arXiv:2301.04109v1 [stat.ME] 10 Jan 2023

匹配卡尺与指标估计的精度

Ben B. Hansen*

2023年1月11日

摘要

本文阐明了指标估计的精度如何影响匹配精度。在假设协变量服从高斯分布，并对匹配质量作出最有利假设的模型中，本文精确刻画了真实指标值与估计指标值的配对差值之间的平均差异和最坏情况差异。在这一乐观设定下，如果 $p = o[n/(\log n)]$, ，则真实指标差值和估计指标差值的最坏情况都会趋于零；这与常见指标模型保持一致性所需的模型规模限制相同。即使将高斯假设放宽为次高斯假设，这一结论仍然成立，不过此时对配对指标误差的刻画不那么精确。本文利用高斯假设下推导出的公式作为匹配卡尺的依据。按照这一卡尺进行匹配，即要求估计指标上的配对差值低于该卡尺，其好处是匹配后潜在指标上的最坏情况差异会趋于 0，只要$p = o\{[n/(\log n)]^{2/3}\}$。（对于线性指标模型， $p = o[n/(\log\ n)]$便已足够。）本文提出的卡尺条件改进方案，无需对协变量作次高斯假设，也能带来相同的好处。当强可忽略性成立，且指标是设定正确的倾向得分或预后得分时，以这种方式确保该指标上匹配后的最坏情况差异随着 n 增加而趋于 0，也能确保处理效应匹配估计量的一致性。

*本文受益于 Jake Bowers、Joshua Errickson、Mark Fredrickson、Xuming He、Peter Schochet、Stilian Stoev 和 Lan Wang 提出的意见。对于仍然存在的任何不足，作者承担责任。

关键词与短语：匹配、卡尺、重叠、正值性、倾向得分、预后得分

1 引言

在准备匹配的观察性研究时，治疗倾向的估计会短暂地成为焦点，因为研究者需要选择协变量并确定模型设定。一旦从这些模型中提取出倾向得分估计值，这些模型很快便淡出视野，尽管它们包含着有关这些估计值可能精度的重要信息。对于基于预后评分或主分层评分进行匹配，情况也大同小异：几乎从未评估过匹配指标背后模型的抽样变异。我们似乎理所当然地认为，匹配指标的估计误差不可能大到威胁匹配的可靠性。

倾向得分匹配被视为一种大样本技术，通常用于指标模型估计的逻辑斯蒂回归及其他回归方法也是如此；经典渐近理论似乎也会让人忽视指标估计误差。随着来自任何合理的固定维数 $p+1$分布的处理组/协变量样本$(z_i,\ \vec{x}_i)$不断累积，人们预期指标$\{|\vec{x}_i\hat{}\beta - \vec{x}_i\beta_n|:\ i\}$的估计误差会越来越可以忽略，按 $n^{-1/2}$或近似于 $n^{-1/2}$的速度递减，正如 $|\hat{}\beta - \beta_n|_2 = O(n^{-1/2})$。问题在于，匹配惯例不鼓励倾向得分建模时采用简约模型（Rubin 和 Thomas，1996），而固定-p 大样本理论可能无法准确描述非简约模型。

逻辑回归及类似技术的递增-p 渐近理论已有研究（Portnoy, 1988; He 和 Shao, 2000），只是鲜为人知。鉴于 $p = o[(\log n)/n]$，这些理论给出的是 $|\hat{}\beta - \beta_n|_2 = O_P[(p/n)^{1/2}]$，而非 $O_P(n^{-1/2})$，下文第 2 节将对此进行回顾。这表明，对于形式为 $\vec{X}(\hat{}\beta - \beta_n)$,$\sum_{j=1}^{p} X_j^2 = O(p)$、对应于 $|\vec{X}|_2 = \left(\sum_{j=1}^{p} X_j^2\right)^{1/2} = O(p^{1/2})$ 的误差，其阶可能还要更大，为 $p/n^{1/2}$。这一推论虽在某种程度上简化了问题，但其含义是正确的：例如，如果 p 随 $n^{1/2}$成比例增长，那么即使系数估计误差有所减小，指标估计误差也未必会减小。除固定-p 渐近理论外，指标模型的一致性本身并不能使指标误差在渐近意义下可忽略；这需要更强的假设、专门的匹配技术，或两者兼用。

若要通过更强的假设来控制指标估计误差，请注意，如果假设 p 增长得足够缓慢，那么递增维数回归渐近理论就会类似于维数 p 固定时的情形。遗憾的是，$p \propto n^{1/2}$ 稍大了一些，无法满足这种对应关系，因此在这种情况下，固定-p 的直觉不再适用也就不足为奇了。（例如，β 的渐近正态性要求

$p^2\log(p) = o(n)$，而不是 $p = o(n^{1/2})$ [He 和 Shao，2000]。）不过，假设指标模型具有 次-$\sqrt{n}$ 维数，具体而言， $p = o[(n/\log\ n)^{1/2}]$。本文第 3.1 节表明，对于协变量为次高斯且维数为次-$\sqrt{n}$ 的一致估计指标模型，指标实现值的估计误差趋于零。这种收敛具有强意义，即 l_∞ 意义下的 $\max_{i\le n}|\vec{X}_i(\hat{}\beta - \beta_n)| = o_P(1)$，因此，它为常见做法提供了依据：对估计的倾向得分进行匹配、子分类，或直接修剪，以此作为分析性干预来确保更强类型的重叠假设，即 $c_l \le P(Z = 1|\vec{X}) \le c_u$ with$[c_l, c_u] \subset (0,1)$，并将其应用于修剪后剩余的可用观测值子集。

ˆ 第 3.3 节接着研究普通 Cov(β) 估计量能否适用于刻画指标估计误差 $\vec{x}_i(\beta - \beta_n)$ 可能达到的大小。拟合的指标模型的 ˆ ˆ Fisher 信息给出 Cov(β) 的一个估计 C，既可以直接获得，也可以作为 Eicker-Huber-White 夹心估计量的一部分。在次-$n^{1/2}$ 维数情形下，相应的标准误 s. e.$(\vec{x}\hat{}\beta) = [\vec{x}\hat{}C\vec{x}']^{1/2}$ 将被证明能很好地估计抽样变异性 $\mathrm{Var}^{1/2}[(\vec{x} - \bar{\vec{x}})(\hat{}\beta - \beta_n)]$。当次-$\sqrt{n}$ 条件放宽为次-n ($p = o[n/\log(n)]$) 时，基于信息和夹心估计的估计量会低估 ˆCov(β)。这限制了它们在关于 β_n 的推断中的作用，而它们在次-$\sqrt{n}$ 情形之外的表现也较少受到研究。然而，这并不意味着它们不适合为基于指标的匹配提供信息。我们 ˆ 探讨解析 Cov(β) 估计量在何种条件下仍能刻画 β 线性化的抽样变异性，从而为构建能够捕捉 $\mathrm{Var}[\vec{x}(\beta - \beta_n)]$ 大部分变异性的估计量提供依据。在这两种情形下， $\{\mathrm{Var}^{1/2}[\vec{x}_i(\hat{}\beta - \beta_n)] : i\}$ 的最大值都可能与其余值相距甚远，单是这些极端值就会显著增大 $\mathrm{E}[\max_{i\le n}|\vec{X}_i(\hat{}\beta - \beta_n)|]$。第4节的结果有助于识别表现最差的研究对象，即 s. e.$[\vec{x}_i\beta]$ 较大的研究对象 i 。次高斯条件和次-$\sqrt{n}$ 条件可以放宽其中之一或两者，但此时若要控制指标估计误差，就必须将这类研究对象从样本中剔除。

在这种情况下，匹配提供了其他切实可行的补救方法，能够保留更多样本。首先，它将注意力从特定的 ˆ ˆ$\{\vec{x}_i\beta : i\}$ 或 $\{\vec{x}_i\beta_n : i\}$ 转向指标的配对差值，即 $(\vec{x}_i - \vec{x}_j)\beta$ 或 $(\vec{x}_i - \vec{x}_j)\beta_n$，正如下文第 4 节所做的那样。虽然 $\mathrm{Var}^{1/2}[(\vec{x}_i - \vec{x}_j)\hat{}\beta]$ 的低估无助于对 $(\vec{x}_i - \vec{x}_j)\beta_n$进行推断，但它们确实可以为匹配程序提供参考。ˆ 第 3.2 节使用 $C_{\hat{\beta}}$，将刻意缩小的标准误与配对对比（例如 $(\vec{x}_i - \vec{x}_j)\beta$）关联起来；这些对比经过构造，使其均值——配对指数对比汇总标准误（PIC标准误）——能够以较低成本近似匹配差异在估计得分上的均方根

[1]对于 $|\vec{v}|_2 \ne 0$, $|\vec{v}|_2^{-1}\{\text{s. e.}[\vec{v}(\hat{}\beta - \beta_n)] - \mathrm{Var}^{1/2}[\vec{v}(\hat{}\beta - \beta_n)]\}$ 为 $o_P[(p/n)^{1/2}]$, 而$|\vec{v}|_2^{-1}$s. e.$[\vec{v}(\hat{}\beta - \beta_n)]$ 和 $|\vec{v}|_2^{-1}\mathrm{Var}^{1/2}[\vec{v}(\hat{}\beta - \beta_n)]$ 均为 $O_P[(p/n)^{1/2}]$。

ˆ 均方根；这些匹配差异是在各对之间$\{i, j\}$ 真实得分差异很小或没有差异的情况下，对估计得分 $(\vec{x}_i - \vec{x}_j)\beta$计算得出的， $\vec{x}_i\beta \approx \vec{x}_j\beta$。

ˆ PIC 标准误以与 $|\beta - \beta_n|_2$ 相同的速率趋于零，因此可用作匹配的标尺。对于次高斯数据和次-$\sqrt{n}$ 模型维数，利用它来设定指标上的卡尺宽度——只有当 $|(\vec{x}_i - \vec{x}_j)\beta| \leq c_n$pic se$(\beta)$（例如， $c_n = 2$）时，才允许 i 与 j 配对——会迫使真实指标上的匹配差异在渐近极限下趋于零：

$$\max_{1 \leq i \sim j \leq n} |(\vec{x}_i - \vec{x}_j)\beta_n| \xrightarrow{P} 0 \quad (\text{where } \text{“}i \sim j\text{”} \text{ means } \text{“}i \text{ is matched to } j\text{”}), \qquad (1)$$

这是因为 pic se$(\hat{\beta}) = O_P[(p/n)^{1/2}]$。事实上，在常见指标模型一致性所需的 $p = o(n^{-1} \log\ n)$ 增长条件下，只要 $c_n = O_P[(\log\ n)^{1/2}]$，即使 c_n 为非恒定值，式 (1) 仍成立。与可比的修剪规则相比，这些匹配要求通常不太可能排除研究对象，因为只要对照研究组中存在位置相似的研究对象，即使某些研究对象远离分布中心，也可以将其纳入匹配。如果 i 因缺少在卡尺距离范围内的对应对象 j 而被排除在匹配样本之外，那么它与比较组之间的距离必定超过了指标估计的分辨率。正如数据示例所示，这种分辨率可能相当大，远大于现有的卡尺宽度建议（Rosenbaum 和 Rubin，1985年；Rubin 和 Thomas，2000；Austin，2011年；Wang 等人，2013年）；在许多情况下，这种方法排除匹配样本中研究对象的数量会少得多。

这是值得庆幸的，因为本文将建议对要求 $|(\vec{x}_i - \vec{x}_j)\beta|$ 小于指定卡尺宽度这一条件作非标准的强化，从而排除潜在配对：要么是因为 $|(\vec{x}_i - \vec{x}_j)\beta|$ 过大，要么是因为 ˆ s. e.$[(\vec{x}_i - \vec{x}_j)\beta]$ 过大。换言之， $\vec{x}_i$ 和 $\vec{x}_j$ 在指标本身或某个依赖于指标估计量的马哈拉诺比斯距离上相距过远。按照这一准则仍符合条件的配对 $\mathcal{E} \subseteq \{\{i, j\}: 1 \leq i \neq j \leq n\}$ 满足 $\max_{\{i,j\} \in \mathcal{E}} |(\vec{x}_i - \vec{x}_j)(\hat{\beta} - \beta_n)| = o_P(1)$，即使 $\max_{i \leq n} |\vec{X}_i(\hat{\beta} - \beta_n)|$ 可能不再趋于 0。因此，即使将 $\vec{X}$ 的次高斯假设放宽为四阶矩条件，并将 $p = o\{[n/\log(n)]^{1/2}\}$ 放宽为 $p = o\{[n/\log(n)]^{2/3}\}$，条件 (1) 仍然成立。在倾向得分匹配中，从这样的 $\mathcal{E}$ 中选取配对，可确保能够依据估计得分评估重叠准则，因为 $\max_{i \sim j} |(\vec{x}_i - \vec{x}_j)\beta - (\vec{x}_i - \vec{x}_j)\beta_n| = o_P(1)$。由于这一建议要求可视为对 $|(\vec{x}_i - \vec{x}_j)\beta|$施加的、随 s. e.$[(\vec{x}_i - \vec{x}_j)\beta]$取值变化的限制，我仍将其称为卡钳。在许多情况下，这些 PICSE 卡尺仍比 Rosenbaum 和 Rubin（1985年）的经典 $|(\vec{x}_i - \vec{x}_j)\beta| \leq 0.2$ s. d.$(\mathrm{x}\beta)$ 要求更宽松；而在所有情况下，PICSE 卡尺

只有当我们对某个单位 i 到每个潜在匹配对象 j 之间的指标距离所作的最佳估计超过该指标估计的分辨率时，才会将其排除。

2 背景

2.1 可估计的指标分数

在同时包含处理组和对照条件的观察性研究中，倾向得分是一个函数 $\vec{x} \mapsto g^{-1}[$ P($Z = 1 \mid \vec{X} = \vec{x}$)], , 其中 $g:\ \Re \to [0,1]$ 是连续递增的（Rosenbaum 和 Rubin，1983年）。通常假设 z对$\vec{x}$ 回归服从广义线性模型，例如逻辑斯蒂回归，P($Z = 1 \mid \vec{X} = \vec{x}$) $= [1 + e^{-\vec{x}\beta}]^{-1}$。将 g 作为该模型的连接函数，即可将倾向得分表示为指标 $\vec{x}\beta$。Rosenbaum 和 Rubin（1985年）建议根据 $\vec{x}$^β进行匹配，而非根据 g^{-1}($\vec{x}$^β)。同样，预后评分、混杂因素和风险评分，以及主分层评分，都会拟合（参数）回归模型，以提取^指标 $\vec{x}\beta$ ，用于匹配、子分类或样本修剪；也可在选定分析样本后，用于加权或协方差调整。

令 R 表示指标模型的因变量，例如， Z 表示倾向性得分，或 Y 表示预后得分。令 β 为以下方程的解

$$\left[\sum_{i=1}^{n} \psi(r_i, \vec{x}_i; \beta)\right] + \alpha(\beta) = 0 \tag{2}$$

关于 β，其中 ψ(r, $\vec{x}$; β) 是标量函数 ρ(r, $\vec{x}$; β) 关于 β 的 $\Re^p$ 值梯度，且该函数关于 β 为凸函数。（在最大似然估计中， $\alpha(\cdot) = 0$ ；而贝叶斯估计 [，例如 Gelman 等人， 2008], 以及某些频率学派偏差缩减方法[（Firth，1993；Kosmidis 和 Firth， 2009] ）会最小化惩罚目标函数，在这种情况下， $\alpha(\cdot)$ 是惩罚项的梯度。）如果指数得分（IS）可估计，则

$$\mathrm{E}\left\{[\sum_{i=1}^{n} \psi(R_i, \vec{x}_i; \beta)] + \alpha(\beta)\right\} = 0 \tag{3}$$

存在唯一的根 β_n，且 $\sup_{\gamma:|\gamma-\beta_n|_2 \le 1} |\alpha(\gamma)|_2 = o_P(n^{1/2})$。根据 He 和 Shao (2000)，当方程没有精确解时，(2) 和/或 (3) 的近似根也可以接受，前提是存在唯一的最近根，但假设这些精确根或最近根满足 $\sum_{j=1}^{p}\left[\sum_{i=1}^{n} \psi_j(r_i,\ \vec{x}_i;\ \hat{}\beta)\right]^2 = o_P$ ($n^{1/2}$) and$\sum_{j=1}^{p}[\mathrm{E}\ [\sum_{i=1}^{n} \psi_j(R_i,\ \vec{x}_i;\ \beta_n)]]^2 = o(n^{1/2})$。假设存在函数 $c_0(\cdot,\ \cdot,\ \cdot)$ and$w(\cdot)$，使得

$$\psi(r, \vec{x}, \beta) = c_0(r, \vec{x}, \beta_0 + \vec{x}\beta) w(\vec{x}) (1, \vec{x})'. \tag{4}$$

这种结构既适用于稳健（Cantoni 和 Ronchetti, 2001）和广义（Liang 和 Zeger, 1986）估计方程，也适用于诸如逻辑回归得分函数这类函数，$\psi(r,\ \vec{x},\ \beta) = \{r - [1+\exp(-\beta_0 - \vec{x}\beta)]^{-1}\}(1,\ \vec{x})'$。如果已有层 $1, \ldots, L$，并且要在各层内进行匹配，还可以用子类进一步划分这些层，那么可将 (4) 中的 β_0 替换为各层特定的截距 $\beta_1,\ \ldots,\ \beta_L$。第 2.3 节将回顾 IS 的一致性充分条件。

2.2 次高斯随机变量

实值随机变量 V 若其尾部不比中心正态分布更重，则称为次高斯变量：存在有限常数 s_V，使得对于$t > 0$,

$$\mathrm{P}(V < t) \le \exp[-t^2/(2s_V^2)], \quad \mathrm{P}(V > t) \le \exp[-t^2/(2s_V^2)].$$

在此条件下，s_V 可以取为 $\|V\|_{\psi_2}$ 的常数倍，其中 $\|V\|_{\psi_2}$ 是 V 的次高斯范数。该范数定义为满足以下条件的 $\{t > 0$ 的下确界：$\mathrm{E}[\exp(V^2/t^2)] \le 2\}$；对于次高斯 V，这是一个非空集合；若 V不是次高斯变量，则该范数定义为 ∞。对于向量 $\vec{V}$，$\|\vec{V}\|_{\psi_2} = \sup\{\|\vec{V}\gamma\|_{\psi_2}:\ |\gamma|_2 = 1\}$；若$\|\vec{V}\|_{\psi_2} < \infty$，则称 $\vec{V}$ 为次高斯变量（Vershynin，2018，第 2.5、3.4 节）。由这些定义可直接推出，对于固定向量 γ、$\|V\gamma\|_{\psi_2} = |\gamma|_2\|V\|_{\psi_2}$、$\|\vec{V}\gamma\|_{\psi_2} \le |\gamma|_2\|\vec{V}\|_{\psi_2}$和$\|\gamma\|_{\psi_2} =$（$\log 2$）$^{-1/2}|\gamma|_2$；对于固定矩阵 M，$\|\vec{V}M\|_{\psi_2} \le |M|_2\|\vec{V}\|_{\psi_2}$。

次高斯变量的和仍是次高斯的。Hoeffding 不等式根据各加数的次高斯范数平方和，对独立次高斯变量之和的尾部进行界定。次高斯随机变量序列的最大值增长缓慢：对于某个绝对常数 k_0,

$$\mathrm{E}\max_{1\le i\le n} V_i \le k_0\left(\max_{1\le i\le n}\|V_i\|_{\psi_2}\right)(\log n)^{1/2}, \tag{5}$$

其中 $\{V_i:\ i\}$ 独立或相关。对于 $\{G_i:\ i\}$ 方差不超过 1 的正态分布变量，用 $\sqrt{2}$ 替代 $k_0 \max_{i\le n}\|V_i\|_{\psi_2}$，式（5）仍然成立；如果 $\{G_i:\ i\}$ 是独立的 N(0,1)，那么这个界是紧的，其意义是 $\lim_{n\uparrow\infty}(\log\ n)^{-1/2}\,\mathrm{E}\max_{1\le i\le n} G_i = \sqrt{2}$（Boucheron 等，2013，第 2.5 节）。次高斯向量序列的最大值增长也很缓慢：存在绝对常数 k_1 和 k_2，使得对于任意均值为 0、协方差为 I的次高斯 $\{\vec{V}_i\}$，以及列维数与 $\vec{V}$的范围相匹配的任意确定性矩阵 $\{M_i\}$，

$$\mathrm{E}\left(\max_{1\le i\le n}|M_i\vec{V}_i'|_2\right) \le \max_{i\le n}\|\vec{V}_i\|_{\psi_2}\left\{k_1\max_{i\le n}|M_i|_F + k_2\left(\max_{i\le n}|M_i|_2\right)(\log n)^{1/2}\right\} \tag{6}$$

其中 $|M|_F = \text{tr}(M'M)^{1/2}$（Vershynin，2018，例6.3.5）。由此可知，若对每个 n $\{\vec{X}_{ni}: 1 \le i \le n\}$ 都是长度为 p_n 的独立随机向量，满足

$$O_P\Big[\frac{|\vec{X}_{ni}|_2}{p_n} \quad \frac{\|\vec{X}_{ni}\|_{\psi_2}}{n^{1/2}}\Big] \qquad \max_{1\le i\le n}|\vec{X}_i|_2 = \text{Cov}(\,) \text{ 且}$$

一致有界，则 max max(,log) 。次高斯协变量的这一性质（第 c_0 节中的 A）将在第 3 节中假设成立，但之后会在第 4 节中放宽。另一方面，上文式（4）中的标量$c_0(R, \vec{x}, \vec{x}\beta)$ 将始终被要求满足次高斯条件，这通过第 2.3 节的 A5 来保证。

2.3 一致可估计的指数得分

将数据和模型参数排成三角阵列，其中样本和模型 n 具有 n 个观测值，涉及 p 个自变量 x_i，p（严格来说，p_n）随 n递增。对于（β_n ： n），β 的一致性是指 $p = o$（n）且$\hat{}\beta - \beta_n|_2^2 = O_P$（$p/n$）。文献中已有关于此类一致性的条件。我们将这些条件与刻画$\hat{}\beta$和其最近线性近似之间关系的相应条件一并列出。

对于可估计的 β_n，定义 $A_n = A_n$（β_n）和 $A_n = A_n$（β），其中

$$A_n(\gamma) = \frac{1}{n}\left\{\sum_{i=1}^{n} \nabla_\beta \,\mathrm{E}\,[\psi(R_i, \vec{x}_i; \beta)]|_{\beta=\gamma} + \alpha(\gamma)\right\}, \tag{7}$$

如果某些 β 值处不存在通常的偏导数，则“∇_β”按弱微分理解。对于可逆的 A_n，估计量$\hat{}\beta$ 的线性化为

$$\tilde{\beta}_n = A_n^{-1}\frac{1}{n}\{[\sum_{i=1}^{n}\psi(r_i, \vec{x}_i; \beta_n)] + \alpha(\beta_n)\}. \tag{8}$$

随机向量 $n^{-1}\sum_{i=1}^{n}\psi$（$R_i, \vec{x}_i$； β_n）的协方差为 $n^{-1}B_n$，其中

$$B_n = B_n(\beta_n) \text{ and } B_n(\gamma) = n^{-1}\,\mathrm{E}\sum_{i=1}^{n}\psi(R_i, \vec{x}_i; \gamma)\psi(R_i, \vec{x}_i; \gamma)'. \tag{9}$$

Pr命题 1 (He 和 Shao, 2000)。 在下文所述的 *A2*、*A4*、*A6* 和 *A9* 条件下：

1. $|\hat{\beta} - \beta_n|_2 \xrightarrow{P} 0$, *with rate* $O_P[(p/n)^{1/2}]$;
2. $|\hat{\beta} - \tilde{\beta}_n|_2 \xrightarrow{P} 0$, *with rate* $O_P[(p/n)(\log n)^{1/2}]$.

