0.937	0.08	1.82	1.79	0.953	0.08	1.68	1.78	0.960
	cal_rflin	0.06	1.88	1.76	0.943	0.05	1.82	1.77	0.943	0.08	1.69	1.75	0.953
	cal_rf_g	0.01	1.87	1.75	0.933	0.06	1.76	1.75	0.957	0.11	1.67	1.74	0.953
	cal_nn_g	0.04	2.12	2.01	0.950	0.04	1.94	2.02	0.963	0.15	2.03	2.01	0.963
	cal_lin_EL	0.02	1.96	1.82	0.943	0.05	1.87	1.82	0.953	0.04	1.76	1.81	0.967
	aipw_rf	0.00	2.04	1.87	0.943	0.01	1.91	1.88	0.947	0.01	1.81	1.86	0.957
	aipw_nn	0.20	2.98	2.66	0.920	0.01	2.97	2.70	0.950	0.05	2.76	2.66	0.950
	aipw_lin	0.00	1.98	1.83	0.940	0.08	1.89	1.83	0.947	0.04	1.79	1.82	0.953
	sdim	0.07	2.69	2.81	0.967	0.05	2.77	2.81	0.950	0.00	2.77	2.80	0.967
2000	cal_rf	0.10	1.31	1.22	0.943	0.01	1.20	1.21	0.947	0.06	1.41	1.24	0.937
	cal_nn	0.08	1.42	1.36	0.947	0.08	1.37	1.36	0.957	0.11	1.59	1.39	0.943
	cal_rfnn	0.10	1.29	1.21	0.943	0.03	1.21	1.21	0.947	0.09	1.43	1.24	0.940
	cal_rflin	0.10	1.29	1.20	0.940	0.01	1.20	1.20	0.943	0.06	1.41	1.23	0.940
	cal_rf_g	0.12	1.28	1.19	0.950	0.03	1.16	1.19	0.950	0.08	1.40	1.22	0.943
	cal_nn_g	0.10	1.34	1.28	0.940	0.07	1.30	1.28	0.953	0.13	1.55	1.31	0.940
	cal_lin_EL	0.07	1.34	1.24	0.930	0.04	1.24	1.23	0.957	0.02	1.46	1.26	0.953
	aipw_rf	0.05	1.36	1.25	0.930	0.03	1.26	1.25	0.950	0.01	1.46	1.27	0.940
	aipw_nn	0.08	1.54	1.45	0.930	0.05	1.48	1.45	0.963	0.06	1.77	1.49	0.933
	aipw_lin	0.07	1.31	1.23	0.933	0.03	1.25	1.23	0.953	0.04	1.57	1.26	0.930
	sdim	0.07	2.14	1.98	0.927	0.05	2.06	1.98	0.943	0.01	2.17	2.00	0.947

缩写：随机化、随机化；标准差、标准差；标准误：标准误；覆盖概率：覆盖概率。

7 实证应用

本节中，我们将校准方法应用于Dupas 等人(2018)的实验数据。该研究在三个国家——乌干达、马拉维和智利——开展了协变量自适应随机实验，以评估补贴银行账户获取（A_i）对个人总储蓄（Y_i）的影响。Dupas 等人(2018)估计了平均

补贴的平均处理效应（ATE）和分位数处理效应（QTE）。近期，Jiang 等人(2023) 将回归调整估计量应用于该数据集，以分析乌干达的 QTE。

本节旨在估计乌干达和马拉维银行账户补贴所带来的平均处理效应（ATE）。乌干达样本包含 2,159 个观测值，依据性别、职业和银行分行划分为 41 层，并采用分层区组随机化设计。马拉维样本包含 2,108 个观测值，依据职业、性别、婚姻状况、识字能力，以及受访者来自家庭还是市场划分为 78 层，同样采用分层区组随机化设计。为避免层内样本过少，我们剔除少于六个样本的层。最终，乌干达样本包含 2,115 个观测值，分布于 37 层；马拉维样本包含 1,987 个观测值，分布于 67 层。在这两个国家，每个家庭被随机分配至接受银行账户补贴的组或对照组，两组各占一半。所有货币变量均在第九十九百分位数处进行缩尾处理，以减轻异常值的影响。随机化和干预之后，Dupas 等人(2018) 在乌干达和马拉维开展了 3 轮后续调查。与Jiang 等人(2023) 一致，我们重点关注第一轮后续调查，以评估银行账户补贴对总储蓄的影响。

遵循 Dupas 等人（2018），我们考虑一个基线协变量 X：总储蓄的基线值。在估计乌干达补贴银行账户获取的平均处理效应时，我们利用马拉维的数据来获取校准估计量中使用的 $\boldsymbol{\xi}_n(X)$ 的某些组成部分；估计马拉维的平均处理效应时则反之。对于这两个国家，我们考虑以下估计量：(i) `cal_X`：通过在 (2.1) 中取 $\boldsymbol{\xi}_n(X) = X$ 得到该估计量；(ii) `cal_X`$^\beta$：通过在 (2.1) 中取 $\boldsymbol{\xi}_n(X) = (X+1)^\beta$ 得到该估计量；(iii) `cal_X_X`$^\beta$：通过在 (2.1) 中取 $\boldsymbol{\xi}_n(X) = (X, (X+1)^\beta)^\top$ 得到该估计量；(iv) `cal_info_X`：通过在 (2.1) 中取 $\boldsymbol{\xi}_n(X) = (X, \widehat{g}_{\text{rf}}^{\text{info}}(X))^\top$ 得到该估计量，其中 $\widehat{g}_{\text{rf}}^{\text{info}}(X)$ 是利用另一国家的数据得到的 $\mathbb{E}[Y \mid X]$ 的随机森林估计量；(v) `cal_info_X`$^\beta$：通过在 (2.1) 中取 $\boldsymbol{\xi}_n(X) = ((X+1)^\beta, \widehat{g}_{\text{rf}}^{\text{info}}(X))^\top$ 得到该估计量；

(vi) `cal_info_X_X`$^{\beta}$：通过在 (2.1) 中取 $\boldsymbol{\xi}_n(X) = (X, (X+1)^{\beta}, \hat{g}_{\mathrm{rf}}^{\mathrm{info}}(X))^{\top}$ 得到该估计量；(vii) `sdim`：分层均值差估计量 $\hat{\tau}_{\mathrm{sdim}}$。我们纳入项 $(X+1)^{\beta}$，因为两个国家的 log($Y+1$) 与 log($X+1$) 之间均观察到近似线性关系。我们拟合如下形式的线性回归：log($Y+1$) $= \alpha + \beta \log(X+1) + \epsilon$，得到乌干达和马拉维的估计 β 分别为 0.481 和 0.408。图2和表5展示了使用这些方法得到的 95%置信区间和点估计。

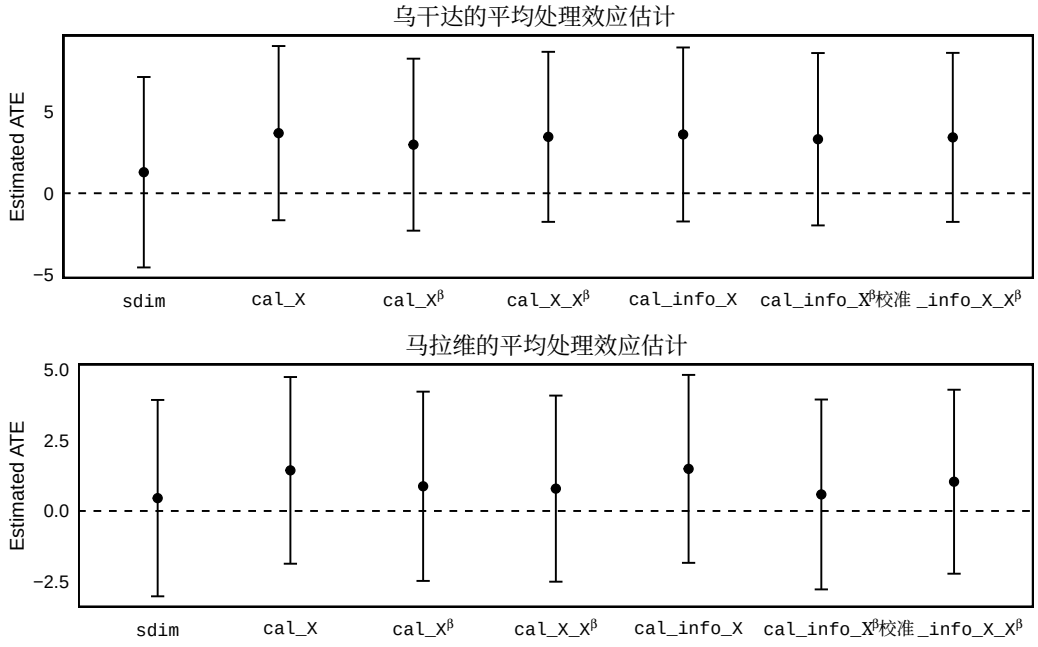


图2：银行账户补贴对乌干达和马拉维家庭总储蓄的平均处理效应的95%置信区间。上图：乌干达；下图：马拉维。名称中包含“X”（或“X$^{\beta}$”）的估计量表示，我们将分量 X（或 $(X+1)^{\beta}$）加入 $\boldsymbol{\xi}(X)$，以得到校准估计量。名称中包含“info”的估计量表示，我们纳入了另一个国家的信息，以构成 $\boldsymbol{\xi}(X)$ 的一个分量。

图 2 和表 5 所呈现的结果引出了两个关键观察。首先，与理论结果（定理 2）一致，`cal_info_X_X`$^{\beta}$估计量在乌干达和马拉维的所有估计量中标准误最低。例如，`cal info X X`$^{\beta}$ 平均处理效应估计的标准误分别比乌干达和马拉维的分层均值差ATE估计值小 11.2% 和 6.3%_ _ _。其次，在这两个国家，所有平均处理效应估计在统计上均不显著，这表明

表5：乌干达和马拉维银行账户补贴对家庭总储蓄影响的平均处理效应估计，标准误报告于括号中。

	sdim	cal_X	cal_X^{β}	cal_X_X^{β}	cal_info_X	cal_info_X^{β}	cal_info_X_X^{β}
乌干达	1.289	3.687	2.979	3.459	3.604	3.313	3.426
	(2.980)	(2.726)	(2.691)	(2.661)	(2.722)	(2.697)	(2.645)
马拉维	0.452	1.431	0.870	0.787	1.484	0.582	1.031
	(1.766)	(1.679)	(1.702)	(1.674)	(1.690)	(1.708)	(1.654)

注： 名称中包含“X”（或“X^{β}”）的估计量表示，我们将分量 X （或（$X+1$）$^{\beta}$）加入 $\boldsymbol{\xi}$（X），以得到校准估计量。名称中包含“info”的估计量表示，我们利用来自另一国家的信息，构成 $\boldsymbol{\xi}$（X）的一个分量。

扩大基本银行账户的可及性并不会显著提高平均总储蓄。这一发现与 Dupas 等人 (2018) 的研究结果一致。

参考文献

Bai, Y. (2022年),《配对设计在随机对照试验中的最优性》， 《美国经济评论》 **112**(12), 3911–3940。

Bai, Y., Jiang, L., Romano, J. P., Shaikh, A. M. & Zhang, Y. (2024),《配对实验中的协变量调整》， *Journal of Econometrics* **241**(1), 105740。

Bannick, M. S., Shao, J., Liu, J., Du, Y., Yi, Y. & Ye, T. (2025年),《协变量自适应随机化下临床试验中协变量调整的一般形式》， 《生物计量学》**112**(3), asaf029。

Boyd, S. P. & Vandenberghe, L. (2004年), 《凸优化》, 剑桥大学出版社，英国剑桥；纽约。

Breiman, L. (2001),《随机森林》， 《机器学习》**45**(1), 5–32。

Bugni, F. A.、Canay, I. A. 与 Shaikh, A. M.（2018）， 《协变量自适应随机化下的推断》， *Journal of the American Statistical Association* **113**(524)，1784–1796。

Bugni, F. A., Canay, I. A. & Shaikh, A. M. (2019), ‘多重处理下协变量自适应随机化的推断’, 《定量经济学》**10**(4),1747–1785.

Callegaro, A., Galwey, N. & Abellan, J. J. (2023), ‘临床试验中的历史对照：关于将 Pocock 模型与稳健混合先验联系起来的说明’, 生物统计学24(2), 443–448. Chan, K. C. G., Yam, S. C. P. & Zhang, Z. (2016), ‘通过经验平衡校准加权对平均处理效应进行全局有效的非参数推断’, 英国皇家统计学会杂志 *B*辑：统计方法学78(3), 673–700. Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018), ‘针对处理效应和结构参数的双重／去偏机器学习’, 计量经济学杂志21(1), C1–C68.

Chi, C.-M., Vossler, P., Fan, Y. & Lv, J. (2022), ‘高维随机森林的渐近性质’, *The Annals of Statistics* 50(6), 3415–3438.

Cohen, P. L. & Fogarty, C. B. (2024), ‘广义 Oaxaca–Blinder 估计量的无损害校准’, 《生物计量学》 111(1), 331–338.

Deville, J.-C. & Särndal, C.-E. (1992), ‘调查抽样中的校准估计量’,*Journalof the American Statistical Association*87(418), 376–382.

Dupas, P., Karlan, D., Robinson, J. & Ubfal, D. (2018), ‘为无银行账户者提供银行服务？来自三个国家的证据’, 美国经济学杂志：应用经济学 10(2), 257–297.

Efron, B. (1971), ‘使序贯试验达到平衡’, 《生物计量学》 58(3), 403– 417.

Farrell, M. H.、Liang, T. 与 Misra, S.（2021），《用于估计与推断的深度神经网络》，*Econometrica* 89(1)，181–213。

美国食品药品监督管理局（2019），《罕见病：用于药物开发的自然史研究：行业指南草案》，美国卫生与公众服务部。**网址：***https://www.fda.gov/media/122425/download*

Gu, Y., Liu, H. 与 Ma, W.（2023），《协变量自适应随机化下基于回归的多重处理效应估计》，《生物统计学》 79（4），2869–2880。

Gu, Y., Liu, H. 与 Ma, W.（2024），《纳入外部数据以分析随机临床试验：一种迁移学习方法》。arXiv 预印本 arXiv:2409.04126。

Gu, Y., Liu, L. & Ma, W.（2025年），《当 $p = o（n）$ 时基于假设精简的协变量调整》。arXiv 预印本 arXiv:2512.20046。

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J. & Sargent, D. J.（2011年），《临床试验中自适应纳入历史信息的层级可通约模型与幂先验模型》，《生物统计学》**67(3)**，1047–1056。

Ibrahim, J. G., Chen, M.-H., Gwon, Y. & Chen, F.（2015），《幂先验：理论与应用》，《医学统计学》**34(28)**，3724–3749。

Imbens, G. W. & Rubin, D. B.（2015），《统计学、社会科学与生物医学科学中的因果推断：导论》，剑桥大学出版社，纽约。

Jiang, L., Li, L., Miao, K. & Zhang, Y.（2025年），'协变量自适应随机化下多回归变量的调整'，*Journal of Econometrics* **249**，105991。

Jiang, L., Phillips, P. C. B., Tao, Y. & Zhang, Y.（2023），'协变量自适应随机化下分位数处理效应的回归调整估计'，*Journal ofEconometrics***234(2)**，758–776。

Jiao, Y., Shen, G., Lin, Y. & Huang, J.（2023），'近似流形上的深度非参数回归：带多项式前置因子的非渐近误差界'，*The Annals ofStatistics***51(2)**，691–716。

Kitamura, Y. & Stutzer, M.（1997年），'广义矩方法估计的一种信息论替代方法'，*Econometrica* **65(4)**，861–874。

Kwon, Y., Kim, J. K. 和 Qiu, Y.（2025年），《在调查抽样中使用广义熵的去偏校准估计》，*Journal of the American Statistical Association* **0(0)**，1–12。已接受，即将发表。

LeCun, Y.、Bengio, Y. 与 Hinton, G.（2015），"深度学习"，*Nature* **521**（7553），436–444。

Lei, L. 与 Ding, P.（2021），“具有发散数量的协变量的完全随机化实验中的回归调整”，《生物计量学》 108（4），815–828。

Lin, W.（2013年），“关于实验数据回归调整的无偏见笔记：重新审视Freedman的批评”，《应用统计年鉴》7(1)，295–318。

Liu, H.、Tu, F. 与 Ma, W.（2023），“协变量自适应随机化下的Lasso调整的处理效应估计”，《生物计量学》110(2)，431–447。

Lu, X.、Yang, F. 与 Wang, Y.（2025年），“具有中等高维协变量的完全随机化实验中的去偏回归调整”，*The Annals ofStatistics*53(4)，1535–1558。

Ma, W.、Tu, F. 与 Liu, H.（2022），“协变量自适应随机化的回归分析：稳健且有效的推断视角”，医学统计学41(29)，5645–5661。

Ma, W.、Wu, Z. 与 Zhang, Z.（2026年），“《在随机化实验中整合异质信息：统一校准框架》的补充材料”。

Newey, W. K. & Smith, R. J. (2004年), ‘广义矩估计和广义经验似然估计量的高阶性质’, *Econometrica*72(1), 219–255。

Pocock, S. J. & Simon, R. (1975), ‘对照临床试验中基于预后因素平衡的序贯处理分配’, 《生物统计学》 31(1), 103–115。

Qin, J. & Lawless, J. (1994年), ‘经验似然与广义估计方程’, *TheAnnals of Statistics*22(1), 300–325。

Qin, J. & Zhang, B. (2007年), ‘基于经验似然的响应缺失问题推断及其在观察性研究中的应用’, 英国皇家统计学会杂志 *B*辑：统计方法学69(1), 101–122。

Rafi, A. (2023), ‘协变量自适应随机化下平均处理效应的有效半参数估计’. arXiv预印本 arXiv:2305.08340。

Robins, J. M., Rotnitzky, A. & Zhao, L. P. (1994年), ‘当

某些回归元并非总是可观测时回归系数的估计’，*Journal of the American Statistical Association*89(427), 846–866。

Stewart, G. W.（1977），《伪逆、投影和线性最小二乘问题的扰动》，*SIAM*评论 19(4)，634–661。

Tan, Z.（2014），《抽样和缺失数据问题中校准估计量的二阶渐近理论》，多元分析杂志131，240–253。

Taves, D. R.（1974年），《最小化法：一种将患者分配到治疗组与对照组的新方法》，临床药理学与治疗学 15(5)，443–453。

Tibshirani, R.（1996），《通过 Lasso 进行回归收缩与变量选择》，英国皇家统计学会杂志 *B*辑：统计方法学58(1)，267–288。

Tsiatis, A. A., Davidian, M., Zhang, M. & Lu, X.（2008年），《随机临床试验中两组治疗比较的协变量调整：一种有原则且灵活的方法》，医学统计学27(23)，4658–4677。

Tu, F., Ma, W. & Liu, H. (2024), ‘分层随机化下协变量调整的统一框架’，*Stat* 13(4),e70016。

Wager, S. & Athey, S. (2018), ‘使用随机森林估计和推断异质处理效应’，*Journal of the American Statistical Association* 113(523), 1228– 1242。

Xin, J., Liu, H. & Ma, W. (2024), ‘具有多个层的协变量自适应随机化下的推断’。arXiv 预印本 arXiv:2405.18856。

Ye, T., Shao, J., Yi, Y. & Zhao, Q. (2023), ‘改进随机临床试验分析中协变量调整实践’，*Journal of the American Statistical Association*118(544), 2370–2382。

Ye, T., Yi, Y. 和 Shao, J.（2022），《最小化及其他协变量自适应随机化方法下平均处理效应的推断》，《生物计量学》109(1), 33–47。

Zelen, M.（1974年）, 《患者进入临床试验的随机化与分层》, 《慢性病杂志》 27(7), 365–375。

Zhang, M., Tsiatis, A. A. & Davidian, M. (2008年), ‘利用辅助协变量提高随机临床试验中推断的效率’,《生物统计学》 64(3), 707–715.

《在随机化实验中整合异质信息：统一校准框架》的补充材料

Wei Ma、Zeqi Wu、Zheng Zhang 中国人民大学统计学与大数据研究院

本补充材料包含若干附录，其中提供了额外的模拟结果以及正文的证明。

附录A 若干有用的引理

引理 1. 假设*1*-*2* 成立，且 $\log(2K)\, n\inf_{1\le k\le K} p_{[k]} \to 0$ （当 ~~$n \to \infty$~~）。则

$$\left|\frac{n_{[k]}}{n} - p_{[k]}\right| = \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(B_i = k) - p_{[k]}\right| \le \sqrt{p_{[k]}} O_P\left(\sqrt{\frac{\log(2K)}{n}}\right) = p_{[k]} o_P(1)$$

and

$$\left|\frac{n}{n_{[k]}} - \frac{1}{p_{[k]}}\right| = \frac{1}{p_{[k]}} o_P(1),$$

其中随机变量 $O_P\left(\sqrt{\frac{\log(2K}{n}}\right)$ 和 ~~$o_P(1)$~~ 不依赖于 k。此外，我们还有

$$\left|\frac{1}{n}\sum_{i=1}^{n} A_i \mathbb{1}(B_i = k) - \pi_{[k]} p_{[k]}\right| \le p_{[k]} o_P(1),$$

其中随机变量 $o_P(1)$ 不依赖于 k。

证明。 根据 Chernoff 不等式（参见 Vershynin(2018，习题 2.3.5），我们有

$$P\left(\frac{\left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(B_i = k) - p_{[k]}\right|}{\sqrt{p_{[k]}}} \ge \delta\right) \le 2\exp\left(-c\delta^2 n\right)$$

对于任意 $0 < \delta < \sqrt{p_{[k]}}$。于是，根据并集界可得

$$P\left(\sup_{1\le k\le K}\frac{\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)-p_{[k]}\right|}{\sqrt{p_{[k]}}}\ge\delta\right)\le 2K\exp\left(-c\delta^2 n\right)$$

对于任意 $0 < \delta < \min_{1\le k\le K}\sqrt{p_{[k]}}$，由此可得

$$P\left(\sup_{1\le k\le K}\frac{\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)-p_{[k]}\right|}{\sqrt{p_{[k]}}}\ge\delta\sqrt{\frac{\log(2K)}{n}}\right)\le\exp\left(\log(2K)\left\{1-c\delta^2\right\}\right)$$

对于任意 $0 < \delta < \sqrt{\frac{n}{\log(2K)}}\min_{1\le k\le K}\sqrt{p_{[k]}}$。设 $0 < \epsilon < 2$ 为任意数。由于 $\frac{\log(2K)}{n\inf_{1\le k\le K}p_{[k]}} \to 0$，存在整数 N_0 使得

$$\sqrt{\frac{1-\log\left(\frac{\epsilon}{2}\right)}{c}} < \sqrt{\frac{n}{\log(2K)}}\min_{1\le k\le K}\sqrt{p_{[k]}}$$

对于所有 $n \ge N_0$。设 $\delta_\epsilon = \sqrt{\frac{1-\log\left(\frac{\epsilon}{2}\right)}{c}}$，则有 $1 - c\delta_\epsilon^2 = \log\left(\frac{\epsilon}{2}\right) < 0$ 且

$$P\left(\sup_{1\le k\le K}\frac{\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)-p_{[k]}\right|}{\sqrt{p_{[k]}}}\ge\delta_\epsilon\sqrt{\frac{\log(2K)}{n}}\right)$$
$$\le\exp\left(\log(2K)\left\{1-c\delta_\epsilon^2\right\}\right)\le\exp\left(\log(2)\left\{1-c\delta_\epsilon^2\right\}\right)=\epsilon.$$

这表明

$$\sup_{1\le k\le K}\frac{\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)-p_{[k]}\right|}{\sqrt{p_{[k]}}}=O_P\left(\sqrt{\frac{\log(2K)}{n}}\right)$$

因此

$$\left|\frac{n_{[k]}}{n}-p_{[k]}\right|=\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)-p_{[k]}\right|\le\sqrt{p_{[k]}}O_P\left(\sqrt{\frac{\log(2K)}{n}}\right),$$

其中随机变量 $O_P\left(\sqrt{\frac{\log(2K}{n}}\right)$ 不依赖于 k。注意，

$$\begin{aligned}&\frac{1}{\sqrt{p_{[k]}}}O_P\left(\sqrt{\frac{\log(2K)}{n}}\right)\leq\frac{1}{\min_{1\leq k\leq K}\sqrt{p_{[k]}}}O_P\left(\sqrt{\frac{\log(2K)}{n}}\right)\\=&O_P\left(\sqrt{\frac{\log(2K)}{\min_{1\leq k\leq K}\sqrt{p_{[k]}}n}}\right)=o_P(1).\end{aligned}$$

我们还有

$$\sqrt{p_{[k]}}O_P\left(\sqrt{\frac{\log(2K)}{n}}\right)=p_{[k]}\frac{1}{\sqrt{p_{[k]}}}O_P\left(\sqrt{\frac{\log(2K)}{n}}\right)=p_{[k]}o_P(1).$$

此外，我们有

$$\left|\frac{n}{n_{[k]}}-\frac{1}{p_{[k]}}\right|=\frac{1}{p_{[k]}}\left|\frac{\frac{n_{[k]}}{n}-p_{[k]}}{\frac{n_{[k]}}{n}}\right|=\frac{1}{p_{[k]}}\left|\frac{p_{[k]}o_P(1)}{p_{[k]}+p_{[k]}o_P(1)}\right|=\frac{1}{p_{[k]}}\left|\frac{o_P(1)}{1+o_P(1)}\right|=\frac{1}{p_{[k]}}o_P(1)$$

其中，随机变量 $o_P(1)$ 与 k 无关，第二步则由.....得出

$\left|\frac{n_{[k]}}{n}-p_{[k]}\right|=p_{[k]}o_P(1).$

最后一个结论由.....得出

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)-\pi_{[k]}p_{[k]}\right|=\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\frac{\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)}{\sum_{i=1}^{n}\mathbb{1}(B_i=k)}-\pi_{[k]}p_{[k]}\right|\\\leq&\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)-p_{[k]}\right|\frac{\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)}{\sum_{i=1}^{n}\mathbb{1}(B_i=k)}+\left|\frac{\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)}{\sum_{i=1}^{n}\mathbb{1}(B_i=k)}-\pi_{[k]}\right|p_{[k]}\\\leq&p_{[k]}o_P(1),\end{aligned}$$

其中最后一步由 $\left|\frac{n_{[k]}}{n}-p_{[k]}\right|\leq p_{[k]}o_P(1)$ 和假设 2 . ■

引理 2. 假设 *1-2* 和 *4* 成立。那么对于每个 $k=1,\ldots,K$，均有 ***(i)***

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right]\right\|\leq\sqrt{p_{[k]}}O_P\left(\sqrt{\frac{r_n\log(2Kd)}{n}}\right)$$

其中，随机变量 $O_P\left(\sqrt{\frac{r_n \log(2Kd)}{n}}\right)$ 不依赖于 k；*(ii)*

$$\left\|\frac{1}{n}\sum_{i=1}^{n} A_i \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\| \leq \sqrt{p_{[k]}} O_P\left(\sqrt{\frac{r_n \log(2Kd)}{n}}\right)$$

其中，随机变量 $O_P\left(\sqrt{\frac{r_n \log(2Kd)}{n}}\right)$ 不依赖于 k；*(iii)*

$$\left\|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top - \mathbb{E}\left[\mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right]\right\| \leq p_{[k]} o_P(1),$$

其中，随机变量 $o_P(1)$ 不依赖于 k；并且 *(iv)*

$$\left\|\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]}^* \boldsymbol{\Xi}_{i,[k]}^{*\top} - \boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right\| \leq p_{[k]} o_P(1),$$

其中，随机变量 $o_P(1)$ 不依赖于 k；

证明。**(i)**令

$$\boldsymbol{Z}_i = \mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i) - \mathbb{E}\left[\mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right]$$

and

$$\boldsymbol{S}_n = \sum_{i=1}^{n} \boldsymbol{Z}_i.$$

根据假设 4，有

$$\|\boldsymbol{Z}_i\| \leq 2\zeta_n$$

对所有 $1 \leq i \leq n$ 和

$$\begin{aligned} V &:= \max\left\{\left\|\mathbb{E}\left[\boldsymbol{S}_n \boldsymbol{S}_n^\top\right]\right\|, \left\|\mathbb{E}\left[\boldsymbol{S}_n^\top \boldsymbol{S}_n\right]\right\|\right\} = \max\left\{\left\|\sum_{i=1}^{n}\mathbb{E}\left[\boldsymbol{Z}_i \boldsymbol{Z}_i^\top\right]\right\|, \left\|\sum_{i=1}^{n}\mathbb{E}\left[\boldsymbol{Z}_i^\top \boldsymbol{Z}_i\right]\right\|\right\} \\ &\leq \sum_{i=1}^{n}\mathbb{E}\left[\|\boldsymbol{Z}_i\|^2\right] \leq \sum_{i=1}^{n}\mathbb{E}\left[\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^2\right] = np_{[k]}\mathrm{tr}\left\{\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right]\right\} \\ &\leq Cnp_{[k]} r_n. \end{aligned}$$

由矩阵 Bernstein 不等式（参见 Tropp （2015，定理 1.6.2））可得

$$P\left(\|\boldsymbol{S}_n\| \geq t\right) \leq (d+1)\exp\left(-\frac{t^2/2}{V+2\zeta_n t/3}\right) \leq (d+1)\exp\left(-\frac{t^2/2}{Cnp_{[k]}r_n+2\zeta_n t/3}\right)$$

对所有 $t \geq 0$。因此，

$$\begin{aligned}
&P\left(\frac{\left\|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right]\right\|}{\sqrt{p_{[k]}}} \geq t\right) \\
\leq&(d+1)\exp\left(-\frac{np_{[k]}t^2/2}{Cp_{[k]}r_n+2\zeta_n\sqrt{p_{[k]}}t/3}\right) \\
\leq&(d+1)\exp\left(-\frac{nt^2}{4Cr_n}\right)
\end{aligned}$$

对所有 $0 \leq t \leq 3Cr_n\sqrt{p_{[k]}}/(2\zeta_n)$。根据并集界，我们有

$$\begin{aligned}
&P\left(\sup_{1\leq k\leq K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right]\right\|}{\sqrt{p_{[k]}}} \geq \delta\sqrt{\frac{r_n\log(K(d+1))}{n}}\right) \\
\leq&K(d+1)\exp\left(-\frac{\delta^2}{4C}\log(K(d+1))\right) = \exp\left(\log(K(d+1))\left\{1-\frac{\delta^2}{4C}\right\}\right)
\end{aligned}$$

对于任意

$$0 \leq \delta \leq \sqrt{\frac{n}{r_n\log(K(d+1))}}\frac{3Cr_n\min_{1\leq k\leq K}\sqrt{p_{[k]}}}{2\zeta_n} = \frac{3C}{2}\sqrt{\frac{nr_n\inf_{1\leq k\leq K}p_{[k]}}{\zeta_n^2\log(K(d+1))}}.$$

设 $0<\epsilon<2$ 为任意数。由于

$$\frac{3C}{2}\sqrt{\frac{nr_n\inf_{1\leq k\leq K}p_{[k]}}{\zeta_n^2\log(K(d+1))}} \geq \frac{3C}{2}\sqrt{\frac{n\inf_{1\leq k\leq K}p_{[k]}}{\zeta_n^2\log(2Kd)}} \to \infty,$$

存在一个整数 N_0 使得

$$\sqrt{4C\left\{1-\log\left(\frac{\epsilon}{2}\right)\right\}} < \frac{3C}{2}\sqrt{\frac{nr_n\inf_{1\leq k\leq K}p_{[k]}}{\zeta_n^2\log(K(d+1))}}$$

对于所有 $n \geq N_0$。令 $\delta_\epsilon = \sqrt{4C\left\{1 - \log\left(\frac{\epsilon}{2}\right)\right\}}$，则有 $1 - \frac{\delta_\epsilon^2}{4C} = \log\left(\frac{\epsilon}{2}\right) < 0$ 且

$$P\left(\sup_{1\leq k\leq K} \frac{\left\|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i) - \mathbb{E}\left[\mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right]\right\|}{\sqrt{p_{[k]}}} \geq \delta_\epsilon\sqrt{\frac{r_n \log(K(d+1))}{n}}\right)$$
$$\leq \exp\left(\log(K(d+1))\left\{1 - \frac{\delta_\epsilon^2}{4C}\right\}\right) \leq \exp\left(\log(2)\log\left(\frac{\epsilon}{2}\right)\right) = \epsilon.$$

这意味着

$$\sup_{1\leq k\leq K} \frac{\left\|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i) - \mathbb{E}\left[\mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right]\right\|}{\sqrt{p_{[k]}}} = O_P\left(\sqrt{\frac{r_n \log(K(d+1))}{n}}\right)$$

因此

$$\left\|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i) - \mathbb{E}\left[\mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right]\right\| \leq \sqrt{p_{[k]}} O_P\left(\sqrt{\frac{r_n \log(2Kd)}{n}}\right),$$

其中，随机变量 $O_P\left(\sqrt{\frac{r_n \log(2Kd)}{n}}\right)$ 不依赖于 k。

(ii) 令 $\mathcal{C}_n$ 表示由 $(A^{(n)}, B^{(n)})$ 生成的 σ-代数，并令 $\mathbb{E}_{\mathcal{C}_n}[\cdot] := \mathbb{E}[\cdot \mid \mathcal{C}_n]$, $P_{\mathcal{C}_n}(\cdot) := P(\cdot \mid \mathcal{C}_n)$。那么，$\boldsymbol{\Xi}_{i,[k]}^* \boldsymbol{\Xi}_{i}^{*\top}{}_{,[k]}$, $i = 1, \ldots, n$ 在 $\mathcal{C}_n$ 条件下相互独立。现在，我们固定 $(A^{(n)}, B^{(n)})$ 的任意一个取值。令

$$\boldsymbol{Z}_i = A_i \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) - \mathbb{E}_{\mathcal{C}_n}\left[A_i \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right] = A_i \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)$$

and

$$\boldsymbol{S}_n = \sum_{i=1}^n \boldsymbol{Z}_i.$$

则由假设4可知

$$\|\boldsymbol{Z}_i\| \leq 2\zeta_n$$

对所有 $1 \le i \le n$ 和

$$
\begin{aligned}
V &:= \max\left\{\left\|\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{S}_n\boldsymbol{S}_n^\top\right]\right\|, \left\|\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{S}_n^\top\boldsymbol{S}_n\right]\right\|\right\} = \max\left\{\left\|\sum_{i=1}^n \mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{Z}_i\boldsymbol{Z}_i^\top\right]\right\|, \left\|\sum_{i=1}^n \mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{Z}_i^\top\boldsymbol{Z}_i\right]\right\|\right\} \\
&\le \sum_{i=1}^n \mathbb{E}_{\mathcal{C}_n}\left[\|\boldsymbol{Z}_i\|^2\right] = \sum_{i=1}^n \mathbb{E}_{\mathcal{C}_n}\left[A_i^2\mathbb{1}(B_i=k)\left\|\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|^2\right] \le \sum_{i=1}^n \mathbb{1}(B_i=k)\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|^2 \mid B_i=k\right] \\
&\le n_{[k]}\mathbb{E}\left[\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\|^2 \mid B_i=k\right] = n_{[k]}\mathrm{tr}\left\{\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i=k\right]\right\} \le C r_n n_{[k]}
\end{aligned}
$$

几乎必然。由矩阵 Bernstein 不等式（参见Tropp （2015，定理 1.6.2））

that

$$
P_{\mathcal{C}_n}\left(\|\boldsymbol{S}_n\| \ge t\right) \le 2d\exp\left(-\frac{t^2/2}{V + 2\zeta_n t/3}\right) \le 2d\exp\left(-\frac{t^2/2}{Cr_n n_{[k]} + 2\zeta_n t/3}\right)
$$

对所有 $t \ge 0$ 几乎必然成立。因此，

$$
\begin{aligned}
&P_{\mathcal{C}_n}\left(\frac{\left\|\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|}{\sqrt{n_{[k]}/n}} \ge t\right) \\
\le& 2d\exp\left(-\frac{nt^2/2}{Cr_n + 2(\sqrt{n_{[k]}/n})^{-1}\zeta_n t/3}\right) \le 2d\exp\left(-\frac{nt^2}{4Cr_n}\right)
\end{aligned}
$$

对所有几乎必然成立

$$
0 \le t \le \frac{3}{2}C\frac{r_n}{\zeta_n}\sqrt{\frac{n_{[k]}}{n}}.
$$

根据并集界，我们有

$$
\begin{aligned}
&P_{\mathcal{C}_n}\left(\sup_{1\le k\le K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|}{\sqrt{n_{[k]}/n}} \ge \delta\sqrt{\frac{r_n\log(2Kd)}{n}}\right) \\
\le& 2Kd\exp\left(-\frac{\delta^2}{4C}\log(2Kd)\right) \le \exp\left(\log(2Kd)\left\{1 - \frac{\delta^2}{4C}\right\}\right)
\end{aligned}
$$

对所有几乎必然成立

$$
0 \le \delta \le \frac{3}{2}C\sqrt{\frac{nr_n\min_{1\le k\le K}(n_{[k]}/n)}{\zeta_n^2\log(2Kd)}}.
$$

设 $0 < \epsilon < 2$ 为任意数。根据引理 1，我们有 $\frac{n_{[k]}}{np_{[k]}} = 1 + o_P(1)$，其中随机变量 $o_P(1)$ 不依赖于 k。因此，事件 $E_n := \left\{1/4 \le \text{下确界}_{1\le k\le K}\frac{n_{[k]}}{np_{[k]}}\right\}$

以概率 $1-\epsilon$ 对所有 $n \geq N_1$发生，其中 N_1 是有限整数。那么在事件 E_n 上，我们有

$$\begin{aligned}
&\frac{3}{2}C\sqrt{\frac{nr_n \min_{1\leq k\leq K}(n_{[k]}/n)}{\zeta_n^2 \log(2Kd)}} \geq \frac{3}{2}C\sqrt{\frac{n \min_{1\leq k\leq K}\left\{(n_{[k]}/np_{[k]})p_{[k]}\right\}}{\zeta_n^2 \log(2Kd)}} \\
\geq &\frac{3}{2}C\sqrt{\frac{n \inf_{1\leq k\leq K} p_{[k]}}{\zeta_n^2 \log(2Kd)}}\sqrt{\min_{1\leq k\leq K}(n_{[k]}/np_{[k]})} \geq \frac{3}{4}C\sqrt{\frac{n \inf_{1\leq k\leq K} p_{[k]}}{\zeta_n^2 \log(2Kd)}}.
\end{aligned}$$

根据假设 4，我们有 $\sqrt{n \text{下确界} \frac{\zeta_n^2}{2Kd} \log()} \to \infty$，因此存在一个整数 N_0，使得在事件 E_n 上

$$\sqrt{4C\left\{1-\log\left(\frac{\epsilon}{2}\right)\right\}} \leq \frac{3}{4}C\sqrt{\frac{n \inf_{1\leq k\leq K} p_{[k]}}{\zeta_n^2 \log(2Kd)}} \leq \frac{3}{2}C\sqrt{\frac{nr_n \min_{1\leq k\leq K}(n_{[k]}/n)}{\zeta_n^2 \log(2Kd)}}$$

