arXiv:2302.08097v1 [math.ST] 16 Feb 2023

新型 $\sqrt{n}$ 一致、数值稳定的高阶影响函数估计量

Lin Liu*[1] 和 Chang Li[2]

[2]自然科学研究院、教育部统计科学研究中心、数学科学学院、中国气象局上海台风研究所、上海交通大学-耶鲁大学生物统计与数据科学联合中心、上海交通大学及上海人工智能实验室，中国上海[1]统计系，弗吉尼亚大学，夏洛茨维尔，弗吉尼亚州，美国

2023年2月17日

摘要

高阶影响函数（HOIF）为构造一大类低维（光滑）统计泛函/参数（有时甚至是无穷维函数）的速率最优估计量提供了统一理论。这些统计泛函/参数源自流行病学、经济学和社会科学等实质性领域。自Robins 等人(2008）或Robins 等人(2016）[1], 提出 HOIF 以来，它们大多被视为理论基准，而非统计实践中的实用工具。旨在改变这一局面的研究寥寥无几，但近期有少数论文Liu 等人(2017、2021b)取得了一些进展。本文尝试以全新的方式实现这一目标，构造新型、数值稳定的 HOIF 估计量（简称 sHOIF 估计量，其中“s”代表“稳定”），并为其提供可证明的统计和计算保证。这类新型 sHOIF 估计量（最高至二阶）的构想，已在Liu 等人(2020a)开展的合成实验中初现端倪。

关键词：因果推断、泛函估计、高阶影响函数、半参数理论、组合学

*通讯作者：linliu@sjtu.edu.cn。作者谨向高凤南、Zhenyu Liao、Rajarshi Mukherjee、Jamie Robins和Zheng Zhang 致谢，感谢他们就本文提出的宝贵意见。作者衷心感谢国家自然科学基金委员会项目（编号：12101397 和 12090024）、上海市科学技术重大专项（编号：2021SHZDZX0102）、上海市科学技术委员会项目（编号：21JC1402900）以及上海市自然科学基金项目（编号：21ZR1431000）提供的资助。[1]Robins 等人(2016年)是Robins 等人(2008)的完整版，其中包含更多结果和证明。因此，下文仅引用Robins 等人(2016年)。

1 引言

高阶影响函数（HOIFs）（Robins 等人，2016年）是一阶影响函数（IFs）的高阶推广；一阶影响函数是半参数统计理论中的重要工具（Newey，1990；Bickel 等人，1998；van der Vaart，2002）。HOIFs 是一种强大且统一的方法，可用于为一类统计泛函/参数（有时甚至是函数；参见 Kennedy 等人（2022））构造极小极大速率最优估计量。这些统计泛函/参数见于（生物）统计学、流行病学、经济学和社会科学。Robins 等人（2016年，2017）[2] 最初提出的 HOIF 估计量，至今仍是针对上述学科中具有实质性研究意义的统计泛函/参数的唯一已知极小极大速率最优估计量，包括强可忽略性假设[3]下的平均处理效应（ATE），以及给定第三个随机变量 X时两个随机变量 A 和 Y 的条件协方差的期望；尽管近年来统计学和计量经济学界对此进行了大量研究（Newey 和 Robins，2018；Kennedy，2020；Hirshberg 和 Wager，2021；Yu 和 Wang，2020）。近期的研究（Kennedy 等人，2022；Bonvini 和 Kennedy，2022）也开始将 HOIF 方法应用于条件平均处理效应（CATE）函数或剂量反应曲线的极小极大最优估计。这些成果为个体化决策问题（例如个性化医疗）奠定了重要的理论基础。这是首次证明 HOIF 估计量至少在理论上也适用于函数估计问题，或者更准确地说，适用于“混合函数与泛函估计问题”。类似思路也已应用于剂量反应曲线估计（Bonvini 和 Kennedy，2022）。关于 HOIF 的入门级综述，感兴趣的读者可参阅 van der Vaart（2014）和 Liu 等（2020b）第 1 节。相对更具技术性的 HOIF 综述见第 1.3 节。

过去十年间，Robins 及其同事启动了一项研究计划，旨在为一类统计泛函/参数建立 HOIF 及基于 HOIF 的估计量的理论基础（Robins, 2004；van der Vaart, 2014；Robins 等人（2016）, 2017；Liu 等人, 2017）。这类泛函/参数最近由 Rotnitzky 等人（2021）加以刻画，本文将其称为双重稳健泛函（DRF）。我们采用这一术语，是为了体现其非参数一阶影响函数可导出双重稳健估计量这一事实（Scharfstein 等人, 1999；Robins 和 Rotnitzky, 2001；Chernozhukov 等人, 2018a）。这类双重稳健泛函包含了 Robins 等人（2016）和 Chernozhukov 等人（2018b）研究的泛函类。在对干扰参数作出标准霍尔德正则性假设（简称霍尔德干扰模型）的条件下，Robins 等人（2017）为一类双重稳健泛函构造了极小极大最优但非自适应的高阶影响函数估计量。Liu 等人（2021a）在霍尔德干扰模型的一个严格子模型中，利用著名的 Lepskii 自适应方案（Lepskii, 1991），为双重稳健泛函构造了自适应二阶影响函数估计量。然而，即使在 $\sqrt{n}$ 可估的情形下，这两种估计量仍需估计可能具有高维度的协变量 X 的密度。当协变量的维度 d 为

[2]关于 Robins 等人（2022）对 Robins 等人（2017）中证明的更正，请参见。[3]在 Liu 等人（2021b）中，我们推导了平均处理效应泛函的高阶影响函数，即使强可忽略性假设不成立也是如此，前提是我们能够获得处理和结局的有效代理变量；这项工作承接了一系列关于近端因果学习的研究（Tchetgen Tchetgen 等人，2020年）。

仅为中等程度地大（例如 $d = 10$）时，非参数密度估计在计算和统计方面都已是一项艰巨任务。

为克服上述问题，Liu 等人（2017）提出了经验HOIF（简称 eHOIF）估计量。这类估计量通过求逆协变量的向量值基变换所对应的样本/经验 Gram 矩阵，避免了多维密度估计；该矩阵基于一个独立于用于构造双重稳健泛函估计量的样本的另一份样本计算。采用这种样本分割策略，主要是为了简化数学分析，从而能够相当直接地分析 eHOIF 估计量的统计性质。具体而言，在最弱的霍尔德正则性假设下，eHOIF 估计量仍是唯一一类能够实现 $\sqrt{n}$-相合性并对双重稳健泛函达到半参数效率的估计量（Robins 等人，2009年）。eHOIF 估计量的这些优良统计性质，也促使人们开发了一类较少依赖假设的假设检验统计量，旨在证伪标准的 $(1-\alpha) \times 100\%$ 双重稳健泛函 Wald 置信区间具有其宣称覆盖概率的说法（Liu 等人，2020a，2021b）。此时，敏锐的读者必定会想：我们为何还需要一类新的经验 HOIF 估计量？这正是本文要探讨的问题。

1.1 动机与主要贡献

尽管Liu 等人（2017）做出了努力，但根据我们过去在实践中使用 eHOIF 估计量的经验（Liu 等人，2017、2020a、2021b；Wanis 等人，2023），大规模模拟实验揭示了其有限样本性能中的若干突出问题[4]：

(i) 数值不稳定性：在Liu 等人（2017）中，尽管 eHOIF 估计量的有限样本性能优于 Robins 等人（2017）中的原始HOIF估计量，但模拟仅限于条件数非常低的情况 k/n：例如 $k \approx 1,000$ 和 $n \approx 10,000$。在Liu 等人（2020a）的模拟研究中，当 k 接近 n 时，eHOIF 估计量在数值上会发散（见Liu 等人（2020a）的 S3.1 节），而且在二阶时就已出现。更令人瞩目的是，高阶 eHOIF 估计量本应纠正偏差，却只会加剧数值发散。

(ii) 非单调偏差降低：Liu 等人（2017）中的理论结果表明，提高估计量的阶数原则上应当降低偏差。然而，我们发现，在实践中，这通常并不适用于 eHOIF 估计量（例如，见原则上，参见（Liu 等人，2021b）的第5节）。有趣的是，sHOIF 估计量在模拟中似乎并未遭受这一问题，使得理论结果从单纯的原则更接近于经验事实；有关二阶或三阶的模拟，参见Liu 等人（2020a）或Wanis 等人（2023）。

我们的贡献主要有三点。

[4]对于感兴趣的读者，这些模拟实验还被用于揭示（非参数）统计理论中的深度神经网络（DNN）与其在实践中的表现之间的差距，详见Xu 等人（2022）。可通过此处访问生成此类模拟的计算机代码。

- 方法论与实际相关性：本文提出了一类适用于双重稳健泛函的全新数值稳定的 sHOIF 估计量，克服了上述 eHOIF 估计量的两个主要局限。稳定二阶IF 估计量最早出现在 Liu 等人 (2020a) 的模拟研究中，但其统计性质仍不明确。

- 理论与证明策略：要更深入地从理论上阐明这一现象，需要进行细致的计算，而非使用粗略的上界。这是相较于其他高阶影响函数相关研究的一项关键技术创新。具体而言，我们大量运用了以下证明技巧：留出分析、矩阵值泰勒展开和组合计算（二项式恒等式的推论）。本文提出的证明策略本身也可能具有独立的研究价值。

- 将高阶平滑影响函数推广到平均处理效应设定之外：我们还将 sHOIF 估计量推广至所有双重稳健泛函，从而能够处理当前因果推断（或计量经济学）文献中更多的结构参数。

1.2 符号说明

在继续讨论之前，我们先汇总本文中经常使用的一些符号。我们将观测数据随机向量记为 $O \in \mathsf{O}$，其中 O 是其对应的样本空间。令 $\bar{\mathsf{z}}_k := (z_1, \cdots z_k)^\top$ 表示一组 k 不同的函数，每个函数的输入域均为 X。固定某个 $\theta' \in \Theta$。 $\mathbb{E}_{\theta'}$、$\mathsf{var}_{\theta'}$和 $\mathsf{cov}_{\theta'}$ 分别表示概率律 $\mathbb{P}_{\theta'}$下的期望、方差和协方差算子。对于任意可测函数 $h : \mathsf{X} \to \mathbb{R}$，令 $\|h\|_\infty := \operatorname{ess\,sup}_{x \in \mathcal{X}} h(x)$，并且对于任意 $p \geq 1$，令 $\|h\|_{\theta',p} := \{\mathbb{E}_{\theta'}[h(X)^p]\}^{1/p}$。我们采用标准的（随机）渐近记号 $.$, $\gtrsim$, $\asymp$, $\gg$, $\ll$, $o(\cdot)$, $\omega(\cdot)$, $O(\cdot)$, $\Omega(\cdot)$, $o_{\mathbb{P}_{\theta'}}(\cdot)$,$O_{\mathbb{P}_{\theta'}}(\cdot)$。对于任意实值向量 $\bar{v}$ 和任意 $q \in \mathbb{R}$，令 v^q 表示 v的逐元素 q次幂。

此外，定义 $\Sigma_{\theta'} := \mathbb{E}_{\theta'}[Q]$，其中 $Q := A \bar{\mathsf{z}}_k(X) \bar{\mathsf{z}}_k(X)^\top$，将其作为 $\bar{\mathsf{z}}_k(X)$的（按A加权）总体格拉姆矩阵；这一记号沿用至第 5.1 节。在该节中，我们将所有结果从强可忽略性条件下处理组结局 Y的均值 $\psi(\theta) := \mathbb{E}_\theta[Y(a = 1)]$[5], 推广至双重稳健泛函的所有成员。同样，定义 $\widehat{\Sigma} := \mathbb{P}_n[Q_k] \equiv n^{-1} \sum_{i=1}^n Q_i$ 为 $\bar{\mathsf{z}}_k(X)$的（按A加权）样本格拉姆矩阵，这一记号也沿用至第 5.1 节。这里， $\mathbb{P}_n[\cdot]$ 表示样本均值算子。为进一步简化符号，对任意多重指标集 $I \subseteq [n]$，记 $Q_I := \sum_{i \in I} Q_i$。为方便起见，我们还将多重指标集$\{i_1, i_2, \cdots i_j\} \subseteq [n]$ 记作$\bar{i}_j$，其中 $j \leq n$。$\Omega_{\theta'} \equiv \Sigma_{\theta'}^{-1}$ 和 $\widehat{\Omega} \equiv \widehat{\Sigma}^{-1}$（若存在）分别是总体格拉姆矩阵和样本格拉姆矩阵的逆矩阵。由 $\bar{\mathsf{z}}_k$ 构造的核记作$K_{\theta',k}(x, x') := \bar{\mathsf{z}}_k(x)^\top \Omega_{\theta'} \bar{\mathsf{z}}_k(x')$和 $\widehat{K}_k(x, x') = \bar{\mathsf{z}}_k(x)^\top \widehat{\Omega} \bar{\mathsf{z}}_k(x')$。给定一组函数$\bar{\mathsf{v}}_k : \mathsf{X} \to \mathbb{R}^k$ 以及任意函数 $L_2 : \mathsf{O} \to \mathbb{R}$， $\Pi[h|\bar{\mathsf{v}}_k]$ 表示将 h 投影到 $\bar{\mathsf{v}}_k$的线性张成空间上的线性投影算子：形式化地，

$$\Pi[h|\bar{\mathsf{v}}_k](x) := \bar{\mathsf{v}}_k(x)^\top \mathbb{E}[\bar{\mathsf{v}}_k(X) \bar{\mathsf{v}}_k(X)^\top]^{-1} \mathbb{E}[\bar{\mathsf{v}}_k(X) h(O)].$$

[5]我们直接使用潜在结果记号，而不作介绍，这不会影响对本文主旨的理解。

除非另有说明，我们用 $\mathbb{P}_\theta$ 表示真实的数据生成律。当参考测度是真实律 $\mathbb{P}_\theta$时，我们通常省略对 θ的依赖：例如，我们写作$\mathbf{k}h\mathbf{k}_p \equiv \mathbf{k}h\mathbf{k}_{\theta,\,p}$, $\Sigma \equiv \Sigma_\theta$, $\Omega \equiv \Omega_\theta$ 和 $\mathbb{P}$, $\mathbb{E}$, var, cov分别对应 $\mathbb{P}_\theta$, $\mathbb{E}_\theta$, var_θ, cov_θ。请注意，Ω不应与渐近记号$\Omega(\cdot)$混淆，这一点可从上下文中辨明。如果某个统计量依赖于未知的真实数据生成律 $\mathbb{P}_\theta$ 的某些部分（例如Ω），则称其为“预言机”统计量；否则称其为“可行”统计量。我们还引入 Diag，作为提取矩阵对角元素的算子。

最后，令 $\mathbb{U}_{n,m}[\cdot]$ 表示 m 阶 U-统计量算子：对于任意函数 $h:\ \mathbb{R}^m \to \mathbb{R}$

$$\mathbb{U}_{n,m}[h(O_1,\cdots,O_m)] \coloneqq \frac{(n-m)!}{n!} \sum_{1\le i_1 \ne \cdots \ne i_m \le n} h(O_{i_1},\cdots,O_{i_m}).$$

当 $m=1$, $\mathbb{U}_{n,m}[\cdot]$ 时，化简为样本均值算子 $\mathbb{P}_n[\cdot]$。类似地，令 $\mathbb{V}_{n,m}$ 为相应的 V-统计量算子[6]:

$$\mathbb{V}_{n,m}[h(O_1,\cdots,O_m)] \coloneqq \frac{(n-m)!}{n!} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} h(O_{i_1},\cdots,O_{i_m}).$$

在本文后文中，对于 $m \ge 2$，我们将定义使用字典 $\bar{\mathbf{z}}_k$ 构造的“预言机” m 阶影响函数估计量 $\widehat{\mathbb{IF}}_{m,m,k}(\Omega) \equiv \widehat{\mathbb{IF}}_{m,m,k} = \widehat{\mathbb{U}_{n,m}[\mathrm{IF}_{m,m,k,\bar{i}_m}]}$，其 U-统计量核为 $\mathrm{IF}_{m,m,k,\bar{i}_m} \equiv \mathrm{IF}_{m,m,k,\bar{i}_m}(\Omega)$。其稳定的可行版本记为 $\widehat{\mathbb{IF}}_{m,m,k}(\Omega)$，相应的核为 $\mathrm{IF}_{m,m,k,\bar{i}_m}(\Omega)$。

