基于自适应收集的数据的半参数推断

LicongLin[†], Koulik Khamaru[⋆], Martin J. Wainwright[◇,†,‡]

电气工程与计算机科学系[◇]统计系[†]加州大学伯克利分校

统计系[⋆]罗格斯大学

信息与决策系统实验室[‡]统计与数据科学中心[‡]电气
工程与计算机科学及数学系 麻省理工学院

2025年3月4日

arXiv:2303.02534v2 [math.ST] 2 Mar 2025

摘要

许多标准估计量应用于自适应收集的数据时，并不具备渐近正态性，从而使置信区间的构造更加复杂。我们在半参数背景下应对这一挑战：估计受到非参数干扰成分污染的广义线性回归模型的参数向量。我们构造了适当加权的估计方程，以考虑数据收集中的自适应性，并给出相关估计值具备渐近正态性的条件。我们的结果刻画了渐近正态性成立所需的"可探索性"程度。对于估计线性泛函这一更简单的问题，我们在弱得多的假设下给出了类似的保证。我们以包括标准线性老虎机和稀疏广义老虎机在内的各种问题为例，说明了我们的通用理论所带来的具体结论，并通过模拟研究与其他方法进行了比较。

1 引言

半参数统计中的一个典型问题，是在存在高维或非参数干扰成分的情况下估计低维参数。一个标准目标是获得同时满足 $\sqrt{n}$一致性和渐近正态性的估计量；这些性质简化了构造渐近有效置信区间和假设检验的任务。关于这一主题，目前已有丰富的文献（例如，[9, 10, 52, 4, 51, 3, 59, 13]）；然而，这些研究成果大多涉及由独立同分布（或弱相关）样本组成的数据集，在这种情况下，中心极限定理等标准渐近结果成立。

本文关注的是这类假设不再成立的情形。具体而言，我们考虑一种模型，该模型允许数据集以自适应方式收集；具体来说，第（i+1）个数据点的分布可以依赖于此前的 i 样本。这类自适应收集的数据集出现在各种应用中，

包括老虎机实验 [39], 主动学习 [22], 时间序列建模 [12], 自适应随机逼近方案 [17, 38],以及动态治疗方案。

本文的主要贡献是提出并分析一类估计量，即使在允许存在相当一般的序贯依赖的数据收集模型下，这类估计量仍具有渐近正态性。我们在广义部分线性回归的半参数框架内开展研究。在这类模型中，标量响应变量 y 通过方程与协变量向量 $x \in \mathbb{R}^{d_T}$ 和辅助向量 $z \in \mathbb{R}^{d_N}$ 相关联:

$$y_i = g\big(\langle x_i, \theta^* \rangle + h^*(z_i)\big) + \varepsilon_i. \tag{1}$$

这里， $\{\varepsilon_i\}_{i\geq 1}$ 是一个独立同分布的噪声序列；函数 $g : \mathbb{R} \to \mathbb{R}$ 称为逆链接函数；向量 $\theta^* \in \mathbb{R}^{d_T}$ 是我们感兴趣的目标参数；而 $h^* : \mathbb{R}^{d_N} \to \mathbb{R}$ 是一个高维（或非参数）干扰分量。我们假设，在第 i 轮，协变量与辅助变量对 (x_i, z_i)可以依赖于此前观测值的集合 $\big\{\big(x_j, z_j, y_j\big)\big\}_{j=1}^{i-1}$。

作为一个示例，部分线性回归模型是一般设定（1）的一个特例，它出现在治疗分配问题中（例如， [56, 64, 21, 62, 57]）。给定一组 d_T 药物，目标是确定其中最有效的一种。为此，我们开展一项涉及一组 n 患者的序贯实验，在每一轮中，我们都要决定是分配一种 d_T 药物，还是不予治疗（这可能对应于对照组）。对于给定的患者索引 $i \in [n] := \{1, \ldots, n\}$，分配药物 $k \in [d_T]$ 的决定通过设置回归向量 $x_i = e_k$来编码；该二元指示向量只有第 k个位置为1。另一方面，分配到对照组则编码为设置 $x_i = \mathbf{0}$，对应全零向量。采用这些设定后，如果我们为患者分配药物 k，响应变量 y_i 就是 θ_k^* 的一个含噪版本；如果分配到对照组，则响应变量仅包含噪声。在这一设定中，选择协变量向量的各种自适应程序都很自然。例如，医生可能会根据患者 i 的个人信息 z_i以及先前患者的历史数据 $\{ (x_j, z_j, y_j) \}_{j=1}^{i-1}$来决定对其采取何种治疗。

1.1 可视化自适应性下的失效

为了说明我们提出的方法，直观展示在独立同分布抽样下有效的经典保证在数据点以自适应方式收集时如何失效，会很有帮助。一个简单的例子便足以说明这一现象：具体而言，我们考虑如下线性模型

$$y_i = \langle x_i, \theta^* \rangle + \langle z_i, \eta^* \rangle + \varepsilon_i, \tag{2}$$

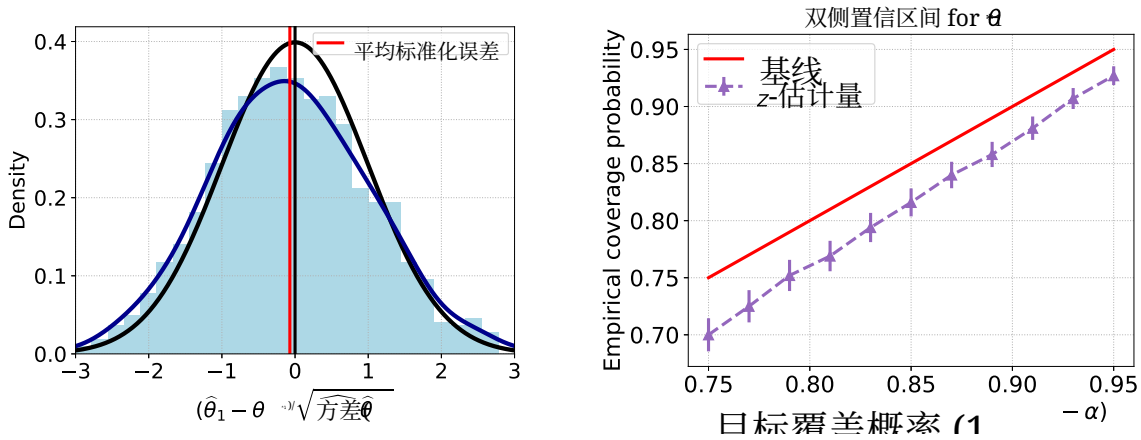
其中涉及目标参数 $\theta^* \in \mathbb{R}^{d_T}$ 和干扰参数 $\eta \in \mathbb{R}^{d_N}$。这是一般设定的一个特例，其中链接函数为 $g(x) = x$ ，干扰函数为$h(z) = \langle z, \eta^* \rangle$。给定干扰向量 η^* 的一个估计量 $\widehat{\eta}$，可通过定义得分函数，得到目标参数的标准 Z-估计量 θ

$$\psi_i(y_i, x_i, z_i, \theta, \eta) := (x_i - p_i)\big\{y_i - \langle x_i, \theta \rangle - \langle z_i, \eta \rangle\big\}, \tag{3a}$$

然后求解估计方程

$$\sum_{i=1}^{n} \psi_i(y_i, x_i, z_i, \widetilde{\theta}, \widehat{\eta}) = 0. \tag{3b}$$

在定义中　得分函数 ψ_i 的方程 (3a) 中，向量 p_i 是给定



(a) (b)

图 1。 (a)： Z-估计量（3b）第一个坐标的标准化估计误差θ_1^*；图中显示的是基于 1000 次试验的直方图。(b)：在参数为（d_T， d_N， n） = （2， 1000， 950）的模拟中， θ_1^*的双侧置信区间的经验覆盖概率。详情请参见第4.1节。

过去数据点的条件均值。该 Z-估计量是一种经过充分研究的方法 [52]；更多详情请参阅第2.1节和方程 (40)。当数据点独立同分布时，可以证明 [13] 该估计量 θ 是 $\sqrt{n}$-一致的，且渐近正态。

然而，当数据以自适应方式收集时，这些有吸引力的保证可能不再成立。为展示这种失效情况，我们在参数为（d_T， d_N） = （2， 1000）的线性模型（2）上进行了实验；为了应用 LASSO 多臂老虎机算法 [48]，我们假设干扰向量 $\eta \in \mathbb{R}^{1000}$ 是 4-稀疏的。我们使用 LASSO 多臂老虎机程序生成了一条包含 $n = 950$个样本的路径，以自适应方式选择协变量，并将其应用于目标向量 $\theta^* = [2,\ 2]^T \in \mathbb{R}^2$。

图 1 的面板 (a) 显示，与 $\widehat{\theta}_1$ 并非标准正态分布；相反，该分布具有向下偏差，这反映在标准化误差的负均值 -0.07上。因此，我们看到，对于自适应收集的数据，渐近正态性可能不成立。图 1 的面板 (b) 显示，基于未加权 Z-估计量构造的置信区间无法达到预期的目标覆盖率；具体而言，这些区间覆盖真实参数的比例始终低于目标覆盖率。鉴于标准化误差偏离高斯性，这种覆盖不足是意料之中的。

需要明确的是，此类分布异常十分普遍：它们既不专属于 我们在此模拟的特定 Z-估计量 ， 也不专属于 LASSO 多臂老虎机算法。时间序列和预测文献中早已充分记录了类似的失效现象，最早可追溯至 Dickey 和 Fuller 的经典研究 [18], White [63], 以及 Lai 和 Wei [38]。较新的研究 [17, 67, 34]指出，在多臂老虎机问题中，使用汤普森采样、上置信界（UCB）和 ϵ-贪心选择等常见选择算法时，也存在类似现象。

1.2 相关工作

本节综述与我们的问题相关的自适应收集的数据推断和半参数推断方面的现有文献。

1.2.1 基于自适应收集的数据进行推断

在开创性研究中，Lai 和 Wei [38, 37] 研究了多种回归模型，其中协变量-响应对是以自适应方式收集的。除其他结果外，他们还给出了普通最小二乘法（OLS）估计量渐近正态的条件。然而，他们的结果要求协变量矩阵满足稳定性条件。在多种情形下，这一稳定性条件并不成立，例如某些类型的自回归模型 [18, 63, 38], 用于老虎机问题的 UCB 及相关在线程序 [39], , 以及针对自适应收集数据的多臂老虎机问题的离线程序（例如， [17, 67]）。

为应对这些挑战，Hadad等人 [25]针对多臂老虎机提出了增强型逆倾向加权（AIPW, [42]）估计量的自适应加权版本。他们提出了若干自适应权重的选取方法，以确保方差稳定，从而能够应用鞅中心极限定理。Zhan等人 [65] 和Bibaut等人 [8] 的后续研究将这一方法拓展到上下文老虎机，构造出渐近正态的估计量。Zhang等人 [68] 分析了用于上下文老虎机问题的加权 M估计量。Syrgkanis等人 [55]提出了一种加权 Z估计量，用于估计结构均值嵌套模型中的结构参数。所有这些关于老虎机问题的研究都假设数据收集算法已知，因此能够基于每个臂的选择概率构造加权估计量。另一方面，当老虎机算法未知时，Deshpande等人 [17]和Khamaru等人 [34]提出了在线去偏程序，使估计结果呈现渐近正态性。

1.2.2 半参数推断中的 Neyman 正交性

半参数统计研究的是在存在高维或非参数干扰参数的情况下，如何估计低维参数；相关研究文献丰富且不断发展（例如， [9, 50, 10, 52, 4, 51, 3, 59, 13]）。其中一个关键概念是得分函数的 Neyman 正交性 [45], ，它形式化了干扰项扰动对目标估计量的一阶影响。Neyman 正交性在半参数估计 [4, 44]、目标学习 [59]以及高维线性模型的推断 [66, 6, 7, 30]中发挥了重要作用。样本分割方法也在相关文献中得到普遍应用（例如， [9, 53, 19, 31]）；在这些方法中，数据集的不同部分分别用于估计非参数成分和参数成分。

Chernozhukov 等人 [13] 将 Neyman 正交性的概念与样本分割相结合，构造出渐近正态的 Z 估计量；他们将这一方法称为双重/去偏机器学习（DML）。样本分割降低了对干扰参数估计量满足 Donsker 类条件的要求，因此可以采用更复杂的非参数方法。此后，人们开发了其他基于 DML 方法或与之密切相关的方法，用于估计异质性处理效应 [46, 33, 20, 35, 54]、连续处理效应 [15, 54]、基于树的方法 [60, 5, 49]、含干扰参数的统计学习 [23]，以及动态处理效应 [40, 11, 14]。其中一些研究将适用范围扩展到独立同分布设定之外，允许样本来自稳定马尔可夫链，但并未涉及本文关注的一般自适应设定。

在独立同分布设定下，Belloni 等人 [7]研究了含干扰参数的广义线性模型中的推断，并建立了一个通用框架，用于在存在高维干扰项时对一维参数进行推断。Liu 等人 [41] 提出了一种适用于部分逻辑回归模型的估计量。这两项工作都利用了 Neyman 正交性，并且

其方法都涉及求解某个估计方程，与本文类似。本文关注的是一个类似的问题设定，但主要将其作为研究自适应数据收集影响的载体。

1.2.3 非渐近置信区间

不同于渐近保证，另一种方法是利用集中不等式构造在时间上均匀有效的非渐近置信域。例如，Abbasi 等人 [1] 证明了适用于老虎机问题的任意时刻自归一化集中不等式。这些界随后被进一步应用于多臂老虎机 [29, 32] 和一般序贯实验 [28]。一方面，这些方法具有非渐近保证，并且对模型错误设定相对稳健。然而，另一方面，在许多情形下，与基于渐近正态估计量构造的置信区间相比，这些程序得到的置信区间过于保守；例如，参见论文中的图2 [25]，其中给出了这类比较。

1.3 我们的贡献与论文结构

本文研究了在存在（可能是非参数的）干扰成分以及一般自适应数据收集模型的情况下，如何估计与广义线性回归模型相关的目标参数 θ^* 。由于自适应数据收集所诱导的序贯依赖，许多标准 Z-估计量可能表现出非正态的渐近行为，而我们的主要贡献正是纠正这一问题。为此，我们提出并分析了一族用于 θ^* 的估计量，并证明在温和条件下，这些估计量是渐近无偏且渐近正态的。这些程序基于对两阶段 Z-估计量进行自适应重新加权，因此我们将其称为 AdapTZ方法。在定理 1 中，我们讨论了针对部分线性模型设计的 AdapTZ-PL程序；而定理 3 则为更一般的程序（AdapTZ-GLM）提供了保证，该程序适用于广义线性模型。在某些正则条件下，这两个定理都能为参数向量 θ^*给出渐近有效的置信域。接下来，我们考虑估计形如 $u^\top\theta^*$的线性泛函的问题，其中， u 是 $\mathbb{R}^{d_T}$中的任意固定单位向量。在定理 2 和 4 中，我们证明，对于这一更简单的问题，与定理 1 和 3 相比，只需弱得多的条件即可获得渐近正态性。最后，在第 3 节中，我们通过推导一般理论对于一些具体半参数模型类别的结果，展示了该理论的实用性。

记号 对于任意满足 $n = n_1 + n_2$的数值 n、 n_1、 $n_2 \geq 1$，以及随机变量序列 $\{W_i\}_{i=1}^n$，我们使用简写

$$\widehat{\mathbb{E}}_{n_2} f_i(W_i) = \frac{1}{n_2} \sum_{i=n_1+1}^{n} f_i(W_i) \quad \text{and} \quad \widetilde{\mathbb{E}}_{n_2} f_i(W_i) = \frac{1}{n_2} \sum_{i=n_1+1}^{n} \mathbb{E} f_i(W_i \mid \mathcal{F}_{i-1})$$

我们用 $\|\cdot\|_2$ 表示向量的 2-范数；对于矩阵，我们分别用 $|||\cdot|||_{\text{op}}$ 和 $|||\cdot|||_F$ 表示其算子范数和 Frobenius 范数。对于向量 $a,\ b \in \mathbb{R}^d$，我们用 $\langle a,\ b\rangle = \sum_{j=1}^d a_j b_j$ 作为其欧几里得内积的简写。

2 主要结果

本节首先设定本文要研究的问题类别。我们关注在存在非参数干扰成分时，广义线性回归模型参数的渐近保证。我们的主要结果是对带有干扰参数的自适应广义模型 (4) 中两种估计算法的分析。我们推导出将这些方法应用于感兴趣参数时的若干渐近正态性保证。我们的第一种算法 (AdapTZ-PL) 专为部分线性模型（即特殊情形 $g(x) = x$）设计，而第二种算法 (AdapTZ-GLM) 则适用于更一般的非线性链接函数 g。

2.1 问题设定

假设标量响应变量 y 通过以下方程与协变量向量 $x \in \mathbb{R}^{d_T}$ 和辅助向量 $z \in \mathbb{R}^{d_N}$ 相联系：

$$y = g\big(\langle x, \theta^* \rangle + h^*(z)\big) + \varepsilon, \tag{4}$$

其中，ε 是均值为零的噪声变量。这里，$g : \mathbb{R} \to \mathbb{R}$ 是已知的链接函数，而 $\theta^* \in \Theta \subset \mathbb{R}^{d_T}$ 是未知的 目标参数，函数 $h^* : \mathbb{R}^{d_N} \to \mathbb{R}$ 也未知。我们假设目标参数空间 Θ 是 $\mathbb{R}^{d_T}$ 的一个有界开子集，而 h^* 属于某个函数类 $\mathcal{H}$，其中的函数在上确界范数下均一致有界。

模型 (4) 是一种特殊的 半参数模型，因为它同时包含参数成分和非参数成分。我们主要关注参数成分 θ^*：我们的目标是构造点估计及其对应的置信集。在此背景下，未知函数 h^* 起到 干扰参数 的作用。要获得 θ^* 的良好估计，需要对其加以控制，但它本身并非我们真正关心的对象。

2.1.1 自适应数据收集的允许形式

为了估计目标参数 θ^*，我们观察一组 n 样本，每个样本的形式为 $(x_i,\ y_i,\ z_i)$，其中 $i = 1, \ldots\ldots, n$。我们允许数据收集以如下方式呈现序贯依赖性。这些样本定义了一个嵌套的 σ-域序列，其中 $\mathcal{F}_0 = \emptyset$，并且

$$\mathcal{F}_{i-1} = \sigma\Big(\{x_j, y_j, z_j\}_{j=1}^{i-1}\Big) \qquad \text{for each } i = 2, \ldots, n. \tag{5}$$

设 $\mathcal{P}$为定义在 $\mathbb{R}^{d_N}$上的一个分布族。[1] 在阶段 $i = 1, \ldots\ldots, n$，我们假设：

- 给定 $\mathcal{F}_{i-1}$ 时，干扰向量 z_i 的分布属于 $\mathcal{P}$。
- 回归变量 x_i 的选择由一个已知的选择函数决定，该函数将成对的 $(z_i,\ \mathcal{F}_{i-1})$ 映射为概率 $p_i(z_i,\ \mathcal{F}_{i-1}) \in [0, 1]$。

稍微滥用一下记号，我们常用简写 p_i 表示函数值$p_i(z_i, \mathcal{F}_{i-1})$。在本文中，我们假设选择函数是已知的；例如，在上下文老虎机的设定中，这些函数可以对应于策略。

[1]例如，$\mathcal{P}$可以是定义在 $[0, 1]^{d_N}$上的所有分布的集合。

结构性假设 我们的分析涉及关于链接函数 g以及协变量所在空间 $\mathcal{X} \subset \mathbb{R}^{d_T}$ 的一些结构性假设。

(a) 在定理 1 和定理 2 中，我们为 $g(x) = x$ 提供了保证；在这种情况下，我们的一般设定 (4) 简化为部分线性回归的设定。

(b) 在定理 3 和定理 4 中，我们允许函数 g 为非线性，只要求满足某些光滑性和可识别性条件。(c) 在全文中，我们假设 d_T 维回归向量 x 取值于一个离散集合，该集合由 $\mathbb{R}^{d_T}$ 的一个标准正交基以及全零向量构成。不失一般性，通过适当旋转，我们可以假设该标准正交基为标准基 $\{e_1, \ldots, e_{d_T}\}$，其中 $e_j \in \mathbb{R}^{d_T}$是第 j 个坐标为 1（其余坐标为 0）的向量。这种特定设定自然出现于多臂老虎机和治疗分配问题中。

鉴于协变量的假设结构，选择函数很自然地可视为选择概率——也就是说，对于每个 $i = 1, \ldots\ldots, n$ 和 $j = 1, \ldots\ldots, d_T$

$$p_{ij} := \mathbb{E}\big[x_{ij} \mid \mathcal{F}_{i-1}, z_i\big] \tag{6}$$

是 $x_i = e_j$ 的条件概率。因此，$x_i = 0$ 的条件概率为 $p_{i0} := 1 - \sum_{j=1}^{d_T} p_{ij}$。

2.2 部分线性模型的保证

本节专门讨论一般设定中的一个特例：选择 $g(x) = x$ 会得到部分线性回归模型

$$y_i = \langle x_i, \theta^* \rangle + h^*(z_i) + \varepsilon_i. \tag{7}$$

我们假设目标参数 θ^* 位于有界开子集 $\Theta \subset \mathbb{R}^{d_T}$ 中，而干扰函数 h^* 属于函数类 $\mathcal{H}$，且其 ℓ_∞- 范数有界。

2.2.1 目标参数 θ^* 的估计

我们的方法是一种特殊类型的 Z-估计量，即通过求解一组基于 d_T 维得分函数的方程来计算解。下面引入定义该得分函数所需的一些记号。当以回归向量 x_i 所组成的向量对 $(\mathcal{F}_{i-1}, z_i)$ 为条件时，其条件协方差为

$$\mathbf{\Sigma}_i := \mathbb{E}\big[(x_i - p_i)(x_i - p_i)^\top \mid \mathcal{F}_{i-1}, z_i\big], \tag{8}$$

其中，$p \in \mathbb{R}^{d_T}$是前面定义的选择概率向量。请注意，由于选择机制已知，因此可以在每个时刻 i计算这个矩阵。随后，我们利用这个随机矩阵构造得分函数[2]

$$\phi_i(\theta, h) := \mathbf{\Sigma}_i^{-1/2}(x_i - p_i)\big\{y_i - \langle x_i, \theta\rangle - h(z_i)\big\}, \tag{9}$$

[2]严格来说，这个得分函数 ϕ_i 也取决于四元组 $(y_i, x_i, z_i, \mathcal{F}_{i-1})$，但为简化记号，我们省略了这种依赖关系。

ϕ_i 的一个重要性质是其条件均值为零，即：

$$\mathbb{E}\big[\phi_i(\theta^*, h^*) \mid \mathcal{F}_{i-1}\big] = 0. \tag{10a}$$

此外，它满足 Neyman 正交条件，

$$\mathbb{E}\big[\partial_h \phi_i(\theta^*, h^*)\{h - h^*\} \mid \mathcal{F}_{i-1}\big] = 0 \qquad \text{for any } h \in \mathcal{H}, \tag{10b}$$

其中 $\partial_h \phi_i$ 是盖托导数。关于该导数及相关正交条件的更多详情，请参见附录 C。

条件均值性质 (10a) 对于确保总体层面的一致性是必要的，而正交条件 (10b) 则保证——同样是在总体层面——扰动干扰参数所产生的一阶效应消失。基于这一认识，我们介绍 AdapTZ-PL 算法，这是 自适应两阶段Z部分线性模型估计的简称。

算法 1 AdapTZ-PL：部分线性模型

1: 给定来自部分线性模型 (7) 的 n 个样本 $\{(x_i,\ z_i,\ y_i)\}_{i=1}^n$。2: 定义指标集 $I_1 := \{1, 2, \ldots, n_1\}$ 和 $I_2 := \{n_1 + 1, \ldots, n\}$，并令 $n_2 := n - n_1$。3: 基于样本 $\{(y_i,\ x_i,\ z_i)\}_{i \in I_1}$，计算干扰函数 h^* 的一个估计量 $\widehat{h}$。

4: 基于样本 $\{(y_i,\ x_i,\ z_i)\}_{i \in I_2}$，构建估计方程

$$\frac{1}{n_2} \sum_{i \in I_2} \phi_i(\theta, \widehat{h}) = 0, \tag{11}$$

并求解得到解 $\widetilde{\theta}$。

注意： 根据得分函数 ϕ_i 的定义 (9)，估计方程 (11) 关于参数 θ 是线性的；此外，我们在证明定理 1 时的分析表明，随着 n 增大，这个线性方程组以趋近于 1 的概率存在唯一解 $\widetilde{\theta}$。

2.2.2 渐近正态性

本节的主要结果是，为使用 AdapTZ-PL 算法计算得到的向量 θ 提供渐近正态性保证。我们首先陈述假设，并讨论这些假设在定理中的作用。

(NOI($\nu,\ \sigma^2$)) 在给定 $(x_i,\ z_i,\ \mathcal{F}_{i-1})$ 的条件下，零均值噪声序列$\{\varepsilon_i\}_{i=1}^n$的每个元素都是参数为 ν 的次高斯变量，且其条件方差为 $\sigma^2 := \mathbb{E}[\varepsilon_i^2 \mid x_i, z_i, \mathcal{F}_{i-1}]$。

(SEL(t)) 每一轮的选择概率p_{ij} i 均满足下界

$$p_{ij} \geq \frac{c_0}{i^{2t}} \quad \text{for all } j = 0, 1, \ldots, d_T \text{ and } i = 1, 2, \ldots, \tag{12}$$

对于某个常数 $c_0 > 0$ 和指数 $t \in [0, \frac{1}{2})$。

(NUI) 设 $\mathcal{P}$ 是一个足够丰富的分布族，包含对于所有 $i \geq 1$，在给定 $\mathcal{F}_{i-1}$的条件下， z_i 的所有可能分布 $\widehat{}$。通过 AdapTZ-PL程序的步骤 3 得到的估计量 h 满足

$$\sup_{P \in \mathcal{P}} (\mathbb{E}_{v \sim P} |\widehat{h}(v) - h^*(v)|^2)^{1/2} = o_p(1). \tag{13}$$

我们来阐明这些假设的含义和重要性。噪声条件 **(NOI(ν, σ^2))**使我们能够控制噪声的尾部行为，虽然相对标准，但也可以放宽[3]。更值得关注的是选择条件 **(SEL(t))**，它允许最小选择概率以最快 n^{-2t} 的速度下降（对某个 $t \in [0, 1/2$而言）。这比一些以往研究中的条件略宽松，例如要求选择概率统一地有正下界 [68]，或收敛到某个非随机极限 [25, 65]。$\widehat{}$最后，干扰条件 **(NUI)**保证基于留出集得到的估计量 h 是真实干扰函数 h^*的弱相合估计量。实践中，可以采用各种方法估计 h^* （例如， k近邻估计量、随机森林、提升法、核方法和神经网络）。

在这一设定下，我们现在陈述第一个主要结果：

定理 1. 假设条件（**NOI**(ν, σ^2)）、（**SEL**(t)）和（**NUI**）均成立。那么，由 *AdapTZ-PL*（算法 *1*）得到的$\widetilde{}$估计量 θ 满足

$$(\sqrt{n_2}\widehat{\mathbb{E}}_{n_2} \mathbf{\Sigma}_i^{1/2})(\widetilde{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \mathbf{I}_{d_T}). \tag{14}$$

证明见附录 A.1。

关于这一结论，有几点需要说明。

独立同分布干扰项 一般而言，要找到一个合适的 $\mathcal{P}$ 来验证条件（13）并非易事。然而，当干扰项 z_i 是独立同分布的，且与 $\mathcal{F}_{i-1}$相互独立时，该条件可简化为（$\mathbb{E}|\widehat{h}(z_i) - h^*(z_i)|^2)^{1/2} = o_p(1)$，因此具体而明确。

线性干扰函数 假设干扰函数关于 z是线性的——也就是说，设$h^*(z) = \langle z, \eta^* \rangle$ 其中 $\eta^* \in \mathbb{R}^{d_N}$，并且对所有 $i \geq 1$均有 $\mathbb{E}\|z_i\|_2^2 \leq M_z < \infty$。在这些条件下，给定满足 $\|\widehat{\eta} - \eta^*\|_2 = o_p(1)$的估计量 $\widehat{\eta}$，可知假设（**NUI**）$\widehat{}$成立，其中 $h(z) = \langle z, \widehat{\eta} \rangle$，而 $\mathcal{P}$ 为所有二阶矩不超过 M_z的分布所构成的集合。

扩展到连续回归变量 如前所述，定理 1 适用于取值于有限基数集合 $\{0, e_1, \ldots, e_{d_T}\}$ 的回归变量 x_i。然而，对于连续取值的回归变量，同样可以证明类似的结果。具体而言，假设回归变量根据某个已知概率密度取值于 ℓ_2-球 $\{x \in \mathbb{R}^{d_T} \mid \|x\|_2 \leq 1\}$ 中。回顾 $\mathbf{\Sigma}_i$ 表示方程 (8) 中 x_i 的条件协方差矩阵，假设 **(SEL(t))** 被替换为以下条件：对于所有 i，$\mathbf{\Sigma}_i \succeq c_0 i^{-2t}$ （对于某个指数 $t \in [0, 1/2)$ 以及前置因子 $c_0 > 0$。在这些条件下，定理 1 的结论仍然成立。更深入的讨论请参见附录 A.1。

[3]更多详情请参见附录 F.1。

计算复杂度 注意，矩阵 $\mathbf{\Sigma}_i$ 是多项分布的协方差，而此类矩阵的乔列斯基分解可以在 $\mathcal{O}(d_T^2)$ 时间内完成。因此，建立估计方程 (11) 的时间复杂度为 $\mathcal{O}(nd_T^2)$。求解线性方程组至多需要 $\mathcal{O}(d_T^2)$ 时间。

目标参数的推断根据定理 1，我们可以为整个参数向量构造一个置信域（例如，通过 a χ^2-检验）。此外，如果随机矩阵序列 $\widehat{\mathbb{E}}_{n_2}[\mathbf{\Sigma}_i^{1/2}]$ 收敛于某个非随机且可逆的矩阵——记为 $\mathbf{\Gamma}^{1/2}$——那么方程 (14) 表明，$\sqrt{n_2}(\widehat{\theta}-\theta^*)$ 具有渐近正态性，其协方差为 $\sigma^2\mathbf{\Gamma}^{-1}$。

方差估计 σ^2 当 σ^2 未知时，需要对其进行估计。如果样本量满足某个常数 $c>0$对应的下界 $n_2 \geq cn$，则可通过插件估计得到相合估计量

$$\widehat{\sigma}^2 := \widehat{\mathbb{E}}_{n_2}(y_i - x_i^\top \widetilde{\theta} - \widehat{h}(z_i))^2. \tag{15}$$

更准确地说，只要除了满足定理 1 中的条件外，四阶矩 $\sup_{P\in\mathcal{P}} \mathbb{E}_{v\sim P}|\widehat{h}(v) - h^*(v)|^4$ 和 $\mathbb{E}[\varepsilon_i^4 \mid x_i, z_i, \mathcal{F}_{i-1}]$ 还由某个常数有界，就有 $\widehat{\sigma}^2 \to \sigma^2$ 依概率成立。有关这一结论的证明，请参见附录 A.1 的末尾。

干扰函数的自适应估计 此处所述的过程基于样本分割，使用前 n_1 个样本来估计干扰项 h^*。另一种 $\widehat{}$ 方法是依次更新估计量 h，以实现更高的样本效率。$\widetilde{}$具体而言，我们不求解方程 (11)，而是通过求解以下方程来得到 θ

$$\frac{1}{n}\sum_{i=1}^{n} \phi_i(\theta, \widehat{h}_i) = 0,$$

其中，$\widehat{h}_i$ 是使用样本 $\{(y_j, x_j, z_j)\}_{j=1}^{i-1}$ 得到的干扰项估计。可以证明，在定理 1 的条件下，若干扰项估计序列满足极限关系 $\sum_{i=1}^{n} \mathbb{E}_{\widehat{h}_i,\, z_i}(\widehat{h}_i(z_i) - h^*(z_i))^2/n \to 0$，则有

$$(\sqrt{n}\widehat{\mathbb{E}}_n \mathbf{\Sigma}_i^{1/2})(\widetilde{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \mathbf{I}_{d_T}).$$

直观而言，人们可能会预期这种方法具有良好的经验表现，因为 $\sim\theta$ 的方差按 $1/n$ 缩放；但另一方面，它的计算成本可能更高。更多详情请参阅附录 D。

选择概率未知时的推断 定理 1 也可以推广到选择概率的确切值 p_i 未知、但只有相合估计值 $\widehat{p}_i$ 可用的情形。详情请参见附录 E。

2.3 部分线性模型的固定方向推断

在许多应用中，人们只关心估计目标参数向量的线性泛函。具体而言，给定一个单位范数向量 $u \in \mathbb{R}^{d_T}$，考虑为标量目标 θ_u^* 提供置信区间的问题：$= \langle u, \theta^* \rangle$；常见示例包括第一个坐标 θ_1^*，或两个坐标之差 $\theta_1^* - \theta_2^*$。我们将证明

对于这类标量，推断保证可以在远弱于定理 1 的条件下获得。具体而言，我们只要求假设 **(SEL(t))** 对坐标 j 成立，其中 u_j 非零。

2.3.1 构造得分函数

假设我们使用数据集 $\{(x_i,\ z_i,\ y_i)\}_{i=1}^{n_1}$ 来计算一组初始的"粗略"估计值$\widehat{\ }\theta$ 和 h。回想一下，$\langle\cdot,\ \cdot\rangle$ 表示欧几里得内积，我们考虑一维得分函数

$$\phi_{i1}(\theta_u,\widehat{\theta},\widehat{h}) = \langle A_{i1},\, x_i - p_i\rangle\big\{y_i - \langle x_i,\, u\rangle\theta_u - x_i^\top(\mathbf{I}_{d_T} - uu^\top)\widehat{\theta} - \widehat{h}(z_i)\big\}, \tag{16a}$$

其中向量 A_{i1} 定义为

$$A_{i1} = \big(\mathbf{\Sigma}_i^{-1}u\big)\frac{1}{\sqrt{u^\top\mathbf{\Sigma}_i^{-1}u}},$$

而逆协方差矩阵 $\mathbf{\Sigma}_i^{-1}$具有显式表达式

$$\mathbf{\Sigma}_i^{-1}(z_i,\mathcal{F}_{i-1}) = \begin{pmatrix} \frac{1}{p_{i1}}+\gamma_i & \gamma_i & \gamma_i & \cdots & \gamma_i \\ \gamma_i & \frac{1}{p_{i2}}+\gamma_i & \gamma_i & \cdots & \gamma_i \\ \gamma_i & \gamma_i & \frac{1}{p_{i3}}+\gamma_i & \cdots & \gamma_i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_i & \gamma_i & \cdots & \gamma_i & \frac{1}{p_{id_T}}+\gamma_i \end{pmatrix} \quad \text{其中 } \gamma_i = 1/p_{i0}。 \tag{16b}$$

这种对 A_{i1} 的选择使我们能够稳定得分函数的方差：具体而言，我们有$\mathbb{E}|\langle A_{i1},\ x_i - p_i\rangle|^2 = 1$。下一步是通过求解线性方程组来求出 $\widetilde{\theta}_u$

$$\frac{1}{n_2}\sum_{i=n_1+1}^{n_2}\phi_{i1}(\widetilde{\theta}_u,\widehat{\theta},\widehat{h}) = 0 \tag{16c}$$

2.3.2 渐近正态性的保证

现在我们可以为估计量 θ_u^*建立保证。我们在以下较弱的先前选择条件变体（**SEL(t)**)下进行：

(SEL*(t,u,S_u)) 对于某个 $t\in[0,\frac{1}{2})$，选择概率有如下下界：

$$p_{ij} \succeq \frac{c_0}{i^{2t}} \quad \text{for all } j\in S_u\cup\{0\}, \text{ and for all } i=1,2\ldots, \tag{17}$$

其中 $S_u := \{j \mid u_j \neq 0\}$ 是 u的支集。

与条件（**SEL(t)**）相比，假设（**SEL*(t,u,S_u)**)较弱，因为下界条件仅 施加于向量 u的支撑集以及参考点（全零向量）。例如，当我们的目标是估计单个坐标或两个坐标之差时，这一区别就很重要。

定理 2. 假设以下假设均成立：**(NOI**(ν,σ^2)**)**、**(SEL**$^*(t,u,S_u)$**)**和**(NUI)**。$\widetilde{\ }$ $\widehat{\ }$那么，使用对 θ 的任意相合估计量 θ，根据(16c) 计算得到的 Z-估计量 θ_u 满足

$$\Big(\widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top\mathbf{\Sigma}_i^{-1}u}}\Big)(\widetilde{\theta}_u - \theta_u^*) \xrightarrow{d} \mathcal{N}(0,\sigma^2).$$

证明见附录 A.2。

下面对定理 2 作几点说明。首先，尽管需要假设先验 $\widehat{}$存在 θ的相合估计量，但其保证在关于选择概率的较弱假设下仍然成立。不过，由于在部分线性模型中，我们通常会同时估计 θ^* 和 h^*，如果能够找到 h^* 的相合估计量（参见条件 (13)），那么我们也能得到 θ^*的相合估计量。

其次，假设干扰函数是线性的——即，对于某个$\eta^* \in \mathbb{R}^{d_N}$，$h^*(z) = \langle z,\ \eta^* \rangle$。与定理 1 类似，设 $\widehat{\eta}$是 η^* 的估计量，且满足 $\|\widehat{\eta} - \eta^*\|_2 = o_p(1)$，并假设 $\sup_i \mathbb{E}\|z_i\|_2^2 \le M_z < \infty$，则假设（**NUI**）在 $\widehat{h}(z) = \langle z,\ \widehat{\eta} \rangle$下成立，并且 $\mathcal{P}$ 是二阶矩小于 M_z的分布集合。

注意，定理 2 允许我们构造一个关于 θ_u^*的渐近有效、置信水平为α 的置信区间。具体而言，我们有

$$\lim_{n\to\infty} \mathbb{P}\left[\widetilde{\theta}_u - \tfrac{q_{1-\alpha/2}\sigma}{\sqrt{n_2}}\Big(\widehat{\mathbb{E}}_{n_2}\tfrac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big)^{-1} \le \theta_u^* \le \widetilde{\theta}_u + \tfrac{q_{1-\alpha/2}\sigma}{\sqrt{n_2}}\Big(\widehat{\mathbb{E}}_{n_2}\tfrac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big)^{-1}\right] = 1-\alpha,$$

其中，$q_{1-\alpha/2}$ 是标准正态分布的 $1-\alpha/2$ 分位数。特别地，如果我们关注第一个坐标 θ_1^*，那么令 $u = e_1$ 并应用定理 2，可得

$$\Big(\widehat{\mathbb{E}}_{n_2}\sqrt{\frac{p_{i0}p_{i1}}{p_{i0}+p_{i1}}}\Big)(\widetilde{\theta}_1 - \theta_1^*) \xrightarrow{d} \mathcal{N}(0,\sigma^2).$$

$\widehat{}$第三，尽管我们陈述结果时假设 θ 对完整向量 θ 是一致的，但事实上，我们只要求它在与 u 正交的任意方向上一致，也就是说，只需满足稍弱的一致性条件（$\mathbf{I}_{d_T} - uu^\top$）（$\widehat{\theta} - \theta^*$）$\xrightarrow{p} 0$。

最后，定理 2 也可以推广到连续回归变量的情形，如下所示。假设回归变量按照某个已知的概率密度取值于 ℓ_2-球 $\{x \in \mathbb{R}^{d_T} \mid \|x\|_2 \le 1\}$。那么，只要将假设（**SEL**$^*(t,u,S_u)$）替换为这样的条件：存在某个指数 $t \in [0,1/2)$ 和前置因子$c_0 > 0$，使得 $\|v\|_2/\sqrt{v^\top \mathbf{\Sigma}_i^{-1} v} \succeq c_0 i^{-t}$ 对于~~$v \in$~~ $\{u, \mathbf{\Sigma}_i^{-1/2}u\}$ 对于 $i = 1,2,\ldots$，便可得到相同的保证。更详细的讨论请参见附录 A.2。

2.4 广义线性模型

现在我们回到一般情形，其中模型的形式为

$$y_i = g\big(\langle x_i,\, \theta^* \rangle + h^*(z_i)\big) + \varepsilon_i, \tag{18}$$

对于一般的逆链接函数 g。我们假设参数 $(\theta^*,\ h^*) \in \Theta \times \mathcal{H}$，其中参数空间 Θ 是 $\mathbb{R}^{d_T}$ 中的一个有界开集，而 $\mathcal{H}$ 是一组具有有界 ℓ_∞-范数的函数。

2.4.1 目标参数 θ^* 的估计

我们首先构造一个不同的得分函数。

引入一个辅助干扰向量 θ，并定义得分函数

$$\phi_i(\theta, \bar{\theta}, h) \equiv \mathbf{\Omega}_i(x_i - m_i)\left\{y_i - g\big(\langle x_i,\, \theta \rangle + h(z_i)\big)\right\} \tag{19}$$

其中

$$m_i \equiv \mathbb{E}((x_i g'\big(\langle x_i, \bar{\theta}\rangle + h(z_i)\big)|z_i, \mathcal{F}_{i-1})[\mathbb{E}(g'\big(\langle x_i, \bar{\theta}\rangle + h(z_i)\big)|z_i, \mathcal{F}_{i-1})]^{-1}, \tag{20a}$$

$$\begin{aligned}\mathbf{\Omega}_i &\equiv [\mathbb{E}(\varepsilon_i^2(x_i - m_i)(x_i - m_i)^\top|z_i, \mathcal{F}_{i-1})]^{-1/2} \\ &= [\mathbb{E}(\nu^2\big(g\big(\langle x_i, \bar{\theta}\rangle + h(z_i)\big)\big)(x_i - m_i)(x_i - m_i)^\top|z_i, \mathcal{F}_{i-1})]^{-1/2}\end{aligned} \tag{20b}$$

且 $\nu^2(x) \equiv \mathbb{E}(\varepsilon_i^2|g\big(\langle x_i, \theta^*\rangle + h^*(z_i)\big) = x)$ 是噪声 ε_i 的条件方差。当 $\theta = \theta^*$（或 θ）且 $h = h^*$（或 h）时，我们分别用 m_i 和 $\mathbf{\Omega}_i$ 表示相应的 m_i^*（或 $\widehat{m}_i$）和 $\mathbf{\Omega}_i^*$（或 $\widehat{\mathbf{\Omega}}_i$）。直观地说，向量 m_i 可以看作回归变量 x_i 的加权条件期望，而矩阵 $\mathbf{\Omega}_i$ 可以看作 x_i 的加权条件协方差矩阵的逆平方根。当 $g(x) = x$ 时，m，$\mathbf{\Omega}$ 不依赖于 θ，且方程（19）中的得分函数退化为部分线性模型方程（9）中的早期得分函数。

与部分线性模型的情形类似，该得分函数满足某种 Neyman 正交条件。更具体地说，我们有

$$\mathbb{E}(\phi(\theta^*, \theta^*, h^*)|\mathcal{F}_{i-1}) = 0, \tag{21a}$$

$$\mathbb{E}(\partial_\theta \phi_i(\theta^*, \theta^*, h^*)|\mathcal{F}_{i-1}) = 0, \text{and} \tag{21b}$$

$$\mathbb{E}(\partial_h \phi_i(\theta^*, \theta^*, h^*)[h - h^*]|\mathcal{F}_{i-1}) = 0 \quad \text{for all} \quad h \in \mathcal{H}. \tag{21c}$$

这些方程的证明见附录 C。

在此设置下，我们使用以下 AdapTZ-GLM 程序估计目标参数 θ^*，即自适应两阶段 Z 广义线性模型估计。

算法 2 AdapTZ-GLM：广义线性模型

1: 给定来自部分线性模型（18）的 n 个样本 $\{(x_i, z_i, y_i)\}_{i=1}^n$。

2: 定义 th 指标集 $I_1 := \{1, 2, \ldots, n_1\}$ 和 $I_2 := \{n_1+1, \ldots, n\}$，并定义 $n_2 = n - n_1$.