命题1重述了 He 和 Shao 的定理2.1和2.2，并将其应用于仅系数参数 β_n 随 n 增大而维数增长的模型，同时略微加强了速率条件。（他们假设 $p$log(p) $= o(n)$，而 A9

则指出 $p\log(n) = o(n)$。）他们的定理 2.2 仅刻画了次-$\sqrt{n}$ 维模型中线性化误差的衰减情况，但直接改写其证明即可得到该命题的第 2 部分。另请参见 He 和 Shao 的例 3。

我们的正则性假设如下。

A1. $\mathbf{x}$ 的各列均已中心化。可能存在预先划分的层 $1, \ldots, L$，其中 $L/n \to 0$。在这种情况下，$\mathbf{x}$ 的各列也按层中心化：对于分层变量 v, $\sum_{i:\ v_i=\ell} \vec{x}_i = 0$, $\ell = 1, \ldots, L$。

A2. IS 是可估计的（定义见 § 2.1），且关于 $\mathbf{x}$ 是线性的。

A3. A_n 可逆。此外，存在 $\delta > 0$，使得当 $|\gamma - \beta_n|_2 < \delta$ 时，$A_n(\gamma)$ 可逆，且 $\sup_{\gamma:|\gamma-\beta_n|_2<\delta} |A_n(\gamma)^{-1}|_2$ 有界。

A4. $\psi(r, \mathbf{x}\ \beta)$ 的形式为 (4)。函数 $c_0(r, \vec{x}, \cdot)$（即 $\eta \mapsto c_0(r, \vec{x}, \eta)$）是利普希茨连续的，且具有共同的利普希茨常数；$(\partial/\partial\eta)$ E $c_0(R, \vec{x}, \eta)$ 也如此。

A5. 随机变量 $c_0(R_i,\ \vec{x}_i,\ \vec{x}_i\beta_n)$ 的次高斯范数有界。

A6. 对于 $\ell = 0$、2 和 4，

$$\max_{\substack{\gamma,\delta:|\gamma|_2= \\ |\delta|_2=1}} \sum_{i=1}^{n} w(\vec{x}_i)^\ell (\vec{x}_i\gamma)^2 (\vec{x}_i\delta)^2 = O(n).$$

A7. $p^{-1}|\beta_n|_2^2 = p^{-1}\sum_k \beta_{nk}^2$ 有界。

A8. $s^2(\mathbf{x}\beta_n) = \beta_n' S^{(x)} \beta_n$ 收敛到 $(0,\ \infty]$ 中的一个极限。

A9. $p = o[n/(\log\ n)]$, 即 $(p\log\ n)/n \to 0$（次-n 模型维数）。

虽然假设每个 n 的连接函数 g 相同，但系数向量β 的长度不断增加，而 β_n 不必收敛。事实上，根据 A8 $s(\mathbf{s}\beta_n)$，其可以发散（但不能趋于 0）。

在这里，假设 A_n 可逆（如 A3 所要求的）比在其他回归情境中负担更小，因为对于当前应用，可以自由改变设计矩阵的基底；我们并不关心 β 的特定元素或对比。根据柯西－施瓦茨不等式，A6 意味着 $\max_{\gamma:|\gamma|_2=1} n^{-1}\sum_{i=1}^n w(\vec{x}_i)^\ell(\vec{x}_i\gamma)^2 = O(1)$ 对 $\ell \in \{0, 1, 2\}$ 成立，进而得到 $|S^{(x)}|_2 = O(1)$。这些假设是否恰当，可在进行 IS 估计之前评估；而 A3 则要求在估计 β 后检查模型拟合产物。改善 A3 拟合效果（以及下文的 A12 ）的一个简单方法，是剔除对指标模型拟合贡献相对较小的解释变量，常见的模型选择准则可用于识别这类变量。

在满足一个或多个次要条件时，某些结果的形式会更强。

A10. $p^2 = o[n/(\log n)]$ (次高斯模型维数)。A11. $\max_{i\le n} |\vec{x}_i|_2^2 = O[\max(p,\ \log\ n)$], 以及 $\max_{i\le n} w(\vec{x}_i)^2 = O(\log\ n)$（次高斯协变量）。A12. 每个 $S^{(x)} = (n - L)^{-1}\mathbf{x}'\mathbf{x}$ 和 B_n 均为满秩，且 $|(S^{(x)})^{-1}|_2$ 和$|B_n^{-1}|_2$ 有界（满秩协方差）。

如果 $\vec{x}$es（以及 $w(\vec{x})$）是次高斯的，即它们是 $\vec{X}$的实现值，且 $\mathrm{E}\,|\vec{X}|_{\psi_2}$一致有界，同时 A6 也成立，即 $|\mathrm{Cov}(\vec{X})|_2$（因此 $p^{-1}\,\mathrm{E}\,|\vec{X}|_2^2$）也一致有界，那么如 § 2.2 中 (6) 之后所述，A11 可由 (6) 推出。根据下文的命题 4，Cov(β) 的 sandwich ˆ 估计量通常要求协变量为次高斯，且 A10 要求次-$\sqrt{n}$ˆ ˆ 模型维数；而基于 A_n 、但不基于 B_n 的协方差估计量，通常只要求次-n 模型维数和 A9。就本文而言，即便它不具备费舍尔一致性，与 sandwich 估计量相比仍同样有益。根据命题 6，满秩协方差（A12）使指标抽样变异 Var($\vec{x}$ˆβ) 和 Var $\left[(\vec{x}_i - \vec{x}_j)\,\hat{}\,\beta\right]$ 可估计，无需考虑 $\vec{x}$ 或 $\vec{x}_i - \vec{x}_j$的大小。不过，我们的渐近精确匹配方法并不要求这一点，即使 S 或 B_n 的秩不足，也仍然有效。

3 使用高斯-$\vec{}$或次高斯 X进行渐近精确匹配

对于每个 n，令 $\mathcal{S}_n$为 $\{1,\dots,n\}$的一个随机划分。我们不假定$\mathcal{S}_n$会细化或扩展任何先前的划分 $\mathcal{S}_1$、.....、 $\mathcal{S}_{n-1}$。用 $[i]_{\mathcal{S}_n}$表示包含 $i \le n$的唯一 $\mathcal{S}_n$ 元素；若存在$\mathbf{s} \in \mathcal{S}_n$ 使得 $i,\ j \in \mathbf{s}$，则记为 $i \overset{\mathcal{S}_n}{\sim} j$。如果所指的是哪个 n 或哪个划分序列并无歧义，则分别将这些符号写作“$i\sim j$”或“$[i]$”。若沿潜在指标方向的 $\mathcal{S}_n$层宽度 $\max\{|(\vec{x}_i - \vec{x}_j)\beta_n| :\ i \overset{\mathcal{S}_n}{\sim} j\}$依概率趋于 0，则序列 $\{\mathcal{S}_n\}$ 构成渐近精确的指标后分层。本节针对沿估计指标方向的 $\mathcal{S}_n$层宽度 $\max\{|(\vec{x}_i - \vec{x}_j)\,\hat{}\,\beta| :\ i \overset{\mathcal{S}_n}{\sim} j\}$提出一个容差；对于协变量为次高斯的特殊情形（A11），该容差足够窄，能够确保渐近精确性。此外，在第 3.2 节将要介绍的另一种特殊情形中，该容差也足够宽，使得任何值得被置于与对比组具有代表性的后分层中的对象，都不会仅仅因为相对于 $\{\vec{x}_j\beta : z_j \ne z_i\}$而言 $\vec{x}_i\beta$ ˆ 孤立而被排除在这种安排之外。

当 β_n 需要估计时，差值 $(\vec{x}_i - \vec{x}_j)\beta_n$, $1 \le i,\ j \le n$的可观测对应量，即形式为 $(\vec{x}_i - \vec{x}_j)\beta$的对比，称为配对指标对比（PICs）。PIC 与其所估计的配对对比 $(\vec{x}_i - \vec{x}_j)(\beta - \beta_n)$ 之间的差异称为*PIC*误差。要确保后分层渐近精确，必须分别关注 PIC误差与 PICs 本身。我们的层宽度容差将涉及一种新的 PIC误差大小估计，即*PIC*标准误。

3.1 次高斯情形下的 PIC 误差

回顾一下，指数系数估计的误差 $\beta - \beta_n$ 可分解为 $(\beta - \beta) + (\beta - \beta_n)$，其中 β 是式（8）中定义的线性化。指标误差和 PIC 误差也有类似的分解。

推论 (命题 1 的第 2 部分)。 若满足 *A2*、*A4*、*A6*、*A9* 和 *A11*，则 $\max_i |\vec{x}_i(\beta - \tilde{\beta}_n)| = O_P[p^{3/2}(\log n)^{1/2}/n]$。（除非 $p = o(\log n)$，此时 $\max_i |\vec{x}_i(\beta - \beta_n)| = O_P[p(\log n)/n]$）。

命题 2。 若满足 *A1*、*A2*、*A5* 和 *A6*，则 $\|\tilde{\beta}_n - \beta_n\|_{\psi_2} = O(n^{-1/2})$。

推论（命题 2 的推论）。 若满足 *A1*、*A2*、*A5*、*A6* 和 *A11*，则 $\mathrm{E}\max_{i \le n} |\vec{x}_i(\beta_n - \beta_n)| = O\left\{[(p \log n)/n]^{1/2}\right\}$.（除非 $p = o(\log n)$，此时 $\mathrm{E}\max_{i \le n} |\vec{x}_i(\tilde{\beta}_n - \beta_n)| = O\{[(\log n)/n^{1/2}]\}$。）

命题1的推论可由其第2部分结合 A11 直接得出。命题 2 在附录中得到证明；其推论由 A11 和（5）得出。它们沿用 D'Amour等（2021）的做法，假设协变量是次高斯的（按 A11 的定义），并将在下文第4节放宽这一假设。

这些推论刻画的是指标值的估计误差，而非其配对对比的估计误差，但可直接推广，分别得出 $\max_{i,j \le n} |(\vec{x}_i - \vec{x}_j)(\beta - \beta_n)|$ 和 $\mathrm{E}\max_{i,j \le n} |(\vec{x}_i - \vec{x}_j)(\beta_n - \beta_n)|$具有相同的衰减速率。因为对于任意一组长度为-$p$ 的 $\mathcal{W}$ 行向量，

$$\sup_{\vec{w} \in \mathcal{W}} |\vec{w}(\hat{\beta} - \beta_n)| \le \sup_{\vec{w} \in \mathcal{W}} |\vec{w}(\hat{\beta} - \tilde{\beta})| + \sup_{\vec{w} \in \mathcal{W}} |\vec{w}(\tilde{\beta} - \beta_n)|, \tag{10}$$

由此可知，对于次高斯协变量，只要 $(p^{3/2}/n)\log n$ 趋于 0，即若 $p = o\{[n/(\log n)]^{2/3}\}$，最坏情况下的 PIC 或指标误差就会趋于 0。

当协变量的次$\sqrt{n}$维数较低时，推论表明，在大样本中，$\{|\vec{x}_i(\beta - \beta)| : i \le n\}$ 和 $\{|(\vec{x}_i - \vec{x}_j)(\beta - \beta)| : i,\ j \le n\}$ 的上确界将比 $\{|\vec{x}_i(\tilde{\beta} - \beta_n)| : i\}$和 $\{|(\vec{x}_i - \vec{x}_j)(\tilde{\beta} - \beta_n)| : i,\ j\}$（二者均为 $O_P\left[((p/n)^{1/2}\log n)\right]$）的上确界小一个数量级，即 $p/n^{1/2} = o_P(1)$。式（10）右侧的两个误差中，$\sup |\vec{w}(\beta - \beta_n)|$ 项通常占主导地位；我们将注意力转向这一项。

3.2 一个思想实验

在一种特殊情形下，若 X 和 β 均服从高斯分布，则 PIC误差的大小可以用易于估计的量进行具体刻画。Ghosh 和 Cort´es（2019），以及其他研究者，在高斯 X的假设下探讨了相关问题。对于向量 $v \in \Re^m$，令 $|v|_2$ 和 $|v|_\infty$ 分别具有通常的含义，即 $(m^{-1}\sum_{i=1}^m v_i^2)^{1/2}$ 和 $\sum_{i\le m}|v_i|$。对于维数相同的矩阵 M 和 N，$\langle M,N\rangle_F$ 表示弗罗贝尼乌斯内积 $\mathrm{tr}(M'N)$。

命题 *3.* 设 $\mathcal{S}$ 是 $\{1,\ldots,n\}$ 的一个分割，并配有映射 $\vec{\mu}:\mathcal{S}\to\Re^p$，且令 $E_{\mathcal{S}}(\cdot)$ 和 $Cov_{\mathcal{S}}(\cdot)$ 表示在固定 $\mathcal{S}$ 和 $\{\mu_s: s\in\mathcal{S}\}$ 时计算的期望。给定 $\mathcal{S}$ 和 $\{\mu_s: s\in\mathcal{S}\}$，令 $(\tilde{\beta}_n-\beta_n;\ \vec{X}_i-\vec{\mu}([i]_{\mathcal{S}});\ \vec{X}_j-\vec{\mu}([j]_{\mathcal{S}}))$ 对每个 i、$j\le n$ 均服从联合多元正态分布，均值为零，$Cov_{\mathcal{S}}(\beta)=C$，$Cov_{\mathcal{S}}(\vec{X}_i)=Cov_{\mathcal{S}}(\vec{X}_j)=\Sigma$，且若 $i\ne j$，则 $Cov_{\mathcal{S}}(\vec{X}_i,\vec{X}_j)=0$。于是有

$$\mathrm{E}_{\mathcal{S}}\left[\left|[(\vec{X}_i-\vec{X}_j)(\tilde{\beta}-\beta_n): i\overset{\mathcal{S}}{\sim}j, i<j]\right|_2^2\right]=\langle 2\Sigma,C\rangle_F \quad and \tag{11}$$

$$\mathrm{E}_{\mathcal{S}}\left[\left|[(\vec{X}_i-\vec{X}_j)(\tilde{\beta}-\beta_n): i\overset{\mathcal{S}}{\sim}j]\right|_\infty\right]\le z^*_{n_{\mathcal{S}}}\langle 2\Sigma,C\rangle_F^{1/2}, \tag{12}$$

其中 $n_{\mathcal{S}}:=\mathrm{card}(\{\{i,j\}: i\overset{\mathcal{S}}{\sim}j, i<j\})$ 且 $z^*_{n_{\mathcal{S}}}:=(2\log 2n_{\mathcal{S}})^{1/2}$ 给$\mathrm{E}\max_{1\le i\le n}|G_i|$($G_i$：$i$) 独立N(0,1)设定了界，如第*2.2*节所述。

鉴于命题 3 对协变量的假设很强，它在 PIC 误差控制方面的实际用途有限；我们将提出用于限制$[(\vec{X}_i-\vec{X}_j)(\tilde{\beta}-\beta_n): i\overset{\mathcal{S}}{\sim}j]$ 的方法，将这些假设放宽为矩条件A6。但这些方法要求对可容忍的 PIC误差大小设定上限，而（12）在这方面将会提供有用的具体指引。

着手选择匹配样本的统计学家，其操作性假设是：焦点组中的每个成员，在 $\vec{x}\beta_n$方面，都能在替代组中找到足够接近的对应成员。为简化问题，暂且搁置“足够接近”这一循环论证式的说法，先假设对于每个焦点组成员 i，现有样本中至少包含一名对照组成员 j，其潜在指标上与之完美匹配，记为 $\vec{x}_i\beta_n=\vec{x}_j\beta_n$。继续这一思想实验，假设 β 和 X 如命题 3 所述，并将较小的处理组或对照组作为焦点组，这意味着至少有 $\min(n_0,\ n_1)$ 个完美配对。令 $\mathcal{S}$ 为由如下关系所诱导的等价类集合：当且仅当 $\vec{x}_i\beta_n=\vec{x}_j\beta_n$时，$i\overset{\mathcal{S}}{\sim}j$。于是，该命题刻画了使潜在指标上的差异 $|(\vec{x}_i-\vec{x}_j)\beta_n|$为 0 的 PICs $(\vec{x}_i-\vec{x}_j)\beta|$。迄今为止，为了应用该命题所作的每一项简化，都应朝着低估紧密匹配配对中 PIC 最大值的方向

$\vec{x}\beta_n$；但即使我们继续调整思想实验，尽量压低这个量，也会发现它大得几乎无法实际运用，具体含义下文将述。

$\vec{}$如目前所规定的那样，我们的完美配对思想实验不允许 $\mathcal{S}$的层内存在任何 $X\beta_n$变异。这意味着，分层协方差 $\mathrm{Cov}_{\mathcal{S}}(\vec{X})$ 必须满足 $\beta_n'\Sigma\beta_n = 0$。根据这一约束调整可用的协变量数据，令 $\vec{X}$按照命题 3 中的分布，其中 $\Sigma = S^{\perp\beta_n}$：$=(n-L)^{-1}\mathbf{x}^{\perp\mathbf{x}\beta_n\prime}\mathbf{x}^{\perp\mathbf{x}\beta_n}$，即观测协变量的协方差在指标的正交补上的投影，$\mathbf{x}^{\perp\mathbf{x}\beta_n}$。（这里，x 是观测到的协变量 $n\times p$ 矩阵；$\mathbf{x}^{\perp v}$ 表示由将 x 的各列对 n-v向量进行 p 回归所产生的 $n\times p$ 残差矩阵； L 是显式存在的预先分层的数量（如果存在此类分层），否则为 1。遵循 A1，假定 x 已居中，或在适当情况下按层居中，并且 $\hat{}\,\hat{}\,\hat{}\,\mathbf{x}^{\perp\mathbf{x}\beta_n}$ 继承这一居中处理。）自然估计量 $S^{\perp\beta} = (n-L)^{-1}\mathbf{x}^{\perp\mathbf{x}\beta\prime}\mathbf{x}^{\perp\mathbf{x}\beta}$适当地一致估计 $S^{\perp\beta_n}$，正如下面的命题 5 所述。$\hat{}$ 注意，使用 $S^{\perp\beta}$ （而不是 S）再次简化了（11）和（12）；如果在递增p 的情形下它保持二者的阶数不变，情况亦然。

当焦点组和比较组中较小一组的每个成员都恰好有一个完美匹配的对应成员时，最大 PIC 界 (12) 达到最小值，即 $n_{\mathcal{S}} = \min(n_0, n_1)$。在这类配置中，当各配对互不重叠时（如无放回匹配），该界最为紧致。（在给定 $\tilde{\beta}$ 的条件下，同时给定 $\mathcal{S}$和 $\{\vec{\mu}(\mathbf{s})\colon \mathbf{s}\in\mathcal{S}\}$，$(\vec{X}_i - \vec{X}_j)(\tilde{\beta}-\beta_n)$ 独立于 $(\vec{X}_{i'} - \vec{X}_{j'})(\tilde{\beta}-\beta_n)$，其中 i,j,i',j' 且 $\{i,j\}\neq\{i',j'\}$；根据第 2.2 节中 (5) 后的讨论，这使得附录 B.2 中的不等式 (31) 达到紧致。）假设完美配对思想实验中的 $n_{\mathcal{S}}$ =min(n_0, n_1) 个配对互不重叠，从而补全该实验的设定。这样，(12) 对 $\mathcal{S}$ 中 $\vec{x}\beta$ 的宽度估计就会更加准确。

在构成该思想实验的各对 $\mathcal{S}$ 中，最大 PIC误差预计为 $z^*_{\min(n_0,n_1)}\langle 2S^{\perp\beta_n}, C\rangle_F^{1/2}$，其阶为 $O\{[(p\log n)/n]^{1/2}\}$。相应的估计值为 $z^*_{\min(n_0,n_1)}\langle 2S^{\perp\hat{\beta}}, \hat{} C\rangle_F^{1/2}$。这些量足够小，会趋于零，但仅仅勉强如此：$[(p\log n)/n]^{1/2}$ 恰好是 A9 要求其降至 0 的速率。对于实际中完全匹配的各对 $\{i,j\}$，$(\vec{x}_i - \vec{x}_j)\beta_n = 0$，即使其他协变量方向上存在温和的高斯变异，也会产生最大 PIC $||(\vec{X}_i - \vec{X}_i)\hat{}\beta : i \overset{\mathcal{S}}{\sim} j]||_\infty$ 其大小可达到实际指标上可容忍的分离程度的最大阶，$||(\vec{X}_i - \vec{X}_i)\beta_n : i \overset{\mathcal{S}}{\sim} j]||_\infty$，前提是匹配要渐近精确。因此，我们将建议把这个数值作为匹配容差；不仅在采用命题 3 的限制性假设时如此，在仅接受较弱的 A1–A9 时也如此。

3.3 PICs 的渐近标准误

˜根据 (8) 和 (9)， β_n 的协方差为

$$C_n = n^{-1} A_n^{-1} B_n (A_n^{-1})'. \tag{13}$$

ˆ 该 C_n 近似于 β 的协方差，尤其是在指标模型的子-$\sqrt{n}$ 维数较低时。当 $p = o\{[n/\log(n)]^{1/2}\}$ 时，命题 1 表明 $|\tilde{\beta} - \hat{}\beta|_2 = o_P(n^{-1/2})$，其值足够小，因而无需区分 $C_n =$ Cov($\tilde{\beta}$) 与 Cov(ˆβ)。例如，下面的引理 1 推出 s. d.($\vec{x}\tilde{\beta}$) $= [\vec{x} C_n \vec{x}']^{1/2}$ 与 $\vec{x}$(ˆ$\beta - \beta_n$) 具有相同的阶 $O_P(n^{-1/2}|\vec{x}|_2)$；而命题 1 的第 2 部分表明 $\vec{x}(\tilde{\beta} - \hat{}\beta) = o_P(n^{-1/2}|\vec{x}|_2)$。引理后的命题将表明，更大的阶 $O_P(n^{-1/2}|\vec{x}|_2)$ 也同样适用于 s. e.($\vec{x}\tilde{\beta}$) $= [\vec{x}\hat{}C_n\vec{x}']^{1/2}$。

引理 1。 在 *A3*、*A5* 和 *A6* 条件下， $|B_n|_2 = O(1)$ 和 $|C_n|_2 = O(n^{-1})$。

如果 ψ 是（设定正确的）对数似然的梯度，那么 $A_n = B_n$，且 $n^{-1}\hat{}A_n^{-1}$ 估计 C_n。更一般地， C_n 由 $n^{-1}\hat{}A_n^{-1}\hat{}B_n(\hat{}A_n^{-1})'$ 估计，其中

$$\hat{B}_n = \hat{B}_n(\hat{\beta}), \quad \hat{B}_n(\beta) = n^{-1}\sum_{i=1}^{n} \psi(r_i, \vec{x}_i; \beta)\psi(r_i, \vec{x}_i; \beta)'.$$

接下来的命题确立了自然协方差估计量及相关量的一致性。

命题 4。 在 *A1–A9* 条件下， $|\hat{}A_n^{-1} - A_n^{-1}|_2 \xrightarrow{P} 0$。如果还满足 *A10*（次-$\sqrt{n}$ 维）和 *A11*（次高斯协变量），那么 $|\hat{}A_n^{-1}\hat{}B_n(\hat{}A_n^{-1})' - nC_n|_2 \xrightarrow{P} 0$。

ˆ ˆ**命题 5.** 设 *A1–A9* 成立， C_n 是 C_n *(* $|C_n - \hat{}C_n|_2 = o_P(n^{-1})$*)* 的一致估计，且 S、 $S^{\perp\beta_n}$ 和 $S^{\perp\beta}$ 如第 *3.2* 节所定义。那么

$$|\langle S, \hat{C}\rangle_F - \langle S, C\rangle_F| = o_P(p/n) \quad and \quad |\langle S^{\perp\hat{\beta}}, \hat{C}\rangle_F - \langle S^{\perp\beta_n}, C\rangle_F| = o_P(p/n),$$

而 $\langle S, C\rangle_F = O_P(p/n)$ 和 $\langle S^{\perp\beta_n}, C\rangle_F = O_P(p/n)$。

量 $\langle 2S^{\perp\hat{}\beta}, \hat{}C_{\hat{\beta}}\rangle_F^{1/2}$ 将称为*PIC* 标准误 （PIC SE）；命题 5 表明，它能一致地估计命题 3 中公式 (11) 和 (12) 右侧出现的类似参数。命题 4 在递增-p 情形下的应用是新的；命题 5 则完全是新的。其证明见附录 B.3，其中还给出了包括引理 1 在内的中间结果的证明。

3.4 PIC标准误卡尺

我们建议在 $z^*_{\min(n_0,n_1)}$ 倍 PIC 标准误 $\langle S^{\perp\hat{\beta}}, \hat{C}\rangle_F^{1/2}$ 的范围内进行匹配，无论命题 3 的高斯模型是否适用。如果适用，这就确保潜在指标上的匹配差异也由同一倍数的 PIC 标准误来刻画（第 3.2 节和第 3.3 节）。如果不适用，但协变量是次高斯的，那么 (12) 可能不再限制 PIC 误差的大小，但只要 $p = o\{[n/(\log n)]^{2/3}\}$ （第 3.1 节），这些误差仍会趋于 0。如果高斯或次高斯建模假设都不适用，就需要满足下文将要说明的其他匹配要求，才能使 PIC 误差趋于 0；但$z^*_{\min(n_0,n_1)}$ 倍 PIC 标准误仍是 PIC 的适当容差。它趋于 0，因此将其用作卡尺宽度会使 PIC 趋于 0；它以与命题 1 要求指数模型一致性所需相同的 $[(p\log\ n)/n]^{1/2}$ 速率趋于 0，因此不会比使 PIC 最大值趋于 0 所必需的条件更严格。在第 3.2 节研究的理想化情形中，它在更具体的定量意义上也被证明是最小的，因为它对假想实验最大 PIC 误差的刻画是精确的：如果实际数据中潜藏着这样一个配对实验，那么在 $\mathbf{x}\beta$ 上设定任何小于 $z^*_{\min(n_0,n_1)}$ 倍 PIC 标准误的匹配容差，都会排除实际上在$\mathbf{x}\beta_n$上完全匹配的配对。