1.3 HOIF 的设定与理论回顾

有了刚刚引入的符号，我们现在可以陈述问题设定，并简要回顾本文相关的 HOIFs 理论，尤其是 eHOIFs 理论。

假设给定了 N 个独立同分布观测 $\{O_i\}_{i=1}^N \sim \mathbb{P}_\theta$，其中 $\theta \in \Theta$ 是所谓的干扰参数，Θ 是其底层参数空间。令 $\mathbf{P} \coloneqq \{\mathbb{P}_\theta : \theta \in \Theta\}$ 为数据生成概率测度空间。我们的主要目标是估计光滑统计泛函 $\psi(\theta) :\to \mathbb{R}$，并据此进行统计推断，其含义见 van der Vaart（1991）。我们将 $\psi(\theta)$ 限定为 Rotnitzky 等人（2021）定义的双重稳健泛函。如前所述，我们贯穿全文的示例是 $\psi(\theta) = \mathbb{E}_\theta[Y(a=1)]$ 处理组 $A=1$ 中结局 Y 的均值。此处，观测数据具体为 $O=(X,A,Y)$：依次表示属于紧子集 $\mathbf{X} \equiv [-B,B]^d$（$\mathbb{R}^d$ 的子集）的 d 维协变量、二元治疗分配和有界结局变量。在无混杂假设下（在近端因果推断情境中，可使用 Liu 等人（2021b）的 HOIFs 放宽该假设），$\psi(\theta)$ 可由观测数据分布的以下两个统计泛函之一识别：

$$\psi(\theta) \equiv \mathbb{E}\left[Aa(X)Y\right] \equiv \mathbb{E}[b(X)] \tag{1}$$

[6]为方便记号，这里我们采用缩放方式$\frac{(n-m)!}{n!}$，而非更常用的$\frac{1}{n^m}$。

其中 $a(x) := \{\mathbb{E}[A|X=x]\}^{-1}$，而 $b(x) := \mathbb{E}[Y|X=x, A=1]$，第 5.1 节除外。对于该泛函 $\psi(\theta)$，干扰参数为 $\theta \equiv (a, b, g)$，其中 $g(x)$ 是给定 $A=1$ 时协变量 X 的概率密度/质量函数。因此，干扰参数空间为 $\Theta = \mathsf{A} \times \mathsf{B} \times \mathsf{G}$，其中 A、B、G 分别是 a、b、g 所属的空间。我们进一步将全部 N 个数据点分为两部分：一部分样本量为 n，称为估计样本；另一部分样本量为 $N-n$，称为干扰样本，用于估计干扰参数 θ。本文始终以干扰样本数据为条件，将其或由其计算出的任何量视为固定值。

对于一个光滑统计泛函 $\psi(\theta)$，依照 vander Vaart (1991年的定义，其一阶影响函数 $\mathbb{IF}_1(\theta)$ 是一个均值为零的一阶 U-统计量，满足以下泛函方程

$$\left.\frac{\mathrm{d}\psi(\theta_t)}{\mathrm{d}t}\right|_{t=0} = \mathbb{E}\left[\mathbb{IF}_1(\theta) \cdot \mathbb{S}_1\right]$$

其中 $\mathbb{P}_{\theta_t}$ 是 $\{\mathbb{P}_\theta, \theta \in \Theta\}$中的任意参数子模型，并且当 $t=0$, $\mathbb{P}_{\theta_t} \equiv \mathbb{P}_\theta$时，表示真实的数据生成律；$\mathbb{S}_1$ 是其一阶得分向量，定义见 Waterman 和 Lindsay(1996)；另见 Robins 等人 (2016年）。这里 $\mathbb{IF}_1(\theta)$ 的形式如下（Robins 等人，1994）：

$$\mathbb{IF}_1(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathsf{IF}_{1,i}(\theta), \text{ where } \mathsf{IF}_1(\theta) = Aa(X)(Y - b(X)) + b(X) - \psi(\theta). \tag{2}$$

通常，经典半参数理论（Newey，1990；Bickel 等人，1998）构造半参数有效的一阶估计量 $\widehat{\ }\ \psi_1$ 的 $\psi(\theta)$，其依据其一阶影响函数，形式如下：

$$\widehat{\psi}_1 = \frac{1}{n}\sum_{i=1}^{n} A_i \widehat{a}(X_i)(Y_i - \widehat{b}(X_i)) + \widehat{b}(X_i)$$

$\widehat{\ }$其中 $\widehat{a}$，b 是利用干扰样本计算得到的干扰参数估计值。特别地，$\widehat{\ }$地，ψ_1 的偏差为

$$\begin{aligned} \mathsf{bias}(\widehat{\psi}_1) &= \mathbb{E}[\widehat{\psi}_1 - \psi(\theta)] = \mathbb{E}[\mathsf{IF}_1(\widehat{\theta}) - \mathsf{IF}_1(\theta)] \\ &= \mathbb{E}\left[\left(\frac{\widehat{a}(X)}{a(X)} - 1\right)(b(X) - \widehat{b}(X))\right]. \end{aligned} \tag{3}$$

形式上，偏差$(\widehat{\psi}_1)$ 是两个干扰参数估计误差的乘积[7]，，因此具有双重稳健(Scharfstein 等人，1999)。

尽管具有双重稳健性，但当干扰参数 θ 的复杂度较高时，基于一阶估计量（如$\widehat{\ }\psi_1$）进行推断的可靠性仍可能存疑：例如，函数光滑度较低或不具备稀疏性。举例来说，当 a，b 属于光滑度为 s_a，s_b 的霍尔德函数，且 g 复杂度任意高时，迄今尚无已知的一阶估计量能在整个范围内以$\sqrt{n}$-相合的方式估计 $\psi(\theta)$

$$\{(s_a, s_b) : (s_a + s_b)/2 \geq d/4\} \tag{4}$$

[7]Rotnitzky 等人(2021) 实际上定义了基于这种二阶偏差性质、可使用双重稳健估计量的一般统计泛函类；参见第5.1节。

但若进一步假设 g 是光滑度为 $s_g > 0$ 的霍尔德函数，则 Liu 等人（2017）的 eHOIF 估计量或 Robins 等人（2016）的原始HOIF估计量可以做到。事实上，Robins 等人（2009）还证明，在霍尔德型干扰建模假设下，条件（4）是存在以 $\sqrt{n}$-相合方式估计 $\psi(\theta)$ 的估计量的最小条件。在条件（4）之外，$\psi(\theta)$ 不可$\sqrt{n}$估计；而在极小极大意义下，唯一已知具有最优收敛速率的估计量仍是 HOIF 估计量（Robins 等人，2016、2017、2022）。对于高度光滑的 g，Liu 等人（2021a）通过将 HOIF 估计量与著名的 Lepskii 自适应方案（Lepskii，1991）相结合，构造出用于估计$\psi(\theta)$的极小极大最优自适应估计量。

本文讨论的是 $\sqrt{n}$-可估计情形（4），因此我们将重点关注经验高阶影响函数估计量。首先，我们选取一组 k维函数 $\bar{\mathsf{z}}_k \equiv (z_1, \cdots z_k)^\top : \mathcal{X} \to \mathbb{R}^k$，它们满足稍后将在第 2节给出的某些正则性条件。参数 $\psi(\theta)$ 的二阶影响函数（SOIF）估计量是以下二阶 U-统计量：

$$\widehat{\psi}_{2,k}(\Omega) \coloneqq \widehat{\psi}_1 + \widehat{\mathbb{IF}}_{2,2,k} \text{ where } \widehat{\mathbb{IF}}_{2,2,k} \equiv \widehat{\mathbb{IF}}_{2,2,k}(\Omega) \coloneqq \mathbb{U}_{n,2}\left[\mathsf{IF}_{2,2,k;1,2}\right] \tag{5}$$

and

$$\begin{aligned}\mathsf{IF}_{2,2,k;1,2} \equiv \mathsf{IF}_{2,2,k;1,2}(\Omega) &\coloneqq (A_1\widehat{a}(X_1) - 1)\,\bar{\mathsf{z}}_k(X_1)^\top \Omega \bar{\mathsf{z}}_k(X_2) A_2(Y_2 - \widehat{b}(X_2)) \\ &\equiv (A_1\widehat{a}(X_1) - 1)\,K_k(X_1, X_2) A_2(Y_2 - \widehat{b}(X_2)).\end{aligned}$$

根据 HOIF 的定义（Robins 等人，2016年），$-\widehat{\mathbb{IF}}_{2,\,2,\,k}$ 实际上是偏差$(\widehat{\psi}_1)$[8]的二阶影响函数。更直观的解释如下：$-\widehat{\mathbb{IF}}_{2,\,2,\,k}$ 是下列量的无偏估计量：

$$\begin{aligned}\mathsf{bias}_k(\widehat{\psi}_1) &= \mathbb{E}\left[\left(\frac{\widehat{a}(X)}{a(X)} - 1\right)\bar{\mathsf{z}}_k(X)^\top\right] \Omega \mathbb{E}\left[A\bar{\mathsf{z}}_k(X)(b(X) - \widehat{b}(X))\right] \\ &= \mathbb{E}\left[\left(\frac{\widehat{a}(X_1)}{a(X_1)} - 1\right) K_k(X_1, X_2) A_2(b(X_2) - \widehat{b}(X_2))\right]\end{aligned} \tag{6}$$

这只是将式（3）中的估计误差 $\widehat{a}/a - 1$ 和 $b - b$ 替换为

$$\Pi\left[\frac{\widehat{a}}{a} - 1 \middle| \bar{\mathsf{z}}_k\right] \text{ and } \Pi\left[b - \widehat{b} \middle| A\bar{\mathsf{z}}_k\right].$$

[8]因此 $\widehat{\mathbb{IF}}_{2,\,2,\,k}$ 可以解释为一个偏差校正项，它部分地消除了偏差(ψ_1)。

然而，在实践中评估 Ω 需要知道 g，而分析者通常并不知道这一点。最初，Robins 等人 (2016年) 和 Robins 等人 (2017) 试图从干扰样本中通过 $\widehat{g}$估计 g，导致统计性质受到 $g - \widehat{g}$的影响，因此需要对 $\mathcal{G} \ni g$作出降低复杂度的假设。为彻底消除这种依赖，Liu 等人 (2017) 选择使用干扰样本通过其经验类比量来估计 Ω，记为 $\widehat{\Omega}_{\text{nuis}} = \widehat{\Sigma}^{-1}_{\text{nuis}}$。由此得到的估计核记为 $\widehat{K}^{\text{nuis}}_k(x,\ x')$，类似于

[8]此处与 $\widehat{\mathbb{IF}}_{2,\,2,\,k}$ 之间的符号差异Robins 等人(2016年)并不重要。

$\widehat{\quad}$ $\widehat{\quad}$ $\widehat{\quad}$第 K_k 1.2节中定义的 . 然后，经验SOIF（eSOIF）估计量 $\mathbb{IF}_{2,\ 2,\ k}(\Omega_{\text{nuis}})$ 的$\widehat{\quad}$偏差$_k(\psi_1)$ 为

$$\widehat{\mathbb{IF}}_{2,2,k}(\widehat{\Omega}_{\text{nuis}}) \equiv \mathbb{U}_{n,2}\left[\mathsf{IF}_{2,2,k;1,2}(\widehat{\Omega}_{\text{nuis}})\right],$$

$\widehat{\quad}$与 $\mathbb{IF}_{2,2,k}$ 不同，它会产生核估计偏差

$$\begin{aligned}\mathbb{E}[\widehat{\mathbb{IF}}_{2,2,k}(\widehat{\Omega}_{\text{nuis}}) - \widehat{\mathbb{IF}}_{2,2,k}] &= \mathbb{E}[(A_1\widehat{a}(X_1)-1)\bar{\mathsf{z}}_k(X_1)^\top](\widehat{\Omega}_{\text{nuis}} - \mathbb{I})\mathbb{E}[\bar{\mathsf{z}}_k(X_2)A_2(Y_2-\widehat{b}(X_2))] \\ &= \mathbb{E}[(A_1\widehat{a}(X_1)-1)(\widehat{K}_k^{\text{nuis}}(X_1,X_2) - K_k(X_1,X_2))A_2(Y_2-\widehat{b}(X_2))],\end{aligned}$$

Liu 等人证明，其阶数至多为 $\sqrt{k\log k/n}$，见 Liu 等人 (2017）。为进一步降低核估计偏差，可以考虑以下 m 阶 eHOIF 估计量，它是一个m 阶 U统计量：

$$\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega}_{\text{nuis}}) := \sum_{j=2}^{m}\widehat{\mathbb{IF}}_{j,j,k}(\widehat{\Omega}_{\text{nuis}})$$

$$\text{where } \widehat{\mathbb{IF}}_{j,j,k}(\widehat{\Omega}_{\text{nuis}}) := \mathbb{U}_{n,j}\left[\widehat{\mathsf{IF}}_{j,j,k;1,\cdots,j}(\widehat{\Omega}_{\text{nuis}})\right]$$

and

$$\widehat{\mathsf{IF}}_{j,j,k;1,\cdots,j}(\widehat{\Omega}_{\text{nuis}}) = (-1)^j(A_1\widehat{a}(X_1)-1)\bar{\mathsf{z}}_k(X_1)^\top\widehat{\Omega}_{\text{nuis}}\prod_{s=3}^{j}\left\{(Q_s-\widehat{\Sigma}_{\text{nuis}})\widehat{\Omega}_{\text{nuis}}\right\}\bar{\mathsf{z}}_k(X_2)A_2(Y_2-\widehat{b}(X_2)).$$

$\widehat{\quad}$ $\widehat{\quad}$Liu 等人（2017）证明，$\mathbb{IF}_{(2,2)\to(m,m),k}(\Omega_{\text{nuis}})$ 的核估计偏差阶数至多为 $(k\log k/n)^{m/2}$，方差阶数至多为 $1/n \vee k/n^2$。因此，取 $m \asymp \sqrt{\quad n}$ $k \asymp n/\quad^c n$ $c > 0$ $_k\ \widehat{\psi}_1$ log 和 log 为某个绝对常数时，我们可以在基本无偏的情况下估计偏差 ( ) $\widehat{\quad}$，且不会增加 ψ_1 的方差阶数。此外，在 A $\times$B 上的霍尔德干扰模型下，Liu 等人（2017）证明，在某些额外的温和假设下，采用上述 m 和 k 取值的 sHOIF 估计量 $\widehat{\quad}$ $\widehat{\quad}$ $\widehat{\quad}\psi_1 + \mathbb{IF}_{(2,2)\to(m,m),k}(\Omega_{\text{nuis}})$ 在 (4) 中是 $\sqrt{n}$-一致的，并且在 (4) 的内部是半参数有效的。$\widehat{\quad}$本文将在第3节介绍 sHOIF 估计量，它们只需将 eHOIF 估计量中的 Σ_{nuis} 和$\widehat{\quad}$ $\widehat{\quad}$ $\widehat{\quad}\Omega_{\text{nuis}}$ 替换为 Σ 和 Ω，即从估计样本中计算得到的 Σ 和 Ω 的经验对应物。显然，由于 Ω 与估计样本相关，分析 sHOIF 估计量的统计性质会变得更具挑战性。

1.4 计划

本文其余部分安排如下。第 2 节定义稳定二阶IF（sSOIF）估计量，并初步研究其统计性质和数值性质。第 3节介绍完整版本的 sHOIF 估计量，以及它们的统计、数值和计算性质。随后，我们将 sHOIF 估计量及其统计性质应用于两个具体问题。第 4节：其一，证明 sHOIF 估计量在经典霍尔德干扰模型的最弱条件下 $\psi(\theta)$

达到半参数效率；其二，使用 sHOIF 估计量检验以一阶去偏机器学习估计量为中心的名义（$1-\alpha$）$\times 100\%$ Wald 置信区间是否具有所声称的覆盖率。这是一种新颖的弱假设统计程序，近期由 Liu 等人(2020a提出，并在后续工作中得到进一步发展，见 Liu 等人(2021b)。为展示 sHOIF 估计量的普适性，第 5节从若干方面扩展了此前的结果。最后，第 6 节总结全文，并讨论若干尚未解决的问题和未来可能的研究方向。附录给出了有助于深入理解证明策略的技术细节。其余技术细节见补充材料（Liand Liu，2023）。

2 假设与预备：稳定二阶影响函数估计量

本节将阐述主要假设，并以稳定二阶影响函数（sSOIF）估计量 $\mathbb{IF}_{2,2,k}(\Omega)$ 为例，预先展示主要结果。

除非另有说明，本文始终采用以下假设。

假设1（关于初始第一步干扰参数估计值的条件）。干扰参数估计量 $\hat{a}$ 和 b 由单独的、独立的冻结干扰参数样本获得。为简便起见，我们假设该样本的大小也为 n。$\hat{a}$ 和 b 在另行说明之前还满足以下性质：

(i) $\|\hat{a}-a\|_2 = o(1)$ 和 $\|b-b\|_2 = o(1)$，即两个干扰参数估计量均为L_2-一致；(ii) $\|a\|_\infty$、$\|b\|_\infty$、$\|\hat{a}\|_\infty$ 和 $\|b\|_\infty$ 均由某个绝对常数 $B>0$界定。(iii) 对于强可忽略性下的 $\psi(\theta) = \mathbb{E}[Y(1)]$ 情形，我们还需要 $1/a$ 和$1/\hat{a}$ 被限制在 $(c, 1-c)$ 之间，其中 $0<c<0.5$ 为某个绝对常数。

假设2（关于 $\bar{\mathsf{z}}_k$ 相关量的条件）。假定基函数 $\bar{\mathsf{z}}_k$ 及其对应的（逆）格拉姆矩阵 Σ, Σ, Ω, Ω以及投影核$\wedge K_k$ 和 K_k满足以下条件：