3: 使用样本 $\{(y_i, x_i, z_i)_{i\in I_1}\}$ 得到干扰函数 h^* 的估计量 $\widehat{h}$，以及目标参数 θ^* 的估计量 $\widetilde{\theta}$。4: 通过求解方程来求得 θ

$$\frac{1}{n_2}\sum_{i\in I_2} \phi_i(\widetilde{\theta}, \widehat{\theta}, \widehat{h}) = 0 \tag{22}$$

.

2.4.2 渐近正态性

现在我们转向一项关于使用算法 2 所述的 AdapTZ-GLM 程序计算的估计量 θ 的渐近正态性的结果。首先介绍其基本假设。

(SEL′(t, δ)) 对于某些 $t \in [0, \frac{1}{4}]$ 和 $\delta > 0$，选择概率 p_{ik} 有下界

$$p_{ik} \geq c_i := \frac{c_0}{i^{2(t-\delta)}} \quad \text{for all } k = 1, \ldots, d_T, \text{ and } i = 1, 2, \ldots. \tag{23}$$

此外 因此，选择零向量的概率满足 $p_{i0} \geq \tilde{c}_0$，其中某个 $\tilde{c}_0 > 0$.

(**NUI′**) 假设 $\mathcal{P}$ 中的所有分布都以集合 $\mathrm{dom}(\mathcal{P}) \subseteq \mathbb{R}^{d_N}$ 为支撑集。步骤 3 中由算法 2 得到的估计量 $\widehat{\theta}, \widehat{h}$ 满足 $\|\widehat{\theta} - \theta^*\|_2 = o_p(n^{-1/4})$，且 $\sup_{v \in \mathrm{dom}(\mathcal{P})} |\widehat{h}(v) - h^*(v)| = o_p(n^{-1/4})$.

(IDE) 在我们选择的得分函数下，模型是可辨识的，具体而言，

$$\|\widetilde{\mathbb{E}}_{n_2}(\phi_i(\theta, \theta^*, h^*) - \phi_i(\theta^*, \theta^*, h^*))\|_2 \geq c_\phi \|\widetilde{\mathbb{E}}_{n_2} \partial_\theta \phi_i(\theta^*, \theta^*, h^*)(\theta - \theta^*)\|_2 \wedge c_\phi n^{-1/4}$$

对任意 $\theta \in \mathbb{R}^{d_T}$，几乎必然成立，且存在某个常数 $c_\phi > 0$。

(EIG) 最小奇异值梯度 $\mathbb{E}_{n_2} \partial_\theta \phi_i(\theta^*, \theta^*, h^*)$ 不会太小，即

$$\lim_{n \to \infty} \mathbb{P}(\sigma_{\min}(\widetilde{\mathbb{E}}_{n_2} \mathbf{\Omega}_i^* (x_i - m_i^*) g'\big(\langle x_i, \theta^* \rangle + h^*(z_i)\big)(x_i - m_i^*)^\top) \geq m_\phi n^{\delta - t}) \to 1$$

存在某个常数 $m_\phi > 0$。

2.4.3 其他广义线性模型假设

除上述四个假设外，我们还对广义线性模型作出以下假设。

- 存在常数 M_θ 和 $D_x > 0$，使得 $\sup_{\theta \in \Theta} \|\theta\|_2 \leq M_\theta$, $\|x_i\|_2 \leq D_x$，且 h^*满足 $\|h^*(z_i)\|_\infty \leq M_h$，其中某个 $M_h > 0$。
- 条件方差 $\nu^2(x)$ 三次可微，$\nu^2(x)$，$(\nu^2)'(x)$ 分别是 L_ε，L_ε-利普希茨连续（对于 $|x| \leq M_h + D_x M_\theta$），且存在某些 M_ε，$m_\varepsilon > 0$，使得$M_\varepsilon \geq \nu^2(x) \geq m_\varepsilon$，对于 $|x| \leq M_h + D_x M_\theta$。此外，我们假设零均值噪声 ε_i 在给定 x_i，z_i，$\mathcal{F}_{i-1}$的条件下是参数为 ν的次高斯变量。

- 逆链接函数 g 三次可微且单调，函数g，g'，g'' 分别是 L_g，$L_{g'}$，$L_{g''}$-利普希茨连续。此外，对于某个 $l_g > 0$，$\inf_{|x| \leq M_h + D_x M_\theta} |g'(x)| \geq l_g$。

假设（**SEL′**(t, δ)）略强于假设（**SEL**(t)），具体而言，我们需要将 t 替换为 $t - \delta$，其中 δ 是某个较小的常数，并对 $t \in [0, 1/4]$加以限制。假设（**NUI′**）针对初始估计量的表现提出要求。我们假设 $\theta - \theta^*$，$\sup |h(v) - h^*(v)| = o_p(n^{-1/4})$，是为了确保逆链接函数$g$ 的泰勒展开中的二阶项趋于零。相比之下，在部分线性模型中，假设（**NUI**）仅要求干扰项估计量是一致的。由于在线性情形下，g 的一阶泰勒近似是精确的，因此无需对收敛速度作出假设来消除泰勒展开中的近似误差项。

条件 **(IDE)** 和 **(EIG)** 确保当 θ 远离 θ^* 时，得分函数的期望充分远离零。在 $g(x) = x$ 的简单情形下，假设 **(IDE)** 和 **(EIG)** 可由其余假设推出。此外，可以证明，在逻辑回归中，假设 **(EIG)** 成立；当 $d_T = 1$ 时，假设 **(IDE)** 成立（详细推导见附录 C.3）。然而，由于收集数据的自适应性，通常很难验证这两个假设。为解决这一问题，在实践中，我们建议验证时将所有 $\mathbb{E}_{n_2}$ 替换为 $\widetilde{\mathbb{E}}_{n_2}$。由于经验均值集中于条件期望附近，当假设本身成立时，经验版本的假设 **(IDE)**

和 **(EIG)** 以高概率成立。因此，我们可以用经验版本作为原始假设的替代。

最后，在第 2.4.3 节中，我们列出广义线性模型的一些标准假设。第一个条件假设回归变量、参数空间和真实非线性函数均有界。第二和第三个条件分别对条件方差函数 $\nu^2(\cdot)$和链接函数 g 提出了一些光滑性条件。

定理 3. 假设假设 ***(SEL′(t,δ))—(EIG)*** 以及第 *2.4.3* 节中的广义线性模型标准假设均成立。则由 *AdapTZ-GLM* 得到的估计量 θ （参见算法 *2* ）满足

$$(\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\,\widehat{\theta}\rangle+\widehat{h}(z_i)\big)(x_i-\widehat{m}_i)^\top)\sqrt{n_2}(\widetilde{\theta}-\theta^*)\xrightarrow{d}\mathcal{N}(0,\mathbf{I}_{d_T}). \tag{24}$$

证明见附录A.3。

一个简单的情形是干扰函数为线性函数，即 $h^*(z)=\langle z,\ \eta^*\rangle$ 其中 $\eta^*\in\eta$，且 $\mathcal{H}$ 是 $\mathbb{R}^{d_N}$ 中的有界集合。在这种情形下，若存在 $M_\eta,\ D_x>0$ 使得 $\sup_{\eta\in\mathcal{H}}\|\eta\|_2\le M_\eta$ 和 $\|(x_i^\top,\ z_i^\top)^\top\|_2\le D_x$，则第 2.4.3 节中的假设成立。此外，若还存在估计量 $\widehat{\eta}\in\mathcal{H}$ 使得 $\|\widehat{\eta}-\eta^*\|_2=o_p(n^{-1/4})$，则假设 **(NUI′)** 成立。

与第 2.2 节讨论的部分线性模型类似，我们可以自适应地估计干扰函数 h^*，以提高样本效率。此外，定理 3 允许我们通过 χ^2- 检验为参数向量 θ^* 构造一个置信域。此外，如果方程 (24) 左侧的 $\widetilde{}$ 加权矩阵收敛，那么 $\sqrt{n_2}(\theta-\theta^*)$ 渐近正态，我们就可以为参数向量 θ^* 的任意子集构造置信域（区间）。若不存在这种收敛，则很难为 θ^* 的固定方向构造置信域，即一般而言，$\langle u,\ \theta^*\rangle$，除非依赖额外假设（例如，方程 24 的强高斯近似版本）；详细论证请参见论文 [34] 的第3.2.2节。

尽管如此，我们仍可使用本节讨论的估计量的一种变体，为 $\langle u,\ \theta^*\rangle$ 给出渐近正态的估计量。有趣的是，如果我们只关注固定方向 u 上 $\langle u,\ \theta^*\rangle$ 的置信区间，就可以放宽定理 3 的条件。下一节将详细讨论这些条件。

2.5 广义线性模型的固定方向推断

对于任意满足 $\|u\|_2=1$ 的方向 $u\in\mathbb{R}^{d_T}$，我们可以构造一个一维得分函数，并得到关于 θ_u^* 的渐近正态估计量：$=\langle u,\ \theta^*\rangle$。我们的构造遵循与方程 (16a) 相同的思路。具体而言，我们考虑一个一维得分函数

$$\phi_{i1}(\theta_u,\bar{\theta},h)\equiv A_{i1}(x_i-m_i)(y_i-g\big(\langle x_i,\,u\rangle\theta_u+x_i^\top(\mathbf{I}_{d_T}-uu^\top)\bar{\theta}+h(z_i))), \tag{25}$$

其中，

$$A_{i1}:=u^\top\mathbf{\Omega}_i^2/\sqrt{u^\top\mathbf{\Omega}_i^2u}.$$

$\widehat{\ }$ $\widehat{\ }$类似地，将 $(\theta,h)=(\theta,h)$（或 (θ^*,h^*)）代入即可定义 A_{i1}（或 A_{i1}^*）。我们指出，$\mathbb{E}|A_{i1}(x_i-m_i)|^2=1$，这将在后续章节中发挥作用。根据这些定义，

另一方面，我们使用算法 2 估计参数 θ_u，但将步骤 4 替换为求解满足下式的 θ_u

$$\frac{1}{n_2}\sum_{i\in I_2}\phi_{i1}(\widetilde{\theta}_u,\widehat{\theta},\widehat{h})=0. \tag{26}$$

同样，在以下假设变体（**SEL′**(t,δ)）、（**IDE**）和（**EIG**）下，我们对 θ_u 有渐近保证。

(**SEL′**$^*(t,\delta,S_u)$) 每一轮的选择概率 p_{ik} 满足下界

$$p_{ik}\geq c_i:=\frac{c_0}{i^{2(t-\delta)}}\quad \text{for all } i\geq 1, k\in S_u \tag{27}$$

对于某个常数 $c_0>0$、$t\in[0,\frac{1}{4}]$ 和 $\delta>0$。此外，选择 0 的概率满足 $p_{i0}\geq\tilde{c}_0$，其中某个 $\tilde{c}_0>0$。

(**IDE***) 在我们选择的得分函数下，模型是可识别的，具体而言，

$$\|\widetilde{\mathbb{E}}_{n_2}(\phi_{i1}(\theta_u,\theta^*,h^*)-\phi_{i1}(\theta_u^*,\theta^*,h^*))\|\geq c_\phi\|\widetilde{\mathbb{E}}_{n_2}\partial_{\theta_u}\phi_{i1}(\theta_u^*,\theta^*,h^*)(\theta_u-\theta_u^*)\|_2\wedge c_\phi n^{-1/4}$$

对于任意 $\theta\in\mathbb{R}^{d_T}$ 和某个 $c_\phi>0$，几乎必然成立。

~(**EIG***) 梯度 $\mathbb{E}_{n_2}\partial_{\theta_u}\phi_{i1}(\theta_u^*,\theta^*,h^*)$ 不会太小，即 $\lim_{n\to\infty}\mathbb{P}(|\mathbb{E}_{n_2}A_{i1}^*(x_i-m_i^*)g'(\langle x_i,\theta^*\rangle+h^*(z_i))(x_i-m_i^*)^\top u|\geq m_\phi n^{\delta-t})\to 1$，其中某个常数为 $m_\phi>0$。

关于这些假设，有几点需要说明。假设 (**SEL′**$^*(t,\delta,S_u)$) 弱于假设 (**SEL′**(t,δ))，因为我们没有对向量 u 支撑集之外坐标的选择概率作出假设。假设 (**IDE***) 和 (**EIG***) 是对假设 (**IDE**) 和 (**EIG**) 的调整，即将得分函数 ϕ_i 替换为 ϕ_{i1}。类似地，当 $g(x)=x$ 时，假设 (**IDE***) 和 (**EIG***) 均可由关于广义线性模型的其余假设推出。此外，当 $g(x)$ 为逻辑函数时，可以验证假设 (**IDE***) 和 (**EIG***)（详情见附录 C.3）。

定理 4。 除了第 *2.4.3* 节中的标准广义线性模型条件外，假设~假设 *(***NUI′***)*、*(***SEL′**$^*(t,\ \delta,\ S_u)$*)*、*(***IDE****)* 和 *(***EIG****)* 均成立。那么，方程 (26) 中的估计量 θ_u满足

$$(\widehat{\mathbb{E}}_{n_2}\widehat{A}_{i1}(x_i-\widehat{m}_i)g'(\langle x_i,\widehat{\theta}\rangle+\widehat{h}(z_i))(x_i-\widehat{m}_i)^\top u)\sqrt{n_2}(\widetilde{\theta}_u-\theta_u^*)\xrightarrow{d}\mathcal{N}(0,1). \tag{28}$$

证明见附录 A.4。

定理 4 使我们能够为 θ_u^*构造渐近有效的水平为 α 的置信区间。将 $(\widehat{\mathbb{E}}_{n_2}\widehat{A}_{i1}(x_i-\widehat{m}_i)g'(\langle x_i,\ \widehat{\theta}\rangle+\widehat{h}(z_i))(x_i-\widehat{m}_i)^\top u)$ 记为 v_{cov}。具体而言，我们有

$$\lim_{n\to\infty}\mathbb{P}\Big[\widetilde{\theta}_u-\frac{q_{1-\alpha/2}\sigma}{\sqrt{n_2}v_{cov}}\leq\theta_u^*\leq\widetilde{\theta}_u+\frac{q_{1-\alpha/2}\sigma}{\sqrt{n_2}v_{cov}}\Big]=1-\alpha,$$

其中，$q_{1-\alpha/2}$ 是标准正态分布的 $1-\alpha/2$ 分位数。

3 特定模型的一些推论

本节给出若干示例，在这些示例中，我们可以为干扰项（和目标）参数构造合适的初步估计量。利用 AdapTZ-PL 或 AdapTZ-GLM 算法得到此类估计值后，我们可以设计出明确且计算高效的程序，并享有定理 1 至 4 中所述的保证。在本节中，我们假设对于某个 $K \geq 2$, $n_1 = n/K$ 。

3.1 自适应数据收集下的分区线性模型

我们从最简单的情形开始，即形式如下的分区线性模型

$$y_i = \langle x_i,\, \theta^* \rangle + \langle z_i,\, \eta^* \rangle + \varepsilon_i,$$

其中 $\theta^* \in \mathbb{R}^{d_T}$ 和 $\eta^* \in \mathbb{R}^{d_N}$。假设协变量向量以自适应方式收集，取值属于集合 $\{e_1, \ldots, e_{d_T}, 0\}$，第 i 轮的选择概率向量$p_i \in \mathbb{R}^{d_T+1}$ 允许是数对 $(z_i,\ \mathcal{F}_{i-1})$ 的函数。在这种设定下，Lai 和 Wei [38] 证明，即使设计矩阵不满足稳定性条件，普通最小二乘估计量（$\widehat{\theta}_{\mathrm{OLS}},\ \widehat{\eta}_{\mathrm{OLS}}$）仍然是一致的。因此，我们可以使用 AdapTZ-PL 构造 θ^* 的渐近正态估计量，并以普通最小二乘法估计量作为 η^* 的初始估计量。具体而言，我们假设

$$\inf_{P \in \mathcal{P}} \sigma_{\min}(\mathbb{E}_{z_i \sim \mathcal{P}} z_i z_i^\top) > 0 \quad \text{and} \quad \|z_i\|_2 \leq B \quad \text{for some constant } B > 0. \tag{29}$$

最后，回顾记号 $\mathbf{\Sigma}_i$ 表示第 i 步回归变量的条件协方差，我们由定理1推导出以下推论。

推论 1。 假设假设 **(NOI**$(\nu,\ \sigma^2)$**)–(NUI)** 对某个 $t \in [0, 1/2)$ 成立，并且条件 (29) 也成立。那么，使用 $(\widehat{\theta}_{\mathrm{OLS}}、\widehat{\eta}_{\mathrm{OLS}})$ 作为先导估计量，由 *AdapTZ-PL* 获得的估计量 θ 满足

$$(\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})(\widetilde{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \mathbf{I}_{d_T}). \tag{30}$$

证明见附录B.1。

3.2 稀疏高维线性模型

接下来，我们考虑高维线性回归问题

$$y_i = \langle x_i,\, \theta^* \rangle + \langle z_i,\, \eta^* \rangle + \varepsilon_i,$$

其中 $\theta^* \in \mathbb{R}^{d_T}$ 且 $\eta^* \in \mathbb{R}^{d_N}$。我们允许采用部分高维形式的渐近分析，其中目标维度 d_T 保持固定，而干扰维度 d_N 则允许随着 $n \to \infty$趋于无穷。我们假设噪声变量 $\varepsilon_i' s$ 是参数为 ν 的次高斯变量。我们还假设干扰向量 η^* 是稀疏的，且 $|\{\eta_i^* \neq 0\}| = s$。请注意，只要能够获得精确的初始估计量，我们的定理就允许干扰分量发生变化。我们使用 Lasso估计量作为先导估计量，也就是说

$$(\widehat{\theta}_{\mathrm{Lasso}}, \widehat{\eta}_{\mathrm{Lasso}}) := \arg\min_{\theta,\eta} \Big\{ \tfrac{1}{2n_1} \sum_{i=1}^{n_1} (y_i - \langle x_i,\, \theta \rangle - \langle z_i,\, \eta \rangle)^2 + \lambda_{n_1} (\|\theta\|_1 + \|\eta\|_1) \Big\}, \quad \text{where} \tag{31}$$

$$\lambda_{n_1} := 2\nu(B'+1)\sqrt{\tfrac{2[\log(\frac{2}{\delta_{n_1}}) + \log(d_T + d_N)]}{n_1}}, \quad \delta_{n_1} := \min\{(s+d_T) n_1^{2t-1/2}, \tfrac{1}{d_T+d_N}\}$$

对于某个常数 $B' > 0.$ 。在我们的结果中，我们假设稀疏度有界，满足

$$(s + d_T)\sqrt{\log(d_T + d_N)} = o_p(n_1^{1/2-2t}) \qquad \text{for some exponent } t \in [0, 1/4). \tag{32}$$

此外，假设干扰分量满足

$$\inf_{P\in\mathcal{P}} \sigma_{\min}(\mathbb{E}_{z_i\sim\mathcal{P}} z_i z_i^\top) > 0 \quad \text{and} \quad \|z_i\|_\infty \le B' \quad \text{for some constant } B' > 0. \tag{33}$$

在此设定下，我们可以应用定理 1，从而得到以下推论：

推论 2. 假设假设（**NOI**$(\nu,\ \sigma^2)$）–（**NUI**）以及稀疏性条件 (32) 成立~对于某个 $t \in [0, 1/4)$，并且假设 (33) 成立。然后，使用 *AdapTZ-PL* 和 $(\theta_{\text{Lasso}}$、 $\widehat{\eta}_{\text{Lasso}})$ 作为先导估计量而得到的估计量 $\widehat{\theta}$ 满足

$$(\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\boldsymbol{\Sigma}_i^{1/2})(\widetilde{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2\mathbf{I}_{d_T}). \tag{34}$$

证明见附录 B.2。

一般而言，为目标参数和干扰参数构造一个渐近有效的置信域并非易事；然而，推论 2 说明了，为了对固定数量的目标参数进行有效推断，我们能够容忍多少个干扰参数。

3.3 稀疏广义线性模型

现在我们来考察稀疏线性模型的一种扩展。假设我们观测到三元组 $(x_i,\ z_i,\ y_i)$，它们满足以下模型

$$y_i = g\big(\langle x_i,\, \theta^*\rangle + \langle z_i,\, \eta^*\rangle\big) + \varepsilon_i.$$

我们假设链接函数 g 以通常的指数族形式得到，因此存在一个函数 G，使得 $G'(t) = g(t)$。因此，与该模型相关的负对数似然具有如下形式

$$f(\theta, \eta; x_i, z_i, y_i) = G\big(\langle x_i,\, \theta\rangle + \langle z_i,\, \eta\rangle\big) - y_i(\langle x_i,\, \theta\rangle + \langle z_i,\, \eta\rangle).$$

作为 AdapTZ-GLM 流程（参见算法 2）的初始估计量，我们计算 ℓ_1-正则化估计量

$$(\widehat{\theta}_{\text{GLMlasso}}, \widehat{\eta}_{\text{GLMlasso}}) := \arg\min_{\theta,\eta} \Big\{\frac{1}{n_1}\sum_{i=1}^{n_1} f(\theta, \eta; x_i, z_i, y_i) + \lambda_{n_1}(\|\theta\|_1 + \|\eta\|_1)\Big\}. \tag{35}$$

取值为

$$\lambda_{n_1} := 2\nu D_x\sqrt{\frac{2[\log(2/\delta_{n_1}) + \log(d_T + d_N)]}{n_1}},\ \text{and}\ \delta_{n_1} := \min\{(s + d_T)n_1^{2t-1/4}, \frac{1}{d_T + d_N}\},$$

其中 D_x 是对所有 $i.$ 的 $\| \begin{pmatrix} x_i^\top & z_i^\top \end{pmatrix} \|_2$ 的上界

在我们的分析中，假设目标维度 d_T 固定，而干扰维度d_N 允许随着 $n \to \infty$ 趋于无穷。同样，我们假设噪声变量 $\{\varepsilon_i\}_{i=1}^n$相互独立，且各自为参数至多为 ν 的次高斯变量，并且真实干扰向量η^* 是稀疏的，满足 $|\{\eta_i^* \neq 0\}| = s$。此外，我们假设稀疏度 s 满足以下条件

$$(s + d_T)\sqrt{\log(d_T + d_N)} = o_p(n_1^{1/4-2t}). \tag{36}$$

W在此设定下，我们可以应用定理 3，从而得到以下保证 e:

推论 3. 假设（**SEL**$'(t,\delta)$）—（**EIG**）对某个 $t \in [0, 1/4)$ 成立，且假设 $\widetilde{\ }$*sumptions* (29) 和 (36) 均成立。那么，使用先导估计量(35)并通过 *AdapTZ-GLM*计算得到的估计量 θ 满足

$$(\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)g'\big(\langle x_i, \widehat{\theta}\rangle + \widehat{h}(z_i)\big)(x_i - \widehat{m}_i)^\top)\sqrt{n_2}(\widetilde{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \mathbf{I}_{d_T}). \tag{37}$$

推论 3 的证明见附录B.3。

与高维线性模型的推论 2 相比，这里我们需要对稀疏度作出更强的假设（即 $(s + d_T)\sqrt{\log(d_T + d_N)} = o_p(n_1^{1/4-2t})$），并限制 $t \in [0, 1/8)$。这是因为需要一个 $o_p(n^{-1/4})$-一致的初始估计量。我们指出，我们关于稀疏度的假设可能并非最优，在更强的假设下（例如数据以 [7] 独立同分布的方式收集时）可以改进。

3.4 带有非参数干扰项的部分线性模型

最后，我们考虑干扰分量为非参数函数的情形，即由下式给出的部分线性模型

$$y_i = \langle x_i, \theta^*\rangle + h^*(z_i) + \varepsilon_i,$$

其中 $\theta^* \in \mathbb{R}^{d_T}$、$z_i \in \mathbb{R}^{d_N}$ 和 $h^* : \mathbb{R}^{d_N} \mapsto \mathbb{R}$ 是某个非参数函数。与第3.1节类似，假设协变量向量 x_i 以概率 p_i 取集合 $\{e_1, \ldots, e_{d_T}, 0\}$ 中的值，其中 p_i 由选择概率向量给出。此外，假设

$$z_i \overset{i.i.d.}{\sim} P \text{ for some distribution } P \text{ on } [0,1]^{d_N} \text{ and } p_i \perp\!\!\!\perp z_i \mid \mathcal{F}_{i-1}, \tag{38a}$$

$$\varepsilon_i \text{ is independent of } (x_i, z_i, \mathcal{F}_{i-1}) \text{ and } \varepsilon_i \overset{i.i.d.}{\sim} Q \text{ for some distribution } Q, \tag{38b}$$

且 h^* 关于参数 $L > 0$ 利普希茨连续，即，

$$|h^*(v_1) - h^*(v_2)| \le L\|v_1 - v_2\|_2 \qquad \text{for all } v_1, v_2 \in [0,1]^{d_N}. \tag{38c}$$

在这些假设下，可以使用各种非参数方法（例如，k 近邻估计量 [24, 58]）来找到一致的初始估计量 h，用于估计 h^*。给定这样的初始估计量，应用定理 1 可得：

推论 4. 假设对于某些 $t \in [0, 1/2)$、$\widetilde{\ }$，假设 ***(NOI**(ν, σ^2)**)*** 和 ***(SEL**(t)**)*** 成立，并且条件 (38a) 和 (38c)也成立。那么，使用 h 作为初始估计量得到的 $\widehat{\ }$ θ *AdapTZ-PL* 估计量满足

$$(\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})(\widetilde{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2\mathbf{I}_{d_T}). \tag{39}$$

有关 k 近邻估计量及推论 4 的证明，详见附录B.3。请注意，在方程 (38a) 中，我们要求选择概率仅取决于历史信息 $\mathcal{F}_{i-1}$，而不取决于干扰项 z_i。在实践中，这对应于这样一种情形：治疗分配方案（用于确定选择概率 p_i）必须在观测个体上下文向量 z_i 之前确定。我们施加这一条件$\widehat{\ }$是为了简化分析，但请注意，只要能够为 h^* 构造一个一致的初始估计量 h，就可以去掉这一条件。

4 数值结果

本节通过一系列数值研究展示我们的理论保证。我们给出了自适应线性模型和自适应逻辑回归模型的结果。除了展示基于我们所提出算法的结果外，我们还与其他现有方法进行比较，包括：(a) 最大似然估计量；(b) 基于集中不等式的方法；以及 (c) 一种由双重机器学习（DML）方法 [13] 推导出的现有 Z 估计量程序。

4.1 自适应线性模型

本节研究（可能是）高维线性模型中的半参数问题。与第 3.1 节和第 3.2 节中的应用一样，我们考虑如下线性模型

$$y_i = \langle x_i,\, \theta^* \rangle + \langle z_i,\, \eta^* \rangle + \varepsilon_i,$$

其中，三元组 (y_i, x_i, z_i) 按以下方式自适应收集：

(1) 干扰分量 z_i 具有独立同分布的 $\mathcal{N}(0, 1)$ 条目，且与历史信息$\mathcal{F}_{i-1}$相互独立。

(2) 给定 $(y_1, x_1, z_1) \ldots, (y_{i-1}, x_{i-1}, z_{i-1})$ 后，我们求解一个 LASSO（或普通最小二乘法）问题，以获得估计值 $\widehat{\theta}^i$ 和 $\widehat{\eta}^i$。

(3) 根据估计量 $\widehat{\theta}^i$，算法选择一个臂 $k_i := \mathsf{argmax}_k \{\widehat{\theta}^i_k + \sqrt{C \log_{n^i_k} n}\}$，其中 $C > \overline{0}$ 是某个常数，n^i_k 是截至时间 $i-1$臂 k 被选择的次数。

(4) 最后，根据臂选择概率 $p_i \in \mathbb{R}^{d_T}$选择回归变量 x_i （即 $\mathbb{P}(x_i = e_k) = p_{ik}$，并定义 $e_0 := 0$），其中我们设定

$$p_{i0} = 0.2, \qquad p_{ik} = \min\{\frac{1}{2i^{2t}}, \frac{0.4}{d_T}\} \quad \text{for } k \neq k_i, \quad \text{and} \quad p_{ik_i} = 1 - \sum_{0 \leq k \leq d_T, k \neq k_i} p_{ij}.$$

我们获得的观测值 $y_i = \langle x_i, \quad \theta^* \rangle + \langle z_i, \quad \eta^* \rangle + \varepsilon_i$受到噪声 $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \quad 1)$的污染。当 $i = 1$时，对所有 $k \in [d_T]$，我们设定 $p_{i0} = 0.2$, $p_{ik} = 0.8/d_T$。

收集数据后，我们应用 AdapTZ-PL 方法（算法 1，其中得分函数由方程 (16a) 定义），对目标参数 θ^*_1的第一个坐标进行推断。

在第一个实验中，我们选择目标维度 $d_T = 2$、干扰维度$d_N = 5$ 和样本数 $n = 500$。我们考虑满足$\theta^*_1 = \theta^*_2 = 2$的无间隔情形。在这些条件下，Zhang等人 [67] 证明，选择概率 p 可能不收敛，因此稳定性条件可能不成立。此外，我们假设干扰参数向量 η^* 是由 $\mathcal{N}(0, \mathbf{I}_{d_N})$ 生成的固定向量；我们选择普通最小二乘法估计量作为 θ^*, η^* 的初始估计量，并使用 $n_1 = n/4 = 125$ 个样本。结果如图2所示。

我们将 AdapTZ-PL 估计量与另外三种方法进行比较：(i) 普通最小二乘法；(ii) 基于未加权得分函数的双重机器学习 Z估计量

$$\phi_i(\theta, h) := (x_i - p_i)(y_i - \langle x_i,\, \theta \rangle - h(z_i)); \tag{40}$$

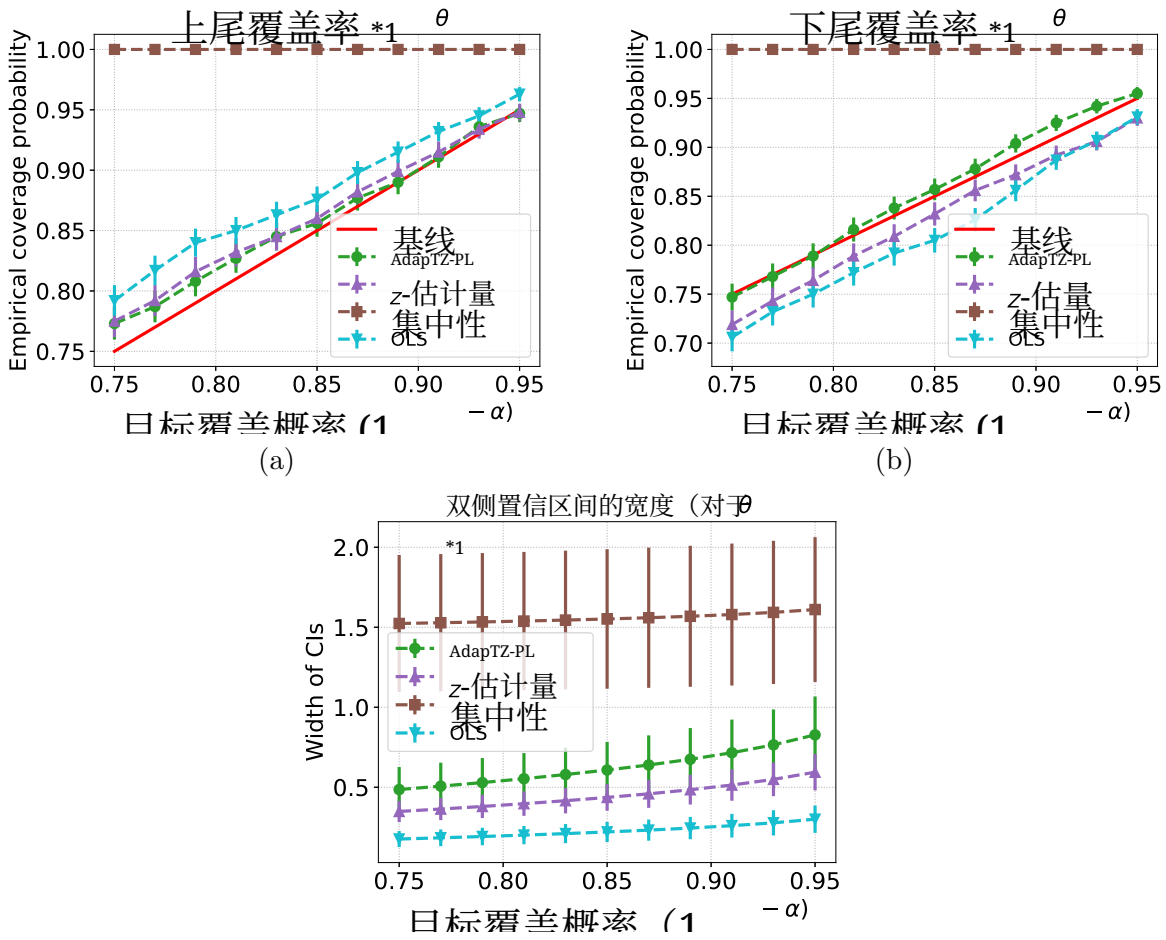


图2。 θ_1^* 在自适应线性模型的 $T = 1000$ 次重复实验中的置信区间平均覆盖率和宽度。误差条表示 ± 1 标准误。参数： d_T =2， $d_N = 5$， $n = 500$， $n_1 = 125$、 $C = 2$ 和 $t = 0.2$。（a）和（b）： θ_1^*的 $1-\alpha$ 水平单侧置信区间的覆盖率。（c）： θ_1^*的 $1-\alpha$ 水平双侧置信区间的宽度。

以及 (iii) 根据标准集中不等式推导出的置信区间（参见 Abbasi-Yadkori 等人 [1]的定理 2。）图2展示了各方法得到的经验覆盖概率和置信区间宽度。我们发现，AdapTZ-PL 方法在所有置信水平下都能提供适当的覆盖率。然而，普通最小二乘估计量和 Z 估计量虽然能提供有效的上尾覆盖率，且置信区间更短，但二者均存在向下偏差 [47]，无法实现适当的下尾覆盖率。

在第二个实验中，我们考虑一个具有高维干扰参数的线性模型。具体而言，取 $(d_T, d_N) = (2, 1000)$，生成 $n = 950$ 个样本。与第一个实验类似，我们考虑无间隔情形，其中 $\theta_1^* = \theta_2^* = 2$。我们还假设该线性模型是稀疏的，即 $\eta_i^* = 0$ 对于 $i > 2$ 成立，并且干扰参数向量 η_1^*, η_2^*的前两个坐标由 $\mathcal{N}(0, \mathbf{I}_2)$ 生成。我们以与第一个实验相同的方式生成样本，但在$\widehat{}$数据生成过程中使用 LASSO 估计量（参见方程 31），并用它来获得参数的初步估计 $\theta, \widehat{\eta}$。我们选择

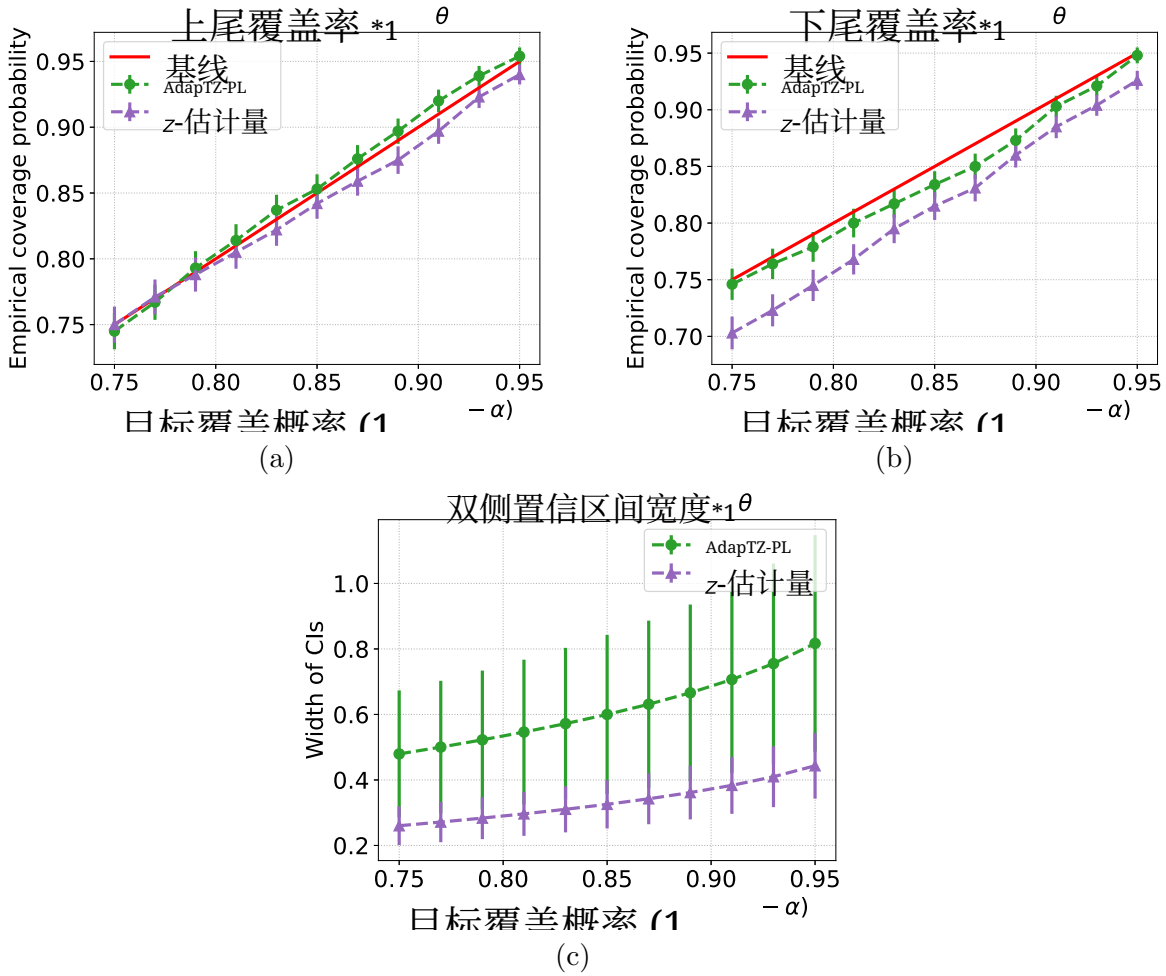


图3。 θ_1^* 自适应线性模型经过1000次重复实验后的置信区间平均覆盖率和宽度。误差条表示 ± 1 标准误。参数： $d_T = 2$， d_N =1000， $n = 950$， $n_1 = 475$、 $C = 16$ 和 $t = 0.2$。（a）和（b）： θ_1^*水平为 $1 - \alpha$ 的单侧置信区间的覆盖率。（c）： θ_1^*水平为 $1 - \alpha$ 的双侧置信区间宽度。