4 解构高斯与次高斯假设

4.1 不受限制的 X条件下的PIC标准误

PIC标准误通过以下两种方式之一对期望PIC误差取平均。首先，如果可用于匹配的样本中包含一组完全匹配的受试者（如第 3.2 节所设想），且其协变量的配对内协方差为 $S^{\perp\beta_n}$，那么PIC标准误平方可估计完全匹配配对中PIC误差平方的期望均值，即 $[(\vec{x}_i - \vec{x}_j)(\hat{\ }\beta - \beta_n)]^2,,\ j$。其次，若直接使用整个样本，但对每位研究对象的协变量关于$\mathbf{x}\beta_n$进行残差化（第 3.2 节也讨论过这一做法），则PIC标准误平方近似等于约化PIC误差平方的期望均值，即 $[(\vec{x}_i^{\perp\mathbf{x}\beta_n} - \vec{x}_j^{\perp\mathbf{x}\beta_n})(\hat{\ }\beta - \beta_n)]^2,,$ 此时涵盖所有配对 $\{i,j\}, 1 \le i < j \le n$。

为说明这一点，对于 $1 \le i,\ j \le n$ 写作 $\vec{d}_{ij}^{\perp\beta_n} := \vec{x}_i^{\perp\mathbf{x}\beta_n} - \vec{x}_j^{\perp\mathbf{x}\beta_n}$，注意到 $\vec{d}_{ij}^{\perp\beta_n} =$

$\vec{x}_i - \vec{x}_j$ 对于指标完全匹配的配对 $\{i, j\}$。注意，

$$[\vec{d}_{ij}^{\perp\beta_n}(\tilde{\beta} - \beta_n)]^2 = [\vec{d}_{ij}^{\perp\beta_n}(\tilde{\beta} - \beta_n)]'[\vec{d}_{ij}^{\perp\beta_n}(\tilde{\beta} - \beta_n)]$$
$$= (\tilde{\beta} - \beta_n)'(\vec{d}_{ij}^{\perp\beta_n\prime}\vec{d}_{ij}^{\perp\beta_n})(\tilde{\beta} - \beta_n).$$

对第一种情形，将所有完全匹配的配对相加；对第二种情形，则将所有 $\binom{n}{2}$ 可能的配对相加。在这两种情形下，令 n_p 表示对总和有贡献的配对数，则有

$$\frac{1}{n_p}\sum[\vec{d}_{ij}^{\perp\beta_n}(\tilde{\beta} - \beta_n)]^2 = (\tilde{\beta} - \beta_n)'\Big(n_m^{-1}\sum \vec{d}_{ij}^{\perp\beta_n\prime}\vec{d}_{ij}^{\perp\beta_n}\Big)(\tilde{\beta} - \beta_n)$$
$$= (\tilde{\beta} - \beta_n)'(2S^{\perp\beta_n})(\tilde{\beta} - \beta_n) \tag{14}$$
$$= \langle 2S^{\perp\beta_n}, (\tilde{\beta} - \beta_n)(\tilde{\beta} - \beta_n)'\rangle_F, \tag{15}$$

其中，(14) 用到了协方差的 U 统计量表示，$(n-1)^{-1}\sum_{i=1}(w_i - \bar{w})(v_i - \bar{v})$ $= \binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{1}{2}(w_i - w_j)(v_i - v_j)$；(15) 则利用矩阵 M 和 N的逐元素乘积之和 $\langle M, N\rangle_F$，对 (14) 进行重新表示。因此

$$\mathrm{E}\left\{\frac{1}{n_p}\sum[\vec{d}_{ij}^{\perp\beta_n}(\tilde{\beta} - \beta_n)]^2\right\} = \langle 2S^{\perp\beta_n}, C_n\rangle_F.$$

类似的论证表明，PIC标准误还可以解释为成对距离 $|\vec{d}_{ij}^{\perp\beta}\hat{}\,C^{1/2}|_2 = (\vec{d}_{ij}^{\perp\beta}\hat{}\,C\vec{d}_{ij}^{\perp\beta\prime})^{1/2}$在配对 $\{i, j\}$上的均方根。依次利用矩阵迹的循环性质、弗罗贝尼乌斯矩阵乘积 $\langle\cdot,\cdot\rangle_F$的定义、弗罗贝尼乌斯乘积的双线性，以及样本协方差的 U 统计量表示：

$$\sum \vec{d}_{ij}^{\perp\hat{\beta}}\hat{C}\vec{d}_{ij}^{\perp\hat{\beta}\prime} = \sum \mathrm{tr}(\vec{d}_{ij}^{\perp\hat{\beta}}\hat{C}\vec{d}_{ij}^{\perp\hat{\beta}\prime}) = \sum \mathrm{tr}(\vec{d}_{ij}^{\perp\hat{\beta}\prime}\vec{d}_{ij}^{\perp\hat{\beta}}\hat{C})$$
$$= \sum\langle\vec{d}_{ij}^{\perp\hat{\beta}\prime}\vec{d}_{ij}^{\perp\hat{\beta}}, \hat{C}\rangle_F = \langle\sum \vec{d}_{ij}^{\perp\hat{\beta}\prime}\vec{d}_{ij}^{\perp\hat{\beta}}, \hat{C}\rangle_F = n_p\langle 2S^{\perp\hat{\beta}}, \hat{C}\rangle_F.$$

当均值仅针对 $(\vec{d}_{ij}^{\perp\beta}\hat{}\,C\vec{d}_{ij}^{\perp\beta\prime})^{1/2}$ 中与指标完全匹配的 $\{i, j\}$ 计算时，它近似于 $\mathrm{s.d.}[(\vec{x}_i - \vec{x}_j)\hat{}\,\beta]$，因为 $\vec{x}_i\beta_n = \vec{x}_j\beta_n$意味着 $\vec{d}_{ij}^{\perp\beta}$ 近似于 $\vec{x}_i - \vec{x}_j$。将其解释为$\mathbf{x}$ 中$\mathbf{x}\hat{}\,\beta$ 的正交补内的成对协变量距离，并经$\hat{}\,C^{1/2}$重新缩放，这种解释既适用于完美配对思想实验，也适用于对所有 $\{i, j\}$、$1 \le i < j \le n$求均值的情形。

这些论证通过命题5隐含地依赖于 A1–A9，但既不要求协变量服从高斯分布，也不要求协变量的次高斯范数有界。同理，这些论证都无法扩展到提供最大值而非平均值的 PIC 误差控制，而这正是渐近精确匹配所必需的。

4.2 基于指标误差距离的卡尺细化

直观来看，当协变量取自重尾分布时，PIC 可能会超过 PIC 标准误，且超出幅度远高于 $z^*_{n_p} = \sqrt{(2 \log 2n_p)}$；这与命题 2 描述的高斯情形形成对比。记 $\delta[x]$为在 x处集中的点质量分布，协变量均值的尾部越重，经验分布 $\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\delta[\mathsf{s}。\ \mathsf{d}.[(\vec{x}_i - \vec{x}_j)\tilde{\beta}]]$的尾部也越重，$\binom{n}{2}^{-}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\delta[(\vec{x}_i - \vec{x}_j)(\tilde{\beta} - \beta_n)]$, 进而 $\binom{n}{2}^{-}\quad{}^{1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\delta[(\vec{x}_i - \vec{x}_j)(\hat{}\beta - \beta_n)]$。

幸运的是，在匹配时，标准差 $\mathsf{s}\,\mathsf{d}(,)()\hat{}$ 的估计值 $|\vec{x}_i - \vec{x}_j\ C^{1/2}|_2 \quad ..ij = |\vec{x}_i - ()\hat{}$ 已经 $\vec{x}_j\ C^{1/2}|_2 \quad \{i,j\}$ 可用：只需避开 $|(\vec{x}_i - \vec{x}_j)\hat{}C^{1/2}|_2$ 过大的配对即可。将区分 i 与 j 的 $|(\vec{x}_i - \vec{x}_j)\hat{}C^{1/2}|_2$ 称为指标误差距离。命题 6 指出，即使对于次 $\sqrt{n}$ 维模型，只要恰当选择估计量 C，指标误差距离也能很好地一致估计成对指标抽样变异性。

命题 6。 *i.* 设 $\psi(r、\ \vec{x}、\ \beta)$ 为支配给定 $\vec{X}$ 条件下 R 的似然函数的梯度，其中 $\hat{}\ C_n = n^{-1}\hat{}A_n^{-1}$。在 *A1–A9* 下，

$$\left|\left(1 - \frac{|\vec{x}_i\hat{C}^{1/2}|_2}{|\vec{x}_i C^{1/2}|_2} : i \le n\right)\right|_\infty, \left|\left[1 - \frac{|(\vec{x}_i - \vec{x}_j)\hat{C}^{1/2}|_2}{|(\vec{x}_i - \vec{x}_j)C^{1/2}|_2} : i,j \le n\right]\right|_\infty \xrightarrow{P} 0. \qquad (16)$$

（此处 $0/0$ 取为 *1*。）

ii. 设 $\{\epsilon_n : \ n\}$ 满足 $\epsilon_n^{-1} = O[n/\max(p、\log n)]$。在 *A1–A11* 且 $\hat{}\ C_n =_{n^{-1}\hat{}}A_n^{-1}\ {}^{\hat{}}B_n\ {}^{\hat{}}A_n^{-1}$ 的条件下，

$$\left|\left[1 - \frac{\max(\epsilon_n^{1/2}, |\vec{x}_i\hat{C}^{1/2}|_2)}{\max(\epsilon_n^{1/2}, |\vec{x}_i C^{1/2}|_2)} : i\right]\right|_\infty \xrightarrow{P} 0.$$

若 x 和 $\mathcal{E}_n \subseteq \{\{i,j\} : \ 1 \le i、\ j \le n\}$ 满足 $|(|\vec{x}_i - \vec{x}_j|_2^2 : \ \{i,j\} \in \mathcal{E}_n)|_\infty = O_P(\max(p、\log n))$，则在 *A1–A10* 下

$$\left|\left\{1 - \frac{\max[\epsilon_n^{1/2}, |(\vec{x}_i - \vec{x}_j)\hat{C}^{1/2}|_2]}{\max[\epsilon_n^{1/2}, |(\vec{x}_i - \vec{x}_j)C^{1/2}|_2]} : \{i,j\} \in \mathcal{E}_n\right\}\right|_\infty \xrightarrow{P} 0.$$

iii. 在 *A1–A12* 条件下，(16) 成立，其中 $\hat{}\ C_n = n^{-1}\hat{}A_n^{-1}\hat{}B_n\ {}^{\hat{}}A_n^{-1}$。

我们关注符合 (i) 或 (iii) 假设的情形，这些情形保证指标误差距离 (16) 一致收敛。

让我们参照第 3.2 节的完美配对思想实验，校准指标误差距离的大小。命题 7 调整了有关高斯混沌的既有结果，以刻画该设定下 $|(\vec{X}_i - \vec{X}_j)C^{1/2}|_2$ 的最大值。

命题 7. 设 $\vec{X}_i,\ i \le n$ 相互独立且服从多元正态分布(μ, Σ)。设 C 是与 Σ 维数相同的另一个半正定矩阵，设 $\mathcal{E} \subseteq \{\{i,j\}:\ 1 \le i \ne j \le n\}$，并设 $n_{\mathcal{E}} := \mathrm{card}(\mathcal{E})$。则

$$\left\{\mathrm{E}\left[\left|\{|(\vec{X}_i - \vec{X}_j)C^{1/2}|_2 : \{i,j\} \in \mathcal{E}\}\right|_\infty^2\right]\right\}^{1/2} \le \langle 2\Sigma, C\rangle_F^{1/2}\left[1 + \left(\frac{\log n_{\mathcal{E}}}{p_{[\Sigma^{1/2}C\Sigma^{1/2}]}}\right)^{1/2}\right], \quad (17)$$

其中 $p_{[M]}$ 表示内在维数，$\mathrm{tr}(M)/|M|_2$，对于半正定的 M。

命题 7 的证明见附录 B.2。借助 $\Sigma = S^{\perp\beta_n}$ 和 $C = C_n$，它明确界定了第 3.2 节完美配对思想实验中的最坏情况成对距离 $|(\vec{X}_i - \vec{X}_j)C^{1/2}|_2$。为使实际实验中的 PIC误差控制在相近水平，我建议采用以下匹配资格要求：

$$|(\vec{x}_i - \vec{x}_j)\hat{C}^{1/2}|_2 \le \langle 2S^{\perp\hat{\beta}}, \hat{C}\rangle_F^{1/2}\left[2 + \left(\frac{\log \min(n_0, n_1)}{p-1}\right)^{1/2}\right], \quad (18)$$

作为第 3.4 节所建议的 PICs $|(\vec{x}_i - \vec{x}_j)\,\hat{}\,\beta|$ 上限 $z^*_{\min(n_0,n_1)}\langle 2S^{\perp\hat{\beta}}, \hat{}C\rangle_F^{1/2}$ 的补充。命题 7 支持约束 (18) 的启发式方法（下文将对此作出解释）更直接地提出，将 $|(\vec{x}_i - \vec{x}_j)\,\hat{}\,C^{1/2}|_2$ 限制在以下更严格的上限：

$$\langle 2S^{\perp\hat{\beta}}, \hat{C}\rangle_F^{1/2}\left[1 + \left(\frac{\log \min(n_0, n_1)}{p-1}\right)^{1/2}\right]; \quad (19)$$

但随后将会证明，对于当前目的而言，满足（18）同时采用较宽松的指标误差距离惩罚（即满足（18）但超过（19））已经足够。

为将（17）与（19）联系起来，首先回顾一下，$\min(n_0, n_1)$ 是完美配对思想实验（第 3.2 节）的规模，而 $\langle 2S^{\perp\hat{\beta}}, \hat{}C\rangle_F$ 对 $\langle 2S^{\perp\beta_n}, C_n\rangle_F$ 是一致的（命题 5）。一般而言，根据定义（另见 Tropp，2015；Vershynin，2018，第 7.8 节），$0 \le p_{[\Sigma^{1/2}C\Sigma^{1/2}]} \le p$。在特殊情形下，$(R, \vec{X})$ 具有线性判别结构且 $\mathrm{Cor}(\vec{X})$ 已知，或者 R 关于 $\vec{X}$ 呈线性关系，并且在两种情形下 $\vec{X}$ 均为高斯变量、β_n 均按相应方式估计，内在维数与外在维数相等：$p_{[\Sigma^{1/2}C_n\Sigma^{1/2}]} = p$ 或 $p-1$，分别取决于 $\Sigma = S^{(x)}$ 或 $S^{\perp\beta_n}$。若假定包含 $\min(n_0, n_1)$ 个不同的配对 $\{i,j\}$，满足 $i \ne j$ 但 $\vec{x}_i\beta_n = \vec{x}_j\beta_n$，那么这两种高斯-$X$ 特殊情形中的任一种都能很好地模拟第 3.2 节所设想的完美配对，其中 $\Sigma = S^{\perp\beta_n}$，而（19）估计协变量距离的期望最大值

$|(\vec{x}_i - \vec{x}_j)C^{1/2}|_2$（完美配对之间）。与此同时，形式为（19）的 $|(\vec{x}_i - \vec{x}_j)\hat{}\, C^{1/2}|_2$ 极限通常不应因指标正交方向上的分离而排除完美配对，因为（17）从上方近似了此类分离的期望最大值。

无论协变量来自何种分布，对协变量距离 $|(\vec{x}_i - \vec{x}_j)\hat{}\, C^{1/2}|_2$ 的限制（19）或（18），也会对 PIC $\hat{}\,\tilde{}$ 误差 $|(\vec{x}_i - \vec{x}_j)(\beta - \beta_n)|$ 施加限制。部分原因在于，如果 $\tilde{\beta}$ 是次高斯的，那么 $(\vec{x}_i - \vec{x}_j)(\beta - \beta_n)$ 也是次高斯的；如果 $\tilde{\beta} \sim$ 服从多元正态分布 $\mathrm{MVN}(\beta_n, C_n)$，那么 $(\vec{x}_i - \vec{x}_j)(\beta - \beta_n)$ 服从 $\mathrm{N}[0, |(\vec{x}_i - \vec{x}_j)C_n^{1/2}|_2^2]$。命题8（此处不附证明）汇集了第2.2节回顾的相关事实。

命题8. 设 $\mathcal{E} \subseteq \{\{i,j\} : 1 \le i, j \le n\}$ 对所有 $\{i,j\} \in \mathcal{E}$ 均满足 $|(\vec{x}_i - \vec{x}_j)C_n^{1/2}|_2 > 0$，并设 $n_{\mathcal{E}} = \mathrm{card}(\mathcal{E})$。若 $\beta \sim \mathrm{MVN}(\beta_n, C_n)$，

$$\mathrm{E}\left[\left|\left\{\frac{(\vec{x}_i - \vec{x}_j)(\tilde{\beta} - \beta_n)}{|(\vec{x}_i - \vec{x}_j)C_n^{1/2}|_2} : \{i,j\} \in \mathcal{E}\right\}\right|_\infty\right] \le z^*_{n_{\mathcal{E}}} = [2\log(2n_{\mathcal{E}})]^{1/2}$$

and

$$\mathrm{E}\left[\left|\left\{\frac{(\vec{x}_i^{\perp x\beta_n} - \vec{x}_j^{\perp x\beta_n})(\tilde{\beta} - \beta_n)}{|(\vec{x}_i^{\perp x\beta_n} - \vec{x}_j^{\perp x\beta_n})C_n^{1/2}|_2} : \{i,j\} \in \mathcal{E}\right\}\right|_\infty\right] \le z^*_{n_{\mathcal{E}}}. \tag{20}$$

~ ~如果 $\beta - \beta_n$ 不是正态分布，但属于次高斯分布，且 $\|\beta - \beta_n\|_{\psi_2}$ 有界，那么这些期望最大值仍为 $(\log n_{\mathcal{E}})^{1/2}$ 量级。

结合命题6，命题8指出，要求每个 $\min(n_0, n_1)$ 对 $\{i,j\}$ 满足 $|(\vec{x}_i - \vec{x}_j)\hat{}\, C_n^{1/2}|_2$ 低于 (19)，就能使相应的 PIC误差低于 $z^*_{\min(n_0,n_1)}$ 倍的 (19)。如果 p 的增长速度快于 $\log n$，那么 (19) 中的比值趋于 0，而 (19) 渐近等价于 PIC标准误。也就是说，将匹配限制在满足 $|(\vec{x}_i - \vec{x}_j)\hat{}\, C_n^{1/2}|_2$ 低于 (19) 的配对 $\{i,j\}$ 上，那么在给定 $\log n = o(p)$、但不对协变量的分布作特殊假设的情况下，~匹配误差的上确界 $|(\vec{x}_i - \vec{x}_j)(\beta - \beta_n)|$ 渐近上将与完美配对思想实验中的情形相同。（如果 p 的增长速度不快于 $\log n$，$p = O(\log n)$，这些误差会略高于相应思想实验中的误差，但由于 p增长缓慢，它们无论如何都会迅速趋于 0。）

这些考量表明，(19) 可作为配对距离 $|(\vec{x}_i - \vec{x}_j)\hat{}\, C^{1/2}|_2$ 的硬性上限；不过，只需满足 (18)，采用一种鼓励匹配满足 $|(\vec{x}_i - \vec{x}_j)\hat{}\, C^{1/2}|_2$ 低于 (19) 的简单策略，也能类似地控制 PIC误差。若满足以下条件，则允许 i 与 j 配对：

$$|(\vec{x}_i - \vec{x}_j)\hat{\beta}| \le z^*_{\min(n_0,n_1)}[\langle 2S^{\perp\hat{\beta}}, \hat{C}\rangle_F^{1/2} - \hat{e}(i,j)] \tag{21}$$

其中 $\hat{e}(i,j) := \left\{ |(\vec{x}_i - \vec{x}_j)\hat{C}^{1/2}|_2 - \langle 2S^{\perp\hat{\beta}}, \hat{C}\rangle_F^{1/2} \left[1 + \left(\frac{\log \min(n_0,\ n_1)}{p-1} \right)^{1/2} \right] \right\}_+$, (22)

$v_+ := \max(0,\ v)$，表示指标误差距离相较于其名义上确界 (19) 的超额。如果 $|(\vec{x}_i - \vec{x}_j)\hat{C}^{1/2}|_2$ 始终不超过该名义上确界，则 (21) 可简化为如下要求：$|(\vec{x}_i - \vec{x}_j)\hat{\beta}| \le z^*_{\min(n_0,n_1)} \langle 2S^{\perp\hat{\beta}}, \hat{C}\rangle_F^{1/2}$，正如第 3.4 节所提出的。对于 $\{i,j\}$ 与 $|(\vec{x}_i - \vec{x}_j)\hat{C}^{1/2}|_2$ 超过 (19) 的潜在配对，第 3.4 节中的 PIC 容许量 $z^*_{\min(n_0,n_1)} \langle 2S^{\perp\hat{\beta}}, \hat{C}\rangle_F^{1/2}$ 会因考虑到该配对的标准误较大而降低。当指标误差距离 $|(\vec{x}_i - \vec{x}_j)\hat{C}^{1/2}|_2$ 大到其超额 $\hat{e}(i,\ j)$ 超过 PIC 标准误时——或者等价地，大到 (18) 不成立时——(21) 将禁止 i 与 j 配对。

与单独使用非变动PIC标准误卡钳，或将其与宽度为 (19) 的成对指标误差距离卡尺结合使用相比，这种选择性收窄的PIC标准误卡钳具有重要优势。非变动PIC标准误卡钳只有在协变量为次高斯时才能保证匹配渐近精确，而选择性收窄卡钳则无需这一假设。将非变动PIC标准误卡钳与成对指标误差距离上限 (19) 结合起来，会将总和限制在

$$||[(\vec{x}_i - \vec{x}_j)\hat{\beta} : i \overset{\mathcal{S}}{\sim} j]||_\infty + ||[(\vec{x}_i - \vec{x}_j)(\tilde{\beta} - \beta_n) : i \overset{\mathcal{S}}{\sim} j]||_\infty \tag{23}$$

乘以 $z^*_{\min(n_0,n_1)}$ 与 (18) 右端相同的值，但代价是绝对不允许成对指标误差距离超过 (19)。相较之下，只要配对的PIC足够小，选择性收窄的卡钳便允许这些配对。这种额外的容许空间很重要，因为对于某些指标模型，即使在高斯 X的情况下，(19) 也会系统性地低估成对指标误差距离的上确界，这是因为它使用了$p-1$，而没有采用命题 7 的 $p_{[(S^{\perp\beta_n})^{1/2}C(S^{\perp\beta_n})^{1/2}]}$。这个小小的缺憾可以通过将 (19) 中的 $p-1$ 替换为 $p_{[(S^{\perp\hat{\beta}})^{1/2}\hat{C}((S^{\perp\hat{\beta}}))^{1/2}]}$ 来弥补，但如此一来，要使 (19) 与 PIC标准误渐近等价，就必须满足假设A12；即使在本文所考虑的模型维数限制中最弱的 A9 下，也需要这种等价关系，才能使 (23) 趋近于 0 的渐近值。下一节将讨论假设A12。

4.3 线性化误差的贡献

公式（23）省略了线性化误差。除非索引回归的估计量关于其因变量 R 呈线性，否则，要估计 $||[(\vec{x}_i - \vec{x}_j)\beta_n :\ i \sim j]||$ 我们必须 $\tilde{}$ $\hat{}$ 关注 $||[(\vec{x}_i - \vec{x}_j)(\beta - \beta) :\ i \sim j]||$ 以及 $||[(\vec{x}_i - \vec{x}_j)\beta :\ i \sim j]||$ 和 $||[(\vec{x}_i -$

$\vec{x}_j)(\beta - \beta_n)$: $i \sim j]|_\infty$。在采用线性回归以及具有固定相关性的线性判别建模这一特殊情况下，指数估计量关于 R 呈线性；但对于通过概率单位回归或逻辑回归估计的指标，则并非如此。

ˆ 回顾第 3.1 节，在次高斯协变量（A11）的情况下，$|[(\vec{x}_i - \vec{x}_j)(\beta - \tilde{\beta})$: $i, j \le n]|_\infty$ 趋于 0，前提是 $p = o\{[n/(\log\ n)]^{2/3}\}$；并且相较于以 PIC 标准误衡量的 $|[(\vec{x}_i - \vec{x}_j)(\beta - \beta_n)$: $\vec{x}_i\beta_n \approx \vec{x}_j\beta_n]|_2$，其量级至少小一个数量级，前提是 $p = o\{[n/(\log n)]^{1/2}\}$。在选择性收窄的 PIC 标准误卡尺内进行匹配，即使在不包含 A11 的条件下，也能确保这些结论成立。不过，具体采用的附加条件和估计程序不同，匹配过程可能还需要遵守其他卡尺限制。