对所有 $n \geq N_0$。令 $\delta_\epsilon = \sqrt{4C\left\{1-\log\left(\frac{\epsilon}{2}\right)\right\}}$，则我们有 $1-\frac{\delta_\epsilon^2}{4C} = \log\left(\frac{\epsilon}{2}\right) < 0$，并且

$$\begin{aligned}
&P_{\mathcal{C}_n}\left(\sup_{1\leq k\leq K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n A_i \mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|}{\sqrt{n_{[k]}/n}} \geq \delta_\epsilon\sqrt{\frac{r_n \log(2Kd)}{n}}\right) \\
\leq &\exp\left(\log(2Kd)\left\{1-\frac{\delta_\epsilon^2}{4C}\right\}\right) \leq \exp\left(\log(2)\log\left(\frac{\epsilon}{2}\right)\right) = \epsilon
\end{aligned}$$

对于 $(A^{(n)}, B^{(n)})$ 的任意固定值，使得 E_n 发生[1]。然后，对于所有 $n \geq \max\{N_0, N_1\}$，我们有

$$\begin{aligned}
&P\left(\sup_{1\leq k\leq K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n A_i \mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|}{\sqrt{n_{[k]}/n}} \geq \delta_\epsilon\sqrt{\frac{r_n \log(2Kd)}{n}}\right) \\
= &\mathbb{E}\left[P_{\mathcal{C}_n}\left(\sup_{1\leq k\leq K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_{i,[k]}^* \boldsymbol{\Xi}_{i,[k]}^{*\top} - \boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right\|}{\sqrt{n_{[k]}/n}} \geq \delta_\epsilon\sqrt{\frac{r_n \log(2Kd)}{n}}\right)\right] \\
\leq &\mathbb{E}\left[\epsilon \mathbb{1}(E_n \text{ happens})\right] + \mathbb{E}\left[\mathbb{1}(E_n \text{does not happen})\right] \leq \epsilon + \epsilon = 2\epsilon.
\end{aligned}$$

[1]这是可行的，因为 E_n 关于 $\mathcal{C}_n$ 可测。

这意味着

$$\sup_{1\le k\le K}\frac{\left\|\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|}{\sqrt{n_{[k]}/n}}=O_P\left(\sqrt{\frac{r_n\log(2Kd)}{n}}\right),$$

其中随机变量 $O_P\left(\sqrt{\frac{r_n\log(2Kd}{n}}\right)$ 不依赖于 k。由引理 1，我们有

$$\sqrt{n_{[k]}/n}=\sqrt{p_{[k]}}O_P(1),$$

其中随机变量 $O_P(1)$ 不依赖于 k。因此,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|\le\sqrt{n_{[k]}/n}O_P\left(\sqrt{\frac{r_n\log(2Kd)}{n}}\right)=\sqrt{p_{[k]}}O_P\left(\sqrt{\frac{r_n\log(2Kd)}{n}}\right),$$

其中随机变量 $O_P\left(\sqrt{\frac{r_n\log(2Kd}{n}}\right)$ 不依赖于 k。

(iii) 设

$$\boldsymbol{Z}_i=\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top-\mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right]$$

and

$$\boldsymbol{S}_n=\sum_{i=1}^{n}\boldsymbol{Z}_i.$$

则由假设 4可知

$$\|\boldsymbol{Z}_i\|\le 2\zeta_n^2$$

对所有 $1\le i\le n$ 以及

$$\begin{aligned}V&:=\max\left\{\left\|\mathbb{E}\left[\boldsymbol{S}_n\boldsymbol{S}_n^\top\right]\right\|,\left\|\mathbb{E}\left[\boldsymbol{S}_n^\top\boldsymbol{S}_n\right]\right\|\right\}\\&=\max\left\{\left\|\sum_{i=1}^{n}\mathbb{E}\left[\boldsymbol{Z}_i\boldsymbol{Z}_i^\top\right]\right\|,\left\|\sum_{i=1}^{n}\mathbb{E}\left[\boldsymbol{Z}_i^\top\boldsymbol{Z}_i\right]\right\|\right\}=\left\|\sum_{i=1}^{n}\mathbb{E}\left[\boldsymbol{Z}_i^2\right]\right\|\\&=\left\|\sum_{i=1}^{n}\left\{\mathbb{E}\left[\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right\}^2\right]-\left(\mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right]\right)^2\right\}\right\|\end{aligned}$$

$$
\begin{aligned}
&\leq n\left\|\mathbb{E}\left[\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right\}^2\right]\right\| + n\left\|\mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right]\right\|^2 \\
&\leq 2n\left\|\mathbb{E}\left[\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right\}^2\right]\right\| \\
&= 2np_{[k]}\left\|\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i=k\right]\right\| \\
&\leq 2np_{[k]}\zeta_n^2\left\|\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i=k\right]\right\| \leq 2Cp_{[k]}n\zeta_n^2.
\end{aligned}
$$

由矩阵 Bernstein 不等式可知（参见Tropp （2015，定理 1.6.2））：

$$
P\left(\|\boldsymbol{S}_n\| \geq t\right) \leq 2d\exp\left(-\frac{t^2/2}{V+2\zeta_n^2 t/3}\right) \leq 2d\exp\left(-\frac{t^2/2}{2Cp_{[k]}n\zeta_n^2+2\zeta_n^2 t/3}\right)
$$

对所有 $t \geq 0$。因此，

$$
\begin{aligned}
&P\left(\frac{\left\|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top - \mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right]\right\|}{\sqrt{p_{[k]}}} \geq t\right) \\
\leq& 2d\exp\left(-\frac{nt^2/2}{2C\zeta_n^2+2\sqrt{p_{[k]}^{-1}}\zeta_n^2 t/3}\right) \leq 2d\exp\left(-\frac{nt^2}{8C\zeta_n^2}\right).
\end{aligned}
$$

对所有 $0 \leq t \leq 3C\sqrt{p_{[k]}}.$，由并集界可得

$$
\begin{aligned}
&P\left(\sup_{1\leq k\leq K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top - \mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right]\right\|}{\sqrt{p_{[k]}}} \geq \delta\sqrt{\frac{\zeta_n^2\log(2Kd)}{n}}\right) \\
\leq& 2Kd\exp\left(-\frac{\delta^2}{8C}\log(2Kd)\right) \leq \exp\left(\log(2Kd)\left\{1-\frac{\delta^2}{8C}\right\}\right)
\end{aligned}
$$

对所有

$$
0 \leq \delta \leq 3C\sqrt{\frac{n\inf_{1\leq k\leq K} p_{[k]}}{\zeta_n^2\log(2Kd)}}.
$$

设 $0<\epsilon<2$ 为任意实数。由于

$$
3C\sqrt{\frac{n\inf_{1\leq k\leq K} p_{[k]}}{\zeta_n^2\log(2Kd)}} \to \infty,
$$

存在一个整数 N_0 使得

$$\sqrt{8C\left\{1-\log\left(\frac{\epsilon}{2}\right)\right\}} < 3C\sqrt{\frac{n\inf_{1\le k\le K} p_{[k]}}{\zeta_n^2\log(2Kd)}}$$

对所有 $n \ge N_0$。令 $\delta_\epsilon = \sqrt{8C\left\{1-\log\left(\frac{\epsilon}{2}\right)\right\}}$，则有 $1-\frac{\delta_\epsilon^2}{8C} = \log\left(\frac{\epsilon}{2}\right) < 0$ 以及

$$P\left(\sup_{1\le k\le K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top - \mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right]\right\|}{\sqrt{p_{[k]}}} \ge \delta\sqrt{\frac{\zeta_n^2\log(2Kd)}{n}}\right)$$
$$\le \exp\left(\log(2Kd)\left\{1-\frac{\delta_\epsilon^2}{8C}\right\}\right) \le \exp\left(\log(2)\log\left(\frac{\epsilon}{2}\right)\right) = \epsilon.$$

这意味着

$$\sup_{1\le k\le K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top - \mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right]\right\|}{\sqrt{p_{[k]}}}$$
$$= \sqrt{\frac{\zeta_n^2\log(2Kd)}{n}}O_P(1)$$

从而

$$\left\|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top - \mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right]\right\|$$
$$\le p_{[k]}\sqrt{\frac{\zeta_n^2\log(2Kd)}{n\inf_{1\le k\le K} p_{[k]}}}O_P(1) = p_{[k]}o_P(1),$$

其中，随机变量 $o_P(1)$ 不依赖于 k，且不等式由下式得出

$$\zeta_n^2\log(2Kd)/(n\inf_{1\le k\le K} p_{[k]}) \to 0$$

当 $n\to\infty$ 时。

(iv) 令$\mathcal{C}_n$ 表示由（$A^{(n)}$, $B^{(n)}$）生成的 σ代数，以及 $\mathbb{E}_{\mathcal{C}_n}[\cdot] := \mathbb{E}[\cdot \mid \mathcal{C}_n]$, $P_{\mathcal{C}_n}(\cdot) := P(\cdot \mid \mathcal{C}_n)$。则，在$\mathcal{C}_n$条件下，$\boldsymbol{\Xi}^*_{i,\,[k]}\boldsymbol{\Xi}^{*\top}_{i,\,[k]}$, $i=1$, ..., n相互独立。现在，我们固定

（$A^{(n)}$, $B^{(n)}$）的任意一个值。令

$$\boldsymbol{Z}_i = \boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]} - \mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\right]$$

and

$$\boldsymbol{S}_n = \sum_{i=1}^{n} \boldsymbol{Z}_i.$$

由假设 4可知

$$\|\boldsymbol{Z}_i\| \le 4\zeta_n^2$$

对所有 $1 \le i \le n$ 和

$$\begin{aligned}
V =& \max\left\{\left\|\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{S}_n\boldsymbol{S}_n^\top\right]\right\|, \left\|\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{S}_n^\top\boldsymbol{S}_n\right]\right\|\right\} = \left\|\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{Z}_i^2\right]\right\| \\
=& \Bigg\|\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\mathbb{1}(B_i=k)(A_i-\pi_{n[k]})^2\left\{\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\right\}^2\right] \\
&- \sum_{i=1}^{n}\left\{\mathbb{E}_{\mathcal{C}_n}\left[(A_i-\pi_{n[k]})\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\right]\right\}^2\Bigg\| \\
\le& \sum_{i=1}^{n}\left\|\mathbb{E}_{\mathcal{C}_n}\left[\mathbb{1}(B_i=k)(A_i-\pi_{n[k]})^2\left\{\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\right\}^2\right]\right\| \\
&+ \sum_{i=1}^{n}\left\|\mathbb{E}_{\mathcal{C}_n}\left[(A_i-\pi_{n[k]})\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\right]\right\|^2 \\
\le& 2\sum_{i=1}^{n}\left\|\mathbb{E}_{\mathcal{C}_n}\left[\mathbb{1}(B_i=k)(A_i-\pi_{n[k]})^2\left\{\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\right\}^2\right]\right\| \\
\le& 2\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\mathbb{E}\left[\left\{\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\right\}^2 \mid B_i=k\right]\right\| \\
\le& 8\zeta_n^2\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top \mid B_i=k\right]\right\| \\
\le& 8\zeta_n^2\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\mathbb{E}\left[\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)^\top \mid B_i=k\right]\right\| \\
\le& 8C\zeta_n^2\sum_{i=1}^{n}\mathbb{1}(B_i=k) = 8C\zeta_n^2 n_{[k]}
\end{aligned}$$

几乎必然。由此可得 m 矩阵 Bernstein 不等式（参见 Tropp (2015，定理 1.6.2））

that

$$P_{\mathcal{C}_n}\left(\|\boldsymbol{S}_n\| \geq t\right) \leq 2d\exp\left(-\frac{t^2/2}{V+4\zeta_n^2 t/3}\right) \leq 2d\exp\left(-\frac{t^2/2}{8C\zeta_n^2 n_{[k]}+4\zeta_n^2 t/3}\right)$$

对所有 $t \geq 0$，几乎必然成立。因此，

$$P_{\mathcal{C}_n}\left(\frac{\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_{i,[k]}^* \boldsymbol{\Xi}_{i,[k]}^{*\top} - \boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right\|}{\sqrt{n_{[k]}/n}} \geq t\right) = P_{\mathcal{C}_n}\left(\frac{\left\|\frac{1}{n}\sum_{i=1}^n \left\{\boldsymbol{\Xi}_{i,[k]}^* \boldsymbol{\Xi}_{i,[k]}^{*\top} - \mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^* \boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\right\}\right\|}{\sqrt{n_{[k]}/n}} \geq t\right)$$

$$\leq 2d\exp\left(-\frac{nt^2/2}{8C\zeta_n^2+4(\sqrt{n_{[k]}/n})^{-1}\zeta_n^2 t/3}\right) \leq 2d\exp\left(-\frac{nt^2}{32C\zeta_n^2}\right)$$

对所有情况几乎必然成立

$$0 \leq t \leq 6C\sqrt{n_{[k]}/n}.$$

由并集界可得

$$P_{\mathcal{C}_n}\left(\sup_{1\leq k\leq K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_{i,[k]}^* \boldsymbol{\Xi}_{i,[k]}^{*\top} - \boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right\|}{\sqrt{n_{[k]}/n}} \geq \delta\sqrt{\frac{\zeta_n^2\log(2Kd)}{n}}\right)$$

$$\leq 2Kd\exp\left(-\frac{\delta^2}{32C}\log(2Kd)\right) \leq \exp\left(\log(2Kd)\left\{1-\frac{\delta^2}{32C}\right\}\right)$$

对所有情况几乎必然成立

$$0 \leq \delta \leq 6C\sqrt{\frac{n\min_{1\leq k\leq K}(n_{[k]}/n)}{\zeta_n^2\log(2Kd)}}.$$

令 $0<\epsilon<2$ 为任意数。由引理 1 可得 $\frac{n_{[k]}}{np_{[k]}} = 1+o_P(1)$，其中随机变量 $o_P(1)$ 不依赖于 k。因此，事件 $E_n := \left\{1/4 \leq \inf_{1\leq k\leq K}\frac{n_{[k]}}{np_{[k]}}\right\}$ 以概率 $1-\epsilon$ 对所有 $n \geq N_1$ 都成立，其中 N_1 是一个有限整数。于是，在事件 E_n 上，有

$$6C\sqrt{\frac{n\min_{1\leq k\leq K}(n_{[k]}/n)}{\zeta_n^2\log(2Kd)}}$$

$$=6C\sqrt{\frac{n\min_{1\leq k\leq K}\left\{(n_{[k]}/np_{[k]})p_{[k]}\right\}}{\zeta_n^2\log(2Kd)}} \geq 6C\sqrt{\frac{n\inf_{1\leq k\leq K}p_{[k]}}{\zeta_n^2\log(2Kd)}}\sqrt{\min_{1\leq k\leq K}(n_{[k]}/np_{[k]})}$$

$$\geq 3C\sqrt{\frac{n\inf_{1\leq k\leq K} p_{[k]}}{\zeta_n^2\log(2Kd)}}.$$

根据假设 4，有 $\sqrt{n\inf\zeta_n^2}^{1\leq k\leq K\, p_{[k]}}_{2Kd\ \log()\to\infty}$，因此存在一个整数 N_0，使得在事件 E_n上

$$\sqrt{32C\left\{1-\log\left(\frac{\epsilon}{2}\right)\right\}} \leq 3C\sqrt{\frac{n\inf_{1\leq k\leq K} p_{[k]}}{\zeta_n^2\log(2Kd)}} \leq 6C\sqrt{\frac{n\min_{1\leq k\leq K}(n_{[k]}/n)}{\zeta_n^2\log(2Kd)}}$$

对所有 $n\geq N_0$。令 $\delta_\epsilon=\sqrt{32C\left\{1-\log\left(\frac{\epsilon}{2}\right)\right\}}$，则我们有 $1-\frac{\delta_\epsilon^2}{32C}=\log\left(\frac{\epsilon}{2}\right)<0$且

$$\begin{aligned}
&P_{\mathcal{C}_n}\left(\sup_{1\leq k\leq K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}-\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]}\right\|}{\sqrt{n_{[k]}/n}}\geq\delta_\epsilon\sqrt{\frac{\zeta_n^2\log(2Kd)}{n}}\right)\\
\leq&\exp\left(\log(2Kd)\left\{1-\frac{\delta_\epsilon^2}{32C}\right\}\right)\leq\exp\left(\log(2)\log\left(\frac{\epsilon}{2}\right)\right)=\epsilon
\end{aligned}$$

对于满足 E_n 发生[2] 的任意固定值 $(A^{(n)},\ B^{(n)})$。那么，对于所有 $n\geq\max\{N_0,\ N_1\}$，我们有

$$\begin{aligned}
&P\left(\sup_{1\leq k\leq K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}-\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]}\right\|}{\sqrt{n_{[k]}/n}}\geq\delta_\epsilon\sqrt{\frac{\zeta_n^2\log(2Kd)}{n}}\right)\\
=&\mathbb{E}\left[P_{\mathcal{C}_n}\left(\sup_{1\leq k\leq K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}-\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]}\right\|}{\sqrt{n_{[k]}/n}}\geq\delta_\epsilon\sqrt{\frac{\zeta_n^2\log(2Kd)}{n}}\right)\right]\\
\leq&\mathbb{E}\left[\epsilon\mathbb{1}(E_n\text{ happens})\right]+\mathbb{E}\left[\mathbb{1}(E_n\text{ does not happen})\right]\leq\epsilon+\epsilon=2\epsilon.
\end{aligned}$$

这意味着

$$\sup_{1\leq k\leq K}\frac{\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}-\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]}\right\|}{\sqrt{n_{[k]}/n}}=O_P\left(\sqrt{\frac{\zeta_n^2\log(2Kd)}{n}}\right),$$

[2]这是可行的，因为 E_n 相对于$\mathcal{C}_n$可测。

其中，随机变量 $O_P\left(\sqrt{\frac{\zeta_n^2 \log(2Kd}{n}}\right)$ 不依赖于 k。根据引理1，我们有

$$\sqrt{n_{[k]}/n} = \sqrt{p_{[k]}} O_P(1),$$

其中，随机变量 $O_P(1)$ 不依赖于 k。因此，

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}^*_{i,[k]} \boldsymbol{\Xi}^{*\top}_{i,[k]} - \boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} \right\| \leq \sqrt{n_{[k]}/n} O_P\left(\sqrt{\frac{\zeta_n^2 \log(2Kd)}{n}} \right)$$
$$\leq p_{[k]} \sqrt{\frac{\zeta_n^2 \log(2Kd)}{n \inf_{1\leq k\leq K} p_{[k]}}} O_P(1) = p_{[k]} o_P(1),$$

其中，随机变量 $o_P(1)$ 不依赖于 k，最后一步由.....得出

$$\zeta_n^2 \log(2Kd)/(n \inf_{1\leq k\leq K} p_{[k]}) \to 0$$

因为 $n \to \infty$。证明完毕。 ■

引理3. 设 $(\Omega, F\ P)$ 是一个概率空间，且 $\mathcal{C}_n$ 是 F 的一列子-σ-代数。设 $\left\{ \boldsymbol{X}_{n,i} \in \mathbb{R}^d : 1 \leq i \leq n, n \geq 1 \right\}$ 是一个三角阵，使得对于每个 n，$\boldsymbol{X}_{n,1}, \ldots, \boldsymbol{X}_{n,n}$ 在 $\mathcal{C}_n$ 条件下相互独立。令$V_n := \sup_{n\geq 1, 1\leq i\leq n} \mathbb{E}\left[\|\boldsymbol{X}_{n,i}\|^q \mid \mathcal{C}_n \right]$，其中某个 $1 < q \leq 2$。若对于所有$1 \leq i \leq n$ 和 $n \geq 1$，$\mathbb{E}\left[\boldsymbol{X}_{n,i} \mid \mathcal{C}_n \right] = 0$，则有

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{n,i} \right\| = O_P(V_n^{1/q} n^{1/q-1}).$$

特别地，如果 $V_n = O_P(1)$，那么

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{n,i} \right\| = O_P(n^{1/q-1}) = o_P(1).$$

证明。 对于任意随机向量 $\boldsymbol{X}$，令 $\|\boldsymbol{X}\|_{L^q,\ \mathcal{C}_n}$ 表示其在$\mathcal{C}_n$条件下的 L^q-范数，即$\|\boldsymbol{X}\|_{L^q,\ \mathcal{C}_n} := \left(\mathbb{E}\left[\|\boldsymbol{X}\|^q \mid \mathcal{C}_n \right]\right)^{1/q}$，其中 $\|\cdot\|$ 表示欧几里得范数。根据定理 3.1

在Burkholder （1988）中，我们有

$$\left\|\sum_{i=1}^{n}\boldsymbol{X}_{n,i}\right\|_{L^q,\mathcal{C}_n}\leq\frac{1}{q-1}\left\|\left(\sum_{i=1}^{n}\|\boldsymbol{X}_{n,i}\|^2\right)^{1/2}\right\|_{L^q,\mathcal{C}_n}.$$

注意，对于任意非负实数 $a_1,\ldots,a_n$和 $r\geq q>1$，我们有$(\sum_{i=1}^n a_i^q)^{1/q}\geq(\sum_{i=1}^n a_i^r)^{1/r}$ (Bernstein Bernstein (2009)，第599页，命题 9.1.5）。因此，由上述两个不等式可得

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{n,i}\right\|_{L^q,\mathcal{C}_n}&\leq\frac{1}{n}\frac{1}{q-1}\left\|\left(\sum_{i=1}^{n}\|\boldsymbol{X}_{n,i}\|^2\right)^{1/2}\right\|_{L^q,\mathcal{C}_n}\leq\frac{1}{n}\frac{1}{q-1}\left\|\left(\sum_{i=1}^{n}\|\boldsymbol{X}_{n,i}\|^q\right)^{1/q}\right\|_{L^q,\mathcal{C}_n}\\&=\frac{1}{n}\frac{1}{q-1}\left(\mathbb{E}\left[\sum_{i=1}^{n}\|\boldsymbol{X}_{n,i}\|^q\mid\mathcal{C}_n\right]\right)^{1/q}=\frac{1}{n}\frac{1}{q-1}\left(\sum_{i=1}^{n}\|\boldsymbol{X}_{n,i}\|_{L^q,\mathcal{C}_n}^q\right)^{1/q}\\&\leq\frac{1}{n}\frac{1}{q-1}n^{1/q}\sup_{n\geq1,1\leq i\leq n}\left(\mathbb{E}\left[\|\boldsymbol{X}_{n,i}\|^q\mid\mathcal{C}_n\right]\right)^{1/q}=O_P(n^{1/q-1}V_n^{1/q}).\end{aligned}$$

根据条件马尔可夫不等式，对于所有 $\delta>0$，我们有

$$\begin{aligned}&P\left(\left\|\frac{1}{V_n^{1/q}n^{1/q-1}}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{n,i}\right\|>\delta\mid\mathcal{C}_n\right)\\\leq&\frac{\left\|\frac{1}{V_n^{1/q}n^{1/q-1}}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{n,i}\right\|_{L^q,\mathcal{C}_n}^q}{\delta^q}=\frac{\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{n,i}\right\|_{L^q,\mathcal{C}_n}^q}{\left\{V_n^{1/q}n^{1/q-1}\right\}^q\delta^q}=\frac{R_n}{\delta^q},\end{aligned}$$

其中，第二步是因为 V_n 关于$\mathcal{C}_n$ 可测；最后一步中，我们令 $R_n:=\left\{\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{n,i}\right\|_{L^q,\mathcal{C}_n}/(V_n^{1/q}n^{1/q-1})\right\}^q=O_P(1)$。对于任意 $\epsilon>0$，存在$N,M>1$ 使得

$$P\left(R_n\geq M\right)<\epsilon/2$$

对于所有的 $n>N$。令 $\delta:=(\frac{2M}{\epsilon})^{1/q}$，则有

$$\begin{aligned}&P\left(\left\|\frac{1}{n^{1/q-1}}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{n,i}\right\|>(\frac{2M}{\epsilon})^{1/q}\right)=\mathbb{E}\left[P\left(\left\|\frac{1}{n^{1/q-1}}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{n,i}\right\|>(\frac{2M}{\epsilon})^{1/q}\mid\mathcal{C}_n\right)\right]\\\leq&P\left(P\left(\left\|\frac{1}{n^{1/q-1}}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{n,i}\right\|>(\frac{2M}{\epsilon})^{1/q}\mid\mathcal{C}_n\right)>\frac{\epsilon}{2}\right)+\frac{\epsilon}{2}\end{aligned}$$

$$\leq P\left(\frac{R_n}{2M/\epsilon} > \frac{\epsilon}{2}\right) + \frac{\epsilon}{2} = P\left(R_n \geq M\right) + \frac{\epsilon}{2} < \epsilon$$

对于所有的 $n > N$。因此，$\frac{1}{V_n^{1/q} n^{1/q-1}} \frac{1}{n} \sum_{i=1}^n \boldsymbol{X}_{n,i} = O_P(1)$，所需结果得证。 ■

引理 4. 设 $(\Omega, \boldsymbol{F}\ P)$ 是一个概率空间，且 $\boldsymbol{C}_n$ 是 $\boldsymbol{F}$ 的一列子-σ-代数。设 $\{X_{n,i} \in \mathbb{R}: 1 \leq i \leq n, n \geq 1\}$ 是一个三角阵，使得对于每个 n, $X_{n,1}, \ldots, X_{n,n}$ 在 $\boldsymbol{C}_n$ 条件下相互独立。设 $S_n := \sum_{i=1}^n X_{n,i}$ 和 $\sigma_n^2(\boldsymbol{C}_n) := \mathrm{Var}(S_n \mid \boldsymbol{C}_n)$。我们假设

$$P\left(\{\omega : \mathrm{Var}(X_{n,i} \mid \mathcal{C}_n) < \infty,\ \forall 1 \leq i \leq n\} \bigcap \left\{\omega : \sigma_n^2(\mathcal{C}_n) > 0\right\}\right) \to 1 \tag{A.1}$$

当 $n \to \infty$ 时。如果林德伯格条件成立：对任意 $\delta > 0$

$$\frac{1}{\sigma_n^2(\mathcal{C}_n)} \sum_{i=1}^n \mathbb{E}\left[\{X_{n,i} - \mathbb{E}\left[X_{n,i} \mid \mathcal{C}_n\right]\}^2 \mathbb{1}\left\{|X_{n,i} - \mathbb{E}\left[X_{n,i} \mid \mathcal{C}_n\right]| \geq \delta \sigma_n(\mathcal{C}_n)\right\} \mid \mathcal{C}_n\right] = o_P(1)$$

当 $n \to \infty$ 时，则对每个 $t \in \mathbb{R}$，*(i)*

$$\mathbb{E}\left[\exp\left\{it\frac{S_n - \mathbb{E}\left[S_n \mid \mathcal{C}_n\right]}{\sigma_n(\mathcal{C}_n)}\right\} \mid \mathcal{C}_n\right] \xrightarrow{p} \exp\left(-t^2/2\right)$$

当 $n \to \infty$ 时，且 *(ii)*

$$\mathbb{E}\left[\left|\mathbb{E}\left[\exp\left\{it\frac{S_n - \mathbb{E}\left[S_n \mid \mathcal{C}_n\right]}{\sigma_n(\mathcal{C}_n)}\right\} \mid \mathcal{C}_n\right] - \exp\left(-t^2/2\right)\right|\right] \to 0$$

当 $n \to \infty$ 时。

证明。 为得到第一个结论，我们对 Bulinski (2017，定理1）的证明稍作修改。令

$$E_n := \{\omega : \mathrm{Var}(X_{n,i} \mid \mathcal{C}_n) < \infty,\ \forall 1 \leq i \leq n\} \bigcap \left\{\omega : \sigma_n^2(\mathcal{C}_n) > 0\right\},$$

那么 $E_n \in \mathcal{C}_n$ 和 $P(E_n) \to 1$ 作为 $n \to \infty$。我们的证明与Bulinski (2017，定理1)的证明唯一的区别在于，我们将注意力限制在 $\omega \in E_n$，这并

不会破坏所需的结果，因为 $P(E_n) \to 1$。我们给出证明中的几个关键步骤。

令

$$Z_{n,i}(\omega) := \begin{cases} \frac{X_{n,i} - \mathbb{E}[X_{n,i}|\mathcal{C}_n]}{\sigma_n(\mathcal{C}_n)}, & \omega \in E_n \\ 0, & \omega \in \Omega \backslash E_n \end{cases}, \quad i = 1, \dots n, \quad n \geq 1,$$

and

$$T_n(\omega) := \sum_{i=1}^{n} Z_{n,i}(\omega) = \begin{cases} \frac{S_n - \mathbb{E}[S_n|\mathcal{C}_n]}{\sigma_n(\mathcal{C}_n)}, & \omega \in E_n, \\ 0, & \omega \in \Omega \backslash E_n. \end{cases}$$

根据 $E_n \in \mathcal{C}_n$ 和引理 7， $\mathbb{E}[Z_{n,i} \mid \mathcal{C}_n] = 0$ 几乎必然，

$$\mathrm{Var}(Z_{n,i} \mid \mathcal{C}_n) = \mathbb{E}\left[Z_{n,i}^2 \mid \mathcal{C}_n\right] = \begin{cases} \frac{\mathrm{Var}(X_{n,i}|\mathcal{C}_n)}{\sigma_n^2(\mathcal{C}_n)}, & \omega \in E_n, \\ 0, & \omega \in \Omega \backslash E_n, \end{cases} \quad \text{a.s.}$$

and

$$\sum_{i=1}^{n} \mathrm{Var}(Z_{n,i} \mid \mathcal{C}_n) = \sum_{i=1}^{n} \mathbb{E}\left[Z_{n,i}^2 \mid \mathcal{C}_n\right] = \begin{cases} 1, & \omega \in E_n, \\ 0, & \omega \in \Omega \backslash E_n, \end{cases} \quad \text{a.s.}$$

对于 $\omega \in E_n$。现在，对任意 $\epsilon > 0$，我们定义

$$\begin{aligned} L_n^{\mathcal{C}_n}(\epsilon) &:= \sum_{i=1}^{n} \mathbb{E}\left[Z_{n,i}^2 \mathbb{1}(|Z_{n,i}| \geq \epsilon) \mid \mathcal{C}_n\right] \\ &= \begin{cases} \frac{1}{\sigma_n^2(\mathcal{C}_n)} \sum_{i=1}^{n} \mathbb{E}\left[\{X_{n,i} - \mathbb{E}[X_{n,i} \mid \mathcal{C}_n]\}^2 \mathbb{1}\{|X_{n,i} - \mathbb{E}[X_{n,i} \mid \mathcal{C}_n]| \geq \epsilon \sigma_n(\mathcal{C}_n)\} \mid \mathcal{C}_n\right], & \omega \in E_n, \\ 0, & \omega \in \Omega \backslash E_n. \end{cases} \end{aligned}$$

根据林德伯格条件和 $P(E_n) \to 1$，我们有

$$L_n^{\mathcal{C}_n}(\epsilon) \xrightarrow{p} 0$$

对任意 $\epsilon > 0$。对任意 $n \geq 1$，有

$$\max_{1 \leq i \leq n} \mathbb{E}\left[Z_{n,i}^2 \mid \mathcal{C}_n\right] = \max_{1 \leq i \leq n} \left\{\mathbb{E}\left[Z_{n,i}^2 \mathbb{1}(|Z_{n,i}| < \epsilon) \mid \mathcal{C}_n\right] + \mathbb{E}\left[Z_{n,i}^2 \mathbb{1}(|Z_{n,i}| \geq \epsilon) \mid \mathcal{C}_n\right]\right\}$$

$$\leq \epsilon^2 + L_n^{\mathcal{C}_n}(\epsilon) \xrightarrow{p} \epsilon^2.$$

由于 $\epsilon > 0$ 是任意的，我们有$\mathbb{E}\left[Z_{n,\ i}^2 \mid \mathcal{C}_n\right] \xrightarrow{p} 0$ 当 $n \to \infty$时。类似于Bulinski (2017，式(15))的推导，我们可以得到，对任意 $t \in \mathbb{R}$

$$\max_{1\leq i\leq n} \left|\varphi_{Z_{n,i}}^{\mathcal{C}_n}(t) - 1\right| \leq \frac{t^2}{2} \max_{1\leq i\leq n} \mathbb{E}\left[Z_{n,i}^2 \mid \mathcal{C}_n\right] \xrightarrow{p} 0,$$

其中 $\varphi_{Z_{n,\ i}}^{\mathcal{C}_n} := \mathbb{E}\left[\exp\left\{\mathrm{i}tZ_{n,\ i}\right\} \mid \mathcal{C}_n\right]$。现在，我们已证明 Bulinski (2017，式（13）).
引入事件 $D_n := \left\{\max_{1\leq i\leq n}\left|\varphi_{Z_{n,\ i}}^{\mathcal{C}_n}(t) - 1\right| \leq 1/2\right\} \bigcap E_n$， $n \geq 1$。那么 $P(D_n) \to 1$，因为 $n \to \infty$。此外，我们可以证明 Bulinski (2017，式（17）—（20）在我们的情形下成立。

Let

$$G_n^{\mathcal{C}_n}(t) := \begin{cases} \sum_{i=1}^n \left(\varphi_{Z_{n,i}}^{\mathcal{C}_n}(t) - 1\right) + \frac{t^2}{2}, & \omega \in D_n. \\ 0, & \omega \in \Omega \backslash D_n. \end{cases}$$

对于 $\omega \in D_n$，有 $\sum_{i=1}^n \mathbb{E}\left[Z_{n,\ i}^2 \mid \mathcal{C}_n\right] = 1$ 因此

$$\begin{aligned} G_n^{\mathcal{C}_n}(t) &= \sum_{i=1}^n \mathbb{E}\left[g_t(Z_{n,i}) \mid \mathcal{C}_n\right] \\ &= \sum_{i=1}^n \mathbb{E}\left[g_t(Z_{n,i})\mathbb{1}(|Z_{n,i}| < \epsilon) \mid \mathcal{C}_n\right] + \sum_{i=1}^n \mathbb{E}\left[g_t(Z_{n,i})\mathbb{1}(|Z_{n,i}| \geq \epsilon) \mid \mathcal{C}_n\right], \end{aligned}$$

其中 $g_t(x) := \exp\{\mathrm{i}tx\} - 1 - \mathrm{i}tx - (\mathrm{i}tx)^2/2$， $x \in \mathbb{R}$。应用 Bulinski (2017，式（14）），对于$\omega \in D_n$ 我们有

$$\left|G_n^{\mathcal{C}_n}(t)\right| \leq \frac{|t|^3 \epsilon}{6} + t^2 L_n^{\mathcal{C}_n}(\epsilon)$$

对于任意 $\epsilon > 0$。由于 $G_n^{\mathcal{C}_n}(t) = 0$ 对于 $\omega \in \Omega \backslash D_n$，我们有

$$\left|G_n^{\mathcal{C}_n}(t)\right| \leq \frac{|t|^3 \epsilon}{6} + t^2 L_n^{\mathcal{C}_n}(\epsilon), \quad \forall \omega \in \Omega.$$

根据 $L_n^{\mathcal{C}_n}(\epsilon) \xrightarrow{p} 0$，我们有 $G_n^{\mathcal{C}_n}(t) \xrightarrow{p} 0$，因为对于每个 $t \in \mathbb{R}$， $n \to \infty$。按照相同步骤

与Bulinski (2017年定理1的证明中一样，我们可以得到

$$\mathbb{E}\left[\exp\left\{\mathrm{i}tT_n\right\} \mid \mathcal{C}_n\right] \xrightarrow{p} \exp(-t^2/2)$$

如 $n \to \infty$。由引理 7 和 $E_n \in \mathcal{C}_n$，我们有

$$\mathbb{E}\left[\exp\left\{\mathrm{i}tT_n\right\} \mid \mathcal{C}_n\right] = \begin{cases} \mathbb{E}\left[\exp\left\{\mathrm{i}t\frac{S_n - \mathbb{E}[S_n|\mathcal{C}_n]}{\sigma_n(\mathcal{C}_n)}\right\} \mid \mathcal{C}_n\right], & \omega \in E_n \\ 1, & \omega \in \Omega \backslash E_n. \end{cases}$$

注意到 $P(E_n) \to 1$ 如 $n \to \infty$，第一个结论得证。
为了得到第二个结论，我们注意到 $\left|\mathbb{E}\left[\exp\left\{\mathrm{i}t\frac{S_n - \mathbb{E}[S_n|\mathcal{C}_{\sigma_n \quad n}]}{\sigma_n(\mathcal{C})}\right\} \mid \mathcal{C}_n\right] - \exp\left(-t^2/2\right)\right| \leq 2$ 有界。然后，结合 $\left|\mathbb{E}\left[\exp\left\{\mathrm{i}t\frac{S_n - \mathbb{E}[S_n|\mathcal{C}_{\sigma_n \quad n}]}{\sigma_n(\mathcal{C})}\right\} \mid \mathcal{C}_n\right] - \exp\left(-t^2/2\right)\right| = o_P(1)$ 和 Billingsley（1968，第 1 章定理 5.4），即可得到所需结果。■

引理 5. 设 (Ω, F, P) 为一个概率空间，且 $\mathcal{C}_n$ 为 F 的一列子-σ-代数。假设 X_n 是一列满足以下条件的随机变量

$$\mathbb{E}\left[\left|\mathbb{E}\left[\exp\left\{\mathrm{i}tX_n\right\} \mid \mathcal{C}_n\right] - \exp\left(-t^2/2\right)\right|\right] \to 0$$

对于任意 $t \in \mathbb{R}$，Y_n 是一列 $\mathcal{C}_n$-可测随机变量，满足 $n \to \infty$

$$\mathbb{E}\left[\exp\left\{\mathrm{i}sY_n\right\}\right] \to \exp\left(-s^2/2\right)$$

对于任意 $s \in \mathbb{R}$，均有 $n \to \infty$。然后

$$(X_n, Y_n)^\top \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

证明。 注意到我们有

$$
\begin{aligned}
&\left|\mathbb{E}\left[\exp\left\{\mathrm{i}tX_n+\mathrm{i}sY_n\right\}\right]-\exp\left(-t^2/2-s^2/2\right)\right| \\
=&\left|\mathbb{E}\left[\mathbb{E}\left[\exp\left\{\mathrm{i}tX_n+\mathrm{i}sY_n\right\}\mid\mathcal{C}_n\right]\right]-\exp\left(-t^2/2-s^2/2\right)\right| \\
=&\left|\mathbb{E}\left[\exp\left\{\mathrm{i}sY_n\right\}\mathbb{E}\left[\exp\left\{\mathrm{i}tX_n\right\}\mid\mathcal{C}_n\right]\right]-\exp\left(-t^2/2-s^2/2\right)\right| \\
\leq&\left|\mathbb{E}\left[\left|\exp\left\{\mathrm{i}sY_n\right\}\right|\left|\mathbb{E}\left[\exp\left\{\mathrm{i}tX_n\right\}\mid\mathcal{C}_n\right]-\exp(-t^2/2)\right|\right]\right| \\
&+\left|\exp(-t^2/2)\mathbb{E}\left[\exp\left\{\mathrm{i}sY_n\right\}-\exp\left(-s^2/2\right)\right]\right| \\
\leq&\mathbb{E}\left[\left|\mathbb{E}\left[\exp\left\{\mathrm{i}tX_n\right\}\mid\mathcal{C}_n\right]-\exp(-t^2/2)\right|\right]+\exp(-t^2/2)\left|\mathbb{E}\left[\exp\left\{\mathrm{i}sY_n\right\}-\exp\left(-s^2/2\right)\right]\right| \\
\to&0
\end{aligned}
$$