(i) 存在一个绝对常数 $B>0$，使得 $\sup_{x\in\mathcal{X}} K_k(x,x) \le Bk$ 且 $\sup_{x\in\mathcal{X}} K_k(x,x) \le Bk$；

(ii) Σ 和 Σ 的谱均有界；

(iii) 投影核满足以下 L_∞-稳定性条件：对于任意可测函数 $h: \mathsf{X} \to \mathbb{R}$,

$$\|\Pi[h|\bar{\mathsf{z}}_k](\cdot)\|_\infty \lesssim \|h\|_\infty. \tag{7}$$

注1 （关于假设的评论 1 和 2）。

（i）鉴于假设 2（ii），不失一般性，可以假设 $\Sigma \equiv \Omega \equiv \mathbb{I}$，它是与 Σ 或 Ω 大小相同的单位矩阵。除非另有说明，本文始终采用这一简化。

(ii) 假设 1 中关于干扰参数及其估计量的假设相当宽松。特别是，我们不假设 b 在 L_2 范数下以任何代数速率收敛到 a。事实上，如果只要求 $\sqrt{n}$ 相合性，而不追求半参数效率，那么 $\|\hat{a}-a\|_2 = o(1)$ 和 $\|b-\hat{b}\|_2 = o(1)$ 甚至可以放宽为 $\|\hat{a}-a\|_2 = O(1)$ 和 $\|b-\hat{b}\|_2 = O(1)$；参见 Liu 等人（2017）。

(iii) 关于字典2 的假设 $\bar{z}_k$ 也见于Robins 等人(2017)；Liu 等人(2017、2020a、2021b)；另见Liu 等人(2020b)中的评论。L_∞-稳定性条件（iii）已针对 Cohen-Daubechies-Vial 小波、B样条和局部多项式分割级数得到验证（Belloni 等人，2015）。这种条件可以放宽为高概率版本，但本文不再进一步探讨。

■

以下关于 sSOIF估量 $\mathbb{IF}_{2,\,2,\,k}(\Omega)$ 的结果是第 1节中将要揭示的定理 3 的一个特例。

命题 1 （$\mathbb{IF}_{2,\,2,\,k}(\Omega)$ 的偏差和方差界）。 在假设 *1* – *2*下，当$k = o\,(n)$ 时，有如下结论：

（i）的核估计偏差$\mathbb{IF}_{2,\,2,\,k}(\Omega)$ 满足

$$\begin{aligned}&\text{kern-bias}_{2,k}(\widehat{\psi}_1) := \mathbb{E}\left[\widehat{\mathbb{IF}}_{2,2,k}(\widehat{\Omega})\right] - \text{bias}_{\theta,k}(\widehat{\psi}_1) \equiv \mathbb{E}\left[\widehat{\mathbb{IF}}_{2,2,k}(\widehat{\Omega}) - \widehat{\mathbb{IF}}_{2,2,k}\right] \\ &\lesssim \frac{k}{n}\left\{\left\|\frac{\widehat{a}-1}{a}\right\|_2 \|\widehat{b}-b\|_2 + \left\|\frac{\widehat{a}-a}{a}\right\|_2 \|\widehat{b}-b\|_2 + \left(\left\|\frac{\widehat{a}-1}{a}\right\|_2 \left\|\widehat{b}-b\right\|_\infty \wedge \left\|\frac{\widehat{a}-1}{a}\right\|_\infty \left\|\widehat{b}-b\right\|_2\right)\right\}. \end{aligned} \tag{8}$$

（ii）的方差$\mathbb{IF}_{2,\,2,\,k}(\Omega)$ 满足

$$\text{var}\left[\widehat{\mathbb{IF}}_{2,2,k}(\widehat{\Omega})\right] \lesssim \frac{1}{n}\left\{\frac{k}{n} + \left(\left\|\frac{\widehat{a}-1}{a}\right\|_2 \left\|\widehat{b}-b\right\|_\infty \wedge \left\|\frac{\widehat{a}-1}{a}\right\|_\infty \left\|\widehat{b}-b\right\|_2\right)\right\}. \tag{9}$$

备注 2。 eSOIF估计量的核估计偏差 k/n上界中对条件数 $\widehat{\mathbb{IF}}_{2,\,2,\,k}(\widehat{\Omega}_{\text{nuis}})$ 的依赖性，在 Liu 等人 (2017) ($\sqrt{k}\log\,k/n$) 中比本文报告的 sSOIF估计量 $\mathbb{IF}_{2,\,2,\,k}(\Omega)$ (k/n) 的依赖性更差。■

2.1 命题1的证明概要

2.1.1 核估计偏差界

核偏差$_{2,k}(\widehat{\psi}_1)$可以通过反复使用矩阵恒等式$(A-B)^{-1}-A^{-1}=-A^{-1}B(A-B)^{-1}$加以控制，其中$A\equiv\widehat{\Sigma}^{\dagger}:=n^{-1}\sum_{i=3}^{n}Q_i$和$B=n^{-1}Q_{1,2}$:

$$
\begin{aligned}
&\mathsf{kern\text{-}bias}_{2,k}(\widehat{\psi}_1)\\
&=\mathbb{E}\left[(A_1\widehat{a}(X_1)-1)\bar{\mathsf{z}}_k(X_1)^{\top}(\widehat{\Omega}^{-1}-\mathbb{I})\bar{\mathsf{z}}_k(X_2)A_2(Y_2-\widehat{b}(X_2))\right]\\
&=\sum_{j=1}^{J-1}\mathbb{E}\left[(A_1\widehat{a}(X_1)-1)\bar{\mathsf{z}}_k(X_1)^{\top}\left(\mathbb{I}-\widehat{\Sigma}^{\dagger}-\frac{Q_{1,2}}{n}\right)^{j}\bar{\mathsf{z}}_k(X_2)A_2(Y_2-\widehat{b}(X_2))\right]\\
&\quad+\mathbb{E}\left[(A_1\widehat{a}(X_1)-1)\bar{\mathsf{z}}_k(X_1)^{\top}\left(\mathbb{I}-\widehat{\Sigma}^{\dagger}-\frac{Q_{1,2}}{n}\right)^{J}\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)A_2(Y_2-\widehat{b}(X_2))\right].
\end{aligned}
$$

取$J\asymp\log n$后，可以证明上述表达式的第二项为$o(n^{-1/2})$。

对于第一项，我们只考察$j=1$，2，正文其余部分的分析是附录1 中定理 C 的证明的一个特例。

对于$j=1$，我们有

$$
\begin{aligned}
&\mathbb{E}\left[(A_1\widehat{a}(X_1)-1)\bar{\mathsf{z}}_k(X_1)^{\top}\left(\mathbb{I}-\widehat{\Sigma}^{\dagger}-\frac{Q_{1,2}}{n}\right)\bar{\mathsf{z}}_k(X_2)A_2(Y_2-\widehat{b}(X_2))\right]\\
&=\frac{2}{n}\mathbb{E}\left[\left(\frac{\widehat{a}(X_1)}{a(X_1)}-1\right)\bar{\mathsf{z}}_k(X_1)^{\top}\right]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)(b(X_2)-\widehat{b}(X_2))\right]\\
&\quad-\frac{1}{n}\mathbb{E}\left[(A_1\widehat{a}(X_1)-1)\,\bar{\mathsf{z}}_k(X_1)^{\top}Q_{1,2}\bar{\mathsf{z}}_k(X_2)A_2\left(Y_2-\widehat{b}(X_2)\right)\right] \qquad (10)\\
&=O\left(\frac{1}{n}\right)-\frac{1}{n}\mathbb{E}\left[\left(\frac{\widehat{a}(X_1)-1}{a(X_1)}\right)\bar{\mathsf{z}}_k(X_1)^{\top}\bar{\mathsf{z}}_k(X_1)\bar{\mathsf{z}}_k(X_1)^{\top}\right]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)(b(X_2)-\widehat{b}(X_2))\right]\\
&\quad-\frac{1}{n}\mathbb{E}\left[\left(\frac{\widehat{a}(X_1)}{a(X_1)}-1\right)\bar{\mathsf{z}}_k(X_1)^{\top}\right]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)\bar{\mathsf{z}}_k(X_2)^{\top}\bar{\mathsf{z}}_k(X_2)(b(X_2)-\widehat{b}(X_2))\right]\\
&\lesssim\frac{1}{n}+\frac{k}{n}\left\|\frac{\widehat{a}-1}{a}\right\|_2\|b-\widehat{b}\|_2+\frac{k}{n}\left\|\frac{\widehat{a}}{a}-1\right\|_2\|b-\widehat{b}\|_2
\end{aligned}
$$

最后一行由三角不等式、柯西-施瓦茨不等式以及假设 1、2(i) 和2(ii) 得出。

对于$j=2$，我们有

$$
\begin{aligned}
&\mathbb{E}\left[(A_1\widehat{a}(X_1)-1)\bar{\mathsf{z}}_k(X_1)^{\top}\left(\mathbb{I}-\widehat{\Sigma}^{\dagger}-\frac{Q_{1,2}}{n}\right)^{2}\bar{\mathsf{z}}_k(X_2)A_2(Y_2-\widehat{b}(X_2))\right]\\
&=\mathbb{E}\left[\left(\frac{\widehat{a}(X_1)}{a(X_1)}-1\right)\bar{\mathsf{z}}_k(X_1)^{\top}\right]\mathbb{E}\left[\left(\mathbb{I}-\widehat{\Sigma}^{\dagger}\right)^{2}\right]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)(b(X_2)-\widehat{b}(X_2))\right]\\
&\quad-\frac{2}{n^2}\mathbb{E}\left[(A_1\widehat{a}(X_1)-1)\,\bar{\mathsf{z}}_k(X_1)^{\top}Q_{1,2}\bar{\mathsf{z}}_k(X_2)A_2\left(Y_2-\widehat{b}(X_2)\right)\right]
\end{aligned}
$$

$$+\frac{1}{n^2}\mathbb{E}\left[(A_1\widehat{a}(X_1)-1)\,\bar{\mathsf{z}}_k(X_1)^\top Q_{1,2}^2\bar{\mathsf{z}}_k(X_2)A_2\left(Y_2-\widehat{b}(X_2)\right)\right] \tag{11}$$
$$=: \text{(I)}+\text{(II)}+\text{(III)}.$$

由于 (II) 由关于 $j=1$ 的项主导，我们只需进一步分析 (I) 和 (III)。(III) 可由下式界定

$$\text{(III)} \lesssim \left(\frac{k}{n}\right)^2\left\{\left\|\frac{\widehat{a}-1}{a}\right\|_2\|b-\widehat{b}\|_2+\left\|\frac{\widehat{a}}{a}-1\right\|_2\|b-\widehat{b}\|_2+\left(\left\|\frac{\widehat{a}-1}{a}\right\|_\infty\|b-\widehat{b}\|_2\right)\wedge\left(\left\|\frac{\widehat{a}-1}{a}\right\|_2\|b-\widehat{b}\|_\infty\right)\right\} \tag{12}$$

其中，前两项来自（非交换）展开式中的前三项

$$Q_{1,2}^2=Q_1^2+Q_2^2+Q_1Q_2+Q_2Q_1, \tag{13}$$

第三项则来自上述展开式中的最后一项。 L_∞-范数中出现估计误差，是因为在“肉” Q_2Q_1 与“夹心”结构 $\bar{\mathsf{z}}_k(X_1)^\top[\cdots]\bar{\mathsf{z}}_k(X_2)$ 的“面包片”之间，索引为 1 和 2 的样本点顺序相反。

对于 (I)，我们需要展开 $(\mathbb{I}-\widehat{\Sigma}^\dagger)^2$。

$$\begin{aligned}
\text{(I)} &= \mathbb{E}\left[\left(\frac{\widehat{a}(X_1)}{a(X_1)}-1\right)\bar{\mathsf{z}}_k(X_1)^\top\right]\mathbb{E}\left[\mathbb{I}-2\widehat{\Sigma}^\dagger+\widehat{\Sigma}^{\dagger 2}\right]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)(b(X_2)-\widehat{b}(X_2))\right]\\
&= \left(\frac{4}{n}-1\right)\mathbb{E}\left[\left(\frac{\widehat{a}(X_1)}{a(X_1)}-1\right)\bar{\mathsf{z}}_k(X_1)^\top\right]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)(b(X_2)-\widehat{b}(X_2))\right]\\
&\quad+\frac{(n-2)(n-3)}{n^2}\mathbb{E}\left[\left(\frac{\widehat{a}(X_1)}{a(X_1)}-1\right)\bar{\mathsf{z}}_k(X_1)^\top\right]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)(b(X_2)-\widehat{b}(X_2))\right]\\
&\quad+\frac{n-2}{n^2}\mathbb{E}\left[\left(\frac{\widehat{a}(X_1)}{a(X_1)}-1\right)\bar{\mathsf{z}}_k(X_1)^\top\right]\mathbb{E}\left[Q_3^2\right]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)(b(X_2)-\widehat{b}(X_2))\right]\\
&= \left(\frac{6}{n^2}-\frac{1}{n}\right)\mathbb{E}\left[\left(\frac{\widehat{a}(X_1)}{a(X_1)}-1\right)\bar{\mathsf{z}}_k(X_1)^\top\right]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)(b(X_2)-\widehat{b}(X_2))\right]\\
&\quad+\left(\frac{1}{n}-\frac{2}{n^2}\right)\mathbb{E}\left[\left(\frac{\widehat{a}(X_1)}{a(X_1)}-1\right)\bar{\mathsf{z}}_k(X_1)^\top\right]\mathbb{E}\left[Q_3^2\right]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)(b(X_2)-\widehat{b}(X_2))\right].
\end{aligned} \tag{14}$$

容易看出，上述公式最后一个等式中的第一项由关于 $j=1$ 的项所控制，而第二项可证明有界于

$$\frac{k}{n}\left\|\frac{\widehat{a}}{a}-1\right\|_2\|b-\widehat{b}\|_2.$$

综上， $j=1$ 和 $j=2$ 的各项给出了核-偏差 $_{2,\ k}(\psi_1)$ 在（8）中的所需界。还需证明 $j\geq 3$ 的各项阶数更小；一般情形下的证明推迟至附录C.2 。对于 $j\geq 3$，对应项的阶数为

$$\left(\frac{k}{n}\right)^{j-1}\left\|\frac{\widehat{a}}{a}-1\right\|_2\|b-\widehat{b}\|_2.$$

注 3。现在正是比较 sSOIF 的核估计偏差分析与 Liu 等人 （2017）对 eSOIF 的分析有何不同的好时机。eSOIF 与 sSOIF 估计量之间唯一的区别，是用于估计 Ω 的样本不同。改用干扰样本$\widehat{\quad}$ $\widehat{\quad}$时，给定干扰样本数据的$\mathbb{IIF}_{2,\,2,\,k}(\Omega_{\text{nuis}})$ 的（条件）核估计偏差为

$$\begin{aligned}&\mathbb{E}\left[\widehat{\mathbb{IIF}}_{2,2,k}(\widehat{\Omega}_{\text{nuis}})-\widehat{\mathbb{IIF}}_{2,2,k}\right]\\&=\mathbb{E}\left[(A_1\hat{a}(X_1)-1)\bar{\mathsf{z}}_k(X_1)^\top\right]\left(\widehat{\Omega}_{\text{nuis}}-\mathbb{I}\right)\mathbb{E}\left[A_2\bar{\mathsf{z}}_k(X_2)(Y_2-\hat{b}(X_2))\right].\end{aligned}$$

由此，我们可以得出结论

$$\left|\mathbb{E}\left[\widehat{\mathbb{IIF}}_{2,2,k}(\widehat{\Omega}_{\text{nuis}})-\widehat{\mathbb{IIF}}_{2,2,k}\right]\right|\lesssim\left(\frac{k\log k}{n}\right)^{1/2}\left\|\frac{\hat{a}}{a}-1\right\|_2\|\hat{b}-b\|_2$$

通过使用矩阵 Bernstein 不等式或欣钦不等式（Rudelson，1999；Bandeira 等人，2021）；另见Couillet 和 Liao （2022）。Liu 等人（2017）进一步证明

$$\left|\mathbb{E}\left[\widehat{\mathbb{IIF}}_{(2,2)\to(m,m),k}(\widehat{\Omega}_{\text{nuis}})-\widehat{\mathbb{IIF}}_{2,2,k}\right]\right|\lesssim\left(\frac{k\log k}{n}\right)^{m/2}\left\|\frac{\hat{a}}{a}-1\right\|_2\|\hat{b}-b\|_2.$$

然而，正如第1.1节所指出的，这些上界并不能充分反映 eHOIF 估计量的有限样本性能，因此有必要开发 sHOIF 估计量。

■

2.1.2 方差界

方差界的推导在技术上较为复杂。缺失的步骤可见附录A。关键步骤是证明

$$\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)A_3\bar{\mathsf{z}}_k(X_3)^\top\widehat{\Omega}\bar{\mathsf{z}}_k(X_4)\right]-\left(\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)\right]\right)^2=O\left(\frac{1}{n}\right).\tag{15}$$