首先考虑 A1–A9 成立，且 C_n 由 $n^{-1}\hat{A}_n^{-1}$估计的情形。匹配要求 (18) 是 (21) 的推论，它保证 $|[(\vec{x}_i - \vec{x}_j)\ \hat{C}^{1/2}|_2 : i \sim j]|_\infty = O_P\{[\max(p, \log\ n)/n]^{1/2}\}$，因为（根据命题 5）PIC 标准误为$O_P[(p/n)^{1/2}]$，且 $(p-1)^{-1}\log\min(n_0, n_1) = O[\max(1, p^{-1}\log\ n)]$。命题 6 和 $\hat{C} = n^{-1}\hat{A}_n^{-1}$进而给出 $|[(\vec{x}_i - \vec{x}_j)A^{-1/2}|_2 : i \sim j]|_\infty = O_P[\max(p, \log\ n)^{1/2}]$。由于 A3 和附录 B.3 中的引理 3 意味着 $|A_n|_2 = O(1)$，因此 $|[(\vec{x}_i - \vec{x}_j)|_2 : i \sim j]|_\infty = O_P[\max(p, \log\ n)^{1/2}]$。无论 $\vec{X}$ 是否来自次高斯分布，在选择性收窄的 PIC 标准误卡尺内选出的配对，其 $\vec{x}$ 上的分离程度都不会超过次高斯抽样下的程度，而命题1意味着

$$|[(\vec{x}_i - \vec{x}_j)(\hat{\beta} - \tilde{\beta}) : i \sim j]|_\infty = O_P[pn^{-1}(\log n)^{1/2}\max(p, \log n)^{1/2}]. \qquad (24)$$

在匹配对内，线性化误差的描述与第 3.1 节相同，即使没有 A11 或额外的匹配限制也是如此。

若改为用 $n^{-1}\hat{A}_n^{-1}B_n\hat{A}_n^{-1}$ 估计 C_n，则除 A1–A9 外，还需要 A10 才能保证 $\hat{C}_n$ 的一致性（根据命题 4）。于是，按照与上文类似的论证，$|[(\vec{x}_i - \vec{x}_j)\ \hat{C}^{1/2}|_2 : i \sim j]|_\infty = O_P\{[\max(p, \log\ n)/n]^{1/2}\}$。接着，命题 6 的情形（ii）表明，$|[(\vec{x}_i - \vec{x}_j)C^{1/2}|_2 : i \sim j]|_\infty = O_P\{[\max(p, \log\ n)/n]^{1/2}\}$，前提是 $\langle 2S^{\perp\hat{\beta}}, \hat{C}\rangle_F \left[1 + \left(\log\min(\frac{n_0\, n_1}{p-1})\right)^{1/2}\right]$,) 与 $\max(p, \log\ n)/n$ 同阶。（命题 5 表明其为 $O_P[\max(p, \log\ n)/n]$, 但这里还要求其倒数为 $O_P\{[\max(p, \log\ n)/n]^{-1}\}$。）此时还必须满足 A12，以确保 $|B_n^{-1/2}|_2 = O_P(1)$，进而确保 $|C_n^{-1/2}|_2 = O_P(n^{1/2})$。若如此，命题 1 再次给出 (24)。

然而，满秩协方差条件 A12 在实践中值得仔细考量。它无法很好地描述某些原本无可指摘的指标模型，因为长期以来，分区者一直被鼓励在这类模型中加入协变量，而不考虑它们之间的相关性（Rubin 和 Thomas，1996）。幸运的是

A12 很容易诊断，只需检查 S 和 B_n 是否都不存在病态问题即可。若该条件成立，它便能（结合引理 1 和 A6）保证 PIC标准误的下降速度不会快于 $(p/n)^{1/2}$。如果 A12 无法成立，但假设 $\langle S^{\perp\beta_n}, C_n\rangle_F = O(p/n)$ 可以成立，我们则可以将这一较弱假设与额外的匹配限制结合，以确保 $\max_{i\sim j}|\vec{x}_i - \vec{x}_j| = O_P[\max(p, \log\ n)^{1/2}]$。可以要求匹配对象之间的欧氏距离不超过宽度为 $\mathrm{tr}(S)^{1/2}\{1 + [(\log n)/(p-1)]^{1/2}\}$ 的卡尺，或者分别要求每个维度 j =1,…p 上的距离不超过宽度为 $s(x_j)\{1 + [(\log n)/(p-1)]^{1/2}\}$ 的卡尺。无论采用哪种方式，额外假设与匹配限制相结合，都能确保 (24) 成立，并确保 PIC误差的主导部分 $\hat{\ }\tilde{\ }\hat{\ }\urcorner[(\vec{x}_i - \vec{x}_j)(\beta - \beta_n)]|_\infty$ 是 $|[(\vec{x}_i - \vec{x}_j)(\beta - \beta_n)]|_\infty$，而不是 $|[(\vec{x}_i - \vec{x}_j)(\beta - \beta)]|_\infty$。

5 渐近精确匹配与影响估计的一致性

在指标模型满足温和条件的情况下，在 PIC 标准误卡尺内进行匹配，可使指标的配对差值一致地趋于零。对于倾向得分及某些其他指标模型，这种收敛恰恰是必要条件，能够确保在不存在未测量混杂时，匹配结构具有与指标配对差值一致且恒为零时相同的一致性性质。本节假设 $Z \in \{0,1\}$；并假设满足以下非常宽松的重叠条件

$$\mathrm{P}[0 < \mathrm{P}(Z = 1 \mid \mathbf{X}\beta_n) < 1] = 1; \tag{25}$$

即对于某个 $\delta > 0$，$\mathrm{E}\ |Y_C|^{1+\delta}, \mathrm{E}\ |Y_T|^{1+\delta} < \infty$；且映射 $v \mapsto \mathrm{logit}[\mathrm{P}(Z = 1|\mathbf{X}\beta = v)]$ 是利普希茨连续的。如果 $\mathbf{X}\beta$ 是基于 logit 尺度建模的倾向得分，则该映射为恒等映射，且式（25）由 Rosenbaum 和 Rubin（1983年）的重叠条件 $0 < \mathrm{P}(Z = 1 \mid \mathbf{X}) < 1$ 得出；如果 $\mathbf{X}\beta$ 是风险得分或预后得分，则式（25）的限制性低于他们已相当宽松的重叠要求。

如果一个分割 $\mathcal{S}_n$ 将 $\{1,\ldots,n\}$划分为分割元素 $\mathbf{s} \in \mathcal{S}_n$，并且这些元素满足 $\sum_{i\in\mathbf{s}} [\![z_i = z]\!] \le 1$ 中的任一条件或两者兼具，那么它就是一种精细分层设计 （Fogarty, 2018）。如果自然数 m_0, m_1不满足 $m_1 : m_0$，则这些匹配组可以是 $1 : m_0$ 或 $m_1 : 1$ 的形式，且区块同时包含 $m_0, m_1 \ge 2$；大小为 $n_\mathbf{s} = 1$ 的单元素 s 表示未匹配单位。此类**分割** 可能来自成对匹配，其中每个$\mathbf{s} \in \mathcal{S}_n$要么是一个 1: 1 匹配对，即 $\sum_{i\in\mathbf{s}}[\![z_i = 1]\!] = \sum_{i\in\mathbf{s}} [\![z_i = 0]\!] = 1$，要么是一个未匹配的 0 : 1 或 1 : 0 单元素；也可能来自允许 $1 : m$, $m \ge 1$ 匹配以及单元素的多个对照组匹配；来自 1-最近邻匹配，其中可能出现 m: 1、但不会出现 1:m 匹配组；或来自完全匹配（Rosenbaum，1991），在这种匹配中两种形式均可出现

$m:1$ 和 $1:m$ 配置，适用于任意 m；或来自带有对称限制的完全匹配（Stuart 和 Green，2006；Fredrickson 等，2020），允许两种形式的匹配组：$m:1$ 和 $1:m$，但仅限于低于指定 m_0 的 m。当只考虑分割 $\mathcal{S}_n$ 时，表示包含 i 的分割元素 $\mathbf{s} \in \mathcal{S}_n$ 的记号 $[i]_{\mathcal{S}_n}$ 可简写为 $[i]$。

考虑由 $\psi_{\mathcal{S}_n}(\cdot) = 0$ 的根所定义的估计量，其中

$$\psi_{\mathcal{S}_n}(\eta) := \frac{\sum_{\mathbf{s}\in\mathcal{S}_n}\sum_{i\in\mathbf{s}}\psi_{\mathbf{s}i}(\eta)}{\sum_{\mathbf{s}\in\mathcal{S}_n} w_{\mathbf{s}} n_{\mathbf{s}} \bar{Z}_{\mathbf{s}}(1-\bar{Z}_{\mathbf{s}})}, \quad \psi_{\mathbf{s}i}(\eta) := w_{\mathbf{s}}[Y_i - \eta(Z_i - \bar{Z}_{\mathbf{s}})](Z_i - \bar{Z}_{\mathbf{s}}), \tag{26}$$

$\bar{z}_{\mathbf{s}} := n_{\mathbf{s}}^{-1}\sum_{j\in\mathbf{s}} z_j$，且 $w_{\mathbf{s}}$ 是由 $z_{\mathbf{s}}$ 和/或 $n_{\mathbf{s}}$ 决定的非负权重。例如，对结果变量 y 关于 z 和匹配集指示变量进行普通回归时，z 的系数可表示为方程 $\psi_{\mathcal{S}_n}(\tau) = 0$ 的解 $\hat{\tau}$，其中 $w_{\mathbf{s}} \equiv 1$，因为 $(z_i - \bar{z}_{[i]} : i)$ 是 z 对匹配集指示变量进行普通回归所得的残差。第二个例子是处理组效应估计量的效应

$$n_{\{i:Z_i=1, n_{[i]}>1\}}^{-1} \sum_{\{i:Z_i=1, n_{[i]}>1\}} Y_i - \text{avg}(Y_j : Z_j = 0, j \sim i)$$

是方程 $\psi_{\mathcal{S}_n}(\cdot) = 0$ 的唯一解，其中，对于未匹配的 i，$w_{[i]_{\mathcal{S}_n}} = 0$，而对于带有 $n_{\mathbf{s}} > 1$ 的 $\mathbf{s}$，$w_{\mathbf{s}} = (1 - \bar{Z}_s)^{-1}$。

推断将通过对各层内的处理分配进行条件化来体现 $\mathcal{S}_n$，也就是说，条件化于包含 $\mathcal{F}_n := \sigma\left(\sum_{i\in\mathbf{s}} Z_i : \mathbf{s} \in \mathcal{S}_n\right)$ 的 σ-代数。尽管有这种记号，$\{\mathcal{F}_n : n\}$ 并非嵌套滤过：通常 $\mathcal{S}_m \not\subseteq \mathcal{S}_m$，因为严格包含并不意味着最大指标偏差 $|\{\vec{x}_i\beta_n - \vec{x}_j\beta_n : i \overset{\mathcal{S}_n}{\sim} j\}|_\infty$ 会随着 n 的增加而减小。定义 $\mathcal{F}_n$ 的统计量 $[\sum_{i\in\mathbf{s}} Z_i : \mathbf{s} \in \mathcal{S}_n]$ 本身并无信息，对定义为 $\eta \mapsto \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)]$ 的根的参数而言，它是 S-辅助的（Severini，2000；Lehmann 和 Romano，2022），其中 $\psi_{\mathcal{S}_n}$ 如式（26）所定义。

命题 9 指出，在关于 $\{\mathcal{S}_n$ 的正则性的温和假设下：$n\}$和 (Y_T, Y_C, Z)，式 (26) 的解趋于一个概率极限。为陈述¯正则性假设，记 $Y_{\mathbf{s}(z)} = 0$，若 $\sum_{i\in\mathbf{s}}[\![Z_i = z]\!] = 0$，其中 $z = 0$ 取 0 或 1；否则记为 (
_ $\sum_{i\in}$ _$[\![Z_i = z]\!]^{-1}\sum_{i\in} Y_i[\![Z_i = z]\!]$ V^{n} $= \bar{Y}_{\mathbf{s}(1)} - \bar{Y}_{\mathbf{s}(0)} -$ **s)s**；并令 (, s) : $\mathrm{E}(Y_{\mathbf{s}(1)} - Y_{\mathbf{s}(0)} \mid \mathcal{F}_n)$。

命题 9. 设 $\{(\vec{X}_i, Y_{Ci}, Y_{Ti}, Z_i) : i\}$ 为独立同分布，设 $\{\mathcal{S}_n : n\}$ 为精细分层设计，并设 $\mathcal{F}_n = \sigma((\sum_{i\in\mathbf{s}} Z_i : \mathbf{s} \in \mathcal{S}_n))$。假设满足以下矩条件：对于某个 $\delta > 0$，满足以下任一条件：*(i)* $\mathrm{E}\,|Y_C|^{1+\delta}$，$\mathrm{E}\,|Y_T|^{1+\delta} < \infty$，且 $n_{\mathbf{s}}$ 有界；或 *(ii)* 存在一个 V 和一个 $\delta > 0$，使得 $\mathrm{E}\,|V|^{1+\delta} < \infty$，从而对每个 n 和 $\mathbf{s} \in \mathcal{S}_n$，$|V|$

在给定 $\mathcal{F}_n$ 的条件下，随机占优 $|V^{(n,\mathbf{s})}|$ 。 [2] 设 $\{w_\mathbf{s} : \mathbf{s} \in \mathcal{S}_n\}$ 为非负的、关于 $\mathcal{F}_n$-可测的权重，并设 $\psi_{\mathcal{S}_n}(\cdot)$ 如 (26) 中所定义。假设几乎必然有：$m_n \to \infty$，其中 $m_n := \sum_{\mathbf{s}\in\mathcal{S}_n} [\![w_\mathbf{s} Z_\mathbf{s}(1-Z_\mathbf{s}) > 0]\!]$ 是 $\mathcal{S}_n$ 中除去未匹配单例和权重为零的层后的基数；$w_\mathbf{s} n_\mathbf{s} Z_\mathbf{s}(1-Z_\mathbf{s})$ 有上界；且 $m_n^{-1} \sum_{\mathbf{s}\in\mathcal{S}_n} w_\mathbf{s} n_\mathbf{s} \bar{Z}_\mathbf{s}(1-\bar{Z}_\mathbf{s})$ 远离 0。在给定 $\mathcal{F}_n$ 的条件下，对于任意满足 $\mathcal{F}_n \subseteq \mathcal{G}_n$ 的 σ-代数 $\{\mathcal{G}_n : n\}$，我们有：

i. 对于每个 η、$\psi_{\mathcal{S}_n}(\eta) - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)|\mathcal{G}_n] \to 0$，其在概率意义下以及在 L_1 中收敛；并且

ii. $\eta \mapsto \psi_{\mathcal{S}_n}(\eta)$ 和 $\eta \mapsto \mathrm{E}[\psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n]$ 具有唯一的根 $\hat{}\tau_n$ 和 τ_n。

iii. 此外，如果存在 $\tau_0 \in (-\infty, \infty)$ 使得 $\tau_n \xrightarrow{P} \tau_0$，则 $\hat{}\tau_n \xrightarrow{P} \tau_0$。

与独立同分布观测的经典相合性原理相比（Huber，1964；Serfling，1980，第 7.2.1 节引理 A），命题 9 将矩条件从估计方程贡献项 $\psi(W;\theta)$ 为 L_1 提升至 $V_{(n,\mathbf{s})}$ 为 $L_{1+\delta}$，并满足某些 $\delta > 0$。这使得结论不仅可以用 L_1 表示，还可以用概率收敛表示；反过来，这也允许将关于 $\mathcal{F}_n$ 的条件化细化为对更细 σ-代数的条件化，以反映指数得分中的匹配变异。具体而言，考虑

$$\mathcal{G}_n := \sigma\Big((\vec{X}_i\beta_n : i \le n) \cup \mathcal{F}_n\Big).$$

生成统计量 $(\vec{X}_i\beta_n,\ 1 \le i \le n; \sum_{j\in\mathbf{s}} Z_j,\ \mathbf{s} \in \mathcal{S}_n)$ 对匹配处理-对照差异（例如 τ_n）再次具有 S-辅助性。

由于禁止使用关于 $(\vec{X}_i,\ Z_i)$ 的剩余信息 $1 \le i \le n$，变换后的指数得分 $\theta_i := \mathrm{logit}[\mathrm{P}(Z = 1|\mathbf{X}\beta_n = \mathbf{x}_i\beta_n)]$ 按如下方式确定 $\mathcal{G}_n$-条件分配概率。若 $\mathbf{s} \in \mathcal{S}_n$ 且 $\zeta : \mathbf{s} \to \{0,1\}$ 满足 $\sum_{i\in\mathbf{s}} \zeta_i = \sum_{i\in\mathbf{s}} z_i$，则

$$\begin{aligned}\pi_\mathbf{s}(\zeta) &:= \mathrm{P}(Z_i = \zeta_i \text{ all } i \in \mathbf{s} \mid \mathcal{G}_n) \\ &= \begin{cases} \frac{\exp(\theta_i)}{\sum_{j\in\mathbf{s}}\exp(\theta_j)}, & \text{any } i \in \mathbf{s} \text{ s.t. } \zeta_i = 1, \text{ and } \zeta_j = 0 \text{ for all } j \in \mathbf{s}\setminus\{i\} \\ \frac{\exp(-\theta_i)}{\sum_{j\in\mathbf{s}}\exp(-\theta_j)}, & \text{any } i \in \mathbf{s} \text{ s.t. } \zeta_i = 0, \text{ and } \zeta_j = 1 \text{ for all } j \in \mathbf{s}\setminus\{i\}. \end{cases}\end{aligned} \tag{27}$$

（因为我们假设对所有 i，(25)、$\theta_i \in (-\infty, \infty)$ 都成立。当 $\mathbf{s} = \{i\}$ 是一个未匹配单例时，$\pi_\mathbf{s}(\zeta) = 1$ 对唯一允许的 $\zeta, \{i \mapsto z_i\}$ 成立。当 $\mathbf{s}$ 是一个 1 : 1 匹配对 $\{i_1, i_2\}$ 时，(27) 中的一个条件在 $i = i_1$ 下成立，而另一个条件成立

[2] 也就是说，$|V^{(n,\mathbf{s})}|$ 的 $\mathcal{F}_n$-条件分布在通常随机序下小于或等于 $|V|$ 的无条件分布，其中当且仅当对所有 a，$\mathrm{P}(W > a) \le \mathrm{P}(V > a)$ 时，$W \preceq V$。

在 $i = i_2$ 下，因此 (27) 给出了 $\pi_{\mathbf{s}}(\zeta)$ 的两个不同表达式。但对于每个 ζ，这些表达式随后都为 $\pi_{\mathbf{s}}(\zeta)$ 赋予相同的值： $\{i_1, i_2\} \to \{0, 1\}$。）现在定义

$$\tilde{\psi}_{\mathbf{s}}(\eta) := \frac{\sum_{i \in \mathbf{s}} \psi_{\mathbf{s}i}(\eta)}{n_{\mathbf{s}} \pi_{\mathbf{s}}(Z_{\mathbf{s}})}; \quad \tilde{\psi}_{\mathcal{S}_n}(\eta) := \frac{\sum_{\mathbf{s} \in \mathcal{S}_n} \tilde{\psi}_{\mathbf{s}}(\eta)}{\sum_{\mathbf{s} \in \mathcal{S}_n} w_{\mathbf{s}} n_{\mathbf{s}} \bar{Z}_{\mathbf{s}}(1 - \bar{Z}_{\mathbf{s}})}. \tag{28}$$

˜与 $\psi_{\mathbf{s}}(\eta)$ 相比：$= \sum_{i \in \mathbf{s}} \psi_{\mathbf{s}i}(\eta)$， $\psi_{\mathbf{s}}(\eta)$ 在实践中无法计算，因为˜其随机分母涉及未知的 β_n。因此， $\psi_{\mathcal{S}_n}(\cdot)$ 对效应估计缺乏直接应用。不过，它有助于分析基于 $\psi_{\mathcal{S}_n}(\cdot)$ 的估计量。

˜**命题 10**。 *i.* $\eta \mapsto \mathrm{E}[\psi_{\mathcal{S}_n}(\eta) | \mathcal{G}_n]$ 的唯一根为

$$\left(\sum_{\mathbf{s} \in \mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}} \right)^{-1} \sum_{\mathbf{s} \in \mathcal{S}_n} \tilde{w}_{\mathbf{s}} \sum_{i \in \mathbf{s}} \mathrm{E}(Y \mid Z = 1, \vec{X}\beta_n = \vec{x}_i \beta_n) - \mathrm{E}(Y \mid Z = 0, \vec{X}\beta_n = \vec{x}_i \beta_n), \tag{29}$$

其中 $\tilde{w}_{\mathbf{s}} := w_{\mathbf{s}} \bar{z}_{\mathbf{s}}(1 - \bar{z}_{\mathbf{s}})$。

ii. 对所有 η 和 n,

$$\left| \mathrm{E}[\tilde{\psi}_{\mathcal{S}_n}(\eta) - \psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n] \right| \leq [\exp(4|\{\theta_i - \theta_j : i \sim j\}|_\infty) - 1] \cdot \frac{\sum_{\mathbf{s} \in \mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}} \, \mathrm{E} \, |V^{(n,\mathbf{s})}|}{\sum_{\mathbf{s} \in \mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}}, \tag{30}$$

其中 $\theta_i = \text{logit函数}[\mathrm{P}(Z = 1 | \mathbf{X}\beta_n = \mathbf{x}_i \beta_n)]$，且 $V^{(n,\mathbf{s})}$ 的定义见上文第 *§ 5* 节。

iii. 如果 $v \mapsto \text{logit函数}[\mathrm{P}(Z = 1 | \mathbf{X}\beta = v)]$ 是利普希茨函数，且命题 *9* 的条件成立，则 $|\{(\vec{x}_i - \vec{x}_j)\beta_n : \ i \sim j\}|_\infty \to 0$ 意味着， τ_n 与 (29) 的差以概率趋于 *0*，其中 τ_n 是$\eta \mapsto \mathrm{E}[\psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n]$的唯一根。

iv. 如果平均值(29) 依概率收敛到有限极限 τ_0，则ˆ$\tau_n \xrightarrow{P} \tau_0$。

如果该指标消除了处理组分配的混杂因素，$(Y_C, Y_T) \perp Z | \mathbf{X}\beta_n$，那么 (29) 就等于平均因果效应

$$\frac{\sum_{\mathbf{s} \in \mathcal{S}_n} \tilde{w}_{\mathbf{s}} \sum_{i \in \mathbf{s}} \mathrm{E}(Y_{Ti} - Y_{Ci} \mid \vec{X}\beta_n = \vec{x}_i \beta_n)}{\sum_{\mathbf{s} \in \mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}}.$$

如果指标是倾向得分，那么这种去混杂源于 Rosenbaum 和 Rubin（1983年）意义上的强可忽略性；如果$\mathbf{X}\beta$ 是预后得分，那么在附加的“无效应修饰”条件下，强可忽略性意味着指标强可忽略性（Hansen，2008，命题3）。

参考文献

Austin, P. C.（2011），《观察性研究中估计均值差和比例差时倾向评分匹配的最优卡钳宽度》，《药学统计学》，10，150–161。Boucheron, S., Lugosi, G. 和 Massart, P.（2013），《集中不等式：独立性的非渐近理论》，牛津大学出版社。Cantoni, E. 和 Ronchetti, E.（2001），《广义线性模型的稳健推断》，《美国统计学会杂志》，96，1022–1030。Durrett, R.（2019），《概率论：理论与例题》，第49卷，剑桥，第5版。D'Amour, A., Ding, P., Feller, A., Lei, L. 和 Sekhon, J.（2021），《高维协变量观察性研究中的重叠》，《计量经济学杂志》，221，644–654。Firth, D.（1993），《最大似然估计值的偏差缩减》，《生物统计学》，80，27–38。Fogarty, C. B.（2018），《论缓解精细分层实验的分析局限》，《英国皇家统计学会杂志：B辑（统计方法学）》，80，1035–1056。Fredrickson, M. M., Errickson, J. 和 Hansen, B. B.（2020），《评论：源自大型行政数据库的观察性研究匹配方法》，《Statist. Sci.》，35，361–366。Gelman, A., Jakulin, A., Pittau, M. G. 和 Su, Y.-S.（2008），《适用于逻辑斯蒂回归及其他回归模型的弱信息默认先验分布》，《应用统计年刊》，1360–1383。Gentle, J. E.（2007），《矩阵代数》，《统计学施普林格丛书》，施普林格，纽约，NY，doi，10，978–0。Ghosh, D. 和 Cort´es, E. C.（2019），《用于高维协变量重叠与因果效应估计的高斯过程框架》，《因果推断杂志》，7。Hansen, B. B.（2008），《倾向得分的预后评分对应物》，《生物统计学》，95，481–488，doi：10.1093/biomet/asn004。

He, X. 和 Shao, Q.-M.（2000），《维数递增的参数》，《多元分析杂志》，73，120–135。Huber, P. J.（1964），《位置参数的稳健估计》，《数理统计年鉴》，35，73–101。Kosmidis, I. 和 Firth, D.（2009），《指数族非线性模型中的偏差缩减》，《生物统计学》，96，793–804。Lehmann, E. L. 和 Romano, J. P.（2022），《统计假设检验》，施普林格科学与商业媒体，第4版。Liang, K.-Y. 和 Zeger, S. L.（1986），《使用广义线性模型进行纵向数据分析》，《生物统计学》，73，13–22。Lunt, M.（2013），《选择合适的卡钳对于通过倾向得分匹配实现良好平衡至关重要》，《美国流行病学杂志》，179，226–235。Portnoy, S.（1988），《当参数数量趋于无穷时指数族似然方法的渐近行为》，《统计年鉴》，356–366。Rosenbaum, P. R.（1991），《观察性研究最优设计的特征》，《皇家统计学会杂志》，53，597–610。Rosenbaum, P. R. 和 Rubin, D. B.（1983），《倾向得分在观察性研究因果效应中的核心作用》，《生物统计学》，70，41–55。—（1985），《利用纳入倾向得分的多元匹配抽样方法构建对照组》，《美国统计学家》，39，33–38。Rubin, D. B. 和 Thomas, N.（1996），《使用估计的倾向得分进行匹配：将理论与实践联系起来》，《生物统计学》，52，249–64。—（2000），《将倾向得分匹配与预后协变量的额外调整相结合》，《美国统计学会杂志》，95，573–585。Serfling, R. J.（1980），《数理统计逼近定理》，约翰·威利父子公司。Severini, T.（2000），《统计学中的似然方法》，牛津大学出版社（美国）。

Stuart, E. A. 和 Green, K. M.（2006），《使用完全匹配估计非实验研究中的因果效应：考察青少年大麻使用与成年期结局之间的关系》，技术报告，约翰斯·霍普金斯大学。

Tropp, J. A. (2015), “矩阵集中不等式导论”，机器学习基础与趋势,8, 1–230.

van der Vaart, A. W. (1998),渐近统计, 剑桥大学出版社。

Vershynin, R.(2018), 高维概率论：数据科学应用导论, 第47卷，剑桥大学出版社。

Wang, L. (2011), “协变量数量渐增时聚类二元数据的广义估计方程分析”，《统计年鉴》,39, 389–417.