对于任意 $t\in\mathbb{R}$ 和 $s\in\mathbb{R}$，其中第二个等式源于以下事实： Y_n 关于 $\mathcal{C}_n$ 可测，最后一步则由 $\mathbb{E}\left[\left|\mathbb{E}\left[\exp\{\mathrm{i}tX_n\}\mid\mathcal{C}_n\right]-\exp\left(-t^2/2\right)\right|\right]\to 0$ 和 $\mathbb{E}\left[\exp\left\{\mathrm{i}sY_n\right\}\right]\to\exp\left(-s^2/2\right)$ 得出。根据 Levy连续性定理，结论成立。■

Lemma 6. 设 X_n 和 Y_n 是满足以下条件的两个随机变量序列

$$
(X_n,Y_n)^\top\xrightarrow{d}(X,Y)^\top\overset{d}{=}N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right).
$$

设 a_n 和 b_n 为 满足 $a_n^2+b_n^2=1$ 对所有 $n\geq 1$均成立的两个实数序列。则

$$
a_nX_n+b_nY_n\xrightarrow{d}N(0,1).
$$

证明。 根据 Levy连续性定理，我们有

$$
\phi_n(t,s):=\mathbb{E}\left[\exp\left\{\mathrm{i}tX_n+\mathrm{i}sY_n\right\}\right]\to\mathbb{E}\left[\exp\left\{\mathrm{i}tX+\mathrm{i}sY\right\}\right]=\exp\left(-t^2/2-s^2/2\right)=:\phi(t,s)
$$

作为 $n\to\infty$，对于任意 $(t,\ s)\in\mathbb{R}^2$。我们证明， $\phi_n(t,\ s)$ 在任意紧集 $\mathcal{G}\subset\mathbb{R}^2$ 上均匀收敛于 $\phi(t,\ s)$。我们固定任意一点 $(t_0,\ s_0)\in\mathbb{R}^2$ 以及任意一个数 $\epsilon>0$。我们采用

3ϵ 论证。

第一 ϵ。 注意到 $\phi(t,s)$ 是连续的，因此存在 $\delta_1 > 0$，使得

$$|\phi(t_0, s_0) - \phi(t, s)| \leq \epsilon$$

对于任意 $(t,s) \in \mathcal{B}((t_0,s_0);\ \delta_1) := \{(t,s) :\ \|(t,s) - (t_0,s_0)\| < \delta_1\}$。

第二 ϵ。 由于 $\phi_n(t_0,s_0) \to \phi(t_0,s_0)$，存在 $N_2 \geq 1$，使得

$$|\phi_n(t_0, s_0) \to \phi(t_0, s_0)| \leq \epsilon$$

对于任意 $n \geq N_2$。

第三 ϵ。 根据 Prokhorov定理（van der Vaart （1998，定理2.4））， $\{(X_n,\ Y_n) :\ n \geq 1\}$是一致紧的。于是存在 $R > 0$ 使得

$$\sup_{n \geq 1} P\left(\|(X_n, Y_n)\| \geq R\right) < \epsilon/4.$$

Note 使 $\exp\{\mathrm{i}t\}$ 在 $t = 0$ 处连续，因此存在 $\delta_3 > 0$，使得 at

$$|\exp\{\mathrm{i}(t - t_0)x + \mathrm{i}(s - s_0)y\} - 1| \leq \epsilon/2$$

对于所有 $(t,s) \in \mathcal{B}((t_0,s_0);\ \delta_3)$ 和 $(x,y) \in \mathcal{B}(\mathbf{0};\ R)$。现在我们有

$$\begin{aligned}
|\phi_n(t,s) - \phi_n(t_0,s_0)| &\leq \left|\mathbb{E}\left[\exp\left\{\mathrm{i}tX_n + \mathrm{i}sY_n\right\}\right] - \mathbb{E}\left[\exp\left\{\mathrm{i}t_0X_n + \mathrm{i}s_0Y_n\right\}\right]\right| \\
&\leq \left|\mathbb{E}\left[\exp\left\{\mathrm{i}t_0X_n + \mathrm{i}s_0Y_n\right\}\left(\exp\left\{\mathrm{i}(t-t_0)X_n + \mathrm{i}(s-s_0)Y_n\right\} - 1\right)\right]\right| \\
&\leq \mathbb{E}\left[\left|\exp\left\{\mathrm{i}(t-t_0)X_n + \mathrm{i}(s-s_0)Y_n\right\} - 1\right|\right] \\
&= \int_{\mathbb{R}^2} \left|\exp\left\{\mathrm{i}(t-t_0)x + \mathrm{i}(s-s_0)y\right\} - 1\right| dF_{X_n,Y_n}(x,y) \\
&= \int_{\mathcal{B}(\mathbf{0};R)} \left|\exp\left\{\mathrm{i}(t-t_0)x + \mathrm{i}(s-s_0)y\right\} - 1\right| dF_{X_n,Y_n}(x,y) \\
&\quad + \int_{\mathbb{R}^2 \setminus \mathcal{B}(\mathbf{0};R)} \left|\exp\left\{\mathrm{i}(t-t_0)x + \mathrm{i}(s-s_0)y\right\} - 1\right| dF_{X_n,Y_n}(x,y)
\end{aligned}$$

$$\leq \epsilon/2 + 2 \times \epsilon/4 = \epsilon$$

对于所有 $(t, s) \in \mathcal{B}((t_0, s_0); \delta_3)$ 以及 $n \geq 1$。

令 $\delta(t_0, s_0) := \min\{\delta_1, \delta_3\} > 0$ 且 $N(t_0, s_0) := N_2 \geq 1$，我们有

$$\begin{aligned} |\phi_n(t,s) - \phi(t,s)| &\leq |\phi_n(t,s) - \phi_n(t_0,s_0)| + |\phi_n(t_0,s_0) - \phi(t_0,s_0)| + |\phi(t_0,s_0) - \phi(t,s)| \\ &\leq 3\epsilon \end{aligned} \tag{A.2}$$

对于所有 $(t, s) \in \mathcal{B}((t_0, s_0); \delta(t_0, s_0))$ 以及 $n \geq N(t_0, s_0)$。

设 G 是 $\mathbb{R}^2$ 的任意紧子集。则 $\cup_{(t_0,s_0)\in G} \mathrm{B}((t_0, s_0); \delta(t_0, s_0))$ 是 G 的一个开覆盖。根据海涅–博雷尔定理，存在一个有限常数 $G \geq 1$，使得 $\bigcup_{g=1}^{G} \mathrm{B}((t_g, s_g); \delta(t_g, s_g))$ 是 G 的一个有限子覆盖。令 $\delta := \min_{1\leq g\leq G}\{\delta(t_g, s_g)\} > 0$ 且 $N := \max_{1\leq g\leq G}\{N(t_g, s_g)\} < \infty$，则由 (A.2) 可得

$$|\phi_n(t,s) - \phi(t,s)| \leq 3\epsilon$$

对于所有 $(t, s) \in \mathcal{G}$ 和 $n \geq N$。因此，$\phi_n(t, s)$ 在任意紧集 $\mathcal{G} \subset \mathbb{R}^2$ 上一致收敛于 $\phi(t, s)$。

对于任意 $u \in \mathbb{R}$，令 $\mathcal{G}_u := \{(t, s) \in \mathbb{R}^2 : t^2 + s^2 = u^2\}$。则 $\mathcal{G}_u$ 是 $\mathbb{R}^2$ 的紧子集。现在我们有

$$\begin{aligned} &\left|\mathbb{E}\left[\exp\left\{\mathrm{i}u\left\{a_n X_n + b_n Y_n\right\}\right\}\right] - \mathbb{E}\left[\exp\left\{-u^2/2\right\}\right]\right| \\ =&|\phi_n(ua_n, ub_n) - \phi(ua_n, ub_n)| \leq \sup_{(t,s)\in\mathcal{G}_u} |\phi_n(t,s) - \phi(t,s)| \\ \to&0 \end{aligned}$$

当 $n \to \infty$ 时。由于 $u \in \mathbb{R}$ 是任意的，结论由 Levy连续性定理得出。■

引理7. 设$(\Omega, F\ P)$为一个概率空间，且 C_n 是 F 的一列子-σ-代数。

设$\{X_n \in \mathbb{R}, n \geq 1\}$和$\{Y_n \in \mathbb{R}, n \geq 1\}$是两个随机变量序列。设 E_n 是Ω 的一个子集，并且

$$Z_n(\omega) := \begin{cases} X_n(\omega), & \omega \in E_n, \\ Y_n(\omega), & \omega \in \Omega \backslash E_n. \end{cases}$$

若 $E_n \in \mathcal{C}_n$，则 $\mathbb{E}[Z_n \mid \mathcal{C}_n] = R_n$ 几乎必然，其中

$$R_n(\omega) := \begin{cases} \mathbb{E}[X_n \mid \mathcal{C}_n](\omega), & \omega \in E_n, \\ \mathbb{E}[Y_n \mid \mathcal{C}_n](\omega), & \omega \in \Omega \backslash E_n. \end{cases}$$

证明。对于任意 $H \in \mathcal{C}_n$，我们有

$$\begin{aligned} \int_H R_n dP &= \int_{H \cap E_n} R_n dP + \int_{H \cap (\Omega \backslash E_n)} R_n dP \\ &= \int_{H \cap E_n} \mathbb{E}[X_n \mid \mathcal{C}_n] dP + \int_{H \cap (\Omega \backslash E_n)} \mathbb{E}[Y_n \mid \mathcal{C}_n] dP \\ &= \int_{H \cap E_n} X_n dP + \int_{H \cap (\Omega \backslash E_n)} Y_n dP = \int_{H \cap E_n} Z_n dP + \int_{H \cap (\Omega \backslash E_n)} Z_n dP \\ &= \int_H Z_n dP, \end{aligned}$$

其中第三个不等式由 $E_n \in \mathcal{C}_n$ 以及 $\mathbb{E}[X_n \mid \mathcal{C}_n]$ 和 $\mathbb{E}[Y_n \mid \mathcal{C}_n]$的定义得出。根据条件期望的定义，所需结果随即成立。 ■

附录 B 主要定理的证明

B.1 符号约定

我们给出渐近方差的表达式:

$$\varsigma^2_{\widetilde{Y}} := \sum_{k=1}^{K} \frac{p_{[k]}}{1 - \pi_{[k]}} \mathrm{Var}\left(\widetilde{Y}_i(0) \mid B_i = k\right) + \sum_{k=1}^{K} \frac{p_{[k]}}{\pi_{[k]}} \mathrm{Var}\left(\widetilde{Y}_i(1) \mid B_i = k\right)$$

$$\varsigma^2_H := \sum_{k=1}^{K} p_{[k]} \left\{\mathbb{E}[Y_i(1) - Y_i(0) \mid B_i = k]\right\}^2 - \left\{\sum_{k=1}^{K} p_{[k]} \mathbb{E}[Y_i(1) - Y_i(0) \mid B_i = k]\right\}^2,$$

$$\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(a)} := \mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(a) \mid B_i = k\right],\ 1 \le k \le K, a \in \{0,1\},$$

$$\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*} := \mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right],\ 1 \le k \le K,$$

$$\begin{aligned}\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*} := \sum_{k=1}^K \frac{p_{[k]}}{\pi_{[k]}(1-\pi_{[k]})} &\left\{(1-\pi_{[k]})\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)} + \pi_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}\right\}^\top \\ &\times \boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\left\{(1-\pi_{[k]})\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)} + \pi_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}\right\},\end{aligned}$$

and

$$\varsigma^2_{\widetilde{Y}|\widetilde{h}^*} := \sum_{k=1}^K \frac{p_{[k]}}{\pi_{[k]}(1-\pi_{[k]})}\mathbb{E}\left[\left\{(1-\pi_{[k]})\widetilde{h}^*_{1[k]}(\boldsymbol{X}_i) + \pi_{[k]}\widetilde{h}^*_{0[k]}(\boldsymbol{X}_i)\right\}^2 \mid B_i = k\right].$$

渐近方差的估计量定义如下:

$$\begin{aligned}\widehat{\varsigma}^2_{\widetilde{Y}} := &\sum_{k=1}^K \frac{n_{[k]}}{n_{[k]} - \widehat{r}_{[k]} - 1}\frac{p_{n[k]}}{1-\pi_{n[k]}}\frac{1}{n_{0[k]}}\sum_{i\in[k]}(1-A_i)(Y_i - \overline{Y}_{0[k]})^2 \\ &+ \sum_{k=1}^K \frac{n_{[k]}}{n_{[k]} - \widehat{r}_{[k]} - 1}\frac{p_{n[k]}}{\pi_{n[k]}}\frac{1}{n_{1[k]}}\sum_{i\in[k]}A_i(Y_i - \overline{Y}_{1[k]})^2, \qquad \text{(B.1)}\end{aligned}$$

$$\widehat{\varsigma}^2_H := \sum_{k=1}^K \frac{n_{[k]}}{n_{[k]} - \widehat{r}_{[k]} - 1}p_{n[k]}\left\{\overline{Y}_{1[k]} - \overline{Y}_{0[k]} - (\overline{Y}_1 - \overline{Y}_0)\right\}^2, \qquad \text{(B.2)}$$

$$\widehat{\boldsymbol{\Gamma}}_{[k]} := \frac{1}{n}\sum_{i\in[k]}\left\{\frac{(1-\pi_{n[k]})A_i}{\pi_{n[k]}}(\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]})(Y_i - \overline{Y}_{1[k]}) + \frac{\pi_{n[k]}(1-A_i)}{1-\pi_{n[k]}}(\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]})(Y_i - \overline{Y}_{0[k]})\right\},$$

$$\widehat{\boldsymbol{\Sigma}}_{[k]} := \frac{1}{n}\sum_{i\in[k]}(A_i - \pi_{n[k]})^2(\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]})(\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]})^\top,\ 1 \le k \le K,$$

and

$$\widehat{\varsigma}^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*} := \sum_{k=1}^K \frac{n_{[k]}}{n_{[k]} - \widehat{r}_{[k]} - 1}\widehat{\boldsymbol{\Gamma}}^\top_{[k]}\widehat{\boldsymbol{\Sigma}}^+_{[k]}\widehat{\boldsymbol{\Gamma}}_{[k]}. \qquad \text{(B.3)}$$

注意，在（B.1）—（B.3）中，我们将每一层的自由度乘以因子$n_{[k]}/(n_{[k]} - \widehat{r}_{n,[k]} - 1)$进行调整，其中 $\widehat{r}_{[k]}$ 是$r_{[k]}$ =秩$\left\{\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right]\right\}$的相合估计量，即当 $n \to \infty$ 时，$\sup_{1\le k\le K}\left\{\left|\widehat{r}_{[k]} - r_{[k]}\right|/r_{[k]}\right\} = o_P(1)$。如果 $\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right]$ 非奇异，那么我们可以直接取 $\widehat{r}_{[k]} = d$。

表6 列出了本节将使用的符号约定。令$\mathcal{C}_n$ 表示由 σ生成的$(A^{(n)},\ B^{(n)})$ 与 $\mathbb{E}_{\mathcal{C}_n}[\cdot]:=\mathbb{E}[\cdot\mid \mathcal{C}_n]$。

表6：B节使用的符号列表B。

符号约定	表达式	符号约定	表达式
$\widetilde{Y}_i(a)$	$Y_i(a)-\mathbb{E}[Y_i(a)\mid B_i]$	$\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)$	$\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)-\mathbb{E}[\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\mid B_i]$
$\boldsymbol{\Xi}_{i,[k]}$	$\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}\right\}$	$\boldsymbol{\Xi}_i$	$\left(\boldsymbol{\Xi}^\top_{i,[1]},\ldots,\boldsymbol{\Xi}^\top_{i,[K]}\right)^\top$
$\boldsymbol{\Xi}^*_{i,[k]}$	$\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\ \widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)$	$\boldsymbol{\Xi}^*_i$	$\left(\boldsymbol{\Xi}^{*\top}_{i,[1]},\ldots,\boldsymbol{\Xi}^{*\top}_{i,[K]}\right)^\top$
$\widetilde{Y}_{i,\ [k]}$	$\mathbb{1}(B_i=k)\left\{\frac{A_i}{\pi_{n[k]}}(Y_i-\overline{Y}_{1[k]})-\frac{1-A_i}{1-\pi_{n[k]}}(Y_i-\overline{Y}_{0[k]})\right\}$	$\widetilde{Y}_i$	$\sum_{k=1}^K\widetilde{Y}_{i,\ [k]}$
$\widetilde{Y}^*_{i,\ [k]}$	$\mathbb{1}(B_i=k)\left\{\frac{A_i}{\pi_{n[k]}}\widetilde{Y}_i(1)-\frac{1-A_i}{1-\pi_{n[k]}}\widetilde{Y}_i(0)\right\}$	$\widetilde{Y}^*_i$	$\sum_{k=1}^K\widetilde{Y}^*_{i,[k]}$
$\boldsymbol{\Sigma}_{[k]}$	$\pi_{[k]}\left(1-\pi_{[k]}\right)p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{\boldsymbol{\xi}}^*_n}$	$\boldsymbol{\Sigma}$	对角线$\left\{\boldsymbol{\Sigma}_{[k]}:\ k=1,\ldots,K\right\}$
$\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]}$	$\frac{1}{n}\sum_{i=1}^n\left\{A_i-\pi_{n[k]}\right\}^2\mathbb{1}(B_i=k)\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{\boldsymbol{\xi}}^*_n}$	$\boldsymbol{\Sigma}^{\mathcal{C}_n}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}$	$\frac{1}{n}\sum_{i=1}^n\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}\right]$
$\boldsymbol{\beta}^*_{[k]}$	$\boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{\boldsymbol{\xi}}^*_n}\left\{\frac{1}{\pi_{[k]}}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}(1)}+\frac{1}{1-\pi_{[k]}}\boldsymbol{\Sigma}^{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}(0)}\right\}$	$\boldsymbol{\beta}^*$	$\left(\boldsymbol{\beta}^{*\top}_{[1]},\ldots,\boldsymbol{\beta}^{*\top}_{[K]}\right)^\top$
$\boldsymbol{\beta}_{[k],\mathcal{C}_n}$	$\left\{\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]}\right\}^+\left\{\frac{1}{n}\sum_{i=1}^n\mathbb{E}_{\mathcal{C}_n}\left[\widetilde{Y}^*_{i,[k]}\boldsymbol{\Xi}^*_{i,[k]}\right]\right\}$	$\boldsymbol{\beta}_{\mathcal{C}_n}$	$\left(\boldsymbol{\beta}^\top_{[1],\mathcal{C}_n},\ \ldots\ldots,\ \boldsymbol{\beta}^\top_{[K],\mathcal{C}_n}\right)^\top$
$\widehat{\boldsymbol{\beta}}_{[k]}$	$\left\{\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}^\top_{i,[k]}\right\}^+\frac{1}{n}\sum_{i=1}^n\widetilde{Y}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}$	$\widehat{\boldsymbol{\beta}}$	$\left(\widehat{\boldsymbol{\beta}}^\top_{[1]},\ldots,\widehat{\boldsymbol{\beta}}^\top_{[K]}\right)^\top$
$\epsilon_{i,[k]}$	$\widetilde{Y}_{i,[k]}-\boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\boldsymbol{\Xi}_{i,[k]}$	ϵ_i	$\left(\epsilon_{i,[1]},\ldots,\epsilon_{i,[K]}\right)^\top$
$\epsilon^*_{i,[k]}$	$\widetilde{Y}^*_{i,\ [k]}-\boldsymbol{\beta}^\top_{[k],\ \mathcal{C}_n}\boldsymbol{\Xi}^*_{i,\ [k]}$	ϵ^*_i	$\left(\epsilon^*_{i,[1]},\ldots,\epsilon^*_{i,[K]}\right)^\top$

B.2 证明定理 1

推导校准权重的表达式。我们首先推导这些表达式

对于校准权重 $\widehat{w}_i$'s。对于所有 $i\in\{1,\ldots,n\}$，令

$$\boldsymbol{\Xi}_i:=\begin{pmatrix}\left\{A_i-\pi_{n[1]}\right\}\mathbb{1}(B_i=1)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[1]}\right\}\\ \left\{A_i-\pi_{n[2]}\right\}\mathbb{1}(B_i=2)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[2]}\right\}\\ \vdots\\ \left\{A_i-\pi_{n[K]}\right\}\mathbb{1}(B_i=K)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[K]}\right\}\end{pmatrix}\in\mathbb{R}^{Kd}.$$

令 $D(\boldsymbol{w}) = \sum_{i=1}^{n} D(w_i,\ 1)$。然后，校准问题可以改写为

$$\min_{w_i:1\le i\le n} D(\boldsymbol{w}) \ \text{ s.t. } \ (\boldsymbol{\Xi}_1,\ldots,\boldsymbol{\Xi}_n)\,\boldsymbol{w}=0. \tag{B.4}$$

注意， $D(\boldsymbol{w})$ 的共轭函数为

$$D^*(\boldsymbol{z}) = \sum_{i=1}^{n}\left\{z_i\cdot(D')^{-1}(z_i) - D\left\{(D')^{-1}(z_i)\right\}\right\} = \sum_{i=1}^{n} -\rho(-z_i), \quad \forall \boldsymbol{z}=(z_1,\ldots,z_n)^\top,$$

其中 $\rho(v) := D\{(D')^{-1}(-v)\} + v\cdot(D')^{-1}(-v)$ 对任意 $v\in\mathbb{R}$成立。根据 Tseng & Bertsekas (1991)（另见 Boyd & Vandenberghe (2004年)），(B.4) 的对偶问题为

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^{Kd}}\left\{-D^*((\boldsymbol{\Xi}_1,\ldots,\boldsymbol{\Xi}_n)^\top\boldsymbol{\lambda})\right\} = \max_{\boldsymbol{\lambda}\in\mathbb{R}^{Kd}}\sum_{i=1}^{n}\rho(-\boldsymbol{\lambda}^\top\boldsymbol{\Xi}_i) = \max_{\boldsymbol{\lambda}\in\mathbb{R}^{Kd}}\frac{1}{n}\sum_{i=1}^{n}\rho(\boldsymbol{\lambda}^\top\boldsymbol{\Xi}_i).$$

令 $\hat{\boldsymbol{\lambda}} := \arg\max_{\boldsymbol{\lambda}\in\mathbb{R}^{Kd}}\frac{1}{n}\sum_{i=1}^{n}\rho(\boldsymbol{\lambda}^\top\boldsymbol{\Xi}_i)$。接下来，考虑特殊情形 $D(v) = (v-1)^2/2$，我们有 $\rho(v) = -v^2/2+v$, $\rho'(v) = -v+1$ 以及 $\rho''(v) = -1$。于是

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}\in\mathbb{R}^{Kd}}\frac{1}{n}\sum_{i=1}^{n}\left\{-\frac{1}{2}\boldsymbol{\lambda}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\boldsymbol{\lambda}+\boldsymbol{\lambda}^\top\boldsymbol{\Xi}_i\right\} = \arg\max_{\boldsymbol{\lambda}\in\mathbb{R}^{Kd}}\left\{-\frac{1}{2}\boldsymbol{\lambda}^\top\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\boldsymbol{\lambda}+\boldsymbol{\lambda}^\top\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\right\}. \tag{B.5}$$

我们声称

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i \in \operatorname{range}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right), \tag{B.6}$$

other否则，优化问题 (B.5) 无界，这与

$$-\frac{1}{2}\boldsymbol{\lambda}^\top\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\boldsymbol{\lambda}+\boldsymbol{\lambda}^\top\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i = \frac{1}{n}\sum_{i=1}^{n}\left\{-\frac{1}{2}\left(\boldsymbol{\lambda}^\top\boldsymbol{\Xi}_i\right)^2+\boldsymbol{\lambda}^\top\boldsymbol{\Xi}_i\right\} \le \frac{1}{2}.$$

由（B.5）的一阶条件可得

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\boldsymbol{\lambda} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i.$$

由（B.6）可知，该方程至少有一个解。我们始终取该解为

$$\widehat{\boldsymbol{\lambda}} = \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^\top \right\}^+ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i, \tag{B.7}$$

其中，$\left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^\top \right\}^+$ 是 $\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^\top$ 的摩尔-彭若斯伪逆。

注意，

$$\sum_{k=1}^{K} p_{n[k]} \left(\overline{Y}_{1[k]} - \overline{Y}_{0[k]} \right) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \left\{ \frac{A_i}{\pi_{n[k]}} - \frac{1 - A_i}{1 - \pi_{n[k]}} \right\} \mathbb{1}(B_i = k) \cdot Y_i \tag{B.8}$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \left\{ \frac{A_i}{\pi_{n[k]}} \left(Y_i - \overline{Y}_{1[k]} \right) - \frac{1 - A_i}{1 - \pi_{n[k]}} \left(Y_i - \overline{Y}_{0[k]} \right) \right\} \mathbb{1}(B_i = k) = 0. \tag{B.9}$$

我们有

$$\begin{aligned}
&\widehat{\tau}_{\text{cal}} - \tau \\
=& \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \left\{ \frac{A_i}{\pi_{n[k]}} - \frac{1 - A_i}{1 - \pi_{n[k]}} \right\} \mathbb{1}(B_i = k) \cdot Y_i - \tau \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} (\widehat{w}_i - 1) \sum_{k=1}^{K} \left\{ \frac{A_i}{\pi_{n[k]}} \left(Y_i - \overline{Y}_{1[k]} \right) - \frac{1 - A_i}{1 - \pi_{n[k]}} \left(Y_i - \overline{Y}_{0[k]} \right) \right\} \mathbb{1}(B_i = k) \\
=& \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \left\{ \frac{A_i}{\pi_{n[k]}} - \frac{1 - A_i}{1 - \pi_{n[k]}} \right\} \mathbb{1}(B_i = k) \cdot Y_i - \tau - \frac{1}{n} \sum_{i=1}^{n} \widehat{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i \sum_{k=1}^{K} \widetilde{Y}_{i,[k]} \\
=& \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \left\{ \frac{A_i}{\pi_{n[k]}} - \frac{1 - A_i}{1 - \pi_{n[k]}} \right\} \mathbb{1}(B_i = k) \cdot Y_i - \tau - \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \boldsymbol{\Xi}_i^\top \widetilde{Y}_{i,[k]} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^\top \right\}^+ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i \\
=& \sum_{k=1}^{K} \frac{1}{\pi_{n[k]}} \frac{1}{n} \sum_{i=1}^{n} A_i \mathbb{1}(B_i = k) \widetilde{Y}_i(1) - \sum_{k=1}^{K} \frac{1}{1 - \pi_{n[k]}} \frac{1}{n} \sum_{i=1}^{n} (1 - A_i) \mathbb{1}(B_i = k) \widetilde{Y}_i(0) \\
&\quad + \sum_{k=1}^{K} \mathbb{E}\left[Y_i(1) \mid B_i = k \right] \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{1}(B_i = k) - p_{[k]} \right) - \sum_{k=1}^{K} \mathbb{E}\left[Y_i(0) \mid B_i = k \right] \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{1}(B_i = k) - p_{[k]} \right) \\
&\quad - \sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]}^\top \widetilde{Y}_{i,[k]} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}_{i,[k]}^\top \right\}^+ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]}
\end{aligned}$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\widetilde{Y}^{*}_{i,[k]}}_{R_1} + \underbrace{\left\{-\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\widehat{\boldsymbol{\beta}}^{\top}_{[k]}\boldsymbol{\Xi}_{i,[k]}\right\}}_{R_2} + \underbrace{\left\{\sum_{k=1}^{K}\mathbb{E}\left[Y_i(1)\mid B_i=k\right]\cdot\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}(B_i=k)-p_{[k]}\right)\right\}}_{R_3}$$
$$+ \underbrace{\left\{-\sum_{k=1}^{K}\mathbb{E}\left[Y_i(0)\mid B_i=k\right]\cdot\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}(B_i=k)-p_{[k]}\right)\right\}}_{R_4}, \tag{B.10}$$

其中第一个等式利用了（B.8）和（B.9），第三个等式由 $\widehat{\boldsymbol{\lambda}} = \left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\right\}^{+}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i$得到，第四个等式由

$$\sum_{k=1}^{K}\boldsymbol{\Xi}_i^{\top}\widetilde{Y}_{i,[k]} = \left(\boldsymbol{\Xi}_{i,[1]}^{\top}\sum_{k=1}^{K}\widetilde{Y}_{i,[k]},\ldots,\boldsymbol{\Xi}_{i,[K]}^{\top}\sum_{k=1}^{K}\widetilde{Y}_{i,[k]}\right) = \left(\boldsymbol{\Xi}_{i,[1]}^{\top}\widetilde{Y}_{i,[1]},\ldots,\boldsymbol{\Xi}_{i,[K]}^{\top}\widetilde{Y}_{i,[K]}\right)$$

以及$\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k) = \pi_{n[k]}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)$。

(B.10) 中的最后两项（R_3 和 R_4）可以使用经典中心极限定理处理。我们聚焦于 (B.10) 中的前两项（R_1 和 R_2）。首先，我们证明 $\widehat{\boldsymbol{\beta}}_{[k]}$的一致性。只需证明对于每个 $k=1,\ldots,K$，$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}$ 和$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{\top}\widetilde{Y}_{i,[k]}$ 的一致性结果。

$\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\right\}^{+}$**的概率极限**。我们有

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\\
=&\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\bar{\boldsymbol{\xi}}_{n[k]}^{*}\right\}\left\{\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\bar{\boldsymbol{\xi}}_{n[k]}^{*}\right\}^{\top}-\\
&\frac{2}{n}\sum_{i=1}^{n}\pi_{n[k]}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\bar{\boldsymbol{\xi}}_{n[k]}^{*}\right\}\left\{\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\bar{\boldsymbol{\xi}}_{n[k]}^{*}\right\}^{\top}+\\
&\frac{1}{n}\sum_{i=1}^{n}\pi_{n[k]}^{2}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\bar{\boldsymbol{\xi}}_{n[k]}^{*}\right\}\left\{\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\bar{\boldsymbol{\xi}}_{n[k]}^{*}\right\}^{\top}+\\
&\underbrace{\frac{1}{n}\sum_{i=1}^{n}(A_i-\pi_{n[k]})^{2}\mathbb{1}(B_i=k)\left[\left(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\bar{\boldsymbol{\xi}}_{n[k]}\right)\left(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\bar{\boldsymbol{\xi}}_{n[k]}\right)^{\top}-\left(\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\bar{\boldsymbol{\xi}}_{n[k]}^{*}\right)\left(\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\bar{\boldsymbol{\xi}}_{n[k]}^{*}\right)^{\top}\right]}_{R_5},
\end{aligned} \tag{B.11}$$

其中，$\bar{\boldsymbol{\xi}}_{n[k]}^{*} := \frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)$ 是 $\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)$ 的分层特定样本均值。

对于每个 $1 \leq k \leq K$ 和 $\boldsymbol{\alpha} \in \mathbb{R}^d$，根据柯西–施瓦茨不等式，有

$$\begin{aligned}
\boldsymbol{\alpha}^\top R_5 \boldsymbol{\alpha} &= \frac{1}{n}\sum_{i=1}^{n}\left(A_i - \pi_{n[k]}\right)^2 \mathbb{1}(B_i = k)\left[\left(\boldsymbol{\alpha}^\top \boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\alpha}^\top \overline{\boldsymbol{\xi}}_{n[k]}\right)^2 - \left(\boldsymbol{\alpha}^\top \boldsymbol{\xi}_n^*(\boldsymbol{X}_i) - \boldsymbol{\alpha}^\top \overline{\boldsymbol{\xi}}_{n[k]}^*\right)^2\right] \\
&\leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\{\boldsymbol{\alpha}^\top \boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\alpha}^\top \overline{\boldsymbol{\xi}}_{n[k]} - \left(\boldsymbol{\alpha}^\top \boldsymbol{\xi}_n^*(\boldsymbol{X}_i) - \boldsymbol{\alpha}^\top \overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\}^2} \times \\
&\quad \sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\{\boldsymbol{\alpha}^\top \boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\alpha}^\top \overline{\boldsymbol{\xi}}_{n[k]} + \left(\boldsymbol{\alpha}^\top \boldsymbol{\xi}_n^*(\boldsymbol{X}_i) - \boldsymbol{\alpha}^\top \overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\}^2} \\
&\leq \|\boldsymbol{\alpha}\|^2 \sqrt{\underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]} - \left(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\|^2}_{R_6}} \times \\
&\quad \sqrt{\underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]} + \left(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\|^2}_{R_7}},
\end{aligned}$$

从而得到

$$\|R_5\| \leq \sqrt{R_6} \times \sqrt{R_7}.$$

我们分别控制 R_6 和 R_7。根据假设3，可得

$$\begin{aligned}
\left\|\overline{\boldsymbol{\xi}}_{n[k]} - \overline{\boldsymbol{\xi}}_{n[k]}^*\right\| &= \left\|\frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}\right\| \\
&\leq \sqrt{\frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^2} = o_P(1) \qquad \text{(B.12)}
\end{aligned}$$

对所有 $k = 1, \ldots K$均成立。随后，根据假设3、大数弱律以及 (B.12) that

$$R_6 \leq 2 \times \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^2 + 2 \times \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\overline{\boldsymbol{\xi}}_{n[k]} - \overline{\boldsymbol{\xi}}_{n[k]}^*\right\|^2 = o_P(1). \qquad \text{(B.13)}$$

根据引理 3 和假设 3，我们有

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^2 - \mathbb{E}\left[\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^2\right]\right\} = o_P(1)$$

and

$$\begin{aligned}
&\overline{\boldsymbol{\xi}}^*_{n[k]} - \mathbb{E}\left[\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) \mid B_i = k\right] \\
=&\frac{n}{n_{[k]}}\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{1}(B_i = k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \mathbb{E}\left[\mathbb{1}(B_i = k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right]\right\} + o_P(1) = o_P(1). \qquad \text{(B.14)}
\end{aligned}$$

因此，我们有

$$\begin{aligned}
R_7 &\leq 2 \times R_6 + 8 \times \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}^*_{n[k]}\right\|^2 \\
&\leq o_P(1) + \frac{16}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right\|^2 + \frac{16}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\overline{\boldsymbol{\xi}}^*_{n[k]}\right\|^2 \\
&= o_P(1) + \{O(1) + o_P(1)\} + O_P(1)\{O(1) + o_P(1)\}^2 = O_P(1).
\end{aligned}$$

现在，我们可以得到 $\|R_5\| \leq \sqrt{R_6} \overline{\times} \sqrt{R_7} \overline{=} o_P(1)$，因此

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top \\
=&\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i = k)\left\{\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}^*_{n[k]}\right\}\left\{\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}^*_{n[k]}\right\}^\top \\
&- \frac{2}{n}\sum_{i=1}^{n}\pi_{n[k]}A_i\mathbb{1}(B_i = k)\left\{\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}^*_{n[k]}\right\}\left\{\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}^*_{n[k]}\right\}^\top \\
&+ \frac{1}{n}\sum_{i=1}^{n}\pi^2_{n[k]}\mathbb{1}(B_i = k)\left\{\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}^*_{n[k]}\right\}\left\{\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}^*_{n[k]}\right\}^\top + o_P(1).
\end{aligned}$$

令 C_n 表示由 $(A^{(n)}, B^{(n)})$ 生成的 σ-代数。根据假设 1 和 2，我们有 $(Y_i(1)\ Y_i(0)\ \boldsymbol{X}_i^\top)^\top$，$i = 1, \ldots, n$ 在 C_n 条件下相互独立。为简化起见，我们令 $\mathbb{E}_{\mathcal{C}_n}[\cdot] := \mathbb{E}[\cdot \mid \mathsf{C}_n]$。

注意到

$$\mathbb{E}_{\mathcal{C}_n}\left[A_i\mathbb{1}(B_i = k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)^\top\right] = A_i\mathbb{1}(B_i = k)\mathbb{E}\left[\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)^\top \mid B_i = k\right]$$

and

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{C}_n}\left[\left\{\left\|A_i\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top-\mathbb{E}_{\mathcal{C}}\left[A_i\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right]\right\|_F\right\}^{1+\epsilon/2}\right]\\
\leq&\mathbb{E}_{\mathcal{C}_n}\left[\left\{\left\|A_i\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right\|_F\right\}^{1+\epsilon/2}\right]\\
\leq&\mathbb{E}\left[\left\{\sqrt{\operatorname{tr}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right\}}\right\}^{1+\epsilon/2}\mid B_i=k\right]\\
=&\mathbb{E}\left[\left\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^{2+\epsilon}\mid B_i=k\right]=O(1),
\end{aligned}
$$

其中最后一步由假设3得出。由引理3，我们有[3]

$$
\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top=\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(B_i=k)\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\mid B_i=k\right]+o_P(1).
$$

同样地，我们还可以得到

$$
\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)=\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(B_i=k)\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]+o_P(1),
$$

$$
\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top=\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\mid B_i=k\right]+o_P(1),
$$

$$
\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)=\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]+o_P(1),
$$

and

$$
\overline{\boldsymbol{\xi}}_{n[k]}^*=\frac{n}{n_{[k]}}\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)=\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]+o_P(1). \tag{B.15}
$$

由于 $\pi_{n[k]}=O_P(1)$ 和 $\overline{\boldsymbol{\xi}}_{n[k]}^*=O_P(1)$，我们可以推导出

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top\\
=&o_P(1)+\frac{1}{n}\sum_{i=1}^n\left\{A_i-2\pi_{n[k]}A_i+\pi_{n[k]}^2\right\}\mathbb{1}(B_i=k)\times
\end{aligned}
$$

[3]将 $A_i\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top$ 视为一个向量，并利用以下事实：

$$
\left\|\operatorname{vec}\left\{A_i\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right\}\right\|=\left\|A_i\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right\|_F.
$$

$$
\begin{aligned}
&\left\{\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i=k\right]-\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top \mid B_i=k\right]\right\}\\
=&\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^*\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]+o_P(1)\\
=&\frac{1}{n}\sum_{i=1}^n\left\{A_i-\pi_{n[k]}\right\}^2\mathbb{1}(B_i=k)\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}+o_P(1)=\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}+o_P(1).
\end{aligned}
$$

由引理 mma 1，可得 $\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(B_i=k)=\pi_{[k]}p_{[k]}+o_P(1)$。接下来由

$$
\left\|\mathbb{E}\left[\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\mid B_i=k\right]\right\|\le\mathbb{E}\left[\left\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^2\mid B_i=k\right]<\infty
$$

that

$$
\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}=\frac{1}{n}\sum_{i=1}^n\left\{A_i-\pi_{n[k]}\right\}^2\mathbb{1}(B_i=k)\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}=\boldsymbol{\Sigma}_{[k]}+o_P(1).
$$

因此，我们有

$$
\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top=\boldsymbol{\Sigma}_{[k]}+o_P(1). \tag{B.16}
$$

根据假设3 以及摩尔-彭若斯伪逆的推导，我们有

$$
\left\|\boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\right\|\le c^{-1}
$$

对所有 $k=1$、.....、K 以及 $n\ge 1$，都有。于是我们有

$$
\left\|\boldsymbol{\Sigma}^+_{[k]}\right\|=\left\|\pi_{[k]}^{-1}(1-\pi_{[k]})^{-1}p_{[k]}^{-1}\boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\right\|\le\pi_{[k]}^{-1}(1-\pi_{[k]})^{-1}p_{[k]}^{-1}c^{-1}.
$$