用于数据生成和初步估计的 Lasso 正则化分别为 $\lambda = 0.05$,0.15。

图3比较了AdapTZ与标准双重机器学习 Z-估计量的覆盖概率。可以看到，AdapTZ程序在大多数水平上都实现了适当的经验覆盖概率。与低维情形类似，尽管 Z-估计量的方差较低，但它存在向下偏差，因而无法实现适当的覆盖率。此处我们不提供由Abbasi-Yadkori 等人 [1] 基于集中不等式推导出的置信区间，因为由于界中包含 $\sqrt{d/n}$ 因子，该区间过宽。

4.2 自适应逻辑回归模型

接下来，我们将演示如何将 AdapTZ-GLM 方法应用于具有自适应收集的数据的逻辑回归模型。我们按照第4.1节所述的过程生成数据，并进行以下更改：

（1）va　变量 z_i 由自回归过程 $z_i = \gamma z_{i-1} + W_i$生成，其中　e

$z_0 := 0$，$\gamma = 0.5$ 和 W_i 是服从 $\mathcal{N}(\mathbf{0}, \mathbf{I}_{d_N})$ 的独立同分布随机变量。

(2) 作为先导估计量，我们分别计算未知参数 θ^* 和 η^*的最大似然估计 $\widehat{\theta}$ 和 $\widehat{\eta}$。

(3) 响应变量 y_i 是均值为 $g\big(\langle x_i, \theta^* \rangle + \langle z_i, \eta^* \rangle\big)$ 给定条件下的伯努利随机变量$\mathcal{F}_{i-1}$，x_i，z_i。

在这个例子中，我们研究一个维度为 $d_T = 2$ 和 $d_N = 20$ 的低维实例，以及样本量为 $n = 2000$ 的情形。同样，我们设定 $\theta_1^* = \theta_2^* = 2$，并令 η^* 为从 $\mathcal{N}(\mathbf{0}, \mathbf{I}_{d_N})$ 生成的固定向量。此外，我们使用最大似然估计，根据 $n_1 = n/2 = 1000$ 个样本为 θ^* 和 η^* 生成初步估计。

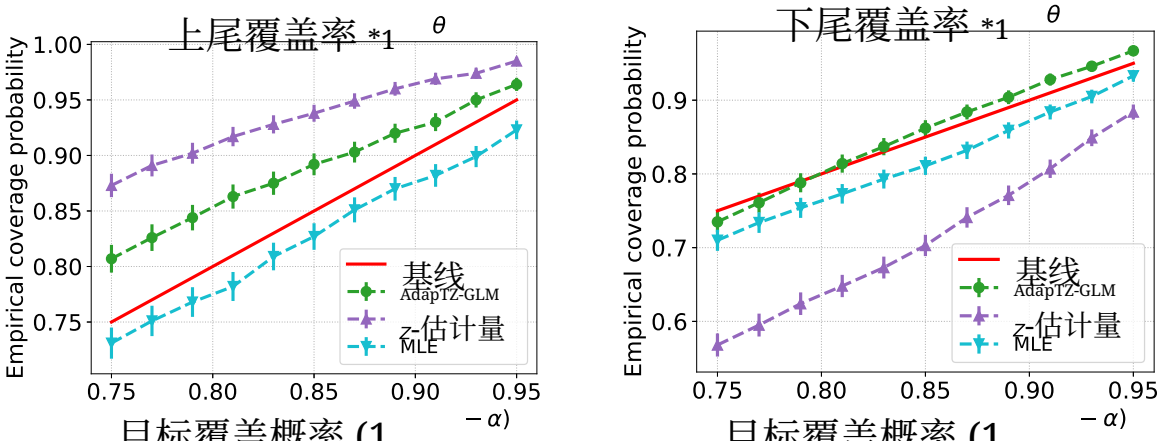
从图4中可以看出，AdapTZ 和 Z-估计量的上尾覆盖率均高于预设水平。然而， Z-估计量和最大似然估计都未能达到适当的下尾覆盖率。这与我们此前在线性模型中的观察一致。此外，值得注意的是，AdapTZ 的经验覆盖概率与基线并非完全吻合，这可能是由于样本量相对较小。

最后，我们还针对具有稀疏高维干扰分量的逻辑回归模型进行了实验。具体而言，我们生成 $n = 950$ 个样本，取值为 (d_T, d_N) =(2,1000)。我们考虑无间隔情形，其中 $\theta_i^* = \theta_2^* = 2$ 与先前实验相同。我们假设逻辑回归模型是稀疏的，即对于 $i > 5$ 和 $\eta_{1:5}^* = \mathbf{1}_5$，$\eta_i^* = 0$。我们假设数据的生成过程与第一个逻辑回归模型实验相同，但在数据生成中使用逻辑回归的 LASSO 估计量（参见方程 35），惩罚项为 $\lambda = 0.0025$，并据此获得初步估计 $\widehat{\theta}, \widehat{\eta}$。此外，我们假设自回归过程中随机向量 $W_i \in \mathbb{R}^{d_N}$ 的生成方式改为：对于 $k \le 5$，$W_{ik} \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_5)$；对于 $6 \le k \le d_N$，$W_{ik} \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, 1/d_N)$。我们选择 W_i 的异方差，以确保干扰分量的范数为 $\mathcal{O}_p(1)$ 量级，并使非稀疏干扰项对响应变量 y_i 产生不可忽略且可检测的信号。

在图5中，我们看到 AdapTZ-GLM 的上尾和下尾覆盖率均高于预先指定的水平，而朴素的 Z 估计量虽然方差较小，却存在向下偏差，且下尾覆盖率不足。同样，AdapTZ-GLM 与基线并未完全吻合，这可能是因为逻辑回归问题中的样本量相对较小。

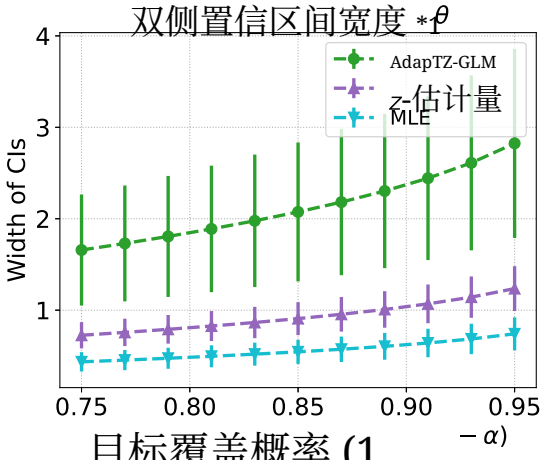
5 讨论

本文研究了存在高维或非参数干扰成分时，如何为低维目标参数构造置信区间。我们工作的主要创新在于，应对数据经过自适应收集时构造此类区间所面临的挑战。我们提出了一类称为 AdapTZ方法 的程序，其基础是对两阶段 Z 估计量进行自适应重加权。我们针对部分线性模型以及更一般的、具有半参数干扰项的广义线性模型，分别开发了这些程序的相应版本。主要结果表明，在某些正则条件下，这类估计量的某些版本具有渐近正态性。我们分析的显著特点包括：(a) 对自适应数据收集过程，我们仅假设较为温和的“可探索性”条件；(b) 与先前最先进的 [37, 38]，方法不同，我们不要求任何形式的稳定性条件。



(a)

(b)



(c)

图4. θ_1^* 自适应逻辑回归模型重复1000次后，置信区间的平均覆盖率和宽度。误差条表示 ± 1 标准误。参数： d_T =2, $d_N = 20, n = 2000, n_1 = 1000$, $C = 8$ 和 $t = 0.1$。面板(a)和(b)给出了 θ_1^*的$1 - \alpha$ 单侧置信区间覆盖率。面板(c)显示了 θ_1^*的 $1 - \alpha$ 双侧置信区间宽度。

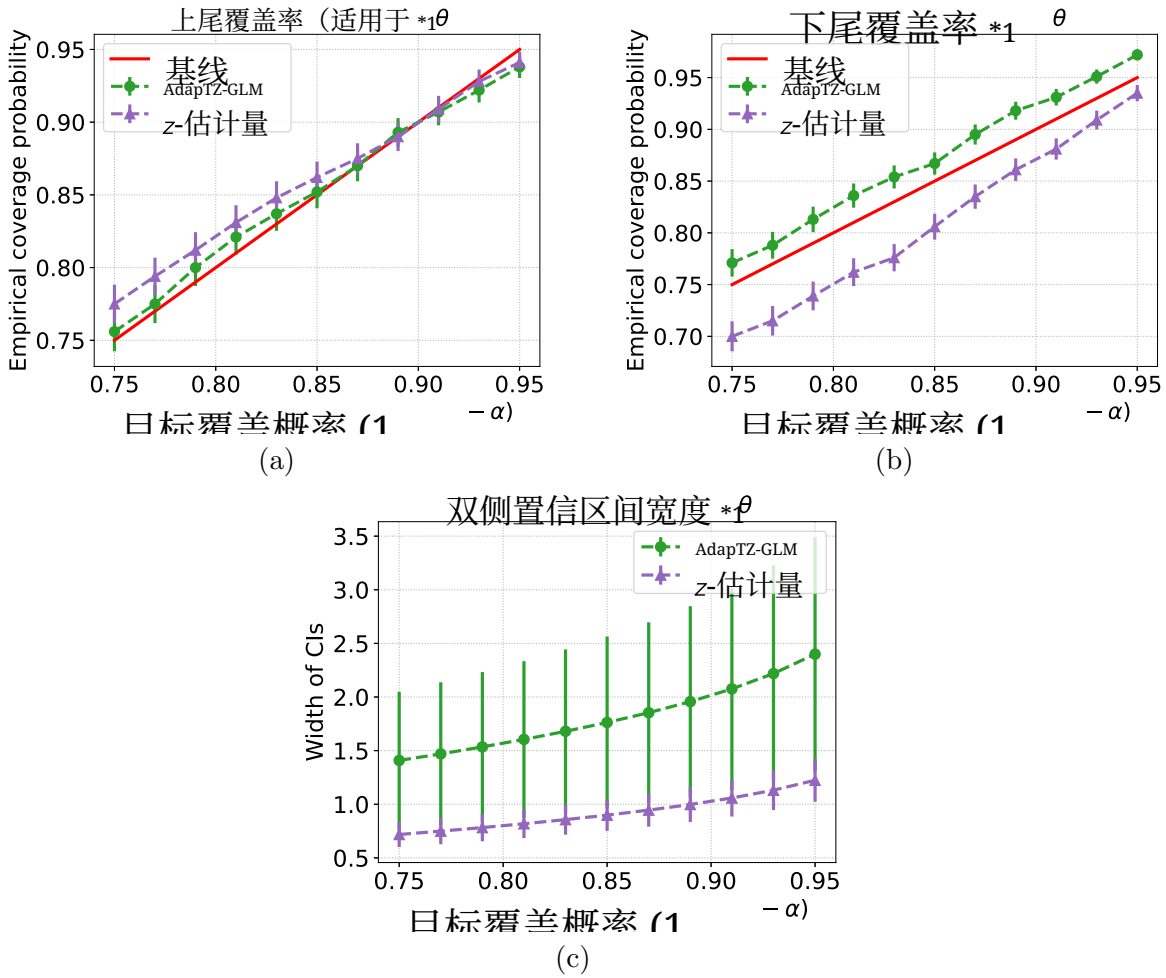


图5. 自适应逻辑回归模型经过 1000 次重复实验后，θ_1^* 的平均覆盖率和置信区间宽度。误差条表示 ± 1 标准误。参数：d_T =2, $d_N = 1000, n = 950, n_1 = 475$, $C = 100$ 和 $t = 0.1$。(a) 和 (b)：θ_1^*的 $1-\alpha$ 水平单侧置信区间的覆盖率。(c)：θ_1^*的 $1-\alpha$ 水平双侧置信区间宽度。

我们的工作为未来研究指明了若干方向。首先，本文的结果为半参数模型中的固定维度参数向量提供了推断保证（其中干扰量可能是高维的或非参数的）。若能扩展我们的结果，使其也允许目标参数为高维，或更一般地适用于具有非参数特征的目标，将会很有意义。其次，我们给出了渐近正态性保证，其方差取决于具体问题实例。在独立同分布数据的半参数文献中，已有研究刻画了依赖具体实例的最优性概念——即对于$\sqrt{n}$相合估计量而言，具有最小方差（例如，[43, 26]）。而在本文所讨论的更具挑战性的自适应数据情形下，这些最优性概念尚未得到充分理解。为本文研究的自适应模型推导紧下界，并提出能够达到这些界的估计量，将是很有意义的。

第三，我们构造自适应加权Z估计量时，依赖于已知每一轮次的选择概率。在一些应用中，包括实验设计和

老虎机实验中，这一假设是合理的。然而，对于各种观察性研究而言，这一假设不太现实，因此，设计无需此类信息即可运行的最优程序是一项重要的研究方向。

6 致谢

本研究部分获得美国海军研究办公室项目 ONR-N00014-21-1-2842 和美国国家科学基金会项目 DMS-2311072 对 MJW 的资助，以及 UCB 伯克利 Howard Friesen 工程学讲席的经费支持。

参考文献

[1] Yasin Abbasi-yadkori、D´avid P´al 和 Csaba Szepesv´ari。线性随机老虎机的改进算法。载于神经信息处理系统进展，第24卷，2011年。[2] Alekh Agarwal、Daniel Hsu、Satyen Kale、John Langford、Lihong Li 和 Robert Schapire。驯服怪兽：一种快速简便的上下文老虎机算法。载于国际机器学习会议，第1638–1646页。PMLR，2014。[3] Chunrong Ai 和 Xiaohong Chen。含未知函数的条件矩约束模型的有效估计。《计量经济学》，71(6)：1795–1843，2003年。[4] Donald WK Andrews。通过随机等度连续性研究半参数计量经济模型的渐近性质。《计量经济学》（计量经济学会期刊），第43–72页，1994。[5] Susan Athey、Julie Tibshirani 和 Stefan Wager。广义随机森林。《统计年刊》，47(2)：1148–1178，2019年。[6] Alexandre Belloni、Victor Chernozhukov 和 Lie Wang。通过平方根 LASSO 进行非参数回归的枢轴估计。《统计年刊》，42(2)：757–788，2014。[7] Alexandre Belloni、Victor Chernozhukov 和 Ying Wei。针对含有众多控制变量的广义线性模型进行后选择推断。《商业与经济统计杂志》，34(4)：606–619，2016。[8] Aur´elien Bibaut、Maria Dimakopoulou、Nathan Kallus、Antoine Chambaz 和 Mark van der Laan。上下文老虎机后推断。神经信息处理系统进展，34：28548–28559，2021。[9] Peter J Bickel。论自适应估计。《统计年刊》，第647–671页，1982。[10] Peter J Bickel、Chris AJ Klaassen、Peter J Bickel、Ya'acov Ritov、J Klaassen、Jon A Well-ner 和 Ritov。半参数模型的有效与自适应估计，第4卷。施普林格，1993。[11] Hugo Bodory、Martin Huber 和 Luk´aˇs Laff´ers。通过双重机器学习评估（加权）动态处理效应。《计量经济学期刊》，2022年。[12] George EP Box、Gwilym M Jenkins、Gregory C Reinsel 和 Greta M Ljung。时间序列分析：预测与控制。John Wiley & Sons，2015年。

[13] Victor Chernozhukov、Denis Chetverikov、Mert Demirer、Esther Duflo、Christian Hansen、Whitney Newey 和 James Robins。用于处理效应和结构参数的双重/去偏机器学习。《计量经济学期刊》，21(1)：C1–C68，2018年。[14] Victor Chernozhukov、Whitney Newey、Rahul Singh 和 Vasilis Syrgkanis。用于动态处理效应及一般嵌套泛函的自动去偏机器学习。*arXiv* 预印本 *arXiv:2203.13887*，2022年。[15] Kyle Colangelo 和 Ying-Ying Lee。连续处理下双重去偏机器学习的非参数推断。*arXiv* 预印本 *arXiv:2004.03036*，2020。[16] Pierre Del Moral 和 Angele Niclas。矩阵平方根函数的泰勒展开。《数学分析与应用杂志》，465(1)：259–266，2018年。[17] Yash Deshpande、Lester Mackey、Vasilis Syrgkanis 和 Matt Taddy。自适应线性模型的准确推断。载于国际机器学习会议，第1194–1203页。PMLR，2018年。[18] David A. Dickey 和 Wayne A. Fuller。含单位根的自回归时间序列估计量的分布。《美国统计学会杂志》，74(366)：427–431，1979年。[19] Jianqing Fan、Shaojun Guo 和 Ning Hao。利用重新拟合交叉验证进行超高维回归的方差估计。《英国皇家统计学会杂志：*B*辑（统计方法论）》，74(1)：37–65，2012年。[20] Qingliang Fan、Yu-Chin Hsu、Robert P. Lieli 和 Yichong Zhang。利用高维数据估计条件平均处理效应。《商业与经济统计杂志》，40(1)：313–327，2022年。[21] Caroline A Figueroa、Adrian Aguilera、Bibhas Chakraborty、Arghavan Modiri、Jai Aggar-wal、Nina Deliu、Urmimala Sarkar、Joseph Jay Williams 和 Courtney R Lyles。用于优化行为健康移动应用的自适应学习算法：设计决策指南。《美国医学信息学会杂志》，28(6)：1225–1234，2021年。[22] Xavier Fontaine、Pierre Perrault、Michal Valko 和 Vianney Perchet。在线A最优设计与主动线性回归。载于国际机器学习会议，第3374–3383页。PMLR，2021年。[23] Dylan J Foster 和 Vasilis Syrgkanis。正交统计学习。*arXiv* 预印本 *arXiv:1901.09036*，2019年。[24] L´aszl´o Gy¨orfi、Michael Kohler、Adam Krzyzak、Harro Walk 等。非参数回归的无分布理论，第1卷。施普林格，2002年。[25] Vitor Hadad、David A Hirshberg、Ruohan Zhan、Stefan Wager 和 Susan Athey。自适应实验中政策评估的置信区间。《美国国家科学院院刊》，118(15)：e2014602118，2021年。[26] Jinyong Hahn。倾向得分在平均处理效应有效半参数估计中的作用。《计量经济学》，66(2)：315–331，1998年。

[27] Roger A Horn 和 Charles R Johnson。《矩阵分析》。剑桥大学出版社，2012年。[28] Steven R Howard、Aaditya Ramdas、Jon McAuliffe 和 Jasjeet Sekhon。时间一致的非参数非渐近置信序列。《统计年刊》，49(2):1055–1080，2021。[29] Kevin Jamieson、Matthew Malloy、Robert Nowak 和 S´ebastien Bubeck。lil'ucb：一种用于多臂老虎机的最优探索算法。载于《学习理论会议》，第423–439页。PMLR，2014。[30] Adel Javanmard 和 Andrea Montanari。高维回归的置信区间与假设检验。《机器学习研究期刊》，15(1):2869–2909，2014。[31] Masahiro Kato、Kenichiro McAlinn 和 Shota Yasui。自适应双重稳健估计量及与日志策略相关的悖论。载于《神经信息处理系统进展》，第34卷，第1351–1364页，2021。[32] Emilie Kaufmann、Olivier Capp´e 和 Aur´elien Garivier。多臂老虎机模型中最佳臂识别的复杂度。《机器学习研究期刊》，17(1):1–42，2016。[33] Edward H. Kennedy。迈向异质性因果效应的最优双重稳健估计。《arXiv预印本 arXiv:2004.14497》，2020。[34] Koulik Khamaru、Yash Deshpande、Lester Mackey 和 Martin J Wainwright。自适应线性回归中的近最优推断。《arXiv预印本 arXiv:2107.02266》，2021。[35] Michael C Knaus。无混杂条件下基于双重机器学习的项目评估。《计量经济学期刊》，2022年06月。[36] Thomas Koshy。《具有应用的卡特兰数》。牛津大学出版社，2008年。[37] Tze Leung Lai。随机回归模型中非线性最小二乘估计值的渐近性质。《统计年刊》，第1917–1930页，1994。[38] Tze Leung Lai 和 Ching Zong Wei。随机回归模型中的最小二乘估计及其在动态系统辨识与控制中的应用。《统计年刊》，10(1):154–166，1982。[39] Tor Lattimore 和 Csaba Szepesv´ari。《老虎机算法》。剑桥大学出版社，2020。[40] Greg Lewis 和 Vasilis Syrgkanis。用于动态处理效应的双重/去偏机器学习。载于《神经信息处理系统进展》，第34卷，第22695–22707页。Curran Associates, Inc.，2021。[41] Molei Liu、Yi Zhang 和 Doudou Zhou。逻辑部分线性模型的双重/去偏机器学习。《计量经济学期刊》，24(3):559–588，2021。[42] Shie Mannor 和 John N Tsitsiklis。多臂老虎机问题中探索的样本复杂度。《机器学习研究期刊》，5:623–648，2004。

[43] Whitney K. Newey。半参数效率界。《应用计量经济学杂志》，5(2):99–135，1990年。[44] Whitney K Newey。半参数估计量的渐近方差。《计量经济学》（计量经济学会期刊），第1349–1382页，1994。[45] J. Neyman。复合统计假设的最优渐近检验。载于 U. Grenan-der（编），《概率与统计》，第416–44页，1959年。[46]X Nie 和 S Wager。异质性处理效应的准预言机估计。《生物统计学》，108(2):299–319，2020年09月。[47] Xinkun Nie、Xiaoying Tian、Jonathan Taylor 和 James Zou。为什么自适应收集的数据具有负偏差，以及如何纠正这一偏差。载于《神经信息处理系统进展》，84:1261–1269，2018年。[48] Min-hwan Oh、Garud Iyengar 和 Assaf Zeevi。稀疏性无关的 Lasso 多臂老虎机。载于《国际机器学习会议》，第8271–8280页。PMLR，2021。[49] Miruna Oprescu、Vasilis Syrgkanis 和 Zhiwei Steven Wu。用于因果推断的正交随机森林。载于《国际机器学习会议》，第4932–4941页。PMLR，2019年。[50] Johann Pfanzagl。《一般渐近统计理论的贡献》，第13卷。施普林格科学与商业媒体，2012年。[51] James M Robins 和 Andrea Rotnitzky。带有缺失数据的多元回归模型中的半参数效率。《美国统计学会杂志》，90(429):122–129，1995年。[52] Peter M Robinson。根号 n 一致的半参数回归。《计量经济学》（计量经济学会期刊），第931–954页，1988年。[53] Anton Schick。关于半参数模型中的渐近有效估计。《统计年刊》，第1139–1151页，1986年。[54] Vira Semenova 和 Victor Chernozhukov。条件平均处理效应及其他因果函数的去偏机器学习。《计量经济学期刊》，24(2):264–289，2021。[55] Vasilis Syrgkanis 和 Ruohan Zhan。回合后强化学习推断，2023年。[56] Ambuj Tewari 和 Susan A Murphy。从广告到干预：移动健康中的上下文老虎机。载于《移动健康：传感器、分析方法与应用》，第495–517页，2017年。[57] Anna L Trella、Kelly W Zhang、Inbal Nahum-Shani、Vivek Shetty、Finale Doshi-Velez 和 Susan A Murphy。支持口腔自我护理的在线强化学习算法的奖励设计。《arXiv预印本 arXiv:2208.07406》，2022年。[58] A. Tsybakov。非参数估计导论。载于《施普林格统计学丛书》，2008年。

[59] Mark J Van der Laan、Sherri Rose 等。《目标学习：观察性和实验数据的因果推断》，第10卷。施普林格，2011年。[60]Stefan Wager 和 Susan Athey。使用随机森林估计和推断异质性治疗效应。《美国统计学会杂志》，113(523):1228–1242，2018年。[61] Martin J Wainwright。《高维统计：非渐近视角》，第48卷。剑桥大学出版社，2019年。[62] Shihan Wang、Karlijn Sporrel、Herke van Hoof、Monique Simons、R´emi DD de Boer、Dick Ettema、Nicky Nibbeling、Marije Deutekom 和 Ben Kr¨ose。利用强化学习在智能手机运动应用中适时发送提醒：一项可行性研究。《国际环境研究与公共卫生杂志》，18(11):6059，2021年。[63] John S. White。爆炸情形II下序列相关系数的极限分布。《数理统计年刊》，30(3):831–834，1959年。[64] Elad Yom-Tov、Guy Feraru、Mark Kozdoba、Shie Mannor、Moshe Tennenholtz 和 Irit Hochberg。鼓励糖尿病患者进行身体活动：使用强化学习系统进行干预。《医学互联网研究杂志》，19(10):e338，2017年。[65] Ruohan Zhan、Vitor Hadad、David A Hirshberg 和 Susan Athey。利用上下文老虎机数据通过自适应加权进行离策略评估。载于《第27届ACM SIGKDD知识发现与数据挖掘会议论文集》，第2125–2135页，2021年。[66] Cun-Hui Zhang 和 Stephanie S Zhang。高维线性模型中低维参数的置信区间。《英国皇家统计学会杂志：B辑（统计方法论）》，76(1):217–242，2014年。[67] Kelly Zhang、Lucas Janson 和 Susan Murphy。批处理老虎机的推断。《神经信息处理系统进展》，33:9818–9829，2020年。[68] Kelly Zhang、Lucas Janson 和 Susan Murphy。基于自适应收集数据的m估计量统计推断。《神经信息处理系统进展》，34:7460–7471，2021年。

目录

A 定理证明 **32**
A.1 定理 1 的证明.32 A.2 定理2的证明.33 A.3 定理 3 的证明.35 A.4 定理 4 的证明.38

B 推论证明 **38**
B.1 推论 1 的证明.39 B.2 推论 2 的证明.39 B.3 推论 3 的证明.40 B.4 推论 4 的证明.40
B.5 引理 1 的证明.41

C Neyman 正交性及其他假设 42
C.1 线性模型.42 C.2 广义线性模型.43 C.3 关于逻辑回归假设的评论.43

D 干扰函数的自适应估计 46
D.1 推论 5 的证明.47 D.2 推论 6 的证明.48

E 推断 当 p_i 未知时 48
E.1 推论 7 的证明.49

F 辅助引理 **51**
F.1 定理 1 的辅助引理.51 F.2 定理 2 的辅助引理.54 F.3 定理 3 的辅助引理.56

G 技术引理及其证明 **71**
G.1 鞅差序列.71 G.2 假设（**SEL**(t)）的等价条件.72

A 定理证明

我们给出四个一般性结果的证明，A.1 至 A.4 节分别证明定理 1 至定理 4。

A.1 定理 1 的证明

回顾得分函数 ϕ_i的定义（9），注意到它关于参数向量 θ 和 h呈线性，并且我们有如下方便的分解

$$\widehat{\mathbb{E}}_{n_2}\phi_i(\theta,h)=(\widehat{\mathbb{E}}_{n_2}v_ix_i^\top)(\theta-\theta^*)+\widehat{\mathbb{E}}_{n_2}v_i\varepsilon_i-\widehat{\mathbb{E}}_{n_2}v_i(h(z_i)-h^*(z_i)),$$

其中 $v_i:=\mathbf{\Sigma}_i^{-1/2}(x_i-p_i)$。根据我们的 Z-估计量的定义，$(\widetilde{\theta},\widehat{h})$ 满足条件 $\mathbb{E}_{n_2}\phi_i(\theta,h)=0$。重新整理该等式，并将两边乘以 $\sqrt{n_2}$，得到

$$\sqrt{n_2}(\widehat{\mathbb{E}}_{n_2}v_ix_i^\top)(\widetilde{\theta}-\theta^*)=\sqrt{n_2}\big\{\widehat{\mathbb{E}}_{n_2}v_i\varepsilon_i-\widehat{\mathbb{E}}_{n_2}v_i(\widehat{h}(z_i)-h^*(z_i))\big\}. \tag{41}$$

接下来，我们通过以下三个结果分析方程（41）；这些结果将在引理 3、4 和 5 中证明（详见附录）。

$$\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}v_i\varepsilon_i\xrightarrow{d}\mathcal{N}(0,\sigma^2\mathbf{I}_{d_T}) \tag{42a}$$

$$\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}v_i(\widehat{h}(z_i)-h^*(z_i))\xrightarrow{p}0 \tag{42b}$$

$$\|\widehat{\mathbb{E}}_{n_2}v_ix_i^\top-\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\|_{\mathrm{op}}=o_p(\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})) \tag{42c}$$

有了这三个结果，证明的其余部分便十分直接。事实上，将条件（42a）和（42b）代入方程（41），并应用斯卢茨基定理，得到

$$(\widehat{\mathbb{E}}_{n_2}v_ix_i^\top)\sqrt{n_2}(\widetilde{\theta}-\theta^*)\xrightarrow{d}\mathcal{N}(0,\sigma^2\mathbf{I}_{d_T}). \tag{43}$$

根据连续映射定理、弱收敛的定义以及依概率有界性，这种分布收敛还意味着 $\|(\widehat{\mathbb{E}}_{n_2}v_ix_i^\top)\sqrt{n_2}(\widetilde{\theta}-\theta^*)\|_2=\mathcal{O}_p(1)$。将最后一个推论与界 (42c) 结合可得

$$\begin{aligned}\|(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\sqrt{n_2}(\widetilde{\theta}-\theta^*)\|_2&\le\|(\widehat{\mathbb{E}}_{n_2}v_ix_i^\top)\sqrt{n_2}(\widetilde{\theta}-\theta^*)\|_2+\|(\widehat{\mathbb{E}}_{n_2}(v_ix_i^\top-\mathbf{\Sigma}_i^{1/2})\sqrt{n_2}(\widetilde{\theta}-\theta^*)\|_2\\&=\mathcal{O}_p(1)+o_p(\|(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\sqrt{n_2}(\widetilde{\theta}-\theta^*)\|_2).\end{aligned}$$

综合各部分可得 $\|(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\sqrt{n_2}(\widetilde{\theta}-\theta^*)\|_2=\mathcal{O}_p(1)$，进而推出

$$\|(\widehat{\mathbb{E}}_{n_2}(v_ix_i^\top-\mathbf{\Sigma}_i^{1/2})\sqrt{n_2}(\widetilde{\theta}-\theta^*)\|_2=o_p(1)$$

最后，将最后一个结果与收敛结论 (43) 结合，并应用斯卢茨基定理，我们得出

$$(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})\sqrt{n_2}(\widetilde{\theta}-\theta^*)\xrightarrow{d}\mathcal{N}(0,\sigma^2\mathbf{I}_{d_T}).$$

定理 1 的证明至此完成。

对连续回归变量的推广： 我们注意到，即使回归变量 x_i 取连续值，也可以建立三个关键引理（引理3、4和5）的相应版本；因此，定理 1 可以推广到这一连续情形。具体而言，这些引理证明中的一个关键环节是对协方差 $\mathbf{\Sigma}_i$使用下界条件——特别参见附录G中的假设 **(A2b)** ，以控制某些辅助量的 ℓ_2范数。虽然我们证明了在离散回归变量的情形下，假设（**SEL(**t**)**)与**(A2b)**等价，但正如我们之前所述，在连续回归变量的情形下，假设**(A2b)** 本身已被假定。因此，所有推导都遵循相同的论证。有关这些论证的更多细节，请参见附录F。

一致性 $\widehat{\sigma}^2$ **在公式(15)** 在此，我们证明公式（15）中的估计量 $\widehat{\sigma}^2$ 是噪声方差 σ^2的相合估计。注意，

$$\begin{aligned}
|\widehat{\sigma}^2 - \sigma^2| &= |\widehat{\mathbb{E}}_{n_2}(y_i - x_i^\top \widetilde{\theta} - \widehat{h}(z_i))^2 - \sigma^2| \\
&= |\widehat{\mathbb{E}}_{n_2}(\varepsilon_i + x_i^\top \theta^* + h^*(z_i) - x_i^\top \widetilde{\theta} - \widehat{h}(z_i))^2 - \sigma^2| \\
&\le |\widehat{\mathbb{E}}_{n_2}\varepsilon_i^2 + \widehat{\mathbb{E}}_{n_2}(x_i^\top(\theta^* - \widetilde{\theta}) + (h^*(x_i) - \widehat{h}(z_i)))^2 - \sigma^2| \\
&\quad + 2\sqrt{\widehat{\mathbb{E}}_{n_2}\varepsilon_i^2} \cdot \sqrt{\widehat{\mathbb{E}}_{n_2}(x_i^\top(\theta^* - \widetilde{\theta}) + (h^*(z_i) - \widehat{h}(z_i)))^2},
\end{aligned}$$

其中第三行由柯西–施瓦茨不等式得出。

我们声称

$$\widehat{\mathbb{E}}_{n_2}\varepsilon_i^2 \xrightarrow{p} \sigma^2, \quad \text{and} \quad \widehat{\mathbb{E}}_{n_2}(x_i^\top(\theta^* - \widetilde{\theta}) + (h^*(z_i) - \widehat{h}(z_i)))^2 \xrightarrow{p} 0. \tag{44}$$

E公式（15）可由前述界和这两个辅助命题立即推出。为证明公式（44）中的第一个命题，注意到

$$\widehat{\mathbb{E}}_{n_2}\varepsilon_i^2 = \widetilde{\mathbb{E}}_{n_2}\varepsilon_i^2 + (\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2})\varepsilon_i^2 = \sigma^2 + o_p(1) \xrightarrow{p} \sigma^2,$$

其中第二个不等式用到了假设（**NUI**）以及 ε_i的有限四阶矩条件。为证明第二个命题，注意到

$$\begin{aligned}
&\widehat{\mathbb{E}}_{n_2}(x_i^\top(\theta^* - \widetilde{\theta}) + (h^*(z_i) - \widehat{h}(z_i)))^2 \\
&\le 2\widehat{\mathbb{E}}_{n_2}\big[x_i^\top(\theta^* - \widetilde{\theta})^2\big] + 2\widehat{\mathbb{E}}_{n_2}\big[(h^*(z_i) - \widehat{h}(z_i))^2\big] \\
&\le 2\|\theta^* - \widetilde{\theta}\|_2^2 + 2\widetilde{\mathbb{E}}_{n_2}(h^*(z_i) - \widehat{h}(z_i))^2 + 2(\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2})(h^*(z_i) - \widehat{h}(z_i))^2 \\
&= 2\|\theta^* - \widetilde{\theta}\|_2^2 + o_p(1) \\
&= o_p(1),
\end{aligned}$$

其中第三行用到了 $\|x_i\|_2 \le 1$，第四行由假设（**NUI**）、引理 18 以及关于 $h^*(z_i) - h(z_i)$的有限四阶矩条件得出；最后一行用到了公式（41）—（42c）、 $\widehat{\mathbb{E}}_{n_2}\overline{\mathbf{\Sigma}}_i^{1/2} \gtrsim \sqrt{n_2 n^{-t}} \gtrsim n_2^{1-t} \to \infty$ 这一事实（其中某个 $t < 1/2$, ），并注意到公式（42a）中的$\mathbb{E}_{n_2}\sqrt{n_2}v_i\varepsilon_i$ 具有有限方差。

A.2 定理2的证明

该定理的证明与定理 1 的证明基本相同，但使用了不同的加权向量。下面引入简写

$$w_{i1} := \langle A_{i1},\, x_i - p_i(z_i, \mathcal{F}_{i-1})\rangle \in \mathbb{R}.$$

按照类似于方程 (41) 的分解，我们有

$$(\widehat{\mathbb{E}}_{n_2} w_{i1}\langle x_i,\, u\rangle)\sqrt{n_2}(\widetilde{\theta}_u - \theta_u^*) + [\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top (\mathbf{I}_{d_T} - uu^\top)]\sqrt{n_2}(\widehat{\theta} - \theta^*)$$
$$= [\widehat{\mathbb{E}}_{n_2}\sqrt{n_2} w_{i1}\varepsilon_i - \widehat{\mathbb{E}}_{n_2}\sqrt{n_2} w_{i1}(\widehat{h}(z_i) - h^*(z_i))]. \tag{45}$$

我们分别在引理 6、7 和 8 中证明以下三个结果，它们分析了最后一个分解中的三个项。

$$\widehat{\mathbb{E}}_{n_2}\sqrt{n_2} w_{i1}\varepsilon_i \xrightarrow{d} \mathcal{N}(0, \sigma^2) \tag{46a}$$
$$\widehat{\mathbb{E}}_{n_2}\sqrt{n_2} w_{i1}(\widehat{h}(z_i) - h^*(z_i)) \xrightarrow{p} 0 \tag{46b}$$
$$\Big|\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top u - \widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big| = o_p\Big(\Big|\widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big|\Big). \tag{46c}$$

暂且假设这三个结果已知，将方程 (46a) 代入 (45)，可得 (46b)

$$(\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top u)\sqrt{n_2}(\widetilde{\theta}_u - \theta_u^*) + [\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top (\mathbf{I}_{d_T} - uu^\top)]\sqrt{n_2}(\widehat{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{47}$$

此外，利用界 (46c)，可得

$$\Big(\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top u - \widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big)\sqrt{n_2}(\widetilde{\theta}_1 - \theta_1^*) = o_p\Big(\widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\sqrt{n_2}(\widetilde{\theta}_1 - \theta_1^*)\Big) \tag{48}$$

结合方程 (47) 和 (48)，接下来只需证明

$$\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top (\mathbf{I}_{d_T} - uu^\top)\sqrt{n_2}(\widehat{\theta} - \theta^*) = o_p(1). \tag{49}$$

方程的证明 (49) 注意到 $\mathbb{E}(w_{i1} x_i^\top \mid \mathcal{F}_{i-1}) = 0$，这意味着 $\{w_{i1} x_i^\top\}_{i\geq 1}$ 构成一个鞅差序列。我们有

$$\mathbb{E}\|\sqrt{n_2}\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top\|_2^2 = \frac{1}{n_2}\sum_{i=n_1+1}^{n}\mathbb{E}\|w_{i1} x_i^\top\|_2^2 \;\leq\; \frac{1}{n_2}\sum_{i=n_1+1}^{n}\mathbb{E} w_{i1}^2 = 1,$$

其中最后一行由界 $\|x_i^\top\|_2 \leq 1$ 得出，并注意到 $\mathbb{E} w_{i1}^2 = 1$。因此，我们得出 $\sqrt{n_2}\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top = \mathcal{O}_p(1)$。结合这一事实以及假设 $(\mathbf{I}_{d_T} - uu^\top)(\widehat{\theta} - \theta^*) \xrightarrow{p} 0$，即可得到结论。

推广到连续回归变量 同样，当回归变量 x_i 取连续值时，也可以证明引理 6、7 和 8，因此定理 1 可以推广到这种情形。值得注意的是，这些引理证明中的一个关键环节是求得量 $u^\top \mathbf{\Sigma}_i^{-1} u / u^\top \mathbf{\Sigma}_i^{-2} u$ 和 $1/\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}$的下界；参见方程 (92) 和 (94)。在离散回归变量的情形下，我们利用方程 (16b) 直接计算来界定这些项；而在这里，我们明确假设它们有下界。

A.3 定理 3 的证明

为简化记号，我们仅证明干扰函数为线性的情形，即$h^*(z) = \langle z, \eta^* \rangle$ 其中 $\eta^* \in \mathcal{H} \in \mathbb{R}^{d_N}$。对于一般干扰函数 h 的证明本质上相同，只需将 $\langle z, \eta \rangle$ 替换为 $h(z)$。更多细节参见 A.3.3 节。

对于线性干扰函数 $h^*(z) = \langle z, \eta^* \rangle$，广义线性模型假设和假设（**NUI′**）可以由以下两个简化版本替代:

（a）（有界协变量和干扰项）存在 $M_\eta, M_\omega > 0$ 使得 $\sup_{\eta \in \mathcal{H}} \|\eta\|_2 \le M_\eta$ 且 $\|(x_i^\top, z_i^\top)\|_2 \le D_x$;

（b）（初步估计值的准确性）AdapTZ-GLM 的步骤 3 中的初始估计量 $\widehat{\eta} \in \mathcal{H}$ 满足 $\|\widehat{\eta} - \eta^*\|_2 = o_p(n^{-1/4})$。

A.3.1 主要论证

将方程（19）中 ϕ_i 的定义代入估计方程（22），我们得到

$$\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)\varepsilon_i \tag{50a}$$
$$= \sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)(g(x_i^\top \widetilde{\theta} + z_i^\top \widehat{\eta}) - g(x_i^\top \theta^* + z_i^\top \eta^*)). \tag{50b}$$

聚焦于前面的方程，我们现在对 g 在点（θ, $\widehat{\eta}$）处进行二阶泰勒级数展开 $\widehat{\ }$，由此得到

$$\begin{aligned}
&\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)\varepsilon_i \\
&= \sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)g'\big(\langle x_i, \widehat{\theta}\rangle + \langle z_i, \widehat{\eta}\rangle\big)(x_i - \widehat{m}_i)^\top(\widetilde{\theta} - \theta^*) \\
&\quad + \sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)\Big[Q_1 + Q_2 + Q_3 + Q_4\Big],
\end{aligned} \tag{51}$$