Wang, Y., Cai, H., Li, C., Jiang, Z., Wang, L., Song, J., and Xia, J. (2013), “三治疗组倾向得分匹配的最优卡钳宽度：一项蒙特卡罗研究”，*PloS one*, 8, e81045.

数学符号综述

符号 $|\cdot|_2$ 和 $|\cdot|_\infty$ 分别表示通常意义下的欧几里得范数和上确界范数（§ 3.2）。对于标量或向量随机变量 V， $\|V\|_{\psi_2}$ 是 V的次高斯范数；对于固定矩阵 M，$|M|_2$和 $|M|_F$分别是 M的算子范数和弗罗贝尼乌斯范数（§ 2.2）。对于维数相同的矩阵 M和 N， $\langle M,N\rangle_F$表示弗罗贝尼乌斯内积 tr($M'N$)（§ 3.2）。

符号$\hat{\beta}(r,\vec{x},\beta)$ 和 $c_0(r,\vec{x},\beta_0+\vec{x}\beta)$ 在第 2.1 节中定义，而第 2.3 节定义了 β_n, β, A_n 和 B_n。对于 $\{1,\ldots,n\}$ 的划分 $\mathcal{S}$， $[i]_{\mathcal{S}}$表示属于 $\mathcal{S}$ 且包含 i 的 $\{1,\ldots,n\}$ 子集； $i \overset{\mathcal{S}}{\sim} j$表示存在某个 s $\in \mathcal{S}$，使得其中同时有 $i \in$s 和 $j \in$ s（§ 3）。第 3.2 节定义了 n_S （即 card(S)）和z_m^*，其中 m 为正整数。 $[\![\mathcal{A}]\!]$ 是事件 $\mathcal{A}$的指示函数。第 5 节定义了 s $\subseteq \{1,\ldots,n\}$ 对应的 v_j s $\bar{v}_{\mathbf{s}} = n_{\mathbf{s}}^{-1}\sum_{j\in}$；给出了估计函数 $\psi_{\mathcal{S}_n}(\cdot)$ 和$\tilde{\psi}_{\mathcal{S}_n}(\cdot)$，以及 σ-代数 $\mathcal{F}_n$ 和 $\mathcal{G}_n$，它们对应于 $\{1,\ldots,n\}$ 的划分 $\mathcal{S}_n$；此外，还定义了 s $\subseteq \{1,\ldots,n\}$ 对应的 $z \in \{0,1\}$ $\bar{v}_{\mathbf{s}(z)}$。

B 第 3 节的证明

B.1 第 3.1 节

命题 *2* 的证明。根据 (8) 和 IS 可估性 (A2, A3)，($\tilde{\beta}_n - \beta_n$) 与 $A_n^{-1}\frac{1}{n}[\sum_{j=1}^n \psi(R_j, \vec{x}_j; \beta_n)]$ 之间的差的欧几里得范数的阶小于 $n^{-1/2}$。由于该差是非随机的，其次高斯范数也为 $o(n^{-1/2})$。因此，只需证明 $\|A_n^{-1}\frac{1}{n}[\sum_{j=1}^n \psi(R_j, \vec{x}_j; \beta_n)]\|_{\psi_2} = O(n^{-1/2})$。由 A3 可知，为此只需进一步证明 $\|\sum_{j=1}^n \psi(R_j, \vec{x}_j; \beta_n)\|_{\psi_2} = O(n^{1/2})$。

根据（4），

$$\begin{aligned}\|\sum_{i=1}^n \psi(R_i, \vec{x}_i; \beta_n)\|_{\psi_2} &= \sup_{\gamma:|\gamma|_2=1} \|\sum_{i=1}^n \gamma'\psi(R_i, \vec{x}_i; \beta_n)\|_{\psi_2} \\ &= \sup_{\gamma:|\gamma|_2=1} \|\sum_{1}^n c_0(R_i, \vec{x}_i, \vec{x}_i\beta_n)w(\vec{x}_i)\vec{x}_i\gamma\|_{\psi_2}.\end{aligned}$$

根据 A5，设 k_1 是 $\|c_0(R_i, \vec{x}_i, \vec{x}_i\beta_n)\|_{\psi_2}$ 的一个界。根据一般的 Ho- effding 不等式（Vershynin，2018，§ 2.6），存在一个通用的 k_0，使得

$$\begin{aligned}\|\sum_{1}^n c_0(R_i, \vec{x}_i, \vec{x}_i\beta)w(\vec{x}_i)\vec{x}_i\gamma\|_{\psi_2}^2 &\le k_0 \sum_{1}^n \|c_0(R_i, \vec{x}_i, \vec{x}_i\beta)w(\vec{x}_i)\vec{x}_i\gamma\|_{\psi_2}^2 \\ &\le k_0 k_1 \sum_{1}^n |w(\vec{x}_i)\vec{x}_i\gamma|_2^2. \text{ So} \\ \sup_{\gamma:|\gamma|_2=1} \|\sum_{1}^n c_0(R_i, \vec{x}_i, \vec{x}_i\beta)w(\vec{x}_i)\vec{x}_i\gamma\|_{\psi_2}^2 &\le k_0 k_1 n\left(\frac{1}{n}\sum_{1}^n w(\vec{x}_i)^2\right)\cdot \\ &\quad \sup_{\gamma:|\gamma|_2=1}\left(\frac{1}{n}\sum_{1}^n |\vec{x}_i\gamma|_2^2\right).\end{aligned}$$

左侧等于 $\|\sum_1^n c_0(R_i, \vec{x}_i, \vec{x}_i\beta)\vec{x}_i\|_{\psi_2}$ 的平方，而 A6 表明右侧的乘积为 $O(n)$。证明完毕。 □

B.2 第 3.2 节的证明

以下引理有助于证明命题 3。

引理 2. 在命题 *3* 的条件下，对于所有 $1 \le i < j < n$ 我们有 $(\vec{X}_i - \vec{X}_j)$ $\perp \tilde{\beta} - \beta_n \mid \mathcal{S}, \{\vec{\mu}(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$。

证明。 回想一下 $\tilde{\beta} - \beta_n = A_n^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi(R_i, \vec{X}_i; \beta_n)$。在记号中省略对 $\mathcal{S}, \{\mu_{\mathbf{s}} : \mathbf{s} \in \mathcal{S}\}$ 的条件化，

$$\begin{aligned}\mathrm{Cov}\left[A_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\psi(R_i,\vec{X}_i;\beta_n),\vec{X}_1-\vec{X}_2\right] = \; & \tfrac{1}{n}\mathrm{Cov}\left[A_n^{-1}\psi(R_1,\vec{X}_1;\beta_n),\vec{X}_1\right] - \\ & \tfrac{1}{n}\mathrm{Cov}\left[A_n^{-1}\psi(R_2,\vec{X}_2;\beta_n),\vec{X}_2\right] \\ & = 0.\end{aligned}$$

根据 $\{[\vec{X}_i$ 的联合正态性，$\psi'(R_i, \vec{X}_i; \beta_n)] : i\}$、$\frac{1}{n} A_n^{-1} \psi(R_1, \vec{X}_1; \beta_n)$ 和 $(\vec{X}_i - \vec{X}_j)$ 联合服从正态分布；它们不相关这一事实意味着它们相互独立。

□

命题 *3* 的证明。对于固定的 $\gamma \in \Re^p$，经过一些我省略的代数运算可得，

$$\begin{aligned}\mathrm{E}[(\vec{X}_1 - \vec{X}_2)\gamma]^2 &= 2\gamma'\Sigma\gamma \quad \text{and} \\ \mathrm{E}\left|\{(\vec{X}_i - \vec{X}_j)\gamma : i \overset{\mathcal{S}}{\sim} j, i \neq j\}\right|_2^2 &= 2\gamma'\Sigma\gamma.\end{aligned}$$

（在整个证明中，我用“$\mathrm{E}[\cdot]$”表示“$\mathrm{E}_{\mathcal{S}}[\cdot]$。”）根据引理 2 中得出的条件独立性，可知

$$\mathrm{E}\left|\{(\vec{X}_i - \vec{X}_j)(\tilde{\beta} - \beta_n) : i \overset{\mathcal{S}}{\sim} j, i \neq j\}\right|_2^2 = \langle 2\Sigma, C\rangle_F.$$

对于固定的 γ，将 (5) 应用于正态分布变量，也可得

$$\mathrm{E}\left|\{(\vec{X}_i - \vec{X}_j)\gamma : i \overset{\mathcal{S}}{\sim} j, i \neq j\}\right|_\infty \leq [4\gamma'\Sigma\gamma \log 2n_{\mathcal{S}}]^{1/2}. \tag{31}$$

由于对于每个 $1 \leq i < j \leq n$，$\tilde{\beta}$ 与 $\vec{X}_i - \vec{X}_j$ 相互独立（引理 2），因此可 (31) 得

$$\begin{aligned}\mathrm{E}\left\{\mathrm{E}\left[\left|\{(\vec{X}_i - \vec{X}_j)(\tilde{\beta} - \beta_n) : i \overset{\mathcal{S}}{\sim} j, i \neq j\}\right|_\infty \mid \tilde{\beta}\right]^2\right\} \leq & \\ \mathrm{E}\left\{4(\tilde{\beta} - \beta_n)'\Sigma(\tilde{\beta} - \beta_n)\log 2n_{\mathcal{S}}\right\} &= 4\langle\Sigma, C_n\rangle_F \log 2n_{\mathcal{S}}.\end{aligned}$$

将这一事实与条件期望的詹森不等式结合，

$$\begin{aligned}\left\{\mathrm{E}\left[\left|\{(\vec{X}_i - \vec{X}_j)(\tilde{\beta} - \beta_n) : i \overset{\mathcal{S}}{\sim} j, i \neq j\}\right|_\infty\right]\right\}^2 \leq & \\ \mathrm{E}\left\{\mathrm{E}\left[\left|\{(\vec{X}_i - \vec{X}_j)(\tilde{\beta} - \beta_n) : i \overset{\mathcal{S}}{\sim} j, i \neq j\}\right|_\infty \mid \tilde{\beta}\right]^2\right\} &\leq 4\langle\Sigma, C_n\rangle_F \log 2n_{\mathcal{S}}.\end{aligned}$$

□

B.3 第 3.3 节

引理 1 的证明。令 $k < \infty$ 表示 $\{c_0(R_i, \vec{x}_i, \vec{x}_i\beta_n)$ 的次高斯范数的上确界（条件 A5）：对所有 i，$\mathrm{E}[c_0(R_i, \vec{x}_i, \vec{x}_i\beta_n)^4] \leq (2k)^4$（例如，Vershynin，2018，第 2.5.2 节）。因此，$\sum_i \mathrm{E}\left[c_0(R_i, \vec{x}_i, \vec{x}_i\beta_n)^4\right] = O(n)$，进而$\sum_i \{\mathrm{E}\left[c_0(R_i, \vec{x}_i, \vec{x}_i\beta_n)^2\right]\}^2 = O(n)$。将其与 $\sup_{\gamma:|\gamma|_2=1} \sum_i w(\vec{x}_i)^4(\vec{x}_i\gamma)^4 = O(n)$（条件 A6）结合，柯西-施瓦茨不等式给出 $\sup_{\gamma:|\gamma|_2=1} \sum_i \mathrm{E}\left[c_0(R_i, \vec{x}_i, \vec{x}_i\beta_n)^2\right] w(\vec{x}_i)^2(\vec{x}_i\gamma)^2 = O(n)$。根据式 (4) 重排各项可得，$\sup_{\gamma:|\gamma|_2=1} \sum_i \mathrm{E}\left\{[\gamma'\psi(R_i, \vec{x}_i, \beta)]^2\right\} = O(n)$，或者 $\sup_{\gamma:|\gamma|_2=1} \gamma' n B_n \gamma = O(n)$；因此 $|B_n|_2 = O(1)$。条件 A3 给出$|A_n^{-1}|_2 = O(1)$，所以也有 $|C_n^{-1}|_2 = O(n^{-1})$。 □

我们对命题 4 的证明依赖于以下三个辅助引理。

引理 3. 在 *A4* 和 *A6* 条件下，$\sup_\beta |A_n(\beta)|_2 = O(1)$。

证明。记 $c_1(\vec{x}, \eta) := (\partial/\partial\eta)\,\mathrm{E}[c_0(R, \vec{x}, \eta) \mid \vec{X} = \vec{x}]$，使得 $\nabla_\beta \mathrm{E}[\psi(R, \vec{x}, \beta) \mid \vec{X} = \vec{x}] = c_1(\vec{x}, \vec{x}\beta)\, w(\vec{x})\vec{x}'\vec{x}$。由 A4，存在 $K_1 < \infty$，使得对任意 β，$|c_1(\vec{x}, \beta_0 + \vec{x}\beta)| < K_1$。由柯西-施瓦茨不等式，A6 给出 $\sup_{\gamma:|\gamma|_2=1} \sum_{i=1}^n w(\vec{x}_i)(\vec{x}_i\gamma)^2 = O(n)$。由于

$$n|A_n(\beta)|_2 = \sup_{\gamma:|\gamma|_2=1} \gamma' \left[\sum_i c_1(\vec{x}_i, \vec{x}_i\beta) w(\vec{x}_i)\vec{x}'\vec{x}\right]\gamma \leq K_1 \sup_{\gamma:|\gamma|_2=1} \sum_i w(\vec{x}_i)(\vec{x}\gamma)^2$$

对于任意 β，结论成立。 □

引理 4. 在 *A1–A9* 条件下，$|A_n(\hat{}\beta) - A_n(\beta_n)|_2 \xrightarrow{P} 0$ 以及 $|\hat{}B_n(\hat{}\beta) - \hat{}B_n(\beta_n)|_2 \xrightarrow{P} 0$。

引理 5. 在 *A1–A9*、*A11* 和次-$\sqrt{n}$ 维数条件下，$|B_n(\beta_n) - B_n(\beta_n)|_2 \xrightarrow{P} 0$。

引理 4 和引理 5 的证明见命题 4 的证明之后。

命题 *4* 的证明。由于 $\hat{} A_n \equiv A_n(\hat{}\beta)$，$|\hat{}A_n - A_n|_2 \xrightarrow{P} 0$ 由引理 4 得出。由于 $\hat{} B_n \equiv \hat{}B_n(\hat{}\beta)$，$|\hat{}B_n - B_n|_2 \xrightarrow{P} 0$ 由 $|\hat{}B_n(\hat{}\beta) - \hat{}B_n(\beta_n)|_2 \xrightarrow{P} 0$ (引理 4) 和 $|\hat{}B_n(\beta_n) - B_n|_2 \xrightarrow{P} 0$ (引理 5) 得出。

由于 $|A_n^{-1}|_2 = O(1)$ 且 $|\hat{}A_n^{-1}|_2 = O_P(1)$（条件 A3 和命题 1），对右侧 $(\hat{} A_n^{-1} - A_n^{-1}) = \hat{}A_n^{-1}(\hat{}A_n - A_n)A_n^{-1}$ 应用谱范数的次乘性可得 $|\hat{}A_n^{-1} - A_n^{-1}|_2 \xrightarrow{P} 0$。又由于 $|B_n|_2 = O_P(1)$（引理 1），$\hat{} A_n^{-1}\hat{}B_n\hat{}A_n^{-1}$ 中第二项和第三项的 2-范数

$$A_n^{-1}B_nA_n^{-1} + (\hat{A}_n^{-1} - A_n^{-1})B_nA_n^{-1} + \hat{A}_n^{-1}B_n(\hat{A}_n^{-1} - A_n^{-1}) + \hat{A}_n^{-1}(\hat{B}_n - B_n)\hat{A}_n^{-1}$$

必然依概率趋于 0。因此，$|A_n^{-1}B_nA_n^{-1} - \hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}|_2$ 的随机阶不可能高于 $|\hat{A}_n^{-1}(\hat{B}_n - B_n)\hat{A}_n^{-1}|_2$ 的随机阶。但由于 A4 和 A6 意味着 $|\hat{A}_n|_2 = O_P(1)$，根据引理 3，$O_P(|\hat{A}_n^{-1}(\hat{B}_n - B_n)\hat{A}_n^{-1}|_2) = O_P(|\hat{B}_n - B_n|_2) = o_P(1)$；这意味着 $|C_n - \hat{C}_n|_2 = n^{-1}|A_n^{-1}B_nA_n^{-1} - \hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}|_2 = o_P(n^{-1})$。 □

这个引理 4 的证明部分基于 Wang 对广义估计方程中一个类似原理的证明（2011年，定理 3.10）。

引理 4 的证明。为证明 $|A_n(\hat{\beta}) - A_n(\beta_n)|_2 = \sup_{\gamma:|\gamma|_2=1} \gamma'[A_n(\hat{\beta}) - A_n(\beta_n)]\gamma \xrightarrow{P} 0$，设 $K_1 < \infty$ 为 $\eta \mapsto c_1(r, \vec{x}, \eta)$ 关于每个 r 和 $\vec{x}$ 的利普希茨常数，其中 $c_1(\cdot)$ 的定义如上文引理 3 的证明所述。（根据条件 A4。）于是

$$|\gamma'\{\nabla_\beta \mathrm{E}[\psi(R, \vec{x}_i, \beta) \mid \vec{X} = \vec{x}_i]_{\beta=\hat{\beta}} - \nabla_\beta \mathrm{E}[\psi(R, \vec{x}_i, \beta) \mid \vec{X} = \vec{x}_i]_{\beta=\beta_n}\}\gamma| \\ \leq K_1[\vec{x}_i(\hat{\beta} - \beta_n)]w(\vec{x}_i)(\vec{x}_i\gamma)^2. \quad (32)$$

对 i 求和并应用柯西-施瓦茨不等式，

$$(\gamma'[A_n(\hat{\beta}) - A_n(\beta_n)]\gamma)^2 \leq K_1^2 \left[\frac{1}{n}\sum_{i=1}^n w(\vec{x}_i)^2(\vec{x}_i\gamma)^4\right] \left[\frac{1}{n}\sum_{i=1}^n \left(\vec{x}_i \frac{\hat{\beta} - \beta_n}{|\hat{\beta} - \beta_n|_2}\right)^2\right] \times |\hat{\beta} - \beta_n|_2^2,$$

当 $|\delta|_2 = 0$ 时，将“$\vec{x}(\delta/|\delta|_2)$”解释为 0。由此可得

$$|A_n(\hat{\beta}) - A_n(\beta_n)|_2^2 \leq K_1^2 \left[\sup_{\gamma:|\gamma|_2=1} \frac{1}{n}\sum_{i=1}^n w(\vec{x}_i)^2(\vec{x}_i\gamma)^4\right] \left[\sup_{\delta:|\delta|_2=1} \frac{1}{n}\sum_{i=1}^n (\vec{x}_i\delta)^2\right] |\hat{\beta} - \beta_n|_2^2.$$

注意到，命题1的条件可由命题 4的条件推出，因此我们可以假设 $|\hat{\beta} - \beta_n|_2 = O_P[(p/n)^{1/2}]$。结合条件 A6，可得 $|A_n(\hat{\beta}) - A_n(\beta_n)|_2 = O(1)O(1)O_P[(p/n)^{1/2}]$，根据 A9，这就是 $o_P(1)$。

至于 $|\hat{B}_n(\hat{\beta}) - \hat{B}_n(\beta_n)|_2$，

$$\begin{aligned}
&\gamma'\{\Psi(R_i, \vec{x}_i, \hat{\beta})\Psi(R_i, \vec{x}_i, \hat{\beta})' - \Psi(R_i, \vec{x}_i, \beta_n)\Psi(R_i, \vec{x}_i, \beta_n)'\}\gamma \quad (33)\\
&\quad = (c_0^2(R_i, \vec{x}_i, \hat{\eta}_i) - c_0^2(R_i, \vec{x}_i, \eta_i))w(\vec{x}_i)^2(\vec{x}_i\gamma)^2\\
&\quad = (c_0(R_i, \vec{x}_i, \hat{\eta}_i) - c_0(R_i, \vec{x}_i, \eta_i))(c_0(R_i, \vec{x}_i, \hat{\eta}_i) + c_0(R_i, \vec{x}_i, \eta_i))w(\vec{x}_i)^2(\vec{x}_i\gamma)^2\\
&\quad = (c_0(R_i, \vec{x}_i, \hat{\eta}_i) - c_0(R_i, \vec{x}_i, \eta_i))^2 w(\vec{x}_i)^2(\vec{x}_i\gamma)^2\\
&\qquad + (c_0(R_i, \vec{x}_i, \hat{\eta}_i) - c_0(R_i, \vec{x}_i, \eta_i)) \cdot 2c_0(R_i, \vec{x}_i, \eta_i)w(\vec{x}_i)^2(\vec{x}_i\gamma)^2\\
&\quad =: V_i(\gamma) + W_i(\gamma). \quad (34)
\end{aligned}$$

根据 $c_0(r,\ \vec{x},\ \cdot)$的利普希茨性质 (A4),

$$\begin{aligned}\sup_{\gamma:|\gamma|_2=1}\frac{1}{n}\sum_i|V_i(\gamma)| \leq & K_1^2\sup_{\gamma:|\gamma|_2=1}\frac{1}{n}\sum_i(\hat{\eta}_i-\eta_i)^2w(\vec{x}_i)^2(\vec{x}_i\gamma)^2\\ \leq & K_1^2|\hat{\beta}-\beta_n|_2^2\sup_{\delta,\gamma:|\delta|_2=|\gamma|_2=1}\frac{1}{n}\sum_i w(\vec{x}_i)^2(\vec{x}_i\delta)^2(\vec{x}_i\gamma)^2\\ = & O(1)O_P(p/n)O(1)=o_P(1),\end{aligned}\tag{35}$$