根据Stewart (1977)中的定理 3.3以及假设3，我们有

$$
\begin{aligned}
\left\|\left\{\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top\right\}^+-\boldsymbol{\Sigma}^+_{[k]}\right\|&\le 3\max\left\{\left\|\left\{\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top\right\}^+\right\|^2,\left\|\boldsymbol{\Sigma}^+_{[k]}\right\|^2\right\}\left\|\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top-\boldsymbol{\Sigma}^+_{[k]}\right\|\\
&=O_P(1)\times o_P(1)=o_P(1).
\end{aligned} \tag{B.17}
$$

概率 **.....的极限**$\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}_{i,\ [k]}^\top\widetilde{Y}_{i,\ [k]}$。注意，对于每个 $k=1$，.....，K，我们有 th e

根据以下分解:

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{1[k]})\\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})Y_i(1)\\
&-\overline{Y}_{1[k]}\times\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})\\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*)(Y_i(1)-\overline{Y}_{1[k]})\\
&\underbrace{+\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\left(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\}Y_i}_{R_8}\\
&\underbrace{-\overline{Y}_{1[k]}\times\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*)\right\}}_{R_9}. \qquad \text{(B.18)}
\end{aligned}
$$

我们分别控制 R_8 和 R_9。我们有

$$
\begin{aligned}
|R_8|&\le\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}Y_i^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\left(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\|^2}\\
&=O_P(1)\times\sqrt{R_6}=o_P(1)
\end{aligned}
$$

对每个 $k=1$，.....，K，其中，第一步由柯西–施瓦茨不等式得出，第二步由假设 1 得出，最后一步由 $R_6=o_P(1)$得出。同理，注意到 $Y_{1[k]}=\frac{1}{n_{1[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)A_iY_i(1)$ $=O_P(1)$ 可得 _

$$
|R_9|\le O_P(1)\times\sqrt{R_6}=o_P(1).
$$

将 R_8 和 R_9 的结果合并，我们得到

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{1[k]})\\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*)(Y_i(1)-\overline{Y}_{1[k]})+o_P(1) \qquad \text{(B.19)}
\end{aligned}
$$

注意，我们有

$$\mathbb{E}_{\mathcal{C}_n}\left[\mathbb{1}(B_i=k)A_i\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)Y_i(1)\right]=\mathbb{1}(B_i=k)A_i\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)Y_i(1)\mid B_i=k\right]$$

and

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}_n}\left[\left\|\mathbb{1}(B_i=k)A_i\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)Y_i(1)\right\|^{1+\widetilde{\epsilon}}\right]\leq\mathbb{E}\left[\left\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)Y_i(1)\right\|^{1+\widetilde{\epsilon}}\mid B_i=k\right]\\
\leq&\left\{\mathbb{E}\left[\left\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^{(1+\widetilde{\epsilon})\times\frac{2}{1-\epsilon}}\mid B_i=k\right]\right\}^{\frac{1-\epsilon}{2}}\left\{\mathbb{E}\left[\left|Y_i(1)\right|^{(1+\widetilde{\epsilon})\times\frac{2}{1+\epsilon}}\mid B_i=k\right]\right\}^{\frac{1+\epsilon}{2}}\\
=&\left\{\mathbb{E}\left[\left\|\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^{2+\epsilon}\mid B_i=k\right]\right\}^{\frac{1}{2+\epsilon/2}}\left\{\mathbb{E}\left[\left|Y_i(1)\right|^2\mid B_i=k\right]\right\}^{\frac{1+\epsilon/2}{2+\epsilon/2}}\\
=&O(1),
\end{aligned}$$

其中，第二个不等式由赫尔德不等式得出，第一个等式由 $\tilde{\epsilon}$得出：$=\frac{\epsilon/4}{1+\epsilon/4}$ 最后一步则由假设 1 和 3得出。根据引理 3 ，我们有

$$\frac{1}{n}\sum_{i=1}^n\mathbb{1}(B_i=k)A_i\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)Y_i(1)=\frac{1}{n}\sum_{i=1}^n\mathbb{1}(B_i=k)A_i\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)Y_i(1)\mid B_i=k\right]+o_P(1).$$

结合 $\frac{1}{n}\sum_{i=1}^n\mathbb{1}(B_i=k)A_i=\pi_{[k]}p_{[k]}+o_P(1)$ 可得

$$\frac{1}{n}\sum_{i=1}^n\mathbb{1}(B_i=k)A_i\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)Y_i(1)=\pi_{[k]}p_{[k]}\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)Y_i(1)\mid B_i=k\right]+o_P(1).$$

同理可得

$$\frac{1}{n}\sum_{i=1}^n\mathbb{1}(B_i=k)A_iY_i(1)=\pi_{[k]}p_{[k]}\mathbb{E}\left[Y_i(1)\mid B_i=k\right]+o_P(1),$$

$$\frac{1}{n}\sum_{i=1}^n\mathbb{1}(B_i=k)A_i\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)=\pi_{[k]}p_{[k]}\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]+o_P(1)$$

and

$$\overline{Y}_{1[k]}=\frac{1}{n_{1[k]}}\sum_{i=1}^n\mathbb{1}(B_i=k)A_iY_i(1)=\mathbb{E}\left[Y_i(1)\mid B_i=k\right]+o_P(1).$$

回顾式B.15和式B.19，我们可以得到

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{1[k]})\\
=&\frac{1-\pi_{[k]}}{\pi_{[k]}}\pi_{[k]}p_{[k]}\mathbb{E}\left[\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]\}\{Y_i(1)-\mathbb{E}\left[Y_i(1)\mid B_i=k\right]\}\mid B_i=k\right]+o_P(1)\\
=&(1-\pi_{[k]})p_{[k]}\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(1)\mid B_i=k\right]+o_P(1)=(1-\pi_{[k]})p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)}+o_P(1).
\end{aligned}
$$

通过类似的论证，我们还可以得到

$$
\frac{1}{n}\sum_{i=1}^{n}\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\mathbb{1}(B_i=k)(1-A_i)(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{0[k]})=\pi_{[k]}p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}+o_P(1).
$$

于是我们有

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{\top}\widetilde{Y}_{i,[k]}\\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{1[k]})\\
&+\frac{1}{n}\sum_{i=1}^{n}\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\mathbb{1}(B_i=k)(1-A_i)(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{0[k]})\\
=&(1-\pi_{[k]})p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)}+\pi_{[k]}p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}+o_P(1).
\end{aligned}
\tag{B.20}
$$

结合 $\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,\ [k]}\boldsymbol{\Xi}_{i,\ [k]}^{\top}\right\}_{-}^{+}$ 和 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,\ [k]}^{\top}\widetilde{Y}_{i,\ [k]}$ 的结果可得

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{[k]}&=\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\right\}^{+}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{\top}\widetilde{Y}_{i,[k]}\\
&=\left\{\boldsymbol{\Sigma}_{[k]}^{+}+o_P(1)\right\}\left\{(1-\pi_{[k]})p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)}+\pi_{[k]}p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}+o_P(1)\right\}\\
&=\boldsymbol{\beta}_{[k]}^{*}+o_P(1).
\end{aligned}
\tag{B.21}
$$

此外，根据假设1-3和柯西–施瓦茨不等式，可得 $\left\|\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(a)}\right\|=$

$O(1)$ 对所有 $a=0,1$。于是我们有

$$\left\|\boldsymbol{\beta}_{[k]}^{*}\right\| \leq \left\|\boldsymbol{\Sigma}_{[k]}^{+}\right\| \left\|(1-\pi_{[k]})p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^{*}\widetilde{Y}(1)} + \pi_{[k]}p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^{*}\widetilde{Y}(0)}\right\| = O(1). \tag{B.22}$$

对 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}$ **的分析。** 令 $\boldsymbol{\xi}_{[k]}^{*} := \mathbb{E}\left[\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i) \mid B_i=k\right]$。由 $\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k) = \pi_{n[k]}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)$ 与 $\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i) = \mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i) - \boldsymbol{\xi}_{[k]}^{*}\right\}$，我们有

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\\
=&\frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}\right\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_{[k]}^{*}\right\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i)+\frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)\right\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}+\frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)\right\}
\end{aligned} \tag{B.23}$$

由假设 3，我们有

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)\right\}\\
=&\frac{n_{[k]}}{n}\pi_{n[k]}(1-\pi_{n[k]})\times\\
&\left\{\frac{1}{n_{1[k]}}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)\right\}-\frac{1}{n_{0[k]}}\sum_{i=1}^{n}(1-A_i)\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)\right\}\right\}\\
=&o_P(n^{-1/2}).
\end{aligned}$$

因此我们有

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*} + o_P(n^{-1/2}).$$

注意，$\boldsymbol{\Xi}^*_{i,\ [k]}$，　$i=1$，……，　n在$\mathcal{C}_n$条件下独立，

$$\begin{aligned}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^*_{i,[k]}\right] &= \mathbb{E}_{\mathcal{C}_n}\left[\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\right]\\ &= \left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\mid B_i=k\right]=0\end{aligned}$$

and

$$\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}^*_{i,[k]}\right\|^2\right]=\mathbb{E}_{\mathcal{C}_n}\left[\left\|\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\right\|^2\right]\le\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\right\|^2\mid B_i=k\right]=O(1),$$

其中最后一步由假设3得出。应用引理3可得

$$\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}^*_{i,[k]}=O_P(n^{-1/2}).$$

因此，我们可以得出结论

$$\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}_{i,[k]}=\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}^*_{i,[k]}+o_P(n^{-1/2})=O_P(n^{-1/2}).\tag{B.24}$$

将其与（B.21）和（B.22）结合可得

$$\widehat{\boldsymbol{\beta}}^\top_{[k]}\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}_{i,[k]}=\left\{\boldsymbol{\beta}^*_{[k]}+o_P(1)\right\}^\top\left\{\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}^*_{i,[k]}+o_P(n^{-1/2})\right\}=\boldsymbol{\beta}^{*\top}_{[k]}\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}^*_{i,[k]}+o_P(n^{-1/2})$$

因此，R_1 和 R_2（即（B.10）中的前两项）满足

$$R_1+R_2=\frac{1}{n}\sum_{i=1}^n\left\{\sum_{k=1}^K\left(\widetilde{Y}^*_{i,[k]}-\boldsymbol{\beta}^{*\top}_{[k]}\boldsymbol{\Xi}^*_{i,[k]}\right)\right\}+o_P(n^{-1/2})\tag{B.25}$$

注意到 $\sum_{k=1}^K\left(\widetilde{Y}^*_{i,[k]}-\boldsymbol{\beta}^{*\top}_{[k]}\boldsymbol{\Xi}^*_{i,[k]}\right)$ 在$\mathcal{C}_n$条件下独立，我们将应用引理4。直接计算可得

$$\mathbb{E}_{\mathcal{C}_n}\left[\sum_{k=1}^K\left(\widetilde{Y}^*_{i,[k]}-\boldsymbol{\beta}^{*\top}_{[k]}\boldsymbol{\Xi}^*_{i,[k]}\right)\right]$$

$$
\begin{aligned}
&= \sum_{k=1}^{K} \mathbb{1}(B_i = k) \left\{ \frac{A_i}{\pi_{n[k]}} \mathbb{E}\left[\widetilde{Y}_i(1) \mid B_i = k\right] - \frac{1 - A_i}{1 - \pi_{n[k]}} \mathbb{E}\left[\widetilde{Y}_i(0) \mid B_i = k\right] \right\} \\
&\qquad - \sum_{k=1}^{K} \mathbb{1}(B_i = k)(A_i - \pi_{n[k]}) \boldsymbol{\beta}_{[k]}^{*\top} \mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) \mid B_i = k\right] \\
&= 0,
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Var}\left(\sum_{k=1}^{K} \left(\widetilde{Y}_{i,[k]}^* - \boldsymbol{\beta}_{[k]}^{*\top} \boldsymbol{\Xi}_{i,[k]}^* \right) \mid \mathcal{C}_n \right) = \sum_{k=1}^{K} \mathbb{E}\left[\left(\widetilde{Y}_{i,[k]}^* - \boldsymbol{\beta}_{[k]}^{*\top} \boldsymbol{\Xi}_{i,[k]}^* \right)^2 \mid \mathcal{C}_n \right] \\
&= \sum_{k=1}^{K} \mathbb{1}(B_i = k) \left\{ \frac{A_i}{\pi_{n[k]}^2} \mathbb{E}\left[\widetilde{Y}_i^2(1) \mid B_i = k\right] + \frac{1 - A_i}{(1 - \pi_{n[k]})^2} \mathbb{E}\left[\widetilde{Y}_i^2(0) \mid B_i = k\right] \right\} \\
&\quad - 2 \sum_{k=1}^{K} \mathbb{1}(B_i = k) A_i \frac{1 - \pi_{n[k]}}{\pi_{n[k]}} \mathbb{E}\left[\widetilde{Y}_i(1) \boldsymbol{\beta}_{[k]}^{*\top} \widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) \mid B_i = k\right] \\
&\quad - 2 \sum_{k=1}^{K} \mathbb{1}(B_i = k)(1 - A_i) \frac{\pi_{n[k]}}{1 - \pi_{n[k]}} \mathbb{E}\left[\widetilde{Y}_i(0) \boldsymbol{\beta}_{[k]}^{*\top} \widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) \mid B_i = k\right] \\
&\quad + \sum_{k=1}^{K} \mathbb{1}(B_i = k)(A_i - \pi_{n[k]})^2 \mathbb{E}\left[(\boldsymbol{\beta}_{[k]}^{*\top} \widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i))^2 \mid B_i = k \right] \\
&= \sum_{k=1}^{K} \mathbb{1}(B_i = k) \left\{ \frac{A_i}{\pi_{n[k]}^2} \mathbb{E}\left[\widetilde{Y}_i^2(1) \mid B_i = k\right] + \frac{1 - A_i}{(1 - \pi_{n[k]})^2} \mathbb{E}\left[\widetilde{Y}_i^2(0) \mid B_i = k\right] \right\} \\
&\quad - 2 \sum_{k=1}^{K} \mathbb{1}(B_i = k) \boldsymbol{\beta}_{[k]}^{*\top} \left\{ A_i \frac{1 - \pi_{n[k]}}{\pi_{n[k]}} \boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{Y}(1)} + (1 - A_i) \frac{\pi_{n[k]}}{1 - \pi_{n[k]}} \boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{Y}(0)} \right\} \\
&\quad + \sum_{k=1}^{K} \mathbb{1}(B_i = k)(A_i - \pi_{n[k]})^2 \boldsymbol{\beta}_{[k]}^{*\top} \boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{\boldsymbol{\xi}}_n^*} \boldsymbol{\beta}_{[k]}^*
\end{aligned} \tag{B.26}
$$

and

$$
\begin{aligned}
\sigma_n^2(\mathcal{C}_n) &:= \mathrm{Var}\left(\sum_{i=1}^{n} \sum_{k=1}^{K} \left(\widetilde{Y}_{i,[k]}^* - \boldsymbol{\beta}_{[k]}^{*\top} \boldsymbol{\Xi}_{i,[k]}^* \right) \mid \mathcal{C}_n \right) = \sum_{i=1}^{n} \mathrm{Var}\left(\sum_{k=1}^{K} \left(\widetilde{Y}_{i,[k]}^* - \boldsymbol{\beta}_{[k]}^{*\top} \boldsymbol{\Xi}_{i,[k]}^* \right) \mid \mathcal{C}_n \right) \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbb{1}(B_i = k) \left\{ \frac{A_i}{\pi_{n[k]}^2} \mathbb{E}\left[\widetilde{Y}_i^2(1) \mid B_i = k\right] + \frac{1 - A_i}{(1 - \pi_{n[k]})^2} \mathbb{E}\left[\widetilde{Y}_i^2(0) \mid B_i = k\right] \right\} \\
&\quad - 2 \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbb{1}(B_i = k) \boldsymbol{\beta}_{[k]}^{*\top} \left\{ A_i \frac{1 - \pi_{n[k]}}{\pi_{n[k]}} \boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{Y}(1)} + (1 - A_i) \frac{\pi_{n[k]}}{1 - \pi_{n[k]}} \boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{Y}(0)} \right\} \\
&\quad + \sum_{i=1}^{n} \sum_{k=1}^{K} \mathbb{1}(B_i = k)(A_i - \pi_{n[k]})^2 \boldsymbol{\beta}_{[k]}^{*\top} \boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{\boldsymbol{\xi}}_n^*} \boldsymbol{\beta}_{[k]}^*.
\end{aligned}
$$

根据引理 1 和 $\pi_{n[k]} \xrightarrow{p} \pi_{[k]}$，我们有

$$\begin{aligned}
\frac{\sigma_n^2(\mathcal{C}_n)}{n} &= \sum_{k=1}^{K}\left\{\frac{p_{[k]}}{\pi_{[k]}}\mathbb{E}\left[\widetilde{Y}_i^2(1) \mid B_i = k\right] + \frac{p_{[k]}}{1-\pi_{[k]}}\mathbb{E}\left[\widetilde{Y}_i^2(0) \mid B_i = k\right]\right\} \\
&\quad - 2\sum_{k=1}^{K} p_{[k]}\boldsymbol{\beta}_{[k]}^{*\top}\left\{(1-\pi_{[k]})\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)} + \pi_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}\right\} \\
&\quad + \sum_{k=1}^{K} p_{[k]}\pi_{[k]}(1-\pi_{[k]})\boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\beta}_{[k]}^{*} + o_P(1) \\
&= \varsigma_{\widetilde{Y}}^2 - \varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2 + o_P(1) \qquad \text{(B.27)} \\
&= O_P(1),
\end{aligned}$$

其中第二个等式由 $\boldsymbol{\beta}_{[k]}^*$ 的定义得到。根据（B.26）、（B.27）和假设 3，引理4 中的条件（A.1）得到满足。

还需验证林德伯格条件。根据 Loève 的 c_r 不等式（Davidson 2021，定理 9.32）以及假设1 和 3，我们有

$$\begin{aligned}
&\mathbb{E}\left[\left|\sum_{k=1}^{K}\left(\widetilde{Y}_{i,[k]}^* - \boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^*\right)\right|^{2+\epsilon} \mid \mathcal{C}_n\right] \\
\leq &\frac{(3K)^{1+\epsilon}}{\min_{1\leq k\leq K}\left\{\pi_{n[k]}^{2+\epsilon}, (1-\pi_{n[k]})^{2+\epsilon}\right\}} \times \\
&\left\{2\max_{a\in\{0,1\}}\mathbb{E}\left[\left|\widetilde{Y}_i(a)\right|^{2+\epsilon} \mid B_i = k\right] + \left\|\boldsymbol{\beta}_{[k]}^*\right\|^{2+\epsilon}\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|^{2+\epsilon} \mid B_i = k\right]\right\} \\
= &O_P(1).
\end{aligned}$$

那么，引理4中的林德伯格条件由此可得

$$\begin{aligned}
&\frac{1}{\sigma_n^2(\mathcal{C}_n)}\sum_{i=1}^{n}\mathbb{E}\left[\left\{\sum_{k=1}^{K}\left(\widetilde{Y}_{i,[k]}^* - \boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^*\right)\right\}^2 \mathbb{1}\left\{\left|\sum_{k=1}^{K}\left(\widetilde{Y}_{i,[k]}^* - \boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^*\right)\right| \geq \delta\sigma_n(\mathcal{C}_n)\right\} \mid \mathcal{C}_n\right] \\
\leq &\frac{1}{\sigma_n^2(\mathcal{C}_n)}\sum_{i=1}^{n}\frac{\mathbb{E}\left[\left|\sum_{k=1}^{K}\left(\widetilde{Y}_{i,[k]}^* - \boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^*\right)\right|^{2+\epsilon} \mid \mathcal{C}_n\right]}{\{\delta\sigma_n(\mathcal{C}_n)\}^{\epsilon}} = \frac{1}{\varsigma_{\widetilde{Y}}^2 - \varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2 + o_P(1)}\frac{O_P(1)}{O_P(n^{\epsilon/2})}\frac{1}{\delta^{\epsilon}} = o_P(1)
\end{aligned}$$

对任意 $\delta > 0$，其中第二个不等式由（B.27）得出。应用引理 4 可得

$$\mathbb{E}\left[\left|\mathbb{E}\left[\exp\left\{\mathrm{i}t\frac{\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\widetilde{Y}_{i,[k]}^{*}-\boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right)}{\sigma_n(\mathcal{C}_n)}\right\}\mid\mathcal{C}_n\right]-\exp\left(-t^2/2\right)\right|\right]\to 0$$

当 $n\to\infty$ 对所有 $t\in\mathbb{R}$ 均成立时。

回顾 R_3 和 R_4 在（B.10）中的定义，根据经典中心极限定理，我们有

$$\mathbb{E}\left[\left|\mathbb{E}\left[\exp\left\{\mathrm{i}s\frac{\sqrt{n}(R_3+R_4)}{\varsigma_H}\right\}\mid\mathcal{C}_n\right]-\exp\left(-s^2/2\right)\right|\right]\to 0$$

对任意 $s\in\mathbb{R}$。注意到 R_3+R_4 关于 $\mathcal{C}_n$可测，根据引理 5 我们有

$$\left(\frac{\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\widetilde{Y}_{i,[k]}^{*}-\boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right)}{\sigma_n(\mathcal{C}_n)},\frac{\sqrt{n}(R_3+R_4)}{\varsigma_H}\right)^{\top}\xrightarrow{d}N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right).$$

结合（B.10）和（B.25），我们有

$$\begin{aligned}
&\frac{\sqrt{n}\left(\widehat{\tau}_{\mathrm{cal}}-\tau\right)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}=\frac{\sqrt{n}\left(R_1+R_2\right)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}+\frac{\sqrt{n}\left(R_3+R_4\right)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\\
=&\frac{\sigma_n(\mathcal{C}_n)}{\sqrt{n}\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\widetilde{Y}_{i,[k]}^{*}-\boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right)}{\sigma_n(\mathcal{C}_n)}+\frac{\varsigma_H}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sqrt{n}(R_3+R_4)}{\varsigma_H}+o_P(1)\\
=&\frac{\sqrt{\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}+o_P(1)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\widetilde{Y}_{i,[k]}^{*}-\boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right)}{\sigma_n(\mathcal{C}_n)}+\frac{\varsigma_H}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sqrt{n}(R_3+R_4)}{\varsigma_H}+o_P(1)\\
=&\frac{\sqrt{\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\widetilde{Y}_{i,[k]}^{*}-\boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right)}{\sigma_n(\mathcal{C}_n)}+\frac{\varsigma_H}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sqrt{n}(R_3+R_4)}{\varsigma_H}+o_P(1)\\
\xrightarrow{d}&N(0,1),
\end{aligned}$$

其中，第三个等号由 (B.27) 得出，第四个等号由 $\sum_{i=1}^{n}\sum_{k=1}^{K}(\widetilde{Y}_{i,\ [k]}^{*}-\boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,\ [k]}^{*})/\sigma_n(\mathcal{C}_n)$ 一致紧以及最后一步由引理 6 和 Slutsky定理得出。现在已证明 $\widehat{\tau}_{\mathrm{cal}}$ 的渐近正态性。

方差估计量的一致性。对 $\widehat{\varsigma}^2_{\widetilde{Y}}$ 和 $\widehat{\varsigma}^2_H$ 的一致性可采用与 Bugni 等人 (2018)、Ma 等人 (2022) 类似的方法证明，因此此处省略细节。剩下需要证明 $\widehat{\varsigma}^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}^*_n}$ 的一致性。

由（B.17），可得

$$\widehat{\boldsymbol{\Sigma}}^+_{[k]} = \pi^{-1}_{[k]}(1-\pi_{[k]})^{-1}p^{-1}_{[k]}\boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{\boldsymbol{\xi}}^*_n} + o_P(1),\ \forall 1 \le k \le K.$$

同理，由（B.20），可得

$$\widehat{\boldsymbol{\Gamma}}_{[k]} = (1-\pi_{[k]})p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}(1)} + \pi_{[k]}p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}(0)} + o_P(1),\ \forall 1 \le k \le K.$$

结合 $\left|n_{[k]}/(n_{[k]} - \widehat{r}_{[k]} - 1)\ - 1\right| = o_P(1)$，可得

$$\widehat{\varsigma}^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}^*_n} = \sum_{k=1}^{K}\frac{n_{[k]}}{n_{[k]} - \widehat{r}_{[k]} - 1}\widehat{\boldsymbol{\Gamma}}^\top_{[k]}\widehat{\boldsymbol{\Sigma}}^+_{[k]}\widehat{\boldsymbol{\Gamma}}_{[k]} = \varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}^*_n} + o_P(1).$$

证明完毕。

建立半参数效率。若对每个 $k = 1$，……，K，都存在一个非随机向量 $\boldsymbol{\alpha}_k$ 使得

$$\left\{\sqrt{\frac{1-\pi_{[k]}}{\pi_{[k]}}}\widetilde{h}^*_{1[k]}(\boldsymbol{X}) + \sqrt{\frac{\pi_{[k]}}{1-\pi_{[k]}}}\widetilde{h}^*_{0[k]}(\boldsymbol{X})\right\}\mathbb{1}(B=k) = \boldsymbol{\alpha}^\top_k\left\{\boldsymbol{\xi}^*_n(\boldsymbol{X}) - \mathbb{E}\left[\boldsymbol{\xi}^*_n(\boldsymbol{X}) \mid B=k\right]\right\}\mathbb{1}(B=k),$$

then

$$\begin{aligned}
&\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}^*_n}\\
&= \sum_{k=1}^{K} p_{[k]}\boldsymbol{\alpha}^\top_k\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top \mid B_i=k\right]\left\{\mathrm{Var}\left(\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i) \mid B_i=k\right)\right\}^+\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top \mid B_i=k\right]\boldsymbol{\alpha}_k\\
&= \sum_{k=1}^{K} p_{[k]}\boldsymbol{\alpha}^\top_k\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\mathbb{1}(B_i=k) \mid B_i=k\right]\boldsymbol{\alpha}_k\\
&= \sum_{k=1}^{K} p_{[k]}\mathbb{E}\left[\left\{\sqrt{\frac{1-\pi_{[k]}}{\pi_{[k]}}}\widetilde{h}^*_{1[k]}(\boldsymbol{X}) + \sqrt{\frac{\pi_{[k]}}{1-\pi_{[k]}}}\widetilde{h}^*_{0[k]}(\boldsymbol{X})\right\}^2 \mid B=k\right]
\end{aligned}$$

$$=\varsigma^2_{\widetilde{Y}|\widetilde{h}^*}$$

从而

$$\sqrt{n}\left(\widehat{\tau}_{\text{cal}}-\tau\right)\xrightarrow{d}\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma^2_{\widetilde{Y}|\widetilde{h}^*}.$$

The半参数有效界在Rafi （2023，定理 3.1）中由下式给出

$$V^*=\mathbb{E}\left[\frac{\text{Var}\left(Y(1)\mid\boldsymbol{X},B\right)}{\pi_{[B]}}+\frac{\text{Var}\left(Y(0)\mid\boldsymbol{X},B\right)}{1-\pi_{[B]}}+\left\{h^*_{1[B]}(\boldsymbol{X})-h^*_{0[B]}(\boldsymbol{X})-\tau\right\}^2\right].$$

通过直接计算，我们得到

$$\begin{aligned}
&\mathbb{E}\left[\frac{\text{Var}\left(Y(1)\mid\boldsymbol{X},B\right)}{\pi_{[B]}}+\frac{\text{Var}\left(Y(0)\mid\boldsymbol{X},B\right)}{1-\pi_{[B]}}\right]\\
=&\sum_{k=1}^K p_{[k]}\mathbb{E}\left[\frac{\text{Var}\left(Y(1)\mid\boldsymbol{X},B=k\right)}{\pi_{[k]}}+\frac{\text{Var}\left(Y(0)\mid\boldsymbol{X},B=k\right)}{1-\pi_{[k]}}\mid B=k\right]\\
=&\sum_{k=1}^K\frac{p_{[k]}}{\pi_{[k]}}\left\{\text{Var}\left(Y(1)\mid B=k\right)-\text{Var}\left(\mathbb{E}\left[Y(1)\mid\boldsymbol{X},B=k\right]\right)\right\}\\
&+\sum_{k=1}^K\frac{p_{[k]}}{1-\pi_{[k]}}\left\{\text{Var}\left(Y(0)\mid B=k\right)-\text{Var}\left(\mathbb{E}\left[Y(0)\mid\boldsymbol{X},B=k\right]\right)\right\}\\
=&\varsigma_{\widetilde{Y}}^2-\sum_{k=1}^K\frac{p_{[k]}}{\pi_{[k]}}\text{Var}\left(\widetilde{h}^*_{1[k]}(\boldsymbol{X})\right)-\sum_{k=1}^K\frac{p_{[k]}}{1-\pi_{[k]}}\text{Var}\left(\widetilde{h}^*_{0[k]}(\boldsymbol{X})\right)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\left[\left\{h^*_{1[B]}(\boldsymbol{X})-h^*_{0[B]}(\boldsymbol{X})-\tau\right\}^2\right]\\
=&\sum_{k=1}^K p_{[k]}\mathbb{E}\left[\left\{h^*_{1[k]}(\boldsymbol{X})-h^*_{0[k]}(\boldsymbol{X})-\tau\right\}^2\mid B=k\right]\\
=&\sum_{k=1}^K p_{[k]}\mathbb{E}\left[\left\{\widetilde{h}^*_{1[k]}(\boldsymbol{X})-\widetilde{h}^*_{0[k]}(\boldsymbol{X})+\mathbb{E}\left[h^*_{1[k]}(\boldsymbol{X})\mid B=k\right]-\mathbb{E}\left[h^*_{0[k]}(\boldsymbol{X})\mid B=k\right]-\tau\right\}^2\mid B=k\right]\\
=&\sum_{k=1}^K p_{[k]}\mathbb{E}\left[\left\{\widetilde{h}^*_{1[k]}(\boldsymbol{X})-\widetilde{h}^*_{0[k]}(\boldsymbol{X})\right\}^2\mid B=k\right]\\
&+\sum_{k=1}^K p_{[k]}\left\{\mathbb{E}\left[h^*_{1[k]}(\boldsymbol{X})\mid B=k\right]-\mathbb{E}\left[h^*_{0[k]}(\boldsymbol{X})\mid B=k\right]-\tau\right\}^2\\
=&\sum_{k=1}^K p_{[k]}\left\{\text{Var}\left(\widetilde{h}^*_{1[k]}(\boldsymbol{X})\right)+\text{Var}\left(\widetilde{h}^*_{0[k]}(\boldsymbol{X})\right)-2\mathbb{E}\left[\widetilde{h}^*_{1[k]}(\boldsymbol{X})\widetilde{h}^*_{0[k]}(\boldsymbol{X})\mid B=k\right]\right\}
\end{aligned}$$

$$+\sum_{k=1}^{K} p_{[k]} \left\{\mathbb{E}\left[Y(1) \mid B=k\right]-\mathbb{E}\left[Y(0) \mid B=k\right]-\tau\right\}^2$$
$$=\sum_{k=1}^{K} p_{[k]} \left\{\mathrm{Var}\left(\widetilde{h}^*_{1[k]}(\boldsymbol{X})\right)+\mathrm{Var}\left(\widetilde{h}^*_{0[k]}(\boldsymbol{X})\right)-2\mathbb{E}\left[\widetilde{h}^*_{1[k]}(\boldsymbol{X})\widetilde{h}^*_{0[k]}(\boldsymbol{X}) \mid B=k\right]\right\}+\varsigma_H^2.$$

因此，我们得到

$$\begin{aligned}
V^* &= \mathbb{E}\left[\frac{\mathrm{Var}\left(Y(1) \mid \boldsymbol{X}, B\right)}{\pi_{[B]}}+\frac{\mathrm{Var}\left(Y(0) \mid \boldsymbol{X}, B\right)}{1-\pi_{[B]}}+\left\{h^*_{1[B]}(\boldsymbol{X})-h^*_{0[B]}(\boldsymbol{X})-\tau\right\}^2\right]. \\
&= \varsigma^2_{\widetilde{Y}}-\sum_{k=1}^{K}\frac{p_{[k]}}{\pi_{[k]}}\mathrm{Var}\left(\widetilde{h}^*_{1[k]}(\boldsymbol{X})\right)-\sum_{k=1}^{K}\frac{p_{[k]}}{1-\pi_{[k]}}\mathrm{Var}\left(\widetilde{h}^*_{0[k]}(\boldsymbol{X})\right) \\
&\quad +\sum_{k=1}^{K} p_{[k]} \left\{\mathrm{Var}\left(\widetilde{h}^*_{1[k]}(\boldsymbol{X})\right)+\mathrm{Var}\left(\widetilde{h}^*_{0[k]}(\boldsymbol{X})\right)-2\mathbb{E}\left[\widetilde{h}^*_{1[k]}(\boldsymbol{X})\widetilde{h}^*_{0[k]}(\boldsymbol{X}) \mid B=k\right]\right\}+\varsigma_H^2 \\
&= \varsigma^2_{\widetilde{Y}}+\varsigma_H^2-\varsigma^2_{\widetilde{Y}|\widetilde{h}^*}.
\end{aligned}$$

证明完成。 ■

定理2的证明。根据定理1 ， 我们有

$$\frac{\sqrt{n}\left(\widehat{\tau}_{\mathrm{cal}}(\boldsymbol{\xi}_n)-\tau\right)}{\sqrt{\varsigma_H^2+\varsigma^2_{\widetilde{Y}}-\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}^*_n}}} \xrightarrow{d} N(0,1).$$

注意，假设 3仍然成立，只需将 $\boldsymbol{\xi}_n$ 替换为 $\boldsymbol{\Lambda}\boldsymbol{\xi}_n$。由定理1可知

$$\frac{\sqrt{n}\left(\widehat{\tau}_{\mathrm{cal}}(\boldsymbol{\Lambda}\boldsymbol{\xi}_n)-\tau\right)}{\sqrt{\varsigma_H^2+\varsigma^2_{\widetilde{Y}}-\varsigma^2_{\widetilde{Y}|\boldsymbol{\Lambda}\widetilde{\boldsymbol{\xi}}^*_n}}} \xrightarrow{d} N(0,1).$$

现在，剩下需要证明的是 $\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}^*_n} \geq \varsigma^2_{\widetilde{Y}|\boldsymbol{\Lambda}\widetilde{\boldsymbol{\xi}}^*_n}$。注意，我们有

$$\boldsymbol{\Sigma}_{[k]\boldsymbol{\Lambda}\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}(a)}=\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}(a)} \text{ and } \boldsymbol{\Sigma}^{+}_{[k]\boldsymbol{\Lambda}\widetilde{\boldsymbol{\xi}}^*_n\boldsymbol{\Lambda}\widetilde{\boldsymbol{\xi}}^*_n}=\left\{\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{\boldsymbol{\xi}}^*_n}\boldsymbol{\Lambda}^\top\right\}^{+}.$$

因此，

$$\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}^*_n}-\varsigma^2_{\widetilde{Y}|\boldsymbol{\Lambda}\widetilde{\boldsymbol{\xi}}^*_n}$$

$$= \sum_{k=1}^{K} \frac{p_{[k]}}{\pi_{[k]}(1-\pi_{[k]})} \left\{(1-\pi_{[k]})\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)} + \pi_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}\right\}^\top \times$$
$$\left\{\boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*} - \boldsymbol{\Lambda}^\top\left(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Lambda}^\top\right)^+\boldsymbol{\Lambda}\right\} \times$$
$$\left\{(1-\pi_{[k]})\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)} + \pi_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}\right\}.$$

我们声称

$$\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(a)} \in \text{range}(\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}),\ \forall k = 1,\ldots,K \text{ and } a \in \{0,1\}. \tag{B.28}$$

由于 $\mathbb{R}^d = \text{range}(\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}) \oplus \text{核}(\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*})$，只需证明

$$\boldsymbol{v}^\top\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(a)} = 0,\ \forall \boldsymbol{v} \in \text{kernel}(\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}).$$

对于任意 $\boldsymbol{v} \in \text{核}(\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*})$，我们有

$$\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{v} = 0 \Longrightarrow \mathbb{E}\left[\boldsymbol{v}^\top\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top\boldsymbol{v} \mid B_i = k\right] = 0 \Longrightarrow \mathbb{E}\left[\left(\boldsymbol{v}^\top\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right)^2 \mathbb{1}(B_i = k)\right] = 0,$$

这进一步表明 $\boldsymbol{v}^\top\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\mathbb{1}(B_i = k) = 0$ 几乎必然成立。因此，有

$$\boldsymbol{v}^\top\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(a)} = \mathbb{E}\left[\boldsymbol{v}^\top\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(a) \mid B_i = k\right] = \frac{1}{p_{[k]}}\mathbb{E}\left[\boldsymbol{v}^\top\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\mathbb{1}(B_i = k)\widetilde{Y}_i(a)\right] = 0.$$

因此，(B.28)成立。
现在，为了证明 $\varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*} \geq \varsigma^2_{\widetilde{Y}|\boldsymbol{\Lambda}\widetilde{\boldsymbol{\xi}}_n^*}$，只需证明对于所有 $\boldsymbol{x} \in \text{range}(\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*})$，
$\boldsymbol{x}^\top\left\{\boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*} - \boldsymbol{\Lambda}^\top\left(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Lambda}^\top\right)^+\boldsymbol{\Lambda}\right\}\boldsymbol{x} \geq 0$。由于 $\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}$ 是半正定的，因此存在一个半正定矩阵 $\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}$

$$\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*} = \boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}.$$

对于任意 $\boldsymbol{x} \in \text{range}(\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*})$，存在一个向量 $\boldsymbol{w}$ 使得 $\boldsymbol{x} = \boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{w}$。因此，

$$\boldsymbol{x}^\top\boldsymbol{\Sigma}^{+}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{x} = \boldsymbol{w}^\top\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Sigma}^{+}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{w}.$$

根据奇异值分解，我们有 $\boldsymbol{\Sigma}^{+}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*} = \left(\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\right)^{+}\left(\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\right)^{+}$，由此可得

$$\begin{aligned}\boldsymbol{x}^\top\boldsymbol{\Sigma}^{+}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{x} &= \boldsymbol{w}^\top\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\left(\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\right)^{+}\left(\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\right)^{+}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{w}\\ &= \boldsymbol{w}^\top\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\mathbf{P}^2\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{w}\end{aligned}$$

其中$\mathbf{P}：= \boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\left(\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\right)^{+}$是到range($\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}$)上的投影矩阵，并且我们利用了$\mathbf{P} = \mathbf{P}^\top$这一事实。令 $\boldsymbol{u} = \boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{w} \in$range($\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}$)，则

$$\boldsymbol{x}^\top\boldsymbol{\Sigma}^{+}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{x} = \boldsymbol{u}^\top\mathbf{P}^2\boldsymbol{u} = \boldsymbol{u}^\top\boldsymbol{u}. \tag{B.29}$$

另一方面，我们有

$$\begin{aligned}\boldsymbol{x}^\top\boldsymbol{\Lambda}^\top\left(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Lambda}^\top\right)^{+}\boldsymbol{\Lambda}\boldsymbol{x} &= \boldsymbol{u}^\top\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Lambda}^\top\left(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Lambda}^\top\right)^{+}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{u}\\ &= \boldsymbol{u}^\top\mathbf{Q}^\top\left(\mathbf{Q}\mathbf{Q}^\top\right)^{+}\mathbf{Q}\boldsymbol{u},\end{aligned} \tag{B.30}$$