要证明（15），只需给出

$$\begin{aligned}&\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top\bar{\mathsf{z}}_k(X_2)A_3\bar{\mathsf{z}}_k(X_3)^\top\left(\widehat{\Omega}-\mathbb{I}\right)\bar{\mathsf{z}}_k(X_4)\right]\\&-\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top\bar{\mathsf{z}}_k(X_2)\right]\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top\left(\widehat{\Omega}-\mathbb{I}\right)\bar{\mathsf{z}}_k(X_2)\right]=O\left(\frac{1}{n}\right)\end{aligned}\tag{16}$$

and

$$\begin{aligned}&\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top\left(\widehat{\Omega}-\mathbb{I}\right)\bar{\mathsf{z}}_k(X_2)A_3\bar{\mathsf{z}}_k(X_3)^\top\left(\widehat{\Omega}-\mathbb{I}\right)\bar{\mathsf{z}}_k(X_4)\right]\\&-\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top\left(\widehat{\Omega}-\mathbb{I}\right)\bar{\mathsf{z}}_k(X_2)\right]\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top\left(\widehat{\Omega}-\mathbb{I}\right)\bar{\mathsf{z}}_k(X_2)\right]=O\left(\frac{1}{n}\right).\end{aligned}\tag{17}$$

回顾 $\widehat{\Omega} = \widehat{\Sigma}^{-1}$ 和 $\widehat{\Sigma} = n^{-1}\sum_{i=1}^{n} Q_i$。我们引入独立的"幽灵副本" Q_1', Q_2' 用于 Q_1, Q_2 并记 $\widehat{\Omega}' = (\widehat{\Sigma}')^{-1}$ 和 $\widehat{\Sigma}' = n^{-1}\sum_{i=3}^{n} Q_i + n^{-1}(Q_1' + Q_2')$。那么（16）等价于

$$\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top \bar{\mathsf{z}}_k(X_2)A_3\bar{\mathsf{z}}_k(X_3)^\top \left(\widehat{\Omega} - \widehat{\Omega}'\right)\bar{\mathsf{z}}_k(X_4)\right] = O\left(\frac{1}{n}\right). \tag{18}$$

令 $\bar{\Omega} \equiv \Sigma$ 且 $\bar{\Sigma} \equiv n^{-1}\sum_{i=5}^{n} Q_i$。在（18）上通过令 $A = \Sigma$ 和 $B = n^{-1}(Q_{1,2} + Q_{3,4})$ 或 $B = n^{-1}(Q'_{1,2} + Q_{3,4})$，重复使用矩阵恒等式 $(A+B)^{-1} - A^{-1} = -A^{-1}B(A+B)^{-1}$，可得

$$\begin{aligned}
&\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top \bar{\mathsf{z}}_k(X_2)A_3\bar{\mathsf{z}}_k(X_3)^\top \left(\widehat{\Omega} - \widehat{\Omega}'\right)\bar{\mathsf{z}}_k(X_4)\right] \\
&= \sum_{j=1}^{J-1}(-1)^j \mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top \bar{\mathsf{z}}_k(X_2)A_3\bar{\mathsf{z}}_k(X_3)^\top \left\{\left(\bar{\Omega}\frac{Q_{1,2}+Q_{3,4}}{n}\right)^j - \left(\bar{\Omega}\frac{Q'_{1,2}+Q_{3,4}}{n}\right)^j\right\}\bar{\Omega}\bar{\mathsf{z}}_k(X_4)\right] \\
&\quad + (-1)^J \mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top \bar{\mathsf{z}}_k(X_2)A_3\bar{\mathsf{z}}_k(X_3)^\top \left\{\left(\bar{\Omega}\frac{Q_{1,2}+Q_{3,4}}{n}\right)^J - \left(\bar{\Omega}\frac{Q'_{1,2}+Q_{3,4}}{n}\right)^J\right\}\widehat{\Omega}\bar{\mathsf{z}}_k(X_4)\right].
\end{aligned}$$

令 $J \asymp$ 对数 n。上述表达式中的第二项可证明为 $o(1/n)$。接下来考察第一项，容易根据假设 2(iii) 看出，与 $j = 1$对应的项为：

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}\left[A_1\bar{\mathsf{z}}_k(X_1)^\top \bar{\mathsf{z}}_k(X_2)A_3\bar{\mathsf{z}}_k(X_3)^\top \bar{\Omega}(Q_{1,2} - Q'_{1,2})\bar{\Omega}\bar{\mathsf{z}}_k(X_4)\right] \\
&= \frac{1}{n}\left(\begin{array}{c}
\mathbb{E}\left\{A_1\bar{\mathsf{z}}_k(X_1)^\top \mathbb{E}[\bar{\mathsf{z}}_k(X_2)]\mathbb{E}[A_3\bar{\mathsf{z}}_k(X_3)]^\top \bar{\Omega}\bar{\mathsf{z}}_k(X_1)\bar{\mathsf{z}}_k(X_1)^\top \bar{\Omega}\mathbb{E}[\bar{\mathsf{z}}_k(X_4)]\right\} \\
+ \mathbb{E}\left\{\mathbb{E}[A_1\bar{\mathsf{z}}_k(X_1)]^\top \bar{\mathsf{z}}_k(X_2)\mathbb{E}[A_3\bar{\mathsf{z}}_k(X_3)]^\top \bar{\Omega}A_2\bar{\mathsf{z}}_k(X_2)\bar{\mathsf{z}}_k(X_2)^\top \bar{\Omega}\mathbb{E}[\bar{\mathsf{z}}_k(X_4)]\right\} \\
- \mathbb{E}\left\{\mathbb{E}[A_1\bar{\mathsf{z}}_k(X_1)]^\top \bar{\mathsf{z}}_k(X_2)\mathbb{E}[A_3\bar{\mathsf{z}}_k(X_3)]^\top \bar{\Omega}\mathbb{E}[Q'_{1,2}]\bar{\Omega}\mathbb{E}[\bar{\mathsf{z}}_k(X_4)]\right\}
\end{array}\right) = O\left(\frac{1}{n}\right).
\end{aligned}$$

类似地，与 $j \geq 2$ 对应的项可证明为 $O\left(\frac{1}{n}\left(\frac{2k}{n}\right)^{j-1}\right)$，这是下文引理 1 的推论。注意，由于展开 $\left(\frac{Q_{1,2}+Q_{3,4}}{n}\right)^j$ 和 $\left(\frac{Q'_{1,2}+Q_{3,4}}{n}\right)^j$ 后总共有 $O(2^j)$ 项，因此会出现额外因子 2。

引理 1。给定一个正整数 j。对于满足 $j_1 + j_2 = j$ 的任意一对整数 $j_1 \geq 0$、$j_2 > 0$，对于 j 维布尔超立方体的任意子集 $\mathsf{j} \subseteq \{0, 1\}^j$，其中 $\|\mathsf{j}\|_1 = j_1$，有

$$\mathbb{E}\left[\bar{\mathsf{z}}_k(X_1)^\top \bar{\mathsf{z}}_k(X_2)\bar{\mathsf{z}}_k(X_3)^\top \left(\prod_{\ell=1}^{j} Q_{1,2}^{\mathsf{j}_\ell} Q_{3,4}^{(1-\mathsf{j}_\ell)}\right)\bar{\mathsf{z}}_k(X_4)\right] \lesssim k^{j-1}. \tag{19}$$

然而，如果 $j_1 = 0$ 和 $j_2 = j$，则有

$$\mathbb{E}\left[\bar{\mathsf{z}}_k(X_1)^\top \bar{\mathsf{z}}_k(X_2)\bar{\mathsf{z}}_k(X_3)^\top Q_{3,4}^j \bar{\mathsf{z}}_k(X_4)\right] \lesssim k^j. \tag{20}$$

引理1 的证明见附录A.1。最后，式（17）的证明将留待在线补充材料中给出，该证明可以用类似的方法完成。

2.2 $\widehat{\mathbb{IF}}_{2,2,k}(\Omega)$ 的数值稳定性与时间复杂度

如第1节所述，提出 sHOIF 估计量的主要动机是 eHOIF 估计量中观察到的数值不稳定性。作为预备，我们将在本节严格证明数值稳定性，并计算$\widehat{\mathbb{IF}}_{2,2,k}(\Omega)$ 的时间复杂度。

$\widehat{\mathbb{IF}}_{2,2,k}(\Omega)$ 可能出现数值不稳定的原因是，当 k 接近 n 时，$\lambda_{\min}(\Sigma) \approx 0$ 的可能性很高，因此 $\lambda_{\max}(\Omega)$ 接近无穷大。但是：

命题 2。 $\widehat{\mathbb{IF}}_{2,2,k}(\Omega)$ 不依赖于Ω的特征值。

为便于说明，以下令

- $\bar{\mathsf{Z}}_{n,k} := (\bar{\mathsf{z}}_k(X_1), \cdots \bar{\mathsf{z}}_k(X_n))^\top$ 作为所有 n个样本的字典向量的 $n \times k$矩阵；
- $\bar{\mathsf{Z}}^A_{n,k} := (A_1 \bar{\mathsf{z}}_k(X_1), \cdots A_n \bar{\mathsf{z}}_k(X_n))^\top$ 作为所有 n 个样本的 A加权字典向量的 $n \times k$ 矩阵；
- $\boldsymbol{\mathcal{E}}_{n,a}(\widehat{a}) := (\mathcal{E}_a(\widehat{a}, O_1), \cdots, \mathcal{E}_a(\widehat{a}, O_n))^\top$ and $\boldsymbol{\mathcal{E}}_{n,b}(\widehat{b}) := \left(\mathcal{E}_b(\widehat{b}, O_1), \cdots, \mathcal{E}_b(\widehat{b}, O_n)\right)^\top$.

证明。 我们可以将$\widehat{\mathbb{IF}}_{2,2,k}(\Omega)$改写为

$$\widehat{\mathbb{IF}}_{2,2,k}(\widehat{\Omega}) = \frac{1}{n-1} \boldsymbol{\mathcal{E}}_{n,a}(\widehat{a})^\top [\mathbb{I} - \mathsf{Diag}] \left\{ \bar{\mathsf{Z}}_{n,k} \left(\bar{\mathsf{Z}}^\top_{n,k} \bar{\mathsf{Z}}^A_{n,k} \right)^{-1} \bar{\mathsf{Z}}^{A\top}_{n,k} \right\} \boldsymbol{\mathcal{E}}_{n,b}(\widehat{b}). \tag{21}$$

现在对矩阵 $\bar{\mathsf{Z}}^A_{n,k}$ 进行奇异值分解（SVD）：

$$\bar{\mathsf{Z}}^A_{n,k} = U^A \mathsf{Diag}(D^A) V^{A\top}.$$

Then

$$\bar{\mathsf{Z}}_{n,k} \left(\bar{\mathsf{Z}}^\top_{n,k} \bar{\mathsf{Z}}^A_{n,k} \right)^{-1} \bar{\mathsf{Z}}^{A\top}_{n,k} = \bar{\mathsf{Z}}^A_{n,k} \left(\bar{\mathsf{Z}}^{A\top}_{n,k} \bar{\mathsf{Z}}^A_{n,k} \right)^{-1} \bar{\mathsf{Z}}^{A\top}_{n,k} = U^A U^{A\top}.$$

So

$$(21) = \frac{1}{n-1} \boldsymbol{\mathcal{E}}_{n,a}(\widehat{a})^\top [\mathbb{I} - \mathsf{Diag}] \left\{ U^A U^{A\top} \right\} \boldsymbol{\mathcal{E}}_{n,b}(\widehat{b}),$$

这完全独立于 $\widehat{\Omega}$ ($(D^A)^2$ 的特征值（相差一个常数因子）。 □

因此，$\widehat{\mathbb{IF}}_{2,2,k}(\Omega)$ 即使在 $k \to n$ 时仍数值稳定，这并不令人意外。

注 4. 从某种意义上说，$\widehat{\mathbb{IF}}_{2,2,k}(\Omega)$ 可以看作是自归一化版本的$\widehat{\mathbb{IF}}_{2,2,k}$。通常认为，自归一化统计量的统计性质可能优于非自归一化统计量（Peña 等人, 2008）。然而，自归一化视角是否有助于确立 $\widehat{\mathbb{IF}}_{2,2,k}(\Omega)$，我们尚不清楚；这值得作为一个研究问题进一步探讨。■

此外，替代公式（21）不仅直接体现了$\widehat{\mathbb{IF}}_{2,\,2,\,k}(\Omega)$的数值稳定性，也揭示了计算$\widehat{\mathbb{IF}}_{2,\,2,\,k}(\Omega)$的复杂度。除去奇异值分解的时间复杂度（$O(nk^2)$），$\widehat{\mathbb{IF}}_{2,\,2,\,k}(\Omega)$的时间复杂度随 kn 线性增长，而非二次增长。这可由（21）看出，其中仅涉及两个向量-矩阵乘积，每个乘积需要 $O(nk)$ 次运算。因此，我们有

命题 3. 计算 $\widehat{\mathbb{IF}}_{2,2,k}(\Omega)$ 的时间复杂度为 $O(nk^2)$，其主要开销来自奇异值分解。

备注 5. 得出上述结论的另一种方法是注意到，$\mathbb{IF}_{2,2,k}(\Omega)$ 的U-统计量核，记为 $\mathrm{IF}_{2,2,k,\bar{i}_2}(\Omega)$，是可分离的，具体来说，存在一对（但不一定唯一的一对）函数 h_1, h_2，使得

$$\widehat{\mathsf{IF}}_{2,2,k,\bar{i}_2}(\widehat{\Omega}) \equiv h_1(O_{i_1}, \widehat{\Omega}) \cdot h_2(O_{i_2}, \widehat{\Omega}).$$

■

3 sHOIF 估计量的层级结构

如第 1.3 节所示，m 阶 sHOIF 估计量 $\widehat{\mathbb{IF}}_{m,m,k}(\Omega)$ 与 m 阶 eHOIF 估计量 $\widehat{\mathbb{IF}}_{m,m,k}(\Omega_{\mathrm{nuis}})$ 具有相同的形式，唯一的区别是 Ω_{nuis} 被 Ω 替代。形式上，m 阶 sHOIF 以及相应的 m 阶 $\psi(\theta)$ 估计量如下所示：

$$\begin{aligned}
\widehat{\mathbb{IF}}_{m,m,k}(\widehat{\Omega}) &:= (-1)^m \mathbb{U}_{n,m}\left[\mathcal{E}_a(\widehat{a}; O_1)\bar{\mathsf{z}}_k(X_1)^\top \widehat{\Omega} \prod_{s=3}^{m}\left\{\left(Q_s - \widehat{\Sigma}\right)\widehat{\Omega}\right\}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b}; O_2)\right], \\
\widehat{\psi}_{m,k}(\widehat{\Omega}) &:= \widehat{\psi}_1 + \sum_{j=2}^{m}\widehat{\mathbb{IF}}_{j,j,k}(\widehat{\Omega}) \equiv \widehat{\psi}_1 + \widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega}).
\end{aligned} \tag{22}$$

备注 6. 上述 sHOIF 统计量与 eHOIF 统计量相同，区别仅在于 Ω 是由估计样本而非干扰样本构造的。■

本节首先以启发式方式解释为何 $\widehat{\mathbb{IF}}_{m,m,k}(\Omega)$ 足以校正核估计偏差（见第 3.1 节），随后正式陈述 $\widehat{\mathbb{IF}}_{m,m,k}(\Omega)$ 的统计、数值和计算性质。

3.1 启发式解释

接下来，我们将从启发式角度解释，为何通过添加以下项，可以进一步校正核估计偏差$\widehat{\mathbb{IF}}_{2,\,2,\,k}(\Omega)$：

$$\widehat{\mathbb{IF}}_{3,3,k}(\widehat{\Omega}) := \frac{n-2}{n}\mathbb{U}_{n,3}\left[-\mathcal{E}_a(\widehat{a}; O_1)\bar{\mathsf{z}}_k(X_1)^\top \widehat{\Omega}\left(Q_3 - \widehat{\Sigma}\right)\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b}; O_2)\right]$$

and

$$\widehat{\mathbb{IF}}_{4,4,k}(\widehat{\Omega}) := \frac{n-3}{n}\mathbb{U}_{n,4}\left[(A_1\widehat{a}(X_1)-1)\bar{\mathsf{z}}_k(X_1)^\top\widehat{\Omega}\left(Q_3-\widehat{\Sigma}\right)\widehat{\Omega}\left(Q_4-\widehat{\Sigma}\right)\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)A_2(Y_2-\widehat{b}(X_2))\right].$$

简言之，我们定义 $\widehat{\mathbb{IF}}_{(2,2)\to(3,3),\,k}(\widehat{\Omega}) := \widehat{\mathbb{IF}}_{2,\,2,\,k}(\widehat{\Omega})+\widehat{\mathbb{IF}}_{3,\,3,\,k}(\widehat{\Omega})$ 和 $\widehat{\mathbb{IF}}_{(2,2)\to(4,4),\,k}(\widehat{\Omega}) := \sum_{j=2}^{4}\widehat{\mathbb{IF}}_{j,\,j,\,k}(\widehat{\Omega})$。另请注意，本节大部分内容是为保证数学严谨性而撰写的。