ˆ 根据假设A6以及 β 在（35）处的一致性。$c_0(r,\ \vec{x},\ \cdot)$的利普希茨性质也给出

$$\begin{aligned}\sup_{\gamma:|\gamma|_2=1}\frac{1}{n}\sum_i|W_i(\gamma)| \leq & 2K_1\sup_{\gamma:|\gamma|_2=1}\frac{1}{n}\sum_i|c_0(R_i,\vec{x}_i,\eta_i)(\hat{\eta}_i-\eta_i)|w(\vec{x}_i)^2(\vec{x}_i\gamma)^2\\ \leq & 2K_1|\hat{\beta}-\beta_n|_2\sup_{\gamma,\delta:|\gamma|_2=|\delta|_2=1}\frac{1}{n}\sum_i c_0(R_i,\vec{x}_i,\eta_i)(\vec{x}_i\delta)w(\vec{x}_i)^2(\vec{x}_i\gamma)^2\\ \leq & 2K_1|\hat{\beta}-\beta_n|_2[\frac{1}{n}\sum_i c_0^4(R_i,\vec{x}_i,\eta_i)]^{1/4}\\ & \times\sup_{\delta:|\delta|_2=1}[\frac{1}{n}\sum_i w(\vec{x}_i)^4(\vec{x}_i\delta)^4]^{1/4}\sup_{\gamma:|\gamma|_2=1}[\frac{1}{n}\sum_i w(\vec{x}_i)^4(\vec{x}_i\gamma)^4]^{1/2}\\ = & O_P(p/n)O_P(1)O(1)O(1)=o_P(1).\end{aligned}\tag{36}$$

此处我们在（36）处两次应用柯西-施瓦茨不等式；要推导到下一行，则需依据命题1使用 β 的一致性，并结合假设A5、马尔可夫不等式和假设A6。由于

$$\begin{aligned}|\hat{B}_n(\hat{\beta})-\hat{B}_n(\beta_n)|_2 = & \sup_{\gamma:|\gamma|_2=1}\gamma'(\hat{B}_n(\hat{\beta})-\hat{B}_n(\beta_n))\gamma\\ \leq & \sup_{\gamma:|\gamma|_2=1}\frac{1}{n}\sum_i|V_i(\gamma)|+\sup_{\gamma:|\gamma|_2=1}\frac{1}{n}\sum_i|W_i(\gamma)|,\end{aligned}$$

结果由此得出。 □

引理 *5* 的证明。为了对照组 $|\gamma[\hat{}B_n(\beta_n)-B_n(\beta_n)]\gamma'|$, 取满足 $|\gamma|_2=1$ 的 γ，并考虑

$$\begin{aligned}\gamma[\hat{B}_n(\beta_n)-B_n(\beta_n)]\gamma' = & \gamma'\big(n^{-1}\sum_{i\leq n}\{c_0^2(R_i,\vec{x}_i,\vec{x}_i\beta_n)-\mathrm{E}[c_0^2(R,\vec{x}_i,\vec{x}_i\beta_n)\mid\vec{X}=\vec{x}_i]\}w^2(\vec{x}_i)\vec{x}_i'\vec{x}_i\big)\gamma\\ = & n^{-1}\sum_{i\leq n}\{c_{0i}^2(R_i)-\mathrm{E}[c_{0i}^2(R)\mid\vec{X}=\vec{x}_i]\}w^2(\vec{x}_i)(\vec{x}_i\gamma)^2.\end{aligned}$$

注意，A5 意味着随机变量 $c_{0i}^2(R_i) - \mathrm{E}[c_{0i}^2(R) \mid \vec{X} = \vec{x}_i]$ 是次指数随机变量，且其次指数范数一致有界（Vershynin，2018，§ 2.7）。应用伯恩斯坦不等式（Vershynin，2018，§ 2.8），

$$\mathrm{P}\{|\gamma[\hat{B}_n(\beta_n) - B_n(\beta_n)]\gamma'| \geq t\} \leq 2\exp\left\{-k_2 \min\left(\frac{n^2t^2}{k_e^2\sum_{i=1}^n w(\vec{x}_i)^4(\vec{x}_i\gamma)^4}, \frac{nt}{k_e \max_{i\leq n} w(\vec{x}_i)^2(\vec{x}_i\gamma)^2}\right)\right\}, \quad (37)$$

其中，k_e 是 $c_{0i}^2(R_i) - \mathrm{E}[c_{0i}^2(R_i)], i \geq 1$的次指数范数的有限上界，而 k_2 是一个通用常数。

现在令 $\mathcal{N}$ 为 p 维球面的一个 $1/4$-网，即 $\{\gamma : |\gamma|_2 = 1\}$ 的一个有限子集，满足 $\{\gamma : |\gamma|_2 = 1\}$ 可由以$\mathcal{N}$ 为中心、半径为 $1/4$的球覆盖，因此（Vershynin，2018，§ 4.4.1）

$$\sup_{\gamma:|\gamma|_2=1} \gamma[\hat{B}_n(\beta_n) - B_n(\beta_n)]\gamma' \leq 2\sup_{\gamma\in\mathcal{N}} \gamma[\hat{B}_n(\beta_n) - B_n(\beta_n)]\gamma'. \quad (38)$$

我们可以选取该 $\mathcal{N}$，使其基数不超过 9^p（Vershynin，2018，推论 4.2.13）。

由于（37）对单位球面上的任意 γ 都成立，因此可得

$$\frac{1}{2}\mathrm{P}\{\sup_{\gamma:|\gamma|_2=1} |\gamma[\hat{B}_n(\beta_n) - B_n(\beta_n)]\gamma'| \geq t\} \leq \exp\left\{p\log(9) - nk_2 \min\left(\frac{t^2}{k_e^2 n^{-1}\sum_{i=1}^n w(\vec{x}_i)^4(\vec{x}_i\gamma)^4}, \frac{t}{k_e \max_{i\leq n} w(\vec{x}_i)^2(\vec{x}_i\gamma)^2}\right)\right\} = \exp\left\{-k_2 \min\left(\frac{nt^2}{k_e^2 n^{-1}\sum_{i=1}^n w(\vec{x}_i)^4(\vec{x}_i\gamma)^4} - pk_3, \frac{nt}{k_e \max_{i\leq n} w(\vec{x}_i)^2(\vec{x}_i\gamma)^2} - pk_3\right)\right\}, \quad (39)$$

其中 $k_3 = \log(9)/k_2$。回想一下，$n^{-1}\sum_{i=1}^n w(\vec{x}_i)^4(\vec{x}_i\gamma)^4 = O(1)$（由 A6），$t^2/[k_e^2 n^{-1}\sum_{i=1}^n w(\vec{x}_i)^4(\vec{x}_i\gamma)^4]$ 具有正的下极限和有限的上极限（对任意 t）；因此，取最小值的两个量中第一个趋于 ∞，因为 $p = o(n)$（A9）。由于 $\max_{i\leq n} w(\vec{x}_i)^2(\vec{x}_i\gamma)^2 = O\{\max[p\log n, (\log n)^2]\}$（A11），$p^2\log n = o(n)$（次-$\sqrt{n}$ 维数）意味着第二个量也必然趋于 ∞。（如果 $p_n \leq \log$ 无穷多次成立，那么在满足这一条件的子序列上，相关项有下界 $(\log n)\{(t/k_e)n/[(\log n)\max_{i\leq n} w(\vec{x}_i)^2(\vec{x}_i\gamma)^2] - k_3\}$，它趋于 ∞，且 $(\log n)[n/(\log n)^3 - 1]$。否则 $p_n > \log n$，因此 $\max[p_n\log n, (\log n)^2] = p_n\log n$。相关项等于 $p\{(t/k_e)n/[p\max_{i\leq n} w(\vec{x}_i)^2(\vec{x}_i\gamma)^2] - k_3\}$，它趋于 ∞，因为 $p\uparrow\infty$，$\max_{i\leq n} w(\vec{x}_i)^2(\vec{x}_i\gamma)^2 = O(p\log n)$，且 $p^2\log n = o(n)$。）因此，(39) 中的最小值无限增大，而 (39) 本身趋于 0。

□

命题 5 的证明使用了两个辅助引理。

引理 6。 在 *A1* 和 *A6* 下， $s(\mathbf{x}\hat{\beta}) - s(\mathbf{x}\beta_n) = O_P(|\hat{\beta} - \beta_n|_2)$。

证明。根据 A1,

$$\begin{aligned} s^2(\mathbf{x}\hat{\beta}) - s^2(\mathbf{x}\beta_n) &= \hat{\beta}' S\hat{\beta} - \beta_n' S\beta_n \\ &= (\hat{\beta} + \beta_n)' S(\hat{\beta} - \beta_n) \\ &= (S^{1/2}\hat{\beta} + S^{1/2}\beta_n)' S^{1/2}(\hat{\beta} - \beta_n), \end{aligned}$$

其中， $S^{1/2}$ 表示 S的矩阵平方根。注意到 A6 意味着$|S^{1/2}|_2 = O(1)$,

$$\begin{aligned} |s(\mathbf{x}\hat{\beta}) - s(\mathbf{x}\beta_n)| =& [s(\mathbf{x}\hat{\beta}) + s(\mathbf{x}\beta_n)]^{-1} |(S^{1/2}\hat{\beta} + S^{1/2}\beta_n)' S^{1/2}(\hat{\beta} - \beta_n)| \\ &\leq (|S^{1/2}\hat{\beta}|_2 + |S^{1/2}\beta_n|_2)^{-1} |S^{1/2}\hat{\beta} + S^{1/2}\beta_n|_2 |S^{1/2}|_2 |\hat{\beta} - \beta_n|_2 \\ &\leq |S^{1/2}|_2 |\hat{\beta} - \beta_n|_2 = O(1) O_P(|\hat{\beta} - \beta_n|_2). \end{aligned}$$

□

在下面引理 7 的证明中，令 $(I, J) \subseteq \{1, \ldots, n\}$ 为大小为 2 的随机排序简单随机样本，并令 $\vec{D}$ （或 $\vec{D}^{\perp w}$， w一个 n维向量）为表示配对差 $\vec{x}_I - \vec{x}_J$ （或 $\vec{x}_I^{\perp w} - \vec{x}_J^{\perp w}$）的 $1 \times p$ 随机向量。于是

$$\mathrm{s.e.}_r^2(I, J) = \vec{D}^{\perp \mathbf{x}\hat{\beta}} \hat{C} \vec{D}^{\perp \mathbf{x}\hat{\beta}\prime} \tag{40}$$

由于协方差的 U-统计量表示， $\mathrm{Cov}(\vec{D}) = 2S^{(x)}$。由 (I, J) 分布的对称性， $\mathrm{E}\left(\vec{D}\right) = 0$， 因此 $\mathrm{E}\left(\vec{D}'\vec{D}\right) = \mathrm{Cov}(D) = 2S^{(x)}$。$(I, J)$ 的选取不考虑处理组和对照组之间的区别，因此 $\vec{D}$与 $\{Z_i\}_{i=1}^n$ 独立，进而与 $\hat{\ } \beta$和 $\hat{\ }C$独立。因此，其条件矩和边际矩相同： $\mathrm{E}\left(\vec{D}|\hat{\beta}, \hat{C}\right) = \mathrm{E}\left(\vec{D}\right) = 0$； $\mathrm{E}\left(\vec{D}'\vec{D}|\hat{\beta}, \hat{C}\right) = \mathrm{Cov}(\vec{D}|\hat{\beta}, \hat{C}) = \mathrm{Cov}(\vec{D}) = 2S^{(x)}$。

引理 7。 令 $S = S^{(x)}$ *or* $S^{\perp v} = (n - L)^{-1}\mathbf{x}^{\perp v\prime}\mathbf{x}^{\perp v}$ 为某个具有 L 个类别的分类变量 v。那么 $\langle S^{\perp\hat{\beta}}, \hat{C}\rangle_F = \langle S, \hat{C}\rangle_F - s^{-2}(\mathbf{x}\hat{\ }\beta)\langle S\hat{\ }\beta\hat{\ }\beta' S、\hat{C}\rangle_F$ 和 $\langle S^{\perp\beta_n}, C\rangle_F = \langle S, \hat{C}\rangle_F - s^{-2}(\mathbf{x}\beta_n)\langle S\beta_n\beta_n' S, C\rangle_F$。

证明。 如果 $S = S^{(x)}$， 令 $\vec{D}$ 如上所定义。否则，如果 $S = S^{\perp v}$， 则令 $(I, J) \subseteq \{1, \ldots, n\}$为从 v 中至少包含两个元素的某个类别里抽取的、容量为 2 的随机排序分层简单随机样本；抽取时，先按与类别大小成比例的概率选定其中一个类别，并令 $\vec{D}$ （或 $\vec{D}^{\perp w}$， w 一个

n-向量）为表示配对差 $\vec{x}_I - \vec{x}_J$（或$\vec{x}_I^{\perp w} - \vec{x}_J^{\perp w}$）的一个随机向量。

利用样本协方差的 U 统计量表示法，计算 $\mathrm{E}\left(\vec{D}^{\perp \mathbf{x}\hat{\beta}\prime}\vec{D}^{\perp \mathbf{x}\hat{\beta}}\middle|\hat{\beta},\hat{C}\right) = \mathrm{Cov}(\vec{D}^{\perp \mathbf{x}\hat{\beta}}|\hat{\beta},\hat{C})$，得到 $\mathrm{Cov}(\vec{D}^{\perp \mathbf{x}\hat{\beta}}|\hat{\beta},\hat{C}) = 2(n-1)^{-1}\mathbf{x}^{\perp \mathbf{x}\hat{\beta}\prime}\mathbf{x}^{\perp \mathbf{x}\hat{\beta}}$。现在与下式比较：

$$\begin{aligned}\frac{1}{2}\mathrm{E}\left(\vec{D}^{\perp \mathbf{x}\hat{\beta}\prime}\vec{D}^{\perp \mathbf{x}\hat{\beta}}|\hat{\beta},\hat{C}\right) =& (I - \hat{\beta}\hat{\beta}'S^{(x)}/s^2(\mathbf{x}\hat{\beta}))'S^{(x)}(I - \hat{\beta}\hat{\beta}'S^{(x)}/s^2(\mathbf{x}\hat{\beta})) \\ =& S^{(x)} - 2s^{-2}(\mathbf{x}\hat{\beta})S^{(x)}\hat{\beta}\hat{\beta}'S^{(x)} + \\ & s^{-4}(\mathbf{x}\hat{\beta})S^{(x)}\hat{\beta}\cdot\hat{\beta}'S^{(x)}\hat{\beta}\cdot\hat{\beta}'S^{(x)} \\ =& S^{(x)} - s^{-2}(\mathbf{x}\hat{\beta})S^{(x)}\hat{\beta}\hat{\beta}'S^{(x)} =: S^{\perp\hat{\beta}}.\end{aligned}$$

□

命题 *5* 的证明。由 A6 可知 $\|S\|_F = \mathrm{tr}(S'S)^{1/2} = O(p^{1/2})$；由 $\hat{C}_n$ 的一致性假设可知 $\|\hat{C}_n - C_n\|_F = o_P(p^{1/2}/n)$。因此

$$|\langle S, \hat{C}_n\rangle_F - \langle S, C_n\rangle_F| = |\langle S, \hat{C}_n - C_n\rangle_F| = o_P(p/n).$$

根据引理 7,

$$\begin{aligned}\langle S^{\perp\hat{\beta}}, \hat{C}_n\rangle_F - \langle S^{\perp\beta_n}, C_n\rangle_F =& \langle S, \hat{C}_n - C_n\rangle_F \\ & + s^{-2}(\mathbf{x}\hat{\beta})\langle S\hat{\beta}\hat{\beta}'S, \hat{C}_n\rangle_F - s^{-2}(\mathbf{x}\beta_n)\langle S\beta_n\beta_n'S, C_n\rangle_F \\ =& o_P\left(\frac{p}{n}\right) + [s^{-2}(\mathbf{x}\hat{\beta}) - s^{-2}(\mathbf{x}\beta_n)]\langle S\hat{\beta}\hat{\beta}'S, \hat{C}_n\rangle_F \qquad (41) \\ & + s^{-2}(\mathbf{x}\beta_n)\Big(\langle S\hat{\beta}\hat{\beta}'S, \hat{C}_n\rangle_F - \langle S\beta_n\beta_n'S, C_n\rangle_F\Big). \qquad (42)\end{aligned}$$

要分析（41）式最右侧的加项，首先注意到 $s^2(\mathbf{x}\beta_n) = \beta_n'S\beta_n$，因此 $|s(\mathbf{x}\beta_n)| \le |S|_2|\beta_n|_2 = O(p^{1/2})$。结合命题1，$\hat{}$ 引理 6 得出 $O_P\ p^{1/2}$ $|s\ \ \beta_n| = O_P\ p^{1/2}$ $s\ \ \beta + s\ \ \beta_n = O_P\ p^{1/2}$

$(\mathbf{x}\beta)\,(\mathbf{x}\beta_n) = O_P(p/n)$。因此，A9 表明 $(\mathbf{x}\beta)\,(\,)$ 与 $(\mathbf{x}\,)\,(\,)$ 一样，故 $(\mathbf{x}\hat{})\,(\mathbf{x}\,)\,(\,)$。综合这些事实，$s^2(\mathbf{x}\hat{}\beta) - s^2(\mathbf{x}\beta_n) = [s(\mathbf{x}\hat{}\beta) - s(\mathbf{x}\beta_n)][s(\mathbf{x}\hat{}\beta) + s(\mathbf{x}\beta_n)] = O_P(p/n^{1/2})$。根据 A8，可得 $|s^{-2}(\mathbf{x}\hat{}\beta) - s^{-2}(\mathbf{x}\beta_n)| = O_P\{p/[n^{1/2}s^4(\mathbf{x}\beta_n)]\}$。

关于 $\langle S \quad \hat{\beta}\hat{\beta}'S, \hat{C}_n\rangle_F$ 式 (41) 的因子，

$$\begin{aligned}|\langle S\hat{\beta}\hat{\beta}'S, \hat{C}_n\rangle_F| &= \mathrm{tr}(S\hat{\beta}\hat{\beta}'S\hat{C}_n) \\ &= |\,\mathrm{tr}(\hat{\beta}'S\hat{C}_nS\hat{\beta})| \\ &= \hat{\beta}'S\hat{C}_nS\hat{\beta} \le \|\hat{\beta}'S^{1/2}\|_2\|S^{1/2}\|_2\|\hat{C}_n\|_2\|S^{1/2}\|_2\|S^{1/2}\hat{\beta}\|_2 \\ &= O_P(\frac{s^2(\mathbf{x}\beta_n)}{n}),\end{aligned}$$

ˆ 根据 A8、A6、 C_n 的一致性以及引理 1。现在我们有

$$|[s^{-2}(\mathbf{x}\hat{\beta}) - s^{-2}(\mathbf{x}\beta_n)]\langle S\hat{\beta}\hat{\beta}'S, \hat{C}_n\rangle_F| = O_P\left(\frac{p}{n^{3/2}s^2(\mathbf{x}\beta_n)}\right) = o_P\left(\frac{p}{n}\right).$$

为界定 (42)，根据 A8，我们关注第二个因子：

$$|\langle S\hat{\beta}\hat{\beta}'S, \hat{C}_n\rangle_F - \langle S\beta_n\beta_n{}'S, C_n\rangle_F| \leq \\ |\langle S\hat{\beta}\hat{\beta}'S - S\beta_n\beta_n{}'S, \hat{C}_n\rangle_F| + |\langle S\beta_n\beta_n{}'S, \hat{C}_n - C_n\rangle_F|. \quad (43)$$

我们利用柯西-施瓦茨不等式处理 (43) 中的左侧项。根据引理 1 以及 ˆC_n，$\|\hat{C}_n\|_2 = O_P(n^{-1})$ 的一致性，可得 $\|\hat{C}_n\|_F = O_P(p^{1/2}/n)$。由于 $S\hat{\beta}\hat{\beta}'S - S\beta_n\beta_n{}'S = S(\hat{\beta}\hat{\beta}' - \beta_n\beta_n{}')S$，因此 $\|S\hat{\beta}\hat{\beta}'S - S\beta_n\beta_n{}'S\|_F \leq \|S^{1/2}\|_2\|S^{1/2}\hat{\beta}\hat{\beta}'S^{1/2} - S^{1/2}\beta_n\beta_n{}'S^{1/2}\|_F\|S^{1/2}\|_2 = O_P(1)O_P\Big[\|S^{1/2}\hat{\beta}\hat{\beta}'S^{1/2} - S^{1/2}\beta_n\beta_n{}'S^{1/2}\|_F\Big]O_P(1)$。这可分解为

$$\|S^{1/2}\hat{\beta}\hat{\beta}'S^{1/2} - S^{1/2}\beta_n\beta_n{}'S^{1/2}\|_F \leq \|S^{1/2}(\hat{\beta} - \beta_n)(\hat{\beta} + \beta_n)'S^{1/2}\|_F + \\ \|S^{1/2}\hat{\beta}\beta_n'S^{1/2} - S^{1/2}\beta_n\hat{\beta}'S^{1/2}\|_F$$

其中：

$$\begin{aligned}
\|S^{1/2}(\hat{\beta} - \beta_n)(\hat{\beta} + \beta_n)'S^{1/2}\|_F \leq & \|S^{1/2}\|_2\|(\hat{\beta} - \beta_n)(\hat{\beta} + \beta_n)'S^{1/2}\|_F \\
&= O_P(1)\,\mathrm{tr}[S^{1/2}(\hat{\beta} + \beta_n)(\hat{\beta} - \beta_n)'(\hat{\beta} - \beta_n)(\hat{\beta} + \beta_n)'S^{1/2}]^{1/2} \\
&= O_P(1)\,\mathrm{tr}[(\hat{\beta} + \beta_n)'S(\hat{\beta} + \beta_n)(\hat{\beta} - \beta_n)'(\hat{\beta} - \beta_n)]^{1/2} \\
&= O_P(1)[(\hat{\beta} + \beta_n)'S(\hat{\beta} + \beta_n)\|\hat{\beta} - \beta_n\|_2^2]^{1/2} \\
&= O_P(1)[(\hat{\beta} + \beta_n)'S(\hat{\beta} + \beta_n)]^{1/2}O_P[(\frac{p}{n})^{1/2}] \\
&= O_P[(\frac{p}{n})^{1/2}][(\hat{\beta} - \beta_n + 2\beta_n)'S(\hat{\beta} - \beta_n + 2\beta_n)]^{1/2} \\
&= O_P[(\frac{p}{n})^{1/2}][\|S^{1/2}(\hat{\beta} - \beta_n)\|_2^2 + 4(\hat{\beta} - \beta_n)'S\beta_n + 4\|S^{1/2}\beta_n\|_2^2]^{1/2} \\
&= O_P[(\frac{p}{n})^{1/2}]\left\{O_P\left(\frac{p}{n}\right) + O_P\left[\left(\frac{p}{n}\right)^{1/2}s(\mathbf{x}\beta_n)\right] + O_P[s^2(\mathbf{x}\beta_n)]\right\}^{1/2} \\
&= O_P\left[\left(\frac{p}{n}\right)^{1/2}\right]O_P\left\{\left[\left(\frac{p}{n}\right)^{1/2} + s(\mathbf{x}\beta_n)\right]^{2\cdot 1/2}\right\} \\
&= O_P\left\{\max\left[\frac{p}{n}, \left(\frac{p}{n}\right)^{1/2}s(\mathbf{x}\beta_n)\right]\right\};
\end{aligned}$$

and

$$\begin{aligned}\|S^{1/2}\hat{\beta}\beta_n' S^{1/2} - S^{1/2}\beta_n\hat{\beta}' S^{1/2}\|_F =& \|S^{1/2}(\hat{\beta}-\beta_n)\beta_n' S^{1/2} - S^{1/2}\beta_n(\hat{\beta}-\beta_n)' S^{1/2}\|_F \\ \leq& \|S^{1/2}(\hat{\beta}-\beta_n)\beta_n' S^{1/2}\|_F + \|S^{1/2}\beta_n(\hat{\beta}-\beta_n)' S^{1/2}\|_F \\ =& \, 2\,\mathrm{tr}[S^{1/2}\beta_n(\hat{\beta}-\beta_n)' S(\hat{\beta}-\beta_n)\beta_n' S^{1/2}]^{1/2} \\ =& \, 2[(\hat{\beta}-\beta_n)' S(\hat{\beta}-\beta_n)\beta_n' S\beta_n]^{1/2} \\ =& \, O_P\left[\left(\frac{p}{n}\right)^{1/2} s(\mathbf{x}\beta_n)\right].\end{aligned}$$