其中 $\mathrm{Q} := \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}$。注意，$\left\{\mathrm{Q}^\top\left(\mathrm{Q}\mathrm{Q}^\top\right)^{+}\mathrm{Q}\right\}^2 = \mathrm{Q}^\top\left(\mathrm{Q}\mathrm{Q}^\top\right)^{+}\mathrm{Q}$ 且 $\mathrm{Q}^\top\left(\mathrm{Q}\mathrm{Q}^\top\right)^{+}\mathrm{Q}$ 是对称的。我们有 $\mathrm{Q}^\top\left(\mathrm{Q}\mathrm{Q}^\top\right)^{+}\mathrm{Q}$ 是幂等的，因此 $\left\|\mathrm{Q}^\top\left(\mathrm{Q}\mathrm{Q}^\top\right)^{+}\mathrm{Q}\right\| \leq 1$。结合 (B.29) 和 (B.30) 可得

$$\begin{aligned}\boldsymbol{x}^\top\boldsymbol{\Sigma}^{+}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{x} - \boldsymbol{x}^\top\boldsymbol{\Lambda}^\top\left(\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\Lambda}^\top\right)^{+}\boldsymbol{\Lambda}\boldsymbol{x} &= \boldsymbol{u}^\top\boldsymbol{u} - \boldsymbol{u}^\top\mathbf{Q}^\top\left(\mathbf{Q}\mathbf{Q}^\top\right)^{+}\mathbf{Q}\boldsymbol{u}\\ &\geq \|\boldsymbol{u}\|^2 - \|\boldsymbol{u}\|^2 = 0.\end{aligned}$$

由于 $\boldsymbol{x} \in$range($\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}$) 是任意的，证明完成。 ■

B.3 定理证明 3

我们沿用定理证明中的符号约定1。回顾一下 (B.10)，我们有

$$\widehat{\tau}_{\text{cal}} - \tau = R_1 + R_2 + R_3 + R_4. \tag{B.31}$$

我们首先处理 $R_1 + R_{2\circ}$

处理 $R_1 + R_{2\circ}$ 根据 $\epsilon_{i,\ [k]}$ 我们有

$$\begin{aligned}
R_1 + R_2 &= \sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \left\{ \widetilde{Y}^*_{i,[k]} - \widehat{\boldsymbol{\beta}}^{\top}_{[k]} \boldsymbol{\Xi}_{i,[k]} \right\} \\
&= \sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \left\{ \widetilde{Y}^*_{i,[k]} - \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}^{\top}_{i,[k]} \widetilde{Y}_{i,[k]} \right\} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}^{\top}_{i,[k]} \right\}^{+} \boldsymbol{\Xi}_{i,[k]} \right\} \\
&= \sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \left\{ \widetilde{Y}^*_{i,[k]} - \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}^{\top}_{i,[k]} \left(\boldsymbol{\beta}^{\top}_{[k],\mathcal{C}_n} \boldsymbol{\Xi}_{i,[k]} + \epsilon_{i,[k]} \right) \right\} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}^{\top}_{i,[k]} \right\}^{+} \boldsymbol{\Xi}_{i,[k]} \right\} \\
&= \sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \left\{ \widetilde{Y}^*_{i,[k]} - \left\{ \boldsymbol{\beta}^{\top}_{[k],\mathcal{C}_n} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}^{\top}_{i,[k]} + \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}^{\top}_{i,[k]} \epsilon_{i,[k]} \right\} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}^{\top}_{i,[k]} \right\}^{+} \boldsymbol{\Xi}_{i,[k]} \right\} \\
&= \sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \widetilde{Y}^*_{i,[k]} - \sum_{k=1}^{K} \boldsymbol{\beta}^{\top}_{[k],\mathcal{C}_n} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}^{\top}_{i,[k]} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}^{\top}_{i,[k]} \right\}^{+} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \\
&\quad - \sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}^{\top}_{i,[k]} \epsilon_{i,[k]} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}^{\top}_{i,[k]} \right\}^{+} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \\
&= \underbrace{\sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \left\{ \widetilde{Y}^*_{i,[k]} - \boldsymbol{\beta}^{\top}_{[k],\mathcal{C}_n} \boldsymbol{\Xi}_{i,[k]} \right\}}_{Q_2} + \underbrace{- \sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}^{\top}_{i,[k]} \epsilon_{i,[k]} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}^{\top}_{i,[k]} \right\}^{+} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]}}_{Q_1},
\end{aligned} \tag{B.32}$$

其中最后一步由 (B.6) 以及

$$\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^{\top} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^{\top} \right\}^{+} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i.$$

我们分别给出 Q_1 和 Q_2 的界。

为了界定 Q_1，我们推导 $\left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,\ [k]} \boldsymbol{\Xi}^{\top}_{i,\ [k]} \right\}^{+}$ 和 $\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}^{\top}_{i,\ [k]} \epsilon_{i,\ [k]}$

– $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}$ 在 $k=1$、.....、K 上一致成立。在此证明中，$o_P(\cdot)$、$O_P(\cdot)$、$O(1)$ 和 $o(1)$ 所隐藏的随机变量不依赖于 k。对于任意随机向量 $\boldsymbol{Z}_n$ 以及一列子-σ-代数 $\mathcal{C}_n$、$\boldsymbol{Z}_n = O_{L^2,\,\mathcal{C}_n}(a_n)$ 表示当 $n\to\infty$ 时 $\mathbb{E}\left[\|\boldsymbol{Z}_n/a_n\|_2 \,|\mathcal{C}_n\right] = O_P(1)$，其中随机变量 $O_P(1)$ 不依赖于 k，且 a_n 是一列非负数。

$\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\right\}^{+}$。回忆一下

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\\
=&\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top}-\\
&\frac{2}{n}\sum_{i=1}^{n}\pi_{n[k]}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top}+\\
&\frac{1}{n}\sum_{i=1}^{n}\pi_{n[k]}^2\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top}+\\
&\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\left[\left(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}\right)\left(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}\right)^{\top}-\left(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\left(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right)^{\top}\right]}_{R_5},
\end{aligned}
\tag{B.33}
$$

其中，$\overline{\boldsymbol{\xi}}_{n[k]}^*:=\frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)$ 是 $\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)$ 的分层特定样本均值。令 $S^{d-1}:=\{\boldsymbol{\alpha}\in\mathbb{R}^d:\ \|\boldsymbol{\alpha}\|=1\}$ 表示 $\mathbb{R}^d$ 中所有单位向量构成的集合。对于每个$1\le k\le K$ 和 $\boldsymbol{\alpha}\in\mathbb{R}^d$，根据柯西–施瓦茨不等式，我们有

$$
\begin{aligned}
&\sup_{\boldsymbol{\alpha}\in S^{d-1}}\left|\boldsymbol{\alpha}^{\top}R_5\boldsymbol{\alpha}\right|\\
=&\sup_{\boldsymbol{\alpha}\in S^{d-1}}\left|\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\left[\left(\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}\right)^2-\left(\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}^*\right)^2\right]\right|\\
\le&\sup_{\boldsymbol{\alpha}\in S^{d-1}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\{\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}-\left(\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\}^2}\times\\
&\qquad\sup_{\boldsymbol{\alpha}\in S^{d-1}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\{\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}+\left(\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\}^2}\\
\le&\sqrt{\underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\left(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\|^2}_{R_6}}\times
\end{aligned}
$$

$$\sqrt{\underbrace{\sup_{\boldsymbol{\alpha}\in S^{d-1}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\{\boldsymbol{\alpha}^\top\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\alpha}^\top\overline{\boldsymbol{\xi}}_{n[k]}+\left(\boldsymbol{\alpha}^\top\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\boldsymbol{\alpha}^\top\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\}^2}_{R_7}},$$

由此可得

$$\|R_5\|\le\sqrt{R_6}\times\sqrt{R_7}.$$

我们分别控制 R_6 和 R_7。根据假设4，我们有

$$\begin{aligned}\left\|\overline{\boldsymbol{\xi}}_{n[k]}-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\|&=\left\|\frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}\right\|\\&\le\sqrt{\frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^2}=O_P(n^{-1/4})\end{aligned}\tag{B.34}$$

在 $k=1,\ \ldots K$ 上一致成立。于是我们有

$$\begin{aligned}R_6&\le 2\times\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^2+2\times\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\overline{\boldsymbol{\xi}}_{n[k]}-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\|^2\\&\le\frac{n_{[k]}}{n}\frac{2}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\|^2+\frac{n_{[k]}}{n}\times 2\left\|\overline{\boldsymbol{\xi}}_{n[k]}-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\|^2\\&=\frac{n_{[k]}}{n}O_P(n^{-1/2})=p_{[k]}n^{-1/2}O_P(1),\end{aligned}\tag{B.35}$$

其中最后一步由引理1得出。对于 R_7，我们注意到

$$\begin{aligned}R_7&\le 2\sup_{\boldsymbol{\alpha}\in S^{d-1}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\{\boldsymbol{\alpha}^\top\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\alpha}^\top\overline{\boldsymbol{\xi}}_{n[k]}-\left(\boldsymbol{\alpha}^\top\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\boldsymbol{\alpha}^\top\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\}^2\\&\quad+2\sup_{\boldsymbol{\alpha}\in S^{d-1}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\{2\left(\boldsymbol{\alpha}^\top\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\boldsymbol{\alpha}^\top\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\}^2\\&\le 2R_6+16\sup_{\boldsymbol{\alpha}\in S^{d-1}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\{\boldsymbol{\alpha}^\top\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^2+16\sup_{\boldsymbol{\alpha}\in S^{d-1}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\{\boldsymbol{\alpha}^\top\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}^2\\&\le 2R_6+16\underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\overline{\boldsymbol{\xi}}_{n[k]}^*\right\|^2}_{R_{7,1}}+16\underbrace{\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)^\top\right\|}_{R_{7,2}}.\end{aligned}$$

我们处理 $R_{7,\ 1}$。由引理 2 可得

$$\left\|\frac{n_{[k]}}{n}\overline{\boldsymbol{\xi}}^*_{n[k]} - \mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right]\right\| = \left\|\frac{1}{n}\sum_{i=1}^n\left\{\mathbb{1}(B_i=k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right]\right\}\right\|$$
$$= \sqrt{\frac{p_{[k]}r_n\log(2Kd)}{n}}O_P(1). \tag{B.36}$$

另一方面，根据假设 4 我们有

$$\begin{aligned}
&\left\|\mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right]\right\| \\
=&p_{[k]}\sup_{\boldsymbol{\alpha}\in S^{d-1}}\left|\mathbb{E}\left[\boldsymbol{\alpha}^\top\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\mid B_i=k\right]\right| \le p_{[k]}\sup_{\boldsymbol{\alpha}\in S^{d-1}}\sqrt{\mathbb{E}\left[\{\boldsymbol{\alpha}^\top\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\}^2\mid B_i=k\right]} \\
=&p_{[k]}\sup_{\boldsymbol{\alpha}\in S^{d-1}}\sqrt{p_{[k]}\boldsymbol{\alpha}^\top\mathbb{E}\left[\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)^\top\mid B_i=k\right]\boldsymbol{\alpha}} \le p_{[k]}\sqrt{C} = p_{[k]}O(1).
\end{aligned} \tag{B.37}$$

因此，由引理 1 可得

$$\left\|\overline{\boldsymbol{\xi}}^*_{n[k]}\right\| = \frac{n}{n_{[k]}}\left\{p_{[k]}O(1) + \sqrt{\frac{p_{[k]}r_n\log(2Kd)}{n}}O_P(1)\right\} = O_P(1) + \sqrt{\log(2Kd)r_n/(np_{[k]})}O_P(1) = O_P(1),$$

其中，最后一步由 $\log(2Kd)r_n/(n\inf^{1\le k\le K}p_{[k]}) \to 0$ 因为 $n\to\infty$。回顾 $R_{7,\ 1}$的定义，则由引理 1 可知

$$R_{7,1} = \frac{1}{n}\sum_{i=1}^n\mathbb{1}(B_i=k)\left\|\overline{\boldsymbol{\xi}}^*_{n[k]}\right\|^2 = \left\{p_{[k]}+p_{[k]}o_P(1)\right\}\left\{O_P(1)\right\}^2 = p_{[k]}O_P(1).$$

由引理 2 可得

$$\left\|\frac{1}{n}\sum_{i=1}^n\mathbb{1}(B_i=k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)^\top - \mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)^\top\right]\right\| \le p_{[k]}o_P(1).$$

因此，

$$\begin{aligned}
R_{7,2} &\le \left\|\mathbb{E}\left[\mathbb{1}(B_i=k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)^\top\right]\right\| + p_{[k]}o_P(1) \\
&= p_{[k]}O(1) + p_{[k]}o_P(1) = p_{[k]}O_P(1).
\end{aligned}$$

结合（B.35）以及关于 $_{,\ 1}$ 和 $R_{7,\ 2}$ 的结果可得

$$R_7 \leq 2R_6 + 16R_{7,1} + 16R_{7,2} = p_{[k]}n^{-1/2}O_P(1) + p_{[k]}O_P(1) = p_{[k]}O_P(1).$$

因此，我们有

$$\|R_5\| \leq \sqrt{R_6} \times \sqrt{R_7} = p_{[k]}n^{-1/4}O_P(1). \tag{B.38}$$

回顾（B.33），我们有

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top} \\
=&\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top} \\
&-\frac{2}{n}\sum_{i=1}^{n}\pi_{n[k]}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top} \\
&+\frac{1}{n}\sum_{i=1}^{n}\pi_{n[k]}^2\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top}+p_{[k]}n^{-1/4}O_P(1) \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\left\{\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)-\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\right\}\left\{\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)-\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\right\}^{\top}+p_{[k]}n^{-1/4}O_P(1) \\
=&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\boldsymbol{\Xi}_{i,[k]}^{*\top}-2\times\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\times\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^{*\top}(\boldsymbol{X}_i) \\
&+\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\times\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^{*\top}\times\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)+p_{[k]}n^{-1/4}O_P(1)
\end{aligned} \tag{B.39}$$

其中 $\widetilde{\boldsymbol{\xi}}_{n[k]}^*:=\frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)$。由（B.36）可得，

$$\begin{aligned}
\|\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]\| &= \sup_{\boldsymbol{\alpha}\in S^{d-1}}\left|\mathbb{E}\left[\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]\right| \\
\leq \sup_{\boldsymbol{\alpha}\in S^{d-1}}&\sqrt{\mathbb{E}\left[\{\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\}^2\mid B_i=k\right]}\leq C
\end{aligned}$$

以及 $\frac{n_{[k]}}{n}=p_{[k]}+\sqrt{p_{[k]}}O_P(\sqrt{\log(2K)/n})$，由此可知

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|=\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\frac{n_{[k]}}{n}\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]\right\|$$

$$\leq \left\| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(B_i = k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i) - p_{[k]}\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i) \mid B_i = k\right] \right\| + \left| \frac{n_{[k]}}{n} - p_{[k]} \right| \left\| \mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i) \mid B_i = k\right] \right\|$$
$$= \sqrt{\frac{p_{[k]} r_n \log(2Kd)}{n}} O_P(1) + \sqrt{\frac{p_{[k]} \log(2K)}{n}} O_P(1). \tag{B.40}$$

因此，根据引理1 可知

$$\left\| \widetilde{\boldsymbol{\xi}}_{n[k]}^* \right\| = \frac{n}{n_{[k]}} \left\| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) \right\| = p_{[k]}^{-1}(1 + o_P(1)) \left\| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) \right\|$$
$$= \sqrt{r_n \log(2Kd)/(np_{[k]})} O_P(1) + \sqrt{\log(2K)/(np_{[k]})} O_P(1). \tag{B.41}$$

此外，由引理2 我们有

$$\left\| \frac{1}{n}\sum_{i=1}^{n} A_i \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) \right\| \leq \sqrt{\frac{p_{[k]} r_n \log(2Kd)}{n}} O_P(1).$$

则由假设2 和（B.40）可得

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \left(A_i - \pi_{n[k]}\right)^2 \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) \right\|$$
$$= \left\| \frac{1}{n}\sum_{i=1}^{n} A_i \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) - 2\pi_{n[k]} \frac{1}{n}\sum_{i=1}^{n} A_i \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) + \pi_{n[k]}^2 \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(B_i = k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) \right\|$$
$$= \sqrt{\frac{p_{[k]} r_n \log(2Kd)}{n}} O_P(1) + \sqrt{\frac{p_{[k]} r_n \log(2Kd)}{n}} O_P(1) + \sqrt{\frac{p_{[k]} \log(2K)}{n}} O_P(1)$$
$$= \sqrt{\frac{p_{[k]} r_n \log(2Kd)}{n}} O_P(1) + \sqrt{\frac{p_{[k]} \log(2K)}{n}} O_P(1).$$

这一点结合（B.39）、（B.41）、（B.38）以及

$$\frac{1}{n}\sum_{i=1}^{n} \left(A_i - \pi_{n[k]}\right)^2 \mathbb{1}(B_i = k) \leq \frac{n_{[k]}}{n} = p_{[k]}\left(1 + o_P(1)\right) = p_{[k]} O_P(1)$$

意味着

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}_{i,[k]}^\top - \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]}^* \boldsymbol{\Xi}_{i,[k]}^{*\top} \right\|$$

$$
\begin{aligned}
&= \left\{ \sqrt{\frac{p_{[k]} r_n \log(2Kd)}{n}} + \sqrt{\frac{p_{[k]} \log(2K)}{n}} \right\} \left\{ \sqrt{\frac{r_n \log(2Kd)}{np_{[k]}}} + \sqrt{\frac{\log(2K)}{np_{[k]}}} \right\} O_P(1) \\
&\quad + \left\{ \frac{r_n \log(2Kd)}{np_{[k]}} + \frac{\log(2K)}{np_{[k]}} \right\} p_{[k]} O_P(1) + p_{[k]} n^{-1/4} O_P(1) \\
&= \left\{ \frac{r_n \log(2Kd)}{n} + \frac{\log(2K)}{n} \right\} O_P(1) + p_{[k]} n^{-1/4} O_P(1). \qquad \text{(B.42)}
\end{aligned}
$$

根据引理 2 我们有

$$
\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}^*_{i,[k]} \boldsymbol{\Xi}^{*\top}_{i,[k]} - \boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} \right\| = p_{[k]} o_P(1). \qquad \text{(B.43)}
$$

此外，根据 $\left\| \mathbb{E}\left[\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i) \widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top \mid B_i = k \right] \right\| \leq C < \infty$ 和引理 1 可得

$$
\left\| \boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} - \boldsymbol{\Sigma}_{[k]} \right\| = p_{[k]} o_P(1). \qquad \text{(B.44)}
$$

结合（B.42）、（B.43）和（B.44）可得

$$
\begin{aligned}
&\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}^{\top}_{i,[k]} - \boldsymbol{\Sigma}_{[k]} \right\| \\
&= \left\{ \frac{r_n \log(2Kd)}{n} + \frac{\log(2K)}{n} \right\} O_P(1) + p_{[k]} n^{-1/4} O_P(1) + p_{[k]} o_P(1). \qquad \text{(B.45)}
\end{aligned}
$$

B根据假设 4 以及摩尔-彭若斯伪逆的推导，我们有

$$
\left\| \boldsymbol{\Sigma}^{+}_{[k]\widetilde{\boldsymbol{\xi}}^*_n \widetilde{\boldsymbol{\xi}}^*_n} \right\| \leq c^{-1}
$$

在以下各项上一致成立： $k = 1, \ldots\ldots, K$ 以及 $n \geq 1$。然后根据假设2，我们有

$$
\left\| \boldsymbol{\Sigma}^{+}_{[k]} \right\| = \left\| \pi^{-1}_{[k]} (1 - \pi_{[k]})^{-1} p^{-1}_{[k]} \boldsymbol{\Sigma}^{+}_{[k]\widetilde{\boldsymbol{\xi}}^*_n \widetilde{\boldsymbol{\xi}}^*_n} \right\| \leq p^{-1}_{[k]} O(1). \qquad \text{(B.46)}
$$

由Stewart (1977) 中的定理4 和引理1 可知

$$
\begin{aligned}
&\left\|\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\right\}^{+}-\boldsymbol{\Sigma}_{[k]}^{+}\right\| \\
\leq & 3\max\left\{\left\|\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\right\}^{+}\right\|^{2},\left\|\boldsymbol{\Sigma}_{[k]}^{+}\right\|^{2}\right\}\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}-\boldsymbol{\Sigma}_{[k]}\right\| \\
= & \max\left\{p_{[k]}^{-2}O(1),\left(\frac{n}{n_{[k]}}\right)^{2}O_P(1)\right\}\times\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}-\boldsymbol{\Sigma}_{[k]}\right\| \\
= & p_{[k]}^{-1}\left\{\left\{\frac{r_n\log(2Kd)}{np_{[k]}}+\frac{\log(2K)}{np_{[k]}}\right\}O_P(1)+n^{-1/4}O_P(1)+o_P(1)\right\} \qquad \text{(B.47)} \\
= & p_{[k]}^{-1}o_P(1). \qquad \text{(B.48)}
\end{aligned}
$$

其中最后一步由以及 $r_n\log(2Kd)/(n\inf_{1\leq k\leq K}p_{[k]})\to 0$ 当 $n\to\infty$ 时得出。

对 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\epsilon_{i,[k]}$ **的分析**。我们有

$$
\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\epsilon_{i,[k]}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\widetilde{Y}_{i,[k]}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}.
$$

我们逐一处理 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\widetilde{Y}_{i,[k]}$ 和 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}$。

对 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\widetilde{Y}_{i,[k]}$ **的分析如下。我们有以下分解：**

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{1[k]}) \\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})Y_i(1) \\
&-\overline{Y}_{1[k]}\times\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}) \\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^{*})(Y_i(1)-\overline{Y}_{1[k]}) \\
&\underbrace{+\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\left(\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^{*}\right)\right\}Y_i}_{R_8} \\
&\underbrace{-\overline{Y}_{1[k]}\times\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-(\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^{*})\right\}}_{R_9}.
\end{aligned}
$$

我们分别控制 R_8 和 R_9。注意，根据假设 1 和引理 1 ，我们有

$$\mathbb{E}_{\mathcal{C}_n}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)Y_i^2\right]=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\mathbb{E}\left[Y_i^2\mid B_i=k\right]=\frac{n_{[k]}}{n}O(1)=p_{[k]}O_P(1).$$

由柯西-施瓦茨不等式、假设1 和（B.35）可得

$$\begin{aligned}
&|R_8|\\
\leq&\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)Y_i^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\left(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\|^2}\\
=&O_P(1)\sqrt{p_{[k]}}O_{L^2,\mathcal{C}_n}(1)\times\sqrt{R_6}=O_P(1)\sqrt{p_{[k]}}O_{L^2,\mathcal{C}_n}(1)\times\sqrt{p_{[k]}}n^{-1/4}O_P(1)\\
=&p_{[k]}n^{-1/4}O_P(1)O_{L^2,\mathcal{C}_n}(1).
\end{aligned}$$

同理，注意到

$$\begin{aligned}
\left|\overline{Y}_{1[k]}\right|&=\frac{n}{n_{[k]}}\frac{n_{[k]}}{n_{1[k]}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)A_iY_i(1)\right|\leq\frac{n}{n_{[k]}}\frac{n_{[k]}}{n_{1[k]}}\sqrt{\frac{n_{[k]}}{n}}\sqrt{p_{[k]}}O_{L^2,\mathcal{C}_n}(1)\\
&=\sqrt{p_{[k]}^{-1}(1+o_P(1))}\left\{\frac{1}{\pi_{[k]}}+o_P(1)\right\}\sqrt{p_{[k]}}O_{L^2,\mathcal{C}_n}(1)=O_P(1)O_{L^2,\mathcal{C}_n}(1)
\end{aligned}$$

我们有

$$|R_9|\leq O_P(1)O_{L^2,\mathcal{C}_n}(1)\times\sqrt{\frac{n_{[k]}}{n}}\times\sqrt{R_6}=p_{[k]}n^{-1/4}O_P(1)O_{L^2,\mathcal{C}_n}(1).$$

结合 R_8 和 R_9的结果，我们得到

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{1[k]})\\
=&\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\times\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*)(Y_i(1)-\overline{Y}_{1[k]})+p_{[k]}n^{-1/4}O_P(1)O_{L^2,\mathcal{C}_n}(1)\\
=&\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\times\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)A_i(\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)-\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*)(\widetilde{Y}_i(1)-\overline{\widetilde{Y}}_{1[k]})+p_{[k]}n^{-1/4}O_P(1)O_{L^2,\mathcal{C}_n}(1)\\
=&\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\times\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)A_i\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(1)-\overline{\widetilde{Y}}_{1[k]}\times\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)A_i\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)
\end{aligned}$$

$$- \widetilde{\overline{\xi}}^*_{n[k]} \frac{1-\pi_{n[k]}}{\pi_{n[k]}} \frac{1}{n} \sum_{i=1}^n \mathbb{1}(B_i = k) A_i \widetilde{Y}_i(1) + \widetilde{\overline{\xi}}^*_{n[k]} \overline{\widetilde{Y}}_{1[k]} \frac{1-\pi_{n[k]}}{\pi_{n[k]}} \frac{1}{n} \sum_{i=1}^n \mathbb{1}(B_i = k)$$

$$+ p_{[k]} n^{-1/4} O_P(1) O_{L^2, \mathcal{C}_n}(1), \tag{B.49}$$

其中 $\widetilde{\xi}^*_{n[k]} := \frac{1}{n_{[k]}} \sum_{i=1}^n \mathbb{1}(B_i = k) \widetilde{\xi}^*_n(\boldsymbol{X}_i) = \xi^*_{n[k]} - \mathbb{E}\left[\xi^*_n(\boldsymbol{X}_i) \mid B_i = k\right]$ 和 $\widetilde{Y}_{1[k]} := \frac{1}{n_{1[k]}} \sum_{i=1}^n \mathbb{1}(B_i = k) A_i \widetilde{Y}_i(1) = Y_{1[k]} - \mathbb{E}\left[Y_i(1) \mid B_i = k\right]$。注意，

$$\mathbb{E}_{\mathcal{C}_n}\left[\mathbb{1}(B_i = k) A_i \widetilde{Y}_i(1)\right] = \mathbb{1}(B_i = k) A_i \mathbb{E}\left[\widetilde{Y}_i(1) \mid B_i = k\right] = 0$$

and

$$\begin{aligned}\mathbb{E}_{\mathcal{C}_n}\left[\left\{\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i = k) A_i \widetilde{Y}_i(1)\right\}^2\right] &= \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}_{\mathcal{C}_n}\left[\mathbb{1}(B_i = k) A_i \widetilde{Y}_i(1)^2\right] \\ &\leq \frac{1}{n}\frac{n_{[k]}}{n}\mathbb{E}\left[\widetilde{Y}_i(1)^2 \mid B_i = k\right] = \frac{1}{n} p_{[k]} O_P(1)\end{aligned}$$

我们有

$$\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i = k) A_i \widetilde{Y}_i(1) = \sqrt{p_{[k]}/n}\, O_{L^2, \mathcal{C}_n}(1)$$

从而

$$\begin{aligned}\overline{\widetilde{Y}}_{1[k]} &= \frac{n}{n_{1[k]}}\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i = k) A_i \widetilde{Y}_i(1) = \frac{n}{n_{1[k]}}\sqrt{p_{[k]}/n}\, O_{L^2, \mathcal{C}_n}(1) = \frac{n}{n_{[k]}}\frac{n_{[k]}}{n_{1[k]}}\sqrt{p_{[k]}/n}\, O_{L^2, \mathcal{C}_n}(1) \\ &= p_{[k]}^{-1}(1 + o_P(1)) O_P(1)\sqrt{p_{[k]}/n}\, O_{L^2, \mathcal{C}_n}(1) = \sqrt{1/(np_{[k]})}\, O_P(1) O_{L^2, \mathcal{C}_n}(1),\end{aligned}$$

w此处第四步由假设2。根据引理2，我们有

$$\frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i = k) A_i \widetilde{\xi}^*_n(\boldsymbol{X}_i) = \sqrt{\frac{r_n p_{[k]} \log(2Kd)}{n}} O_P(1).$$

结合 上述结果结合（B.41）、（B.49）、假设2 和引理1，可得

$$\left\| \frac{1}{n}\sum_{i=1}^n \frac{1-\pi_{n[k]}}{\pi_{n[k]}} \mathbb{1}(B_i = k) A_i (\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]})(Y_i - \overline{Y}_{1[k]}) - \frac{1-\pi_{n[k]}}{\pi_{n[k]}} \times \frac{1}{n}\sum_{i=1}^n \mathbb{1}(B_i = k) A_i \widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i) \widetilde{Y}_i(1) \right\|$$
$$\leq \frac{\sqrt{r_n \log(2Kd)}}{n} O_P(1) O_{L^2, \mathcal{C}_n}(1) + p_{[k]} n^{-1/4} O_P(1) O_{L^2, \mathcal{C}_n}(1).$$

同理，我们也有

$$\begin{aligned}
&\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)(1-A_i)(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{0[k]})-\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)(1-A_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(0)\right|\\
\leq&\frac{\sqrt{r_n\log(2Kd)}}{n}O_P(1)O_{L^2,\mathcal{C}_n}(1)+p_{[k]}n^{-1/4}O_P(1)O_{L^2,\mathcal{C}_n}(1).
\end{aligned}$$

注意到

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\widetilde{Y}_{i,[k]}\\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{1[k]})\\
&+\frac{1}{n}\sum_{i=1}^{n}\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\mathbb{1}(B_i=k)(1-A_i)(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{0[k]}).
\end{aligned}$$

我们有

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\widetilde{Y}_{i,[k]}\\
=&\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\times\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)A_i\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(1)\\
&+\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\times\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)(1-A_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(0)\\
&+\frac{\sqrt{r_n\log(2Kd)}}{n}O_P(1)O_{L^2}(1)+p_{[k]}n^{-1/4}O_P(1)O_{L^2}(1)\\
=&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\widetilde{Y}_{i,[k]}^*+\frac{\sqrt{r_n\log(2Kd)}}{n}O_P(1)O_{L^2,\mathcal{C}_n}(1)+p_{[k]}n^{-1/4}O_P(1)O_{L^2,\mathcal{C}_n}(1). \qquad \text{(B.50)}
\end{aligned}$$

对$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,-[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}$**的分析**。根据 $\boldsymbol{\beta}_{[k],\mathcal{C}_n}$ 的定义，我们有

$$\boldsymbol{\beta}_{[k],\mathcal{C}_n}=\left\{\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right\}^{+}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\widetilde{Y}_{i,[k]}^*\boldsymbol{\Xi}_{i,[k]}^*\right]\right\}.$$

注意

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\widetilde{Y}_{i,[k]}^*\boldsymbol{\Xi}_{i,[k]}^*\right]$$

$$=\frac{1-\pi_{n[k]}}{\pi_{n[k]}} \times \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(B_i=k) A_i \boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{Y}(1)} + \frac{\pi_{n[k]}}{1-\pi_{n[k]}} \times \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(B_i=k)(1-A_i) \boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{Y}(0)}.$$

对于任意 $a \in \{0,1\}$，根据假设 4 ，我们有

$$\begin{aligned}
\left\|\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{Y}(a)}\right\| &= \left\|\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}(\boldsymbol{X}_i)\widetilde{Y}_i(a) \mid B_i=k\right]\right\| = \sup_{\boldsymbol{\alpha}\in S^{d-1}} \left|\mathbb{E}\left[\boldsymbol{\alpha}^\top \widetilde{\boldsymbol{\xi}}(\boldsymbol{X}_i)\widetilde{Y}_i(a) \mid B_i=k\right]\right| \\
&\leq \sqrt{\mathbb{E}\left[\left\{\boldsymbol{\alpha}^\top \widetilde{\boldsymbol{\xi}}(\boldsymbol{X}_i)\right\}^2 \mid B_i=k\right] \mathbb{E}\left[\widetilde{Y}_i(a)^2 \mid B_i=k\right]} \\
&\leq \sqrt{\mathbb{E}\left[\widetilde{Y}_i(a)^2 \mid B_i=k\right]} \times \sqrt{\boldsymbol{\alpha}^\top \mathbb{E}\left[\widetilde{\boldsymbol{\xi}}(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}(\boldsymbol{X}_i)^\top \mid B_i=k\right]\boldsymbol{\alpha}} = O(1). \quad \text{(B.51)}
\end{aligned}$$

然后，应用引理1 可得

$$\left\|\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\mathcal{C}_n}\left[\widetilde{Y}_{i,[k]}^* \boldsymbol{\Xi}_{i,[k]}^*\right] - p_{[k]}(1-\pi_{[k]})\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{Y}(1)} - p_{[k]}\pi_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^* \widetilde{Y}(0)}\right\| = p_{[k]} o_P(1). \quad \text{(B.52)}$$

此外，根据引理1 和（B.46）可得

$$\begin{aligned}
\left\{\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right\}^+ &= \left\{\frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}^2 \mathbb{1}(B_i=k)\right\}^{-1} \boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*} \\
&= \left\{p_{[k]}\pi_{[k]}(1-\pi_{[k]}) + p_{[k]}o_P(1)\right\}^{-1} \boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*} \\
&= p_{[k]}^{-1}\left\{\pi_{[k]}^{-1}(1-\pi_{[k]}^{-1}) + o_P(1)\right\} \boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*} \\
&= \boldsymbol{\Sigma}^+_{[k]} + p_{[k]}^{-1}\boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*} o_P(1) = \boldsymbol{\Sigma}^+_{[k]} + p_{[k]}^{-1} o_P(1).
\end{aligned}$$

结合（B.51）和（B.52）可得

$$\begin{aligned}
\boldsymbol{\beta}_{[k],\mathcal{C}_n} &= \left\{\left\{\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right\}^+ + p_{[k]}^{-1}o_P(1)\right\}\left\{p_{[k]}(1-\pi_{[k]})\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)} + p_{[k]}\pi_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)} + p_{[k]}o_P(1)\right\} \\
&= \boldsymbol{\beta}^*_{[k]} + o_P(1) \quad \text{(B.53)} \\
&= O_P(1) \quad \text{(B.54)}
\end{aligned}$$

由（B.42）可知

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right\|\\=&\left\{\frac{r_n\log(2Kd)}{n}+\frac{\log(2K)}{n}\right\}O_P(1)+p_{[k]}n^{-1/4}O_P(1).\end{aligned}$$

回顾（B.50），于是我们得到

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\epsilon_{i,[k]}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}\right\|=\frac{r_n\log(2Kd)}{n}O_P(1)O_{L^2,\mathcal{C}_n}(1)+p_{[k]}n^{-1/4}O_P(1)O_{L^2,\mathcal{C}_n}(1).\tag{B.55}$$

注意,

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}\right]\right\|^2\right]\leq\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}\right\|^2\right]\\\leq&2\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\widetilde{Y}_{i,[k]}^{*}\right\|^2\right]+2\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right\|^2\right]\\\leq&4\max\left\{\frac{1-\pi_{n[k]}}{\pi_{n[k]}},\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\right\}^2\mathbb{1}(B_i=k)\max_{a=0,1}\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i)\widetilde{Y}_i(a)\right\|^2\mid B_i=k\right]\\&+2\mathbb{E}_{\mathcal{C}_n}\left[\left\|\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i)^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right\|^2\right]\mathbb{1}(B_i=k)\\\leq&\zeta_n^2O_P(1)\max_{a=0,1}\mathbb{E}\left[\left\|\widetilde{Y}_i(a)\right\|^2\mid B_i=k\right]\mathbb{1}(B_i=k)\\&+\zeta_n^2O_P(1)\boldsymbol{\beta}_{[k],\mathcal{C}_n}^{\top}\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i)^{\top}\mid B_i=k\right]\boldsymbol{\beta}_{[k],\mathcal{C}_n}\mathbb{1}(B_i=k)\\=&\mathbb{1}(B_i=k)\zeta_n^2O_P(1),\end{aligned}$$

其中 $O_P(1)$ 既不依赖于 k，也不依赖于 i，最后一步由假设 1 和 4以及（B.54）得出。因此，根据引理 1可知

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}_n}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}\right]\right\}\right\|^2\right]\\=&\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}\right]\right\|^2\right]\\\leq&\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\zeta_n^2O_P(1)=\frac{1}{n}\zeta_n^2O_P(1)p_{[k]}.\end{aligned}$$

因此，

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}\right]\right\}\right\|=\sqrt{\zeta_n^2 p_{[k]}/n}O_P(1)O_{L^2,\mathcal{C}_n}(1). \tag{B.56}$$

由于

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\widetilde{Y}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*}\right]\\=&\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\times\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)A_i\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^{*}\widetilde{Y}(1)}+\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\times\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)(1-A_i)\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^{*}\widetilde{Y}(0)}\end{aligned}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]=\frac{1}{n}\sum_{i=1}^{n}(A_i-\pi_{n[k]})^2\mathbb{1}(B_i=k)\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^{*}\widetilde{\boldsymbol{\xi}}_n^{*}}^{+},$$

根据（B.28），我们有

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\widetilde{Y}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*}\right]\in\text{range}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\right).$$

利用摩尔-彭若斯伪逆的性质，我们有

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\right\}^{+}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\widetilde{Y}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*}\right]\right\}=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\widetilde{Y}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*}\right].$$

因此，我们有

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}\right]=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\widetilde{Y}_{i,[k]}^{*}\right]-\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\boldsymbol{\beta}_{[k],\mathcal{C}_n}\\=&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\widetilde{Y}_{i,[k]}^{*}\right]-\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\right\}^{+}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\widetilde{Y}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*}\right]\right\}\\=&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\widetilde{Y}_{i,[k]}^{*}\right]-\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\widetilde{Y}_{i,[k]}^{*}\right]=0.\end{aligned} \tag{B.57}$$

结合（B.56）可得

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}^{*}_{i,[k]}\epsilon^{*}_{i,[k]}\right\| = \sqrt{\zeta_n^2 p_{[k]}/n}O_P(1)O_{L^2,\mathcal{C}_n}(1).$$

回顾（B.55），我们有

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\epsilon_{i,[k]}\right\| = \left\{\frac{r_n\log(2Kd)}{n} + p_{[k]}n^{-1/4} + \sqrt{\frac{\zeta_n^2 p_{[k]}}{n}}\right\}O_P(1)O_{L^2,\mathcal{C}_n}(1). \tag{B.58}$$

对$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}$**的分析**。对每个 $k=1,\ldots\ldots,K$，我们令 $\boldsymbol{\xi}^{*}_{[k]} := \mathbb{E}\left[\boldsymbol{\xi}^{*}_{n}(\boldsymbol{X}_i)\,|\,B_i=k\right]$。由 $\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k) = \pi_{n[k]}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)$ 与 $\mathbf{1}(B_i=k)\widetilde{\boldsymbol{\xi}}^{*}_{n}(\boldsymbol{X}_i) = \mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}^{*}_{n}(\boldsymbol{X}_i) - \boldsymbol{\xi}^{*}_{[k]}\right\}$，我们有

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]} &= \frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}\right\}\\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}^{*}_{[k]}\right\}\\
&= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}^{*}_{i,[k]} + \frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}^{*}_{n}(\boldsymbol{X}_i)\right\}.
\end{aligned}$$