其中

$$\begin{aligned}
Q_1 &:= g'(\langle x_i, \widehat{\theta}\rangle + \langle z_i, \widehat{\eta}\rangle)\langle z_i, \widehat{\eta} - \eta^*\rangle \\
Q_2 &:= g'\big(\langle x_i, \widehat{\theta}\rangle + \langle z_i, \widehat{\eta}\rangle\big)\widehat{m}_i^\top(\widetilde{\theta} - \theta^*) \\
Q_3 &:= \frac{1}{2}\int_0^1\int_0^1 g''\big(\langle x_i, \widehat{\theta} + r_1 r_2(\widetilde{\theta} - \widehat{\theta})\rangle + \langle z_i, \widehat{\eta}\rangle\big)|\langle x_i, \widetilde{\theta} - \widehat{\theta}\rangle|^2 dr_1 dr_2, \quad \text{and} \\
Q_4 &:= -\frac{1}{2}\int_0^1\int_0^1 g''\big(\langle x_i, \widehat{\theta} + r_1 r_2(\theta^* - \widehat{\theta})\rangle + \langle z_i, \widehat{\eta} + r_1 r_2(\eta^* - \widehat{\eta})\rangle\big) \\
&\qquad\qquad \cdot |\langle x_i, \theta^* - \widehat{\theta}\rangle + \langle z_i, \eta^* - \widehat{\eta}\rangle|^2 dr_1 dr_2.
\end{aligned}$$

我们通过证明以下三个结果来完成证明。

$$\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)\varepsilon_i \xrightarrow{d} \mathcal{N}(0, \mathbf{I}_{d_T}) \tag{53a}$$

$\widetilde{\ }$如果 $\theta - \theta^* = o_p(1)$，那么我们有

$$\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)(Q_1 + Q_2) \xrightarrow{p} 0 \tag{53b}$$

~如果 $\theta - \theta^* = o_p(n^{-t})$，那么

$$\begin{aligned}
&\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)(Q_3 + Q_4) \\
&= o_p(1) + o_p(\|\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)g'\big(\langle x_i, \widehat{\theta}\rangle + \langle z_i, \widehat{\eta}\rangle\big)(x_i - \widehat{m}_i)^\top(\widetilde{\theta} - \theta^*)\|_2). \qquad (53c)
\end{aligned}$$

我们分别证明引理 10、11 和 12 中的结论 (53a)、(53b) 和 (53c)。我们还将马上验证

$$\widetilde{\theta} - \theta^* = o_p(n^{-t}). \qquad (54)$$

有了最后四个结果，定理 3 的证明便显而易见。事实上，记 $\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)g'\big(\langle x_i, \widehat{\theta}\rangle + \langle z_i, \widehat{\eta}\rangle\big)(x_i - \widehat{m}_i)^\top(\widetilde{\theta} - \theta^*)$ 为 Q_0，将上述结果代入 (51) 并应用斯卢茨基定理可得

$$Q_0 + o_p(|||Q_0|||_{\mathrm{op}}) \xrightarrow{d} \mathcal{N}(0, \mathbf{I}_{d_T}). \qquad (55)$$

由于 $\mathcal{N}(0, \mathbf{I}_{d_T}) = \mathcal{O}_p(1)$，我们有 $Q_0 = \mathcal{O}_p(1)$，进而 $o_p(|||Q_0|||_{\mathrm{op}}) = o_p(1)$。结合斯卢茨基定理即可得到所需结果。剩下的任务是证明一致性条件 (54)。

A.3.2 一致性条件的证明 (54)

我们对 k使用归纳论证。更具体地说，我们首先证明基本情形 $k = 1$下的 $o_p(n^{-(\delta\wedge t)})$-一致性。在归纳步骤中，假设对于某个 $k \geq 1$， $o_p(n^{-(k\delta\wedge t)})$-一致性成立，然后证明它在步骤（$k+1$）仍然成立——也就是说， $o_p(n^{-((k+1)\delta\wedge t)})$-一致性成立。

基本情形 我们首先证明 $o_p(n^{-(\delta\wedge t)})$-一致性。

^将 AdapTZ-GLM 步骤 3 中计算的估计量简记为 $\omega := (\theta, \eta)$、 $\widehat{\omega} := (\theta, \widehat{\eta})$，以及 $\omega^* := (\theta^*, \eta^*)$。

^ ~根据三角不等式以及关系$\mathbb{E}_{n_2}\phi_i(\theta, \widehat{\omega}) = 0$，我们有

$$\begin{aligned}
\|\widetilde{\mathbb{E}}_{n_2}\phi_i(\widetilde{\theta}, \omega^*)\|_2 &\leq \|(\widetilde{\mathbb{E}}_{n_2}\phi_i(\widetilde{\theta}, \omega^*) - \widetilde{\mathbb{E}}_{n_2}\phi_i(\widetilde{\theta}, \widehat{\omega})\|_2 + \|(\widetilde{\mathbb{E}}_{n_2} - \widehat{\mathbb{E}}_{n_2})\phi_i(\widetilde{\theta}, \widehat{\omega})\|_2 \\
&\leq \sup_{\theta\in\Theta}\|\widetilde{\mathbb{E}}_{n_2}\phi_i(\theta, \omega^*) - \widetilde{\mathbb{E}}_{n_2}\phi_i(\theta, \widehat{\omega})\|_2 + \sup_{\theta\in\Theta}\|(\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2})\phi_i(\theta, \widehat{\omega})\|_2, \\
&=: \mathfrak{R}_1 + \mathfrak{R}_2. \qquad (56)
\end{aligned}$$

~ ~由于$\|\mathbb{E}_{n_2}\phi_i(\theta, \omega^*) - \mathbb{E}_{n_2}\phi_i(\theta, \widehat{\omega})\|_2 \leq L_{\phi,1}\|\widehat{\omega} - \omega^*\|_2$ 对于某个常数 $L_{\phi,1} > 0$ 由引理15可得 $\mathfrak{R}_1 = o_p(n^{-1/4})$。对于 $\mathfrak{R}_2$，由引理 14 可得 $\mathfrak{R}_2 = \mathcal{O}_p(\log n/\sqrt{n})$。

结合上述结果，我们得到 $\|\widetilde{\mathbb{E}}_{n_2}\phi_i(\widetilde{\theta}, \omega^*)\|_2 = o_p(n^{-1/4})$。另一方面，

$$\begin{aligned}
&\|\widetilde{\mathbb{E}}_{n_2}\phi_i(\widetilde{\theta}, \omega^*)\|_2 \\
&= \|\widetilde{\mathbb{E}}_{n_2}(\phi_i(\widetilde{\theta}, \omega^*) - \phi_i(\theta^*, \omega^*))\|_2 \\
&\geq c_\phi\|\widetilde{\mathbb{E}}_{n_2}\partial_\theta\phi_i(\theta^*, \omega^*)(\widetilde{\theta} - \theta^*)\|_2 \wedge c_\phi n^{-1/4} \\
&\geq c_\phi\sigma_{\min}(\widetilde{\mathbb{E}}_{n_2}\partial_\theta\phi_i(\theta^*, \omega^*))\|\widetilde{\theta} - \theta^*\|_2 \wedge c_\phi n^{-1/4} \\
&\geq c_\phi m_\phi n^{\delta - t}\|\widetilde{\theta} - \theta^*\|_2 \wedge c_\phi n^{-1/4}
\end{aligned}$$

其成立的概率趋于 1。这里的不等式由定理 3 中的可辨识性假设得出。因此，

$$o_p(n^{-1/4}) = \|\widetilde{\mathbb{E}}_{n_2}\phi(\widetilde{\theta},\omega^*)\|_2 \geq c_\phi m_\phi n^{\delta-t}\|\widetilde{\theta}-\theta^*\|_2 \wedge c_\phi n^{-1/4} \tag{57}$$

~以趋于 1 的概率。由于 $t \leq 1/4$，可直接推出 $\|\theta-\theta^*\|_2 = o_p(n^{-\delta}) = o_p(n^{-(\delta\wedge t)})$。

归纳步骤 接下来，我们证明给定 $\widetilde{\theta}-\theta^* = o_p(n^{-(k\delta\wedge t)})$ 时，有 $\widetilde{\theta}-\theta^* = o_p(n^{-((k+1)\delta\wedge t)})$。事实上，只需证明对于任意 $t \in (k\delta,\ 1/4]$，$\|\widetilde{\mathbb{E}}_{n_2}\phi_i(\widetilde{\theta},\ \omega^*)\|_2 = o_p(n^{-1/4-(k\delta\wedge t)})$。这是因为将其与方程（57）给出的高概率下界 $m_\phi n^{-t}\|\theta-\theta^*\|_2$ 结合，便可直接得到所需的 $\|\widetilde{\theta}-\theta^*\|_2 = o_p(n^{-((k+1)\delta\wedge t)})$。

按照上界 (56) 中的相同步骤，并利用 $\Re_2 = \mathcal{O}_p(\log n/\sqrt{n})$ 的结果，我们得到

$$\begin{aligned}
\|\widetilde{\mathbb{E}}_{n_2}\phi_i(\widetilde{\theta},\omega^*)\|_2 &\leq \|(\widetilde{\mathbb{E}}_{n_2}\phi_i(\widetilde{\theta},\omega^*) - \widetilde{\mathbb{E}}_{n_2}\phi_i(\widetilde{\theta},\widehat{\omega})\| + \|(\widetilde{\mathbb{E}}_{n_2} - \widehat{\mathbb{E}}_{n_2})\phi_i(\widetilde{\theta},\widehat{\omega})\|_2 \\
&\leq \sup_{\theta\in\Theta,\|\theta-\theta^*\|\leq n^{-(k\delta\wedge t)}} \|\widetilde{\mathbb{E}}_{n_2}\phi_i(\theta,\omega^*) - \widetilde{\mathbb{E}}_{n_2}\phi_i(\theta,\widehat{\omega})\| \\
&\quad + \sup_{\theta\in\Theta}\|(\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2})\phi_i(\theta,\widehat{\omega})\|_2, \\
&= \sup_{\theta\in\Theta,\|\theta-\theta^*\|_2\leq n^{-(k\delta\wedge t)}} \|\widetilde{\mathbb{E}}_{n_2}\phi_i(\theta,\omega^*) - \widetilde{\mathbb{E}}_{n_2}\phi_i(\theta,\widehat{\omega})\|_2 + \mathcal{O}_p(\log n/\sqrt{n}).
\end{aligned} \tag{58}$$

记 $\{\theta\in\Theta, \|\theta-\theta^*\|_2 \leq n^{-(k\delta\wedge t)}\}$ 为 $\mathcal{C}_k$。则

$$\begin{aligned}
&\sup_{\mathcal{C}_k}\|\widetilde{\mathbb{E}}_{n_2}\phi_i(\theta,\omega^*) - \widetilde{\mathbb{E}}_{n_2}\phi_i(\theta,\widehat{\omega})\|_2 \\
&\leq \sup_{\mathcal{C}_k}\|\widetilde{\mathbb{E}}_{n_2}\partial_\omega\phi_i(\theta,\omega^*)(\widehat{\omega}-\omega^*) \\
&+ \frac{1}{2}(\widehat{\omega}-\omega^*)^\top\Big[\int_0^1\int_0^1 \widetilde{\mathbb{E}}_{n_2}\partial_\omega^2\phi_i(\theta,\omega^*+r_1r_2(\widehat{\omega}-\omega^*))dr_1dr_2\Big](\widehat{\omega}-\omega^*)\|_2 \\
&\lesssim \sup_{\mathcal{C}_k}\|[\widetilde{\mathbb{E}}_{n_2}\partial_\omega\phi_i(\theta,\omega^*) - \widetilde{\mathbb{E}}_{n_2}\partial_\omega\phi_i(\theta^*,\omega^*)](\widehat{\omega}-\omega^*)\|_2 + L_{\phi,2}\|\widehat{\omega}-\omega^*\|_2^2 \\
&\leq \sup_{\mathcal{C}_k} L_{\phi,1}\|\widetilde{\theta}-\theta^*\|_2\|\widehat{\omega}-\omega^*\|_2 + L_{\phi,2}\|\widehat{\omega}-\omega^*\|_2^2 = o_p(n^{-(k\delta\wedge t)-1/4}),
\end{aligned} \tag{59}$$

其中，$L_{\phi,\ 1}$，$L_{\phi,\ 2} > 0$ 是引理15中引入的一些常数。第二行使用了~泰勒展开；第二个不等式由 Neyman 正交性 $\mathbb{E}_{n_2}\partial_\omega\phi_i(\theta^*,\ \omega^*) = 0$ 和引理15 得出；最后一行也归因于引理15以及假设 $\widehat{\omega}-\omega^* = o_p(n^{-1/4})$ 和 $\widetilde{\theta}-\theta^* = o_p(n^{-(k\delta\wedge t)})$。由于 $(k\delta\wedge t)+1/4 < 1/2$，结合 (58) 和 (59) 可得 $\|\widetilde{\mathbb{E}}_{n_2}\phi_i(\widetilde{\theta},\ \omega^*)\|_2 = o_p(n^{-(k\delta\wedge t)-1/4})$。

A.3.3 一般干扰项函数 h的证明

我们指出，对于一般干扰函数 h，证明本质上与将 $z^\top\eta$替换为 $h(z)$ 时的证明相同。这是因为在证明中，我们只用到了关于估计 ^误差 $\|h(z) - h^*(z)\|_\infty$ 的假设，并未利用 h的具体形式。不过，通过假设 h满足线性参数化，我们可以避免使用盖托导数，从而简化得分函数的梯度的记号。

A.4 定理 4 的证明

该定理的证明与定理 3 的证明类似，我们只证明线性干扰项的情形，即 $h^*(z) = \langle z,\ \eta^* \rangle$。为简洁起见，我们仅给出证明概要。

将 ϕ_{i1} 的定义代入估计方程 (26)，得到

$$\begin{aligned}
&\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{A}_{i1}(z_i,\mathcal{F}_{i-1})(x_i-\widehat{m}_i(z_i,\mathcal{F}_{i-1}))\varepsilon_i\\
&=\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{A}_{i1}(x_i-\widehat{m}_i)\cdot\\
&\quad\Big[g\big(\langle x_i,\,u\rangle\widetilde{\theta}_u+x_i^\top(\mathbf{I}_{d_T}-uu^\top)\widehat{\theta}+\langle z_i,\,\widehat{\eta}\rangle\big)-g\big(\langle x_i,\,u\rangle\theta_u^*+x_i^\top(\mathbf{I}_{d_T}-uu^\top)\theta^*+\langle z_i,\,\eta^*\rangle\big)\Big].
\end{aligned}$$

在下文中，我们使用简写 $\theta_u := \langle u, \theta\rangle$。在最后一个方程右侧关于 $(\theta_u, \theta, \widehat{\eta})$ 处进行二阶泰勒级数展开，可得

$$\begin{aligned}
&\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{A}_{i1}(z_i,\mathcal{F}_{i-1})(x_i-\widehat{m}_i(z_i,\mathcal{F}_{i-1}))\varepsilon_i &(60)\\
&=\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{A}_{i1}(x_i-\widehat{m}_i)g'\big(\langle x_i,\,\widehat{\theta}\rangle+\langle z_i,\,\widehat{\eta}\rangle\big)(x_i-\widehat{m}_i)^\top u(\widetilde{\theta}_u-\theta_u^*)\\
&\quad+\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{A}_{i1}(x_i-\widehat{m}_i)\Big[\widetilde{Q}_1+\widetilde{Q}_2+\widetilde{Q}_3+\widetilde{Q}_4\Big], &(61)
\end{aligned}$$

其中

$$\begin{aligned}
\widetilde{Q}_1&=g'\big(\langle x_i,\,\widehat{\theta}\rangle+\langle z_i,\,\widehat{\eta}\rangle\big)[x_i^\top(\mathbf{I}_{d_T}-uu^\top)(\widehat{\theta}-\theta^*)+\langle z_i,\,\widehat{\eta}-\eta^*\rangle],\\
\widetilde{Q}_2&=g'\big(\langle x_i,\,\widehat{\theta}\rangle+\langle z_i,\,\widehat{\eta}\rangle\big)\langle\widehat{m}_i,\,u\rangle(\widetilde{\theta}_u-\theta_u^*)\\
\widetilde{Q}_3&=\frac{1}{2}\int_0^1\int_0^1 g''\big(\langle x_i,\,u\rangle(\widehat{\theta}_u+r_1r_2(\widetilde{\theta}_u-\widehat{\theta}_u))+x_i^\top(\mathbf{I}_{d_T}-uu^\top)\widehat{\theta}+\langle z_i,\,\widehat{\eta}\rangle\big)\\
&\qquad\qquad\cdot\big|\langle x_i,\,u\rangle(\widetilde{\theta}_u-\widehat{\theta}_u)\big|^2dr_1dr_2\\
\widetilde{Q}_4&=-\frac{1}{2}\int_0^1\int_0^1 g''\Big(\langle x_i,\,u\rangle(\widehat{\theta}_u+r_1r_2(\theta_u^*-\widehat{\theta}_u))+x_i^\top(\mathbf{I}_{d_T}-uu^\top)(\widehat{\theta}+r_1r_2(\theta^*-\widehat{\theta}))\\
&\qquad\qquad+\langle z_i,\,\widehat{\eta}+r_1r_2(\eta^*-\widehat{\eta})\rangle\Big)\cdot|\langle x_i,\,\theta^*-\widehat{\theta}\rangle+\langle z_i,\,\eta^*-\widehat{\eta}\rangle|^2dr_1dr_2.
\end{aligned}$$

按照与引理 10、11 和 12 类似的论证，可以证明 at

$$\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{A}_{i1}(z_i,\mathcal{F}_{i-1})(x_i-\widehat{m}_i(z_i,\mathcal{F}_{i-1}))\varepsilon_i\xrightarrow{d}\mathcal{N}(0,1), \tag{62a}$$

$$\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{A}_{i1}(x_i-\widehat{m}_i)(\widetilde{Q}_1+\widetilde{Q}_2)\xrightarrow{p}0 \tag{62b}$$

$$\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{A}_{i1}(x_i-\widehat{m}_i)(\widetilde{Q}_3+\widetilde{Q}_4)\xrightarrow{p}0 \tag{62c}$$

这里，结论 (62c) 需要假设 $\theta_u - \theta_u^* = o_p(n^{-t})$。我们可以按照与定理 3 的证明类似的论证，证明这一 $o_p(n^{-t})$-一致性。具体而言，除了将 ϕ_i 替换为 ϕ_{i1} 并证明 $\sim$ $\sim$ $\mathbb{E}_{n_2}\phi_{i1}, \mathbb{E}_{n_2}\partial_\omega\phi_{i1}$ 在 (θ_u, ω) 上有界且利普希茨连续之外，证明基本相同。定理 4 的证明至此完成。

B 推论的证明

我们对推论的证明依赖于以下技术引理。在陈述该引理时，我们使用简写记号 $Q_i := (x_i^\top, z_i^\top)^\top$，并定义 $\mathbf{Q} \equiv \sum_{i=1}^{n_1} Q_i Q_i^\top$。

引理 1. 假设对某个 $t \in [0, 1/4)$，假设 **(NOI**(ν, σ^2)**)–(NUI)** 成立，并且

$$\|z_i\|_2 \le B \quad and \quad \inf_{P\in\mathcal{P}} \sigma_{\min}(\mathbb{E}_{z_i\sim P} z_i z_i^\top) \ge c_P \qquad for\ some\ B < \infty\ and\ c_P > 0.$$

那么，存在某个常数 $c_{\mathbf{Q}} > 0$，使得最小特征值满足

$$\lim_{n\to\infty} \mathbb{P}(\sigma_{\min}(\mathbf{Q}) \ge c_{\mathbf{Q}} n_1^{1-2t}) \to 1. \tag{63}$$

我们将在 B.5 节返回证明引理 1。这里，我们利用引理 1 完成各个推论的证明。

B.1 推论 1 的证明

根据定理 1，只需证明 $\|\widehat{\eta}_{ols} - \eta^*\|_2 = o_p(1)$。为此，我们采用 Lai 和 Wei [38] 的结果。具体而言，记向量 $(x_i^\top,\ z_i^\top)^\top$ 为Q_i，并令 $\mathbf{Q} \equiv \sum_{i=1}^{n_1} Q_i Q_i^\top$。根据 Lai 和 Wei [38] 的定理 1，只需证明 $(\log \sigma_{\max}(\mathbf{Q})/\sigma_{\min}(\mathbf{Q}))^{1/2} = o_p(1)$。由于向量 x_i 和 z_i 在 ℓ_2-范数下都有界，因此 $\log\ \sigma_{\max}(\mathbf{Q}) = \mathcal{O}(\log\ n)$。由此，引理 1 保证，对于任意 $t \in [0, 1/2)$，都有$1/\sigma_{\min}(\mathbf{Q}) = \mathcal{O}_p(n^{2t-1}) = o_p(n^{-\varepsilon})$，其中某个较小的 $\varepsilon > 0$。综合以上结果，我们得出 $\|\widehat{\eta}_{ols} - \eta^*\|_2 = o_p(1)$，这正是推论 1 所述。

B.2 推论 2 的证明

根据定理 1，我们只需证明 $\|\widehat{\eta}_{lasso} - \eta^*\|_2 = o_p(1)$。为此，我们利用 Oh 等人 [48] 的结果。定义指标集

$$S := \{1, 2, \ldots, d_T\} \cup \{i + d_T \mid \eta_i^* \ne 0\},$$

并引入简写记号 $Q_i := (x_i^\top,\ z_i^\top)^\top$，以及

$$\mathbf{Q} := \sum_{i=1}^{n_1} Q_i Q_i^\top, \beta^* := (\theta^{*\top},\qquad \eta^{*\top})^\top, \quad \text{and} \quad \widehat{\beta}_{lasso} := (\widehat{\theta}_{lasso}^\top, \widehat{\eta}_{lasso}^\top)^\top.$$

对于任意向量 $\beta \in \mathbb{R}^{d_T+d_N}$，我们定义向量 β_S，其第 j 个分量为 $\beta_j.\ _S := \beta_j 1_{j\in S}$。利用引理 1 可得

$$\|\beta_S\|_1^2/n_1^{2t} \le |S|\|\beta_S\|_2^2/n_1^{2t} \lesssim \frac{|S|}{n_1} \cdot \beta^\top \mathbf{Q}\beta$$

对所有 β均成立。因此，论文 [48] 中的假设 3 所要求的兼容性条件以趋于 1 的概率成立，其中 $\phi_{n_1}^2 = c n_1^{-2t}$，且某个常数为 $c > 0$。于是，我们可以应用论文 [48] 中的引理 1（注意，当 B' 为 $\|z_i\|_\infty$ 的上界而非 $\|z_i\|_2$时，该引理仍然成立）来断言

$$\begin{aligned}\|\widehat{\eta}_{lasso} - \eta^*\|_2 \le \|\widehat{\beta}_{lasso} - \beta^*\|_1 &\le \frac{4(s + d_T)\lambda_{n_1}}{\phi_{n_1}^2} \\ &= \frac{8(s + d_T)n_1^{2t}\nu(B' + 1)}{c}\sqrt{\frac{2[\log(2/\delta_{n_1}) + \log(d_T + d_N)]}{n_1}} \\ &\lesssim (s + d_T)n_1^{2t-1/2}\sqrt{\log(2/\delta_{n_1}) + \log(d_T + d_N)}\end{aligned}$$

以概率 $1-\delta_{n_1}-\mathbb{P}(\sigma_{\min}(\mathbf{Q}) < cn_1^{1-2t})$。代入 $\delta_{n_1}=\min\{(s+d_T)n_1^{2t-1/2}, 1/(d_T+d_N)\}$ 和 $(s+d_T)\sqrt{\log(d_T+d_N)} = o_p(n_1^{1/2-2t})$，并注意到 d_T 是固定的，可得

$$(s+d_T)n_1^{2t-1/2}\sqrt{\log(2/\delta_{n_1})+\log(d_T+d_N)} = o_p(1).$$

根据我们对 δ_{n_1} 的选择以及引理 1，可知 $1-\delta_{n_1}-\mathbb{P}(\sigma_{\min}(\mathbf{Q}) < cn_1^{1-2t}) \to 1$。因此，我们得出 $\|\widehat{\eta}_{lasso}-\eta^*\|_2 = o_p(1)$，这就完成了推论 2 的证明。

B.3 推论 3 的证明

证明与推论 2 的证明基本相同。回顾推论 2 证明中的记号。应用引理 1 可得

$$\|\beta_S\|_1^2/n_1^{2t} \le |S|\|\beta_S\|_2^2/n_1^{2t} \lesssim \frac{|S|}{n_1}\cdot\beta^\top\mathbf{Q}\beta$$

对所有 β 均成立。因此，Oh 等人 [48] 中的兼容性条件以 $\phi_{n_1}^2 = cn_1^{-2t}$ 得到满足，其中某个常数 $c>0$ 以概率趋近于 1。应用 Oh 等人 [48] 中的引理 1，我们得到

$$\begin{aligned}\max\left\{\|\widehat{\theta}_{lasso}-\theta^*\|_2, \|\widehat{\eta}_{lasso}-\eta^*\|_2\right\} \le \|\widehat{\beta}_{lasso}-\beta^*\|_1 &\le \frac{4(s+d_T)\lambda_{n_1}}{l_g\phi_{n_1}^2}\\ &= \frac{8(s+d_T)n_1^{2t}\nu D_x}{c}\sqrt{\frac{2[\log(2/\delta_{n_1})+\log(d_T+d_N)]}{n_1}}\\ &\lesssim (s+d_T)n_1^{2t-1/2}\sqrt{\log(2/\delta_{n_1})+\log(d_T+d_N)}\end{aligned}$$

其成立的概率至少为 $1-\delta_{n_1}-\mathbb{P}(\sigma_{\min}(\mathbf{Q}) < cn_1^{1-2t})$。代入 $\delta_{n_1}:=\min\{(s+d_T)n_1^{2t-1/4}, 1/(d_T+d_N)\}$ 和 $(s+d_T)\sqrt{\log(d_T+d_N)} = o_p(n_1^{1/4-2t})$，并注意到 d_T 是固定的，我们得到

$$(s+d_T)n_1^{2t-1/2}\sqrt{\log(2/\delta_{n_1})+\log(d_T+d_N)} = o_p(n_1^{-1/4}).$$

根据我们对 δ_{n_1} 的选择和引理 1，可得 $1-\delta_{n_1}-\mathbb{P}(\sigma_{\min}(\mathbf{Q}) < cn_1^{1-2t}) \to 1$。综合各部分，我们得出 $\max\left\{\|\widehat{\theta}_{lasso}-\theta^*\|_2, \|\widehat{\eta}_{lasso}-\eta^*\|_2\right\} = o_p(n_1^{-1/4})$；推论 3 的证明至此完成。

B.4 推论 4 的证明

根据定理 1，只需证明

$$\mathbb{E}_{z_i\sim P}(\widehat{h}(z_i)-h^*(z_i))^2 = o_p(1). \tag{64}$$

由于 $p_{i0} \ge c_0 i^{-2t}$（对于某个 $t<1/2$）且 $n_1 = n/K$（对于某个常数 $K\ge 2$），利用 Freedman 不等式（例如，见 [2]中的引理 9），我们有

$$\sum_{i=1}^{n_1}\mathbf{1}_{\{x_i=\mathbf{0}\}} \ge \sum_{i=1}^{n_1} p_{i0}/2 \ge cn^{1-2t} \tag{65}$$

其概率随着 $n \to \infty$ 趋于 1（对于某个常数 $c > 0$）。由于我们假设选择概率 p_i 仅取决于 $\mathcal{F}_{i-1}$，且 z_i 为独立同分布样本，令 $(\widetilde{z}_1,\ \widetilde{x}_1,\ \widetilde{y}_1)$.....$(\widetilde{z}_{cn^{1-2t}},\ \widetilde{x}_{cn^{1-2t}},\ \widetilde{y}_{cn^{1-2t}})$ 表示前 n_1 个样本中满足相应回归变量 $x_i = \mathbf{0}$[4]的前 cn^{1-2t} 个样本，可以通过归纳法验证，$\{\widetilde{z}_i\}_{i=1}^{cn^{1-2t}}$ 是来自 P的独立同分布样本，并且

$$\widetilde{y}_i = h^*(\widetilde{z}_i) + \widetilde{\varepsilon}_i$$

对于某些独立同分布噪声 $\widetilde{\varepsilon}_i \sim Q$。因此，如 [24], 的定理 6.2 所示，基于样本 $\{(\widetilde{z}_i,\ \widetilde{y}_i)\}_{i=1}^{cn^{1-2t}}$ 的 k-近邻估计量 $\widehat{h}$ 满足 $k \to \infty,\ k/n^{1-2t} \to 0\, \mathbb{E}_{z\sim P}(\widehat{h}(z_i) - h^*(z_i))^2 = o_p(1)$。公式（64）立即成立，因为如公式（65）所示，我们可以以趋近于 1 的概率在前 n_1 个观测样本中找到这样的独立同分布样本。

B.5 引理 1 的证明

只需证明

$$\lim_{n_1\to\infty} \mathbb{P}(\sigma_{\max}(\mathbf{Q}^{-1}) \le n_1^{2t-1}/c_{\mathbf{Q}}) \to 1 \qquad \text{for some constant } c_{\mathbf{Q}}.$$

Usin利用 Sherman-Woodbury公式处理分块矩阵逆，我们有 e

$$\mathbf{Q}^{-1} \equiv \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \\ \mathbf{Q}_3 & \mathbf{Q}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{d_T} & 0 \\ -\mathbf{Q}_4^{-1}\mathbf{Q}_3 & \mathbf{I}_{d_N} \end{bmatrix} \begin{bmatrix} \left(\mathbf{Q}_1 - \mathbf{Q}_2\mathbf{Q}_4^{-1}\mathbf{Q}_3\right)^{-1} & 0 \\ 0 & \mathbf{Q}_4^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{d_T} & -\mathbf{Q}_2\mathbf{Q}_4^{-1} \\ 0 & \mathbf{I}_{d_N} \end{bmatrix},$$

其中 $\mathbf{Q}_1 = \sum_{i=1}^{n_1} x_i x_i^\top$，$\mathbf{Q}_2 = \mathbf{Q}_3^\top = \sum_{i=1}^{n_1} x_i z_i^\top$ 和 $\mathbf{Q}_4 = \sum_{i=1}^{n_1} z_i z_i^\top$。由于向量 $z_i's$ 是具有有界二阶矩的独立同分布向量，由 z_i 的有界性和引理18 可知，$\|\mathbf{Q}_4/n_1 - \widetilde{\mathbb{E}}_{n_1} z_i z_i^\top\|_{\text{op}} = \mathcal{O}_p(n_1^{-1/2})$。将这一事实与下界结合

$$\sigma_{\min}(\widetilde{\mathbb{E}}_{n_1} z_i z_i^\top) \ge \inf_{P\in\mathcal{P}} \sigma_{\min}(\mathbb{E}_{z_i\sim\mathcal{P}} z_i z_i^\top) \ge c_p > 0,$$

我们得到 $\lim_{n_1\to\infty} \mathbb{P}(\|\mathbf{Q}_4/n_1\|_{\text{op}} \ge c_P/2) \to 1$，因此 $\|\mathbf{Q}_4^{-1}\|_{\text{op}} = \mathcal{O}_p(n^{-1})$。此外，由于 x_i 和 z_i 有界，可得 $\|\mathbf{Q}_3\|_{\text{op}}$，$\|\mathbf{Q}_2\|_{\text{op}} = \mathcal{O}_p(n)$。结合上述结果，我们得到 $\|\mathbf{Q}_4^{-1}\mathbf{Q}_3\|_{\text{op}} = \|(\mathbf{Q}_2\mathbf{Q}_4^{-1})^\top\|_{\text{op}} = \mathcal{O}_p(1)$，从而

$$\left\| \begin{bmatrix} \mathbf{I}_{d_T} & 0 \\ -\mathbf{Q}_4^{-1}\mathbf{Q}_3 & \mathbf{I}_{d_N} \end{bmatrix} \right\|_{\text{op}}, \left\| \begin{bmatrix} \mathbf{I}_{d_T} & -\mathbf{Q}_2\mathbf{Q}_4^{-1} \\ 0 & \mathbf{I}_{d_N} \end{bmatrix} \right\|_{\text{op}} = \mathcal{O}_p(1).$$

现在，由于谱范数的次乘性，且 $\|\mathbf{Q}_4^{-1}\|_{\text{op}} = \mathcal{O}_p(n^{-1}) = o_p(n^{2t-1})$，只需证明 $\lim_{n\to\infty} \mathbb{P}(\|\Big(\mathbf{Q}_1 - \mathbf{Q}_2\mathbf{Q}_4^{-1}\mathbf{Q}_3\Big)^{-1}\|_{\text{op}} \lesssim n^{2t-1}) \to 1$，或者等价地，证明 $\lim_{n\to\infty} \mathbb{P}(\sigma_{\min}\Big(\mathbf{Q}_1 - \mathbf{Q}_2\mathbf{Q}_4^{-1}\mathbf{Q}_3\Big) \gtrsim n^{1-2t}) \to 1$。

对于 $\mathbf{Q}_1$,，注意到根据 x_i的独热性质，我们有 $x_i x_i^\top - \text{diag}\{p_{i1}, \ldots, p_{id_T}\}$ 构成一个矩阵值鞅差序列。由于 $x_i,\ p_i$ 有界，由引理18可知 $\|\mathbf{Q}_1 - \text{diag}\{\sum_{i=1}^{n_1} p_{i1}, \ldots, \sum_{i=1}^{n_1} p_{id_T}\}\|_F = \mathcal{O}_p(n_1^{1/2})$。

稍微滥用一下记号，我们将 $(x_1, \ldots, x_{n_1})^\top$ 记为 $\mathbf{X}$，将 $(z_1, \ldots, z_{n_1})^\top$ 记为 $\mathbf{Z}$，并将 $(p_1, \ldots, p_{n_1})^\top$ 记为 $\mathbf{P}$。那么 $\mathbf{Q}_2\mathbf{Q}_4^{-1}\mathbf{Q}_3 = \mathbf{X}^\top\mathbf{H}\mathbf{X}$，其中投影矩阵 $\mathbf{H} := \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top$。类似地，$(x_i - p_i)z_i^\top$构成一个矩阵值鞅差序列，并且 $\mathbb{E}\|(x_i - p_i)z_i^\top\|_F^2$

[4]如果此类样本少于 cn^{1-2t} 个，我们会生成额外的独立样本。

有界。由引理18可得 $\|\|\|(\mathbf{X}\ -\mathbf{P})^\top\mathbf{Z}\|\|_F = \mathcal{O}_p(n_1^{1/2})$。将其代入 $\mathbf{X}^\top\mathbf{H}\mathbf{X}$，得到

$$
\begin{aligned}
&\|\|\mathbf{X}^\top\mathbf{H}\mathbf{X} - \mathbf{P}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{P}\|\|_{\mathrm{op}} \\
&= \|\|\mathbf{X}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{X} - \mathbf{P}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{P}\|\|_{\mathrm{op}} \\
&= \|\|\mathbf{P}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top(\mathbf{X}-\mathbf{P}) + (\mathbf{X}-\mathbf{P})^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{P} + (\mathbf{X}-\mathbf{P})^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top(\mathbf{X}-\mathbf{P})\|\|_{\mathrm{op}} \\
&\le \|\|\mathbf{P}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top(\mathbf{X}-\mathbf{P})\|\|_{\mathrm{op}} + \|\|(\mathbf{X}-\mathbf{P})^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{P}\|\|_{\mathrm{op}} \\
&\quad + \|\|(\mathbf{X}-\mathbf{P})^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top(\mathbf{X}-\mathbf{P})\|\|_{\mathrm{op}} \\
&= \mathcal{O}_p\Big(n_1^{1/2}\Big),
\end{aligned}
$$

其中，最后一行使用了以下关系

$$
\begin{aligned}
\|\|(\mathbf{Z}^\top\mathbf{Z})^{-1}\|\|_{\mathrm{op}} = \|\|\mathbf{Q}_4^{-1}\|\|_{\mathrm{op}} = \mathcal{O}_p(n^{-1}), \quad \|\|\mathbf{Z}^\top\mathbf{P}\|\|_{\mathrm{op}} = \mathcal{O}(n), \text{ and} \\
\|\|(\mathbf{X}-\mathbf{P})^\top\mathbf{Z}\|\|_F = \mathcal{O}_p(n_1^{1/2}).
\end{aligned}
$$

综合各部分可得

$$
\begin{aligned}
\|\|\mathbf{Q}_1 - \mathbf{Q}_2\mathbf{Q}_4^{-1}\mathbf{Q}_3\|\|_{\mathrm{op}} &= \|\|\mathbf{X}^\top(\mathbf{I}-\mathbf{H})\mathbf{X}\|\|_{\mathrm{op}} \\
&= \|\|\operatorname{diag}\Big\{\sum_{i=1}^{n_1} p_{i1}, \ldots, \sum_{i=1}^{n_1} p_{id_T}\Big\} - \mathbf{P}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{P}\|\|_{\mathrm{op}} + \mathcal{O}_p(n^{1/2}) \\
&\ge \|\|\operatorname{diag}\Big\{\sum_{i=1}^{n_1} p_{i1}, \ldots, \sum_{i=1}^{n_1} p_{id_T}\Big\} - \mathbf{P}^\top\mathbf{P}\|\|_{\mathrm{op}} + \mathcal{O}_p(n^{1/2}) \\
&= \|\|\sum_{i=1}^{n_1}\mathbf{\Sigma}_i\|\|_{\mathrm{op}} + \mathcal{O}_p(n^{1/2}) \ge c_0 n^{1-2t} + \mathcal{O}_p(n^{1/2}),
\end{aligned}
$$

其中，第一个不等式利用了界 $\|\|\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\|\|_{\mathrm{op}} \le 1$ 以及正定性；最后一行由假设（**SEL**(t)）得出。由于根据假设有 $t \in (0, 1/4)$，因此 $1-2t > 1/2$，证明至此完成。

C Neyman 正交性及其他假设

本节验证我们构造的得分函数所满足的若干条件，包括 Neyman 正交性，以及关于逻辑模型的假设（**EIG**）、（**EIG***）、（**IDE**）和（**IDE***）。

C.1 线性模型

回顾式 (9) 中 ϕ_i 的定义，我们有

$$
\mathbb{E}(\phi_i(\theta^*, h^*) \mid \mathcal{F}_{i-1}) = \mathbb{E}_{x_i, z_i, \varepsilon_i}[\mathbf{\Sigma}_i^{-1/2}(x_i - p_i(z_i, \mathcal{F}_{i-1}))\varepsilon_i \mid \mathcal{F}_{i-1}] = 0, \tag{66a}
$$

其中第二个等号利用了以下事实：$\mathbb{E}(\varepsilon_i \mid x_i,\ z_i,\ \mathcal{F}_{i-1}) = 0$。接下来注意到

$$
\begin{aligned}
&\mathbb{E}(\partial_h\phi_i(\theta^*, h^*)[\bar{h} - h^*] \mid \mathcal{F}_{i-1}) \\
&= \mathbb{E}_{x_i, z_i}[-\mathbf{\Sigma}_i^{-1/2}(x_i - p_i(z_i, \mathcal{F}_{i-1}))(\bar{h}(z_i) - h^*(z_i)) \mid \mathcal{F}_{i-1}] \\
&= \mathbb{E}_{z_i}\mathbb{E}_{x_i}[-\mathbf{\Sigma}_i^{-1/2}(x_i - p_i(z_i, \mathcal{F}_{i-1})) \mid \mathcal{F}_{i-1}, z_i](\bar{h}(z_i) - h^*(z_i)) = 0,
\end{aligned} \tag{66b}
$$

最后一行由 $\mathbb{E}[x_i - p_i(z_i, \mathcal{F}_{i-1}) \mid \mathcal{F}_{i-1}, z_i] = 0$ 得出。这验证了条件 (10b)。

C.2 广义线性模型

回顾由方程 (19) 得到的得分函数 ϕ_i 定义，我们有

$$\mathbb{E}(\phi_i(\theta^*, \theta^*, h^*) \mid \mathcal{F}_{i-1}) = \mathbb{E}_{x_i, z_i, \varepsilon_i}[\mathbf{\Omega}_i^*(x_i - m_i^*(z_i, \mathcal{F}_{i-1}))\varepsilon_i \mid \mathcal{F}_{i-1}] = 0 \tag{67}$$

我们将 $\phi_i(\theta, h^*, m_i^*, \mathbf{\Omega}_i^*) = \phi_i(\theta, \theta^*, h^*)$ 记作 ϕ_i 对 $\mathbf{\Omega}_i$ 和 m_i 的显式依赖关系。为验证梯度条件 (21b) 和 (21c)，我们首先计算 ϕ_i 关于 $\mathbf{\Omega}_i, m_i$ 和 h 的偏导数

。具体而言，对于任意 $\mathbf{\Omega}_i = \mathbf{\Omega}_i(z_i, \mathcal{F}_{i-1})$，

$$\begin{aligned}
&\mathbb{E}(\partial_{\mathbf{\Omega}_i}\widetilde{\phi}_i(\theta^*, h^*, m_i^*, \mathbf{\Omega}_i^*)[\bar{\mathbf{\Omega}}_i - \mathbf{\Omega}_i^*] \mid z_i, \mathcal{F}_{i-1}) \\
&= \mathbb{E}_{x_i, z_i}[(\bar{\mathbf{\Omega}}_i - \mathbf{\Omega}_i^*)(x_i - m_i^*(z_i, \mathcal{F}_{i-1}))(y_i - g\big(\langle x_i, \theta^*\rangle + h^*(z_i)\big)) \mid z_i, \mathcal{F}_{i-1}] \\
&= \mathbb{E}_{x_i, z_i}[(\bar{\mathbf{\Omega}}_i - \mathbf{\Omega}_i^*)(x_i - m_i^*(z_i, \mathcal{F}_{i-1}))\varepsilon_i \mid z_i, \mathcal{F}_{i-1}] = 0.
\end{aligned} \tag{68}$$