简单的代数运算可得

$$\begin{aligned}
&\widehat{\mathbb{IF}}_{3,3,k}(\widehat{\Omega})\\
\equiv&-\frac{n-2}{n}\frac{(n-3)!}{n!}\sum_{1\le i_1\neq i_2\neq i_3\le n}\mathcal{E}_a(\widehat{a};O_{i_1})\bar{\mathsf{z}}_k(X_{i_1})^\top\widehat{\Omega}\left(Q_{i_3}-\widehat{\Sigma}\right)\widehat{\Omega}\bar{\mathsf{z}}_k(X_{i_2})\mathcal{E}_b(\widehat{b};O_{i_2})\\
=&\frac{1}{n}\mathbb{U}_{n,2}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\widehat{\Omega}Q_{1,2}\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]-\frac{2}{n}\widehat{\mathbb{IF}}_{2,2,k}(\widehat{\Omega})\\
\approx&\frac{1}{n}\mathbb{U}_{n,2}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\widehat{\Omega}Q_{1,2}\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]=:\widetilde{\widehat{\mathbb{IF}}}_{3,3,k}(\widehat{\Omega})
\end{aligned}$$

and

$$\begin{aligned}
&\widehat{\mathbb{IF}}_{4,4,k}(\widehat{\Omega})\\
\equiv&\frac{n-3}{n}\frac{(n-4)!}{n!}\sum_{1\le i_1\neq i_2\neq i_3\neq i_4\le n}\mathcal{E}_a(\widehat{a};O_{i_1})\bar{\mathsf{z}}_k(X_{i_1})^\top\widehat{\Omega}\prod_{s=3}^{4}\left[\left(Q_{i_s}-\widehat{\Sigma}\right)\widehat{\Omega}\right]\bar{\mathsf{z}}_k(X_{i_2})\mathcal{E}_b(\widehat{b};O_{i_2})\\
=&\frac{1}{n(n-2)}\mathbb{U}_{n,2}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\widehat{\Omega}Q_{1,2}\widehat{\Omega}Q_{1,2}\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]\\
&-\frac{1}{n}\mathbb{U}_{n,3}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\widehat{\Omega}Q_3\widehat{\Omega}Q_3\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]\\
&-\frac{1}{n-2}\left(6\widehat{\mathbb{IF}}_{3,3,k}(\widehat{\Omega})-\left(1-\frac{6}{n}\right)\widehat{\mathbb{IF}}_{2,2,k}(\widehat{\Omega})\right)\\
\approx&\frac{1}{n^2}\mathbb{U}_{n,2}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\widehat{\Omega}Q_{1,2}\widehat{\Omega}Q_{1,2}\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]\\
&-\frac{1}{n}\mathbb{U}_{n,3}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\widehat{\Omega}Q_3\widehat{\Omega}Q_3\widehat{\Omega}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]\\
=:&\widetilde{\widehat{\mathbb{IF}}}_{4,4,k}(\widehat{\Omega}).
\end{aligned}$$

$\widetilde{\widehat{\mathbb{IF}}}_{3,3,k}(\widehat{\Omega})$ 首先，注意到 $\mathbb{IF}_{3,3,k}(\Omega)$ 的预言机版本的期望

$$\mathbb{E}\left[\widetilde{\widehat{\mathbb{IF}}}_{3,3,k}\right]=\frac{1}{n}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top Q_{1,2}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]$$

$\widehat{\mathbb{IF}}_{2,2,k}(\widehat{\Omega})$ 恰好抵消了（10），与 $j=1$ 对应的 $\mathbb{IF}_{2,2,k}(\Omega)$ 的核估计偏差的主导项。

接下来，注意到预言机版本的期望 $\widetilde{\widehat{\mathbb{IF}}}_{4,\,4,\,k}(\Omega)$ 是

$$\mathbb{E}\left[\widetilde{\widehat{\mathbb{IF}}}_{4,4,k}\right]=\frac{1}{n^2}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top Q_{1,2}^2\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]$$

$$-\frac{1}{n}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\right]\mathbb{E}[Q_3^2]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]$$

这再次抵消了在 $j=2$ 处截断的 $\mathbb{IF}_{(2,2)\to(3,3),\,k}(\Omega)$ 的核估计偏差，主要由以下因素主导

$$-\frac{1}{n^2}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top Q_{1,2}^2\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]+\frac{1}{n}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\right]\mathbb{E}[Q_3^2]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right] \tag{23}$$

可由（11）、（14）以及 $\mathbb{IF}_{3,\,3,\,k}(\Omega)$ 截断至阶 $j=1$的核估计偏差推导得到；更详细的计算请见附录B 。因此，$\mathbb{IF}_{(3,3)\to(4,4),\,k}(\Omega)$ 进一步降低了 $\mathbb{IF}_{2,\,2,\,k}(\Omega)$ 的核估计偏差。

3.2 sHOIF 估计量的偏差和方差刻画

现在我们陈述本文的主要理论结果。

定理 1。 在假设*1*–*2*下，当 $k\lesssim n/\log^2 n$ 且 $m\gtrsim\log n$时，有如下结论：

（i）$\mathbb{IF}_{(2,2)\to(m,m),\,k}(\Omega)$ 的核估计偏差满足

$$\begin{aligned}\text{kern-bias}_{m,k}(\widehat{\psi}_1) &:= \mathbb{E}\left[\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega})\right]-\text{bias}_{\theta,k}(\widehat{\psi}_1)\equiv\mathbb{E}\left[\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega})-\widehat{\mathbb{IF}}_{2,2,k}\right]\\ &\lesssim\left(\frac{km}{n}\right)^{\lceil\frac{\lceil\frac{m-1}{2}\rceil-1}{2}\rceil\vee 1}\left\{\begin{array}{c}\left\|\frac{\widehat{a}-1}{a}\right\|_2\|\widehat{b}-b\|_2+\left\|\frac{\widehat{a}-a}{a}\right\|_2\|\widehat{b}-b\|_2\\ +\left(\left\|\frac{\widehat{a}-1}{a}\right\|_2\left\|\widehat{b}-b\right\|_\infty\wedge\left\|\frac{\widehat{a}-1}{a}\right\|_\infty\left\|\widehat{b}-b\right\|_2\right)\end{array}\right\}.\end{aligned} \tag{24}$$

（ii）对于 $m\geq 2$，$\mathbb{IF}_{m,\,m,\,k}(\Omega)$ 的方差满足

$$\text{var}\left[\widehat{\mathbb{IF}}_{m,m,k}(\widehat{\Omega})\right]\lesssim\frac{1}{n}\left\{\frac{k}{n}+\left(\left\|\frac{\widehat{a}-1}{a}\right\|_2\left\|\widehat{b}-b\right\|_\infty\wedge\left\|\frac{\widehat{a}-1}{a}\right\|_2\left\|\widehat{b}-b\right\|_\infty\right)\right\}. \tag{25}$$

因此，$\mathbb{IF}_{(2,2)\to(m,m),\,k}(\Omega)$ 的方差满足

$$\text{var}\left[\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega})\right]\lesssim\frac{1}{n}\left\{\frac{k}{n}+\left(\left\|\frac{\widehat{a}-1}{a}\right\|_2\left\|\widehat{b}-b\right\|_\infty\wedge\left\|\frac{\widehat{a}-1}{a}\right\|_2\left\|\widehat{b}-b\right\|_\infty\right)\right\}. \tag{26}$$

上述定理的证明见附录 C （核估计偏差界）及在线补充材料（方差界）。

注 7 （渐近正态性与自助法近似）。 正如 Liu 等人（2020a）所示，预言统计量 $\mathbb{IF}_{2,\,2,\,k}$–偏差$_{\theta,\,k}(\psi_1)$ $\mathsf{se}_\theta(\widehat{\mathbb{IF}}_{2,\,2,\,k})$ 可由 Bhattacharya 和 Ghosh （1992） 的定理 1 推得，因此 $1\ll k\ll n^2$。因此，要证明中心极限定理

对于任意 $m \geq 2$，证明 $\mathbb{IF}_{(2,2)\to(m,m)}$，$k(\Omega)$ 即已足够；要证明满足何种条件时核偏差m，$k(\Omega) \ll \mathrm{se}_\theta(\mathbb{IF}_2,2,k)$。分布的$\mathbb{IF}_{(2,2)\to(m,m)}$，$k(\Omega)$的自助法近似（及其收敛速度），甚至$\mathbb{IF}_2$，2，$k$ 的自助法近似，仍是一个重要的开放问题，尽管 Liu 等人（2021b）已经取得了一些进展。更全面地研究中心极限定理（CLT）或自助法近似成立的条件，超出了本文的范围。 ■

3.3 sHOIF 估计量的数值稳定性和时间复杂度

接下来，我们将考察 sHOIF 估计量的数值与计算性质，这将第 2.2 节的结果推广到更高阶。本节的第一个结果，即定理 2，指出 sHOIF 估计量的“稳定性”体现在其不受 $\hat{\Omega}$ 的特征值影响，而特征值正是 eHOIF 估计量不稳定的根源。

定理 2。$\mathbb{IF}_{m,\,m,\,k}(\Omega)$ 不依赖于Ω的特征值。

证明。 证明与命题 2 的证明非常相似，只需注意到，对于任意$i,\ j \in [n]$,

$$A_i \bar{\mathsf{z}}_k(X_i)^\top \hat{\Omega} \bar{\mathsf{z}}_k(X_j) = U^A_{i,\bullet} U^{A\top} U U_{j,\bullet},$$

这完全不受 Ω 的特征值影响。 □

因此，当我们在实践中令 k 接近 n 时，sHOIF 估计量不会因样本 Gram 矩阵的条件数过大而遭受数值不稳定性。

注 8。 定理 2 还提出了一种更好的 sHOIF 估计量计算方法。我们不应使用数值方法计算样本 Gram 矩阵 Σ 及其逆矩阵 Ω，而应对基矩阵 $\bar{\mathsf{Z}}_{n,\,k}$ 和 $\bar{\mathsf{Z}}^A_{n}$进行奇异值分解，k 然后计算 $\widehat{\mathbb{IF}}_{m,\,m,\,k}(\hat{\Omega})$。事实上，即将推出的用于计算 HOIF 相关统计量的 R 软件包（Wanis 等人，2023）正是采用了这一策略。 ■

由于 sHOIF 估计量具有数值稳定性，因此可能是统计实践中的有用工具（Liu 等人，2020a；Wanis 等人，2023），因此也值得讨论一般阶数下 sHOIF 估计量的计算复杂度 m 。

定理 3。T 计算 $\widehat{\mathbb{IF}}_{m,\,m,\,k}(\hat{\Omega})$ 的时间复杂度为 $O(\max\{(nk)^{\lceil (m-1)/2 \rceil}, nk^2\})$.

证明。 与命题 2 的证明类似，我们需要将 $\mathbb{IF}_{(2,2)\to(m,m),k}(\Omega)$ 改写为 V-统计量的线性组合。不失一般性，我们取 $\mathrm{E}_{a} \equiv (\hat{a}; O)\, \mathrm{E}_b(b; O) \equiv 1$。不过，我们先将 $\mathbb{IF}_{(2,2)\to(m,m),k}(\Omega)$ 表示为以下级数：

$$\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\hat{\Omega}) \equiv \sum_{j=1}^{m} (-1)^j \binom{m-1}{j-1} \mathbb{U}_{n,j} \left[\bar{\mathsf{z}}_k(X_1)^\top \hat{\Omega} \cdot \prod_{s=3}^{m} \left(Q_s \hat{\Omega} \right) \cdot \bar{\mathsf{z}}_k(X_2) \right].$$

$\widehat{\quad}$ $\widehat{\quad}$注意，$\mathbb{IF}_{m,m,k}(\Omega)$ 中求和的次数为

$$n(n-1)\cdots(n-m+1)=\sum_{j=1}^{m}(-1)^{m-j}s(m,j)n^j$$

其中，$s(m,\ j)$ 是第一类无符号斯特林数，也就是由 m 个元素组成且含有 j 个循环的排列数。因此，可以将一个 m 阶 U-统计量写成从 1 阶到 m 阶的 V-统计量的线性组合，其中对于 $j=1,\ \cdots\ m$，V 阶 j-统计量的数量等于 $s(m,\ j)$。

利用 sHOIF 估计量的 U-统计量核的特殊结构即可完成证明。

□

注 9. 综合考虑定理 1和定理 3，可以清楚地看到统计—计算权衡。然而，这种统计—计算权衡是否源于估计某些光滑统计泛函时可能存在的内在计算困难性，仍是一个悬而未决的问题

最后，我们简要谈谈对 sHOIF 估计量实用性的看法。sHOIF 估计量实际上是无限阶 U-统计量，因此，鉴于当前的计算设备，毫无疑问，短期内从业者不会在实践中使用 sHOIF 估计量。这一判断“取决于”硬件是否可用。然而，无论是否有更强大的计算资源，数值稳定性或其缺失都是一个问题。■

注10。 定理 3 也适用于经验高阶影响函数估计量（Liu 等人， 2017）以及 Robins 等人 （2016年）提出的原始 HOIF 估计量；这些估计量需要估计协变量的密度 X，但不计入密度估计所需的时间。■

4 sHOIF估计量统计性质的应用

4.1 在对干扰函数仅作最小霍尔德假设时的半参数效率

在非参数统计学中，统计程序的最优性通常是在霍尔德干扰模型下评估的。

上述计算最终导出了以下定理，它是本文的第二个主要结果。

定理 4。 若 $\mathcal{A}\times\mathcal{B}\subseteq\mathcal{H}(s_a,\mathcal{X})\times\mathcal{H}(s_b,\mathcal{X})$ 满足 $(s_a+s_b)/2\geq d/4$，并选择 $m\asymp\sqrt{\log n}$和 $k\lesssim\overline{n/\log(n)^2}$,

$$\sqrt{n}\left(\widehat{\psi}_{m,k}-\psi(\theta)\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\mathbb{E}[\mathsf{IF}_1(\theta)^2]) \tag{27}$$

其中 $\mathbb{E}[$影响函数$_1(\theta)^2]$ 是 $\psi(\theta)$ 的半参数效率界。

注 11。 根据 Robins 等人(2009年) 在霍尔德干扰模型下给出的下界，$(s_a + s_b)/2 \geq d/4$ 是存在 (θ) 的半参数有效估计量的最小条件。可以合理预期，这一最小条件也适用于大多数（如果不是全部）双重稳健泛函。■

4.2 对Liu 等人(2020a）和Liu 等人(2021b）的无假设偏差检验程序的启示

鉴于人们日益关注基于深度学习的因果推断（Farrell 等人，2021；Chen 等人，2020年）的性能，以及这些理论结果与实证表现之间的差距（Xu 等人，2022），Liu 等人(2020a) 提出了以下预言机无假设、有效且名义水平为 α 的检验统计量：

$$\widehat{\chi} := \mathbb{I}\left\{\frac{\widehat{\mathbb{IF}}_{2,2,k}}{\widehat{\mathsf{se}}[\widehat{\psi}_1]} - z_{\alpha/2}\frac{\widehat{\mathsf{se}}[\widehat{\mathbb{IF}}_{2,2,k}]}{\widehat{\mathsf{se}}[\widehat{\psi}_1]} > \delta\right\} \tag{28}$$

针对以下原假设：

$$\mathsf{H}_0(\delta) : \frac{\mathsf{cs\text{-}bias}(\widehat{\psi}_1)}{\mathsf{se}[\widehat{\psi}_1]} \leq \delta \tag{29}$$

其中

$$\mathsf{cs\text{-}bias}(\widehat{\psi}_1) := \left\{\mathbb{E}\left[\lambda(X)(\widehat{a}(X) - a(X))^2\right]\mathbb{E}\left[\lambda(X)(\widehat{b}(X) - b(X))^2\right]\right\}^{1/2}. \tag{30}$$

Liu 等人(2021b) 随后构造了一个可行、假设较少、有效且名义水平为 α 的检验统计量

$$\widehat{\chi}_{3,k}(\widehat{\Omega}_{\text{nuis}}) := \mathbb{I}\left\{\frac{\widehat{\mathbb{IF}}_{(2,2)\to(3,3),k}(\widehat{\Omega}_{\text{nuis}})}{\widehat{\mathsf{se}}[\widehat{\psi}_1]} - z_{\alpha/2}\frac{\widehat{\mathsf{se}}[\widehat{\mathbb{IF}}_{(2,2)\to(3,3),k}(\widehat{\Omega}_{\text{nuis}})]}{\widehat{\mathsf{se}}[\widehat{\psi}_1]} > \delta\right\} \tag{31}$$

以及以下基于 eHOIF 估计量的高阶检验统计量：

$$\widehat{\chi}_{m,k}(\widehat{\Omega}_{\text{nuis}}) := \mathbb{I}\left\{\frac{\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega}_{\text{nuis}})}{\widehat{\mathsf{se}}[\widehat{\psi}_1]} - z_{\alpha/2}\frac{\widehat{\mathsf{se}}[\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega}_{\text{nuis}})]}{\widehat{\mathsf{se}}[\widehat{\psi}_1]} > \delta\right\}. \tag{32}$$