由假设 4 与引理 1，我们有

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{i=1}^{n}\left\{A_i-\pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}^{*}_{n}(\boldsymbol{X}_i)\right\}\right\|\\
=&\frac{n_{[k]}}{n}\pi_{n[k]}(1-\pi_{n[k]})\times\\
&\left\|\frac{1}{n_{1[k]}}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}^{*}_{n}(\boldsymbol{X}_i)\right\} - \frac{1}{n_{0[k]}}\sum_{i=1}^{n}(1-A_i)\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\xi}^{*}_{n}(\boldsymbol{X}_i)\right\}\right\|\\
=&p_{[k]}O_P(1)\times o_P(n^{-1/2}) = p_{[k]}n^{-1/2}o_P(1),
\end{aligned}$$

由此可得

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]} - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}^{*}_{i,[k]}\right\| = p_{[k]}n^{-1/2}o_P(1). \tag{B.59}$$

注意,

$$\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^{*}_{i,[k]}\right] = (A_i-\pi_{n[k]})\mathbb{1}(B_i=k)\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}^{*}_{n}(\boldsymbol{X}_i)\mid B_i=k\right]$$

and

$$\mathbb{E}_{\mathcal{C}_n}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}^*_{i,[k]}\right\|^2\right] = \frac{1}{n^2}\sum_{i=1}^{n}(A_i-\pi_{n[k]})^2\mathbb{1}(B_i=k)\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\right\|^2 \mid B_i=k\right]$$
$$\leq \frac{1}{n}\frac{n_{[k]}}{n}\mathrm{tr}\left\{\mathbb{E}\left[\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)^\top \mid B_i=k\right]\right\} = p_{[k]}\frac{r_n}{n}O_P(1).$$

我们有

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}^*_{i,[k]}\right\| = \sqrt{\frac{r_n p_{[k]}}{n}}O_P(1)O_{L^2,\mathcal{C}_n}(1)$$

因此，我们可以得出

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\right\| \leq \left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}^*_{i,[k]}\right\| + p_{[k]}n^{-1/2}o_P(1) = \sqrt{\frac{r_n p_{[k]}}{n}}O_P(1)O_{L^2,\mathcal{C}_n}(1), \tag{B.60}$$

其中最后一步由 $p_{[k]} \leq \sqrt{r_n p_{[k]}}$ 得出。 ——

结合（B.46）、（B.48）、（B.58）和（B.60），我们有

$$\begin{aligned}
|Q_1| &\leq \sum_{k=1}^{K}\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}^\top_{i,[k]}\epsilon_{i,[k]}\right\|\left\|\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}^\top_{i,[k]}\right\}^+\right\|\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\right\| \\
&\leq \sum_{k=1}^{K}\left\{\frac{r_n\log(2Kd)}{n} + p_{[k]}n^{-1/4} + \sqrt{\frac{\zeta_n^2 p_{[k]}}{n}}\right\}O_P(1)O_{L^2,\mathcal{C}_n}(1) \\
&\quad \times\left\{p_{[k]}^{-1}O(1) + p_{[k]}^{-1}o_P(1)\right\}\times\left\{\sqrt{\frac{r_n p_{[k]}}{n}}O_P(1)O_{L^2,\mathcal{C}_n}(1)\right\} \\
&\leq O_P(1)\sqrt{\frac{r_n^3\log^2(2Kd)}{n^3\inf_{1\leq k\leq K}p_{[k]}}}\sum_{k=1}^{K}O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1) \\
&\quad + O_P(1)\left(\frac{r_n^2}{n^3}\right)^{1/4}\sum_{k=1}^{K}\sqrt{p_{[k]}}O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1) \\
&\quad + O_P(1)\frac{\sqrt{\zeta_n^2 r_n}}{n}\sum_{k=1}^{K}O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1) \\
&= O_P(n^{-1/2})\left\{\sqrt{\frac{K^2r_n^2}{n}\frac{r_n\log^2(2Kd)}{n\inf_{1\leq k\leq K}p_{[k]}}} + \left(\frac{K^2r_n^2}{n}\right)^{1/4} + \sqrt{\frac{K^2\zeta_n^2 r_n}{n}}\right\} \\
&= o_P(n^{-1/2})
\end{aligned} \tag{B.61}$$

第一个等式　　上述公式成立是因为根据 Chernozhukov 等人(2018，引理

6.1）以及

$$\mathbb{E}_{\mathcal{C}_n}\left[\left|\sum_{k=1}^{K} O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1)\right|\right] \leq \sum_{k=1}^{K}\mathbb{E}_{\mathcal{C}_n}\left[|O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1)|\right]$$
$$\leq \sum_{k=1}^{K}\sqrt{\mathbb{E}_{\mathcal{C}_n}\left[O_{L^2,\mathcal{C}_n}(1)^2\right]}\sqrt{\mathbb{E}_{\mathcal{C}_n}\left[O_{L^2,\mathcal{C}_n}(1)^2\right]} \leq \sum_{k=1}^{K} O_P(1) \leq KO_P(1),$$

$$\mathbb{E}_{\mathcal{C}_n}\left[\sqrt{p_{[k]}}\left|\sum_{k=1}^{K} O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1)\right|\right] \leq \sum_{k=1}^{K}\sqrt{p_{[k]}}\mathbb{E}_{\mathcal{C}_n}\left[|O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1)|\right]$$
$$\leq \sum_{k=1}^{K}\sqrt{p_{[k]}}\sqrt{\mathbb{E}_{\mathcal{C}_n}\left[O_{L^2,\mathcal{C}_n}(1)^2\right]}\sqrt{\mathbb{E}_{\mathcal{C}_n}\left[O_{L^2,\mathcal{C}_n}(1)^2\right]} \leq \sum_{k=1}^{K}\sqrt{p_{[k]}}O_P(1) = O_P(1)\sum_{k=1}^{K}\sqrt{p_{[k]}}$$
$$\leq O_P(1)\sqrt{K}\sqrt{\sum_{k=1}^{K} p_{[k]}} = \sqrt{K}O_P(1),$$

有

$$\left|\sum_{k=1}^{K} O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1)\right| = O_P(K) \text{ and } \left|\sum_{k=1}^{K}\sqrt{p_{[k]}}O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1)\right| = O_P(\sqrt{K}).$$

由（B.53）和（B.59），可得

$$\begin{aligned} Q_2 &= \sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\left\{\widetilde{Y}^*_{i,[k]} - \boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\boldsymbol{\Xi}_{i,[k]}\right\} \\ &= \sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\left\{\widetilde{Y}^*_{i,[k]} - \boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\boldsymbol{\Xi}^*_{i,[k]}\right\} + \sum_{k=1}^{K}\boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\Xi}^*_{i,[k]} - \boldsymbol{\Xi}_{i,[k]}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\epsilon^*_{i,[k]} + n^{-1/2}o_P(1)\sum_{k=1}^{K}p_{[k]} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\epsilon^*_{i,[k]} + o_P(n^{-1/2}), \end{aligned} \tag{B.62}$$

注意到 $\sum_{k=1}^{K}\epsilon^*_{i,\ [k]}$ 在.....条件下相互独立 $\mathcal{C}_n$，我们打算应用引理 4。直接计算可得

$$\mathbb{E}_{\mathcal{C}_n}\left[\sum_{k=1}^{K}\epsilon^*_{i,[k]}\right] = \sum_{k=1}^{K}\mathbb{1}(B_i = k)\left\{\frac{A_i}{\pi_{n[k]}}\mathbb{E}\left[\widetilde{Y}_i(1) \mid B_i = k\right] - \frac{1-A_i}{1-\pi_{n[k]}}\mathbb{E}\left[\widetilde{Y}_i(0) \mid B_i = k\right]\right\}$$
$$- \sum_{k=1}^{K}\mathbb{1}(B_i = k)(A_i - \pi_{n[k]})\boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i) \mid B_i = k\right]$$

$$= 0,$$

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{k=1}^{K}\epsilon^*_{i,[k]} \mid \mathcal{C}_n\right) = \mathbb{E}\left[\left(\sum_{k=1}^{K}\epsilon^*_{i,[k]}\right)^2 \mid \mathcal{C}_n\right] = \sum_{k=1}^{K}\mathbb{E}\left[\epsilon^{*2}_{i,[k]} \mid \mathcal{C}_n\right] \\
=&\sum_{k=1}^{K}\mathbb{1}(B_i=k)\left\{\frac{A_i}{\pi^2_{n[k]}}\mathbb{E}\left[\widetilde{Y}_i^2(1) \mid B_i=k\right] + \frac{1-A_i}{(1-\pi_{n[k]})^2}\mathbb{E}\left[\widetilde{Y}_i^2(0) \mid B_i=k\right]\right\} \\
&- 2\sum_{k=1}^{K}\mathbb{1}(B_i=k)A_i\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\mathbb{E}\left[\widetilde{Y}_i(1)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i) \mid B_i=k\right] \\
&- 2\sum_{k=1}^{K}\mathbb{1}(B_i=k)(1-A_i)\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\mathbb{E}\left[\widetilde{Y}_i(0)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i) \mid B_i=k\right] \\
&+ \sum_{k=1}^{K}\mathbb{1}(B_i=k)(A_i-\pi_{n[k]})^2\boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top \mid B_i=k\right]\boldsymbol{\beta}_{[k],\mathcal{C}_n} \\
=&\sum_{k=1}^{K}\mathbb{1}(B_i=k)\left\{\frac{A_i}{\pi^2_{n[k]}}\mathbb{E}\left[\widetilde{Y}_i^2(1) \mid B_i=k\right] + \frac{1-A_i}{(1-\pi_{n[k]})^2}\mathbb{E}\left[\widetilde{Y}_i^2(0) \mid B_i=k\right]\right\} \\
&- 2\sum_{k=1}^{K}\mathbb{1}(B_i=k)\boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\left\{A_i\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}(1)} + (1-A_i)\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}(0)}\right\} \\
&+ \sum_{k=1}^{K}\mathbb{1}(B_i=k)(A_i-\pi_{n[k]})^2\boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{\boldsymbol{\xi}}^*_n}\boldsymbol{\beta}_{[k],\mathcal{C}_n}
\end{aligned} \tag{B.63}$$

and

$$\begin{aligned}
\sigma^2_n(\mathcal{C}_n) :=& \mathrm{Var}\left(\sum_{i=1}^{n}\sum_{k=1}^{K}\epsilon^*_{i,[k]} \mid \mathcal{C}_n\right) = \sum_{i=1}^{n}\mathrm{Var}\left(\sum_{k=1}^{K}\epsilon^*_{i,[k]} \mid \mathcal{C}_n\right) \\
=& \sum_{i=1}^{n}\sum_{k=1}^{K}\mathbb{1}(B_i=k)\left\{\frac{A_i}{\pi^2_{n[k]}}\mathbb{E}\left[\widetilde{Y}_i^2(1) \mid B_i=k\right] + \frac{1-A_i}{(1-\pi_{n[k]})^2}\mathbb{E}\left[\widetilde{Y}_i^2(0) \mid B_i=k\right]\right\} \\
&- 2\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbb{1}(B_i=k)\boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\left\{A_i\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}(1)} + (1-A_i)\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}(0)}\right\} \\
&+ \sum_{i=1}^{n}\sum_{k=1}^{K}\mathbb{1}(B_i=k)(A_i-\pi_{n[k]})^2\boldsymbol{\beta}^\top_{[k],\mathcal{C}_n}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{\boldsymbol{\xi}}^*_n}\boldsymbol{\beta}_{[k],\mathcal{C}_n}.
\end{aligned}$$

由引理 1, $\left|\pi_{n[k]} - \pi_{[k]}\right| = o_P(1)$, $\left\|\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{\boldsymbol{\xi}}^*_n}\right\| = O(1)$, $\left\|\boldsymbol{\beta}_{[k],\ \mathcal{C}_n}\right\| = O_P(1)$, $\left\|\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}^*_n\widetilde{Y}^{(a)}}\right\| =$

$O(1)$ 和（B.53），我们有

$$\begin{aligned}\frac{\sigma_n^2(\mathcal{C}_n)}{n} &= \sum_{k=1}^{K}\left\{\frac{p_{[k]}}{\pi_{[k]}}\mathbb{E}\left[\widetilde{Y}_i^2(1)\mid B_i=k\right]+\frac{p_{[k]}}{1-\pi_{[k]}}\mathbb{E}\left[\widetilde{Y}_i^2(0)\mid B_i=k\right]\right\}\\ &\quad -2\sum_{k=1}^{K}p_{[k]}\boldsymbol{\beta}_{[k]}^{*\top}\left\{(1-\pi_{[k]})\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)}+\pi_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}\right\}\\ &\quad +\sum_{k=1}^{K}p_{[k]}\pi_{[k]}(1-\pi_{[k]})\boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\boldsymbol{\beta}_{[k]}^{*}+o_P(1)\\ &= \varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2+o_P(1)\\ &= O_P(1).\end{aligned} \tag{B.64}$$

根据（B.63）、（B.64）和假设4，条件（A.1）在引理4 中成立。

接下来还需验证林德伯格条件。利用假设1 和4 以及（B.54），我们有

$$\begin{aligned}&\mathbb{E}\left[\left|\sum_{k=1}^{K}\epsilon_{i,[k]}^*\right|^{2+\epsilon}\mid\mathcal{C}_n\right]\\ \leq& 3^{1+\epsilon}\mathbb{E}\left[\left|\sum_{k=1}^{K}\mathbb{1}(B_i=k)\frac{A_i}{\pi_{n[k]}}\right|^{2+\epsilon}\left|\widetilde{Y}_i(1)\right|^{2+\epsilon}+\left|\sum_{k=1}^{K}\mathbb{1}(B_i=k)\frac{1-A_i}{1-\pi_{n[k]}}\right|^{2+\epsilon}\left|\widetilde{Y}_i(0)\right|^{2+\epsilon}\mid\mathcal{C}_n\right]\\ &+3^{1+\epsilon}\mathbb{E}\left[\left|\sum_{k=1}^{K}\mathbb{1}(B_i=k)(A_i-\pi_{n[k]})\right|^{2+\epsilon}\left|\boldsymbol{\beta}_{[k],\mathcal{C}_n}^{\top}\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right|^{2+\epsilon}\mid\mathcal{C}_n\right]\\ \leq&\frac{2\times 3^{1+\epsilon}}{\min_{1\leq k\leq K}\left\{\pi_{n[k]}^{2+\epsilon},(1-\pi_{n[k]})^{2+\epsilon}\right\}}\left\{\sup_{a\in\{0,1\},1\leq k\leq K}\mathbb{E}\left[\left|\widetilde{Y}_i(a)\right|^{2+\epsilon}\mid B_i=k\right]+\mathbb{E}\left[\left|\boldsymbol{\beta}_{[k],\mathcal{C}_n}^{\top}\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right|^{2+\epsilon}\mid\mathcal{C}_n\right]\right\}\\ \leq&\frac{2\times 3^{1+\epsilon}}{\min_{1\leq k\leq K}\left\{\pi_{n[k]}^{2+\epsilon},(1-\pi_{n[k]})^{2+\epsilon}\right\}}\left\{O(1)+\zeta_n^{\epsilon}\boldsymbol{\beta}_{[k],\mathcal{C}_n}^{\top}\sup_{1\leq k\leq K}\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^{\top}\mid B_i=k\right]\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right\}\\ =&O_P(\zeta_n^{\epsilon}),\end{aligned}$$

其中第一步由 Loève 的 c_r 不等式推出（Davidson 2021，定理 9.32），第二步由

$$\max\left\{\sum_{k=1}^{K}\mathbb{1}(B_i=k)\frac{A_i}{\pi_{n[k]}},\sum_{k=1}^{K}\mathbb{1}(B_i=k)\frac{1-A_i}{1-\pi_{n[k]}}\right\}\leq\frac{1}{\min_{1\leq k\leq K}\left\{\pi_{n[k]},(1-\pi_{n[k]})\right\}}$$

and

$$\mathbb{E}\left[\left|\widetilde{Y}_i(a)\right|^{2+\epsilon} \mid \mathcal{C}_n\right] = \mathbb{E}\left[\left|\widetilde{Y}_i(a)\right|^{2+\epsilon} \mid B_i\right] \leq 2^{2+\epsilon} \sup_{a\in\{0,1\},1\leq k\leq K} \mathbb{E}\left[|Y_i(a)|^{2+\epsilon} \mid B_i = k\right],$$

第三步由

$$\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid \mathcal{C}_n\right] = \mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right] \leq \sup_{1\leq k\leq K} \mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top \mid B_i = k\right].$$

于是，引理4中的林德伯格条件由以下得出

$$\begin{aligned}
&\frac{1}{\sigma_n^2(\mathcal{C}_n)}\sum_{i=1}^n \mathbb{E}\left[\left\{\sum_{k=1}^K \epsilon_{i,[k]}^*\right\}^2 \mathbb{1}\left\{\left|\sum_{k=1}^K \epsilon_{i,[k]}^*\right| \geq \delta\sigma_n(\mathcal{C}_n)\right\} \mid \mathcal{C}_n\right] \\
\leq &\frac{1}{\sigma_n^2(\mathcal{C}_n)}\sum_{i=1}^n \frac{\mathbb{E}\left[\left|\sum_{k=1}^K \epsilon_{i,[k]}^*\right|^{2+\epsilon} \mid \mathcal{C}_n\right]}{\{\delta\sigma_n(\mathcal{C}_n)\}^\epsilon} = \frac{1}{\varsigma_{\widetilde{Y}}^2 - \varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2 + o_P(1)} \frac{O_P(\zeta_n^\epsilon)}{O_P(n^{\epsilon/2})}\frac{1}{\delta^\epsilon} = o_P(1)
\end{aligned}$$

对任意 $\delta > 0$，其中第一步由（B.64）得出，最后一步由$K^2\zeta_n^2 r_n/n \to 0$ 得出，当 $n \to \infty$时。应用引理4 我们有

$$\mathbb{E}\left[\left|\mathbb{E}\left[\exp\left\{\mathrm{i}t\frac{\sum_{i=1}^n\sum_{k=1}^K \epsilon_{i,[k]}^*}{\sigma_n(\mathcal{C}_n)}\right\} \mid \mathcal{C}_n\right] - \exp\left(-t^2/2\right)\right|\right] \to 0$$

当 $n \to \infty$ 时，对所有 $t \in \mathbb{R}$成立。

回顾 R_3 以及 R_4 （定义于（B.10））， 由经典中心极限定理，我们有

$$\mathbb{E}\left[\left|\mathbb{E}\left[\exp\left\{\mathrm{i}s\frac{\sqrt{n}(R_3+R_4)}{\varsigma_H}\right\} \mid \mathcal{C}_n\right] - \exp\left(-s^2/2\right)\right|\right] \to 0$$

对于任意 $s \in \mathbb{R}$。注意到 $R_3 + R_4$ 关于 $\mathcal{C}_n$可测，根据引理 5 可得

$$\left(\frac{\sum_{i=1}^n\sum_{k=1}^K \epsilon_{i,[k]}^*}{\sigma_n(\mathcal{C}_n)}, \frac{\sqrt{n}(R_3+R_4)}{\varsigma_H}\right)^\top \xrightarrow{d} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right).$$

结合（B.31）、（B.32）、（B.61）和（B.62），可得

$$
\begin{aligned}
&\frac{\sqrt{n}\left(\widehat{\tau}_{\text{cal}}-\tau\right)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}} \\
=&\frac{R_1+R_2+R_3+R_4}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}=\frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n\sum_{k=1}^K\epsilon_{i,[k]}^*}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}+\frac{\sqrt{n}(R_3+R_4)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}+o_P(1) \\
=&\frac{\sigma_n(\mathcal{C}_n)}{\sqrt{n}\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sum_{i=1}^n\sum_{k=1}^K\epsilon_{i,[k]}^*}{\sigma_n(\mathcal{C}_n)}+\frac{\varsigma_H}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sqrt{n}(R_3+R_4)}{\varsigma_H}+o_P(1) \\
=&\frac{\sqrt{\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}+o_P(1)}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sum_{i=1}^n\sum_{k=1}^K\epsilon_{i,[k]}^*}{\sigma_n(\mathcal{C}_n)}+\frac{\varsigma_H}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sqrt{n}(R_3+R_4)}{\varsigma_H}+o_P(1) \\
=&\frac{\sqrt{\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sum_{i=1}^n\sum_{k=1}^K\epsilon_{i,[k]}^*}{\sigma_n(\mathcal{C}_n)}+\frac{\varsigma_H}{\sqrt{\varsigma_H^2+\varsigma_{\widetilde{Y}}^2-\varsigma_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2}}\frac{\sqrt{n}(R_3+R_4)}{\varsigma_H}+o_P(1) \\
\xrightarrow{d}&N(0,1),
\end{aligned}
$$

其中，第三个等号由（B.64）得出，第四个等号由$\sum_{i=1}^n\sum_{k=1}^K\epsilon_{i,\ [k]}^*/\sigma_n\left(\mathcal{C}_n\right)$是一致紧的这一点得出，最后一步则由引理6 和Slutsky定理得出。至此，$\widehat{\tau}_{\text{cal}}$ 的渐近正态性得证。

方差估计量的一致性。我们仅证明 $\widehat{\varsigma}_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}^2$ 的一致性，因为 $\widehat{\varsigma}_{\widetilde{Y}}^2$ 的一致性可用类似的方法证明。

由（B.48），可得

$$
\left\|\widehat{\boldsymbol{\Sigma}}_{[k]}^{+}-\pi_{[k]}^{-1}(1-\pi_{[k]})^{-1}p_{[k]}^{-1}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}^{+}\right\|=p_{[k]}^{-1}o_P(1),\ \forall 1\leq k\leq K.
$$

注意到

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^*\widetilde{Y}_{i,[k]}^*-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^*\widetilde{Y}_{i,[k]}^*\right]\right\|^2\right]\leq\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^*\widetilde{Y}_{i,[k]}^*\right\|^2\right] \\
\leq&2\max\left\{\frac{1-\pi_{n[k]}}{\pi_{n[k]}},\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\right\}^2\mathbb{1}(B_i=k)\max_{a=0,1}\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(a)\right\|^2\mid B_i=k\right]
\end{aligned}
$$

$$\leq O_P(\zeta_n^2)\max_{a=0,1}\mathbb{E}\left[\left\|\widetilde{Y}_i(a)\right\|^2 \mid B_i=k\right]\mathbb{1}(B_i=k)=O_P(\zeta_n^2)\mathbb{1}(B_i=k),$$

其中最后一步由假设1 和4 以及随机变量 $O_P(\zeta_n^2)$ 不依赖于 k。因此，由引理1 可得

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}_n}\left[\left\|\frac{1}{n}\sum_{i=1}^n\left\{\boldsymbol{\Xi}^*_{i,[k]}\widetilde{Y}^*_{i,[k]}-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^*_{i,[k]}\widetilde{Y}^*_{i,[k]}\right]\right\}\right\|^2\right]\\
\leq&\frac{1}{n^2}\sum_{i=1}^n\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}^*_{i,[k]}\widetilde{Y}^*_{i,[k]}-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^*_{i,[k]}\widetilde{Y}^*_{i,[k]}\right]\right\|^2\right]\\
\leq&\frac{1}{n}\frac{n_{[k]}}{n}O_P(\zeta_n^2)=\frac{\zeta_n^2}{n}p_{[k]}O_P(1)
\end{aligned}$$

and

$$\left\|\frac{1}{n}\sum_{i=1}^n\left\{\boldsymbol{\Xi}^*_{i,[k]}\widetilde{Y}^*_{i,[k]}-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^*_{i,[k]}\widetilde{Y}^*_{i,[k]}\right]\right\}\right\|=\sqrt{\frac{\zeta_n^2p_{[k]}}{n}}O_{L^2,\mathcal{C}_n}(1)O_P(1).$$

将此与（B.50）和（B.52）结合起来，可得

$$\begin{aligned}
&\left\|\widehat{\boldsymbol{\Gamma}}_{[k]}-(1-\pi_{[k]})p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)}-\pi_{[k]}p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}\right\|\\
=&\left\{\sqrt{\frac{\zeta_n^2p_{[k]}}{n}}+\frac{\sqrt{r_n\log(2Kd)}}{n}+p_{[k]}n^{-1/4}\right\}O_P(1)O_{L^2,\mathcal{C}_n}(1)+p_{[k]}o_P(1)\\
=&p_{[k]}o_P(1)O_{L^2,\mathcal{C}_n}(1).
\end{aligned}$$

于是，由 $\left\|\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}\ (a)}\right\|=O(1)$、$\left\|\boldsymbol{\Sigma}^+_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*}\right\|=O(1)$,

$$\begin{aligned}
\sup_{1\leq k\leq K}\left|\frac{\widehat{r}_{[k]}}{np_{[k]}}\right|&\leq\sup_{1\leq k\leq K}\left|\frac{r_{[k]}}{np_{[k]}}\right|+\sup_{1\leq k\leq K}\left|\frac{r_{[k]}}{np_{[k]}}\right|\left|\frac{\widehat{r}_{[k]}-r_{[k]}}{r_{[k]}}\right|\\
&\leq\frac{r_n}{n\min_{1\leq k\leq K}p_{[k]}}+\frac{r_n}{n\min_{1\leq k\leq K}p_{[k]}}\sup_{1\leq k\leq K}\left|\frac{\widehat{r}_{[k]}-r_{[k]}}{r_{[k]}}\right|\\
&=o_P(1)
\end{aligned}$$

以及假设2 即

$$\widehat{\zeta}^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*}=\sum_{k=1}^K\{1+o_P(1)\}\left\{(1-\pi_{[k]})p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)}+\pi_{[k]}p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)}+p_{[k]}o_P(1)O_{L^2,\mathcal{C}_n}(1)\right\}$$

$$
\begin{aligned}
&\times \left\{\pi_{[k]}^{-1}(1-\pi_{[k]})^{-1}p_{[k]}^{-1}\boldsymbol{\Sigma}^{+}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{\boldsymbol{\xi}}_n^*} + p_{[k]}^{-1}o_P(1)\right\} \\
&\times \left\{(1-\pi_{[k]})p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(1)} + \pi_{[k]}p_{[k]}\boldsymbol{\Sigma}_{[k]\widetilde{\boldsymbol{\xi}}_n^*\widetilde{Y}(0)} + p_{[k]}o_P(1)O_{L^2,\mathcal{C}_n}(1)\right\} \\
&= \varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*} + o_P(1)\sum_{k=1}^{K} p_{[k]}O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1) + o_P(1)\sum_{k=1}^{K} p_{[k]}O_{L^2,\mathcal{C}_n}(1) \\
&= \varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*} + o_P(1)\sum_{k=1}^{K} p_{[k]}O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1) \\
&= \varsigma^2_{\widetilde{Y}|\widetilde{\boldsymbol{\xi}}_n^*} + o_P(1),
\end{aligned}
$$

最后一步成立的原因是，根据Chernozhukov 等人(2018，引理6.1），以及

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{C}_n}\left[\left|\sum_{k=1}^{K} p_{[k]}O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1)\right|\right] \leq \sum_{k=1}^{K} p_{[k]}\mathbb{E}_{\mathcal{C}_n}\left[|O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1)|\right] \\
\leq &\sum_{k=1}^{K} p_{[k]}\sqrt{\mathbb{E}_{\mathcal{C}_n}\left[O_{L^2,\mathcal{C}_n}(1)^2\right]}\sqrt{\mathbb{E}_{\mathcal{C}_n}\left[O_{L^2,\mathcal{C}_n}(1)^2\right]} \leq \sum_{k=1}^{K} p_{[k]}O_P(1) = O_P(1),
\end{aligned}
$$

可得

$$
\left|\sum_{k=1}^{K} p_{[k]}O_{L^2,\mathcal{C}_n}(1)O_{L^2,\mathcal{C}_n}(1)\right| = O_P(1).
$$

证明完毕。

半参数效率界的建立与定理1的建立相同;

因此我们省略证明。 ■

B.4 定理4的证明

推导校准权重的表达式。我们首先推导其表达式

对于校准权重 $\widehat{w}_i$'s。对于所有 $i \in \{1,\ldots,n\}$，令

$$
\boldsymbol{\Xi}_i := \begin{pmatrix}
\left\{A_i - \pi_{n[1]}\right\}\mathbb{1}(B_i = 1)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[1]}\right\} \\
\left\{A_i - \pi_{n[2]}\right\}\mathbb{1}(B_i = 2)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[2]}\right\} \\
\vdots \\
\left\{A_i - \pi_{n[K]}\right\}\mathbb{1}(B_i = K)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[K]}\right\}
\end{pmatrix} \in \mathbb{R}^{Kd}.
$$

令 $D(\boldsymbol{w}) = \sum_{i=1}^{n} D(w_i, 1)$。则校准问题可改写为

$$\min_{w_i:1\le i\le n} D(\boldsymbol{w}) \quad \text{s.t.} \quad (\boldsymbol{\Xi}_1, \ldots, \boldsymbol{\Xi}_n)\, \boldsymbol{w} = 0. \tag{B.65}$$

注意， $D(\boldsymbol{w})$ 的共轭函数为

$$D^*(\boldsymbol{z}) = \sum_{i=1}^{n} \left\{ z_i \cdot (D')^{-1}(z_i) - D\left\{(D')^{-1}(z_i)\right\}\right\} = \sum_{i=1}^{n} -\rho(-z_i), \quad \forall \boldsymbol{z} = (z_1, \ldots, z_n)^\top,$$

其中 $\rho(v) := D\{(D')^{-1}(-v)\} + v \cdot (D')^{-1}(-v)$ 对于任意 $v \in \mathbb{R}$。根据 Tseng & Bertsekas (1991) （另见 Boyd & Vandenberghe (2004年)），问题（B.65）的对偶问题为

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^{Kd}} \left\{ -D^*((\boldsymbol{\Xi}_1, \ldots, \boldsymbol{\Xi}_n)^\top \boldsymbol{\lambda})\right\} = \max_{\boldsymbol{\lambda}\in\mathbb{R}^{Kd}} \sum_{i=1}^{n} \rho(-\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i) = \max_{\boldsymbol{\lambda}\in\mathbb{R}^{Kd}} \frac{1}{n} \sum_{i=1}^{n} \rho(\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i).$$

令 $\widehat{\boldsymbol{\lambda}} := \operatorname{argmax}_{\boldsymbol{\lambda}\in\mathbb{R}^{Kd}} \frac{1}{n} \sum_{i=1}^{n} \rho(\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i)$。那么，根据一阶条件，最小化问题（B.4）的解为

$$\widehat{w}_i = \rho'(\widehat{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i) \text{ for all } i \in \{1, \ldots, n\}, \tag{B.66}$$

其中 $\widehat{\boldsymbol{\lambda}}$ 满足

$$\frac{1}{n} \sum_{i=1}^{n} \rho'(\widehat{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i) \boldsymbol{\Xi}_i = 0 \tag{B.67}$$

and $\rho'(\cdot)$ 是 $\rho(\cdot)$ 的导数。我们首先推导 $\widehat{\boldsymbol{\lambda}}$ 的收敛速率。

$\widehat{\boldsymbol{\lambda}}$ 的收敛速率。 对任意 $\epsilon > 0$，令 $\Upsilon(\epsilon) := \{\boldsymbol{\lambda} \in \mathbb{R}^{Kd} : \|\boldsymbol{\lambda}\| \le C(\epsilon) n^{-1/2}\}$ 以及 $\partial\Upsilon := \{\boldsymbol{\lambda} \in \mathbb{R}^{Kd} : \|\boldsymbol{\lambda}\| = C(\epsilon) n^{-1/2}\}$，其中常数 $C(\epsilon)$ 将在后文确定。令 $\widehat{G}(\boldsymbol{\lambda}) := \frac{1}{n} \sum_{i=1}^{n} \rho(\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i)$，则 $\frac{\partial \widehat{G}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \frac{1}{n} \sum_{i=1}^{n} \rho'(\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i) \boldsymbol{\Xi}_i$ 且

$$\frac{\partial \widehat{G}(\mathbf{0})}{\partial \boldsymbol{\lambda}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Xi}_i = O_P(n^{-1/2}),$$

其中最后一步由(B.24)。于是，存在常数 $M(\epsilon) > 0$ 使得

$$\left\| \frac{\partial \widehat{G}(\mathbf{0})}{\partial \boldsymbol{\lambda}} \right\| \leq M(\epsilon) \times n^{-1/2} \tag{B.68}$$

当 n 足够大时，以至少 $1-\epsilon$ 的概率。

注意，

$$\begin{aligned}
\mathbb{E}\left[\max_{1\leq i\leq n} \|\boldsymbol{\Xi}_i\|^{2+\epsilon}\right] &\leq \sum_{i=1}^{n} \mathbb{E}\left[\|\boldsymbol{\Xi}_i\|^{2+\epsilon}\right] = \sum_{i=1}^{n} \mathbb{E}\left[\left\{\sum_{k=1}^{K}\left(A_i - \pi_{n[k]}\right)^2 \mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}_{n[k]}\right\|^2\right\}^{\frac{2+\epsilon}{2}}\right] \\
&\leq \sum_{i=1}^{n} \mathbb{E}\left[\left\{\sum_{k=1}^{K}\left\{2\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^2 + 2\left\|\overline{\boldsymbol{\xi}}_{n[k]}\right\|^2\right\}\right\}^{\frac{2+\epsilon}{2}}\right] \\
&\leq \sum_{i=1}^{n} (2K)^{\frac{2+\epsilon}{2}} 2^{\frac{2+\epsilon}{2}} \mathbb{E}\left[\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^{2+\epsilon} + \max_{1\leq k\leq K}\left\|\overline{\boldsymbol{\xi}}_{n[k]}\right\|^{2+\epsilon}\right] \\
&\leq \sum_{i=1}^{n} (2K)^{\frac{2+\epsilon}{2}} 2^{\frac{2+\epsilon}{2}} \mathbb{E}\left[\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^{2+\epsilon}\right] + \sum_{i=1}^{n} (2K)^{\frac{2+\epsilon}{2}} 2^{\frac{2+\epsilon}{2}} \mathbb{E}\left[\sum_{k=1}^{K}\frac{1}{n_{[k]}}\sum_{i\in[k]}\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^{2+\epsilon}\right] \\
&= O(n)
\end{aligned}$$

其中第三个不等式使用了幂均值不等式，第四个不等式由

$$\left\|\overline{\boldsymbol{\xi}}_{n[k]}\right\|^{2+\epsilon} = \left\|\frac{1}{n_{[k]}}\sum_{i\in[k]}\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^{2+\epsilon} \leq \left\{\frac{1}{n_{[k]}}\sum_{i\in[k]}\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|\right\}^{2+\epsilon} \leq \frac{1}{n_{[k]}}\sum_{i\in[k]}\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^{2+\epsilon},$$

且最后一步由 $\sup_{n\geq 1}\sup_{1\leq i\leq n}\mathbb{E}\left[\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^{2+\epsilon}\right] < \infty$ 得出。然后，根据马尔可夫不等式，我们有

$$\max_{1\leq i\leq n}\|\boldsymbol{\Xi}_i\| = O_P(n^{\frac{1}{2+\epsilon}}). \tag{B.69}$$

因此，由柯西-施瓦茨不等式可得

$$\max_{1\leq i\leq n}\sup_{\boldsymbol{\lambda}\in\Upsilon(\epsilon)}\left|\boldsymbol{\lambda}^\top\boldsymbol{\Xi}_i\right| \leq O_P(n^{\frac{1}{2+\epsilon}}) \times C(\epsilon) \times n^{-1/2} = o_P(1).$$

由此可知， $\rho''(v)$ 在 $v=0$ 处连续，且 $\rho''(0)<0$，因此

$$2\rho''(0) \le \rho''(\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i) \le \rho''(0)/2 \quad \forall \boldsymbol{\lambda} \in \Upsilon(\epsilon) \text{ and } i=1,\dots,n \tag{B.70}$$

以概率 $1-o(1)$。由（B.16）及 $\boldsymbol{\Sigma}$ 的定义，我们有

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^\top = \boldsymbol{\Sigma} + o_P(1).$$

由韦尔不等式、假设 3 以及 $\boldsymbol{\Sigma}$ 为非奇异的假设，我们有

$$\lambda_{\min}\left\{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^\top\right\} \ge c/2 \tag{B.71}$$

以概率 $1-o(1)$，并且

$$\left\| \left\{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^\top\right\}^{+} - \boldsymbol{\Sigma}^{+} \right\| = o_P(1).$$

B $\tilde{}$y 泰勒展开，对任意 $\boldsymbol{\lambda} \in \partial\Upsilon(\epsilon)$，存在 $\boldsymbol{\lambda}_1 \in \Upsilon(\epsilon)$，使得

$$\begin{aligned}
&\widehat{G}(\boldsymbol{\lambda}) - \widehat{G}(\mathbf{0}) = \boldsymbol{\lambda}^\top \frac{\partial \widehat{G}(\mathbf{0})}{\partial \boldsymbol{\lambda}} + \frac{1}{2}\boldsymbol{\lambda}^\top \frac{\partial^2 \widehat{G}(\tilde{\boldsymbol{\lambda}}_1)}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}^\top}\boldsymbol{\lambda} \\
\le& C(\epsilon) n^{-1/2} \left\| \frac{\partial \widehat{G}(\mathbf{0})}{\partial \boldsymbol{\lambda}} \right\| + \frac{1}{2}\boldsymbol{\lambda}^\top \frac{1}{n}\sum_{i=1}^{n} \rho''(\tilde{\boldsymbol{\lambda}}_1^\top \boldsymbol{\Xi}_i) \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^\top \boldsymbol{\lambda} \\
\le& C(\epsilon) \times M(\epsilon) n^{-1} + \frac{\rho''(0)}{4}\boldsymbol{\lambda}^\top \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^\top \boldsymbol{\lambda} \\
\le& C(\epsilon) \times M(\epsilon) n^{-1} + \frac{\rho''(0)c}{8}\left(C(\epsilon) n^{-1/2}\right)^2 = C(\epsilon) n^{-1}\left(M(\epsilon) + \frac{\rho''(0)c}{8} C(\epsilon)\right)
\end{aligned} \tag{B.72}$$

以概率 $1-o(1)-\epsilon$，其中第二个不等式由（B.68）和（B.70）得到，最后一个由（B.71）得到。若我们选取 $C(\epsilon) = -16M(\epsilon)/(\rho''(0)c)$，则

$$\widehat{G}(\boldsymbol{\lambda}) - \widehat{G}(\mathbf{0}) < 0$$

对任意 $\boldsymbol{\lambda} \in \partial\Upsilon($ $\hat{}$ $\epsilon)$。由于 $G(\boldsymbol{\lambda})$ 连续，因此 $G(\boldsymbol{\lambda})$ 在

$\widehat{}$内部 $\Upsilon(\epsilon)$。注意， $G(\boldsymbol{\lambda})$也严格凹，且具有唯一的全局最大点 $\widehat{\boldsymbol{\lambda}}$，由此可得 $\widehat{\boldsymbol{\lambda}} \in \Upsilon(\epsilon) \backslash \partial \Upsilon(\epsilon)$ ，因此 $\left\|\widehat{\boldsymbol{\lambda}}\right\| \leq C(\epsilon) n^{-1/2}$。由于 (B.72) 以概率 $1-o(1)-\epsilon$成立，因此我们有

$$\left\|\widehat{\boldsymbol{\lambda}}\right\| = O_P(n^{-1/2}).$$

$\widehat{}$**的渐近展开 $\boldsymbol{\lambda}$**。回顾（B.67）。根据泰勒展开，我们有

$$\mathbf{0} = \frac{1}{n}\sum_{i=1}^{n} \rho'(\widehat{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i)\boldsymbol{\Xi}_i = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_i + \frac{1}{n}\sum_{i=1}^{n} \rho''(\tilde{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i)\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top \widehat{\boldsymbol{\lambda}} \tag{B.73}$$