同样，对于任意 $\bar{m}_i = \bar{m}_i(z_i, \mathcal{F}_{i-1})$，

$$\begin{aligned}
&\mathbb{E}(\partial_{m_i}\widetilde{\phi}_i(\theta^*, h^*, m_i^*, \mathbf{\Omega}_i^*)[\bar{m}_i - m_i^*] \mid z_i, \mathcal{F}_{i-1}) \\
&= \mathbb{E}_{x_i, z_i}[-\mathbf{\Omega}_i^*(\bar{m}_i - m_i^*)(y_i - g\big(\langle x_i, \theta^*\rangle + h^*(z_i)\big)) \mid z_i, \mathcal{F}_{i-1}] \\
&= \mathbb{E}_{x_i, z_i}[-\mathbf{\Omega}_i^*(\bar{m}_i - m_i^*)\varepsilon_i \mid z_i, \mathcal{F}_{i-1}] = 0.
\end{aligned} \tag{69}$$

此外，固定 $\mathbf{\Omega}_i, m_i$，对于任意 $h = h(z_i)$

$$\begin{aligned}
&\mathbb{E}(\partial_h\widetilde{\phi}_i(\theta^*, h^*, m_i^*, \mathbf{\Omega}_i^*)[\bar{h} - h^*] \mid z_i, \mathcal{F}_{i-1}) \\
&= \mathbb{E}[\mathbf{\Omega}_i^*(x_i - m_i^*(z_i, \mathcal{F}_{i-1}))g'\big(\langle x_i, \theta^*\rangle + h^*(z_i)\big)[\bar{h} - h^*] \mid z_i, \mathcal{F}_{i-1}] \\
&= \mathbf{\Omega}_i^*\mathbb{E}[(x_i - m_i^*(z_i, \mathcal{F}_{i-1}))g'\big(\langle x_i, \theta^*\rangle + h^*(z_i)\big) \mid \mathcal{F}_{i-1}, z_i](\bar{h}(z_i) - h^*(z_i)) \\
&= \mathbf{\Omega}_i^*\mathbb{E}_{z_i}(0 \mid z_i, \mathcal{F}_{i-1})(\bar{h}(z_i) - h^*(z_i)) = 0.
\end{aligned} \tag{70}$$

将各部分结合起来并应用链式法则，我们得到

$$\begin{aligned}
&\mathbb{E}(\partial_\theta\widetilde{\phi}_i(\theta^*, \theta^*, h^*) \mid \mathcal{F}_{i-1}) \\
&= \mathbb{E}(\mathbb{E}(\partial_{\mathbf{\Omega}_i}\widetilde{\phi}_i\partial_\theta\mathbf{\Omega}_i + \partial_{m_i}\widetilde{\phi}_i\partial_\theta m_i + \partial_h\widetilde{\phi}_i\partial_\theta h \mid z_i, \mathcal{F}_{i-1}) \mid \mathcal{F}_{i-1}) \\
&= \mathbb{E}(\mathbb{E}(\partial_{\mathbf{\Omega}_i}\widetilde{\phi}_i \mid z_i, \mathcal{F}_{i-1})\partial_\theta\mathbf{\Omega}_i + \mathbb{E}(\partial_{m_i}\widetilde{\phi}_i \mid z_i, \mathcal{F}_{i-1})\partial_\theta m_i + \mathbb{E}(\partial_h\widetilde{\phi}_i \mid z_i, \mathcal{F}_{i-1})\partial_\theta h \mid \mathcal{F}_{i-1}) \\
&= 0.
\end{aligned} \tag{71}$$

同样，对于任意 $h = h(z_i)$，我们有

$$\mathbb{E}(\partial_h\phi_i(\theta^*, \theta^*, h^*)[\bar{h} - h^*] \mid \mathcal{F}_{i-1}) = 0. \tag{72}$$

因此，我们得出结论：$\phi_i(\theta, \theta, h^*)$ 是在 $(\theta^*, \theta^*, h^*)$ 处、干扰项为 (θ, h) 的 Neyman 正交得分函数。

C.3 关于逻辑回归假设的评论

本节将说明，在逻辑回归的设定下，假设 **(EIG)**、**(EIG*)** 和 **(IDE*)** 均成立。此外，在特殊情形 $d_T = 1$ 下，假设 **(IDE)** 也成立。

我们首先验证假设（**EIG**）。对于逻辑回归，逆链接函数为 $g(x) = e^x/(1+e^x)$，且有 $g'(x) = e^x/(1+e^x)^2 = \nu^2(x)$。因此，根据 $\mathbf{\Omega}_i^*$ 的定义，我们有

$$\begin{aligned}
&\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^*(x_i - m_i^*)g'\big(\langle x_i, \theta^*\rangle + h^*(z_i)\big)(x_i - m_i^*)^\top \\
&= \widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^*(x_i - m_i^*)\nu^2\big(\langle x_i, \theta^*\rangle + h^*(z_i)\big)(x_i - m_i^*)^\top = \widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*-1} \succeq cn^{\delta - t}\mathbf{I}_{d_T}
\end{aligned}$$

对于某个 $c>0$，其中最后一个不等式由引理 9 得出。令 $m_{\phi,2}=c$，我们看到，关于最小奇异值的假设（**EIG**）成立。

类似地，对于假设（**EIG***），根据 A_{i1}^* 的定义可得

$$\begin{aligned}&\widetilde{\mathbb{E}}_{n_2}A_{i1}^*(x_i-m_i^*)g'\big(\langle x_i,\,\theta^*\rangle+h^*(z_i)\big)(x_i-m_i^*)^\top u\\&=\widetilde{\mathbb{E}}_{n_2}A_{i1}^*(x_i-m_i^*)\nu^2\big(\langle x_i,\,\theta^*\rangle+h^*(z_i)\big)(x_i-m_i^*)^\top u=\widetilde{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top\mathbf{\Omega}_i^{*,2}u}}\geq cn^{\delta-t}\end{aligned}$$

对于某个 $c>0$，其中，最后一个不等式由方程（108）中$\mathbf{\Sigma}_i^{-1}$的显式公式以及假设（$\mathbf{SEL}'^*(t,\delta,S_u)$得出。取$m_{\phi,2}=c$ 即可得到假设（**EIG***）。

为验证假设（**IDE**），我们首先断言

$$\widetilde{\mathbb{E}}_{n_2}(\phi_i(\theta,\theta^*,h^*)-\phi_i(\theta^*,\theta^*,h^*))=[\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\widetilde{\mathbf{B}}_i](\theta^*-\theta),\tag{73}$$

$\sim$ $\sim$其中$\mathbf{B}_i$ 是满足$c_{\mathbf{B},1}\mathbf{I}_{d_T}\preceq\mathbf{B}_i\preceq c_{\mathbf{B},2}\mathbf{I}_{d_T}$的某些对角矩阵，且存在常数$c_{\mathbf{B},1},c_{\mathbf{B},2}>0$，它们可能取决于问题参数。我们将在证明的最后回过头来证明这一断言。另一方面，我们有

$$\begin{aligned}\widetilde{\mathbb{E}}_{n_2}\partial_\theta\phi_i(\theta^*,\theta^*,h^*)(\theta^*-\theta)&=\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^*(x_i-m_i^*)g'\big(\langle x_i,\,\theta^*\rangle+h^*(z_i)\big)x_i^\top(\theta^*-\theta)\\&=\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^*(x_i-m_i^*)g'\big(\langle x_i,\,\theta^*\rangle+h^*(z_i)\big)(x_i-m_i^*)^\top(\theta^*-\theta)\\&=[\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}](\theta^*-\theta).\end{aligned}$$

因此，接下来只需证明

$$\|\|[\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}][\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\widetilde{\mathbf{B}}_i]^{-1}\|\|_{\mathrm{op}}=O_p(1).\tag{74}$$

当 $d_T=1$时，由于$\mathbf{\Omega}_i^{*\,-1}>0$,，我们有

$$\begin{aligned}\|\|[\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}][\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\widetilde{\mathbf{B}}_i]^{-1}\|\|_{\mathrm{op}}&=|[\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}]/[\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\widetilde{\mathbf{B}}_i]|\leq|[\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}]/[\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\cdot c_{\mathbf{B},1}]|\\&=1/c_{\mathbf{B},1}=O_p(1).\end{aligned}$$

因此，选择 $c_\phi=c_{\mathbf{B},1}$ 后，我们验证了带有$d_T=1$的逻辑回归模型满足假设（**IDE**）。

最后，我们验证假设（**IDE***）。通过类似的计算，我们得到

$$\widetilde{\mathbb{E}}_{n_2}(\phi_{i1}(\theta_u,\theta^*,h^*)-\phi_{i1}(\theta_u^*,\theta^*,h^*))=\Big[\widetilde{\mathbb{E}}_{n_2}\frac{u^\top\widetilde{\mathbf{B}}_iu}{\sqrt{u^\top\mathbf{\Omega}_i^{*,2}u}}\Big](\theta_u^*-\theta_u),$$

$\sim$ $\sim$其中 $\mathbf{B}_i$ 是满足 $c'_{\mathbf{B},1}\mathbf{I}_{d_T}\preceq\mathbf{B}^i\preceq c'_{\mathbf{B},2}\mathbf{I}_{d_T}$的对角矩阵；对于某些常数 $c_{\mathbf{B},1},c_{\mathbf{B},2}>0$，它们可能取决于问题参数。此外，我们有

$$\widetilde{\mathbb{E}}_{n_2}\partial_{\theta_u}\phi_{i1}(\theta_u^*,\theta^*,h^*)(\theta_u-\theta_u^*)=\Big[\widetilde{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top\mathbf{\Omega}_i^{*,2}u}}\Big](\theta_u^*-\theta_u).$$

由于

$$\Big|\Big[\widetilde{\mathbb{E}}_{n_2}\frac{u^\top\widetilde{\mathbf{B}}_iu}{\sqrt{u^\top\mathbf{\Omega}_i^{*,2}u}}\Big](\theta_u^*-\theta_u)\Big|\geq\Big|\Big[\widetilde{\mathbb{E}}_{n_2}\frac{c'_{\mathbf{B},1}\|u\|_2^2}{\sqrt{u^\top\mathbf{\Omega}_i^{*,2}u}}\Big](\theta_u^*-\theta_u)\Big|=c'_{\mathbf{B},1}\Big|\Big[\widetilde{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top\mathbf{\Omega}_i^{*,2}u}}\Big](\theta_u^*-\theta_u)\Big|,$$

假设（**IDE***）通过选择 $c_\phi=c'_{\mathbf{B},\ 1}$ 即可直接得出。

命题的证明 (73) 注意，我们有

$$\begin{aligned}
&\widetilde{\mathbb{E}}_{n_2}(\phi_i(\theta,\theta^*,h^*)-\phi_i(\theta^*,\theta^*,h^*))\\
=&\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^*(x_i-m_i^*)\big(g(\langle x_i,\,\theta^*\rangle+h^*(z_i))-g(\langle x_i,\,\theta\rangle+h^*(z_i))\big)\\
=&\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\mathbf{\Omega}_i^{*,2}(x_i-m_i^*)\big(g(\langle x_i,\,\theta^*\rangle+h^*(z_i))-g(\langle x_i,\,\theta\rangle+h^*(z_i))\big).
\end{aligned}$$

定义 $\Delta_k=\Delta_k(\theta):=g(\langle e_k,\ \theta^*\rangle+h^*(z_i))-g(\langle e_k,\ \theta\rangle+h^*(z_i))$，其中 $k\in[d_T]$, $\Delta_0:=0$，并记$\mathbf{\Delta}_i^{\mathrm{vec}}:=(\Delta_1,\ \ldots,\ \Delta_{d_T})^\top$。于是有

$$\begin{aligned}
&\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\mathbf{\Omega}_i^{*,2}(x_i-m_i^*)\big(g(\langle x_i,\,\theta^*\rangle+h^*(z_i))-g(\langle x_i,\,\theta\rangle+h^*(z_i))\big)\\
=&\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\mathbb{E}(\mathbf{\Omega}_i^{*,2}(x_i-m_i^*)\big(g(\langle x_i,\,\theta^*\rangle+h^*(z_i))-g(\langle x_i,\,\theta\rangle+h^*(z_i))\big)|z_i,\mathcal{F}_{i-1})\\
=&\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\mathbf{\Omega}_i^{*,2}\mathbf{D}_p(\mathbf{\Delta}_i^{\mathrm{vec}}-\bar m^*\bar\Delta),
\end{aligned}$$

其中 $\mathbf{D}_p:=\mathrm{diag}\{p_{i1},\ldots,p_{id_T}\}$，$\bar m^*:=\mathbf{D}_p^{-1}m_i^*$且 $\bar\Delta:=\sum_{j=0}^{d_T}p_{ij}\Delta_j$。这里，最后一行是对 x_i 取条件期望得到的。为简洁起见，我们省略了 p_{ij} 中对时间 i 的依赖。此外，根据引理 16 的证明中的方程（108），有

$$\begin{aligned}
\mathbf{\Omega}_i^{*,2}\mathbf{D}_p(\mathbf{\Delta}_i^{\mathrm{vec}}-\bar m^*\bar\Delta)&=(\mathbf{C}_i+\mathbf{\Delta}_i)\mathbf{D}_p(\mathbf{\Delta}_i^{\mathrm{vec}}-\bar m^*\bar\Delta)\\
&=(\mathbf{B}_i+\mathbf{\Delta}_i\mathbf{D}_p)(\mathbf{\Delta}_i^{\mathrm{vec}}-\bar m^*\bar\Delta),
\end{aligned}$$

其中 $\mathbf{B}_i=\mathrm{diag}\{1/\bar\varepsilon_1^*,\ldots,1/\bar\varepsilon_{d_T}^*\}$, $\mathbf{C}_i:=\mathbf{D}_p^{-1}\mathbf{B}_i=\mathrm{diag}\{1/(p_1\bar\varepsilon_1^*)\ldots,\ 1/(p_{d_T}\bar\varepsilon_{d_T}^*)\}$,

$$\mathbf{\Delta}_i:=\mathbf{B}_i\mathbf{K}_i\frac{\begin{pmatrix}-p_0\bar\varepsilon_0^* & \bar m_0^*p_0\\ \bar m_0^*p_0 & \sum_{k=1}^{d_T}p_k\bar m_k^{*2}/\bar\varepsilon_k^*\end{pmatrix}}{(\sum_{k=1}^{d_T}p_k\bar m_k^{*2}/\bar\varepsilon_k^*)p_0\bar\varepsilon_0^*+\bar m_0^{*2}p_0^2}\mathbf{K}_i^\top\mathbf{B}_i=\frac{1_{d_T}1_{d_T}^\top}{p_0\bar\varepsilon_0^*}+\frac{\mathbf{B}_i\mathbf{K}_i\begin{pmatrix}-p_0\bar\varepsilon_0^* & \bar m_0^*p_0\\ \bar m_0^*p_0 & -p_0\bar m_0^{*2}/\bar\varepsilon_0^*\end{pmatrix}\mathbf{K}_i^\top\mathbf{B}_i}{(\sum_{k=1}^{d_T}p_k\bar m_k^{*2}/\bar\varepsilon_k^*)p_0\bar\varepsilon_0^*+\bar m_0^{*2}p_0^2},$$

$\mathbf{K}_i:=\begin{pmatrix}\bar m_1 & \bar m_2 & \cdots & \bar m_{d_T}\\ \bar\varepsilon_1^* & \bar\varepsilon_2^* & \cdots & \bar\varepsilon_{d_T}^*\end{pmatrix}$，$\bar\varepsilon_j^*:=\nu^2\left(g(\theta_j^*+h^*\left(z_i\right)\right)$，以及

$$\bar m_0^*:=g'(h^*(z_i))/\sum_{k=0}^{d_T}p_{ik}g'(\theta_k^*+h^*(z_i)).$$

由于对于逻辑回归模型，$\nu^2\left(g\left(s\right)\right)=g'\left(s\right)$对所有 $s\in\mathbb{R}$均成立，因此可得

$$\bar\varepsilon_j^*/\bar m_j^*=\sum_{k=0}^{d_T}p_{ik}g'(\theta_k^*+h^*(z_i))$$

对于 $0\le j\le d_T$。因此，可以验证 $\mathbf{\Delta}_i=1_{d_T}1_{d_T}^\top/\left(p_0\bar\varepsilon_0^*\right)$，从而 e

$$\begin{aligned}
&(\mathbf{B}_i+\mathbf{\Delta}_i\mathbf{D}_p)(\mathbf{\Delta}_i^{\mathrm{vec}}-\bar m^*\bar\Delta)\\
=&\mathbf{B}_i\mathbf{\Delta}_i^{\mathrm{vec}}+\frac{\bar\Delta}{p_0\bar\varepsilon_0^*}1_{d_T}-\frac{\bar\Delta}{\sum_{k=0}^{d_T}p_{ik}g'(\theta_k^*+h^*(z_i))}1_{d_T}-\frac{(1-p_0\bar m_0^*)\bar\Delta}{p_0\bar\varepsilon_0^*}1_{d_T}\\
=&\mathbf{B}_i\mathbf{\Delta}_i^{\mathrm{vec}}.
\end{aligned}$$

综合各部分可得

$$\widetilde{\mathbb{E}}_{n_2}(\phi_i(\theta,\theta^*,h^*)-\phi_i(\theta^*,\theta^*,h^*))=\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\mathbf{B}_i\mathbf{\Delta}_i^{\mathrm{vec}}.$$

注意，根据泰勒展开，$\Delta_k = g'(\langle e_k, \theta\rangle + h^*(z_i))(\theta_k^* - \theta_k)$对某个 θ 成立。由 g'，ν^2 的有界性假设，我们还可以写出

$$\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\mathbf{B}_i\mathbf{\Delta}_i^{\mathrm{vec}} = \widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\widetilde{\mathbf{B}}_i(\theta^* - \theta) = [\widetilde{\mathbb{E}}_{n_2}\mathbf{\Omega}_i^{*,-1}\widetilde{\mathbf{B}}_i](\theta^* - \theta),$$

其中，$\mathbf{B}_i$ 是满足以下条件的对角矩阵：$c_{\mathbf{B},1}\mathbf{I}_{d_T} \preceq \mathbf{B}_i \preceq c_{\mathbf{B},2}\mathbf{I}_{d_T}$ 其中常数 $c_{\mathbf{B},1}, c_{\mathbf{B},2} > 0$ 可依赖于问题参数。

D 干扰项的自适应估计

本节将讨论一种估计量的替代构造方法 θ，该方法可能具有更高的样本效率。我们原先的过程基于拆分数据集，并使用前 n_1 个数据点来获得干扰项估计 h（以及广义线性模型的目标估计 θ）。相反，假设在每个时刻 $i \in [n]$，我们利用截至时刻 $i-1$ 所收集的数据来构造估计量 h_i（以及广义线性模型的目标估计 θ_i）（从 $\theta_1 = 0_{d_T}$，$h_1 \equiv 0$开始）。对于部分线性模型，我们随后求解

$$\frac{1}{n}\sum_{i=1}^{n}\phi_i(\theta, \widehat{h}_i) = 0 \tag{75}$$

以计算估计量 θ。对于广义线性模型，我们随后根据以下系统计算估计量 θ

$$\frac{1}{n}\sum_{i=1}^{n}\phi_i(\theta, \widehat{\theta}_i, \widehat{h}_i) = 0 \tag{76}$$

由于使用了自适应更新的干扰项估计，可以证明，估计值 θ的样本效率优于算法 1 和 2 得到的估计值。具体而言，我们有以下结果[5]（与定理 1 和 3 相比）。

推论 5. 假设定理 *1* 中的假设成立，并将假设（**NUI**）替换为

（$\mathbf{NUI_{ada}}$）由方程 (75) 得到的估计量序列 h_i 满足 s

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\widehat{h}_i, z_i}(\widehat{h}_i(z_i) - h^*(z_i))^2 \to 0, \tag{77}$$

其中，期望是对 $(\mathcal{F}_{i-1}, z_i)$ 求取的。

那么，由方程 (75) 得到的估计量 θ 满足

$$(\sqrt{n}\widehat{\mathbb{E}}_n\mathbf{\Sigma}_i^{1/2})(\widetilde{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2\mathbf{I}_{d_T}). \tag{78}$$

参见 D.1 节中的证明。

推论 6。 假设定理 *3* 中的假设均成立，并将假设（**NUI′**）替换为

[5]对于固定方向推断，也可以证明类似的结果。

（$\mathbf{NUI}'_{\mathbf{ada}}$）假设 $\mathcal{P}$ 中的所有分布都支撑于集合 $\mathrm{dom}(\mathcal{P})$ 上（可以是 $\mathbb{R}^{d_N}$）。$\widehat{\ }$ $\widehat{\ }$通过方程 (76) 获得的估计量 θ_i、h_i 满足

$$\frac{1}{n}\sum_{i=1}^{n}\|\widehat{\theta}_i-\theta^*\|_2^2=o_p(n^{-1/2}),\quad and$$

$$\frac{1}{n}\sum_{i=1}^{n}\sup_{v\in\mathrm{dom}(\mathcal{P})}|\widehat{h}_i(v)-h^*(v)|^2=o_p(n^{-1/2}).$$

$\widetilde{\ }$由方程（76）得到的估计量 θ 满足

$$(\widehat{\mathbb{E}}_n\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}_i\rangle+\widehat{h}_i(z_i)\big)(x_i-\widehat{m}_i)^\top)\sqrt{n}(\widetilde{\theta}-\theta^*)\xrightarrow{d}\mathcal{N}(0,\mathbf{I}_{d_T}).\tag{79}$$

证明见 D.2 节。

可以验证，$(\mathbf{NUI}'_{\mathbf{ada}})$ 的一个充分条件是

$$\lim_{i\to\infty}\mathbb{E}[i^{1+\delta_0}\|\widehat{\theta}_i-\theta^*\|_2^4]\to 0,\quad \lim_{i\to\infty}\mathbb{E}\big[i^{1+\delta_0}\sup_{v\in\mathrm{dom}(\mathcal{P})}|\widehat{h}_i(v)-h^*(v)|^4\big]\to 0$$

对于某个常数 $\delta_0>0$。此外，我们指出，条件 $(\mathbf{NUI}_{\mathbf{ada}})$、$(\mathbf{NUI}'_{\mathbf{ada}})$ 比条件 $(\mathbf{NUI})$、$(\mathbf{NUI}')$ 更强，因为前者是针对估计量序列提出的，而不是针对通过样本分割得到的单个估计量。

与公式（75）和（76）中的估计量相比，在使用固定比例$\widehat{\ }$（而非递减比例）的数据点来计算先验估计 h（即 $\liminf n_1/n>0$）时，算法 1 和 2 中描述的估计量可能具有更大的渐近方差。另一方面，估计量（75）和（76）要求在每个时间步 i计算$\widehat{\ }$干扰项估计 h_i。当不存在干扰项估计的简单更新规则时，这在计算上可能效率低下。

D.1 推论 5 的证明

$\widehat{\ }$推论 5 的证明可沿用证明定理 1 的相同论证，只需将所有公式中的对象 h、$\widehat{h_i}$、n 和 $\mathbb{E}_n$。主要区别在于需要证明公式（42b）的对应结果，即在假设（$\mathbf{NUI}_{\mathbf{ada}}$）下，我们有 n_2 和 $\mathbb{E}_{n_2}$ 分别替换为 $\widehat{\ }$ $\widehat{\ }$

$$\widehat{\mathbb{E}}_n\sqrt{n}v_i(\widehat{h}_i(z_i)-h^*(z_i))\xrightarrow{p}0.\tag{80}$$

由于其余证明在很大程度上相同，这里我们只证明方程 (80)。

方程的证明 (80) 我们首先观察到

$$\mathbb{E}[v_i(\widehat{h}_i(z_i)-h^*(z_i))|\mathcal{F}_{i-1},z_i]=\mathbb{E}[v_i|\mathcal{F}_{i-1},z_i](\widehat{h}_i(z_i)-h^*(z_i))=0.$$

$\widehat{\ }$因此可知，$\{v_i(h_i(z_i)-h^*(z_i))\}_{i=1}^n$ 构成一个鞅差序列。此外，我们有

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\|v_i(\widehat{h}_i(z_i)-h^*(z_i))\|_2^2&=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(\widehat{h}_i(z_i)-h^*(z_i))^2\cdot\mathbb{E}[\|v_i\|_2^2|\mathcal{F}_{i-1},z_i]]\\&=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(\widehat{h}_i(z_i)-h^*(z_i))^2]\to 0,\end{aligned}$$

其中，最后一行的期望是关于 $(\mathcal{F}_{i-1},\ z_i)$ 取的，收敛性源于假设 $(\mathbf{NUI}_{\mathbf{ada}})$。因此，方程 (80) 由引理18立即得出。

D.2 推论 6 的证明

推论 6 的证明与定理 3 的证明基本相同，只需将先验估计量$\widehat{\omega}$ 替换为自适应估计值 $\widehat{\omega}_i$。同样，我们考虑干扰分量为线性的简单情形，即 $h^*(z_i) = \langle z_i, \eta^* \rangle$，并记作 $\omega = (\theta, \eta)$（ $\widehat{\omega}_i$ 和 ω^*也类似）。注意，假设 $(\mathbf{NUI}'_{\mathbf{ada}})$ 蕴含

$$\frac{1}{n}\sum_{i=1}^{n} \|\widehat{\theta}_i - \theta^*\|_2 = o_p(n^{-1/4}), \quad \frac{1}{n}\sum_{i=1}^{n} \|\widehat{\eta}_i - \eta^*\|_2 = o_p(n^{-1/4})$$

根据柯西-施瓦茨不等式。此外，从引理15的证明中可以看出，$\mathbb{E}(\phi_i(\theta, \omega) | \mathcal{F}_{i-1})$ 和 $\mathbb{E}(\partial_\omega \phi_i(\theta, \omega) | \mathcal{F}_{i-1})$ 在所有 i 上均为一致 Lipschitz。

因此，可以验证：参照定理 3（及相关引理）的证明，将其中的 $\|\widehat{\omega} - \omega^*\|_2$和 $\|\widehat{\omega} - \omega^*\|_2^2$ 分别替换为 $\sum_{i=1}^n \|\widehat{\omega}_i - \omega^*\|_2/n$ 和 $\sum_{i=1}^n \|\widehat{\omega}_i - \omega^*\|_2^2/n$ 后，可以建立相同的结果。根据假设（$\mathbf{NUI}'_{\mathbf{ada}}$），推论 6 随即成立。由于证明本质上相同，为简洁起见，此处略去。

E 当 p_i 未知时的推断

本节研究选择概率 $\{p_i\}_{i=1}^n$ 未知时的推断问题，但我们可以使用一系列相合估计量 $\{\widehat{p}_i\}_{i=1}^n$。在这种情况下，类似地，可以在得分函数 ϕ_i的计算中用 $\widehat{p}_i$ 替换 p_i，从而得到估计值 θ（或 θ_u）。我们证明，当 $\{\widehat{p}_i\}_{i=1}^n$ 与 $\{p_i\}_{i=1}^n$足够接近时，定理 1 的一个修正版仍然成立。定义 $\widehat{\mathbf{\Sigma}}_i := \mathbb{E}((x_i - \widehat{p}_i)(x_i - \widehat{p}_i)^\top | \mathcal{F}_{i-1}, z_i)$，并定义 $\|v\|_{\mathbf{\Sigma}_i^{-1}} := \sqrt{v^\top \mathbf{\Sigma}_i^{-1} v}$，其中 $v \in \mathbb{R}^{d_T}$为任意向量。我们假设估计量序列 $\{\widehat{p}_i\}_{i=1}^n$满足以下一组收敛假设：

（**CON**）（a） $\widehat{p}_i \in [0,1]$ 和 $\widehat{p}_i \in \sigma(z_i, \mathcal{F}_{i-1}) =: \mathcal{G}_{i-1}$ 对所有 $i \in [n]$均成立，即，$\widehat{p}_i$ 使用以下方式计算 g
z_i、前 $i-1$ 个样本以及任何与所收集的 d
样本。

(b) $\widehat{\mathbb{E}}_{n_2}\|\widehat{p}_i - p_i\|^2_{\mathbf{\Sigma}_i^{-1}} \xrightarrow{p} 0$, and $(\widehat{\mathbb{E}}_{n_2}\|\widehat{p}_i - p_i\|^2_{\mathbf{\Sigma}_i^{-1}}) \cdot (\widehat{\mathbb{E}}_{n_2}(\widehat{h}(z_i) - h^*(z_i))^2) = o_p(1/n_2)$.

(c) $\|\|\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i^{1/2}\|\|_{\mathrm{op}} \le B_{\mathbf{\Sigma}}$ for some $B_{\mathbf{\Sigma}} > 0$, and $\widehat{\mathbb{E}}_{n_2}\|\|\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i^{1/2} - \mathbf{I}_{d_T}\|\|^2_{\mathrm{op}} \xrightarrow{p} 0$.

推论 7。假设假设（**NOI**（ν, σ^2））、（**SEL**（t））和（**CON**）均成立。当~选择概率 p_i 未知时，由方程 (75) 获得的估计量 θ 用$\{\widehat{p}_i\}_{i=1}^n$ 替代 $\{p_i\}_{i=1}^n$ 后满足

$$(\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{v}_i x_i^\top)(\widetilde{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \mathbf{I}_{d_T}). \tag{81}$$

证明见 E.1 节。

我们注意到，方程 (81) 与定理 1 中的方程 (14) 唯一的区别在于左侧的预条件矩阵。具体而言，方程 (81) 将 $\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}$ 替换为 $\widehat{\mathbb{E}}_{n_2}\widehat{v}_i x_i^\top$，后者在未知 p_i 的情况下也可计算。我们猜想，当 $\{\widehat{p}_i\}_{i=1}^n$ 与 $\{p_i\}_{i=1}^n$ 足够接近时，类似结论或许也适用于固定方向推断和广义线性模型。我们认为这是一个值得未来深入研究的方向。

在实践中，通常很难找到这样一个由相合估计量 $\{\widehat{p}_i\}_{i=1}^n$ 构成的序列。从理论上说，如果对选择概率一无所知，则不可能做到这一点；因为在最坏情况下， p_i 对 $(z_i,\ \mathcal{F}_{i-1})$ 的依赖关系可能在不同的 $i \in [n]$ 之间任意不同，而对于每个 i.，我们只有一个样本 x_i 来估计 p_i

然而，如果提供额外的先验知识，仍有可能实现一致估计。例如，若选择概率随时间保持不变，即对于所有 $i \in [n]$ 和某个函数 p， $p_i(z_i,\ \mathcal{F}_{i-1}) = p(z_i)$，那么在 p 具有良性参数（或非参数）形式的前提下，标准估计方法（如经验风险最小化）可能能够找到满足假设（**CON**）的 $\{\widehat{p}_i\}_{i=1}^n$。另一种情况是，如果整个选择概率函数集合 $\{p_i(\cdot)\}_{i=1}^n$（我们称该集合为一个选择算法）取自一个已知的有限选择算法集合，那么随着 n 增大，基于观测样本 $\{(x_i,\ z_i)\}_{i=1}^n$ 以趋近于 1 的概率识别出真实选择算法或许是可行的。

超出本文的研究设定，如果我们观测到一批 K 个独立同分布轨迹 $\{(y_i^{(k)},\ x_i^{(k)},\ z_i^{(k)})\}_{i=1}^n$, $k \in [K]$，且 K [67] 足够大，那么可能能够一致估计 $\{p_i\}_{i=1}^n$。在这种情况下，我们有 K 个独立同分布样本 $\{(x_i^{(k)},\ z_i^{(k)},\ \mathcal{F}_i^{(k)})\}_{k=1}^K$ 来估计每个 p_i。因此，当批次大小 $K \to \infty$ 时，或许能够实现一致估计。

E.1 推论 7 的证明

回顾定理 1 的证明中向量 $v_i := \mathbf{\Sigma}_i^{-1/2}(x_i - \widehat{p}_i)$，我们定义向量$\widehat{v}_i := \widehat{\mathbf{\Sigma}}_i^{-1/2}(x_i - \widehat{p}_i)$。与定理 1 的证明类似，我们证明数对 $(\widetilde{\theta},\ \widehat{h})$ 满足方程

$$\sqrt{n_2}(\widehat{\mathbb{E}}_{n_2}\widehat{v}_i x_i^\top)(\widetilde{\theta} - \theta^*) = \sqrt{n_2}\big\{\widehat{\mathbb{E}}_{n_2}\widehat{v}_i\varepsilon_i - \widehat{\mathbb{E}}_{n_2}\widehat{v}_i(\widehat{h}(z_i) - h^*(z_i))\big\}. \tag{82}$$

我们的证明基于以下两个辅助命题：

$$\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}\widehat{v}_i\varepsilon_i \xrightarrow{d} \mathcal{N}(0, \sigma^2\mathbf{I}_{d_T}), \tag{83a}$$

$$\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}\widehat{v}_i(\widehat{h}(z_i) - h^*(z_i)) \xrightarrow{p} 0. \tag{83b}$$

结合公式（83a）和（83b）即可立即得到推论 7。

方程的证明 (83a) 证明与引理 3 的证明基本相同。这里我们只强调其中的差异。

回忆一下，我们将 $\mathcal{G}_{i-1}$ 定义为 σ-域 $\sigma(z_i,\ \mathcal{F}_{i-1})$。注意， $\{\widehat{v}_i\varepsilon_i\}_{i>n_1}$ 关于 $\{\mathcal{G}_i\}_{i>n_1}$ 构成一个鞅差序列，因为 $\widehat{v}_i\varepsilon_i \in \mathcal{G}_i$ 且 $\mathbb{E}(\widehat{v}_i\varepsilon_i \mid \mathcal{G}_{i-1}) = 0$。因此，我们可以通过应用鞅差序列的中心极限定理来证明方程 (83a)。

渐近协方差 注意，

$$\begin{aligned}\mathbb{E}(\varepsilon_i^2\widehat{v}_i\widehat{v}_i^\top \mid \mathcal{G}_{i-1}) &= \mathbb{E}(\widehat{v}_i\widehat{v}_i^\top\mathbb{E}(\varepsilon_i^2 \mid x_i, z_i, \mathcal{F}_{i-1}) \mid \mathcal{G}_{i-1})) \\ &= \mathbb{E}(\sigma^2\widehat{v}_i\widehat{v}_i^\top \mid z_i, \mathcal{F}_{i-1}) = \sigma^2\widehat{\mathbf{\Sigma}}_i^{-1/2}(\mathbf{\Sigma}_i + (p_i - \widehat{p}_i)(p_i - \widehat{p}_i)^\top)\widehat{\mathbf{\Sigma}}_i^{-1/2}.\end{aligned}$$

因此，渐近方差为

$$\begin{aligned}\frac{1}{n_2}\sum_{i=n_1+1}^{n}\mathbb{E}(\varepsilon_i^2\widehat{v}_i\widehat{v}_i^\top\mid\mathcal{G}_{i-1}) &= \sigma^2\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Sigma}}_i^{-1/2}(\mathbf{\Sigma}_i+(p_i-\widehat{p}_i)(p_i-\widehat{p}_i)^\top)\widehat{\mathbf{\Sigma}}_i^{-1/2}\\ &= \sigma^2\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i\widehat{\mathbf{\Sigma}}_i^{-1/2}+\sigma^2\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Sigma}}_i^{-1/2}(p_i-\widehat{p}_i)(p_i-\widehat{p}_i^\top)\widehat{\mathbf{\Sigma}}_i^{-1/2}\\ &\xrightarrow{p}\sigma^2\mathbf{I}_{d_T},\end{aligned}$$

其中，最后一行用到了假设（**CON**）（b）和引理 2。

Lind伯格条件 注意，根据假设 **(A2b)** ，我们有 $\mathbf{\Sigma}_i\succeq c_i\mathbf{I}_{d_T}$，且

$$\begin{aligned}\|\widehat{v}_i\|_2^2 &\le |||\widehat{\mathbf{\Sigma}}_i^{-1/2}|||_{\mathrm{op}}^2\cdot\|x_i-\widehat{p}_i(z_i,\mathcal{F}_{i-1})\|_2^2\\ &\le |||\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i^{1/2}|||_{\mathrm{op}}^2\cdot|||\mathbf{\Sigma}_i^{-1/2}|||_{\mathrm{op}}^2\cdot\|x_i-\widehat{p}_i(z_i,\mathcal{F}_{i-1})\|_2^2\le 4B_{\mathbf{\Sigma}}{}^2/c_i,\end{aligned}\tag{84}$$

其中，最后一个不等式由 $\|x_i-\widehat{p}_i(z_i,\mathcal{F}_{i-1})\|_2\le\|x_i\|_2+\|\widehat{p}_i(z_i,\mathcal{F}_{i-1})\|_2\le 2$ 和假设 **(CON)** 得出。因此，参照引理 3 证明中的类似论证，可以验证 $\{\widehat{v}_i\varepsilon_i\}_{i=n_1+1}^n$ 满足林德伯格条件。

综合各部分并应用鞅中心极限定理，我们得出$\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}\widehat{v}_i\varepsilon_i\xrightarrow{d}\mathcal{N}(0,\sigma^2\mathbf{I}_{d_T})$。

—

公式的证明 (83b) 将关系式 $\widehat{v}_i=\widehat{\mathbf{\Sigma}}_i^{-1/2}(x_i-p_i)+\widehat{\mathbf{\Sigma}}_i^{-1/2}(p_i-\widehat{p}_i)$ 代入公式（83b）的左端，得到分解 $\mathbb{E}_{n_2}\sqrt{n_2}\widehat{v}_i(h(z_i)-h^*(z_i))\equiv T_1+T_2$，其中

$$\begin{aligned}T_1&:=\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}\widehat{\mathbf{\Sigma}}_i^{-1/2}(x_i-p_i)(\widehat{h}(z_i)-h^*(z_i)),\quad\text{and}\\ T_2&:=\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}\widehat{\mathbf{\Sigma}}_i^{-1/2}(p_i-\widehat{p}_i)(\widehat{h}(z_i)-h^*(z_i)).\end{aligned}$$

注意到 $\{\widehat{\mathbf{\Sigma}}_i^{-1/2}(x_i-p_i)(\widehat{h}(z_i)-h^*(z_i))\}_{i=n_1+1}^n$ 是关于 $\{\mathcal{G}_i\}_{i=n_1+1}^n$ 的鞅差序列。结合界 $\mathbb{E}(\|\widehat{\mathbf{\Sigma}}_i^{-1/2}(x_i-p_i)\|_2^2|\mathcal{G}_{i-1})=\mathrm{tr}(\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i\widehat{\mathbf{\Sigma}}_i^{-1/2})\le d_TB_{\mathbf{\Sigma}}{}^2$，根据引理 4 证明中的类似论证可得 $T_1\xrightarrow{p}0$。

接下来考虑第二项 T_2，观察到

$$\begin{aligned}&\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}\widehat{\mathbf{\Sigma}}_i^{-1/2}(p_i-\widehat{p}_i)(\widehat{h}(z_i)-h^*(z_i))\\ &\qquad\le\sqrt{n_2}(\widehat{\mathbb{E}}_{n_2}\|\widehat{\mathbf{\Sigma}}_i^{-1/2}(p_i-\widehat{p}_i)\|_2^2)^{1/2}\cdot(\widehat{\mathbb{E}}_{n_2}|\widehat{h}(z_i)-h^*(z_i)|^2)^{1/2}.\end{aligned}$$

利用假设 **(CON)**，可知 $T_2\xrightarrow{p}0$。综合 T_1 和 T_2 的结果，证明得证。

引理 2. 在假设 **(CON)** 下，我们有以下结果

$$|||\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i\widehat{\mathbf{\Sigma}}_i^{-1/2}-\mathbf{I}_{d_T}|||_{\mathrm{op}}\xrightarrow{p}0,\tag{85a}$$

$$\widehat{\mathbb{E}}_{n_2}\|\widehat{\mathbf{\Sigma}}_i^{-1/2}(p_i-\widehat{p}_i)\|_2^2\le B_{\mathbf{\Sigma}}{}^2\widehat{\mathbb{E}}_{n_2}\|p_i-\widehat{p}_i\|_{\mathbf{\Sigma}_i^{-1}}^2\xrightarrow{p}0.\tag{85b}$$

证明. 我们先证明方程 (85a)。设 $s_{i1} \geq s_{i2} \geq \ldots \geq s_{id_T}$ 为 $\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i^{1/2}$ 的奇异值。注意，假设 (CON) 意味着

$$\widehat{\mathbb{E}}_{n_2}|s_{ik}-1|^2 \xrightarrow{p} 0$$

对所有 $k \in [d_T]$。因此,

$$\widehat{\mathbb{E}}_{n_2}|s_{ik}^2-1| \leq \widehat{\mathbb{E}}_{n_2}|s_{ik}-1|^2 + 2\widehat{\mathbb{E}}_{n_2}|s_{ik}-1| \xrightarrow{p} 0,$$