Liu 等人（2021b）证明，上述检验统计量中的所有标准误都可以通过某种自助法程序进行一致估计。更重要的是，他们证明了以下结论。

命题 4. 令

$$\mathsf{cs\text{-}bias}_k(\widehat{\psi}_1) := \left\{\mathbb{E}\left[\Pi[(A\widehat{a} - 1)|\bar{\mathsf{z}}_k](X)^2\right]\mathbb{E}\left[\Pi[A(\widehat{b} - b)|\bar{\mathsf{z}}_k](X)^2\right]\right\}^{1/2}.$$

在定理 4 的假设下，若有 $k \lesssim n/(\log n)^2$ 并满足以下附加条件：

$$|\mathsf{kern\text{-}bias}_{3,k}(\widehat{\psi}_1)| \ll \mathsf{cs\text{-}bias}_k(\widehat{\psi}_1) \tag{33}$$

^则 $\widehat{\chi}_{3,\ k}(\Omega_{tr})$ 是一个有效的名义 α-水平检验，用于检验 $\mathsf{H}_0(\delta)$（29）。附加条件 (33）可放宽为

$$|\mathsf{kern\text{-}bias}_{m,k}(\widehat{\psi}_1)| \not\ll \mathsf{cs\text{-}bias}_k(\widehat{\psi}_1)\left(\frac{k\log k}{n}\right)^{\frac{m-1}{2}} \tag{34}$$

^ ^如果使用 $\widehat{\chi}_{m,\ k}(\Omega_{tr})$ </style}而不是 ,

$_k(\Omega_{tr})$</style}。

类似地，我们可以定义以下基于 sHOIF 的检验统计量：对于 $m \geq 2$,

$$\widehat{\chi}_{m,k}(\widehat{\Omega}) := \mathbb{I}\left\{\frac{\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega})}{\widehat{\mathsf{se}}[\widehat{\psi}_1]} - z_{\alpha/2}\frac{\widehat{\mathsf{se}}[\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega})]}{\widehat{\mathsf{se}}[\widehat{\psi}_1]} > \delta\right\}.$$

于是，作为定理 1的直接推论，我们有

定理 5。 在定理 4的假设下，令 $k \lesssim n/(\log n)^2$，并采用条件 (34)的另一个放宽版本：

$$|\mathsf{kern\text{-}bias}_{m,k}(\widehat{\psi}_1)| \not\ll \mathsf{cs\text{-}bias}_k(\widehat{\psi}_1)\left(\frac{k}{n}\right)^{m-1} \tag{35}$$

^则 $\widehat{\chi}_{m,\ k}(\Omega)$ 是一个有效的名义 α水平检验，用于检验 $\mathsf{H}_0(\delta)$ (29)。

鉴于上述理论保证，并进一步考虑到 sHOIF 估计量和检验的有限样本性能优于相应的经验高阶影响函数估计量和检验，我们建议在实践中使用 ^量和检验，我们建议使用 $\widehat{\chi}_{m,\ k}(\Omega)$。有关其应用的更多示例，请参见 Wanis 等人(2023)。

5 sHOIF 估计量的进一步扩展

5.1 推广至整个双重稳健泛函类

在本小节中，我们简要说明如何将我们的结果推广至 Rotnitzky 等人 (2021所刻画的整个双重稳健泛函类。双重稳健泛函类包含许多源于（生物）统计学、流行病学、经济学和社会科学中实质性研究的其他泛函，包括：

- 期望条件方差，可用于构造置信集/预测集（Robins 和 van der Vaart，2006）；
- 条件协方差的期望，可同时用于因果推断和条件独立性检验（Shah 和 Peters，2020）；
- 连续治疗的平均因果效应，这对于处理分配至关重要（Ai 等人，2021；Bonvini 和 Kennedy，2022）。

Rotnitzky 等人（2021）将双重稳健泛函定义如下：

定义 1（双重稳健泛函；Rotnitzky 等人（2021）的定义 1）。如果对于每个 $\theta \in \Theta$，存在 $a: x \mapsto a(x) \in \mathsf{A}$，以及 $b: x \mapsto b(x) \in \mathsf{B}$，使得（i） $\theta = (a, b, \theta \setminus \{b, p\})$，且 $\Theta = \mathsf{A} \times \mathsf{B} \times \Theta \setminus \{\mathsf{A}, \mathsf{B}\}$；以及（ii）对于任意 θ、$\theta' \in \Theta$， $\psi(\theta)$ 是双重稳健泛函

$$\psi(\theta) - \psi(\theta') + \mathbb{E}\left[\mathsf{IF}_1(\theta')\right] = \mathbb{E}\left[S(a(X) - a'(X))(b(X) - b'(X))\right] \tag{36}$$

其中， $S \equiv s(O)$ 是一个已知函数，与 θ 和 θ' 无关，并满足 $\mathbb{P}_\theta(S \geq 0) = 1$ 或 $\mathbb{P}_\theta(S \leq 0) = 1$。我们还记 $\lambda(x) := \mathbb{E}[S|X = x]$。那么， $\psi(\theta)$ 的一阶影响函数形式如下：给定 $\theta' \equiv (a',\ b',\ \theta' \setminus \{a',\ b'\})^\top \in \Theta$,

$$\mathsf{IF}_1(\theta') \equiv Sa'(X)b'(X) + m_a(O, a') + m_b(O, b') + S_0 \tag{37}$$

其中， S_0 是某个已知统计量，与 a' 和 b' 无关； $h \mapsto m_a(o,\ h)$（对应$h \in \mathsf{A}$）和 $h \mapsto m_b(o,\ h)$（对应 $h \in \mathsf{B}$）是两个已知的线性映射，满足

$$\begin{aligned}
&\mathbb{E}\left[Sh(X)b(X) + m_a(O, h)\right] \equiv 0,\ \forall\ h \in \mathcal{A}, \\
&\mathbb{E}\left[Sa(X)h(X) + m_b(O, h)\right] \equiv 0,\ \forall\ h \in \mathcal{B}.
\end{aligned}$$

因此， $\psi(\theta) \equiv \mathbb{E}[m_a(O,\ a) + S_0] \equiv \mathbb{E}[m_b(O,\ b) + S_0]$。

注释 12. 对于在强可忽略性条件下的 $\psi(\theta) \equiv \mathbb{E}[Y(1)]$， S、 a、 b、 $m_a(O,\ a)$、 $m_b(O,\ b)$ 和 S_0 分别对应于 $-A$、$\{\mathbb{E}[A|X]\}^{-1}$、 $\mathbb{E}[Y|X, A = 1]$、 $AYa(X)$、 $b(X)$ 和 0。因此，$\lambda(X) = \mathbb{E}[-A|X] = -\ {}_a^{\ 1}{}_{X()}$。我们还有

$$\begin{aligned}
\mathbb{E}[\mathcal{E}_a(\widehat{a}; O)\bar{\mathsf{z}}_k(X)] &= \mathbb{E}\left[\frac{1}{a(X)}(\widehat{a}(X) - a(X))\bar{\mathsf{z}}_k(X)\right], \\
\mathbb{E}[\mathcal{E}_b(\widehat{b}; O)\bar{\mathsf{z}}_k(X)] &= \mathbb{E}\left[\frac{1}{a(X)}(\widehat{b}(X) - b(X))\bar{\mathsf{z}}_k(X)\right].
\end{aligned}$$

■

我们有以下记号对应关系，可将 $\psi(\theta) \equiv \mathbb{E}[Y(1)]$ 在强可忽略性假设下的结果推广到任意双重稳健泛函 $\psi(\theta)$:

- $\mathrm{E}_a(\widehat{a};\ O)^{-}\mathsf{z}_k(X) \Rightarrow \mathrm{E}_a(\widehat{a},\ {}^{-}\mathsf{z}_k;\ O)$ 和 $\mathrm{E}_b(b;\ O)^{-}\mathsf{z}_k(\widehat{X}) \Rightarrow \mathrm{E}_b(b,\ {}^{-}\mathsf{z}_k;\ O)$, 其中 $\mathrm{E}_a(\widehat{a},\ {}^{-}\mathsf{z}_k;\ O)$ 和 $\widehat{\ }\mathrm{E}_b(b,\ {}^{-}\mathsf{z}_k;\ O)$ 满足

$$\begin{aligned}
\mathbb{E}[\mathcal{E}_a(\widehat{a}, \bar{\mathsf{z}}_k; O)] &= \mathbb{E}\left[\lambda(X)(\widehat{a}(X) - a(X))\bar{\mathsf{z}}_k(X)\right], \\
\mathbb{E}[\mathcal{E}_b(\widehat{b}, \bar{\mathsf{z}}_k; O)] &= \mathbb{E}\left[\lambda(X)(\widehat{b}(X) - b(X))\bar{\mathsf{z}}_k(X)\right].
\end{aligned}$$

- $\Sigma = \mathbb{E}[A^{-}{}_k\ X\ {}^{-}{}_k\ X\ {}^{\top}] \Rightarrow \Sigma = \mathbb{E}[S^{-}{}_k\ X\ {}^{-}{}_k\ X\ {}^{\top}]$ $\widehat{\Sigma} = \mathbb{P}_n[A^{-}{}_k\ X\ {}^{-}{}_k\ X\ {}^{\top}] \Rightarrow \mathsf{z}\,(\,)\mathsf{z}\,(\,)\,\mathsf{z}\,(\,)\mathsf{z}\,(\,)$ 和 $\mathsf{z}\,(\,)\mathsf{z}\,(\,)\,\mathsf{z}\,(\,)\mathsf{z}(\,)$。 $\widehat{\Sigma} = \mathbb{P}_n[S^{-}{}_k\ X\ {}^{-}{}_k\ X\ {}^{\top}]$

通过上述映射，本文中针对 $\psi(\theta) \equiv \mathbb{E}[Y(1)]$ 推导出的所有理论结果，均可经适当调整后应用于任意双重稳健泛函 $\psi(\theta)$。

5.1.1 一个特殊情形：条件协方差的期望

在结束本文之前，我们进一步研究迄今发展的稳定高阶影响函数理论对于一类特殊双重稳健泛函的启示：两个随机变量 A 和 Y 在给定第三个随机变量 X 时的 ψ 条件协方差的期望：$=\mathbb{E}[\text{协方差}(A,\ Y|X)]$。当 $A=Y$ 几乎必然成立时，ψ 可化为 A 在给定 X 时的 ψ 期望条件方差：$=\mathbb{E}[\text{协方差}(A|X)]$。关于这两个参数之间的差异，详见Liu 等人(2020a)中的扩展讨论。

使 ψ 区别于许多其他双重稳健泛函的主要特征是 $S=1$，由此得到其二阶影响函数：

$$\widehat{\mathbb{IF}}_{2,2,k}=\frac{1}{n(n-1)}\sum_{1\le i_1\neq i_2\le n}(A_{i_1}-\widehat{a}(X_{i_1}))\bar{\mathsf{z}}_k(X_{i_1})^\top\bar{\Omega}\bar{\mathsf{z}}_k(X_{i_2})(Y_{i_2}-\widehat{b}(X_{i_2})),$$

其中 $\bar{\Omega}\equiv\{\mathbb{E}[S\bar{\mathsf{z}}_k(X)\bar{\mathsf{z}}_k(X)^\top]\}^{-1}\equiv\{\mathbb{E}[\bar{\mathsf{z}}_k(X)\bar{\mathsf{z}}_k(X)^\top]\}^{-1}$ 仅取决于 X 的分布 $\widehat{\ }$。此外，在此情形下，$\widehat{a}$ 和 b 是 $a(x)=\mathbb{E}[A|X=x]$ 和 $b(x)=\mathbb{E}[Y|X=x]$ 的干扰参数估计。由此得到改进后的核估计偏差界：

推论 1. 在假设 *1*–*2*下，若 $k\lesssim n^{\ }/\log^2 n$ 且 $m\gtrsim\sqrt{\log n,}$ 则有：$\widehat{\ }\quad\overline{\quad\quad}\quad\widehat{\ }$ $\mathbb{IF}_{(2,2)\to(m,m)}$的核估计偏差，$k(\Omega)$ 满足

$$\begin{aligned}&\text{kern-bias}_{m,k}(\widehat{\psi}_1):=\mathbb{E}\left[\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega})\right]-\text{bias}_{\theta,k}(\widehat{\psi}_1)\equiv\mathbb{E}\left[\widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega})-\widehat{\mathbb{IF}}_{2,2,k}\right]\\&\lesssim\left(\frac{km}{n}\right)^{\lceil\frac{\lceil\frac{m-1}{2}\rceil-1}{2}\rceil\vee1}\left\{\|\widehat{a}-a\|_2\|\widehat{b}-b\|_2+\left(\|\widehat{a}-a\|_2\left\|\widehat{b}-b\right\|_\infty\wedge\|\widehat{a}-a\|_\infty\left\|\widehat{b}-b\right\|_2\right)\right\}.\end{aligned}\tag{38}$$

注意，方差界也可类似地改进，此处省略。

6 讨论

本文提出了一类新型 HOIF 估计量，即稳定高阶影响函数（sHOIF）估计量，用于估计 Rotnitzky 等人（2021）所刻画的双重稳健泛函（DRF）。在 Robins 等人（2009）的最小 Hölder 光滑性条件 $\frac{s_a+s_b}{2}>\frac{d}{4}$ 下，这类估计量具有半参数有效性，并允许基函数的维数 k 以略慢于样本量 n的速率发散。由定理 1 可见，sHOIF 估计量的收敛速率优于 Liu 等人（2017）提出的经验高阶影响函数（eHOIF）。更重要的是，正如 Liu 等人（2020a）和 Wanis 等人（2023）的模拟研究充分显示的那样，实际应用中，sHOIF 估计量的有限样本性能也显著优于现有的高阶估计量，因此更适用于诸如检验因果效应的一阶去偏机器学习估计量 ψ_1 的偏差是否被其标准误所主导（Liu 等人，2020a，2021b；Wanis 等人，2023）等任务。最后，本文以展望若干未来研究方向作结：

(1) 值得研究的是，能否将 sHOIF 估计量的思路推广到非$\sqrt{n}$可估情形，例如通过某些收缩或正则化算法估计 Ω。正如 Robins 等人 (2016) 所推测的那样，本文研究的泛函的极小极大收敛速率可能取决于协变量密度的正则性 X。因此，收缩或正则化方法预计也会取决于 X的密度。 Liu 等人 (2020a) 的模拟研究表明，Ledoit 和 Wolf (2012) 提出的非线性收缩协方差矩阵估计量可能是一种可行选择。 Liu 等人 (2020a) 和 Wanis 等人 (2023) 的初步模拟研究表明，这些收缩协方差矩阵估计量的性能确实会随着设计密度光滑度的变化而下降。

(2) 正如 Kennedy 等人 (2022) 指出的，他们用于条件平均处理效应的二阶 R 学习器（SORL）也需要在协变量满足额外的降低复杂度假设下，对某些基函数（他们使用勒让德多项式）的格拉姆矩阵求逆 X。值得研究 sHOIF 估计量能否推广到条件平均处理效应估计问题，并稳定其 SORL 乃至 HORL 估计量。

(3) van der Vaart(2014) 也提到了另一个重要的开放问题；Liu 等人(2020a，2021b)。要为双重稳健泛函定义高阶影响函数，需要事先选择一组 k维基函数 $\bar{z}_k$，或选择一个克罗内克 δ 函数的逼近核 K，最好在数据分析之前完成。然而，这种策略似乎与当前的数据分析范式背道而驰；该范式强烈倡导从数据中自适应地学习表示（例如基或核），而不是先验地选择一些固定的基/框架 先验地。典型例子包括深度神经网络、自编码器和生成对抗网络。因此，沿不同基方向构造高阶影响函数估计量，再依据某种最优性准则选择其中一个或将它们全部聚合起来，是很有意义的。我们将这一重要而困难的问题留待未来研究。

（4）此外，推导具有潜变量的图模型中可识别因果效应泛函的高阶影响函数和高阶平滑影响函数（Bhattacharya 等人，2022）以及隐式定义的泛函（Robins 等人，2016年；Ai 等人，2021），以改进一般半参数回归问题中的估计质量和统计推断，也将是很有意义的。然而，这需要将当前工作扩展至 U过程，这是一个困难得多的研究问题，我们正在另一篇论文中对此进行研究。

参考文献

Chunrong Ai、Oliver Linton、Kaiji Motegi 和 Zheng Zhang。一般处理模型的高效估计统一框架。《Quantitative Economics》，12（3）：779–816，2021。

Afonso S Bandeira、March T Boedihardjo 和 Ramon van Handel。《矩阵集中不等式与自由概率》。arXiv 预印本 arXiv:2108.06312，2021。