其中 $\tilde{\boldsymbol{\lambda}}$满足 $\left\|\tilde{\boldsymbol{\lambda}}\right\| \leq \left\|\widehat{\boldsymbol{\lambda}}\right\|$。根据 $\left\|\tilde{\boldsymbol{\lambda}}\right\| \leq \left\|\widehat{\boldsymbol{\lambda}}\right\| = O_P\left(n^{-1/2}\right)$ 、（B.69）以及中值定理，我们有

$$\begin{aligned}
&\sup_{1\leq i\leq n}\left|\rho''(\tilde{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i) - \rho''(0)\right| \\
\leq &\sup_{1\leq i\leq n}\sup_{\boldsymbol{\lambda}:\|\boldsymbol{\lambda}\|\leq\left\|\widehat{\boldsymbol{\lambda}}\right\|}\left|\rho'''(\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i)\right| \sup_{1\leq i\leq n}\left|\tilde{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i\right| \\
\leq &\sup_{1\leq i\leq n}\sup_{\boldsymbol{\lambda}:\|\boldsymbol{\lambda}\|\leq\left\|\widehat{\boldsymbol{\lambda}}\right\|}\left|\rho'''(\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i)\right| \left\|\widehat{\boldsymbol{\lambda}}\right\| \sup_{1\leq i\leq n}\|\boldsymbol{\Xi}_i\| \\
= &\sup_{1\leq i\leq n}\sup_{\boldsymbol{\lambda}:\|\boldsymbol{\lambda}\|\leq\left\|\widehat{\boldsymbol{\lambda}}\right\|}\left|\rho'''(\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i)\right| o_P(1).
\end{aligned}$$

此外，我们有

$$\begin{aligned}
\sup_{1\leq i\leq n}\sup_{\boldsymbol{\lambda}:\|\boldsymbol{\lambda}\|\leq\left\|\widehat{\boldsymbol{\lambda}}\right\|}\left|\rho'''(\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i)\right| &\leq |\rho'''(0)| + \sup_{1\leq i\leq n}\sup_{\boldsymbol{\lambda}:\|\boldsymbol{\lambda}\|\leq\left\|\widehat{\boldsymbol{\lambda}}\right\|}\left|\rho'''(\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i) - \rho'''(0)\right| \\
&\leq |\rho'''(0)| + C_\rho \sup_{1\leq i\leq n}\sup_{\boldsymbol{\lambda}:\|\boldsymbol{\lambda}\|\leq\left\|\widehat{\boldsymbol{\lambda}}\right\|}\left|\boldsymbol{\lambda}^\top \boldsymbol{\Xi}_i\right| = O_P(1).
\end{aligned}$$

因此，我们有

$$\sup_{1\leq i\leq n}\left|\rho''(\tilde{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i) - \rho''(0)\right| = o_P(1) \tag{B.74}$$

and

$$\left|\frac{1}{n}\sum_{i=1}^{n}\rho''(\widetilde{\boldsymbol{\lambda}}^{\top}\boldsymbol{\Xi}_i)\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\widehat{\boldsymbol{\lambda}}-\frac{1}{n}\sum_{i=1}^{n}\rho''(0)\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\widehat{\boldsymbol{\lambda}}\right|\leq o_P(1)\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\right\|\left\|\widehat{\boldsymbol{\lambda}}\right\|$$
$$=o_P(1)\left\|\boldsymbol{\Sigma}_{[k]}+o_P(1)\right\|\left\|\widehat{\boldsymbol{\lambda}}\right\|=o_P(n^{-1/2}).$$

由此可得（B.73）

$$\widehat{\boldsymbol{\lambda}}=-\rho''(0)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i+o_P(n^{-1/2}).$$

由（B.8）-（B.9）可得

$$\begin{aligned}\widehat{\tau}_{\text{cal}}-\tau=&\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\left\{\frac{A_i}{\pi_{n[k]}}-\frac{1-A_i}{1-\pi_{n[k]}}\right\}\mathbb{1}(B_i=k)\cdot Y_i-\tau\\&+\frac{1}{n}\sum_{i=1}^{n}(\widehat{w}_i-1)\sum_{k=1}^{K}\left\{\frac{A_i}{\pi_{n[k]}}\left(Y_i-\overline{Y}_{1[k]}\right)-\frac{1-A_i}{1-\pi_{n[k]}}\left(Y_i-\overline{Y}_{0[k]}\right)\right\}\mathbb{1}(B_i=k)\\=&\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\left\{\frac{A_i}{\pi_{n[k]}}-\frac{1-A_i}{1-\pi_{n[k]}}\right\}\mathbb{1}(B_i=k)\cdot Y_i-\tau+\frac{1}{n}\sum_{i=1}^{n}(\rho'(\widehat{\boldsymbol{\lambda}}^{\top}\boldsymbol{\Xi}_i)-1)\widetilde{Y}_i\\=&\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\widetilde{Y}_i^{*}+\frac{1}{n}\sum_{i=1}^{n}(\rho'(\widehat{\boldsymbol{\lambda}}^{\top}\boldsymbol{\Xi}_i)-1)\widetilde{Y}_i\\&+\sum_{k=1}^{K}\left\{\mathbb{E}\left[Y_i(1)\mid B_i=k\right]-\mathbb{E}\left[Y_i(0)\mid B_i=k\right]\right\}\times\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}(B_i=k)-p_{[k]}\right).\end{aligned}$$

回顾定理 1证明中的（B.10）。如果我们能够证明 $\frac{1}{n}\sum_{i=1}^{n}(\rho'(\widehat{\boldsymbol{\lambda}}^{\top}\boldsymbol{\Xi}_i)-1)\widetilde{Y}_i=-\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\widehat{\boldsymbol{\beta}}_{[k]}^{\top}\boldsymbol{\Xi}_{i,\,[k]}+o_P(n^{-1/2})$，那么定理 1 的结论成立。根据中值定理，我们有

$$\frac{1}{n}\sum_{i=1}^{n}(\rho'(\widehat{\boldsymbol{\lambda}}^{\top}\boldsymbol{\Xi}_i)-1)\widetilde{Y}_i=\widehat{\boldsymbol{\lambda}}^{\top}\frac{1}{n}\sum_{i=1}^{n}\rho''(\boldsymbol{\lambda}_1^{\top}\boldsymbol{\Xi}_i)\widetilde{Y}_i\boldsymbol{\Xi}_i\tag{B.75}$$

其中 $\boldsymbol{\lambda}_1$ 满足 $\|\boldsymbol{\lambda}_1\|\leq\left\|\widehat{\boldsymbol{\lambda}}\right\|$。与（B.74）的推导类似，我们可以证明

$$\sup_{1\leq i\leq n}\left|\rho''(\boldsymbol{\lambda}_1^{\top}\boldsymbol{\Xi}_i)-\rho''(0)\right|=o_P(1).$$

因此，

$$\left|\frac{1}{n}\sum_{i=1}^{n}\rho''(\boldsymbol{\lambda}_1^\top\boldsymbol{\Xi}_i)\widetilde{Y}_i\boldsymbol{\Xi}_i-\frac{1}{n}\sum_{i=1}^{n}\rho''(0)\widetilde{Y}_i\boldsymbol{\Xi}_i\right|\leq\sup_{1\leq i\leq n}\left|\rho''(\boldsymbol{\lambda}_1^\top\boldsymbol{\Xi}_i)-\rho''(0)\right|\frac{1}{n}\sum_{i=1}^{n}\left\|\widetilde{Y}_i\boldsymbol{\Xi}_i\right\|.$$

根据柯西-施瓦茨不等式，我们有

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\widetilde{Y}_i\boldsymbol{\Xi}_i\right\|\leq\sqrt{\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\Xi}_i\|^2}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2}\sqrt{\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^2}.$$

首先，我们有

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}-\boldsymbol{\Xi}_{i,[k]}^*\right\|^2=\frac{1}{n}\sum_{i=1}^{n}(A_i-\pi_{n[k]})^2\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|^2\\
\leq&\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|^2\\
\leq&\frac{2}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\|^2+2\left\|\overline{\boldsymbol{\xi}}_{n[k]}^*-\mathbb{E}\left[\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]\right\|^2\\
\leq&2R_6+o_P(1)=o_P(1),
\end{aligned}$$

其中，后两个不等式由（B.13）和（B.14）得出。其次，我们有

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}^*\right\|^2\leq\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|^2=O_P(1)$$

根据马尔可夫不等式和假设3，因此，我们有

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^2\leq2\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}-\boldsymbol{\Xi}_{i,[k]}^*\right\|^2+2\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}^*\right\|^2=O_P(1).$$

注意，我们有

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2\leq K\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\widetilde{Y}_{i,[k]}^2\leq O(1)\sum_{k=1}^{K}\left\{\frac{1}{n}\sum_{i=1}^{n}Y_i^2+\overline{Y}_{1[k]}^2+\overline{Y}_{0[k]}^2\right\}=O_P(1).$$

我们可以得出结论:

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\widetilde{Y}_i\boldsymbol{\Xi}_i\right\| \leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^2}\sqrt{\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^2} = O_P(1)$$

从而

$$\left|\frac{1}{n}\sum_{i=1}^{n}\rho''(\boldsymbol{\lambda}_1^\top\boldsymbol{\Xi}_i)\widetilde{Y}_i\boldsymbol{\Xi}_i - \frac{1}{n}\sum_{i=1}^{n}\rho''(0)\widetilde{Y}_i\boldsymbol{\Xi}_i\right| \leq \sup_{1\leq i\leq n}\left|\rho''(\boldsymbol{\lambda}_1^\top\boldsymbol{\Xi}_i)-\rho''(0)\right|\frac{1}{n}\sum_{i=1}^{n}\left\|\widetilde{Y}_i\boldsymbol{\Xi}_i\right\| = o_P(1). \tag{B.76}$$

回顾（B.75），则有

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(\rho'(\widehat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i)-1)\widetilde{Y}_i = \widehat{\boldsymbol{\lambda}}^\top\frac{1}{n}\sum_{i=1}^{n}\rho''(0)\widetilde{Y}_i\boldsymbol{\Xi}_i + \widehat{\boldsymbol{\lambda}}^\top\left\{\frac{1}{n}\sum_{i=1}^{n}\rho''(\boldsymbol{\lambda}_1^\top\boldsymbol{\Xi}_i)\widetilde{Y}_i\boldsymbol{\Xi}_i - \frac{1}{n}\sum_{i=1}^{n}\rho''(0)\widetilde{Y}_i\boldsymbol{\Xi}_i\right\}\\
=&\widehat{\boldsymbol{\lambda}}^\top\frac{1}{n}\sum_{i=1}^{n}\rho''(0)\widetilde{Y}_i\boldsymbol{\Xi}_i + o_P(n^{-1/2})\\
=&-\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i\boldsymbol{\Xi}_i^\top\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i + o_P(n^{-1/2})\frac{1}{n}\sum_{i=1}^{n}\rho''(0)\widetilde{Y}_i\boldsymbol{\Xi}_i + o_P(n^{-1/2})\\
=&-\widehat{\boldsymbol{\beta}}^\top\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i + o_P(n^{-1/2}) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\widehat{\boldsymbol{\beta}}_{[k]}^\top\boldsymbol{\Xi}_{i,[k]} + o_P(n^{-1/2}),
\end{aligned}$$

其中第二步由 $\left\|\widehat{\boldsymbol{\lambda}}\right\| = O_P(n^{-1/2})$ 与（B.76）得到，第三步由 $\widehat{\boldsymbol{\lambda}} = -\rho''(0)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i + o_P(n^{-1/2})$ 得到，第四步由 $\left\|\frac{1}{n}\sum_{i=1}^{n}\rho''(0)\,\widetilde{Y}_i\boldsymbol{\Xi}_i\right\| \leq |\rho''(0)|\,\frac{1}{n}\sum_{i=1}^{n}\left\|\widetilde{Y}_i\boldsymbol{\Xi}_i\right\| = O_P(1)$。因此，定理1 的结论成立。

现在，我们继续证明第二个结果。 回顾（B.67）。根据泰勒展开，我们有

$$\begin{aligned}
\mathbf{0} = \frac{1}{n}\sum_{i=1}^{n}\rho'(\widehat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i)\boldsymbol{\Xi}_i &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i + \rho''(0)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\widehat{\boldsymbol{\lambda}}\\
&+\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\rho'''(\widetilde{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i)\boldsymbol{\Xi}_i\left(\boldsymbol{\Xi}_i^\top\widehat{\boldsymbol{\lambda}}\right)^2,
\end{aligned}$$

其中 $\tilde{\boldsymbol{\lambda}}$ 满足 $\left\|\tilde{\boldsymbol{\lambda}}\right\| \leq \left\|\hat{\boldsymbol{\lambda}}\right\|$。于是

$$\begin{aligned}\hat{\boldsymbol{\lambda}} &= -\left\{\rho''(0)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i + \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\rho'''(\tilde{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i)\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\hat{\boldsymbol{\lambda}}\right\} \\ &= -\left\{\rho''(0)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i - \left\{\rho''(0)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right\}^{-1}\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\rho'''(\tilde{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i)\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\hat{\boldsymbol{\lambda}}.\end{aligned}$$

根据 $\left\|\tilde{\boldsymbol{\lambda}}\right\| \leq \left\|\hat{\boldsymbol{\lambda}}\right\| = O_P(n^{-1/2})$、（B.69）和假设 5，我们有

$$\sup_{1\leq i\leq n}\left|\rho'''(\tilde{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i) - \rho'''(0)\right| \leq C_\rho \sup_{1\leq i\leq n}\left|\tilde{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i\right| \leq C_\rho\left\|\hat{\boldsymbol{\lambda}}\right\|\sup_{1\leq i\leq n}\|\boldsymbol{\Xi}_i\| = o_P(1).$$

因此，

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{i=1}^{n}\rho'''(\tilde{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i)\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\hat{\boldsymbol{\lambda}} - \frac{1}{n}\sum_{i=1}^{n}\rho'''(0)\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\hat{\boldsymbol{\lambda}}\right\| \\ \leq &\frac{1}{n}\sum_{i=1}^{n}\left|\rho'''(\tilde{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i) - \rho'''(0)\right|\|\boldsymbol{\Xi}_i\|^3\left\|\hat{\boldsymbol{\lambda}}\right\|^2 \leq o_P(n^{-1})\cdot\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\Xi}_i\|^3 = o_P(n^{-1}),\end{aligned}$$

其中最后一步用到了（B.97）以及以下事实：

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\Xi}_i\|^3 &\leq \left\{\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\Xi}_i\|^4\right\}^{3/4} \leq \left\{\frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{k=1}^{K}\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^2\right\}^2\right\}^{3/4} \\ &\leq \left\{K\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^4\right\}^{3/4} = O_P(1).\end{aligned}$$

Then

$$\hat{\boldsymbol{\lambda}} = \rho''(0)^{-1}\hat{\boldsymbol{\lambda}}_{lin} - \left\{\rho''(0)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right\}^{-1}\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\rho'''(0)\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\hat{\boldsymbol{\lambda}} + o_P(n^{-1}), \qquad \text{(B.77)}$$

其中我们令

$$\hat{\boldsymbol{\lambda}}_{lin} := -\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i.$$

根据（B.71），$\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\Xi}_i\|^3 = O_P(1)$ 以及 $\left\|\hat{\boldsymbol{\lambda}}\right\| = O_P(n^{-1/2})$，我们有

$$\left\{\rho''(0)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right\}^{-1}\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\rho'''(0)\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\hat{\boldsymbol{\lambda}} = O_P(n^{-1})$$

因此 $\hat{\boldsymbol{\lambda}} = \rho''(0)^{-1}\hat{\boldsymbol{\lambda}}_{lin} + O_P(n^{-1})$。将其代入（B.77），并利用 $\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\Xi}_i\|^3 = O_P(1)$ 和 $\left\|\hat{\boldsymbol{\lambda}}\right\| = O_P(n^{-1/2})$，可得

$$\hat{\boldsymbol{\lambda}} = \rho''(0)^{-1}\hat{\boldsymbol{\lambda}}_{lin} - \left\{\rho''(0)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right\}^{-1}\frac{1}{2n}\sum_{i=1}^{n}\frac{\rho'''(0)}{\rho''(0)^3}\hat{\boldsymbol{\lambda}}_{lin}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\hat{\boldsymbol{\lambda}}_{lin} + o_P(n^{-1}). \quad \text{(B.78)}$$

由（B.8）-（B.9）可得

$$\begin{aligned}
\widehat{\tau}_{\text{cal}} - \tau &= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\left\{\frac{A_i}{\pi_{n[k]}} - \frac{1-A_i}{1-\pi_{n[k]}}\right\}\mathbb{1}(B_i = k)\cdot Y_i - \tau \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(\widehat{w}_i - 1)\sum_{k=1}^{K}\left\{\frac{A_i}{\pi_{n[k]}}\left(Y_i - \overline{Y}_{1[k]}\right) - \frac{1-A_i}{1-\pi_{n[k]}}\left(Y_i - \overline{Y}_{0[k]}\right)\right\}\mathbb{1}(B_i = k) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\left\{\frac{A_i}{\pi_{n[k]}} - \frac{1-A_i}{1-\pi_{n[k]}}\right\}\mathbb{1}(B_i = k)\cdot Y_i - \tau + \frac{1}{n}\sum_{i=1}^{n}(\rho'(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i) - 1)\widetilde{Y}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^* + \frac{1}{n}\sum_{i=1}^{n}(\rho'(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i) - 1)\widetilde{Y}_i \\
&\quad + \sum_{k=1}^{K}\left\{\mathbb{E}\left[Y_i(1) \mid B_i = k\right] - \mathbb{E}\left[Y_i(0) \mid B_i = k\right]\right\}\times\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}(B_i = k) - p_{[k]}\right).
\end{aligned} \quad \text{(B.79)}$$

关于 $\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^* + \frac{1}{n}\sum_{i=1}^{n}(\rho'(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i) - 1)\widetilde{Y}_i$ 的渐近展开。 我们首先处理 $\frac{1}{n}\sum_{i=1}^{n}(\rho'(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i) - 1)\widetilde{Y}_i$。根据泰勒展开可得

$$\frac{1}{n}\sum_{i=1}^{n}(\rho'(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i) - 1)\widetilde{Y}_i = \frac{1}{n}\sum_{i=1}^{n}\left\{\rho''(0)\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i + \frac{1}{2}\rho'''(\tilde{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i)\hat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\hat{\boldsymbol{\lambda}}\right\}\widetilde{Y}_i,$$

其中 $\tilde{\boldsymbol{\lambda}}$ 满足 $\left\|\tilde{\boldsymbol{\lambda}}\right\| \le \left\|\hat{\boldsymbol{\lambda}}\right\|$。根据 $\left\|\tilde{\boldsymbol{\lambda}}\right\| \le \left\|\hat{\boldsymbol{\lambda}}\right\| = O_P(n^{-1/2})$、(B.69) 和假设 5，可得

$$\sup_{1\le i\le n}\left|\rho'''(\tilde{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i) - \rho'''(0)\right| \le \sup_{1\le i\le n}\left|\tilde{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i\right| \le \left\|\hat{\boldsymbol{\lambda}}\right\|\sup_{1\le i\le n}\|\boldsymbol{\Xi}_i\| = o_P(1).$$

根据 $\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{\Xi}_i\|^2 \left|\widetilde{Y}_i\right| = O_P(1)$[4]和 $\widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\lambda}}_{lin} + O_P(n^{-1})$，可得

$$\frac{1}{n}\sum_{i=1}^n \frac{1}{2}\rho'''(\widetilde{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i)\widehat{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top \widehat{\boldsymbol{\lambda}}\widetilde{Y}_i = \frac{1}{n}\sum_{i=1}^n \frac{1}{2}\frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^\top \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top \widehat{\boldsymbol{\lambda}}_{lin}\widetilde{Y}_i + o_P(n^{-1}).$$

因此，我们有

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n (\rho'(\widehat{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i) - 1)\widetilde{Y}_i \\
=&\frac{1}{n}\sum_{i=1}^n \left\{\rho''(0)\widehat{\boldsymbol{\lambda}}^\top \boldsymbol{\Xi}_i + \frac{1}{2}\frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^\top \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top \widehat{\boldsymbol{\lambda}}_{lin}\right\}\left\{\boldsymbol{\Xi}_i^\top \boldsymbol{\beta}_{\mathcal{C}_n} + \epsilon_i\right\} + o_P(n^{-1}) \\
=&\widehat{\boldsymbol{\lambda}}_{lin}^\top \frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\left\{\boldsymbol{\Xi}_i^\top \boldsymbol{\beta}_{\mathcal{C}_n} + \epsilon_i\right\} \\
&-\frac{1}{2n}\sum_{i=1}^n \frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^\top \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top \boldsymbol{\Xi}_i^\top \widehat{\boldsymbol{\lambda}}_{lin}\left\{\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right\}^{-1}\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\left\{\boldsymbol{\Xi}_i^\top \boldsymbol{\beta}_{\mathcal{C}_n} + \epsilon_i\right\} \\
&+\rho''(0)\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\widetilde{Y}_i \times o_P(n^{-1}) + \frac{1}{2n}\sum_{i=1}^n \frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^\top \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top \widehat{\boldsymbol{\lambda}}_{lin}\left\{\boldsymbol{\Xi}_i^\top \boldsymbol{\beta}_{\mathcal{C}_n} + \epsilon_i\right\} + o_P(n^{-1}) \\
=&-\frac{1}{n}\sum_{i=1}^n \boldsymbol{\beta}_{\mathcal{C}_n}^\top \boldsymbol{\Xi}_i - \frac{1}{2n}\sum_{i=1}^n \frac{\rho'''(0)}{\rho''(0)^2}\left\{\widehat{\boldsymbol{\lambda}}_{lin}^\top \boldsymbol{\Xi}_i\right\}^2 \boldsymbol{\Xi}_i^\top \boldsymbol{\beta}_{\mathcal{C}_n} + \frac{1}{2n}\sum_{i=1}^n \frac{\rho'''(0)}{\rho''(0)^2}\left\{\widehat{\boldsymbol{\lambda}}_{lin}^\top \boldsymbol{\Xi}_i\right\}^2 \boldsymbol{\Xi}_i^\top \boldsymbol{\beta}_{\mathcal{C}_n} \\
&+\widehat{\boldsymbol{\lambda}}_{lin}^\top \frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\epsilon_i + \frac{1}{2n}\sum_{i=1}^n \frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^\top \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top \widehat{\boldsymbol{\lambda}}_{lin}\epsilon_i \\
&-\frac{1}{2n}\sum_{i=1}^n \frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^\top \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top \boldsymbol{\Xi}_i^\top \widehat{\boldsymbol{\lambda}}_{lin}\left\{\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\right\}^{-1}\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\epsilon_i + o_P(n^{-1}) \\
=&-\frac{1}{n}\sum_{i=1}^n \boldsymbol{\beta}_{\mathcal{C}_n}^\top \boldsymbol{\Xi}_i + \widehat{\boldsymbol{\lambda}}_{lin}^\top \frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\epsilon_i + \frac{1}{2n}\sum_{i=1}^n \frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^\top \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top \widehat{\boldsymbol{\lambda}}_{lin}\epsilon_i + o_P(n^{-1}),
\end{aligned} \tag{B.80}$$

其中，第二步由 $\widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\lambda}}_{lin} + O_P(n^{-1})$得出，第三步由$\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\widetilde{Y}_i = O_P(1)$[5]得出，最后一步由 $\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\epsilon_i = o_P(1)$ 得出（见（B.92）和

[4]这是由以下结果得出的

$$\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{\Xi}_i\|^2 \left|\widetilde{Y}_i\right| \le \sqrt{\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{\Xi}_i\|^4}\left\{\frac{1}{n}\sum_{i=1}^n \left|\widetilde{Y}_i\right|^4\right\}^{1/4} = O_P(1),$$

其中，最后一步由（B.97）和（B.99）得出。

[5]这是因为

$$\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Xi}_i\widetilde{Y}_i\right\| \le \left\{\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{\Xi}_i\|^4\right\}^{1/4}\left\{\frac{1}{n}\sum_{i=1}^n \left|\widetilde{Y}_i\right|^4\right\}^{1/4} = O_P(1),$$

其中最后一步是由（B.97）和（B.99）得出的。

$\left\|\widehat{\boldsymbol{\lambda}}_{lin}\right\| = \left\|\widehat{\boldsymbol{\lambda}}\right\| + O_P\left(n^{-1}\right) = O_P\left(n^{-1/2}\right)$。由（B.23），可得

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i^* + \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\left\{A_i - \pi_{n[1]}\right\}\mathbb{1}(B_i=1)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}\\ \left\{A_i - \pi_{n[2]}\right\}\mathbb{1}(B_i=2)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}\\ \vdots \\ \left\{A_i - \pi_{n[K]}\right\}\mathbb{1}(B_i=K)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}\end{pmatrix}.$$

根据假设6，我们有

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{A_i - \pi_{n[k]}\right\}\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}\\
=&\frac{n_{[k]}}{n}\pi_{n[k]}(1-\pi_{n[k]})\times\\
&\left\{\frac{1}{n_{1[k]}}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\} - \frac{1}{n_{0[k]}}\sum_{i=1}^{n}(1-A_i)\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}_n^*(\boldsymbol{X}_i)\right\}\right\}\\
=&o_P(n^{-1}).
\end{aligned}$$

因此，我们有

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i^* + o_P(n^{-1}). \tag{B.81}$$

将此与（B.80）结合可得

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^* + \frac{1}{n}\sum_{i=1}^{n}(\rho'(\widehat{\boldsymbol{\lambda}}^\top\boldsymbol{\Xi}_i) - 1)\widetilde{Y}_i\\
=&\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^* - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\beta}_{\mathcal{C}_n}^\top\boldsymbol{\Xi}_i^* + \widehat{\boldsymbol{\lambda}}_{lin}^\top\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\epsilon_i + \frac{1}{2n}\sum_{i=1}^{n}\frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\widehat{\boldsymbol{\lambda}}_{lin}\epsilon_i + o_P(n^{-1}).
\end{aligned} \tag{B.82}$$

现在，我们分析偏差项 $\widehat{\boldsymbol{\lambda}}_{lin}^\top\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\epsilon_i + \frac{1}{2n}\sum_{i=1}^{n}\frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^\top\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^\top\widehat{\boldsymbol{\lambda}}_{lin}\epsilon_i$。

对 $\widehat{\boldsymbol{\lambda}}_{lin}^\top\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\epsilon_i$的分析。根据 $\widehat{\boldsymbol{\lambda}}_{lin}$ 的定义，我们有

$$\widehat{\boldsymbol{\lambda}}_{lin}^\top\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\epsilon_i = -\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^\top\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top\right\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\epsilon_{i,[k]}. \tag{B.83}$$

我们将$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\epsilon_{i,[k]}$和分解为

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\epsilon_{i,[k]}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\widetilde{Y}_{i,[k]}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}.$$

我们分别计算$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,\ [k]}\boldsymbol{\Xi}_{i\ ,\ [k]}^{\top}\boldsymbol{\beta}_{[k],\ \mathcal{C}_n}$和$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,\ [k]}\widetilde{Y}_{i,\ [k]}$。

由（B.11）可得

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\\
=&\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top}-\\
&\frac{2}{n}\sum_{i=1}^{n}\pi_{n[k]}A_i\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top}+\\
&\frac{1}{n}\sum_{i=1}^{n}\pi_{n[k]}^2\mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top}+\\
&\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\left[\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}\right\}\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}\right\}^{\top}-\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top}\right]}_{R_5},
\end{aligned}$$

其中，$\overline{\boldsymbol{\xi}}_{n[k]}^*:=\frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)$是$\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)$的分层特定样本均值。对于任意$1\le k\le K$和$\boldsymbol{\alpha}\in\mathbb{R}^d$，根据柯西–施瓦茨不等式，我们有

$$\begin{aligned}
\boldsymbol{\alpha}^{\top}R_5\boldsymbol{\alpha}&=\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\left[\left\{\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}\right\}^2-\left\{\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^2\right]\\
&\le\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\{\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}-\left(\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\}^2}\times\\
&\quad\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\{\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}+\left(\boldsymbol{\alpha}^{\top}\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\boldsymbol{\alpha}^{\top}\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\}^2}\\
&\le\|\boldsymbol{\alpha}\|^2\sqrt{\underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\left(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\|^2}_{R_6}}\times\\
&\quad\sqrt{\underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}+\left(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right)\right\|^2}_{R_7}},
\end{aligned}$$

由此可得

$$\|R_5\| \leq \sqrt{R_6} \times \sqrt{R_7}.$$

我们分别控制 R_6 和 R_7。由假设6，我们有

$$\begin{aligned}\left\|\overline{\boldsymbol{\xi}}_{n[k]} - \overline{\boldsymbol{\xi}}^*_{n[k]}\right\| &= \left\|\frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\{\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right\}\right\| \\ &\leq \sqrt{\frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right\|^2} = O_P(\Delta_n) \qquad \text{(B.84)}\end{aligned}$$

对所有 $k = 1, \ldots K$。然后由假设6、大数弱律和（B.84）可得

$$R_6 \leq 2 \times \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i) - \boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right\|^2 + 2 \times \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\overline{\boldsymbol{\xi}}_{n[k]} - \overline{\boldsymbol{\xi}}^*_{n[k]}\right\|^2 = O_P(\Delta_n^2).$$

由引理3 和假设6，我们有

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right\|^2 - \mathbb{E}\left[\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right\|^2\right]\right\} = o_P(1)$$

and

$$\begin{aligned}&\overline{\boldsymbol{\xi}}^*_{n[k]} - \mathbb{E}\left[\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) \mid B_i = k\right] \\ =&\frac{n}{n_{[k]}}\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{1}(B_i = k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \mathbb{E}\left[\mathbb{1}(B_i = k)\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right]\right\} + o_P(1) \\ =&o_P(1). \qquad \text{(B.85)}\end{aligned}$$

因此，我们有

$$\begin{aligned}R_7 &\leq 2 \times R_6 + 8 \times \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}^*_n(\boldsymbol{X}_i) - \overline{\boldsymbol{\xi}}^*_{n[k]}\right\|^2 \\ &\leq O_P(\Delta_n^2) + \frac{16}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\boldsymbol{\xi}^*_n(\boldsymbol{X}_i)\right\|^2 + \frac{16}{n}\sum_{i=1}^{n}\mathbb{1}(B_i = k)\left\|\overline{\boldsymbol{\xi}}^*_{n[k]}\right\|^2 \\ &= o_P(1) + \{O(1) + o_P(1)\} + O_P(1)\{O(1) + o_P(1)\}^2 = O_P(1).\end{aligned}$$

现在，我们可以得到 $\|R_5\| \leq \sqrt{R_6} \times \sqrt{R_7} = O_P(\Delta_n)$，从而

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top} \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2 \mathbb{1}(B_i=k)\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}\left\{\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*\right\}^{\top}+O_P(\Delta_n) \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2 \mathbb{1}(B_i=k)\left\{\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)-\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\right\}\left\{\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)-\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\right\}^{\top}+O_P(\Delta_n) \\
=&\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2 \mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^{\top} \\
&-2\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2 \mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i) \\
&+\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^{*\top}\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2 \mathbb{1}(B_i=k)+O_P(\Delta_n),
\end{aligned}
\tag{B.86}
$$

其中 $\underline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^* := \frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)$。注意，$\mathbb{E}\left[\frac{n_{[k]}}{n}\underline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\right]=0$ 和

$$
\mathrm{Var}\left(\frac{n_{[k]}}{n}\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\right)=\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right)\leq\frac{1}{n}\mathbb{E}\left[\mathbb{1}(B_i=k)\left\|\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|^2\right]=O(1/n).
$$

根据马尔可夫不等式，我们可以得出

$$
\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*=O_P(n^{-1/2}). \tag{B.87}
$$

回忆一下，$\mathcal{C}_n$ 表示由 $(A^{(n)},\ B^{(n)})$ 生成的 σ-代数。我们有

$$
\mathbb{E}_{\mathcal{C}_n}\left[\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right]=\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\mathbb{E}\left[\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\mid B_i=k\right]=0
$$

and

$$
\mathrm{Var}_{\mathcal{C}_n}\left(\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right)\leq\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\right\|^2\mid B_i=k\right]=O_P(1).
$$

那么，由引理 3 可得

$$\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)=O_P(n^{-1/2}).$$

结合B.86) 和 $\widetilde{\boldsymbol{\xi}}_{n[k]}^*=O_P(n^{-1/2})$ 可得

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top&=\frac{1}{n}\sum_{i=1}^{n}\left(A_i-\pi_{n[k]}\right)^2\mathbb{1}(B_i=k)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)^\top+O_P(\Delta_n)+O_P(n^{-1})\\&=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\boldsymbol{\Xi}_{i,[k]}^{*\top}+O_P(\Delta_n)+O_P(n^{-1}).\end{aligned}\tag{B.88}$$

因此，我们有

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top\boldsymbol{\beta}_{[k],\mathcal{C}_n}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\boldsymbol{\Xi}_{i,[k]}^{*\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}+O_P(\Delta_n)+O_P(n^{-1}).\tag{B.89}$$

现在，我们计算 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,\ [k]}\widetilde{Y}_{i,\ [k]}$。由（B.18）可得

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{1[k]})\\=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*)(Y_i(1)-\overline{Y}_{1[k]})+R_8-R_9\\=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n^*(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^*)(Y_i(1)-\overline{Y}_{1[k]})+O_P(\Delta_n)\\=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)-\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*)(\widetilde{Y}_i(1)-\overline{\widetilde{Y}}_{1[k]})+O_P(\Delta_n)\\=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(1)-\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\widetilde{Y}_i(1)\\&-\overline{\widetilde{Y}}_{1[k]}\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*+\overline{\widetilde{\boldsymbol{\xi}}}_{n[k]}^*\overline{\widetilde{Y}}_{1[k]}\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i+O_P(\Delta_n),\end{aligned}\tag{B.90}$$

其中第二步由 $|R_8|\le O_P(1)\times\sqrt{R_6}=O_P(\Delta_n)$ 和 $|R_9|\le O_P(1)\times\sqrt{R_6}=O_P(\Delta_n)$ 得出，并且在第三步中我们定义 $\widetilde{Y}_{1[k]}:=\frac{1}{n_{1[k]}}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\widetilde{Y}_i(a)$。注意

即 $\mathbb{E}_{\mathcal{C}_n}\left[\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\widetilde{Y}_i(1)\right]=0$ 和

$$\mathrm{Var}_{\mathcal{C}_n}\left(\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\widetilde{Y}_i(1)\right)\leq\left\{\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\right\}^2\mathbb{E}\left[Y_i(a)^2\mid B_i=k\right]=O_P(1).$$

由引理 3 可得

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\widetilde{Y}_i(1)=O_P(n^{-1/2}).$$

类似地，我们也有

$$\overline{\widetilde{Y}}_{1[k]}=\frac{1}{n_{1[k]}}\sum_{i=1}^{n}A_i\mathbb{1}(B_i=k)\widetilde{Y}_i(a)=O_P(n^{-1/2}).\tag{B.91}$$

上述两个式子与（B.87）和（B.90）相结合，可得

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{1[k]})\\=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(1)+O_P(n^{-1})+O_P(\Delta_n).\end{aligned}$$

类似地，我们也可以推出

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\mathbb{1}(B_i=k)(1-A_i)(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{0[k]})\\=&\frac{1}{n}\sum_{i=1}^{n}\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\mathbb{1}(B_i=k)(1-A_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(0)+O_P(n^{-1})+O_P(\Delta_n).\end{aligned}$$

请注意

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\widetilde{Y}_{i,[k]}\\=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{1[k]})+\\&+\frac{1}{n}\sum_{i=1}^{n}\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\mathbb{1}(B_i=k)(1-A_i)(\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]})(Y_i-\overline{Y}_{0[k]}),\end{aligned}$$

我们有

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\widetilde{Y}_{i,[k]} \\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi_{n[k]}}{\pi_{n[k]}}\mathbb{1}(B_i=k)A_i\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(1) \\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\frac{\pi_{n[k]}}{1-\pi_{n[k]}}\mathbb{1}(B_i=k)(1-A_i)\widetilde{\boldsymbol{\xi}}_n^*(\boldsymbol{X}_i)\widetilde{Y}_i(0)+O_P(n^{-1})+O_P(\Delta_n) \\
=&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\widetilde{Y}_{i,[k]}^*+O_P(n^{-1})+O_P(\Delta_n).
\end{aligned}
$$

结合式（B.89），我们得到

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\epsilon_{i,[k]}&=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\widetilde{Y}_{i,[k]}^*-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\boldsymbol{\Xi}_{i,[k]}^{*\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}+O_P(\Delta_n)+O_P(n^{-1}) \\
&=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\epsilon_{i,[k]}^*+O_P(\Delta_n)+O_P(n^{-1}).
\end{aligned}
$$

注意到

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^*\epsilon_{i,[k]}^*-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^*\epsilon_{i,[k]}^*\right]\right\|^2\right]\leq\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^*\epsilon_{i,[k]}^*\right\|^2\right] \\
\leq&O(1)\times\left\{\sqrt{\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}_n^*\right\|^4\mid B_i=k\right]}\sqrt{\max_{a\in\{0,1\}}\mathbb{E}\left[|Y_i(a)|^4\mid B_i=k\right]}+\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}_n^*\right\|^4\mid B_i=k\right]\right\} \\
=&O_P(1).
\end{aligned}
$$

由引理 3 可得

$$
\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\epsilon_{i,[k]}^*-\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^*\epsilon_{i,[k]}^*\right]\right\|=o_P(n^{-1/2}).
$$

此外，由（B.57）可得

$$
\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^*\epsilon_{i,[k]}^*\right]=0
$$

因此，我们有 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}=o_P(n^{-1/2})$，从而

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\epsilon_i=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}+O_P(\Delta_n)+O_P(n^{-1})=O_P(\Delta_n)+o_P(n^{-1/2}). \tag{B.92}$$

回顾（B.88），我们有

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}+O_P(\Delta_n)+O_P(n^{-1}).$$

注意到

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\right\|_F^2\right]\le 2\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right\|_F^2\right]+2\left\|\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\right\|_F^2\\&\le 4\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\right\|^4\right]\le 4\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i)\right\|^4\mid B_i=k\right]=O_P(1),\end{aligned}$$

其中最后一步由假设 6 得出。根据引理 3 以及 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,\ [k]}\boldsymbol{\Xi}_{i,\ [k]}^{\top}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,\ [k]}^{*}\boldsymbol{\Xi}_{i,\ [k]}^{*\top}+O_P(\Delta_n)+O_P(n^{-1})$，我们有

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}-\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right\|=O_P(n^{-1/2})+O_P(\Delta_n), \tag{B.93}$$

其中我们定义 $\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}:=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,\ [k]}^{*}\boldsymbol{\Xi}_{i,\ [k]}^{*\top}\right]$。

现在，由（B.83）、（B.81）、（B.92）和（B.93）可得

$$\begin{aligned}&\widehat{\boldsymbol{\lambda}}_{lin}^{\top}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\epsilon_i=-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i^{\top}\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\right\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\epsilon_i\\&=-\sum_{k=1}^{K}\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*\top}+o_P(n^{-1})\right\}\left\{\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}+O_P(n^{-1/2})+O_P(\Delta_n)\right\}\\&\quad\times\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}+O_P(\Delta_n)+O_P(n^{-1})\right\}\\&=\underbrace{-\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*\top}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}\epsilon_{i,[k]}^{*}}_{\text{bias}_1}+O_P(\Delta_n n^{-1/2})+o_P(n^{-1}),\end{aligned} \tag{B.94}$$

其中最后一步由 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i^{*\top}\cdot_{[k]}=O_P(n^{-1/2})$、$\left\|\left\{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\right\}^{-1}\right\|=O(1)$ 和 (B.92).