其中最后一步用到了詹森不等式。由于 $\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i\widehat{\mathbf{\Sigma}}_i^{-1/2}$ 的特征值为 $\{s_{ik}^2\}_{k=1}^{d_T}$，因此

$$\begin{aligned}
\|\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i\widehat{\mathbf{\Sigma}}_i^{-1/2} - \mathbf{I}_{d_T}\|_{\mathrm{op}} &\leq \widehat{\mathbb{E}}_{n_2}\|\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i\widehat{\mathbf{\Sigma}}_i^{-1/2} - \mathbf{I}_{d_T}\|_{\mathrm{op}} \\
&\leq \widehat{\mathbb{E}}_{n_2}\sum_{k=1}^{d_T}|s_{ik}^2-1| \xrightarrow{p} 0.
\end{aligned}$$

为证明结论 (85b)，我们注意到

$$\begin{aligned}
\widehat{\mathbb{E}}_{n_2}\|\widehat{\mathbf{\Sigma}}_i^{-1/2}(p_i-\widehat{p}_i)\|_2^2 &= \widehat{\mathbb{E}}_{n_2}\|\widehat{\mathbf{\Sigma}}_i^{-1/2}\mathbf{\Sigma}_i^{1/2}\mathbf{\Sigma}_i^{-1/2}(p_i-\widehat{p}_i)\|_2^2 \\
&\leq \widehat{\mathbb{E}}_{n_2}\|\mathbf{\Sigma}_i^{1/2}\widehat{\mathbf{\Sigma}}_i^{-1}\mathbf{\Sigma}_i^{1/2}\|_{\mathrm{op}}\cdot\|p_i-\widehat{p}_i\|_{\mathbf{\Sigma}_i^{-1}}^2 \\
&\leq B_{\mathbf{\Sigma}}{}^2\widehat{\mathbb{E}}_{n_2}\|p_i-\widehat{p}_i\|_{\mathbf{\Sigma}_i^{-1}}^2 \xrightarrow{p} 0,
\end{aligned}$$

其中最后一步用到了假设（**CON**）。

□

F 辅助引理

本节汇集了用于证明定理 1–3 的各个引理的证明。

F.1 定理 1 的辅助引理

本节陈述并证明定理 1 的证明中所使用的辅助引理。

引理 3。 在定理 *1* 的假设下，我们有 $\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}v_i\varepsilon_i \xrightarrow{d} \mathcal{N}(0,\ \sigma^2\mathbf{I}_{d_T})$。

证明。 根据我们的假设，回顾 $\mathbb{E}(\varepsilon_i \mid \mathcal{F}_{i-1},\ x_i,\ z_i) = 0$，并且我们有 $\mathbb{E}(v_i\varepsilon_i \mid \mathcal{F}_{i-1}) = \mathbb{E}(v_i\mathbb{E}(\varepsilon_i \mid x_i,\ z_i,\ \mathcal{F}_{i-1}) \mid \mathcal{F}_{i-1}) = 0$，因此 $\{v_i\varepsilon_i\}_{i\geq n_1}$ 是一个鞅差序列。我们通过对序列 $\{v_i\varepsilon_i\}_{i\geq n_1}$ 应用标准鞅中心极限定理来证明引理 3。

渐近协方差 注意到

$$\mathbb{E}(\varepsilon_i^2 v_i v_i^\top \mid \mathcal{F}_{i-1}) = \mathbb{E}(v_i v_i^\top \mathbb{E}(\varepsilon_i^2 \mid x_i, z_i, \mathcal{F}_{i-1}) \mid \mathcal{F}_{i-1}) = \mathbb{E}(\sigma^2 v_i v_i^\top \mid \mathcal{F}_{i-1}) = \sigma^2\mathbf{I}_d,$$

其中最后一个等式由.....得出

$$\begin{aligned}
\mathbb{E}(v_i v_i^\top \mid \mathcal{F}_{i-1}) &= \mathbb{E}(\mathbf{\Sigma}_i^{-1/2}(x_i - p_i(z_i,\mathcal{F}_{i-1}))(x_i - p_i(z_i,\mathcal{F}_{i-1}))^\top\mathbf{\Sigma}_i^{-1/2} \mid \mathcal{F}_{i-1}) \\
&= \mathbb{E}(\mathbf{\Sigma}_i^{-1/2}\mathbb{E}((x_i - p_i(z_i,\mathcal{F}_{i-1}))(x_i - p_i(z_i,\mathcal{F}_{i-1}))^\top \mid z_i,\mathcal{F}_{i-1})\mathbf{\Sigma}_i^{-1/2} \mid \mathcal{F}_{i-1}) \\
&= \mathbb{E}(\mathbf{\Sigma}_i^{-1/2}\mathbf{\Sigma}_i\mathbf{\Sigma}_i^{-1/2} \mid \mathcal{F}_{i-1}) = \mathbf{I}_{d_T}.
\end{aligned}$$

林德伯格条件 注意，根据假设**(A2b)**，我们有 $\mathbf{\Sigma}_i \succeq c_i \mathbf{I}_{d_T}$，并且

$$\|v_i\|_2^2 = \|\!|v_i v_i^\top|\!\|_{\mathrm{op}} \le \|\!|\mathbf{\Sigma}_i^{-1/2}|\!\|_{\mathrm{op}} \cdot \|x_i - p_i(z_i, \mathcal{F}_{i-1}))\|_2^2 \cdot \|\!|\mathbf{\Sigma}_i^{-1/2}|\!\|_{\mathrm{op}} \le 4/c_i, \tag{86}$$

其中，第二个不等式由 $\|x_i - p_i(z_i, \mathcal{F}_{i-1})\|_2 \le \|x_i\|_2 + \|p_i(z_i, \mathcal{F}_{i-1})\|_2 \le 2$ 得出。因此，我们有 $v_i v_i^\top \preceq 4\mathbf{I}_{d_T}/c_i$，进而得到

$$\begin{aligned} 0 \preceq \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbb{E}(\varepsilon_i^2 v_i v_i^\top 1_{\{\|\!|\varepsilon_i^2 v_i v_i^\top|\!\|_{\mathrm{op}} > \varepsilon n\}} \mid \mathcal{F}_{i-1}) &\preceq \lim_{n\to\infty} \frac{4}{n}\sum_{i=1}^n \frac{1}{c_i}\mathbb{E}(\varepsilon_i^2 1_{\{\varepsilon_i^2 \ge \varepsilon n c_i/4\}} \mid \mathcal{F}_{i-1})\mathbf{I}_{d_T} \\ &=: T_0. \end{aligned} \tag{87}$$

由于 $\varepsilon_i' s$ 几乎必然是具有共同参数 ν 的次高斯随机变量，因此 $\varepsilon_i^{2\prime} s$ 是具有共同参数的次指数随机变量。因此，存在某个常数 $K_1 > 0$，其取决于 ν，使得 $\mathbb{P}(\varepsilon_i^2 \ge s) \le 2\exp(-s/K_1)$，于是

$$\mathbb{E}(\varepsilon_i^2 1_{\{\varepsilon_i^2 \ge \varepsilon n c_i/4\}}) = \int_{\varepsilon n c_i/4}^\infty \mathbb{P}(\varepsilon_i^2 \ge s) ds \le 2\int_{\varepsilon n c_i/4}^\infty \exp(-s/K_1) ds = 2K_1 \exp\Big(\frac{-\varepsilon n c_i}{4K_1}\Big).$$

将其代入方程（87），对于 $t \in (0, 1/2)$，我们有

$$T_0 \lesssim \frac{c_0 K_1}{n}\sum_{i=1}^n i^{2t} \exp\Big(\frac{-\varepsilon v n c_0}{4K_1 i^{2t}}\Big) \le c_0 K_1 n^{2t} \exp\Big(\frac{-\varepsilon n^{1-2t} c_0}{4K_1}\Big) \to 0.$$

注意，这意味着 $\{v_i \varepsilon_i\}_{i=n_1+1}^n$ 满足林德伯格条件。

综合以上各步并应用鞅中心极限定理，我们得出结论 $\widehat{\mathbb{E}}_{n_2}\sqrt{n_2} v_i \varepsilon_i \xrightarrow{d} \mathcal{N}(0, \sigma^2 \mathbf{I}_{d_T})$。

放宽假设 (NOI(ν, σ^2)) 噪声变量 $\{\varepsilon_i\}_{i=1}^n$ 的次高斯性并非定理 1 成立的必要条件。该定理依赖于林德伯格条件，即使将假设 **(NOI(ν, σ^2))** 放宽为以下条件，该条件仍然成立：

> **(NOIw(α, σ^2))** 给定 $(x_i, z_i, \mathcal{F}_{i-1})$ 后，零均值噪声序列中的每个元素的条件方差为 σ^2：$= \mathbb{E}[\varepsilon_i^2 \mid x_i, z_i, \mathcal{F}_{i-1}]$，并且满足
>
> $$\mathbb{P}(|\varepsilon_i| \ge s) \le \frac{c}{s^\alpha}, \quad \text{for all } s \ge 0 \text{ and some constant } c > 0,$$
>
> 对于某个 $\alpha > 2/(1-2t)$。

回顾一下，标量 $t \in [0, 1/2)$ 在假设 **(SEL(t))** 中定义。请注意，这一放宽后的假设允许存在许多非次高斯的重尾噪声分布，包括柯西分布、（对称）帕累托分布等。

下面概述在放宽的假设（**NOIw**(α, σ^2)）下的证明。我们有

$$\mathbb{E}(\varepsilon_i^2 1_{\{\varepsilon_i^2 \ge \varepsilon n c_i/4\}}) = \int_{\varepsilon n c_i/4}^\infty \mathbb{P}(\varepsilon_i^2 \ge s) ds \le 2c \cdot \int_{\varepsilon n c_i/4}^\infty s^{-\alpha/2} ds \lesssim (\varepsilon n c_i)^{1-\alpha/2}$$

因此，

$$T_0 \lesssim \frac{1}{n}\sum_{i=1}^n \frac{1}{c_i}(\varepsilon n c_i)^{1-\alpha/2} \lesssim \sum_{i=1}^n (n c_i)^{-\alpha/2} \varepsilon^{1-\alpha/2} \lesssim \varepsilon^{1-\alpha/2} \frac{1}{n^{\alpha/2}} \sum_{i=1}^n i^{\alpha t} \lesssim \varepsilon^{1-\alpha/2} n^{\alpha t + 1 - \alpha/2} \to 0$$

当 $\alpha > 2/(1-2t)$ 时。因此满足林德伯格条件。 □

引理 4. 在定理 *1* 的假设下，我们有 $\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}v_i(\widehat{h}(z_i)-h^*(z_i)) \xrightarrow{p} 0.$

证明。 该证明是对马尔可夫不等式的标准应用，并利用了假设（**NUI**）。注意到

$$\mathbb{E}(v_i(\widehat{h}(z_i)-h^*(z_i)) \mid \mathcal{F}_{i-1}) = \mathbb{E}[\mathbb{E}(v_i \mid \mathcal{F}_{i-1}, z_i)(\widehat{h}(z_i)-h^*(z_i))] = 0,$$

因此可得， $v_i(h(z_i)-h^*(z_i))$ 是一个鞅差序列。现在，对于任意 $\varepsilon>0$，定义事件

$$\mathcal{C}_{n_1,\varepsilon} := \{\sup_{P\in\mathcal{P}}(\mathbb{E}_{z\sim P}[\widehat{h}(z)-h^*(z)]^2)^{1/2} \le \varepsilon\}.$$

注意到 $\mathcal{C}_{n_1,\ \varepsilon} \in \mathcal{F}_{n_1}$。我们有

$$\begin{aligned}
\mathbb{E}\|\widehat{\mathbb{E}}_{n_2} n_2^{1/2} 1_{\mathcal{C}_{n_1,\varepsilon}} v_i(\widehat{h}(z_i)-h^*(z_i))\|_2^2 &= \mathbb{E}1_{\mathcal{C}_{n_1,\varepsilon}}\mathbb{E}(\widehat{\mathbb{E}}_{n_2}\|v_i(\widehat{h}(z_i)-h^*(z_i))\|_2^2 \mid \mathcal{F}_{n_1}) \\
&= \mathbb{E}1_{\mathcal{C}_{n_1,\varepsilon}}\widehat{\mathbb{E}}_{n_2}\mathbb{E}(\|v_i\|_2^2(\widehat{h}(z_i)-h^*(z_i))^2 \mid \mathcal{F}_{n_1}) \\
&= d_T\mathbb{E}1_{\mathcal{C}_{n_1,\varepsilon}}\mathbb{E}(\widehat{\mathbb{E}}_{n_2}[\widehat{h}(z_i)-h^*(z_i)]^2 \mid \mathcal{F}_{n_1}) \\
&\le d_T\varepsilon^2,
\end{aligned}$$

其中第三行使用了界 $\mathbb{E}(\|v_i\|_2^2 \mid z_i,\ \mathcal{F}_{i-1}) = \mathbb{E}(\mathrm{tr}(v_iv_i^\top) \mid z_i,\ \mathcal{F}_{i-1}) = d_T$，而最后一行则由 $\mathcal{C}_{n_1,\ \varepsilon}$ 的定义得出。因此，对于任意 $\delta>0$，由马尔可夫不等式可得

$$\mathbb{P}(\|\widehat{\mathbb{E}}_{n_2} n_2^{1/2} 1_{\mathcal{C}_{n_1,\varepsilon}} v_i(\widehat{h}(z_i)-h^*(z_i))\|_2 \ge \sqrt{\frac{d_T}{\delta}}\varepsilon) \le \delta.$$

根据假设（**NUI**）， $\mathbb{P}(\mathcal{C}_{n_1,\ \varepsilon}) \to 1$ 当 $n_1\to\infty$ 时，因此对于充分大的 n_2，有 $\mathbb{P}(\|\widehat{\mathbb{E}}_{n_2} n_2^{1/2} v_i(\widehat{h}(z_i)-h^*(z_i))\|_2 \ge \sqrt{\frac{d_T}{\delta}}\varepsilon) \le 2\delta$。综合以上结果，我们得出 $\widehat{\mathbb{E}}_{n_2}\sqrt{n_2}v_i(\widehat{h}(z_i)-h^*(z_i)) \xrightarrow{p} 0$。 □

引理 5. 在定理 *1* 的假设下，我们有

$$\|\|\widehat{\mathbb{E}}_{n_2} v_i x_i^\top - \widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\|\|_{\mathrm{op}} = o_p(\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})).$$

证明. 注意到

$$\|\|\widehat{\mathbb{E}}_{n_2} v_i x_i^\top - \widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\|\|_{\mathrm{op}} \le \|\|\widehat{\mathbb{E}}_{n_2} v_i x_i^\top - \widetilde{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\|\|_{\mathrm{op}} + \|\|\widetilde{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2} - \widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\|\|_{\mathrm{op}}. \tag{88}$$

我们通过证明以下两个界来对上面的两项进行界定

$$\|\|\widehat{\mathbb{E}}_{n_2} v_i x_i^\top - \widetilde{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\|\|_{\mathrm{op}} = o_p(\sigma_{\min}(\widetilde{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})) \tag{89a}$$

$$\|\|\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2} - \widetilde{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\|\|_{\mathrm{op}} = o_p(\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})) \tag{89b}$$

暂且假设最后两个界成立，将它们代入方程（88），即可得到

$$\begin{aligned}
&\|\|\widehat{\mathbb{E}}_{n_2} v_i x_i^\top - \widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\|\|_{\mathrm{op}} \\
&= o_p(\sigma_{\min}(\widetilde{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})) + o_p(\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})) \\
&\overset{(i)}{\le} o_p(\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})) + \|\|\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2} - \widetilde{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2}\|\|_{\mathrm{op}} + o_p(\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})) \\
&\le o_p(\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}\mathbf{\Sigma}_i^{1/2})),
\end{aligned}$$

其中，不等式 (i) 由 Weyl定理（例如，见 Horn and Johnson [27] 中的定理4.3.1）得出，最后一个不等式则由界（89b）得出。接下来还需证明界（89a）和（89b）。

界（89a）的证明 由于 $\mathbb{E}(v_i p_i^\top \mid z_i,\ \mathcal{F}_{i-1}) = \mathbb{E}(\boldsymbol{\Sigma}_i^{-1/2}(x_i - p_i)p_i^\top \mid z_i,\ \mathcal{F}_{i-1}) = 0$，因此 $\{v_i p_i^\top\}_{i=n_1+1}^{n}$ 是关于滤过$\mathcal{F}_{i-1}$的鞅差序列。此外，注意到 $\mathbb{E}\|v_i p_i^\top\|_F^2 = \mathbb{E}\|v_i\|_2^2\|p_i\|_2^2 \le \mathbb{E}\|v_i\|_2^2 = d_T$。因此，根据引理18，我们有 $\|\widehat{\mathbb{E}}_{n_2} v_i p_i^\top\|_F = \mathcal{O}(1)/\sqrt{n} = o_p(n^{-t})$，其中 $0 < t < \frac{1}{2}$。注意，

$$\widetilde{\mathbb{E}}_{n_2} v_i v_i^\top \boldsymbol{\Sigma}_i^{1/2} = \widetilde{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2} \succeq \frac{c_0}{n^t}\mathbf{I}_{d_T}. \tag{90}$$

此外，随机向量 $v_i v_i^\top \boldsymbol{\Sigma}_i^{1/2} - \mathbb{E}(v_i v_i^\top \boldsymbol{\Sigma}_i^{1/2} \mid \mathcal{F}_{i-1})$ 构成一个鞅差序列，因此

$$\begin{aligned}\mathbb{E}\|v_i v_i^\top \boldsymbol{\Sigma}_i^{1/2} - \mathbb{E}(v_i v_i^\top \boldsymbol{\Sigma}_i^{1/2} \mid \mathcal{F}_{i-1})\|_F^2 &\le \mathbb{E}\|v_i v_i^\top \boldsymbol{\Sigma}_i^{1/2}\|_F^2 = \mathbb{E}\|v_i (x_i - p_i)^\top\|_F^2 \\ &= \mathbb{E}\|v_i\|_2^2 \|x_i - p_i\|_2^2 \lesssim \mathbb{E}\|v_i\|_2^2 = d_T. \end{aligned} \tag{91}$$

因此，由引理18可知，当 $t < 1/2$ 时，$\|(\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2}) v_i v_i^\top \boldsymbol{\Sigma}_i^{1/2}\|_F = \mathcal{O}_p(n^{-1/2}) = o_p(n^{-t})$。结合 $\|\widehat{\mathbb{E}}_{n_2} v_i p_i^\top\|_F = o_p(n^{-t})$、公式（90），并注意到 $\widetilde{\mathbb{E}}_{n_2} v_i v_i^\top \boldsymbol{\Sigma}_i^{1/2} = \widetilde{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2}$，我们得到

$$\begin{aligned}\|\widehat{\mathbb{E}}_{n_2} v_i x_i^\top - \widetilde{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2}\|_{\mathrm{op}} &\le \|\widehat{\mathbb{E}}_{n_2} v_i v_i^\top \boldsymbol{\Sigma}_i^{1/2} - \widetilde{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2}\|_{\mathrm{op}} + \|\widehat{\mathbb{E}}_{n_2} v_i p_i^\top\|_{\mathrm{op}} \\ &= o_p(\sigma_{\min}(\widetilde{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2})).\end{aligned}$$

界（89b）的证明 由于 $\boldsymbol{\Sigma}_i^{1/2} - \mathbb{E}(\boldsymbol{\Sigma}_i^{1/2} \mid \mathcal{F}_{i-1})$ 是一个鞅差序列，并且

$$\mathbb{E}\|\boldsymbol{\Sigma}_i^{1/2} - \mathbb{E}(\boldsymbol{\Sigma}_i^{1/2} \mid \mathcal{F}_{i-1})\|_F^2 \le \mathbb{E}\|\boldsymbol{\Sigma}_i^{1/2}\|_F^2 = \mathbb{E}\,\mathrm{tr}(\boldsymbol{\Sigma}_i) \le d_T,$$

我们有界 $\mathbb{E}\|(\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2})\boldsymbol{\Sigma}_i^{1/2}\|_F^2 \le \frac{d_T}{n_2} = \mathcal{O}(n^{-1})$。

由公式（90），我们有

$$\|\widehat{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2} - \widetilde{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2}\|_{\mathrm{op}} = o_p(n^{-1/2}) = o_p(n^{-t}) = o_p(\sigma_{\min}(\widetilde{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2})).$$

由Weyl定理可知，$\|\widehat{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2} - \widetilde{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2}\|_{\mathrm{op}} = o_p(\sigma_{\min}(\widehat{\mathbb{E}}_{n_2} \boldsymbol{\Sigma}_i^{1/2}))$. □

F.2 定理2的辅助引理

本节 专门证明定理 2的证明中使用的辅助引理 .

引理 6. 在定理 *2*的假设下，我们有 $\widehat{\mathbb{E}}_{n_2}\sqrt{n_2} w_{i1}\varepsilon_i \xrightarrow{d} \mathcal{N}(0,\ \sigma^2)$。

证明。我们采用与证明引理 3 时非常相似的论证：我们证明$\{w_{i1}\varepsilon_i\}_{i\ge 1}$ 是一个鞅差序列，从而可以应用标准的鞅中心极限定理。

通过直接计算可知，$w_{i1}\varepsilon_i$ 是一个鞅差序列。此外，我们有

$$\mathbb{E}(\varepsilon_i^2 {w_{i1}}^2 \mid \mathcal{F}_{i-1}) = \mathbb{E}({w_{i1}}^2 \mathbb{E}(\varepsilon_i^2 \mid x_i, z_i, \mathcal{F}_{i-1}) \mid \mathcal{F}_{i-1}) = \mathbb{E}(\sigma^2 {w_{i1}}^2 \mid \mathcal{F}_{i-1}) = \sigma^2,$$

其中，最后一个等式由以下关系得出

$$\mathbb{E}{w_{i1}}^2 = \mathbb{E}\ \frac{u^\top \mathbf{\Sigma}_i^{-1}(x_i - p_i)(x_i - p_i)^\top \mathbf{\Sigma}_i^{-1} u}{u^\top \mathbf{\Sigma}_i^{-1} u} = 1.$$

现在验证林德伯格条件。首先注意到，对于所有 $k \in S_u \cup \{0\}$， $p_{ik} \gtrsim c_i$，因此

$$\begin{aligned} w_{i1}^2 &\le \|A_{i1}\|_2^2 \cdot \|x_i - p_i(z_i, \mathcal{F}_{i-1})\|_2^2 \\ &\lesssim \frac{u^\top \mathbf{\Sigma}_i^{-2} u}{u^\top \mathbf{\Sigma}_i^{-1} u} \\ &= \frac{\sum_{j,k=1}^{d_T} u_j \mathbf{\Sigma}_{i,jk}^{-2} u_k}{\sum_{j,k=1}^{d_T} u_j \mathbf{\Sigma}_{i,jk}^{-1} u_k} \lesssim \frac{1}{c_i}, \end{aligned} \tag{92}$$

其中，第二个不等式由 $\|x_i - p_i(z_i, \mathcal{F}_{i-1})\|_2 \le \|x_i\|_2 + \|p_i(z_i, \mathcal{F}_{i-1})\|_2 \le 2$ 以及 A_{i1} 的定义得到；最后一个不等式则源于以下事实：对于任意 $j, k \in S_u$

$$\frac{\mathbf{\Sigma}_{i,jk}^{-2}}{\mathbf{\Sigma}_{i,jk}^{-1}} = \frac{\sum_{l=1}^{d_T}(\gamma_i + 1_{\{j=l\}}\frac{1}{p_j})(\gamma_i + 1_{\{l=k\}}\frac{1}{p_k})}{\gamma_i + 1_{\{j=k\}}\frac{1}{p_j}} \le d_T(\gamma_i + \frac{1}{p_j} + \frac{1}{p_k}) \lesssim \frac{1}{c_i}$$

由 e xpression (16b)。因此，我们得到界 $\varepsilon_i^2 w_{i1}^2 \le \varepsilon_i^2 / c_i$，且对于任意 $\varepsilon >$ 0,

$$0 \le \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(\varepsilon_i^2 w_{i1}^2 1_{\{|\varepsilon_i^2 w_{i1}^2| > \varepsilon n\}} \mid \mathcal{F}_{i-1}) \lesssim \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{c_i}\mathbb{E}(\varepsilon_i^2 1_{\{\varepsilon_i^2 \ge \varepsilon n c_i\}} \mid \mathcal{F}_{i-1}) \to 0, \tag{93}$$

其中，收敛性由 ε_i 的次高斯性以及证明引理 3 中方程（87）时所用的相同论证得到。这意味着 $\{w_{i1}\varepsilon_i\}_{i=n_1+1}^{n}$ 满足林德伯格条件。

综合以上结果并应用鞅中心极限定理，我们得出结论 $\widehat{\mathbb{E}}_{n_2}\sqrt{n_2} w_{i1}\varepsilon_i \xrightarrow{d} \mathcal{N}(0, \sigma^2)$。

□

引理 7. 在定理 *2* 的假设下，我们有

$$\widehat{\mathbb{E}}_{n_2}\sqrt{n_2} w_{i1}(\widehat{h}(z_i) - h^*(z_i)) \xrightarrow{p} 0.$$

证明. 注意，引理 4 的证明仅利用了有界性条件 $\mathbb{E}\|v_i\|^2 = d_T$。此外，我们已在引理 6 的证明中证明了 $\mathbb{E}w_{i1}^2 = 1$ 。因此，只需完全按照证明引理 4 的相同论证，将 v_i 替换为 w_{i1}，即可证明此引理。

□

引理 8. 在定理 *2* 的假设下，我们有

$$\Big|\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top u - \widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big| = o_p\Big(\Big|\widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big|\Big).$$

证明。我们有

$$\begin{aligned} \Big|\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top u - \widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big| &\le \Big|\widehat{\mathbb{E}}_{n_2} w_{i1} x_i^\top u - \widetilde{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big| \\ &\quad + |\widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}} - \widetilde{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}| = T_1 + T_2 \end{aligned}$$

我们证明， T_1 和 T_2 均由 $o_p\Big(\Big|\widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big|\Big)$ 有界。

的界： 由于 $\mathbb{E}(w_{i1}p_i^\top \mid z_i,\ \mathcal{F}_{i-1}) = \mathbb{E}(A_{i1}(x_i - p_i)p_i^\top \mid z_i,\ \mathcal{F}_{i-1}) = 0$，$\{w_{i1}p_i^\top\}_{i=n_1+1}^n$是关于 $\mathcal{F}_{i-1}$的鞅差序列。由于 $\mathbb{E}|w_{i1}p_i^\top|^2 = \mathbb{E}|w_{i1}|^2|p_i^\top|^2 \le \mathbb{E}|w_{i1}|^2 = 1$，由引理18可直接得出 $\widehat{\mathbb{E}}_{n_2} w_{i1}p_i^\top = \mathcal{O}_p(n^{-1/2})$。在假设$p_{ik} \gtrsim i^{-2t}$ 对所有 $k \in S_u \cup \{0\}$均成立的条件下，由方程 (16b) 中 $\mathbf{\Sigma}_i$ 的表达式可知 $\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}} \gtrsim n^{-t}$，因此

$$\mathcal{O}_p(n^{-1/2}) = o_p(n^{-t}) = o_p\Big(\Big|\widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big|\Big). \tag{94}$$

注意到

$$\widetilde{\mathbb{E}}_{n_2} w_{i1}(x_i - p_i)^\top u = \widetilde{\mathbb{E}}_{n_2}\frac{u^\top \mathbf{\Sigma}_i^{-1}(x_i - p_i)(x_i - p_i)^\top u}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}} = \widetilde{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}. \tag{95}$$

且 $w_{i1}(x_i - p_i)^\top - \mathbb{E}(w_{i1}(x_i - p_i)^\top \mid \mathcal{F}_{i-1})$ 是一个鞅差序列，且满足

$$\begin{aligned}\mathbb{E}\|w_{i1}(x_i - p_i)^\top - \mathbb{E}(w_{i1}(x_i - p_i)^\top \mid \mathcal{F}_{i-1})\|_2^2 &\le \mathbb{E}\|w_{i1}(x_i - p_i)^\top\|_2^2 \\ &= \mathbb{E}\|x_i - p_i\|_2^2 w_{i1}^2 \lesssim 1,\end{aligned} \tag{96}$$

可得 $\mathbb{E}\|(\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2})w_{i1}(x_i - p_i)^\top\|_2^2 = \mathcal{O}(n^{-1})$，从而 $(\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2})w_{i1}(x_i - p_i)^\top = \mathcal{O}_p(n^{-1/2})$。将此与 $\widehat{\mathbb{E}}_{n_2} w_{i1}p_i^\top = \mathcal{O}_p(n^{-1/2})$ 以及 (94)、(95) 式结合，可得

$$\Big|\widehat{\mathbb{E}}_{n_2} w_{i1}x_i^\top u - \widetilde{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big| = \mathcal{O}_p(n^{-1/2}) = o_p\Big(\widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big).$$

关于 T_2 的界。由于 $\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}} - \mathbb{E}(\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}} \mid \mathcal{F}_{i-1})$ 是一个鞅差序列，且

$$\mathbb{E}\Big|\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}} - \mathbb{E}(\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}} \mid \mathcal{F}_{i-1})\Big|^2 \le \mathbb{E}\Big|\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big|^2 = \mathbb{E}\|u\|_2^2\|\mathbf{\Sigma}_i\|_2 \lesssim 1,$$

我们有 $\mathbb{E}|(\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2})\ \frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}|^2 \lesssim \frac{1}{n_2} = \mathcal{O}(n^{-1})$。因此，

$$\Big|\widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}} - \widetilde{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big| = \mathcal{O}_p(n^{-1/2}) = o_p\Big(\Big|\widehat{\mathbb{E}}_{n_2}\frac{1}{\sqrt{u^\top \mathbf{\Sigma}_i^{-1} u}}\Big|\Big),$$

由此完成证明。

□

F.3 定理 3 的辅助引理

引理 9 ($|||\mathbf{\Omega}_i|||_{\mathrm{op}}$ **的上界)。** *在定理 3 的假设下，我们有*

$$|||\widehat{\mathbf{\Omega}}_i|||_{\mathrm{op}} \le \frac{1}{\sqrt{\widetilde{c}_i}} \qquad and \qquad |||\mathbf{\Omega}_i^*|||_{\mathrm{op}} \le \frac{1}{\sqrt{\widetilde{c}_i}},$$

其中 $\widetilde{c}_i = \widetilde{c}_0/i^{2t}$ *和* $\widetilde{c}_0 = m_\varepsilon c_0/(d_T + 2)$。

$\widehat{}$证明。 我们只证明$\boldsymbol{\Omega}_i$的结果。$\boldsymbol{\Omega}_i^*$ 的结果可用类似的方法证明。回顾

$$\widehat{\boldsymbol{\Omega}}_i = [\mathbb{E}(\nu^2\big(g\big(\langle x_i,\, \widehat{\theta}\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)\big)(x_i - \widehat{m}_i)(x_i - \widehat{m}_i)^\top \mid z_i, \mathcal{F}_{i-1})]^{-1/2}$$

根据定义，只需证明

$$\mathbb{E}(\nu^2\big(g\big(\langle x_i,\, \widehat{\theta}\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)\big)(x_i - \widehat{m}_i)(x_i - \widehat{m}_i)^\top \mid z_i, \mathcal{F}_{i-1}) \succeq \widetilde{c}_i \mathbf{I}_{d_T}.$$

注意到

$$\begin{aligned}
&\mathbb{E}(\nu^2\big(g\big(\langle x_i,\, \widehat{\theta}\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)\big)(x_i - \widehat{m}_i)(x_i - \widehat{m}_i)^\top \mid z_i, \mathcal{F}_{i-1}) \\
&\succeq m_\varepsilon \mathbb{E}((x_i - \widehat{m}_i)(x_i - \widehat{m}_i)^\top \mid z_i, \mathcal{F}_{i-1}) \\
&\succeq m_\varepsilon \mathbb{E}((x_i - p_i)(x_i - p_i)^\top \mid z_i, \mathcal{F}_{i-1}),
\end{aligned}$$

其中，第一个不等式由假设 $\nu^2(x) \geq m_\varepsilon$ 得出，第二个不等式则是因为 $\mathbb{E}(x_i \mid z_i,\ \mathcal{F}_{i-1}) = p_i$。在引理 19 中，我们证明 $\mathbb{E}((x_i - p_i)(x_i - p_i)^\top \mid z_i,\ \mathcal{F}_{i-1}) \succeq c_i/(d_T + 2)\mathbf{I}_{d_T}$。综合各部分可得

$$\mathbb{E}(\nu^2\big(g\big(\langle x_i,\, \widehat{\theta}\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)\big)(x_i - \widehat{m}_i)(x_i - \widehat{m}_i)^\top \mid z_i, \mathcal{F}_{i-1}) \succeq \widetilde{c}_i \mathbf{I}_{d_T},$$

其中 $\widetilde{c}_i = m_\varepsilon c_i/(d_T + 2)$。证明完毕。 □

引理 10。 在定理 *3* 的假设下，有

$$\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\boldsymbol{\Omega}}_i(x_i - \widehat{m}_i)\varepsilon_i \xrightarrow{d} \mathcal{N}(0, \mathbf{I}_{d_T}).$$

证明。 与引理 3 类似，本证明的思路是在序列 $\boldsymbol{\Omega}_i(z_i,\ \mathcal{F}_{i-1})(x_i - \widehat{m}_i)\varepsilon_i$ 上应用鞅版本的$\widehat{}$中心极限定理。根据定义，在广义线性模型 $Y = g(X^\top\omega) + \varepsilon$ 中，ε 的分布取决于 $X^\top\omega$ 的取值。由于 $\mathbb{E}(\varepsilon_i \mid x_i,\ z_i) = 0$，$\boldsymbol{\Omega}_i(z_i,\ \mathcal{F}_{i-1})(x_i - \widehat{m}_i)\varepsilon_i$ 是一个鞅差序列。

渐近协方差 请注意

$$\begin{aligned}
&\mathbb{E}(\|\widehat{\boldsymbol{\Omega}}_i(x_i - \widehat{m}_i)\varepsilon_i\|_2^2 \mid \mathcal{F}_{i-1}) \\
&= \mathbb{E}(\widehat{\boldsymbol{\Omega}}_i(x_i - \widehat{m}_i)\nu^2\big(g\big(\langle x_i,\, \theta^*\rangle + \langle z_i,\, \eta^*\rangle\big)\big)(x_i - \widehat{m}_i)^\top\widehat{\boldsymbol{\Omega}}_i^\top \mid \mathcal{F}_{i-1}) \\
&= \mathbb{E}(\widehat{\boldsymbol{\Omega}}_i\mathbb{E}((x_i - \widehat{m}_i)\nu^2\big(g\big(\langle x_i,\, \theta^*\rangle + \langle z_i,\, \eta^*\rangle\big)\big)(x_i - \widehat{m}_i)^\top \mid z_i, \mathcal{F}_{i-1})\widehat{\boldsymbol{\Omega}}_i^\top \mid \mathcal{F}_{i-1}) \\
&= \mathbb{E}(\widehat{\boldsymbol{\Omega}}_i\mathbb{E}((x_i - \widehat{m}_i)\nu^2\big(g\big(\langle x_i,\, \widehat{\theta}\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)\big)(x_i - \widehat{m}_i)^\top \mid z_i, \mathcal{F}_{i-1})\widehat{\boldsymbol{\Omega}}_i^\top \mid \mathcal{F}_{i-1}) \\
&+ \mathcal{O}_p(L_\varepsilon L_g D_x\|\omega^* - \widehat{\omega}\|_2\mathbb{E}(\|\widehat{\boldsymbol{\Omega}}_i(x_i - \widehat{m}_i)\|_2^2 \mid \mathcal{F}_{i-1})) \\
&= \mathbf{I}_{d_T} + \mathcal{O}_p\Big(\frac{L_\varepsilon L_g D_x}{m_\varepsilon}\|\omega^* - \widehat{\omega}\|_2\mathbb{E}(\|\widehat{\boldsymbol{\Omega}}_i(x_i - \widehat{m}_i)\|_2^2\nu^2\big(g\big(\langle x_i,\, \widehat{\theta}\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)\big) \mid \mathcal{F}_{i-1})\Big) \\
&= \mathbf{I}_{d_T} + \mathcal{O}_p(d_T\|\omega^* - \widehat{\omega}\|_2) = \mathbf{I}_{d_T} + o_p(1).
\end{aligned}$$

其中，第三个方程由三角不等式以及 ν^2、g 的利普希茨连续性和 ω 的有界性共同得到。第四个方程源于$\widehat{}\boldsymbol{\Omega}_i$ 的定义以及下界假设，$\nu^2(x) \geq m_\varepsilon$. 最后一行使用了 $\widehat{}\boldsymbol{\Omega}_i$ 的定义以及$\widehat{}\,\theta$ 的$\widehat{\eta}$ 相合性假设。注意，最后一行中的 $o_p(1)$ 对所有 i 都相同。由鞅差序列的性质可知，$\mathbb{E}\|\widehat{\mathbb{E}}_{n_2}\widehat{}\boldsymbol{\Omega}_i(x_i - \widehat{m}_i)\varepsilon_i\|_2^2 \to \mathbf{I}_{d_T}$。$\widehat{}$ $\widehat{}$因此，该引理可由关于 $\mathbb{E}_{n_2}\sqrt{n_2}\boldsymbol{\Omega}_i(x_i - \widehat{m}_i)\varepsilon_i$ 的鞅中心极限定理推出，剩下的工作是验证林德伯格条件。 —

林德伯格条件 我们首先对 $\|\|\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\varepsilon_i^2(x_i-\widehat{m}_i)^\top\widehat{\mathbf{\Omega}}_i^\top\|\|_{\mathrm{op}}$ 进行界定。具体而言，

$$\|\|\varepsilon_i^2\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)(x_i-\widehat{m}_i)^\top\widehat{\mathbf{\Omega}}_i^\top\|\|_{\mathrm{op}}\le\varepsilon_i^2\|\|\widehat{\mathbf{\Omega}}_i\|\|_{\mathrm{op}}^2\|(x_i-m_i^*)\|_2^2\le\frac{4}{\widetilde{c}_i}\varepsilon_i^2,$$

其中 $\widetilde{c}_i=\widetilde{c}_0/i^{2t}$ 和 $\widetilde{c}_0=m_\varepsilon c_0/(d_T+2)$。最后一个不等式由引理 9 以及 $\|x_i-\widehat{m}_i\|_2\le\|x_i\|_2+\|\widehat{m}_i\|_2\le 2$ 这一事实得出。因此 $\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\varepsilon_i^2(x_i-\widehat{m}_i)^\top\widehat{\mathbf{\Omega}}_i^\top\preceq 4\varepsilon_i^2\mathbf{I}_{d_T}/\widetilde{c}_i$，且对任意 $\varepsilon>0$,

$$\begin{aligned}0&\preceq\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\mathbb{E}(\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\varepsilon_i^2(x_i-\widehat{m}_i)^\top\widehat{\mathbf{\Omega}}_i^\top 1_{\{\|\|\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\varepsilon_i^2(x_i-\widehat{m}_i)^\top\widehat{\mathbf{\Omega}}_i^\top\|\|_{\mathrm{op}}>\varepsilon n\}}\mid\mathcal{F}_{i-1})\\&\preceq\lim_{n\to\infty}\frac{4}{n}\sum_{i=1}^n\frac{1}{\widetilde{c}_i}\mathbb{E}(\varepsilon_i^2 1_{\{\varepsilon_i^2\ge\varepsilon n\widetilde{c}_i/4\}}\mid\mathcal{F}_{i-1})\mathbf{I}_{d_T}.\end{aligned}\tag{97}$$

由于 ε_i 是次高斯随机变量（在给定 x_i，z_i，$\mathcal{F}_{i-1}$的条件下），且几乎必然具有相同的参数 ν，因此 ε_i^2 是具有共同参数的次指数随机变量。因此，存在某个常数 $K_1>0$，它依赖于 ν，使得$\mathbb{P}(\varepsilon_i^2\ge s)\le 2\exp(-s/K_1)$，于是

$$\mathbb{E}(\varepsilon_i^2 1_{\{\varepsilon_i^2\ge\varepsilon n\widetilde{c}_i/4\}})=\int_{\varepsilon n\widetilde{c}_i/4}^\infty\mathbb{P}(\varepsilon_i^2\ge s)ds\le 2\int_{\varepsilon n\widetilde{c}_i/4}^\infty\exp(-s/K_1)ds=2K_1\exp\Big(\frac{-\varepsilon n\widetilde{c}_i}{4K_1}\Big).$$

将其代入公式（97），对于任意 $t\in(0,\ 1/2)$，有

$$\begin{aligned}&\lim_{n\to\infty}\frac{4}{n}\sum_{i=1}^n\frac{1}{\widetilde{c}_i}\mathbb{E}(\varepsilon_i^2 1_{\{\varepsilon_i^2\ge\varepsilon n\widetilde{c}_i/4\}}\mid\mathcal{F}_{i-1})\\&\lesssim\frac{\widetilde{c}_0K_1}{n}\sum_{i=1}^n i^{2t}\exp\Big(\frac{-\varepsilon n\widetilde{c}_0}{4K_1i^{2t}}\Big)\\&\le\widetilde{c}_0K_1n^{2t}\exp\Big(\frac{-\varepsilon n^{1-2t}\widetilde{c}_0}{4K_1}\Big)\to 0.\end{aligned}$$

因此，林德伯格条件得到满足，证明完毕。 □

引理 11。 在定理 *3* 的假设下，并且假设 $\|\theta-\theta^*\|_2=o_p(1)$，有

$$\begin{aligned}\|\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)\widehat{m}_i^\top(\widetilde{\theta}-\theta^*)\|_2\xrightarrow{p}0\\\|\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)z_i^\top(\widehat{\eta}-\eta^*)\|_2\xrightarrow{p}0\end{aligned}$$