Alexandre Belloni、Victor Chernozhukov、Denis Chetverikov 和 Kengo Kato。《最小二乘级数的一些新渐近理论：点态与一致性结果》。《计量经济学杂志》，18~~6(2):345–36~~6，2015。Rabi N Bhattacharya 和 Jayanta K Ghosh。《一类 U-统计量与 k-簇数量的渐近正态性》。《多元分析杂志》，43(2):300–330，1992。Rohit Bhattacharya、Razieh Nabi 和 Ilya Shpitser。《具有潜变量的图模型中的因果效应半参数推断》。《机器学习研究杂志》，23:1–76，2022。Peter J Bickel、Chris A J Klaassen、Ya'acov Ritov 和 Jon A Wellner。《半参数模型的高效与自适应估计》。约翰斯·霍普金斯数学科学丛书。施普林格纽约，1998。ISBN 9780387984735。Matteo Bonvini 和 Edward H Kennedy。《剂量-反应估计的快速收敛速率》。arXiv 预印本 arXiv:2207.11825，2022。Xiaohong Chen、Ying Liu、Shujie Ma 和 Zheng Zhang。《使用具有发散数量混杂因素的神经网络进行一般处理效应的因果推断》。arXiv 预印本 arXiv:2009.07055，2020。Victor Chernozhukov、Denis Chetverikov、Mert Demirer、Esther Duflo、Christian Hansen、Whitney Newey 和 James Robins。《用于处理与结构参数的双重/去偏机器学习》。《计量经济学杂志》，21(1):C1–C68，2018a。Victor Chernozhukov、Whitney Newey 和 James Robins。《使用正则化黎斯表示元进行双重/去偏机器学习》。arXiv 预印本 arXiv:1802.08667，2018b。Romain Couillet 和 Zhenyu Liao。《机器学习的随机矩阵方法》。剑桥大学出版社，2022。Max H Farrell、Tengyuan Liang 和 Sanjog Misra。《用于估计与推断的深度神经网络》。《计量经济学》，89(1):181–213，2021。David A Hirshberg 和 Stefan Wager。《增广极小极大线性估计》。《统计年刊》，49(6):3206–3227，2021。Edward H Kennedy。《异质性因果效应的最优双重稳健估计》。arXiv 预印本 arXiv:2004.14497，2020。Edward H Kennedy、Sivaraman Balakrishnan 和 Larry Wasserman。《异质性因果效应估计的极小极大速率》。arXiv 预印本 arXiv:2203.00837，2022。Olivier Ledoit 和 Michael Wolf。《大维协方差矩阵的非线性收缩估计》。《统计年刊》，40(2):1024–1060，2012。

Oleg V Lepskii. 关于高斯白噪声中自适应估计问题. 《概率论及其应用》, 35(3):454–466, 1991. Chang Li and Lin Liu. “新的数值稳定、半参数有效经验高阶影响函数”的补充材料. 2023. Lin Liu, Rajarshi Mukherjee, Whitney K Newey, and James M Robins. 半参数有效经验高阶影响函数估计量. arXiv 预印本 arXiv:1705.07577, 2017. Lin Liu, Rajarshi Mukherjee, and James M Robins. 关于机器学习估计因果参数的名义置信区间覆盖率的近乎无假设检验. 《统计科学》, 35(3):518–539, 2020a. Lin Liu, Rajarshi Mukherjee, and James M Robins. 答辩：关于机器学习估计因果参数的名义置信区间覆盖率的近乎无假设检验. 《统计科学》, 35(3):545–554, 2020b. Lin Liu, Rajarsh Mukherjee, James M Robins, and Eric Tchetgen Tchetgen. 非参数泛函的自适应估计. 《机器学习研究杂志》, 22(99):1–66, 2021a. Lin Liu, Rajarshi Mukherjee, and James M Robins. 我们能否在不作任何假设的情况下，判断双重稳健泛函的 Wald 置信区间有效性的论证可能有误？技术报告，上海交通大学，2021b. 网址 https://linliu-stats.github.io/files/Oct24-Assumption-freetestecon(reftogether).pdf . Whitney K Newey. 半参数效率界. 《应用计量经济学杂志》, 5 (2):99–135, 1990. Whitney K Newey and James M Robins. 交叉拟合与半参数估计的快速余项收敛率. arXiv 预印本 arXiv:1801.09138, 2018. Victor H Peña, Tze Leung Lai, and Qi-Man Shao. 自归一化过程：极限定理与统计应用. Springer Science & Business Media, 2008. James Robins and Aad van der Vaart. 自适应非参数置信集. 《统计年刊》, 34(1):229–253, 2006. James Robins, Lingling Li, Eric Tchetgen Tchetgen, and Aad van der Vaart. 高阶影响函数与非线性泛函的极小极大估计. 载于《概率与统计：纪念 David A. Freedman 文集》，第335–421页. Institute of Mathematical Statistics, 2008. James Robins, Eric Tchetgen Tchetgen, Lingling Li, and Aad van der Vaart. 半参数极小极大收敛率. 《电子统计学杂志》, 3:1305–1321, 2009.

James Robins, Lingling Li, Eric Tchetgen Tchetgen, and Aad van der Vaart. 技术报告：高阶影响函数与非线性泛函的极小极大估计. arXiv 预印本 arXiv:1601.05820, 2016. James M Robins. 最优序贯决策的最优结构嵌套模型. 载于《西雅图生物统计学第二届研讨会论文集》, 第189–326页. Springer, 2004. James M Robins and Andrea Rotnitzky. 对“半参数模型推断：若干问题与一个答案”的评论. 《统计学报》, 11(4):920–936, 2001. James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. 当某些回归变量并非总能被观测到时回归系数的估计. 《美国统计学会会刊》, 89(427):846–866, 1994. James M Robins, Lingling Li, Lin Liu, Rajarshi Mukherjee, Eric Tchetgen Tchetgen, and Aad van der Vaart. 结构化高维模型上泛函的极小极大估计. 《统计年刊》, 45(5):1951–1987, 2017. James M Robins, Lingling Li, Lin Liu, Rajarshi Mukherjee, Eric Tchetgen Tchetgen, and Aad van der Vaart. “结构化高维模型上泛函的极小极大估计”的勘误. 2022. Andrea Rotnitzky, Ezequiel Smucler, and James M Robins. 具有混合偏差性质的参数刻画. Biometrika, 108(1):231–238, 2021. Mark Rudelson. 各向同性位置中的随机向量. 《泛函分析杂志》, 164(1):60–72, 1999. Daniel O Scharfstein, Andrea Rotnitzky, and James M Robins. 使用半参数无响应模型校正不可忽略失访. 《美国统计学会会刊》, 94(448):1096–1120, 1999. Rajen D Shah and Jonas Peters. 条件独立性检验的难度与广义协方差测度. 《统计年刊》, 48(3):1514–1538, 2020. Eric J Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao. 近端因果学习导论. arXiv 预印本 arXiv:2009.10982, 2020. Aad van der Vaart. 关于可微泛函. 《统计年刊》, 19(1):178–204, 1991. Aad van der Vaart. 第三部分：半参数统计. 《概率论与统计学讲义》, 第331–457页, 2002. Aad van der Vaart. 高阶切空间与影响函数. 《统计科学》, 29(4):679–686, 2014.

Kerollos Wanis、Lin Liu、Nelya Melnitchoukc 和 James M Robins。机器学习与因果推断：利用高阶影响函数量化偏差。技术报告，哈佛大学，2023。Richard P Waterman 和 Bruce G Lindsay。用于近似条件得分的投影得分方法。Biometrika，83(1):1–13，1996。Siqi Xu、Lin Liu 和 Zhonghua Liu。DeepMed：利用去偏深度学习进行半参数因果中介分析。神经信息处理系统进展，36，2022。Ruoqi Yu 和 Shulei Wang。利用均匀变换器估计治疗效应。arXiv 预印本 arXiv:2008.03738，2020。

附录 A 命题方差部分的证明1

A.1 引理的证明1

我们先证明第二个陈述（20）。

$$\begin{aligned}
&\left|\mathbb{E}\left[\bar{\mathsf{z}}_k(X_1)^\top \bar{\mathsf{z}}_k(X_2)\bar{\mathsf{z}}_k(X_3)^\top Q_{3,4}^j \bar{\mathsf{z}}_k(X_4)\right]\right| \\
=\ &\left|\mathbb{E}[\bar{\mathsf{z}}_k(X_1)]^\top \mathbb{E}[\bar{\mathsf{z}}_k(X_2)]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_3)^\top Q_{3,4}^j \bar{\mathsf{z}}_k(X_4)\right]\right| \\
\lesssim\ &\left|\mathbb{E}\left[\bar{\mathsf{z}}_k(X_3)^\top Q_{3,4}^j \bar{\mathsf{z}}_k(X_4)\right]\right| \\
\leq\ &\left\{\mathbb{E}\left[\bar{\mathsf{z}}_k(X_4)^\top Q_{3,4}^j \bar{\mathsf{z}}_k(X_3)\bar{\mathsf{z}}_k(X_3)^\top Q_{3,4}^j \bar{\mathsf{z}}_k(X_4)\right]\right\}^{1/2} \lesssim k^j.
\end{aligned}$$

对于第一个陈述（19），由于 $j_1 > 0$，在 $\bar{}\mathsf{z}_k(X_3)$ 与 $\bar{}\mathsf{z}_k(X_4)$ 之间必定至少存在一个 $Q_{1,2}$，使得（19）成立。我们对 j 进行归纳法。$j = 1$ 已在正文中证明。假设（19）对 $j-1$ 成立。那么，应用柯西-施瓦茨不等式，可以很容易地证明（19）对 j成立。

附录 B （(23)的推导

回顾一下，在 $j = 2$ 水平截断的 $\mathbb{IF}_{2,2,k}(\Omega)$ 的核估计偏差 kern-bias$_{2,k}(\psi_1)$ 被以下表达式支配：

$$\frac{1}{n^2}\mathbb{E}\left[\mathcal{E}_a(\hat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top Q_{1,2}^2 \bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\hat{b};O_2)\right] + \frac{1}{n}\mathbb{E}\left[\mathcal{E}_a(\hat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\right]\mathbb{E}[Q_3^2]\mathbb{E}\left[\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\hat{b};O_2)\right] \tag{39}$$

这等于（11）与（14）之和。

我们进一步考察将 $\mathbb{IF}_{3,3,k}(\Omega)$ 作为 (10) 的估计量时的核估计偏差，即截至水平 $j = 1$ 截断的核偏差 $_{2,k}(\psi_1)$ 的主导项：

$$\begin{aligned}
&\mathbb{E}\left[\widehat{\mathbb{IF}}_{3,3,k}(\widehat{\Omega}) - (10)\right] \\
=\ &\mathbb{E}\left[\mathcal{E}_a(\hat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\left(Q_{1,2}^2 - \widehat{\Omega}Q_{1,2}^2\widehat{\Omega}\right)\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\hat{b};O_2)\right] + O(n^{-1}) \\
=\ &-\frac{2}{n^2}\mathbb{E}\left[\mathcal{E}_a(\hat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top Q_{1,2}^2 \bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\hat{b};O_2)\right] + \mathsf{Rem} + O(n^{-1})
\end{aligned}$$

其中，余项 Rem 可以证明由第一项所支配。将这一项加到 (39) 上，我们得出结论：$\mathbb{IF}_{(2,2)\to(3,3),k}(\Omega)$ 的核估计偏差的主导部分与 $\mathbb{E}[\mathbb{IF}_{4,4,k}]$ 相消。

附录C 定理 3.2 中核估计偏差界的证明

我们将定理 3.2 的证明分为几个步骤。首先，在第 C.1 节中，我们给出 m 阶高阶平滑影响函数 $\mathbb{IF}_{(2,2)\to(m,m),\ k}(\Omega)$ 的另一种刻画，以便控制偏差

。

C.1 高阶平滑影响函数的另一种刻画

我们可以通过归纳法证明以下对 m 阶高阶平滑影响函数的另一种刻画：

$$
\begin{aligned}
&\widehat{\mathbb{IF}}_{2,2,k} - \widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega}) \\
\equiv\ & \mathbb{U}_{n,2}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\left\{\sum_{j=0}^{m-1}(-1)^j\binom{m-1}{j}\mathbb{U}_{n-2,j-1}\left(\widehat{\Omega}\prod_{s=3}^{j+1}Q_s\widehat{\Omega}\right)\right\}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right] \\
\equiv\ & \mathbb{U}_{n,2}\left[\mathcal{E}_a(\widehat{a};O_1)\bar{\mathsf{z}}_k(X_1)^\top\left\{\sum_{j=1}^{m-1}(-1)^j\binom{m-1}{j}\mathbb{U}_{n-2,j}\left(\widehat{\Omega}\prod_{s=3}^{j+1}Q_s\widehat{\Omega}-\mathbb{I}\right)\right\}\bar{\mathsf{z}}_k(X_2)\mathcal{E}_b(\widehat{b};O_2)\right]
\end{aligned}
\tag{40}
$$

其中，$\widehat{\Omega}\prod_{s=3}^{2}Q_s\widehat{\Omega}$ 和 $\widehat{\Omega}\prod_{s=3}^{1}Q_s\widehat{\Omega}$ 分别视为 $\widehat{\Omega}$ 和 $\mathbb{I}$。

利用(2.10)，我们可以将核偏差刻画为

引理 2。

$$
\begin{aligned}
&\mathbb{E}\left[\widehat{\mathbb{IF}}_{2,2,k} - \widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega})\right] \\
=\ & \sum_{j=1}^{m-1}(-1)^j\binom{m-1}{j}\sum_{\ell=1}^{j-1}\sum_{\substack{S\subseteq[j-1],\\|S|=\ell}}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top\prod_{s=0}^{j-1}\left\{Q_s(\widehat{\Omega}-\mathbb{I})\mathbb{I}^{\{s\in S\}}\right\}\bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\right]
\end{aligned}
\tag{41}
$$

$$
= \sum_{\mathfrak{c}=1}^{m-1}\sum_{j=\mathfrak{c}}^{m-1}(-1)^j\binom{m-1}{j}\sum_{\substack{S\subseteq[j-1],\\|S|=\mathfrak{c}}}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top\prod_{s=0}^{j-1}\left\{Q_s(\widehat{\Omega}-\mathbb{I})\mathbb{I}^{\{s\in S\}}\right\}\bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\right].
\tag{42}
$$

证明。为避免符号说明过于繁杂，我们引入别名项 $Q_0 := \mathbb{I}$。对于任意非负整数 l，我们还将符号 $[l] := \{0, 1, \cdots l\}$ 重载使用。我们首先证明（41）：

$$
\begin{aligned}
&\mathbb{E}\left[\widehat{\mathbb{IF}}_{2,2,k} - \widehat{\mathbb{IF}}_{(2,2)\to(m,m),k}(\widehat{\Omega})\right] \\
=\ & \sum_{j=1}^{m-1}(-1)^j\binom{m-1}{j}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top\left(\prod_{s=0}^{j-1}(Q_s\widehat{\Omega})-\mathbb{I}\right)\bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\right] \\
=\ & \sum_{j=1}^{m-1}(-1)^j\binom{m-1}{j}\sum_{\ell=1}^{j-1}\sum_{\substack{S\subseteq[j-1],\\|S|=\ell}}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top\prod_{s=0}^{j-1}\left\{Q_s(\widehat{\Omega}-\mathbb{I})\mathbb{I}^{\{s\in S\}}\right\}\bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\right]
\end{aligned}
$$

其中，第二个等式是将每个 $\widehat{\Omega}$ 分解为 $\mathbb{I}-(\mathbb{I}-\widehat{\Omega})$（$j=2,\cdots m-1$）后得到的。

对于（42），我们通过根据乘积中 $\widehat{\Omega}$ 的副本数量 $\mathfrak{c}$ 来 $\mathbb{I}-$ 重新组织（41）中的所有求和项：

$$(41)=\sum_{\mathfrak{c}=1}^{m-1}\sum_{j=\mathfrak{c}}^{m-1}(-1)^j\binom{m-1}{j}\sum_{\substack{S\subseteq[j-1],\\|S|=\mathfrak{c}}}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top\prod_{s=0}^{j-1}\left\{Q_s(\widehat{\Omega}-\mathbb{I})\mathbb{I}^{\{s\in S\}}\right\}\bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\right].$$

□

C.2 通过矩阵展开和组合学进行分析

将核估计偏差中的不同项按（42）的方式重新组织后，我们进行如下展开 $\mathbb{I}-\widehat{\Omega}$：

$$\widehat{\Omega}-\mathbb{I}=\sum_{j=1}^{J}(\mathbb{I}-\widehat{\Sigma})^j+(\mathbb{I}-\widehat{\Sigma})^{J+1}\widehat{\Omega}. \tag{43}$$

我们将上述展开到 J 阶记为

$$[\widehat{\Omega}-\mathbb{I}]_J\coloneqq\sum_{j=1}^{J}\left(\mathbb{I}-\widehat{\Sigma}\right)^j.$$

然后，我们按 $\Sigma-\mathbb{I}$ 的副本数 $\mathfrak{c}$ 将不同项汇集在一起，得到以下引理。

引理 3。将 $\mathbb{I}-\widehat{\Omega}$ 替换为 $[\mathbb{I}-\widehat{\Omega}]_J$ 后，（42）可改写为 $\sum_{\mathfrak{c}=1}^{(m-1)J}\mathfrak{M}_\mathfrak{c}$，其中

$$\mathfrak{M}_\mathfrak{c}\coloneqq\sum_{j=1}^{m-1}(-1)^j\binom{m-1}{j}\sum_{\mathfrak{c}'=1}^{\mathfrak{c}\wedge j}\sum_{\substack{S\subseteq[j-1],\\|S|=\mathfrak{c}'}}\sum_{\substack{\{\ell_s,s\in S\}\subseteq\{1,\cdots,J\}^{\mathfrak{c}'},\\\sum_{s\in S}\ell_s=\mathfrak{c}}}\mathbb{E}\left[\begin{array}{c}\mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top\\\times\prod_{s=0}^{j-1}\left\{Q_s(\mathbb{I}-\widehat{\Sigma})^{\ell_s}\mathbb{I}^{\{s\in S\}}\right\}\\\times\ \bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\end{array}\right]. \tag{44}$$