因此，式 (43) 右侧的左项为 $O_P\{\max[p/n, (p/n)^{1/2}s(\mathbf{x}\beta_n)]\}$。

利用 $\langle\cdot,\cdot\rangle_F$ 的定义和迹的循环性质，(43) 中的剩余项可作如下界定:

$$\begin{aligned}\langle S\beta_n\beta_n' S, \hat{C}_n - C_n\rangle_F =& \,\mathrm{tr}[S\beta_n\beta_n' S(\hat{C}_n - C_n)] \\ =& \,\mathrm{tr}[\beta_n' S(\hat{C}_n - C_n)S\beta_n] \\ =& \beta_n' S(\hat{C}_n - C_n)S\beta_n \\ \leq& |\beta_n|_2^2 \|S\|_2^2 \|\hat{C}_n - C_n\|_2 \\ =& \, O_P(p)O_P(1)o_P(n^{-1}) = o_P(p/n),\end{aligned}$$

利用 A7、A6 以及 $\hat{}\,C_n$ 的一致性。由此可知，(43) 整体为 $O_P\{\max[p/n, (p/n)^{1/2}s(\mathbf{x}\beta_n)]\}$，进而可得 (42) 为 $O_P\{\max[(p/n)s^{-2}(\mathbf{x}\beta_n),\ (p/n)^{1/2}s^{-1}(\mathbf{x}\beta_n)]\} = o_P(p/n)$，从而完成了对该命题一致性结论的证明。该命题的其余部分由 A6 和引理 1 推出。

□

C 第4节的证明

C.1 第4.2节的证明

命题 *6* 的证明。 对于 (i) 和 (iii)，只需证明 $\sup_i |1 - |\vec{x}_i\hat{}\,C^{1/2}|_2^2/|\vec{x}_i C^{1/2}|_2^2|$ and $\sup_{i,j}|1 - |(\vec{x}_i - \vec{x}_j)\hat{}\,C^{1/2}|_2^2/|(\vec{x}_i - \vec{x}_j)C^{1/2}|_2^2|$ 依概率收敛到 0（根据依概率收敛的定义），并且只考虑满足 $\vec{x}_i \neq 0$ and $\vec{x}_i \neq \vec{x}_j$ 的 (i, j)。后者保证 $|\vec{x}_i C^{1/2}|_2 > 0$ and $|(\vec{x}_i - \vec{x}_j)C^{1/2}|_2 > 0$，因为 C_n 满秩。事实上，附录中的引理 3 给出 $|A_n|_2 = O(1)$，涵盖情形 (i)；对于情形 (iii)，A12 给出 $|B_n^{-1}|_2 = O(1)$，进而得到 $|A_n B_n^{-1} A_n|_2 = O(1)$。请注意，这不仅表明 C_n 满秩，还表明 $|C_n^{-1}|_2 = O(1)$。

对于任意非零的 p 向量 v，我们有

$$1 - \frac{v'\hat{C}_n v}{v'C_n v} = \frac{v'(C_n - \hat{C}_n)v}{v'v}\frac{v'v}{v'C_n v}$$

因此

$$\left|1 - \frac{v'\hat{C}_n v}{v'C_n v}\right| \leq \left|\frac{v'(C_n - \hat{C}_n)v}{v'v}\frac{v'C_n^{-1}v}{v'v}\right| \leq |C_n - \hat{C}_n|_2 |C_n^{-1}|_2,$$

利用正定矩阵的一般不等式（例如，Gentle，2007，§ 8.4）。命题 4 给出 $|\hat{} C_n - C_n|_2 = o_P(n^{-1})$，从而完成了第 (i) 部分和第 (iii) 部分的证明。

对于第 (ii) 部分，设 $a \vee b := \max(a, b)$，且 $a \wedge b := \min(a, b)$。对于任意非零的 p-向量 v

$$\left|\frac{v'C_n v \vee \epsilon_n - v'\hat{C}_n v \vee \epsilon_n}{v'v}\right| \leq \left|\frac{v'C_n v - v'\hat{C}_n v}{v'v}\right|, \text{ since}$$
$$\left|v'C_n v \vee \epsilon_n - v'\hat{C}_n v \vee \epsilon_n\right| \leq \left|v'C_n v - v'\hat{C}_n v\right|.$$

（后者成立是因为，当 $v'C_n v$ 和 $v'\hat{} C_n v$ 属于同一个半区间 $(0, \epsilon_n]$ 或 $[\epsilon_n, \infty)$ 时，不等式两边的表达式相等；而若它们被 ϵ_n 分隔开，则左侧表达式较小。）因此

$$\left|\frac{v'C_n v \vee \epsilon_n - v'\hat{C}_n v \vee \epsilon_n}{v'v}\right| \leq |C_n - \hat{C}_n|_2 = o_P(n^{-1}). \tag{44}$$

根据 A11，$\sup_i \vec{x}_i\vec{x}_i' = O[\max(p, \log\ n)]$。因此

$$\sup_i \frac{\vec{x}_i\vec{x}_i'}{\epsilon_n}, \sup_{\{i,j\}\in\mathcal{E}_n} \frac{(\vec{x}_i - \vec{x}_j)(\vec{x}_i - \vec{x}_j)'}{\epsilon_n} = O_P(n).$$

Thus

$$\sup_i \frac{\vec{x}_i\vec{x}_i'}{\vec{x}_i C_n \vec{x}_i' \vee \epsilon_n}, \sup_{\{i,j\}\in\mathcal{E}_n} \frac{(\vec{x}_i - \vec{x}_j)(\vec{x}_i - \vec{x}_j)'}{(\vec{x}_i - \vec{x}_j)C_n(\vec{x}_i - \vec{x}_j)' \vee \epsilon_n} = O_P(n).$$

由（44）可得，

$$\sup_{i:|\vec{x}_i|_2\neq 0} \frac{\vec{x}_i C_n \vec{x}_i' \vee \epsilon_n - \vec{x}_i\hat{C}_n\vec{x}_i' \vee \epsilon_n}{\vec{x}_i\vec{x}_i'} \frac{\vec{x}_i\vec{x}_i'}{\vec{x}_i C_n \vec{x}_i' \vee \epsilon_n} = o_P(1),$$

同理

$$\sup_{\substack{\{i,j\}\in\mathcal{E}_n:\\ \vec{x}_i\neq\vec{x}_j}} \frac{(\vec{x}_i-\vec{x}_j)C_n(\vec{x}_i-\vec{x}_j)'\vee\epsilon_n-(\vec{x}_i-\vec{x}_j)\hat{C}_n(\vec{x}_i-\vec{x}_j)'\vee\epsilon_n}{(\vec{x}_i-\vec{x}_j)(\vec{x}_i-\vec{x}_j)'}\times$$

$$\frac{(\vec{x}_i-\vec{x}_j)(\vec{x}_i-\vec{x}_j)'}{(\vec{x}_i-\vec{x}_j)C_n(\vec{x}_i-\vec{x}_j)'\vee\epsilon_n}=o_P(1).$$

□

命题 7 的证明。关于 $\{\vec{X}_i$ 的假设：$i\le n\}$ 意味着 $(\vec{X}_i-\vec{X}_j)$ 服从 MVN(0, 2Σ), 且 $(\vec{X}_i-\vec{X}_j)C^{1/2}$ 服从 MVN(0, $2C^{1/2}\Sigma C^{1/2}$)。设 $2C^{1/2}\Sigma C^{1/2}$ 的特征分解为 $Q'\Lambda Q$，其中 Λ 是非负实对角矩阵，Q 是正交矩阵。于是 $(\vec{X}_i-\vec{X}_j)$ $C^{1/2}Q'\sim$服从 MVN(0, Λ)。令 $W_{ij1},\ldots,W_{ijp}$ 表示 $(\vec{X}_i-\vec{X}_j)C^{1/2}Q'=$的 p 个坐标：$\vec{W}_{ij}$，可知 $W_{ij1},\ldots,W_{ijp}$ 是相互独立的零均值高斯变量，其方差为 $v_1,\ldots,$ v_p，即 Λ 的对角元素，也是 $2C^{1/2}\Sigma C^{1/2}$ 的特征值，而 $(\vec{X}_i-\vec{X}_j)C(\vec{X}_i-\vec{X}_j)$ $'=(\vec{X}_i-\vec{X}_j)C^{1/2}Q'QC^{1/2}(\vec{X}_i-\vec{X}_j)'=\vec{W}_{ij}\vec{W}'_{ij}$。

显然，对于任意 $i\neq j$ E $\vec{W}_{ij}\vec{W}'_{ij}=\sum_i v_i$，或 tr($2C^{1/2}\Sigma C^{1/2}$) $=2\langle C,\Sigma\rangle_F$。接下来，我们刻画 $\vec{W}_{ij}\vec{W}'_{ij}-$E $\vec{W}_{ij}\vec{W}'_{ij}$ 的矩母函数（MGF），或 $\sum_{k=1}^p(W_{ijk}^2-$ E $W_{ijk}^2)$。具有 MGF $e^{-t}(1-2t)^{-1/2}$ 的中心化 χ_1^2 分布，在 $t<1/2$, W_{ijk}^2- E W_{ijk}^2 时成立，其对数矩母函数为 $(1/2)[-2v_kt-\log(1-2v_kt)]$,，在 $t<1/(2v_k)$ 时成立。应用关系式 $-\log(1-x)-x\le x^2/[2(1-x)]$，即对 $\log(1-x)$ 的二阶泰勒展开进行变换，我们现在得到

$$\begin{aligned}
\log \mathrm{E}\{\exp t[W_{ijk}^2-\mathrm{E}\,W_{ijk}^2]\}\le\quad & \tfrac{(v_kt)^2}{1-2v_kt} && \big(\text{for } t<\frac{1}{2v_k}\big)\\
&\le \tfrac{(v_kt)^2}{1-2(\max_k v_k)t} && \big(t<\frac{1}{2\max_k v_k}\big);\ \text{so}\\
\log \mathrm{E}\exp\{t\sum_{k=1}^p[W_{ijk}^2-\mathrm{E}\,W_{ijk}^2)]\}\ &\le \tfrac{t^2\sum_k v_k^2}{1-2(\max_k v_k)t}, && t<\frac{1}{2\max_k v_k}
\end{aligned}$$

这是由于不相关高斯变量的独立性。也就是说，$\sum_{k=1}^p W_{ijk}^2-$E W_{ijk}^2 在其右尾上服从次伽马分布，方差因子为 $\sum_k v_k^2$，尺度因子为 $\max_k v_k$；用符号表示为 $\sum_{k=1}^p W_{ijk}^2-$E $W_{ijk}^2\in\Gamma_+(\sum_k v_k^2,\max_k v_k)$（Boucheron 等，2013，第 2.4 节）。由于 $\sum_k v_k^2=$tr(Λ^2), 迹的循环性质结合上述定义，可将该方差因子化简为 $4|C^{1/2}\Sigma C^{1/2}|_F^2$，因为 tr($Q'\Lambda QQ'\Lambda Q$) 4 $C^{1/2}\Sigma C^{1/2}C^{1/2}\Sigma C^{1/2}$ $2|C^{1/2}\Sigma C^{1/2}|_2$ tr( ) ;

尺度因子则化简为，因为 $\max_k v_k=|\Lambda|_2$ 且 $|\Lambda|_2=|Q'\Lambda Q|_2$。

若要用更熟悉的量来刻画矩母函数，请注意，$|2C^{1/2}\Sigma C^{1/2}|_F^2 = \sum_{i=1}^p v_i^2 \le (\max_k v_k)(\sum_k v_k) = 2|C^{1/2}\Sigma C^{1/2}|_2(\sum_k v_k)$，而 $\sum_k v_k = \operatorname{tr}(D^2) = \operatorname{tr}(2C^{1/2}\Sigma C^{1/2}) = 2\langle \Sigma, C\rangle_F$。因此，方差因子可取为 $4|C^{1/2}\Sigma C^{1/2}|_2|S^{1/2}C^{1/2}|_F^2$，或 $4|C^{1/2}\Sigma C^{1/2}|_2\langle \Sigma, C\rangle_F$。

这些矩母函数刻画给出了 $\{\vec{W}_{ij}\vec{W}'_{ij} - \mathrm{E}\,\vec{W}_{ij}\vec{W}'_{ij} \colon i \ne j \le n\}$ 上确界的对照组。由于该类包含 $\binom{n}{2}$ 个不同的 $\Gamma_+\big[4|C^{1/2}\Sigma C^{1/2}|_2\langle \Sigma, C\rangle_F, 2|C^{1/2}\Sigma C^{1/2}|_2\big]$ 随机变量，根据 Boucheron 等人（2013）的推论 2.6，可得

$$\mathrm{E}\max_{i\ne j\le n}[\vec{W}_{ij}\vec{W}'_{ij} - \mathrm{E}(\vec{W}_{ij}\vec{W}'_{ij})] \le 2\left[2|C^{1/2}\Sigma C^{1/2}|_2\langle \Sigma, C\rangle_F \log\binom{n}{2}\right]^{1/2} + 2|C^{1/2}\Sigma C^{1/2}|_2 \log\binom{n}{2}.$$

利用 $\mathrm{E}\,\vec{W}_{ij}\vec{W}'_{ij} = 2\langle \Sigma, C\rangle_F$、$i \ne j$ 和 $2a^2+2ab+b^2 = a^2[1+(1+b/a)^2]$，化简

$$\begin{aligned}\mathrm{E}\max_{i\ne j\le n}\vec{W}_{ij}\vec{W}'_{ij} &\le \langle \Sigma, C\rangle_F\left\{1+\left[1+\left(\frac{2|C^{1/2}\Sigma C^{1/2}|_2\log\binom{n}{2}}{\langle \Sigma, C\rangle_F}\right)^{1/2}\right]^2\right\}\\ &= \langle \Sigma, C\rangle_F\left\{1+\left[1+\left(\frac{2\log\binom{n}{2}}{p_{[C^{1/2}\Sigma C^{1/2}]}}\right)^{1/2}\right]^2\right\}. \qquad (45)\end{aligned}$$

接下来注意，在正实数轴上，$x \mapsto \sqrt{2+x}$ 支配 $x \mapsto [1+(1+x)^2]^{1/2}$。（这两个函数在 $x = 0$ 处相等；否则，后者的导数等于 $(1+x)^2/[1+(1+x)^2]$，的平方根，该值处处不大于 1。）因此，由 (45) 得

$$\begin{aligned}\left(\mathrm{E}\max_{i\ne j\le n}\vec{W}_{ij}\vec{W}'_{ij}\right)^{1/2} &\le \langle \Sigma, C\rangle_F^{1/2}\left[\sqrt{2}+\left(\frac{2\log\binom{n}{2}}{p_{[C^{1/2}\Sigma C^{1/2}]}}\right)^{1/2}\right]\\ &= \sqrt{2}\langle \Sigma, C\rangle_F^{1/2}\left[1+\left(\frac{\log\binom{n}{2}}{p_{[C^{1/2}\Sigma C^{1/2}]}}\right)^{1/2}\right].\end{aligned}$$

□

D 第 5 节

命题 *9* 的证明。在整个证明中，期望值均按给定 $\mathcal{F}_n$ 的条件期望理解。（由于每个 $\mathcal{F}_n$ 都是由有限个离散随机变量生成的 σ-代数，因此不会引入此前尚未涉及的测度论考量。）

此外，假设采用矩条件的版本 (ii)，注意到它可由版本 (i) 和 $0 < \mathrm{P}(Z = 1) < 1$ 推出（后者又可由 $0 < \mathrm{P}(Z = 1 \mid \vec{X}\beta_n) < 1$ 得出）：$\mathrm{E}\,|Y_T|^p < \infty$ 表示 $\mathrm{E}(|Y_T|^p \mid Z) < \infty$，同时蕴含 $\mathrm{E}\,|V(n,\mathbf{s})|^p < \infty$。若 $n_{\mathbf{s}}$ 有界，则对于细层 $\mathbf{s}$，可实现的配置 $(\sum_{i\in\mathbf{s}}[\![Z_i = 0]\!], \sum_{i\in\mathbf{s}}[\![Z_i = 1]\!])$ 只有有限种；每种配置都有一个特征分布函数 $x \mapsto \mathrm{P}(V(n,\mathbf{s}) \le x)$，使得 $\mathrm{E}\left(|V(n,\mathbf{s})|^p\right) = p\int_0^\infty y^{p-1}\,\mathrm{P}(|V(n,\mathbf{s})| > y)dy < \infty$（例如，参见 Durrett，2019，引理 2.2.13）。将 $|V|$ 在 $\Re^+$ 上的分布定义为：令 $\mathrm{P}(|V| > y)$ 等于这些分布中 $\mathrm{P}(V^{(n,\mathbf{s})} > y)$ 的最大值，则该分布随机占优相关的 $V^{(n,\mathbf{s})}$ 分布。它还满足 $p\int_0^\infty y^{p-1}\,\mathrm{P}(|V| > y)dy < \infty$，因为有限个函数 $y \mapsto \mathrm{P}(|V^{(n,\mathbf{s})}| > y$ 之和不小于它们的最大值；因此 $V \in L_p$。

命题第 i 部分。 由 (26)，

$$\begin{aligned}
\sum_{i\in\mathbf{s}} \psi_{\mathbf{s}i}(\eta) &= w_{\mathbf{s}}[Y_i - \eta(Z_i - \bar{Z}_{\mathbf{s}})](Z_i - \bar{Z}_{\mathbf{s}}) \\
&= w_{\mathbf{s}}\{\sum_{\substack{i\in\mathbf{s}:\\ Z_i=1}} [Y_i(1-\bar{Z}_{\mathbf{s}}) - \eta(1-\bar{Z}_{\mathbf{s}})^2] + \sum_{\substack{i\in\mathbf{s}:\\ Z_i=0}} [-Y_i\bar{Z}_{\mathbf{s}} - \eta\bar{Z}_{\mathbf{s}}^2]\} \\
&= w_{\mathbf{s}}[(1-\bar{Z}_{\mathbf{s}}) \sum_{\substack{i\in\mathbf{s}:\\ Z_i=1}} Y_i - \eta n_{\mathbf{s}}\bar{Z}_{\mathbf{s}}(1-\bar{Z}_{\mathbf{s}})^2 - \bar{Z}_{\mathbf{s}} \sum_{\substack{i\in\mathbf{s}:\\ Z_i=0}} Y_i - \eta n_{\mathbf{s}}(1-\bar{Z}_{\mathbf{s}})\bar{Z}_{\mathbf{s}}^2] \\
&= w_{\mathbf{s}}n_{\mathbf{s}}\bar{Z}_{\mathbf{s}}(1-\bar{Z}_{\mathbf{s}})\left[\bar{Y}_{\mathbf{s}(1)} - \eta(1-\bar{Z}_{\mathbf{s}}) - \bar{Y}_{\mathbf{s}(0)} - \eta\bar{Z}_{\mathbf{s}}\right] \\
&= w_{\mathbf{s}}n_{\mathbf{s}}\bar{Z}_{\mathbf{s}}(1-\bar{Z}_{\mathbf{s}})(\bar{Y}_{\mathbf{s}(1)} - \bar{Y}_{\mathbf{s}(0)} - \eta), \qquad (46)
\end{aligned}$$

因此 $\mathrm{E}[\sum_{i\in\mathbf{s}} \psi_{\mathbf{s}i}(_\ _\ _\ \eta\ _\ _\ _\ =)$ $] = w_{\mathbf{s}}n_{\mathbf{s}}\bar{z}_{\mathbf{s}}(1-\bar{z}_{\mathbf{s}})\,\mathrm{E}(\bar{Y}_{\mathbf{s}(1)} - \bar{Y}_{\mathbf{s}(0)} - \eta)$。回想 $V(n,\mathbf{s})Y_{\mathbf{s}(1)} - Y_{\mathbf{s}(0)} - \mathrm{E}(Y_{\mathbf{s}(1)} - Y_{\mathbf{s}(0)})$，我们有

$$\psi_{\mathcal{S}_n}(\eta) - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)] = \frac{\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{n,\mathbf{s}}n_{\mathbf{s}}V^{(n,\mathbf{s})}}{\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{n,\mathbf{s}}n_{\mathbf{s}}} \qquad (47)$$

其中 $\tilde{w}_{n,\ \mathbf{s}} := w_{\mathbf{s}}\bar{z}_{\mathbf{s}}(1-\bar{z}_{\mathbf{s})} \in \mathcal{F}_n$。（若无歧义，“$\tilde{w}_{n,\ \mathbf{s}}$”可缩写为“$\tilde{w}_{\mathbf{s}}$。”）可以看出，$\psi_{\mathcal{S}_n}(\eta) - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)|\mathcal{G}_n]$ 的收敛将由 (47) 的适当收敛推出，即 $\psi_{\mathcal{S}_n}(\eta) - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)|\mathcal{F}_n]$。

依概率收敛于 $\psi_{\mathcal{S}_n}(\eta) - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)]$。我们将通过截断增量的方法证明 L_1-弱大数定律的论证推广到独立不同分布情形。按照 Durrett（2019，§ 2.2.3），令 $\bar{V}_{(\mathbf{s})n} := V_{(n,\mathbf{s})}[\![w_{n,\mathbf{s}}|V_{(n,\mathbf{s})}| \le m_n]\!]$ 以及 $\bar{V}_n := V[\![|V| \le m_n]\!]$。（回想 $m_n =$

$\sum_{\mathbf{s}\in\mathcal{S}_n}[\![w_{\mathbf{s}}\bar{z}_{\mathbf{s}}(1-\bar{z}_{\mathbf{s}})>0]\!]$。) 根据定义，$\mathrm{E}\ V_{(n,\mathbf{s})}=0$；而 $\mathrm{E}\ \bar{V}_{(\mathbf{s})n}$ 可能不等于 0。我们的首要任务是证明 $(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{n,\mathbf{s}})^{-1}\,\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{n,\mathbf{s}}\bar{V}_{(\mathbf{s})n}\ \to 0$。我们有

$$(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}})^{-1}\,\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\bar{V}_n^{(\mathbf{s})}=(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}})^{-1}\left(\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\bar{V}_{n+}^{(\mathbf{s})}+\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\bar{V}_{n-}^{(\mathbf{s})}\right)$$

其中，a_+ 和 a_- 分别表示 a的正部和负部，max(a,0)和min(a,0)。现在

$$\begin{aligned}(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}})^{-1}\,\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\bar{V}_{n+}^{(\mathbf{s})}&=(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}})^{-1}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\,\mathrm{E}\,\bar{V}_{n+}^{(\mathbf{s})}\\&=(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}})^{-1}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\int_0^\infty\mathrm{P}(\bar{V}_n^{(\mathbf{s})}>x)dx\\&=\int_0^\infty\frac{\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\,\mathrm{P}(\bar{V}_n^{(\mathbf{s})}>x)}{\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}}dx.\end{aligned}\tag{48}$$

根据随机序假设，P($V^{(n,\mathbf{s})}>x$)被P($V>x$)控制；但P($\bar{V}$(s)$_n\ >\ x)\ \le$P(V($_n$,s)$>\ x$)，因此(48)中的被积函数也被P($V>x$)控制。由于 $\int_0^\infty$P($V>x$)$dx=\mathrm{E}\,V_+<\infty$，控制收敛定理给出$(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}})^{-1}\,\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\bar{V}(s)_{n+}\ -\ (\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}})^{-1}\,\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}V$($_n$,s)$_+\ \ =o$(1)。

类似地，$(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}})^{-1}\,\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\bar{V}_{(\mathbf{s})\,n-}-(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}})^{-1}\,\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}V_{(n,\mathbf{s})_-}\ \ =o$(1)，并且

$$\left(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\right)^{-1}\left(\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\bar{V}_n^{(\mathbf{s})}-\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}V^{(n,\mathbf{s})}\right)=o(1).$$

由于 $\mathrm{E}\ V_{(n,\ \mathbf{s})}=0$，这意味着 $(\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}})^{-1}\,\mathrm{E}\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}\bar{V}_{(\mathbf{s})n}\ \to 0$。

要证明式（47）依概率收敛，现在只需证明

$$\frac{\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}(V^{(n,\mathbf{s})}-\mathrm{E}\,\bar{V}_n^{(\mathbf{s})})}{\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}}=o_P(1),\tag{49}$$

三角阵弱大数定律的结论。现在，我们将验证该原理的前提条件，该原理见于 Durrett（2019）的定理 2.2.11。对于每个 $n\{V_{(n,\mathbf{s})}:\ \mathbf{s}\in\mathcal{S}_n\}$ 相互独立，因为 $\{(Y_{Ti},Y_{Ci},Z_i):\ i\}$ 无条件独立，而对 $\mathbb{F}_n$ 进行条件化会导致层内相关，但不同层之间不相关s。定理的前提条件 (i)，$\sum_{\mathbf{s}\in\mathcal{S}_n}$P($\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}V_{(n,\mathbf{s})}>\sum_{\mathbf{s}\in\mathcal{S}_n}\tilde{w}_{\mathbf{s}}n_{\mathbf{s}}$)$\ \to\ 0$,

根据随机占优和 $\tilde{w}_{\mathbf{s}} n_{\mathbf{s}}$ 有界性的假设，将由以下量收敛至 0 得出

$$\sum_{\mathbf{s}\in\mathcal{S}_n:\tilde{w}_{\mathbf{s}} n_{\mathbf{s}}>0} \mathrm{P}(u_{\tilde{w}} V > \sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}) = m_n \,\mathrm{P}(u_{\tilde{w}} V > \sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}), \tag{50}$$