证明。 由于 $\|\widehat{\eta}-\eta^*\|_2=o_p(n^{-1/4})$、$\|\widetilde{\theta}-\theta^*\|_2=o_p(1)$ 是相合的，只需证明

$$\|\|\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)\widehat{m}_i^\top\|\|_F=\mathcal{O}_p(1)\tag{98}$$

$$\|\|\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)z_i^\top\|\|_F=\mathcal{O}_p(1).\tag{99}$$

由于根据 $\widehat{m}_i$ 的定义，$\mathbb{E}(\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)\mid z_i,\mathcal{F}_{i-1})=0$，因此可直接得出 $\{\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)\widehat{m}_i^\top\}_{i=1}^n$ 是一个鞅差序列。注意，

$$\mathbb{E}\|\|\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)\widehat{m}_i^\top\|\|_F^2\le L_g^2\mathbb{E}\|\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\|_2^2\|\widehat{m}_i^\top\|_2^2\lesssim\frac{L_g^2d_T}{m_\varepsilon}=\mathcal{O}(1),\tag{100}$$

其中，第一个不等式利用了 $|g'| \le L_g$ 这一事实，而这由广义线性模型的标准假设所蕴含。第二个不等式由 $\|\widehat{m}_i\|_2, \|x_i\|_2 \le 1$ 得出，并且，

$$\begin{aligned}\mathbb{E}\|\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\|_2^2 &= \mathbb{E}\|\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\nu^2\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)^{1/2}/\nu^2\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)^{1/2}\|_2^2\\ &\le \frac{1}{m_\varepsilon}\mathbb{E}\|\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\nu^2\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)^{1/2}\|_2^2\\ &= \frac{1}{m_\varepsilon}\mathbb{E}\,\mathrm{tr}(\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\nu^2\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)(x_i-\widehat{m}_i)^\top\widehat{\mathbf{\Omega}}_i) = \frac{d_T}{m_\varepsilon} = \mathcal{O}(1),\end{aligned} \tag{101}$$

其中第二行使用了$\widehat{\mathbf{\Omega}}_i$ 和 $\widehat{m}_i$的定义。界 (100) 立即推出界 (98)。由于我们假设 $\|z_i\|_2$ 有界，界 (99) 可通过与上述类似的论证得到，只需将 $\widehat{m}_i^\top$替换为 $z_i^\top$。

□

引理 12。 除定理 *3* 的假设外，进一步假设 $\|\theta-\theta^*\|_2 = o_p(n^{-t})$。则有

$$\begin{aligned}&\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)(Q_3+Q_4)\\ &= o_p(1)+o_p(\|\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)(x_i-\widehat{m}_i)^\top(\widetilde{\theta}-\theta^*)\|_2).\end{aligned}$$

证明。 根据假设， g' 是 L_g-利普希茨的，因此可得 $|g''| \le L_{g'}$。于是，有

$$\begin{aligned}&\sqrt{n_2}\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)(Q_3+Q_4)\\ &= \frac{\sqrt{n_2}}{2}\widehat{\mathbb{E}}_{n_2}\int_0^1\int_0^1 g''\big(\langle x_i,\widehat{\theta}+r_1r_2(\widetilde{\theta}-\widehat{\theta})\rangle+\langle z_i,\widehat{\eta}\rangle\big)\big|\langle x_i,\widetilde{\theta}-\widehat{\theta}\rangle\big|^2dr_1dr_2\\ &- \frac{\sqrt{n_2}}{2}\widehat{\mathbb{E}}_{n_2}\int_0^1\int_0^1\Big\{g''\big(\langle x_i,\widehat{\theta}+r_1r_2(\theta^*-\widehat{\theta})\rangle+\langle z_i,\widehat{\eta}+r_1r_2(\eta^*-\widehat{\eta})\rangle\big)\\ &\qquad\qquad\big|\langle x_i,\theta^*-\widehat{\theta}\rangle+\langle z_i,\eta^*-\widehat{\eta}\rangle\big|^2\Big\}dr_1dr_2\\ &= \mathcal{O}(L_{g'}\sqrt{n_2}\sup_i|\langle x_i,\widetilde{\theta}-\widehat{\theta}\rangle|^2)+\mathcal{O}(L_{g'}|\sqrt{n_2}\sup_i\langle z_i,\widehat{\eta}-\eta^*\rangle|^2)\\ &\qquad+\mathcal{O}(L_{g'}|\sqrt{n_2}\sup_i\langle x_i,\widehat{\theta}-\theta^*\rangle|^2)\\ &= \mathcal{O}_p(\sqrt{n_2}\|\widetilde{\theta}-\theta^*\|_2^2)+\mathcal{O}_p(\sqrt{n_2}\|\widehat{\eta}-\eta^*\|_2^2)+\mathcal{O}_p(\sqrt{n_2}\|\widehat{\theta}-\theta^*\|_2^2)\\ &= o_p(n^{1/2-t}\|\widetilde{\theta}-\theta^*\|_2)+o_p(1),\end{aligned}$$

其中，第二个方程使用了 $|g''| \le L_{g'}$，第三个方程使用了 x_i, z_i 的有界性假设以及 $\|\widetilde{\theta}-\widehat{\theta}\|_2^2 \le 2($ $\|\widetilde{\theta}-\theta^*\|_2^2+\|\widehat{\theta}-\theta^*\|_2^2)$这一事实。最后一行由 n^{-t}与 $\widetilde{\theta}$的一致性以及 $n^{-1/4}$与 $\widehat{\theta},\widehat{\eta}$的一致性推出。将 $\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'(x_i^\top\widehat{\theta}+z_i^\top\widehat{\eta})(x_i-\widehat{m}_i)^\top$ 记为 Z_0，根据引理 13，对于某个$c_{\min}>0$，有 $\mathbb{P}(\sigma_{\min}(Z_0)$ $\ge c_{\min}n^{\delta-t})\to 1$。因此有 $\mathbb{P}(\|\sqrt{n_2}Z_0(\widetilde{\theta}-\theta^*)\|_2 \ge c_{\min}n^{1/2+\delta-t}\|\widetilde{\theta}-\theta^*\|_2)\to 1$，从而我们得到

$$o_p(n^{1/2-t}\|\widetilde{\theta}-\theta^*\|_2) = o_p(c_{\min}n^{1/2+\delta-t}\|\widetilde{\theta}-\theta^*\|_2) = o_p(\sqrt{n_2}\|Z_0(\widetilde{\theta}-\theta^*)\|_2),$$

至此，证明完成。 □

引理 13. 在定理 *3* 的假设下，对于某个 $c_{\min}>0$ ，有

$$\mathbb{P}(\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)(x_i-\widehat{m}_i)^\top) \ge c_{\min}n^{\delta-t}) \to 1, \tag{102}$$

如 $n\to\infty$。

证明。由于向量 y_i, x_i, z_i 在给定 $\mathcal{F}_{n_1}$ 的条件下与 $\theta, \widehat{\eta}$ 相互独立，因此不失一般性，我们可以将 $\theta, \widehat{\eta}$ 视为非随机变量。接下来注意到，U_i：
$\widehat{m}_i\ g'\big(\langle x_i\ \widehat{\theta}\rangle + \langle z_i\ \widehat{\eta}\rangle\big)\ x_i - \widehat{m}_i\ ^\top - \mathbb{E}\ \widehat{}_i\ x_i - \widehat{m}_i\ g'\big(\langle x_i\ \widehat{\theta}\rangle + \langle z_i\ \widehat{\eta}\rangle\big)\ x_i - \widehat{m}_i\ ^\top \mid \mathcal{F}_{i-1}$ $= \quad _i\ x_i -$ $\mathbf{\Omega}(\ ),,(\ )\ (\mathbf{\Omega}(\ ),,(\ )\ )$ 构成一个鞅差序列，并且我们还有

$$\begin{aligned}\mathbb{E}\|\!|U_i|\!\|_F^2 &\le \mathbb{E}\|\!|\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)(x_i-\widehat{m}_i)^\top\|\!|_F^2\\ &\le\ d_T\mathbb{E}\|\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\|_2^2 g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)^2\|(x_i-\widehat{m}_i)^\top\|_2^2\\ &\lesssim L_g^2 d_T\mathbb{E}\|\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)\|_2^2=\mathcal{O}(1),\end{aligned}$$

其中第三个不等式归功于 $|g'|\le L_g$ 和 $\|x_i-\widehat{m}_i\|_2\le 2$，最后一个等式使用了方程（101）。因此，由引理18可得 $\|\!|\mathbb{E}_{n_2}U_i|\!\|_F=\mathcal{O}_p(1/\sqrt{n})$。利用 Weyl定理，我们有

$$\begin{aligned}&\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)(x_i-\widehat{m}_i)^\top)\\ &\ge \sigma_{\min}(\widetilde{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)(x_i-\widehat{m}_i)^\top)-\|\!|\widehat{\mathbb{E}}_{n_2}U_i|\!\|_{\mathrm{op}}\\ &= \sigma_{\min}(\widetilde{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)(x_i-\widehat{m}_i)^\top)+\mathcal{O}_p(n^{-1/2}).\end{aligned}$$

由于 $\delta-t>-1/2$，因此还需证明存在某个 $\widetilde{c}_{\min}$ 使得

$$\mathbb{P}(\sigma_{\min}(\widetilde{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)(x_i-\widehat{m}_i)^\top)\ge\widetilde{c}_{\min}n^{\delta-t})\to 1 \tag{103}$$

回顾我们的记号 $\omega=(\theta,\ \eta)$、$\omega^*=(\theta^*,\ \eta)$ 和 $\widehat{\omega}=(\theta,\ \widehat{\eta})$。令 $\mathbf{D}^p:=\mathrm{diag}\{p_{i1},\dots,p_{id_T}\}$，

$$U_i(\omega):=\mathbb{E}(\mathbf{\Omega}_i(x_i-m_i)g'\big(\langle x_i,\bar{\theta}\rangle+\langle z_i,\eta\rangle\big)(x_i-m_i)^\top\mid\mathcal{F}_{i-1}).$$

我们声称，$\mathbb{E}_{n_2}U_i(\omega)$ 关于 ω 是利普希茨连续的，其常数参数暂记为 L_U，即对于任意 $\omega^a,\ \omega^b\in\Theta\times\mathcal{H}$，有 $\|\!|\widehat{\mathbb{E}}_{n_2}U_i(\omega^a)-\widehat{\mathbb{E}}_{n_2}U_i(\omega^b)|\!\|_{\mathrm{op}}\le L_U\|\omega^a-\omega^b\|_2$。然后再次由 Weyl定理（例如参见 Horn and Johnson 中的定理4.3.1 [27]）可得

$$\begin{aligned}&\sigma_{\min}(\widetilde{\mathbb{E}}_{n_2}\widehat{\mathbf{\Omega}}_i(x_i-\widehat{m}_i)g'\big(\langle x_i,\widehat{\theta}\rangle+\langle z_i,\widehat{\eta}\rangle\big)(x_i-\widehat{m}_i)^\top)\\ &=\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}U_i(\widehat{\omega}))\\ &\ge\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}U_i(\omega^*))-\|\!|\widehat{\mathbb{E}}_{n_2}U_i(\widehat{\omega})-\widehat{\mathbb{E}}_{n_2}U_i(\omega^*)|\!\|_{\mathrm{op}}\\ &\ge\sigma_{\min}(\widehat{\mathbb{E}}_{n_2}U_i(\omega^*))-L_U\|\widehat{\omega}-\omega^*\|_2\\ &\ge m_{\phi,2}n^{\delta-t}+o_p(n^{-t})\end{aligned}$$

其成立的概率趋于 1。这里，在最后一行中，我们使用了关于得分函数的梯度的假设（假设（**EIG**））。因此，通过选择 $\widetilde{c}_{\min}=m_{\phi,\ 2}/2$，方程（103）成立，从而完成证明。

现在，还需证明 $\mathbb{E}_{n_2}U_i(\omega)$ 关于 ω 是利普希茨连续的。根据定义

$$\begin{aligned}\widehat{\mathbb{E}}_{n_2}U_i(\omega)&=\mathbf{\Omega}_i\mathbb{E}((x_i-m_i)g'\big(\langle x_i,\bar{\theta}\rangle+\langle z_i,\eta\rangle\big)(x_i-m_i)^\top\mid\mathcal{F}_{i-1})\\ &=\mathbf{\Omega}_i\mathbf{D}_p\mathbb{E}(\mathbb{E}(\mathbf{D}_p^{-1}(x_i-m_i)g'\big(\langle x_i,\bar{\theta}\rangle+\langle z_i,\eta\rangle\big)(x_i-m_i)^\top\mid z_i,\mathcal{F}_{i-1})\mid\mathcal{F}_{i-1}).\end{aligned}$$

在引理 16 中，我们将证明 $\mathbf{\Omega}_i\mathbf{D}^p$ 关于 ω 有界且满足利普希茨条件。因此，只需证明 $\mathbb{E}(\mathbf{D}_p^{-1}(x_i-m_i)g'\big(\langle x_i,\theta\rangle+\langle z_i,\eta\rangle\big)(x_i-m_i)^\top\mid z_i,\mathcal{F}_{i-1})$ 关于 ω 有界且满足利普希茨条件，因为

两个有界 Lipschitz 函数的乘积仍然有界且满足利普希茨条件。事实上，这个量可以直接计算。具体而言，我们有

$$\begin{aligned}
&\mathbb{E}(\mathbf{D}_p^{-1}(x_i - m_i)g'\big(\langle x_i,\, \bar{\theta}\rangle + \langle z_i,\, \eta\rangle\big)(x_i - m_i)^\top \mid z_i, \mathcal{F}_{i-1}) \\
&= \begin{pmatrix} g_1' & 0 & 0 & \cdots & 0 \\ 0 & g_2' & 0 & \cdots & 0 \\ 0 & 0 & g_3' & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & g_{d_T}' \end{pmatrix} - \begin{pmatrix} g_1' \\ g_2' \\ \vdots \\ g_{d_T}' \end{pmatrix} m_i^\top - \bar{m}_i \begin{pmatrix} p_1 g_1' \\ p_2 g_2' \\ \vdots \\ p_{d_T} g_{d_T}' \end{pmatrix}^\top + (\sum_{j=0}^{d_T} p_j g_j')\bar{m}_i m_i^\top, \qquad (104)
\end{aligned}$$

其中 $\bar{m}_i := \mathbf{D}_p^{-1} m_i$, $g_j' := g'(\bar{\theta_j} + z_i^\top \eta)$（这里我们还定义了 $\theta_0 = 0$）, $\bar{}$ 以及 $p_j := p_{ij}$ 对于 $j = 0, \ldots, d_T$。根据我们对 g' 的假设、 m_i 的定义以及引理15的证明可立即得出，方程 (104) 中的矩阵关于 ω 有界且满足利普希茨条件。证明完毕。 □

引理 14（经验误差）。 在定理 *3* 的假设下，我们有

$$\sup_{\theta\in\Theta} \|(\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2})\phi(\theta, \widehat{\omega})\|_2 = \mathcal{O}_p(\frac{\log n}{\sqrt{n}}).$$

证明。 由于 $\widehat{\omega} \in \mathcal{F}_{n_1} \in \mathcal{F}_{i-1}$，在给定 $\mathcal{F}_{n_1}$ 的条件下，它与 y_i, x_i, z_i 相互独立。因此，我们可以将 $\widehat{\omega}$ 视为固定，并对所有 $\widehat{\omega}$ 证明所需结果。由于 g 满足利普希茨条件，且 x_i, z_i, θ, η 均有界，因此 g 也有界。我们将 $\sup_{x\in D_x,(\theta,\eta)\in\Theta\times\mathcal{H}} |g|$ 记为 M_g。

定义 $d_i(\theta) := \phi(\theta, \widehat{\omega}) - \mathbb{E}(\phi(\theta, \widehat{\omega}) \mid \mathcal{F}_{i-1})$，并将 $d_i(\theta)$ 分解为 $d_{ai} + d_{bi}(\theta)$，其中

$$\begin{aligned}
d_{ai} &:= \widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)\varepsilon_i, \\
d_{bi}(\theta) &:= \widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)[g\big(\langle x_i,\, \theta^*\rangle + \langle z_i,\, \eta^*\rangle\big) - g\big(\langle x_i,\, \theta\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)], \\
&\quad - \mathbb{E}(\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)[g\big(\langle x_i,\, \theta^*\rangle + \langle z_i,\, \eta^*\rangle\big) - g\big(\langle x_i,\, \theta\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)] \mid \mathcal{F}_{i-1}).
\end{aligned}$$

只需证明 $\|\mathbb{E}_{n_2} d_{ai}\|_2 = \mathcal{O}_p(\log n/\sqrt{n})$ 和 $\sup_{\theta\in\Theta} \|\mathbb{E}_{n_2} d_{bi}(\theta)\|_2 = \mathcal{O}_p(\log n/\sqrt{n})$。注意，对于任意 $\theta \in \Theta$， $d_i(\theta)$ d_{ai}， $d_{bi}(\theta)$ 都是鞅差序列。此外，

$$\begin{aligned}
\mathbb{E}(\|d_{ai}\|_2^2 \mid \mathcal{F}_{i-1}) &= \mathbb{E}(\|\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)\varepsilon_i\|_2^2 \mid \mathcal{F}_{i-1}) \\
&= \mathbb{E}(\|\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)\|_2^2 \nu^2\big(g\big(\langle x_i,\, \theta^*\rangle + \langle z_i,\, \eta^*\rangle\big)\big) \mid \mathcal{F}_{i-1}) \\
&\le \mathbb{E}(\|\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)\|_2^2 \mid \mathcal{F}_{i-1}) M_\varepsilon = \mathcal{O}(1),
\end{aligned}$$

其中，第二行是通过计算给定 x_i， z_i 条件下的期望得到的，最后一个等式则用到了方程 (101)。因此，由引理18可立即得出 $\|\mathbb{E}_{n_2} d_{ai}\|_2 = \mathcal{O}_p(1/\sqrt{n}) = \mathcal{O}_p(\log n/\sqrt{n})$。

类似地，我们有

$$\begin{aligned}
\mathbb{E}(\|d_{bi}(\theta)\|_2^2 \mid \mathcal{F}_{i-1}) &\le \mathbb{E}(\|\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)[g\big(\langle x_i,\, \theta^*\rangle + \langle z_i,\, \eta^*\rangle\big) - g\big(\langle x_i,\, \theta\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)]\|_2^2 \mid \mathcal{F}_{i-1}) \\
&\lesssim \mathbb{E}(\|\widehat{\mathbf{\Omega}}_i(x_i - \widehat{m}_i)\|_2^2 \mid \mathcal{F}_{i-1}) M_g^2 = \mathcal{O}(1),
\end{aligned}$$

and

$$\begin{aligned}
\|d_{bi}(\theta)\|_2 &\le \|\widehat{\mathbf{\Omega}}_i\|_{\mathrm{op}} \|(x_i - \widehat{m}_i)\|_2 |g\big(\langle x_i,\, \theta^*\rangle + \langle z_i,\, \eta^*\rangle\big) - g\big(\langle x_i,\, \theta\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)| \\
&\quad + \mathbb{E}(\|\widehat{\mathbf{\Omega}}_i\|_{\mathrm{op}} \|(x_i - \widehat{m}_i)\|_2 \big|g\big(\langle x_i,\, \theta^*\rangle + \langle z_i,\, \eta^*\rangle\big) - g\big(\langle x_i,\, \theta\rangle + \langle z_i,\, \widehat{\eta}\rangle\big)\big| \mid \mathcal{F}_{i-1}) \\
&\lesssim M_g i^{t-\delta},
\end{aligned}$$

其中，在最后一行中我们用到了 $|g| \le M_g$、$\|x_i - \widehat{m}_i\|_2 \le 2$ 和引理 9。由于 $\|d_{bi}(\theta)\|_2$ 的界为 $i^{t-\delta}$，且其方差有界于某个常数，因此存在一些常数σ_b^2，b_b，使得对于每个分量 $d_{bi}^j(j = 1, 2, \ldots\ldots, d_T)$，$d_{bi}^j(\theta)$ 是一个参数为 $(\sigma_b^2, b_b n^{t-\delta})$ 的伯恩斯坦型随机变量。因此，由（例如，.....中的命题 2.10）可知

Wainwrigh t [61]) 其中 $\mathbb{E}(e^{\lambda d_{bi}^j(\theta)} \mid \mathcal{F}_{i-1}) \le e^{\frac{\lambda^2\sigma_b^2/2}{1-b_b n^{t-\delta}|\lambda|}}$ 对所有 $|\lambda| < 1/(b_b n^{t-\delta})$。这意味着 s

$$\mathbb{E}e^{\lambda\widehat{\mathbb{E}}_{n_2} d_{bi}^j(\theta)} \le \mathbb{E}\prod_{i=n_1+1}^{n}\mathbb{E}(e^{\lambda d_{bi}^j(\theta)/n_2} \mid \mathcal{F}_{i-1}) \le e^{\frac{\lambda^2\sigma_b^2/2}{n_2(1-\widetilde{b_b}n^{t-\delta-1}|\lambda|)}}$$

对所有 $|\lambda| < n_2^{1+\delta-t}/\widetilde{b_b}$，其中 $\widetilde{b_b}$ 是某个取决于 b_b 以及 n_2与 n之比的常数。

令 $\mathcal{C}(\varepsilon)$ 为 $\|\cdot\|_2$ 中 Θ 的一个 ε-覆盖。从标准结果可知，我们可以找到这样一个集合，使得$|\mathcal{C}(\varepsilon)| \lesssim \frac{1}{\varepsilon^{d_T}}$。取$\varepsilon = 1/n_2$，可得 $|\mathcal{C}(1/n_2)| \lesssim n_2^{d_T}$。对于任意 $\theta \in \Theta$，令 $\pi(\theta)$ 表示 $\mathcal{C}(1/n_2)$ 中满足 $\|\theta - \pi(\theta)\|_2 < 1/n_2$ 的一个点。使用离散化论证，我们得到

$$\begin{aligned}\sup_{\theta\in\Theta}|\widehat{\mathbb{E}}_{n_2} d_{bi}^j(\theta)| &\le \sup_{\theta\in\Theta}[|\widehat{\mathbb{E}}_{n_2} d_{bi}^j(\pi(\theta))| + |\widehat{\mathbb{E}}_{n_2}[d_{bi}^j(\pi(\theta)) - d_{bi}^j(\theta)]|] \\ &\le \sup_{\theta\in\mathcal{C}(1/n_2)}|\widehat{\mathbb{E}}_{n_2} d_{bi}^j(\theta)| + \sup_{\|\theta^a-\theta^b\|_2\le 1/n_2}|\widehat{\mathbb{E}}_{n_2}[d_{bi}^j(\theta^a) - d_{bi}^j(\theta^b)]| \qquad (105)\end{aligned}$$

对于方程（105）中的第一项，我们有

$$\begin{aligned}\mathbb{E}\sup_{\theta\in\mathcal{C}(1/n_2)}|\widehat{\mathbb{E}}_{n_2} d_{bi}^j(\theta)| &\le \log(\mathbb{E}e^{\lambda\sup_{\theta\in\mathcal{C}(1/n_2)}|\widehat{\mathbb{E}}_{n_2} d_{bi}^j(\theta)|})/\lambda \\ &\le \log(\mathbb{E}e^{\lambda\sup_{\theta\in\mathcal{C}(1/n_2)}\widehat{\mathbb{E}}_{n_2} d_{bi}^j(\theta)} + e^{-\lambda\sup_{\theta\in\mathcal{C}(1/n_2)}\widehat{\mathbb{E}}_{n_2} d_{bi}^j(\theta)})/\lambda \\ &\le \log\left(2|\mathcal{C}(1/n_2)|e^{\frac{\lambda^2\sigma_b^2/2}{n_2(1-\widetilde{b_b}n^{t-\delta-1}|\lambda|)}}\right)/\lambda \\ &\lesssim d_T\log n_2/\lambda + \frac{\lambda\sigma_b^2/2}{n_2(1-\widetilde{b_b}n^{t-\delta-1}|\lambda|)}\end{aligned}$$

对所有 $|\lambda| < n_2^{1+\delta-t}/\widetilde{b_b}$均成立。取 $\lambda = \sqrt{n_2} < n_2^{1+\delta-t}/\widetilde{b_b}$可得

$$\mathbb{E}\sup_{\theta\in\mathcal{C}(1/n_2)}|\widehat{\mathbb{E}}_{n_2} d_{bi}^j(\theta)| \lesssim d_T\frac{\log n_2}{\sqrt{n_2}} + \frac{1}{\sqrt{n_2}} \lesssim d_T\frac{\log n_2}{\sqrt{n_2}}.$$

因此，我们证明了 $\sup_{\theta\in\mathcal{C}(1/n_2)}|\widehat{\mathbb{E}}_{n_2} d_{bi}^j(\theta)| = \mathcal{O}_p(\log n/\sqrt{n})$。对于离散化误差（方程（105）中的第二项），根据 d_{bi} 的定义，我们得到

$$\begin{aligned}&\left|\widehat{\mathbb{E}}_{n_2}[d_{bi}^j(\theta^a) - d_{bi}^j(\theta^b)]\right| \\ &= \left|(\widehat{\mathbb{E}}_{n_2} - \widetilde{\mathbb{E}}_{n_2})\widehat{\mathbf{\Omega}}_{ij}.(x_i - \widehat{m}_i)\left[g\big(\langle x_i, \theta^b\rangle + \langle z_i, \widehat{\eta}\rangle\big) - g\big(\langle x_i, \theta^a\rangle + \langle z_i, \widehat{\eta}\rangle\big)\right]\right| \\ &\le (\widehat{\mathbb{E}}_{n_2} + \widetilde{\mathbb{E}}_{n_2})|||\widehat{\mathbf{\Omega}}_i|||_{\mathrm{op}}\|(x_i - \widehat{m}_i)\|_2\left|g\big(\langle x_i, \theta^b\rangle + \langle z_i, \widehat{\eta}\rangle\big) - g\big(\langle x_i, \theta^a\rangle + \langle z_i, \widehat{\eta}\rangle\big)\right| \\ &\le (\widehat{\mathbb{E}}_{n_2} + \widetilde{\mathbb{E}}_{n_2})|||\widehat{\mathbf{\Omega}}_i|||_{\mathrm{op}}\|(x_i - \widehat{m}_i)\|_2 L_g D_x\|\theta^a - \theta^b\|_2 \\ &\lesssim (\widehat{\mathbb{E}}_{n_2} + \widetilde{\mathbb{E}}_{n_2})i^{t-\delta}\|\theta^a - \theta^b\|_2 \le n^{t-\delta}\|\theta^a - \theta^b\|_2,\end{aligned}$$

其中，第四行用到了 g 的利普希茨连续性，以及 $\|x_i - \widehat{m}_i\|_2 \le 2$ 和$\|\widehat{\mathbf{\Omega}}_i\|_2 \lesssim i^{t-\delta}$ 这一事实。因此，我们得到界

$$\sup_{\|\theta^a-\theta^b\|_2\le 1/n_2}\left|\widehat{\mathbb{E}}_{n_2}[d_{bi}^j(\theta^a) - d_{bi}^j(\theta^b)]\right| \lesssim n^{t-\delta}\|\theta^a - \theta^b\|_2 \le n^{t-\delta}n_2^{-1} = \mathcal{O}(n^{t-\delta-1}) = o(n^{-1/2}).$$

综合各部分可得 $\sup_{\theta\in\Theta}|\widehat{\mathbb{E}}_{n_2}d_{bi}^{j}(\theta)| = \mathcal{O}_p(\log n/\sqrt{n})$，因此 $\sup_{\theta\in\Theta}|\mathbb{E}_{n_2}d_{bi}(\theta)| = \mathcal{O}_p(\log n/\sqrt{n})$。

□

引理15（$\mathbb{E}_{n_2}\phi, \mathbb{E}_{n_2}\partial_\omega\phi$的利普希茨连续性）。在定理 *3* 中给出的假设下，*rem 3*, $\mathbb{E}_{n_2}\phi_i(\theta,\omega)$ 和 $\mathbb{E}_{n_2}\partial_\omega\phi_i(\theta,\omega)$ 关于 (θ,ω) 一致*Lipschitz*，其参数为 $L_{\phi,1}, L_{\phi,2} > 0$，且仅依赖于定理 *3* 中的常数。

证明。根据 $\mathbb{E}_{n_2}$ 的定义，只需证明 $\mathbb{E}(\phi_i(\theta,\omega) \mid \mathcal{F}_{i-1})$ 和 $\mathbb{E}(\partial_\omega\phi_i(\theta,\omega) \mid \mathcal{F}_{i-1})$ 对所有 i 都是一致Lipschitz。

$\mathbb{E}(\phi_i(\theta,\omega) \mid \mathcal{F}_{i-1})$ 的利普希茨连续性

将 ϕ_i 的定义代入 $\mathbb{E}(\phi_i(\theta,\omega) \mid \mathcal{F}_{i-1})$，得到：

$$\mathbb{E}(\phi_i(\theta,\omega) \mid \mathcal{F}_{i-1}) = \mathbb{E}(\mathbf{\Omega}_i(x_i - m_i)\big(y_i - g\big(\langle x_i,\theta\rangle + \langle z_i,\eta\rangle\big)\big) \mid \mathcal{F}_{i-1}),$$

其中

$$\begin{aligned}
m_i &= m_i(z_i,\mathcal{F}_{i-1}) \equiv \mathbb{E}(x_i g'\big(\langle x_i,\bar\theta\rangle + \langle z_i,\eta\rangle\big) \mid z_i,\mathcal{F}_{i-1})[\mathbb{E}(g'\big(\langle x_i,\bar\theta\rangle + \langle z_i,\eta\rangle\big) \mid z_i,\mathcal{F}_{i-1})]^{-1},\\
\mathbf{\Omega}_i &= \mathbf{\Omega}_i(z_i,\mathcal{F}_{i-1}) \equiv [\mathbb{E}(\varepsilon_i^2(x_i - m_i(z_i,\mathcal{F}_{i-1}))(x_i - m_i(z_i,\mathcal{F}_{i-1}))^\top \mid z_i,\mathcal{F}_{i-1})]^{-1/2}\\
&= [\mathbb{E}(\nu^2\big(g\big(\langle x_i,\bar\theta\rangle + \langle z_i,\eta\rangle\big)\big)(x_i - m_i(z_i,\mathcal{F}_{i-1}))(x_i - m_i(z_i,\mathcal{F}_{i-1}))^\top \mid z_i,\mathcal{F}_{i-1})]^{-1/2}.
\end{aligned}$$

这里我们指出，m_i、$\mathbf{\Omega}_i$ 都依赖于 $\omega = (\theta,\eta)$。由于对*L*-利普希茨函数取期望后仍是 *L*-利普希茨的，因此只需证明 $\mathbb{E}(\phi_i(\theta,\omega) \mid z_i,\mathcal{F}_{i-1})$ 关于 (θ,ω) 是利普希茨连续的，且其参数与 i 和 z_i 无关。从现在起，在本证明中，我们用「关于 (θ,ω) 是利普希茨连续」来表示「关于 (θ,ω) 是利普希茨连续，且其参数与 i 无关」。等价地，只需证明

$$\begin{aligned}
&\mathbb{E}_{x_i,\varepsilon_i}\mathbf{\Omega}_i(x_i - m_i(z_i,\mathcal{F}_{i-1}))\big(y_i - g\big(\langle x_i,\theta\rangle + \langle z_i,\eta\rangle\big)\big)\\
&= \mathbf{\Omega}_i\mathbb{E}_{x_i}(x_i - m_i(z_i,\mathcal{F}_{i-1}))\big(g\big(\langle x_i,\theta^*\rangle + \langle z_i,\eta^*\rangle\big) - g\big(\langle x_i,\theta\rangle + \langle z_i,\eta\rangle\big)\big)
\end{aligned}$$

是利普希茨连续的。这里我们滥用记号 $\mathbb{E}_{x_i,\varepsilon_i},\mathbb{E}_{x_i}$ 来表示在给定 $z_i,\mathcal{F}_{i-1}$ 条件下的条件期望。采用简写记号 $p_j, m_j, \bar\varepsilon_j$ 分别表示 $p_{ij}(z_i,\mathcal{F}_{i-1}), m_{ij}(z_i,\mathcal{F}_{i-1})$，$\nu^2(g\big(\theta_j + \langle z_i,\eta\rangle\big)$ $j = 0,1,\ldots,d_T$（我们还定义 $\theta_0 := 0$）。由于条件期望是对 x_i 和 ε_i 取的，因此可将 $p_j, m_j, \bar\varepsilon_j$ 视为给定 $z_i,\mathcal{F}_{i-1}$ 后的固定量。此外，p_j 不依赖于参数 (θ,ω)，而 $m_j, \bar\varepsilon_j$ 是 ω 的函数。定义 $\mathbf{D}_p = \mathrm{diag}\{p_1, p_2, \ldots, p_{d_T}\}$。经过一些代数计算，我们得到 $m_i(z_i,\mathcal{F}_{i-1})$ 是一个向量，其第 j 个分量等于

$$p_j g'\big(\bar\theta_j + \langle z_i,\eta\rangle\big) / \sum_{k=0}^{d_T} p_k g'\big(\bar\theta_k + \langle z_i,\eta\rangle\big). \tag{106}$$

定义 $\bar m_i \equiv \mathbf{D}_p^{-1} m_i$ 为 m_i 的归一化版本。由于我们假设 $L_g \geq |g'| \geq l_g > 0$，$g'$ 是 $L_{g'}$ 利普希茨函数，且 $\|(x_i^\top,\ z_i^\top)\|_2 \leq D_x$，因此 $g'\big(\theta_j + \langle z_i,\ \eta\rangle\big)$、$\sum_{k=0}^{d_T} p_k g'\big(\bar\theta_k + \langle z_i,\ \eta\rangle\big)$ 均为利普希茨函数，且第二项也有界于 l_g 和 L_g 之间。因此，m_j 和 $\bar m_j$ 在 $(\theta,\ \omega)$ 中均有界且满足利普希茨条件。

此外，可以验证，$\mathbb{E}_{x_i}(x_i - m_i(z_i, \mathcal{F}_{i-1}))\big[g\big(\langle x_i, \theta^*\rangle + \langle z_i, \eta^*\rangle\big) - g\big(\langle x_i, \theta\rangle + \langle z_i, \eta^*\rangle\big)\big]$ 的第 j 个元素等于

$$p_j\big(g\big(\theta_j^* + \langle z_i, \eta^*\rangle\big) - g\big(\theta_j + \langle z_i, \eta\rangle\big)\big) - m_j\mathbb{E}_{x_i}(g\big(\langle x_i, \theta^*\rangle + \langle z_i, \eta^*\rangle\big) - g\big(\langle x_i, \theta\rangle + \langle z_i, \eta\rangle\big))$$
$$= p_j\Big[g\big(\theta_j^* + \langle z_i, \eta^*\rangle\big) - g\big(\theta_j + \langle z_i, \eta\rangle\big) - \bar{m}_j\mathbb{E}_{x_i}[g\big(\langle x_i, \theta^*\rangle + \langle z_i, \eta^*\rangle\big) - g\big(\langle x_i, \theta\rangle + \langle z_i, \eta\rangle\big)]\Big].$$

由于 $g, \bar{m}_j$ 都有界（ g 的有界性由 g 的利普希茨连续性以及 $\Theta \times \mathcal{H}$ 的有界性推出，(x_i, z_i)），且在 (θ, ω) 上满足利普希茨条件，因此第二行方括号内的量显然有界，且在 (θ, ω) 上满足利普希茨条件。因此，$\mathbf{D}_p^{-1}\mathbb{E}_{x_i}(x_i - m_i(z_i, \mathcal{F}_{i-1}))\big[g\big(\langle x_i, \theta^*\rangle + \langle z_i, \eta^*\rangle\big) - g\big(\langle x_i, \theta\rangle + \langle z_i, \eta\rangle\big)\big]$ 有界，且在 (θ, ω) 上满足利普希茨条件。由于引理 16 表明 $\mathbf{\Omega}_i\mathbf{D}_p$ 有界，且在 (θ, ω) 上满足利普希茨条件，因此所需结果成立，因为两个有界 Lipschitz 函数的乘积仍然有界且满足利普希茨条件。

$\mathbb{E}\left(\partial_\omega\phi_i\left(\theta, \omega\right) \mid \mathcal{F}_{i-1}\right)$ 的利普希茨连续性

定义

$$T_1 := \mathbf{D}_p^{-1}\mathbb{E}_{x_i}(x_i - m_i)\big(g\big(\langle x_i, \theta^*\rangle + \langle z_i, \eta^*\rangle\big) - g\big(\langle x_i, \theta\rangle + \langle z_i, \eta\rangle\big)\big)$$
$$T_2 := \mathbf{D}_p^{-1}\partial_\omega m_i\mathbb{E}_{x_i}\big(g\big(\langle x_i, \theta^*\rangle + \langle z_i, \eta^*\rangle\big) - g\big(\langle x_i, \theta\rangle + \langle z_i, \eta\rangle\big)\big)$$
$$T_3 := \mathbf{D}_p^{-1}\mathbb{E}_{x_i}(x_i - m_i)g'(\langle x_i, \theta\rangle + \langle z_i, \eta\rangle)(0_{d_T}, z_i^\top)\partial_\omega\eta.$$

将偏导数的表达式代入 $\mathbb{E}(\partial_\omega\phi_i(\theta, \omega) \mid z_i, \mathcal{F}_{i-1})$，得到

$$\begin{aligned}
&\mathbb{E}(\partial_\omega\phi_i(\theta, \omega) \mid z_i, \mathcal{F}_{i-1}) \\
=\ &\mathbb{E}(\partial_\omega\mathbf{\Omega}_i(x_i - m_i)\big(y_i - g\big(\langle x_i, \theta\rangle + \langle z_i, \eta\rangle\big)\big) \mid z_i, \mathcal{F}_{i-1}) \\
&- \mathbb{E}(\mathbf{\Omega}_i\partial_\omega m_i\big(y_i - g\big(\langle x_i, \theta\rangle + \langle z_i, \eta\rangle\big)\big) \mid z_i, \mathcal{F}_{i-1}) \\
&- \mathbb{E}(\mathbf{\Omega}_i(x_i - m_i)g'\big(\langle x_i, \theta\rangle + \langle z_i, \eta\rangle\big)(z_i^\top, 0_{d_T})\partial_\omega\eta \mid z_i, \mathcal{F}_{i-1}) \\
=\ &\partial_\omega\mathbf{\Omega}_i\mathbf{D}_pT_1 - \mathbf{\Omega}_i\mathbf{D}_pT_2 - \mathbf{\Omega}_i\mathbf{D}_pT_3.
\end{aligned} \tag{107}$$

由于引理 16 表明 $\mathbf{\Omega}_i\mathbf{D}_p$，$\partial_\omega\mathbf{\Omega}_i\mathbf{D}_p$ 有界且关于 (θ, ω) 满足利普希茨连续性，因此只需证明 T_1，T_2，T_3 均有界且关于 (θ, ω) 满足利普希茨连续性。对于 T_1，T_3，经过一些基本的代数运算，我们得到每一项的第 j 个分量

$$T_{1j} = g\big(\theta_j^* + \langle z_i, \eta^*\rangle\big) - g\big(\theta_j + \langle z_i, \eta\rangle\big) - \bar{m}_{ij}\mathbb{E}_{x_i}(g\big(\theta_j^* + \langle z_i, \eta^*\rangle\big) - g\big(\theta_j + \langle z_i, \eta\rangle\big)),$$
$$T_{3j} = \big[g'\big(\theta_j + \langle z_i, \eta\rangle\big) - \bar{m}_{ij}\mathbb{E}_{x_i}g'\big(\langle x_i, \theta\rangle + \langle z_i, \eta\rangle\big)(0_{d_T}, z_i^\top)\big]\partial_\omega\eta.$$

由于函数 g'和 $\bar{m}_i$有界且满足利普希茨连续性，z_i有界且 $\partial_\omega\eta = (\mathbf{0}_{d_N \times d_T}, \mathbf{I}^{d_N})^\top$，因此可直接得出 T_{1j}，T_{3j}关于 (θ, ω) 有界且满足利普希茨连续性。对于 T_2，我们也考虑其第 j 个分量 T_{2j}。分别用 g_k'，g_k''简记 $g'\big(\theta_k + \langle z_i, \eta\rangle\big)$，$g''\big(\theta_k + \langle z_i, \eta\rangle\big)$。由方程 (106) 及一些导数计算可得 $\mathbf{D}_p^{-1}\partial_\eta m_i$ 的第 j 个分量

$$\partial_\eta m_{ij}/p_j = \partial_\eta\bar{m}_{ij} = [(\sum_{k=0}^{d_T} p_k g_k')g_j'' - g_j'(\sum_{k=0}^{d_T} p_k g_k'')]z_i^\top/(\sum_{k=0}^{d_T} p_k g_k')^2.$$

由于对所有 k 都有 $g_k' \geq l_g > 0$，因此可得 $(\sum_{k=0}^{d_T} p_k g_k')^2 \geq l_g^2$。结合假设 g_k' 是利普希茨的，可知 $1/(\sum_{k=0}^{d_T} p_k g_k')^2$ 有界且利普希茨连续。此外，由于 g_k'' 有界且利普希茨连续，并且根据我们的假设 z_i 有界，因此可得 $\partial_\eta\bar{m}_i$