对 $\frac{1}{2n}\sum_{i=1}^{n}\frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^{\top}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\widehat{\boldsymbol{\lambda}}_{lin}\epsilon_i$ 的分析。 根据 $\widehat{\boldsymbol{\lambda}}_{lin}$ 的定义，我们有

$$
\begin{aligned}
&\frac{1}{2n}\sum_{i=1}^{n}\frac{\rho'''(0)}{\rho''(0)^2}\widehat{\boldsymbol{\lambda}}_{lin}^{\top}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\widehat{\boldsymbol{\lambda}}_{lin}\epsilon_i\\
=&\frac{\rho'''(0)}{2\rho''(0)^2}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i^{\top}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\epsilon_i\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\boldsymbol{\Xi}_i^{\top}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_i\\
=&\frac{\rho'''(0)}{2\rho''(0)^2}\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{\top}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\epsilon_{i,[k]}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}.
\end{aligned}
\tag{B.95}
$$

我们来处理这一项 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,\ [k]}\boldsymbol{\Xi}_{i\ ,\ [k]}^{\top}\epsilon_{i,\ [k]}$。注意到

$$
\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\epsilon_{i,[k]}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\widetilde{Y}_{i,[k]}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}.
$$

我们首先分析项 $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,-[k]}\boldsymbol{\Xi}_{i\ ,\ [k]}^{\top}\boldsymbol{\Xi}_{i\ ,\ [k]}^{\top}\boldsymbol{\beta}_{[k],\ \mathcal{C}_n}$。根据柯西-施瓦茨不等式，可得

$$
\begin{aligned}
&\sup_{\boldsymbol{\alpha}\in S^{d-1}}\left|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\alpha}^{\top}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\alpha}\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\alpha}^{\top}\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\boldsymbol{\alpha}\boldsymbol{\Xi}_{i,[k]}^{*\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right|\\
=&\sup_{\boldsymbol{\alpha}\in S^{d-1}}\left|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\alpha}^{\top}\left(\boldsymbol{\Xi}_{i,[k]}-\boldsymbol{\Xi}_{i,[k]}^{*}\right)\boldsymbol{\alpha}^{\top}\left(\boldsymbol{\Xi}_{i,[k]}+\boldsymbol{\Xi}_{i,[k]}^{*}\right)\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right|+\\
&\sup_{\boldsymbol{\alpha}\in S^{d-1}}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\alpha}^{\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right)^2\left\{\boldsymbol{\Xi}_{i,[k]}^{\top}-\boldsymbol{\Xi}_{i,[k]}^{*\top}\right\}\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right|\\
\leq&\sup_{\boldsymbol{\alpha}\in S^{d-1}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{\alpha}^{\top}\boldsymbol{\Xi}_{i,[k]}-\boldsymbol{\alpha}^{\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right\}^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{\alpha}^{\top}\boldsymbol{\Xi}_{i,[k]}+\boldsymbol{\alpha}^{\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right\}^2\left\{\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right\}^2}\\
&+\sup_{\boldsymbol{\alpha}\in S^{d-1}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}-\boldsymbol{\Xi}_{i,[k]}^{*\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right\}^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\alpha}^{\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right)^4}\\
\leq&O_P(1)\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}-\boldsymbol{\Xi}_{i,[k]}^{*}\right\|^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\{\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^4+\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\right\|^2\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^2\right\}}\\
&+O_P(1)\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}-\boldsymbol{\Xi}_{i,[k]}^{*}\right\|^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\right\|^4}.
\end{aligned}
$$

首先，我们有

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}-\boldsymbol{\Xi}_{i,[k]}^{*}\right\|^{2}=\frac{1}{n}\sum_{i=1}^{n}(A_i-\pi_{n[k]})^{2}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i)\right\|^{2}\\
\leq&\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i)\right\|^{2}\\
\leq&\frac{2}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}-\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}^{*}\right\|^{2}+2\left\|\overline{\boldsymbol{\xi}}_{n[k]}^{*}-\mathbb{E}\left[\boldsymbol{\xi}_n^{*}(\boldsymbol{X}_i)\mid B_i=k\right]\right\|^{2}\\
\leq&2R_6+o_P(1)=o_P(1),
\end{aligned}\tag{B.96}$$

其中，最后两个不等式由（B.85）和 $R_6=O_P\left(\Delta_n^2\right)=o_P(\mathbf{1})$ 得出。其次，我们 have

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\right\|^{4}\leq\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\widetilde{\boldsymbol{\xi}}_n^{*}(\boldsymbol{X}_i)\right\|^{4}=O_P(1)$$

根据马尔可夫不等式和假设6。第三，我们有

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^{4}&\leq\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)-\overline{\boldsymbol{\xi}}_{n[k]}\right\|^{4}\\
&\leq\frac{8}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^{4}+\frac{8}{n}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\overline{\boldsymbol{\xi}}_{n[k]}\right\|^{4}\\
&\leq\frac{8}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^{4}+8\left|\frac{1}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|\right|^{4}\\
&\leq\frac{8}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^{4}+\frac{8}{n_{[k]}}\sum_{i=1}^{n}\mathbb{1}(B_i=k)\left\|\boldsymbol{\xi}_n(\boldsymbol{X}_i)\right\|^{4}=O_P(1),
\end{aligned}\tag{B.97}$$

其中，最后一个不等式由幂均值不等式推出，最后一个等式由假设6推出。因此，根据柯西-施瓦茨不等式可得

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\right\|^{2}\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^{2}\leq\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}^{*}\right\|^{4}\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^{4}}=O_P(1).$$

结合这三个结果，我们得到

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\Xi}_{i,[k]}^{\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^{*\top}\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right\|$$

$$
\begin{aligned}
&= \sup_{\boldsymbol{\alpha}\in S^{d-1}} \left| \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\alpha}^\top \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}_{i,[k]}^\top \boldsymbol{\alpha} \boldsymbol{\Xi}_{i,[k]}^\top \boldsymbol{\beta}_{[k],\mathcal{C}_n} - \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\alpha}^\top \boldsymbol{\Xi}_{i,[k]}^* \boldsymbol{\Xi}_{i,[k]}^{*\top} \boldsymbol{\alpha} \boldsymbol{\Xi}_{i,[k]}^{*\top} \boldsymbol{\beta}_{[k],\mathcal{C}_n} \right| \\
&= o_P(1). \qquad \text{(B.98)}
\end{aligned}
$$

接下来，我们分析项 $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_{i,\ [k]} \boldsymbol{\Xi}_{i\ ,\ [k]}^\top \widetilde{Y}_{i,\ [k]}$。根据柯西-施瓦茨不等式，我们得到

$$
\begin{aligned}
&\sup_{\boldsymbol{\alpha}\in S^{d-1}} \left| \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\alpha}^\top \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}_{i,[k]}^\top \boldsymbol{\alpha} \widetilde{Y}_{i,[k]} - \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\alpha}^\top \boldsymbol{\Xi}_{i,[k]}^* \boldsymbol{\Xi}_{i,[k]}^{*\top} \boldsymbol{\alpha} \widetilde{Y}_{i,[k]}^* \right| \\
=& \sup_{\boldsymbol{\alpha}\in S^{d-1}} \left| \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\alpha}^\top \left(\boldsymbol{\Xi}_{i,[k]} - \boldsymbol{\Xi}_{i,[k]}^* \right) \boldsymbol{\alpha}^\top \left(\boldsymbol{\Xi}_{i,[k]} + \boldsymbol{\Xi}_{i,[k]}^* \right) \widetilde{Y}_{i,[k]} \right| \\
&+ \sup_{\boldsymbol{\alpha}\in S^{d-1}} \left| \frac{1}{n}\sum_{i=1}^{n} \left(\boldsymbol{\alpha}^\top \boldsymbol{\Xi}_{i,[k]}^* \right)^2 \left(\widetilde{Y}_{i,[k]} - \widetilde{Y}_{i,[k]}^* \right) \right| \\
\leq& \sup_{\boldsymbol{\alpha}\in S^{d-1}} \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\{ \boldsymbol{\alpha}^\top \boldsymbol{\Xi}_{i,[k]} - \boldsymbol{\alpha}^\top \boldsymbol{\Xi}_{i,[k]}^* \right\}^2} \times \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\{ \boldsymbol{\alpha}^\top \boldsymbol{\Xi}_{i,[k]} + \boldsymbol{\alpha}^\top \boldsymbol{\Xi}_{i,[k]}^* \right\}^2 \left| \widetilde{Y}_{i,[k]} \right|^2} \\
&+ \sup_{\boldsymbol{\alpha}\in S^{d-1}} \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\{ \widetilde{Y}_{i,[k]} - \widetilde{Y}_{i,[k]}^* \right\}^2} \times \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(\boldsymbol{\alpha}^\top \boldsymbol{\Xi}_{i,[k]}^* \right)^4} \\
\leq& O(1) \times \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\| \boldsymbol{\Xi}_{i,[k]} - \boldsymbol{\Xi}_{i,[k]}^* \right\|^2} \times \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\{ \left\| \boldsymbol{\Xi}_{i,[k]} \right\|^2 \left| \widetilde{Y}_{i,[k]} \right|^2 + \left\| \boldsymbol{\Xi}_{i,[k]}^* \right\|^2 \left| \widetilde{Y}_{i,[k]} \right|^2 \right\}} \\
&+ O(1) \times \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left| \widetilde{Y}_{i,[k]} - \widetilde{Y}_{i,[k]}^* \right|^2} \times \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\| \boldsymbol{\Xi}_{i,[k]}^* \right\|^4}.
\end{aligned}
$$

首先，我们有

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \left| \widetilde{Y}_{i,[k]} - \widetilde{Y}_{i,[k]}^* \right|^2 \\
=& \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{A_i}{\pi_{n[k]}} \mathbb{1}(B_i = k)(Y_i - \overline{Y}_{1[k]} - \widetilde{Y}_i(1)) - \frac{1 - A_i}{1 - \pi_{n[k]}} \mathbb{1}(B_i = k)(Y_i - \overline{Y}_{0[k]} - \widetilde{Y}_i(0)) \right\}^2 \\
\leq& \frac{1}{\pi_{n[k]}^2} \frac{2}{n} \sum_{i=1}^{n} \left\{ \overline{Y}_{1[k]} - \mathbb{E}\left[Y_i(1) \mid B_i = k \right] \right\}^2 + \frac{1}{(1 - \pi_{n[k]})^2} \frac{2}{n} \sum_{i=1}^{n} \left\{ \overline{Y}_{0[k]} - \mathbb{E}\left[Y_i(0) \mid B_i = k \right] \right\}^2 \\
=& \frac{2}{\pi_{n[k]}^2} \widetilde{\overline{Y}}_{1[k]}^2 + \frac{2}{(1 - \pi_{n[k]})^2} \widetilde{\overline{Y}}_{0[k]}^2 = O_P(n^{-1/2}),
\end{aligned}
$$

其中最后一步由（B.91）得出。其次，我们有

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left|\widetilde{Y}_{i,[k]}\right|^4 &\leq \frac{8}{\pi_{n[k]}^4}\frac{1}{n}\sum_{i=1}^{n}(Y_i(1)-\overline{Y}_{1[k]})^4+\frac{8}{(1-\pi_{n[k]})^4}\frac{1}{n}\sum_{i=1}^{n}(Y_i(0)-\overline{Y}_{1[k]})^4\\ &\leq \frac{64}{\pi_{n[k]}^4}\frac{1}{n}\sum_{i=1}^{n}Y_i(1)^4+\frac{64}{\pi_{n[k]}^4}\overline{Y}_{1[k]}^4+\frac{64}{(1-\pi_{n[k]})^4}\frac{1}{n}\sum_{i=1}^{n}Y_i(0)^4+\frac{64}{(1-\pi_{n[k]})^4}\overline{Y}_{0[k]}^4\\ &= O_P(1),\end{aligned} \tag{B.99}$$

其中，最后一步由 $\max_{a\in\{0,1\}}\mathbb{E}\left[|Y_i(a)|^4\right]<\infty$, $Y_{1[k]}=O_P(1)$ 和 $Y_{0[k]}=O_P(1)$ 得出。回想一下，

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}^*\right\|^4=O_P(1) \text{ and } \frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^4=O_P(1),$$

由柯西-施瓦茨不等式可得

$$\begin{aligned}&\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\{\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^2\left|\widetilde{Y}_{i,[k]}\right|^2+\left\|\boldsymbol{\Xi}_{i,[k]}^*\right\|^2\left|\widetilde{Y}_{i,[k]}\right|^2\right\}}\\ \leq&\sqrt{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}\right\|^4\frac{1}{n}\sum_{i=1}^{n}\left|\widetilde{Y}_{i,[k]}\right|^4}+\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\Xi}_{i,[k]}^*\right\|^4\frac{1}{n}\sum_{i=1}^{n}\left|\widetilde{Y}_{i,[k]}\right|^4}}=O_P(1).\end{aligned}$$

结合这些结果与（B.96），我们可以得到

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top\widetilde{Y}_{i,[k]}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\boldsymbol{\Xi}_{i,[k]}^{*\top}\widetilde{Y}_{i,[k]}^*\right\|\\ =&\sup_{\boldsymbol{\alpha}\in S^{d-1}}\left|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\alpha}^\top\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top\boldsymbol{\alpha}\widetilde{Y}_{i,[k]}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\alpha}^\top\boldsymbol{\Xi}_{i,[k]}^*\boldsymbol{\Xi}_{i,[k]}^{*\top}\boldsymbol{\alpha}\widetilde{Y}_{i,[k]}^*\right|=o_P(1).\end{aligned}$$

回顾（B.98），于是我们有

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}\boldsymbol{\Xi}_{i,[k]}^\top\epsilon_{i,[k]}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^*\boldsymbol{\Xi}_{i,[k]}^{*\top}\epsilon_{i,[k]}^*\right\|=o_P(1). \tag{B.100}$$

请注意

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}\right]\right\|_F^{4/3}\right]\\
\leq& 2^{4/3-1}\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}\right\|_F^{4/3}+\left\|\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}\right]\right\|_F^{4/3}\right]\\
\leq& 2\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}\right\|_F^{4/3}\right]+2\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}\right\|_F\right]^{4/3}\leq 4\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}\right\|_F^{4/3}\right]\\
\leq& 8\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\widetilde{Y}^*_{i,[k]}\right\|_F^{4/3}\right]+8\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right\|_F^{4/3}\right]\\
\leq& \frac{16}{\pi^{4/3}_{n[k]}}\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\widetilde{Y}_i(1)\right\|_F^{4/3}\mid B_i=k\right]\\
&+\frac{16}{(1-\pi_{n[k]})^{4/3}}\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\widetilde{Y}_i(0)\right\|_F^{4/3}\mid B_i=k\right]\\
&+8\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\right\|^4\mid B_i=k\right]\left\|\boldsymbol{\beta}_{[k],\mathcal{C}_n}\right\|^{4/3}
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\widetilde{Y}_i(a)\right\|_F^{4/3}\mid B_i=k\right]\\
=&\mathbb{E}\left[\mathrm{tr}\left\{\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)^\top\widetilde{Y}_i(a)^2\right\}^{2/3}\mid B_i=k\right]\\
=&\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\right\|^{8/3}\left|\widetilde{Y}_i(a)\right|^{4/3}\mid B_i=k\right]\\
\leq&\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\xi}}^*_n(\boldsymbol{X}_i)\right\|^{(8/3)\times(3/2)}\mid B_i=k\right]^{2/3}\mathbb{E}\left[\left|\widetilde{Y}_i(a)\right|^{(4/3)\times 3}\mid B_i=k\right]^{1/3}\\
<&\infty,
\end{aligned}
$$

其中，最后两步由 Hölder不等式、假设 6 以及 $\mathbb{E}\left[\left|\widetilde{Y}_i(a)\right|^4\right]<\infty$得出。由此可得

$$
\mathbb{E}_{\mathcal{C}_n}\left[\left\|\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}-\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}\right]\right\|_F^{4/3}\right]=O_P(1).
$$

根据引理 3，我们有

$$
\left\|\frac{1}{n}\sum_{i=1}^n\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}-\frac{1}{n}\sum_{i=1}^n\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^*_{i,[k]}\boldsymbol{\Xi}^{*\top}_{i,[k]}\epsilon^*_{i,[k]}\right]\right\|
$$

$$\leq \left\| \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}^{*}_{i,[k]} \boldsymbol{\Xi}^{*\top}_{i,[k]} \epsilon^{*}_{i,[k]} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^{*}_{i,[k]} \boldsymbol{\Xi}^{*\top}_{i,[k]} \epsilon^{*}_{i,[k]} \right] \right\|_F = o_P(1).$$

结合式（B.100）可得

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_{i,[k]} \boldsymbol{\Xi}^{\top}_{i,[k]} \epsilon_{i,[k]} &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^{*}_{i,[k]} \boldsymbol{\Xi}^{*\top}_{i,[k]} \epsilon^{*}_{i,[k]} \right] + o_P(1) \\ &= \boldsymbol{\Sigma}^{\mathcal{C}_n}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]} + o_P(1). \end{aligned}$$

现在，我们回顾一下（B.95）。由（B.81）、（B.93）和 $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}^{*\top}_{i,[k]} = O_P(n^{-1/2})$ 可得

$$\begin{aligned} &\frac{1}{2n}\sum_{i=1}^{n} \frac{\rho'''(0)}{\rho''(0)^2} \hat{\boldsymbol{\lambda}}^{\top}_{lin} \boldsymbol{\Xi}_i \boldsymbol{\Xi}^{\top}_i \hat{\boldsymbol{\lambda}}_{lin} \epsilon_i \\ =&\frac{\rho'''(0)}{2\rho''(0)^2} \sum_{k=1}^{K} \left\{ \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}^{*\top}_{i,[k]} + o_P(n^{-1}) \right\} \left(\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} + o_P(1) \right)^{-1} \\ &\quad \times \left\{ \boldsymbol{\Sigma}^{\mathcal{C}_n}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]} + o_P(1) \right\} \left(\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} + o_P(1) \right)^{-1} \left\{ \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}^{*}_{i,[k]} + o_P(n^{-1}) \right\} \\ =&\underbrace{\frac{\rho'''(0)}{2\rho''(0)^2} \sum_{k=1}^{K} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}^{*\top}_{i,[k]} \left(\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} \right)^{-1} \boldsymbol{\Sigma}^{\mathcal{C}_n}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]} \left(\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} \right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}^{*}_{i,[k]}}_{\text{bias}_2} + o_P(n^{-1}). \end{aligned} \tag{B.101}$$

结合（B.94）和（B.101），我们得到

$$\hat{\boldsymbol{\lambda}}^{\top}_{lin} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Xi}_i \epsilon_i + \frac{1}{2n}\sum_{i=1}^{n} \frac{\rho'''(0)}{\rho''(0)^2} \hat{\boldsymbol{\lambda}}^{\top}_{lin} \boldsymbol{\Xi}_i \boldsymbol{\Xi}^{\top}_i \hat{\boldsymbol{\lambda}}_{lin} \epsilon_i = \text{bias}_1 + \text{bias}_2 + O_P(\Delta_n n^{-1/2}) + o_P(n^{-1}). \tag{B.102}$$

注意到（$\boldsymbol{\Xi}^{*}_{i,[k]}$、$\epsilon^{*}_{i,[k]}$）、$i = 1$、……、$n$ 在$\mathcal{C}_n$条件下相互独立，因此我们有

$$\begin{aligned} \mathbb{E}_{\mathcal{C}_n}\left[\text{bias}_1\right] &= -\sum_{k=1}^{K} \frac{1}{n^2} \mathbb{E}_{\mathcal{C}_n}\left[\sum_{1\leq i,j\leq n} \boldsymbol{\Xi}^{*\top}_{i,[k]} \left(\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} \right)^{-1} \boldsymbol{\Xi}^{*}_{j,[k]} \epsilon^{*}_{j,[k]} \right] \\ &= -\sum_{k=1}^{K} \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}^{*\top}_{i,[k]} \left(\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} \right)^{-1} \boldsymbol{\Xi}^{*}_{i,[k]} \epsilon^{*}_{i,[k]} \right] \\ &= -\sum_{k=1}^{K} \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}_{\mathcal{C}_n}\left[\text{tr}\left\{ \boldsymbol{\Xi}^{*\top}_{i,[k]} \left(\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} \right)^{-1} \boldsymbol{\Xi}^{*}_{i,[k]} \epsilon^{*}_{i,[k]} \right\} \right] \\ &= -\sum_{k=1}^{K} \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}_{\mathcal{C}_n}\left[\text{tr}\left\{ \left(\boldsymbol{\Sigma}^{\mathcal{C}_n}_{[k]} \right)^{-1} \boldsymbol{\Xi}^{*}_{i,[k]} \boldsymbol{\Xi}^{*\top}_{i,[k]} \epsilon^{*}_{i,[k]} \right\} \right] \end{aligned}$$

$$
\begin{aligned}
&= -\sum_{k=1}^{K}\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{tr}\left\{\mathbb{E}_{\mathcal{C}_n}\left[\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\epsilon_{i,[k]}^{*}\right]\right\}\\
&= -\sum_{k=1}^{K}\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{tr}\left\{\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\epsilon_{i,[k]}^{*}\right]\right\}\\
&= -\sum_{k=1}^{K}\frac{1}{n}\mathrm{tr}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\epsilon_{i,[k]}^{*}\right]\right\}\\
&= -\sum_{k=1}^{K}\frac{1}{n}\mathrm{tr}\left\{\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\epsilon_{i,[k]}^{*}\right]\right\}\\
&= -\frac{1}{n}\sum_{k=1}^{K}\mathrm{tr}\left\{\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}^{\mathcal{C}_n}\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}_{\mathcal{C}_n}\left[\mathrm{bias}_2\right] &= \mathbb{E}_{\mathcal{C}_n}\left[\frac{\rho'''(0)}{2\rho''(0)^2}\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*\top}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\epsilon_{i,[k]}^{*}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Xi}_{i,[k]}^{*}\right]\\
&= \frac{\rho'''(0)}{2\rho''(0)^2}\sum_{k=1}^{K}\frac{1}{n^2}\mathbb{E}_{\mathcal{C}_n}\left[\sum_{1\le i,j\le n}\boldsymbol{\Xi}_{i,[k]}^{*\top}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}^{\mathcal{C}_n}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Xi}_{j,[k]}^{*}\right]\\
&= \frac{\rho'''(0)}{2\rho''(0)^2}\sum_{k=1}^{K}\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\mathrm{tr}\left\{\boldsymbol{\Xi}_{i,[k]}^{*\top}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}^{\mathcal{C}_n}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Xi}_{i,[k]}^{*}\right\}\right]\\
&= \frac{\rho'''(0)}{2\rho''(0)^2}\sum_{k=1}^{K}\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\mathrm{tr}\left\{\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}^{\mathcal{C}_n}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right\}\right]\\
&= \frac{\rho'''(0)}{2\rho''(0)^2}\sum_{k=1}^{K}\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{tr}\left\{\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}^{\mathcal{C}_n}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\right\}\\
&= \frac{\rho'''(0)}{2\rho''(0)^2}\sum_{k=1}^{K}\frac{1}{n}\mathrm{tr}\left\{\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}^{\mathcal{C}_n}\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{C}_n}\left[\boldsymbol{\Xi}_{i,[k]}^{*}\boldsymbol{\Xi}_{i,[k]}^{*\top}\right]\right\}\\
&= \frac{\rho'''(0)}{2\rho''(0)^2}\frac{1}{n}\sum_{k=1}^{K}\mathrm{tr}\left\{\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}^{\mathcal{C}_n}\right\}.
\end{aligned}
$$

因此，我们有

$$
\mathbb{E}\left[\mathrm{bias}_1+\mathrm{bias}_2\right] = \mathbb{E}\left[\mathbb{E}_{\mathcal{C}_n}\left[\mathrm{bias}_1+\mathrm{bias}_2\right]\right] = \left\{\frac{\rho'''(0)}{2\rho''(0)^2}-1\right\}\frac{1}{n}\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbb{E}\left[\left(\boldsymbol{\Sigma}_{[k]}^{\mathcal{C}_n}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\Xi}\boldsymbol{\Xi}\epsilon[k]}^{\mathcal{C}_n}\right]\right\}.
$$

Let

$$
\psi_{1,i} := \sum_{k=1}^{K}\left\{\left(\frac{A_i}{\pi_{n[k]}}-\frac{1-A_i}{1-\pi_{n[k]}}\right)\mathbb{1}(B_i=k)\cdot Y_i-\tau-\boldsymbol{\beta}_{[k],\mathcal{C}_n}^{\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right\},
$$

$$\phi_{1,i} := \sum_{k=1}^{K} \{\mathbb{E}[Y_i(1) \mid B_i = k] - \mathbb{E}[Y_i(0) \mid B_i = k]\} \{\mathbb{1}(B_i = k) - p_{[k]}\} + \sum_{k=1}^{K} \left(\widetilde{Y}^*_{i,[k]} - \boldsymbol{\beta}^\top_{[k],\mathcal{C}_n} \boldsymbol{\Xi}^*_i\right)$$

and

$$\psi_2 := n\{\text{bias}_1 + \text{bias}_2\},$$

由 (B.79)、(B.82) 和 (B.102) 可得

$$\begin{aligned}
\widehat{\tau}_{\text{cal}} - \tau &= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} \{\mathbb{E}[Y_i(1) \mid B_i = k] - \mathbb{E}[Y_i(0) \mid B_i = k]\} \{\mathbb{1}(B_i = k) - p_{[k]}\} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} \left(\widetilde{Y}^*_{i,[k]} - \boldsymbol{\beta}^\top_{[k],\mathcal{C}_n} \boldsymbol{\Xi}^*_{i,[k]}\right) \\
&\quad + \text{bias}_1 + \text{bias}_2 + O_P(\Delta_n n^{-1/2}) + o_P(n^{-1}) \\
&= \frac{1}{n}\sum_{i=1}^{n} \phi_{1,i} + \frac{1}{n}\psi_2 + O_P(\Delta_n n^{-1/2}) + o_P(n^{-1}). \\
&= \frac{1}{n}\sum_{i=1}^{n} \psi_{1,i} + \frac{1}{n}\psi_2 + O_P(\Delta_n n^{-1/2}) + o_P(n^{-1}) + \frac{1}{n}\sum_{i=1}^{n} \phi_{1,i} - \frac{1}{n}\sum_{i=1}^{n} \psi_{1,i}
\end{aligned}$$

现在，剩下要证明的是

$$\frac{1}{n}\sum_{i=1}^{n} \phi_{1,i} = \frac{1}{n}\sum_{i=1}^{n} \psi_{1,i}$$

and

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \psi_{1,i}\right] = 0.$$

首先，由 $\pi_{n[k]} = \frac{1}{n_{[k]}} \sum_{i=1}^{n} \mathbb{1}(B_i = k) A_i$ 可得

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \psi_{1,i} \\
=&\sum_{k=1}^{K} \frac{1}{n}\sum_{i=1}^{n} \frac{A_i}{\pi_{n[k]}} \mathbb{1}(B_i = k) \cdot \{Y_i(1) - \mathbb{E}[Y_i(1) \mid B_i = k] + \mathbb{E}[Y_i(1) \mid B_i = k]\} \\
&- \sum_{k=1}^{K} \frac{1}{n}\sum_{i=1}^{n} \frac{1 - A_i}{1 - \pi_{n[k]}} \mathbb{1}(B_i = k) \cdot \{Y_i(1) - \mathbb{E}[Y_i(0) \mid B_i = k] + \mathbb{E}[Y_i(0) \mid B_i = k]\} \\
&- \sum_{k=1}^{K} p_{[k]} \{\mathbb{E}[Y_i(1) \mid B_i = k] - \mathbb{E}[Y_i(0) \mid B_i = k]\} - \sum_{k=1}^{K} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\beta}^\top_{[k],\mathcal{C}_n} \boldsymbol{\Xi}^*_{i,[k]} \\
=&\sum_{k=1}^{K} \frac{1}{n}\sum_{i=1}^{n} \frac{A_i}{\pi_{n[k]}} \mathbb{1}(B_i = k) \widetilde{Y}_i(1) - \sum_{k=1}^{K} \frac{1}{n}\sum_{i=1}^{n} \frac{1 - A_i}{1 - \pi_{n[k]}} \mathbb{1}(B_i = k) \widetilde{Y}_i(0)
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{k=1}^{K}\frac{n_{[k]}}{n}\left\{\mathbb{E}\left[Y_i(1)\mid B_i=k\right]-\mathbb{E}\left[Y_i(0)\mid B_i=k\right]\right\}\\
&-\sum_{k=1}^{K}p_{[k]}\left\{\mathbb{E}\left[Y_i(1)\mid B_i=k\right]-\mathbb{E}\left[Y_i(0)\mid B_i=k\right]\right\}-\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\beta}_{[k],\mathcal{C}_n}^{\top}\boldsymbol{\Xi}_{i,[k]}^{*}\\
=&\frac{1}{n}\sum_{i=1}^{n}\phi_{1,i}.
\end{aligned}
$$

其次，由此可得

$$
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\left\{\mathbb{E}\left[Y_i(1)\mid B_i=k\right]-\mathbb{E}\left[Y_i(0)\mid B_i=k\right]\right\}\left\{\mathbb{1}(B_i=k)-p_{[k]}\right\}\right]=0
$$

and

$$
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\widetilde{Y}_{i,[k]}^{*}-\boldsymbol{\beta}_{[k]}^{*\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right)\right]=\sum_{k=1}^{K}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{E}_{\mathcal{C}_n}\left[\left(\widetilde{Y}_{i,[k]}^{*}-\boldsymbol{\beta}_{[k],\mathcal{C}_n}^{\top}\boldsymbol{\Xi}_{i,[k]}^{*}\right)\right]\right]=0
$$

that

$$
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\phi_{1,i}\right]=0.
$$

因此，$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\psi_{1,\,i}\right]=\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\phi_{1,\,i}\right]=0$。证明完成 . ■

附录 C 附加模拟结果

我们考虑以下模型，该模型假设条件均值函数 $g_0(\boldsymbol{X})$ 和 $g_1(\boldsymbol{X})$ 在不同层之间存在异质性。

模型4。在此模型中，我们设定

$$
\begin{aligned}
g_0(\boldsymbol{X}_i)&=\mu_0+\left(\beta_{01}X_{i1}+\beta_{02}X_{i2}\right)S_i+\beta_{03}\log(X_{i1}+1)\mathbb{1}(S_i=1)\\
g_1(\boldsymbol{X}_i)&=\mu_1+\left(\beta_{11}X_{i1}+\beta_{12}X_{i2}\right)S_i+\beta_{13}\exp(X_{i2})\mathbb{1}(S_i=-1),
\end{aligned}
$$

其中 $\mu_0=5$, $\mu_1=5$, $(\beta_{01},\ \beta_{02},\ \beta_{03})=(20,\ 30,\ 50)$，以及 $(\beta_{11},\ \beta_{12},\ \beta_{13})=(20,\ 30,\ 65)$。此外，变量设定如下：$\epsilon_{0,\,i}\sim N(0,\ 1)$，$\epsilon_{1,\,i}\sim N(0,\ 9)$，$X_{i1}\sim$贝塔分布$(3,\ 4)$，

表7：模型4下不同估计量性能的比较。

n	估计量	简单随机化			分层区组随机化				最小化法				
		Bias	SD	SE	覆盖概率	偏倚	SD	SE	覆盖概率	偏倚	SD	SE	CP
500	cal_rf	0.08	4.67	5.13	0.957	0.80	5.07	5.08	0.940	0.24	5.14	5.11	0.937
	cal_nn	0.18	4.96	5.55	0.977	0.58	5.37	5.49	0.953	0.08	5.50	5.51	0.937
	cal_rfnn	0.13	4.65	5.14	0.967	0.97	5.02	5.09	0.933	0.40	5.11	5.12	0.930
	cal_rflin	0.08	4.67	5.11	0.960	0.73	5.02	5.05	0.943	0.15	5.14	5.08	0.933
	cal_rf_g	0.09	4.64	5.14	0.953	0.82	5.10	5.08	0.937	0.21	5.16	5.11	0.947
	cal_nn_g	0.18	5.03	5.46	0.970	0.72	5.32	5.41	0.963	0.15	5.34	5.43	0.940
	cal_lin_EL	0.37	4.72	5.29	0.967	0.41	5.14	5.24	0.953	0.07	5.43	5.27	0.933
	aipw_rf	0.42	4.82	5.15	0.950	0.35	5.07	5.10	0.947	0.25	5.17	5.13	0.937
	aipw_nn	0.47	6.27	7.02	0.967	0.36	6.80	6.96	0.950	0.59	6.90	6.92	0.953
	aipw_lin	0.34	5.08	5.49	0.973	0.57	5.29	5.42	0.950	0.18	5.68	5.46	0.933
	sdim	0.33	5.58	5.84	0.960	0.31	5.79	5.79	0.963	0.15	6.00	5.81	0.927
1000	cal_rf	0.00	3.40	3.56	0.970	0.19	3.59	3.56	0.940	0.28	3.58	3.55	0.953
	cal_nn	0.05	3.45	3.68	0.973	0.11	3.76	3.69	0.933	0.19	3.74	3.68	0.947
	cal_rfnn	0.15	3.40	3.55	0.963	0.35	3.59	3.55	0.933	0.42	3.57	3.54	0.943
	cal_rflin	0.00	3.39	3.54	0.963	0.18	3.61	3.54	0.937	0.27	3.58	3.54	0.953
	cal_rf_g	0.00	3.42	3.56	0.960	0.18	3.59	3.56	0.943	0.30	3.59	3.55	0.947
	cal_nn_g	0.09	3.40	3.65	0.970	0.12	3.71	3.66	0.933	0.25	3.71	3.65	0.950
	cal_lin_EL	0.23	3.43	3.71	0.980	0.10	3.71	3.71	0.940	0.01	3.69	3.70	0.957
	aipw_rf	0.23	3.40	3.57	0.970	0.09	3.65	3.57	0.933	0.05	3.59	3.56	0.947
	aipw_nn	0.28	3.83	4.02	0.963	0.13	4.07	4.04	0.943	0.05	4.16	4.04	0.957
	aipw_lin	0.23	3.45	3.75	0.977	0.15	3.76	3.75	0.937	0.05	3.73	3.73	0.957
	sdim	0.30	3.89	4.11	0.980	0.15	4.29	4.11	0.930	0.01	4.11	4.10	0.937
2000	cal_rf	0.37	2.57	2.49	0.933	0.35	2.61	2.48	0.940	0.20	2.44	2.50	0.953
	cal_nn	0.37	2.56	2.51	0.933	0.36	2.66	2.50	0.933	0.21	2.49	2.52	0.953
	cal_rfnn	0.46	2.55	2.47	0.930	0.45	2.62	2.47	0.933	0.12	2.42	2.49	0.957
	cal_rflin	0.37	2.57	2.48	0.920	0.35	2.61	2.48	0.940	0.23	2.44	2.49	0.950
	cal_rf_g	0.39	2.57	2.49	0.933	0.37	2.62	2.48	0.940	0.17	2.43	2.50	0.953
	cal_nn_g	0.37	2.57	2.51	0.933	0.37	2.66	2.50	0.930	0.22	2.50	2.52	0.957
	cal_lin_EL	0.29	2.64	2.61	0.920	0.14	2.71	2.60	0.947	0.35	2.62	2.62	0.953
	aipw_rf	0.26	2.58	2.49	0.940	0.19	2.61	2.49	0.937	0.33	2.48	2.50	0.950
	aipw_nn	0.29	2.57	2.56	0.940	0.21	2.65	2.56	0.947	0.35	2.58	2.57	0.957
	aipw_lin	0.32	2.63	2.61	0.930	0.09	2.73	2.61	0.943	0.38	2.62	2.63	0.960
	sdim	0.33	2.86	2.90	0.937	0.10	3.07	2.90	0.940	0.36	2.94	2.91	0.950

缩写：Rand.，随机化；SD，标准差，SE：标准误；CP，覆盖概率。

$X_{i2} \sim$ 均匀分布$(-2,\ 2)$ 这两个变量相互独立。首先按照模型1生成额外协变量 X_{i3}、...、X_{ip}。然后从这些额外协变量中随机选取 $\lfloor p/3 \rfloor$ 个，并以相同概率分别乘以 X_{i1} 或 X_{i2}，从而构成最终的额外协变量。随机化变量S_i 以相同概率取 $\{1,\ -1\}$ 中的值，并且与 X_{ij} 对于 $j = 1$、.....、p相互独立。模型4的模拟结果见表 7。当样本量较大时

（$n = 2000$），调整后的估计量性能几乎相同，其中 `cal_rf_g`略好，而 `aipw_nn`略差。然而，在样本量较小时（$n = 500,$ 1000），会出现两个重要发现：（i）校准估计量优于AIPW估计量；（ii）基于随机森林的估计量优于基于神经网络的估计量，而基于线性回归的估计量表现居于基于随机森林和基于神经网络的估计量之间。

参考文献

Bernstein, D. S. (2009)，矩阵数学：理论、事实与公式，第2版，普林斯顿大学出版社，新泽西州普林斯顿。

Billingsley, P. (1968)，概率测度的收敛，威利出版社，纽约。

Boyd, S. P. & Vandenberghe, L.(2004)，《凸优化》，剑桥大学出版社，英国剑桥；纽约。

Bugni, F. A., Canay, I. A. & Shaikh, A. M. (2018)，‘协变量自适应随机化下的推断’，*Journal of the American Statistical Association* **113**(524), 1784–1796。

Bulinski, A. V. (2017), ‘条件中心极限定理’, 概率论及其应用**61**(4), 613–631.

Burkholder, D. L. (1988), ‘鞅与随机积分的尖锐不等式’, *Astérisque***157**(158), 75–94.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018), ‘用于处理效应和结构参数的双重／去偏机器学习’, 计量经济学杂志**21**(1), C1–C68.

Davidson,J.(2021),随机极限理论：计量经济学家导论，第2版，牛津大学出版社，牛津。

Ma, W., Tu, F. & Liu, H. (2022), ‘协变量自适应随机化的回归分析：稳健且有效的推断视角’，医学统计学 41(29), 5645–5661.

Rafi, A. (2023)，《协变量自适应随机化下平均处理效应的有效半参数估计》。arXiv 预印本 arXiv:2305.08340。

Stewart, G. W. (1977)，《伪逆、投影和线性最小二乘问题的扰动》，*SIAM*评论 19(4), 634–661。

Tropp, J. A. (2015)，《矩阵集中不等式导论》，基础与趋势® 《机器学习》8(1-2), 1–230。

Tseng, P. & Bertsekas, D. P. (1991)，《具有严格凸成本和线性约束的问题的松弛方法》，运筹学数学 16(3), 462–481。van der Vaart, A. W. (1998)，《渐近统计学》，剑桥大学出版社，剑桥。Vershynin, R. (2018)，《高维概率：数据科学中的导论与应用》，剑桥大学出版社，剑桥。