证明。证明只需几行代数运算即可完成。

$$\sum_{\mathfrak{c}'=1}^{m-1}\sum_{j=\mathfrak{c}'}^{m-1}(-1)^j\binom{m-1}{j}\sum_{\substack{S\subseteq[j-1],\\|S|=\mathfrak{c}'}}\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top\prod_{s=0}^{j-1}\left\{Q_s[\widehat{\Omega}-\mathbb{I}]_\mathbb{J}^{\{s\in S\}}\right\}\bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\right]$$

$$=\sum_{\mathfrak{c}'=1}^{m-1}\sum_{j=\mathfrak{c}'}^{m-1}(-1)^j\binom{m-1}{j}\sum_{\substack{S\subseteq[j-1],\\|S|=\mathfrak{c}'}}\mathbb{E}\left[\begin{array}{c}\mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top\\\times\prod_{s=0}^{j-1}\left\{Q_s\left(\sum_{\ell_s=1}^{J}(\mathbb{I}-\widehat{\Sigma})^{\ell_s}\right)^{\mathbb{I}^{\{s\in S\}}}\right\}\\\times\ \bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\end{array}\right]$$

$$= \sum_{\mathfrak{c}=1}^{(m-1)J} \sum_{\mathfrak{c}'=1}^{(m-1)\wedge\mathfrak{c}} \sum_{j=\mathfrak{c}'}^{m-1} (-1)^j \binom{m-1}{j} \sum_{\substack{S\subseteq[j-1],\\ |S|=\mathfrak{c}'}} \sum_{\substack{\{\ell_s, s\in S\}\subseteq\{1,\cdots,J\}^{\mathfrak{c}'},\\ \sum_{s\in S}\ell_s=\mathfrak{c}}} \mathbb{E}\left[\begin{array}{c} \mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top \\ \times \prod_{s=0}^{j-1}\left\{Q_s(\mathbb{I}-\widehat{\Sigma})^{\ell_s \mathbb{I}\{s\in S\}}\right\} \\ \times\ \bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\end{array}\right]$$

$$= \sum_{\mathfrak{c}=1}^{(m-1)J} \sum_{j=1}^{m-1} (-1)^j \binom{m-1}{j} \sum_{\mathfrak{c}'=1}^{\mathfrak{c}\wedge j} \sum_{\substack{S\subseteq[j-1],\\ |S|=\mathfrak{c}'}} \sum_{\substack{\{\ell_s, s\in S\}\subseteq\{1,\cdots,J\}^{\mathfrak{c}'},\\ \sum_{s\in S}\ell_s=\mathfrak{c}}} \mathbb{E}\left[\begin{array}{c} \mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top \\ \times \prod_{s=0}^{j-1}\left\{Q_s(\mathbb{I}-\widehat{\Sigma})^{\ell_s \mathbb{I}\{s\in S\}}\right\} \\ \times\ \bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\end{array}\right].$$

□

为了进一步推进，我们还需要以下初等引理：

引理 4. 给定 j 非负整数 $\ell_0, \cdots \ell_{j-1} \in \mathbb{Z}_{\geq 0}$ 并令 $\ell_h := \sum_{s=0}^{h} \ell_s$ 对于 $h = 0, \cdots j-1$ 满足 $\ell_{-1} \equiv 0$,

$$\mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top \prod_{s=0}^{j-1}\left\{Q_s(\mathbb{I}-\widehat{\Sigma})^{\ell_s}\right\}\bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\right]$$
$$= n^{-\bar{\ell}_{j-1}} \sum_{i_1=1}^{n} \cdots \sum_{i_{\bar{\ell}_{j-1}}=1}^{n} \mathbb{E}\left[\mathcal{E}_a(\widehat{a};O_{m-1})\bar{\mathsf{z}}_k(X_{m-1})^\top \left\{\prod_{s=0}^{j-1} Q_s \prod_{h=\bar{\ell}_{s-1}+1}^{\bar{\ell}_s}(\mathbb{I}-Q_{i_h})\right\}\bar{\mathsf{z}}_k(X_m)\mathcal{E}_b(\widehat{b};O_m)\right]. \tag{45}$$

证明。不失一般性，取 $\mathrm{E}_a(\widehat{a};\, O) \equiv A, \mathrm{E}_b(b;\, O) \equiv 1$ 以简化表述。反复使用恒等式 $\mathbb{I}-\widehat{\Sigma} \equiv n^{-1}\sum_{i=1}^{n}(\mathbb{I}-Q_i)$，并结合对任意非负整数 ℓ 均有约定 $\prod_{h=\ell+1}^{\ell}(\cdot)_h \equiv \mathbb{I}$，可得

$$\mathbb{E}\left[A_{m-1}\bar{\mathsf{z}}_k(X_{m-1})^\top \prod_{s=0}^{j-1}\left\{Q_s(\mathbb{I}-\widehat{\Sigma})^{\ell_s}\right\}\bar{\mathsf{z}}_k(X_m)\right]$$
$$= n^{-\bar{\ell}_{j-1}}\mathbb{E}\left[A_{m-1}\bar{\mathsf{z}}_k(X_{m-1})^\top \left\{\prod_{s=0}^{j-1} Q_s \prod_{h=1}^{\ell_s}\left[\sum_{i_{s,h}=0}^{n}(\mathbb{I}-Q_{i_{s,\ell_s}})\right]\right\}\bar{\mathsf{z}}_k(X_m)\right]$$
$$= n^{-\bar{\ell}_{j-1}} \sum_{i_1=1}^{n} \cdots \sum_{i_{\bar{\ell}_{j-1}}=1}^{n} \mathbb{E}\left[A_{m-1}\bar{\mathsf{z}}_k(X_{m-1})^\top \left\{\prod_{s=0}^{j-1} Q_s \prod_{h=\bar{\ell}_{s-1}+1}^{\bar{\ell}_s}(\mathbb{I}-Q_{i_h})\right\}\bar{\mathsf{z}}_k(X_m)\right].$$

□

完成证明的过程中，经过上述准备步骤，下面的“消去引理”是第一个关键里程碑。

引理 5。对于满足 $\mathfrak{c} < \lceil (m-1)/2 \rceil$ 的副本数，

$$\text{Equation } (44) \equiv 0.$$

^证明。不失一般性地，我们再次设 $\mathcal{E}_a(\widehat{a};\ O) \equiv A$，$\mathcal{E}_b(b;\ O) \equiv 1$。借助引理 4，方程（44）中任意给定 $\mathfrak{c}$ 处的被加项可重写为

$$\sum_{j=1}^{m-1}(-1)^j\binom{m-1}{j}\sum_{\mathfrak{c}'=1}^{\mathfrak{c}\wedge j}\sum_{\substack{S\subseteq[j-1],\\|S|=\mathfrak{c}'}}\sum_{\substack{\{\ell_s,s\in S\}\subseteq\{1,\cdots,J\}^{\mathfrak{c}'},\\\sum_{s\in S}\ell_s=\mathfrak{c}}}\mathbb{E}\left[A_{m-1}\bar{\mathsf{z}}_k(X_{m-1})^\top\prod_{s=0}^{j-1}\left\{Q_s(\mathbb{I}-\widehat{\Sigma})^{\ell_s\mathbb{I}\{s\in S\}}\right\}\bar{\mathsf{z}}_k(X_m)\right] \tag{46}$$

$$= n^{-\mathfrak{c}}\sum_{i_1,\cdots,i_{\mathfrak{c}}=1}^{n}\sum_{j=1}^{m-1}\binom{m-1}{j}\sum_{\mathfrak{c}'=1}^{\mathfrak{c}\wedge j}\sum_{\substack{S\subseteq[j-1],\\|S|=\mathfrak{c}'}}\sum_{\substack{\{\ell_s,s\in S\}\subseteq\{1,\cdots,J\}^{\mathfrak{c}'},\\\sum_{s\in S}\ell_s=\mathfrak{c}}}\mathbb{E}\left[\begin{array}{c}A_{m-1}\bar{\mathsf{z}}_k(X_{m-1})^\top\\\times\prod_{s=0}^{j-1}(-Q_s)\prod_{h=\bar{\ell}_{s-1}+1}^{\bar{\ell}_s}(\mathbb{I}-Q_{i_h})\\\times\ \bar{\mathsf{z}}_k(X_m)\end{array}\right]$$

$$= n^{-\mathfrak{c}}\sum_{i_1,\cdots,i_{\mathfrak{c}}=1}^{n}\sum_{j=1}^{m-1}\binom{m-1}{j}\sum_{\mathfrak{c}'=1}^{\mathfrak{c}\wedge j}\sum_{\substack{S\subseteq[j-1],\\|S|=\mathfrak{c}'}}\sum_{\substack{\{\ell_s,s\in S\}\subseteq\{1,\cdots,J\}^{\mathfrak{c}'},\\\sum_{s\in S}\ell_s=\mathfrak{c}}}\mathbb{E}\left[\begin{array}{c}A_{m-1}\bar{\mathsf{z}}_k(X_{m-1})^\top\\\times\prod_{s=0}^{j-1}\left\{\begin{array}{c}(\mathbb{I}-Q_s)\prod_{h=\bar{\ell}_{s-1}+1}^{\bar{\ell}_s}(\mathbb{I}-Q_{i_h})\\-\prod_{h=\bar{\ell}_{s-1}+1}^{\bar{\ell}_s}(\mathbb{I}-Q_{i_h})\end{array}\right\}\\\times\ \bar{\mathsf{z}}_k(X_m)\end{array}\right]. \tag{47}$$

现在我们引入另一个辅助副本数 $\mathfrak{c}^\dagger$，将展开以下乘积后，上述公式中所有含有 $\mathfrak{c}^\dagger$ 个 $\mathbb{I}-\mathbb{Q}_s$ 的项汇集起来

$$\prod_{s=0}^{j-1}\left\{(\mathbb{I}-Q_s)\prod_{h=\bar{\ell}_{s-1}+1}^{\bar{\ell}_s}(\mathbb{I}-Q_{i_h})-\prod_{h=\bar{\ell}_{s-1}+1}^{\bar{\ell}_s}(\mathbb{I}-Q_{i_h})\right\} \tag{48}$$

在式（47）的期望中。令 $\mathsf{p}(n,k)$ 表示将 n划分为 k 个正整数的所有可能划分数。展开后，对于任意给定的 $\mathfrak{c}^\dagger$，期望均具有以下形式：

$$\mathbb{E}\left[A_{m-1}\bar{\mathsf{z}}_k(X_{m-1})^\top\prod_{s\in S\subseteq[j-1],|S|=\mathfrak{c}^\dagger}(\mathbb{I}-Q_s)\prod_{l=1}^{\mathfrak{c}}(\mathbb{I}-Q_{i_l})\bar{\mathsf{z}}_k(X_m)\right] \tag{49}$$

在不同的 $(\mathbb{I}-Q_s)$ 和 $(\mathbb{I}-Q_{i_l})$ 的阶数进行排列后。因此，对应期望的系数常数（同样，排列顺序不计）为

$$\sum_{j=1}^{m-1}\binom{m-1}{j}\left\{\sum_{\ell=1}^{\mathfrak{c}\wedge j}\binom{j}{\ell}\mathsf{p}(\mathfrak{c},\ell)\right\}(-1)^{j-\mathfrak{c}^\dagger}\binom{j}{\mathfrak{c}^\dagger}$$
$$=(-1)^{-\mathfrak{c}^\dagger}\sum_{j=0}^{m-1}(-1)^j\binom{m-1}{j}\binom{j+\mathfrak{c}-1}{j-1}\binom{j}{\mathfrak{c}^\dagger}$$

$$= \frac{(-1)^{-\mathfrak{c}^\dagger}}{\mathfrak{c}^\dagger!\mathfrak{c}!} \sum_{j=0}^{m-1} (-1)^j \binom{m-1}{j} j \prod_{\ell=-\mathfrak{c}^\dagger+1}^{\mathfrak{c}-1} (j+\ell) \tag{50}$$

其中，第一个等式由引理 S7 得出。系数常数只是展开（47）后各项的数量：在上述公式的第一行中，$\binom{m-1}{j}$ 来自（47）的$\binom{m-1}{j}$，$\left\{\sum_{\mathfrak{c}'=1}^{\mathfrak{c}\wedge j}\binom{j}{\mathfrak{c}'}\mathfrak{p}(\mathfrak{c},\mathfrak{c}')\right\}$ 由（47）的$\binom{m-1}{j}$ 之后的三个求和项产生，而 $(-1)^{j-\mathfrak{c}^\dagger}\binom{j}{\mathfrak{c}^\dagger}$ 则统计了展开（48）后含有 $\mathfrak{c}^\dagger$ 个 $(\mathbb{I}-Q_s)$ 的项数，其中 $s \in \{0,1,\cdots\ j-1\}$。

另一个重要的观察是，展开后，当 $\mathfrak{c}^\dagger > \mathfrak{c}$ 时，式（47）的期望恒等于零，因此无论系数常数为何，各项求和项都为零。基于这一观察，我们只需考虑 $\mathfrak{c}^\dagger \le \mathfrak{c}$ 的情形。通过初等计算可知，存在整数 $\gamma_1, \cdots\ \gamma_{\mathfrak{c}+\mathfrak{c}^\dagger-1}$ 使得

$$(50) = \frac{(-1)^{-\mathfrak{c}^\dagger}}{\mathfrak{c}^\dagger!\mathfrak{c}!} \sum_{j=0}^{m-1} (-1)^j \binom{m-1}{j} \left(j^{\mathfrak{c}+\mathfrak{c}^\dagger} + \sum_{\ell=1}^{\mathfrak{c}+\mathfrak{c}^\dagger-1} \gamma_\ell j^\ell \right).$$

因此，当

$$m-1-(\mathfrak{c}+\mathfrak{c}^\dagger) > 0, \tag{51}$$

（50）$\equiv 0$，通过像引理 S8 中那样对二项式恒等式求导即可得到。由于我们假设 $\mathfrak{c} \ge \mathfrak{c}^\dagger$，$\mathfrak{c} < \lceil (m-1)/2 \rceil$ 足以使式（51）成立。证明至此完成。 □

根据引理 5，接下来的重要观察总结了证明。

引理 6。 对于满足 $\mathfrak{c} \ge \lceil (m-1)/2 \rceil$ 的副本数，

$$|\text{Equation (44)}| \le \left(\frac{km}{n}\right)^{\lceil \frac{\mathfrak{c}-1}{2} \rceil \vee 1}.$$

证明。 本证明沿用引理 5 证明中定义的符号。我们考虑 $\mathfrak{c} \ge \lceil (m-1)/2 \rceil$ 的情形；如前一个引理所示，该式并非恒等于零。我们需要计算非零期望的个数，而使用表示式（46）更容易完成这一计算。

对于任意给定的副本数 $\mathfrak{c}$，形式为 (49) 的非零期望数量为

$$\sum_{j=1}^{m-1} (-1)^j \binom{m-1}{j} \binom{j+\mathfrak{c}-1}{\mathfrak{c}} = (-1)^{m-1} \binom{\mathfrak{c}-1}{(\mathfrak{c}-m+1)\vee 0} \tag{52}$$

通过使用在线补充材料中的引理 S9。当 $\mathfrak{c} < m-1$，(52) $= (-1)^{m-1}$；而当 $\mathfrak{c} \ge m-1$：

- 如果 $m \ge 4$ 且 $\mathfrak{c} \ge 5$

$$|(52)| = \binom{\mathfrak{c}-1}{m-2} \le (\mathfrak{c}-1)^{m-2} \le m^{\mathfrak{c}};$$

- 如果 $m = 4$ 且 $\mathfrak{c} = 4$,

$$|(52)| = \binom{3}{2} = 3 < m^{\mathfrak{c}};$$

- 最后，如果 $m = 3$,

$$|(52)| = \mathfrak{c} - 1 < m^{\mathfrak{c}}.$$

通过利用引理 S1 对这些形如 (49) 的非零期望的绝对值进行界定，可完成证明，其上界为 $k^{\mathfrak{c}}$ ，从而得到所述结论

$$|(44)| \leq \left(\frac{km}{n}\right)^{\lceil \frac{\mathfrak{c}-1}{2} \rceil \vee 1}.$$

□

注13。 可以改进 $m^{\mathfrak{c}}$ 的上界 $|(52)|$。首先，这里使用的二项式系数上界并不紧。其次，计入 (52) 中的项（即项 $\binom{j+\mathfrak{c}-1}{\mathfrak{c}}$）并非全都非零。出于美观考虑，我们决定不进一步改进 $m^{\mathfrak{c}}$ 。
■

最后，结合上述结果，我们得到定理1给出的所需核估计偏差界。