在 $(\theta,\ \omega)$ 中有界且利普希茨连续。由于 g 有界且满足利普希茨连续性，$\mathbb{E}_{x_i}\big(g\big(\langle x_i,\ \theta^*\rangle+\langle z_i,\ \eta^*\rangle\big)-g\big(\langle x_i,\ \theta\rangle+\langle z_i,\ \eta\rangle\big)\big)$ 也有界且利普希茨连续，因此 T_2 在 $(\theta,\ \omega)$ 中有界且利普希茨连续。至此，证明完成。

□

引理 16 （$\mathbf{\Omega}_i\mathbf{D}^p, \partial_\omega\mathbf{\Omega}_i\mathbf{D}^p$的利普希茨连续性）。 在定理 *3* 的假设以及引理*15*中的记号下，$\mathbf{\Omega}_i\mathbf{D}^p$ 和 $\partial_\omega\mathbf{\Omega}_i\mathbf{D}^p$ 在 (θ,ω) 中均有界且利普希茨连续。

证明。 关于 θ 的利普希茨连续性是显然的，因为$\mathbf{\Omega}_i$ 仅依赖于 $\omega=(\theta,\ \eta)$。剩下需要证明其关于 ω 的利普希茨连续性。同样地，如果一个函数关于 ω 是利普希茨的，我们称其为关于 ω 的利普希茨函数，其中利普希茨参数是一个仅依赖于定理 3 中所定义常数、但不依赖于 i的某个常数。定义

$$\begin{aligned}\mathbf{\Sigma}_i &:= \mathbb{E}(\varepsilon_i^2(x_i-m_i)(x_i-m_i)^\top \mid z_i,\mathcal{F}_{i-1})\\ &= \mathbb{E}(\nu^2\big(g\big(\langle x_i,\bar\theta\rangle+\langle z_i,\eta\rangle\big)\big)(x_i-m_i)(x_i-m_i)^\top \mid z_i,\mathcal{F}_{i-1}).\end{aligned}$$

再次说明，$\mathbf{\Sigma}_i$ 隐式地依赖于 ω。由于 $\nu^2(g(x_i^\top\theta+z_i^\top\eta))\ \le\ M_\varepsilon$,$\|x_i\|_2\le 1$, $\|\bar{m_i}\|_2\le 1$，因此有

$$\|\mathbf{\Sigma}_i\|_{\mathrm{op}}\le\mathbb{E}(\nu^2\big(g\big(\langle x_i,\bar\theta\rangle+\langle z_i,\eta\rangle\big)\big)\cdot\|x_i-m_i\|_2^2\mid z_i,\mathcal{F}_{i-1})\le 4M_\varepsilon.$$

经过一些代数计算，我们得到

$$\mathbf{\Sigma}_i=\begin{pmatrix}p_1\bar\varepsilon_1 & 0 & 0 & \cdots & 0\\ 0 & p_2\bar\varepsilon_2 & 0 & \cdots & 0\\ 0 & 0 & p_3\bar\varepsilon_3 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & p_{d_T}\bar\varepsilon_{d_T}\end{pmatrix}-\begin{pmatrix}p_1\bar\varepsilon_1\\ p_2\bar\varepsilon_2\\ \vdots\\ p_{d_T}\bar\varepsilon_{d_T}\end{pmatrix}m_i^\top-m_i\begin{pmatrix}p_1\bar\varepsilon_1\\ p_2\bar\varepsilon_2\\ \vdots\\ p_{d_T}\bar\varepsilon_{d_T}\end{pmatrix}^\top+(\sum_{j=0}^{d_T}p_j\bar\varepsilon_j)m_im_i^\top.$$

M此外，利用 Woodbury恒等式计算 $\mathbf{\Sigma}_i$ 的逆矩阵，可得

$$\begin{aligned}\mathbf{\Omega}_i^2=\mathbf{\Sigma}_i^{-1}&=\mathbf{C}_i+\mathbf{B}_i\mathbf{K}_i\frac{\begin{pmatrix}-p_0\bar\varepsilon_0 & \bar m_0p_0\\ \bar m_0p_0 & \sum_{k=1}^{d_T}p_k\bar m_k^2/\bar\varepsilon_k\end{pmatrix}}{(\sum_{k=1}^{d_T}p_k\bar m_k^2/\bar\varepsilon_k)p_0\bar\varepsilon_0+\bar m_0^2p_0^2}\mathbf{K}_i^\top\mathbf{B}_i\\ &=:\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)\end{aligned}\tag{108}$$

其中 $\mathbf{B}_i\ =\mathrm{diag}\{1/\bar\varepsilon_1,\ldots,1/\bar\varepsilon_{d_T}\}$，$\mathbf{C}_i:=\mathbf{D}_p^{-1}\mathbf{B}_i\ =\mathrm{diag}\{1/(p_1\bar\varepsilon_1)\ldots,1/(p_{d_T}\bar\varepsilon_{d_T})\}$，且$\mathbf{K}_i:=\begin{pmatrix}\bar m_1 & \bar m_2 & \cdots & \bar m_{d_T}\\ \bar\varepsilon_1 & \bar\varepsilon_2 & \cdots & \bar\varepsilon_{d_T}\end{pmatrix}$。由于假设 $p_0\ge\widetilde{c}_0>0$，因此可得

$$(\sum_{k=1}^{d_T}p_k\bar m_k^2/\bar\varepsilon_k)p_0\bar\varepsilon_0+\bar m_0^2p_0^2\ge\bar m_0^2p_0^2\ge(l_g/L_g)^2\tilde c_0^2.$$

因此，$1/[(\sum_{k=1}^{d_T}p_k\bar m_k^2/\bar\varepsilon_k)p_0\bar\varepsilon_0+\bar m_0^2p_0^2]$ 在 ω 中有界且满足利普希茨条件。同理，可以验证 B_i、K_i、$p_0\bar\varepsilon_0$、$\bar m_0p_0$、$\sum_{k=1}^{d_T}p_k\bar m_k^2/\bar\varepsilon_k$均有界且满足利普希茨条件。由此可得，$\mathbf{\Delta}_i(\omega)$ 在 ω 中有界且满足利普希茨条件。遗憾的是，$\mathrm{C}_i(\omega)$ 在 ω 中不一定满足利普希茨条件，因为 p_i 可能没有常数下界。不过，由引理 17 可得，$\sqrt{\mathrm{C}_i(\omega)+\mathbf{\Delta}_i(\omega)}-\sqrt{\mathrm{C}_i(\omega)}$ 在 ω 中有界且满足利普希茨条件。由于 $\mathrm{C}_i(\omega)\mathrm{D}^p=\mathrm{B}_i(\omega)$ 有界且满足利普希茨条件

且在 ω 和 $\|\mathbf{D}_p\|_{\mathrm{op}} \le 1$ 中满足利普希茨条件，因此 $\Omega_i\mathbf{D}_p = \sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)}\mathbf{D}_p = (\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)} - \sqrt{\mathbf{C}_i(\omega)})\mathbf{D}_p + \sqrt{\mathbf{C}_i(\omega)}\mathbf{D}_p$ 在 ω 中有界且满足利普希茨条件。同理，引理 17 表明 $\partial_\omega[\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)} - \sqrt{\mathbf{C}_i(\omega)}]$ 在 ω 中有界且满足利普希茨条件。此外，

$$\partial_\omega \sqrt{\mathbf{C}_i(\omega)}\mathbf{D}_p = \mathrm{diag}\{-\frac{\sqrt{p_1}\bar{\varepsilon}_1'(\omega)}{2\sqrt{\bar{\varepsilon}_1^3}}, \cdots, -\frac{\sqrt{p_{d_T}}\bar{\varepsilon}_{d_T}'(\omega)}{2\sqrt{\bar{\varepsilon}_{d_T}^3}}\}.$$

由于对于所有 k，$p_k \le 1$、$m_\varepsilon \le \bar{\varepsilon}_k \le M_\varepsilon$ 和 $\bar{\varepsilon}_k'$ 关于 ω 都是利普希茨连续的，因此 $\partial_\omega\sqrt{\mathbf{C}_i(\omega)}\mathbf{D}^p$ 有界且关于 ω 是利普希茨连续的。因此，我们得到 $\partial_\omega\mathbf{\Omega}_i\mathbf{D}_p = \partial_\omega[\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)} - \sqrt{\mathbf{C}_i(\omega)}]\mathbf{D}_p + \partial_\omega\sqrt{\mathbf{C}_i(\omega)}\mathbf{D}^p$ 有界且关于 ω 是利普希茨连续的。 □

T以下结果采用引理 15 和 1 中先前引入的记号 6.

引理 17 (利普希茨连续性)。 在定理 *3* 的假设下，以下量

$$\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)} - \sqrt{\mathbf{C}_i(\omega)}, \quad \partial_\omega[\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)} - \sqrt{\mathbf{C}_i(\omega)}]$$

关于 ω 均有界且为利普希茨连续。

证明。 为简化记号，我们省略每个量对 i 的依赖。在本证明中，若某个量可由一个常数界定，且该常数仅依赖于定理 3 中定义的常数而不依赖于 i，则称该量有界。同样地，我们用 $\lesssim$ 表示 $\le$ 在一个常数因子以内相等，该因子可能依赖于定理 3 中定义的量，也可能不依赖。此外，若一个函数关于 ω 的利普希茨参数仅依赖于定理 3 中定义的常数，则称该函数关于 ω 是利普希茨连续的。

F.3.1 $\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)} - \sqrt{\mathbf{C}_i(\omega)}$ 和 $\partial_\omega[\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)} - \sqrt{\mathbf{C}_i(\omega)}]$ 的有界性

根据定义，$\sigma_{\min}(\Omega^2) = \|\mathbf{\Sigma}^{-1}\|_{\mathrm{op}} \ge 1/(4M_\varepsilon)$。结合 $\Omega^2 = \mathbf{C}(\omega) + \mathbf{\Delta}(\omega)$ 有界这一事实，矩阵 $\mathbf{C}(\omega)$ 的最小特征值有一个常数下界，因此存在一个足够大的常数 $c_T > 0$，使得对于某个常数 $c_t > 0$，$c_t\mathbf{I}_{d_T} \preceq \Omega_{i,\mathrm{trun}}^2(c_T)$，其中 $\Omega_{\mathrm{trun}}^2(c_T)$ 是一个与 Ω^2 相同的矩阵，只是将每个对角项 Ω_{kk}^2 替换为 $\Omega_{kk}^2 \wedge c_T$。不失一般性，由于 Ω^2 的非对角项有界，我们可以选择足够大的 c_T，使得 $\|\Omega_{\mathrm{trun}}^2(c_T)\|_{\mathrm{op}} \le \frac{3}{2}c_T$。

现在，定义 $\tilde{\mathbf{C}}(\omega) := \mathrm{diag}\{\mathbf{\Omega}_{11}^2 \vee c_T, \ldots, \mathbf{\Omega}_{d_Td_T}^2 \vee c_T\}$ 以及 $\tilde{\mathbf{\Delta}}(\omega) := \mathbf{\Omega}_{\mathrm{trun}}^2(c_T) - c_T\mathbf{I}_{d_T}$。于是有 $\mathbf{\Omega}^2 = \tilde{\mathbf{C}}(\omega) + \tilde{\mathbf{\Delta}}(\omega)$，并且

$$\|\mathbf{\Delta}(\omega)\|_{\mathrm{op}} \le \text{最大值}\{0.5c_T, c_T - c_t\} \le \text{最大值}\{0.5, (c_T - c_t)/c_T\}\sigma_{\min}(\tilde{\mathbf{C}}(\omega)) =: \gamma\sigma_{\min}(\tilde{\mathbf{C}}(\omega))$$

对于如此 $\tilde{\mathbf{C}}(\omega)$ 常数 $\gamma < 1$。此外，$\mathbf{C}(\omega)\mathbf{D}^p$，$\mathbf{\Delta}(\omega)$ 关于 ω 有界且满足利普希茨连续性。

在 $\tilde{\mathbf{C}}(\omega)$ 处对 $\sqrt{\tilde{\mathbf{C}}(\omega)} + \tilde{\mathbf{\Delta}}(\omega)$ 进行泰勒展开（由于 $\sigma_{\min}(\mathbf{C}(\omega)) > \|\mathbf{\Delta}(\omega)\|_{\mathrm{op}}$，因此可以这样做），我们得到

$$\sqrt{\tilde{\mathbf{C}}(\omega)+\tilde{\mathbf{\Delta}}(\omega)} - \sqrt{\tilde{\mathbf{C}}(\omega)} = \sum_{k=1}^{\infty}\frac{1}{k!}[\nabla^k\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}_i(\omega)],$$

其中 $[\nabla\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}(\omega)] = \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}(\omega)}}\tilde{\mathbf{\Delta}}(\omega)e^{-t\sqrt{\tilde{\mathbf{C}}(\omega)}}dt$，高阶导数通过以下方式递归定义：

$$\begin{aligned}&[\nabla^k\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}(\omega)]\\ &= -\Big[\nabla\tilde{\mathbf{C}}(\omega)\cdot\Big(\sum_{p+q=k-2}\frac{k!}{(p+1)!(q+1)!}[\nabla^{p+1}\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}(\omega)][\nabla^{q+1}\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}(\omega)]\Big)\Big].\end{aligned}$$

根据 Moral and Niclas 的研究结果 [16]（特别参见他们的方程 (4) 和定理1.1的证明），我们建立了 $\|\|\nabla^{k+1}\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}(\omega)\|\|_{\mathrm{op}}\le c_T^{1/2}k!\ \binom{2k}{k}2^{-(2k+1)}\gamma^{k+1/2}$ 对于 $k\ge 0$。此外，定义

$$\mathbf{H}_{k+1}:=\sum_{p+q=k-1}\frac{(k+1)!}{(p+1)!(q+1)!}[\nabla^{p+1}\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}(\omega)][\nabla^{q+1}\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}(\omega)].$$

Then

$$\begin{aligned}\|\|\mathbf{H}_{k+1}\|\|_{\mathrm{op}}&\le c_T(k+1)!\sum_{p+q=k-1}\binom{2p}{p}\binom{2q}{q}2^{-2k}\gamma^k/[(p+1)(q+1)]\\&=c_T\binom{2k}{k}2^{-2k}\gamma^k k!,\end{aligned}$$

第二行由卡特兰数的Segner递推公式 [36] 得出。由于根据斯特林公式有 $\binom{2k}{k}2^{-(2k+1)}\asymp 1/\sqrt{k}$，且 $\gamma<1$，因此我们有

$$\|\|\sum_{k=1}^{\infty}\frac{1}{k!}[\nabla^k\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}(\omega)]\|\|_{\mathrm{op}}\le\sum_{k=0}^{\infty}\frac{1}{(k+1)!}\|\|[\nabla^{k+1}\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}(\omega)]\|\|_{\mathrm{op}}\lesssim c_T^{1/2}\sum_{k=0}^{\infty}k^{-3/2}\gamma^{k+1/2}$$

有一个不依赖于 i 的常数作为其界，因此 $\sqrt{\tilde{\mathbf{C}}(\omega)+\tilde{\mathbf{\Delta}}(\omega)}-\sqrt{\tilde{\mathbf{C}}(\omega)}$ 也有界。事实上，我们有一个更强的结果。注意到

$$\begin{aligned}\|\|\frac{1}{k!}[\nabla^k\tilde{\mathbf{C}}(\omega)\cdot\tilde{\mathbf{\Delta}}(\omega)]\mathbf{D}_p^{-1/2}\|\|_{\mathrm{op}}&=\|\|\frac{1}{k!}\Big[\nabla\tilde{\mathbf{C}}(\omega)\cdot\mathbf{H}_k\Big]\mathbf{D}_p^{-1/2}\|\|_{\mathrm{op}}\\&\le\frac{1}{k!}\int_0^{\infty}\|\|e^{-t\sqrt{\tilde{\mathbf{C}}(\omega)}}\|\|_{\mathrm{op}}\|\|\mathbf{H}_k\|\|_{\mathrm{op}}\|\|e^{-t\sqrt{\tilde{\mathbf{C}}(\omega)}}\mathbf{D}_p^{-1/2}\|\|_{\mathrm{op}}dt\\&\le\sum_{j=1}^{d_T}\frac{1}{k!}\int_0^{\infty}\|\|\mathbf{H}_k\|\|_{\mathrm{op}}e^{-t\sqrt{1/(p_j\bar{\varepsilon}_j)}}(1/p_j)^{-1/2}dt\\&\lesssim\frac{d_T\|\|\mathbf{H}_k\|\|_{\mathrm{op}}}{2k!}\lesssim d_Tc_T\frac{\gamma^{k-1}}{k^{3/2}}.\end{aligned}$$

由泰勒展开可直接得出 $[\sqrt{\tilde{\mathbf{C}}(\omega)+\tilde{\mathbf{\Delta}}(\omega)}-\sqrt{\tilde{\mathbf{C}}(\omega)}]\mathbf{D}_p^{-1/2}$ 有界。因此，对于任意有界函数 F，$[\sqrt{\tilde{\mathbf{C}}(\omega)+\tilde{\mathbf{\Delta}}(\omega)}-\sqrt{\tilde{\mathbf{C}}(\omega)}]\mathbf{D}_p^{-1/2}F(\omega)$ 有界。

$\partial_\omega[\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)}-\sqrt{\mathbf{C}_i(\omega)}]$ 的有界性直接由 $\frac{\partial\omega}{\partial x}$ 的有界性以及 $\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)}-\sqrt{\mathbf{C}_i(\omega)}$ 的利普希茨连续性得出，下面将证明这一点。

F.3.2 $\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)}-\sqrt{\mathbf{C}_i(\omega)}$ 和 $\partial_\omega[\sqrt{\mathbf{C}_i(\omega)+\mathbf{\Delta}_i(\omega)}-\sqrt{\mathbf{C}_i(\omega)}]$ 的利普希茨连续性

注意，$[\tilde{\mathbf{C}}(\omega)\ \tilde{\mathbf{\Delta}}(\omega)]$ 依赖于 ω，其依赖关系通过 $x_i^\top\theta+z_i^\top\eta$ 体现，并且我们假设 $\|(x_i^\top,z_i^\top)\|_2\le D_x$。为简化记号，我们用 x 表示标量 $\langle x_i,\theta\rangle+\langle z_i,\eta\rangle$，并定义函数

$$d(x):=\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}-\sqrt{\tilde{\mathbf{C}}(x)}\tag{109}$$

为了证明所声称的利普希茨性质，现在只需证明函数 $d'(x)$ $d''(x)$ 都被某个常数界定。（注意，我们仍有 $\tilde{\mathbf{C}}(x)\mathbf{D}_p$，$\tilde{\mathbf{\Delta}}(x)$ 在 x 和 $|x|\le M_\omega D_x$ 上是利普希茨函数。）

F.3.3 $d'(x)$ 的有界性

利用一阶导数公式，我们得到

$$
\begin{aligned}
&\Big|\Big|\Big|[\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}-\sqrt{\tilde{\mathbf{C}}(x)}]'\Big|\Big|\Big|_{\mathrm{op}}\\
&=\Big|\Big|\Big|\int e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}(C'(x)+\tilde{\mathbf{\Delta}}'(x))e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}dt-\int e^{-t\sqrt{\tilde{\mathbf{C}}(x)}}C'(x)e^{-t\sqrt{\tilde{\mathbf{C}}(x)}}dt\Big|\Big|\Big|_{\mathrm{op}}\\
&\le\Big|\Big|\Big|\int e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}\tilde{\mathbf{\Delta}}'(x)e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}dt\Big|\Big|\Big|_{\mathrm{op}}\\
&+2\Big|\Big|\Big|\int(e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}-e^{-t\sqrt{\tilde{\mathbf{C}}(x)}})C'(x)e^{-t\sqrt{\tilde{\mathbf{C}}(x)}}dt\Big|\Big|\Big|_{\mathrm{op}}\\
&+\Big|\Big|\Big|\int(e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}-e^{-t\sqrt{\tilde{\mathbf{C}}(x)}})C'(x)(e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}-e^{-t\sqrt{\tilde{\mathbf{C}}(x)}})dt\Big|\Big|\Big|_{\mathrm{op}}\\
&=:\ T_1+2T_2+T_3.
\end{aligned}
$$

现在分别对项 T_1，T_2 和 T_3 进行界定。对于 T_1，有：

$$
T_1\le\int_0^\infty|||e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}|||_{\mathrm{op}}|||\tilde{\mathbf{\Delta}}'(x)|||_{\mathrm{op}}|||e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}|||_{\mathrm{op}}dt\lesssim\frac{|||\tilde{\mathbf{\Delta}}'(x)|||_{\mathrm{op}}}{\sigma_{\min}(\mathbf{\Omega}^2)},
$$

根据我们的假设，它是有界的。

对于 T_2 和 T_3，注意到

$$
\begin{aligned}
&|||(e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}-e^{-t\sqrt{\tilde{\mathbf{C}}(x)}})\tilde{\mathbf{C}}'(x)^{1/2}|||_{\mathrm{op}}\\
&=|||\int_0^1 e^{-st\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}[\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}-\sqrt{\tilde{\mathbf{C}}(x)}]e^{-(1-s)t\sqrt{\tilde{\mathbf{C}}(x)}})ds\tilde{\mathbf{C}}'(x)^{1/2}|||_{\mathrm{op}}\\
&\le\int_0^1\Big\{|||e^{-st\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}}|||_{\mathrm{op}}|||[\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}-\sqrt{\tilde{\mathbf{C}}(x)}]\mathbf{D}_p^{-1/2}|||_{\mathrm{op}} \qquad (110)\\
&\qquad\qquad\times|||\mathbf{D}_p^{1/2}e^{-(1-s)t\sqrt{\tilde{\mathbf{C}}(x)}}\tilde{\mathbf{C}}'(x)^{1/2}|||_{\mathrm{op}}\Big\}ds\\
&\lesssim|||[\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}-\sqrt{\tilde{\mathbf{C}}(x)}]\mathbf{D}_p^{-1/2}|||_{\mathrm{op}}e^{-t\min\{\sqrt{c_T},\sqrt{\sigma_{\min}(\mathbf{\Omega}^2)}\}}=:v_1e^{-v_2t}, \qquad (111)
\end{aligned}
$$

其中，第一个方程源于分解 $e^{-A}-e^{-B}=\int_0^1 e^{-sA}(B-A)e^{-(1-s)B}ds$，最后一行则由$\mathbf{D}_p^{1/2}\tilde{}$ $\mathbf{C}'(x)^{1/2}$ 有界这一事实得出。此外，

$$
\begin{aligned}
&|||e^{-t\sqrt{\tilde{\mathbf{C}}(x)}}\tilde{\mathbf{C}}'(x)^{1/2}|||_{\mathrm{op}}\\
&\le|||e^{-t\sqrt{\tilde{\mathbf{C}}(x)}}\mathbf{D}_p^{-1/2}|||_{\mathrm{op}}|||\mathbf{D}_p^{1/2}\tilde{\mathbf{C}}'(x)^{1/2}|||_{\mathrm{op}}\\
&\lesssim|||e^{-t\sqrt{\tilde{\mathbf{C}}(x)}}\mathbf{D}_p^{-1/2}|||_{\mathrm{op}}\\
&\lesssim\sum_{j=1}^{d_T}e^{-t(1/\sqrt{p_j\tilde{e}_j})}/\sqrt{p_j}. \qquad (112)
\end{aligned}
$$

注。从推导过程可以看出，方程（111）和（112）中的结果通常成立，只需将 $\tilde{}\mathbf{C}(x)^{1/2}$ 替换为某个对角矩阵函数 $F(x)$，且该函数满足 $\mathbf{D}_p^{1/2}F(x)$ 有界这一性质。例如，我们可以令 $F(x)=[\sqrt{\tilde{}\mathbf{C}(x)}]'$。

结合上述两个结果，我们得到

$$
\begin{aligned}
T_2 &\le \int_0^\infty v_1 e^{-v_2 t} \sum_{j=1}^{d_T} e^{-t(1/(p_j \bar{\varepsilon}_j))}/p_j dt \\
&\lesssim \int_0^\infty \sum_{j=1}^{d_T} e^{-t(1/\sqrt{p_j \bar{\varepsilon}_j})}/\sqrt{p_j} dt \le d_T \sqrt{M_\varepsilon} = \mathcal{O}(1) \\
T_3 &\le \int_0^\infty v_1^2 e^{-2v_2 t} dt = v_1^2/(2v_2) = \mathcal{O}(1).
\end{aligned}
$$

因此，我们得出结论：$\|\|[\sqrt{\tilde{\mathbf{C}}(x) + \tilde{\mathbf{\Delta}}(x)} - \sqrt{\tilde{\mathbf{C}}(x)}]'\|\|_{\mathrm{op}}$有界，因此$[\sqrt{\tilde{\mathbf{C}}(x) + \tilde{\mathbf{\Delta}}(x)} - \sqrt{\tilde{\mathbf{C}}(x)}]$是利普希茨的。

F.3.4 $d''(x)$的有界性

接下来，我们证明 $d''(x)$也有界。首先，对于任意矩阵函数$\mathbf{F}(x) \in \mathcal{S}_{d_T}^+$，有

$$
\begin{aligned}
&\sqrt{\mathbf{F}(x)}'' \\
&= \int_0^\infty e^{-t\sqrt{\mathbf{F}(x)}} \mathbf{F}''(x) e^{-t\sqrt{\mathbf{F}(x)}} dt \\
&\qquad - 2\int_0^\infty e^{-t\sqrt{\mathbf{F}(x)}} \Big(\int_0^\infty e^{-t\sqrt{\mathbf{F}(x)}} \mathbf{F}'(x) e^{-t\sqrt{\mathbf{F}(x)}} dt \Big)^2 e^{-t\sqrt{\mathbf{F}(x)}} dt
\end{aligned}
$$

因此，

$$
\begin{aligned}
&\Big\|\Big[\sqrt{\tilde{\mathbf{C}}(x) + \tilde{\mathbf{\Delta}}(x)} - \sqrt{\tilde{\mathbf{C}}(x)}\Big]''\Big\|_{\mathrm{op}} \\
&\le \Big\| \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}} [\tilde{\mathbf{C}}''(x) + \tilde{\mathbf{\Delta}}''(x)] e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}} dt \\
&\qquad\qquad - \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}(x)}} [\tilde{\mathbf{C}}''(x)] e^{-t\sqrt{\tilde{\mathbf{C}}(x)}} dt \Big\|_{\mathrm{op}} \\
&+ 2\Big\| \int_0^\infty \Big\{ e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}} \Big(\int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}} [\tilde{\mathbf{C}}'(x) + \tilde{\mathbf{\Delta}}'(x)] e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}} dt \Big)^2 \\
&\qquad\qquad e^{-t\sqrt{\tilde{\mathbf{C}}(x)+\tilde{\mathbf{\Delta}}(x)}} \Big\} dt \\
&- \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}(x)}} \Big(\int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}(x)}} \tilde{\mathbf{C}}'(x) e^{-t\sqrt{\tilde{\mathbf{C}}(x)}} dt \Big)^2 e^{-t\sqrt{\tilde{\mathbf{C}}(x)}} dt \Big\|_{\mathrm{op}} =: T_4 + 2T_5.
\end{aligned}
$$

对于 T_4，我们可以使用证明$[\sqrt{\tilde{\mathbf{C}}(x) + \tilde{\mathbf{\Delta}}(x)} - \sqrt{\tilde{\mathbf{C}}(x)}]'$有界性时所用的相同论证，证明其有界。唯一的区别在于，我们分别将 $\tilde{\mathbf{C}}'(x)$ $\tilde{\mathbf{\Delta}}'(x)$替换为$\tilde{\mathbf{C}}''(x)$ $\tilde{\mathbf{\Delta}}''(x)$。请注意，在证明中，我们只用到了 $\tilde{\mathbf{\Delta}}'(x)$和$\mathbf{D}_p \tilde{\mathbf{C}}'(x)$有界这一性质。因此，后续步骤同样成立，因为$\mathbf{D}_p \tilde{\mathbf{C}}''(x)$和 $\tilde{\mathbf{\Delta}}''(x)$都有界。

为简 $\tilde{\quad}$ $\tilde{\quad}$性，我们有时会省略$\mathbf{C}$ 和 $\mathbf{\Delta}$ 对 x 的依赖关系，具体取决于其含义是否 s

清晰。定义

$$
\begin{aligned}
T_6 &:= \Big|\Big|\Big| \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} \sqrt{\tilde{\mathbf{C}}}' (\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}' - \sqrt{\tilde{\mathbf{C}}}') e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} dt \Big|\Big|\Big|_{\mathrm{op}} \\
T_7 &:= \Big|\Big|\Big| \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} (\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}' - \sqrt{\tilde{\mathbf{C}}}')^2 e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} dt \Big|\Big|\Big|_{\mathrm{op}} \\
T_8 &:= \Big|\Big|\Big| \int_0^\infty (e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} - e^{-t\sqrt{\tilde{\mathbf{C}}}}) [\sqrt{\tilde{\mathbf{C}}}']^2 e^{-t\sqrt{\tilde{\mathbf{C}}}} dt \Big|\Big|\Big|_{\mathrm{op}} \\
T_9 &:= \Big|\Big|\Big| \int_0^\infty (e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} - e^{-t\sqrt{\tilde{\mathbf{C}}}}) [\sqrt{\tilde{\mathbf{C}}}']^2 (e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} - e^{-t\sqrt{\tilde{\mathbf{C}}}}) dt \Big|\Big|\Big|_{\mathrm{op}}.
\end{aligned}
$$

根据三角不等式，对于 T_5，我们有

$$
\begin{aligned}
T_5 &= \Big|\Big|\Big| \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} \Big[\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}'\Big]^2 e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} dt - \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}}} \Big[\sqrt{\tilde{\mathbf{C}}}'\Big]^2 e^{-t\sqrt{\tilde{\mathbf{C}}}} dt \Big|\Big|\Big|_{\mathrm{op}} \\
&\le \Big|\Big|\Big| \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} \sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}' (\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}' - \sqrt{\tilde{\mathbf{C}}}') e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} dt \Big|\Big|\Big|_{\mathrm{op}} \\
&+ \Big|\Big|\Big| \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} (\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}' - \sqrt{\tilde{\mathbf{C}}}') \sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}' e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} dt \Big|\Big|\Big|_{\mathrm{op}} \\
&+ \Big|\Big|\Big| \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} [\sqrt{\tilde{\mathbf{C}}}']^2 e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} dt - \int_0^\infty e^{-t\sqrt{\tilde{\mathbf{C}}}} [\sqrt{\tilde{\mathbf{C}}}']^2 e^{-t\sqrt{\tilde{\mathbf{C}}}} dt \Big|\Big|\Big|_{\mathrm{op}} \\
&\lesssim T_6 + T_7 + T_8 + T_9.
\end{aligned}
$$

现在我们来求 T_i 关于 $i \in \{6,7,8,9\}$的界。

关于 T_6 和 T_7 的界。从 T_6开始，我们有

$$
\begin{aligned}
T_6 &\le \int_0^\infty |||e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} \sqrt{\tilde{\mathbf{C}}}'|||_{\mathrm{op}} |||(\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}' - \sqrt{\tilde{\mathbf{C}}}')|||_{\mathrm{op}} |||e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}}|||_{\mathrm{op}} dt \\
&\lesssim \int_0^\infty |||(e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} - e^{-t\sqrt{\tilde{\mathbf{C}}}}) \sqrt{\tilde{\mathbf{C}}}'|||_{\mathrm{op}} + |||e^{-t\sqrt{\tilde{\mathbf{C}}}} \sqrt{\tilde{\mathbf{C}}}'|||_{\mathrm{op}} dt \\
&\lesssim \int_0^\infty v_1 e^{-v_2 t} dt + \sum_{j=1}^{d_T} \int_0^\infty e^{-t(1/\sqrt{p_j \tilde{\varepsilon}_j})} / \sqrt{p_j} dt \\
&= d_T \sqrt{M_\varepsilon} + \frac{v_1}{v_2} = \mathcal{O}(1),
\end{aligned}
$$

其中第二行用到了 $|||(\sqrt{\tilde{\mathbf{C}} + \tilde{\mathbf{\Delta}}}' - \sqrt{\tilde{\mathbf{C}}}')|||_{\mathrm{op}}$ 有界这一事实，$|||e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}}|||_{\mathrm{op}} \le e^{-t\sigma_{\min}(\mathbf{\Omega}^2)^{1/2}} \le 1$ 而最后一行则由方程（111）和（112）以及随后的注释得出。类似地，我们有

$$
\begin{aligned}
T_7 &\le \int_0^\infty |||e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}}|||_{\mathrm{op}} |||(\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}' - \sqrt{\tilde{\mathbf{C}}}')|||_{\mathrm{op}}^2 |||e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}}|||_{\mathrm{op}} dt \\
&\lesssim \int_0^\infty e^{-2t\sigma_{\min}(\mathbf{\Omega}^2)^{1/2}} dt = \mathcal{O}(1),
\end{aligned}
$$

其中第二行由 $|||e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}}|||_{\mathrm{op}} \le e^{-t\sigma_{\min}(\mathbf{\Omega}^2)^{1/2}} \le 1$ 以及 $|||(\sqrt{\tilde{\mathbf{C}} + \tilde{\mathbf{\Delta}}}' - \sqrt{\tilde{\mathbf{C}}}')|||_{\mathrm{op}}$ 的有界性得出。

T_8 **和** T_9 对于 T_8，我们有

$$\begin{aligned} T_8 &\le \int_0^\infty \|\|(e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} - e^{-t\sqrt{\tilde{\mathbf{C}}}})\sqrt{\tilde{\mathbf{C}}}'\|\|_{\mathrm{op}} \|\|\sqrt{\tilde{\mathbf{C}}}' e^{-t\sqrt{\tilde{\mathbf{C}}}}\|\|_{\mathrm{op}} dt \\ &\lesssim \int_0^\infty (v_1 e^{-v_2 t})(\sum_{j=1}^{d_T} e^{-t(1/\sqrt{p_j \bar{\varepsilon}_j})}/\sqrt{p_j}) dt \\ &\lesssim \sum_{j=1}^{d_T} \int_0^\infty e^{-t(1/\sqrt{p_j \bar{\varepsilon}_j})}/\sqrt{p_j} dt = \mathcal{O}(1), \end{aligned}$$

其中第二行由方程（111）和（112）得出。最后，我们将 T_9 界定为

$$\begin{aligned} T_9 &\le \int_0^\infty \|\|(e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} - e^{-t\sqrt{\tilde{\mathbf{C}}}})\sqrt{\tilde{\mathbf{C}}}'\|\|_{\mathrm{op}} \|\|\sqrt{\tilde{\mathbf{C}}}' (e^{-t\sqrt{\tilde{\mathbf{C}}+\tilde{\mathbf{\Delta}}}} - e^{-t\sqrt{\tilde{\mathbf{C}}}})\|\|_{\mathrm{op}} dt \\ &\lesssim \int_0^\infty (v_1 e^{-v_2 t})^2 dt = \frac{v_1^2}{v_2} = \mathcal{O}(1), \end{aligned}$$

其中第二行再次使用了方程（111）。因此，我们证明了T_4和 T_5均有界，从而 $\|\|d''(x)\|\|_{\mathrm{op}} = \|\|[\sqrt{\tilde{\mathbf{C}}(x) + \tilde{\mathbf{\Delta}}(x)} - \sqrt{\tilde{\mathbf{C}}(x)}]''\|\|_{\mathrm{op}}$有界，即 $[\sqrt{\tilde{\mathbf{C}}(x) + \tilde{\mathbf{\Delta}}(x)} - \sqrt{\tilde{\mathbf{C}}(x)}]'$ 关于 x 是利普希茨的。证明完毕。 □

G 技术引理及其证明

本节专门介绍证明中使用的几个技术引理。

G.1 鞅差序列

我们首先给出一个关于鞅差序列的辅助结果。该结果既适用于向量，也适用于矩阵；在这两种情况下，我们都使用 $\|\cdot\|_F$表示 Frobenius范数，而在向量情况下，它等价于欧几里得范数。

引理18. 设 $\{D_i\}_{i\ge1}$ 是相对于滤过$\{\mathcal{F}_i\}_{i\ge1}$ 的鞅差序列（即，$\mathbb{E}(D_i \mid \mathcal{F}_{i-1}) = 0$ 对所有 $i \ge 1$ 均成立）。若 $\frac{1}{n}\sum_{i=1}^n \mathbb{E}\|D_i\|_F^2 \,(*)= \mathcal{O}(1)$，则

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n D_i = \mathcal{O}_p(1).$$

作为一种特殊情况，例如，当二阶矩 $\mathbb{E}\|D_i\|_F^2$ 一致有界时，上述陈述中的假设 **(*)** 成立。

证明。 根据依概率有界的性质，只需证明

$$\frac{1}{n}\mathbb{E}\|\sum_{i=1}^n D_i\|_F^2 = \mathcal{O}(1).$$

由于 $\{D_i\}_{i\ge1}$ 是一个鞅差序列，我们有

$$\mathbb{E}\,\mathrm{tr}(D_i D_j^\top) = \mathbb{E}\mathbb{E}(\mathrm{tr}(D_i D_j^\top \mid \mathcal{F}_{j-1}) = \mathbb{E}\,\mathrm{tr}(D_i \mathbb{E}(D_j^\top \mid \mathcal{F}_{j-1})) = 0. \qquad \text{for all } i < j,$$

因此，

$$\mathbb{E}\|\sum_{i=1}^{n} D_i\|_F^2/n = \mathbb{E}\sum_{i,j=1}^{n}\mathrm{tr}(D_i D_j^\top)/n = \mathbb{E}\sum_{i=1}^{n}\|D_i\|_F^2/n = \mathcal{O}(1).$$

□

G.2 假设 (SEL（t）) 的等价条件

假设 **(SEL(t))** 等价于关于协方差矩阵 $\mathbf{\Sigma}_i$的最小奇异值的以下假设。

(A2b) 存在常数 $c_0 > 0$ 和 $t \in [0, \frac{1}{2})$，使得条件协方差矩阵

$$\mathbf{\Sigma}_i := \mathbb{E}\big[(x_i - p_i(z_i, \mathcal{F}_{i-1}))(x_i - p_i(z_i, \mathcal{F}_{i-1}))^\top \mid \mathcal{F}_{i-1}, z_i\big]$$

满足

$$\mathbf{\Sigma}_i \succeq c_i \mathbf{I}_{d_T} = \frac{c_0}{i^{2t}}\mathbf{I}_{d_T} \quad \text{for all } i = 1, 2, \ldots. \tag{113}$$

具体而言，我们有

引理 19（假设 **(SEL(t))** 与 **(A2b)** 的等价性）。给定 p_0，p_1，...，$p_{d_T} > 0$使得 $p_0 + p_1 + \cdots + p_{d_T} = 1$。令，$\mathbf{\Sigma} \in \mathbb{R}^{d_T \times d_T}$其中 $\mathbf{\Sigma}_{jj} = (1 - p_j)p_j$且 $\mathbf{\Sigma}_{jk} = -p_j p_k$对于 $j \neq k$。

(a) 如果存在某个常数 $c_0 > 0$ 使得 $\mathbf{\Sigma} \succeq c_0 \mathbf{I}_{d_T}$，那么 $p_j \geq c_0$ 对所有 j =0、1、.....、d_T 均成立。

(b) 如果存在某个常数 $c_0 > 0$使得 $p_j \geq c_0$对 j = 0、1、.....、d_T均成立，那么 $\mathbf{\Sigma} \succeq c_0 \mathbf{I}_{d_T}/(d_T + 2)$。

假设 **(SEL(t))** 与**(A2b)** 的等价性可直接由引理 19 得出。在辅助引理的证明中，我们也使用假设**(A2b)** 来代替 **(SEL(t))**。

证明。我们将证明分为引理的两个部分。

第 (a) 部分的证明 由于 $\mathbf{\Sigma}_{jj} \geq \lambda_{\min}(\mathbf{\Sigma}) \geq c_0$，可知 $p_j(1 - p_j) \geq c_0$，因此$p_j > c_0$对 $j \geq 1$成立。此外，由于 $d_T c_0 = c_0\|\mathbf{1}\|_2^2 \leq \mathbf{1}^\top \mathbf{\Sigma}\mathbf{1} = p_0(1 - p_0)$，我们有 $p_0 > d_T c_0 > c_0$。

（b）部分的证明 注意

$$\begin{aligned}
\lambda_{\min}(\mathbf{\Sigma}) = \big(|||\mathbf{\Sigma}^{-1}|||_{\mathrm{op}}\big)^{-1} &\geq \big(|||\mathbf{\Sigma}^{-1}|||_F\big)^{-1} \\
&\overset{(j)}{=} \Big(\sqrt{d_T(d_T-1)\frac{1}{p_0^2} + \sum_{j=1}^{d_T}(\frac{1}{p_0} + \frac{1}{p_j})^2}\Big)^{-1} \\
&> 1/\sqrt{d_T(d_T+1)\frac{1}{p_0^2} + 2\sum_{j=1}^{d_T}\frac{1}{p_j^2}} \\
&> \frac{c_0}{d_T + 2},
\end{aligned}$$

其中，步骤（i）由 $\mathbf{\Sigma}^{-1}$ **(16b)** 的显式表达式得出。进而可得 $\mathbf{\Sigma} \succeq c_1 \mathbf{I}_{d_T}$对于 $c_1 = c_0/(d_T + 2)$。

□