其中 $u_{\tilde{w}}$ 是 $\{\tilde{w}_{\mathbf{s}} n_{\mathbf{s}}$ 的上界： n； $\mathbf{s} \in \mathcal{S}_n\}$。根据假设 $m_n/(\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}})$ =O_P(1)，因此（50）=o_P(1)可由 $(\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}})$ P($u_{\tilde{w}} V > \sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}$) $\to 0$推出。由于我们还假设当 $n \uparrow \infty$时 $m_n \to \infty$，因此随着n 增大，$\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}} \to \infty$ 也随之增大；若当 $x \uparrow \infty$时 x P($u_{\tilde{w}} V > x$) $\to 0$，则（50）收敛于 0。根据控制收敛定理，这一点成立，因为当 $x \uparrow \infty$时， x P($u_{\tilde{w}} V > x$) $\leq$ E($u_{\tilde{w}} V[\![u_{\tilde{w}} V > x]\!]$,$u_{\tilde{w}} V[\![u_{\tilde{w}} V > x]\!] \to 0$ 几乎处处成立，且 E ($u_{\tilde{w}} V$) $< \infty$。

Durrett（2019）定理 2.2.11 的前提 (ii) 是

$$(\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}})^{-2} \sum n_{\mathbf{s}}^2 \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}^2 \,\mathrm{E}\left[(\bar{V}_n^{(\mathbf{s})})^2\right] \to 0. \tag{51}$$

为验证这一点，首先注意到

$$\begin{aligned}
\mathrm{E}\left[(w_{\mathbf{s}} \bar{V}_n^{(\mathbf{s})})^2\right] =& 2\int_0^\infty y \,\mathrm{P}(|w_{\mathbf{s}} \bar{V}_n^{(\mathbf{s})}| > y) dy = 2\int_0^{m_n} y \,\mathrm{P}(|w_{\mathbf{s}} \bar{V}_n^{(\mathbf{s})}| > y) dy \\
&\leq 2\int_0^{m_n} y \,\mathrm{P}(|w_{\mathbf{s}} V^{(n,\mathbf{s})}| > y) dy \qquad (52) \\
&\leq 2\int_0^{m_n} y \,\mathrm{P}(|u_{\tilde{w}} V| > y) dy, \qquad (53)
\end{aligned}$$

其中，(53) 由随机占优假设从 (52) 推出。因此，

$$(\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}})^{-2} \sum n_{\mathbf{s}}^2 \tilde{w}_{\mathbf{s}}^2 \,\mathrm{E}\left[(\bar{V}_n^{(\mathbf{s})})^2\right] \leq (\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}})^{-2} 2 m_n \int_0^{m_n} y \,\mathrm{P}(|u_{\tilde{w}} V| > y) dy.$$

由于我们假设 $m_n/\sum_{\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}} = O_P$ 1 51 $m_n^{-1}\int_0^{m_n} y$ $|u_{\tilde{w}} V| >$ y $\to 0$ $n \uparrow \infty$ $m_n \to \infty$ $x^{-1}\int_0^x y$ $|u_{\tilde{w}} V| >$ (s) ( )，对于 ( )，只需证明当 时 P( )。根据假设，这可由 P(y) $\to 0$ 得出，而后者是 E $|u_{\tilde{w}} V| < \infty$ 的推论，正如 Durrett（2019）在定理 2.1.12 和 2.1.14 的证明中所示。至此，我们完成了验证：(47) 依概率收敛到 0。

L_1 收敛于$\psi_{\mathcal{S}_n}(\eta)$ −E[$\psi_{\mathcal{S}_n}(\eta)$]。根据假设，存在 $p > 1$ 使得 $\|V(n,\mathbf{s})\|_{L_p} \leq \|V\|_{L_p}$对所有 n和 s $\in \mathcal{S}_n$均成立。因此， $\|(\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{n,\mathbf{s}})^{-1} \sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{n,\mathbf{s}} V(n,\mathbf{s})\|_{L_p} \leq \|V\|_{L_p}$ 也成立。根据 Durrett（2019）中的定理 1.6.8 和习题 2.3.5 所述的随机变量控制收敛原理，(47) 因此也依概率论收敛到 0。

L_1- **和依概率收敛于** $\psi_{\mathcal{S}_n}(\eta)$ $-\mathrm{E}[\psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n]$。 注意到

$$\begin{aligned}\|\,\mathrm{E}[\psi_{\mathcal{S}_n}(\eta)|\mathcal{G}_n] - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)]\|_{L_1} &= \|\,\mathrm{E}\{\psi_{\mathcal{S}_n}(\eta) - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)] \mid \mathcal{G}_n\}\|_{L_1} \\ &\leq \|\psi_{\mathcal{S}_n}(\eta) - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)]\|_{L_1},\end{aligned}$$

$\mathcal{F}_n$ 由于它是较小的 σ-代数 $\{\mathcal{F}_n, \mathcal{G}_n\}$, $\mathrm{E}[\psi_{\mathcal{S}_n}(\eta)]$ 与 $\mathrm{E}[\psi_{\mathcal{S}_n}(\eta) \mid \mathcal{F}_n]$ 相同，且条件期望算子在 L_1中是压缩映射（Durrett，2019，定理 4.1.11）。结合这一点,

$$\|\psi_{\mathcal{S}_n}(\eta) - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)|\mathcal{G}_n]\|_{L_1} \leq \|\psi_{\mathcal{S}_n}(\eta) - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)]\|_{L_1} + \|\,\mathrm{E}[\psi_{\mathcal{S}_n}(\eta)|\mathcal{G}_n] - \mathrm{E}[\psi_{\mathcal{S}_n}(\eta)]\|_{L_1}$$

这意味着 L_1 $\psi_{\mathcal{S}_n}(\eta)$–$\mathrm{E}[\psi_{\mathcal{S}_n}(\eta)|\mathcal{F}_n]$ 的收敛蕴含 $\psi_{\mathcal{S}_n}(\eta)$–$\mathrm{E}[\psi_{\mathcal{S}_n}(\eta)|\mathcal{G}_n]$也在 L_1中收敛到零。最后， L_1 的收敛蕴含依概率收敛。

命题的第 ii 部分。若 $\sum_{\mathbf{s}} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}$ 为正，则 $\eta \mapsto \psi_{\mathcal{S}_n}(\eta)$ 和 $\eta \mapsto \mathrm{E}[\psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n]$ 处处可微，斜率为 -1，且可见当 η 趋于 $\mp\infty$时，它们趋于 $\pm\infty$ 。由此可知，它们有唯一的根。

命题的第 iii 部分。 若 $\mathrm{E}[\psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n] = 0$ 的解 τ_n 趋于极限 $\tau_0 \in (-\infty, \infty)$, 则以下对 Huber 关于单调估计方程标量解一致性的论证（Huber，1964；van der Vaart，1998，引理5.10）的改编表明，$\hat{}\tau_n \to \tau_0$ 依概率成立。固定 $\epsilon > 0$。于是

$$\mathrm{P}[\psi_{\mathcal{S}_n}(\tau_0 - \epsilon) > \epsilon/2, \psi_{\mathcal{S}_n}(\tau_0 + \epsilon) < -\epsilon/2] \leq \mathrm{P}(\tau_0 - \epsilon < \hat{\tau}_n < \tau_0 + \epsilon).$$

左侧趋于 1，因为 $\psi_{\mathcal{S}_n}(\tau_0 \pm \epsilon) - \mathrm{E}[\psi_{\mathcal{S}_n}(\tau_0 \pm \epsilon) \mid \mathcal{G}_n] = o_P(1)$, $\mathrm{P}(|\tau_n - \tau_0| < \epsilon/2) \to 1$，当 $\eta < \tau_n - \epsilon/2$ 时 $\mathrm{E}[\psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n] > \epsilon/2$ ，而当 $\eta > \tau_n + \epsilon/2$ 时 $\mathrm{E}[\psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n] < -\epsilon/2$ 。因此，右侧也趋于 1。 □

引理 8。 令 $\theta_i = \text{logit函数}[\mathrm{P}(Z = 1|\mathbf{X}\beta_n = \mathbf{x}_i\beta_n)]$。若 $|\theta_i - \theta_j| < \delta$ 在 $i \overset{\mathcal{S}_n}{\sim} j$时成立，则对于所有 $\mathbf{s} \in \mathcal{S}_n$ 和 $z\colon\ \mathbf{s} \to \{0, 1\}$ ，使得 $\sum_{i \in \mathbf{s}} \zeta_i = \sum_{i \in \mathbf{s}} z_i \in \{1, n_{\mathbf{s}} - 1\}$,

$$\left| \frac{\pi_{\mathbf{s}}(\zeta)}{n_{\mathbf{s}}^{-1}} - 1 \right| \leq (1 - n_{\mathbf{s}}^{-1})(e^{2\delta} - 1) \tag{54}$$

and

$$\left| \frac{n_{\mathbf{s}}^{-1}}{\pi_{\mathbf{s}}(\zeta)} - 1 \right| \leq (1 - n_{\mathbf{s}}^{-1})(e^{4\delta} - 1). \tag{55}$$

引理 *8* 的证明。 对于s 个 $1:m$ 匹配组，由 (27) 可知存在某个非负整数 m，使得我们有

$$\begin{aligned}\frac{\pi_{\mathbf{s}}(\zeta)}{n_{\mathbf{s}}^{-1}} - 1 =& \frac{1}{n_{\mathbf{s}}^{-1}} \cdot \frac{\exp\theta_i}{\sum_{j\in[i]}\exp(\theta_j)} - 1 \\ =& \frac{n_{\mathbf{s}} - \sum_{j\in[i]}\exp(\theta_j-\theta_i)}{\sum_{j\in[i]}\exp(\theta_j-\theta_i)} \\ =& \frac{m - \sum_{j\in[i]\setminus\{i\}}\exp(\theta_j-\theta_i)}{\sum_{j\in[i]}\exp(\theta_j-\theta_i)}; \quad \text{so}\end{aligned} \tag{56}$$

$$\frac{m - m\exp(\delta)}{(m+1)\exp(-\delta)} \le \frac{\pi_{\mathbf{s}}(\zeta)}{n_{\mathbf{s}}^{-1}} - 1 \le \frac{m - m\exp(-\delta)}{(m+1)\exp(-\delta)}, \tag{57}$$

$$-\frac{m}{m+1}e^{\delta}(e^{\delta}-1) \le \frac{\pi_{\mathbf{s}}(\zeta)}{n_{\mathbf{s}}^{-1}} - 1 \le \frac{m}{m+1}(e^{\delta}-1) \quad \text{and}$$

$$\left|\frac{\pi_{\mathbf{s}}(\zeta)}{n_{\mathbf{s}}^{-1}} - 1\right| \le \frac{n_{\mathbf{s}}-1}{n_{\mathbf{s}}}e^{\delta}(e^{\delta}-1). \tag{58}$$

现在注意到 $e^{2\delta} - e^{\delta} < e^{2\delta} - 1$ 对于正的 δ；(54) 得证。

如果 s 是一个 m：1 个匹配组， $m \ge 0$，那么将 $-\theta_i$ 和 $-\theta_j$ 分别代入 θ_i 和 θ_j后，导出 (58) 的论证同样适用。(54) 再次得证。

在所有情况下均已证明 (54) 后，由 (54) 的推论可得 (55)，即

$$\left|\frac{\pi_{\mathbf{s}}(\zeta)}{n_{\mathbf{s}}^{-1}}\right| \le e^{2\delta};$$

the 恒等式 $|x^{-1}-1| \le |x^{-1}|\cdot|x-1|$，对 $x \ne 0$成立；且 $e^{2\delta}(e^{2\delta}-1) \le e^{4\delta}-1$。 □

命题 *10* 的证明。断言(i) 由此可得

$$\mathrm{E}[\tilde{\psi}_{\mathbf{s}}(\eta) \mid \mathcal{G}_n] = \tilde{w}_{\mathbf{s}} \sum_{i\in\mathbf{s}} [\mathrm{E}(Y \mid Z=1, \vec{X}\beta_n = \vec{x}_i\beta_n) - \mathrm{E}(Y \mid Z=0, \vec{X}\beta_n = \vec{x}_i\beta_n) - \eta]. \tag{59}$$

为证明这一点，首先注意到 (46) 和 (28) 结合可得

$$\tilde{\psi}_{\mathbf{s}}(\eta) = \frac{\tilde{w}_{\mathbf{s}}}{\pi_{\mathbf{s}}(Z_{\mathbf{s}})}(\bar{Y}_{\mathbf{s}(1)} - \bar{Y}_{\mathbf{s}(0)} - \eta). \tag{60}$$

¯由于假设 $\mathcal{S}_n$ 是精细分层，因此

$Y_{\mathbf{s}(1)} =$ [Y_i $i \in$ $Z_i = 1$] avg (: s)和 $Y_{\mathbf{s}(0)} =$ ¯avg[(Y_i : $i \in$ s $Z_i = 0$)] 其中至少一个实际上是单次观测。(60)取¯ ¯为零， $\psi_{\mathbf{s}}(\eta) = \tilde{w}_{\mathbf{s}} \sum_{i\in\mathbf{s}} \psi_{\mathbf{s}i}(\eta)$ $= \tilde{w}_{\mathbf{s}}(Y_{\mathbf{s}(1)} - Y_{\mathbf{s}(0)} - \eta)$ 也如此，其中 $\tilde{w}_{\mathbf{s}} = 0$ be- ¯cause $Z_{\mathbf{s}} = 0$ 或 1。

˜对于满足$\sum_{i\in\mathbf{s}} Z_i = 1$的$\mathbf{s}$，由(27)可得，(60)的期望值为 $\mathrm{E}[\psi_{\mathbf{s}}(\eta) \mid \mathcal{G}_n] =$

$$\begin{aligned}
&\sum_{i\in\mathbf{s}} \{\mathrm{E}(Y \mid Z=1, \vec{X}\beta_n = \vec{x}_i\beta_n) - \operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}} [\mathrm{E}(Y \mid Z=0, \vec{X}\beta_n = \vec{x}_j\beta_n)] - \eta\} \\
&\qquad\qquad \cdot \frac{\tilde{w}_{\mathbf{s}}}{\pi_{\mathbf{s}}(0^{(\mathbf{s})}_{+i})} \pi_{\mathbf{s}}(0^{(\mathbf{s})}_{+i}) \\
&= \tilde{w}_{\mathbf{s}} \sum_{i\in\mathbf{s}} \{\mathrm{E}(Y \mid Z=1, \vec{X}\beta_n = \vec{x}_i\beta_n) - \operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}} [\mathrm{E}(Y \mid Z=0, \vec{X}\beta_n = \vec{x}_j\beta_n)] - \eta\},
\end{aligned} \tag{61}$$

其中，我们用 $0^{(\mathbf{s})}_{+i}$ 表示定义在$\mathbf{s}$ 上的映射，该映射将 i 映射为 $\mathbf{1}$，并将其余元素˜映射为 $\mathbf{0}$；由此可得（59）。对于满足 $\sum_{i\in\mathbf{s}} [\![z_i = 0]\!] = 1$ 的$\mathbf{s}$，此论证给出 $\mathrm{E}[\psi_{\mathbf{s}}(\eta) \mid \mathcal{G}_n] =$

$$\tilde{w}_{\mathbf{s}} \sum_{i\in\mathbf{s}} \{ \operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}} [\mathrm{E}(Y \mid Z=1, \vec{X}\beta_n = \vec{x}_j\beta_n)] - \mathrm{E}(Y \mid Z=0, \vec{X}\beta_n = \vec{x}_i\beta_n) - \eta\} \tag{62}$$

其中， $1^{(\mathbf{s})}_{-i}$ 表示$\mathbf{s}$ 上的映射，该映射将 i 映射为 $\mathbf{0}$，并将其余元素映射为 $\mathbf{1}$。同样可得（59）。

第 (ii) 部分。 给定$\mathbf{s} \in \mathcal{S}_n$，记 $\mu^{(\mathbf{s})}_0 = \mathrm{E}(\bar{Y}_{\mathbf{s}(0)} \mid \mathcal{F}_n)$ 和 $\mu^{(\mathbf{s})}_1 = \mathrm{E}(\bar{Y}_{\mathbf{s}(1)} \mid \mathcal{F}_n)$。根据对称性，$\mu^{(\mathbf{s})}_z = \mathrm{E}(Y \mid Z_1 = z, \sum_{i=1}^{n_{\mathbf{s}}} Z_i = \sum_{i\in\mathbf{s}} z_i)$，$z = 0, 1$。对于满足 $\sum_{i\in\mathbf{s}} [\![z_i = 1]\!] = 1$的$\mathbf{s} \in \mathcal{S}_n$，有 $\tilde{w}_{\mathbf{s}}^{-1}\mathrm{E}[\psi_{\mathbf{s}}(\eta) \mid \mathcal{G}_n] =$

$$\begin{aligned}
&\sum_{i\in\mathbf{s}} \{[\mathrm{E}(Y \mid Z=1, \vec{X}\beta_n = \vec{x}_i\beta_n) - \operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}} [\mathrm{E}(Y \mid Z=0, \vec{X}\beta_n = \vec{x}_j\beta_n)]] - \eta\} \pi_{\mathbf{s}}(0^{(\mathbf{s})}_{+i}) \\
=&\sum_{i\in\mathbf{s}} \{[\mathrm{E}(Y - \mu^{(\mathbf{s})}_1 \mid Z=1, \vec{X}\beta_n = \vec{x}_i\beta_n) - \operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}} [\mathrm{E}(Y - \mu^{(\mathbf{s})}_0 \mid Z=0, \vec{X}\beta_n = \vec{x}_j\beta_n)]]\} \pi_{\mathbf{s}}(0^{(\mathbf{s})}_{+i}) \\
&+ \mu^{(\mathbf{s})}_1 - \mu^{(\mathbf{s})}_0 - \eta,
\end{aligned}$$

而如果 $\sum_{i\in\mathbf{s}} [\![z_i = 0]\!] = 1$，那么 $\tilde{w}_{\mathbf{s}}^{-1}\mathrm{E}[\psi_{\mathbf{s}}(\eta) \mid \mathcal{G}_n] =$

$$\begin{aligned}
&\sum_{i\in\mathbf{s}} \{[\operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}} [\mathrm{E}(Y \mid Z=1, \vec{X}\beta_n = \vec{x}_j\beta_n)] - \mathrm{E}(Y \mid Z=0, \vec{X}\beta_n = \vec{x}_i\beta_n)] - \eta\} \pi_{\mathbf{s}}(1^{(\mathbf{s})}_{-i}) \\
=&\sum_{i\in\mathbf{s}} \{ \operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}} [\mathrm{E}(Y - \mu^{(\mathbf{s})}_1 \mid Z=1, \vec{X}\beta_n = \vec{x}_j\beta_n)] - \mathrm{E}(Y - \mu^{(\mathbf{s})}_0 \mid Z=0, \vec{X}\beta_n = \vec{x}_i\beta_n)\} \pi_{\mathbf{s}}(1^{(\mathbf{s})}_{-i}) \\
&+ \mu^{(\mathbf{s})}_1 - \mu^{(\mathbf{s})}_0 - \eta.
\end{aligned}$$

同时，由（61）和（62）可得

$$\tilde{w}_{\mathbf{s}}^{-1}\,\mathrm{E}[\psi_{\mathbf{s}}(\eta)\mid\mathcal{G}_n]=\mu_1^{(\mathbf{s})}-\mu_0^{(\mathbf{s})}-\eta+$$
$$\sum_{i\in\mathbf{s}}\{[\mathrm{E}(Y-\mu_1^{(\mathbf{s})}\mid Z=1,\vec{X}\beta_n=\vec{x}_i\beta_n)-\operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}}[\mathrm{E}(Y-\mu_0^{(\mathbf{s})}\mid Z=0,\vec{X}\beta_n=\vec{x}_j\beta_n)]]\}n_{\mathbf{s}}^{-1}$$

or

$$\tilde{w}_{\mathbf{s}}^{-1}\,\mathrm{E}[\psi_{\mathbf{s}}(\eta)\mid\mathcal{G}_n]=\mu_1^{(\mathbf{s})}-\mu_0^{(\mathbf{s})}-\eta+$$
$$\sum_{i\in\mathbf{s}}\{\operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}}[\mathrm{E}(Y-\mu_1^{(\mathbf{s})}\mid Z=1,\vec{X}\beta_n=\vec{x}_j\beta_n)]-\mathrm{E}(Y-\mu_0^{(\mathbf{s})}\mid Z=0,\vec{X}\beta_n=\vec{x}_i\beta_n)\}n_{\mathbf{s}}^{-1},$$

分别取决于 $\sum_{i\in\mathbf{s}}[\![z_i=1]\!]=1$ 或 $\sum_{i\in\mathbf{s}}[\![z_i=0]\!]=1$。对这些表达式作差，

$$\tilde{w}_{\mathbf{s}}^{-1}\,\mathrm{E}[\tilde{\psi}_{\mathbf{s}}(\eta)-\psi_{\mathbf{s}}(\eta)\mid\mathcal{G}_n]=\sum_{i\in\mathbf{s}}\left(\frac{n_{\mathbf{s}}^{-1}}{\pi_{\mathbf{s}}(0_{+i}^{\mathbf{s}})}-1\right)$$
$$\cdot\{[\mathrm{E}(Y-\mu_1^{(\mathbf{s})}\mid Z=1,\vec{X}\beta_n=\vec{x}_i\beta_n)-\operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}}[\mathrm{E}(Y-\mu_0^{(\mathbf{s})}\mid Z=0,\vec{X}\beta_n=\vec{x}_j\beta_n)]]\}\pi_{\mathbf{s}}(0_{+i}^{\mathbf{s}}) \tag{63}$$

若 $\sum_{i\in\mathbf{s}}[\![z_i=1]\!]=1$，且若 $\sum_{i\in\mathbf{s}}[\![z_i=0]\!]=1$，则

$$\tilde{w}_{\mathbf{s}}^{-1}\,\mathrm{E}[\tilde{\psi}_{\mathbf{s}}(\eta)-\psi_{\mathbf{s}}(\eta)\mid\mathcal{G}_n]=\sum_{i\in\mathbf{s}}\left(\frac{n_{\mathbf{s}}^{-1}}{\pi_{\mathbf{s}}(1_{-i}^{\mathbf{s}})}-1\right)$$
$$\cdot\{\operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}}[\mathrm{E}(Y-\mu_1^{(\mathbf{s})}\mid Z=1,\vec{X}\beta_n=\vec{x}_j\beta_n)]-\mathrm{E}(Y-\mu_0^{(\mathbf{s})}\mid Z=0,\vec{X}\beta_n=\vec{x}_i\beta_n)\}\pi_{\mathbf{s}}(1_{-i}^{(\mathbf{s})}). \tag{64}$$

注意到 $n_{\mathbf{s}}\,\mathrm{E}(|V_{(n,\ \mathbf{s})}|\mid\mathcal{G}_n)\ \geq|\mathrm{E}(n_{\mathbf{s}}V_{(n,\ \mathbf{s})}|\mid\mathcal{G}_n)|=$

$$\left|\sum_{i\in\mathbf{s}}\{[\mathrm{E}(Y-\mu_1^{(\mathbf{s})}\mid Z=1,\vec{X}\beta_n=\vec{x}_i\beta_n)-\operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}}[\mathrm{E}(Y-\mu_0^{(\mathbf{s})}\mid Z=0,\vec{X}\beta_n=\vec{x}_j\beta_n)]]\}\pi_{\mathbf{s}}(0_{+i}^{\mathbf{s}})\right|$$

or

$$\left|\sum_{i\in\mathbf{s}}\{\operatorname*{avg}_{j\in\mathbf{s}\setminus\{i\}}[\mathrm{E}(Y-\mu_1^{(\mathbf{s})}\mid Z=1,\vec{X}\beta_n=\vec{x}_j\beta_n)]-\mathrm{E}(Y-\mu_0^{(\mathbf{s})}\mid Z=0,\vec{X}\beta_n=\vec{x}_i\beta_n)\}\pi_{\mathbf{s}}(1_{-i}^{(\mathbf{s})})\right|$$

分别取决于 $\sum_{i\in\mathbf{s}}[\![z_i=1]\!]=1$ 或 $\sum_{i\in\mathbf{s}}[\![z_i=0]\!]=1$，且在相应的相同条件下

$$\left|\frac{n_{\mathbf{s}}^{-1}}{\pi_{\mathbf{s}}(0_{+i}^{\mathbf{s}})}-1\right|\leq\exp(2\sup_{i,j\in\mathbf{s}}|\theta_i-\theta_j|)-1\ \text{ or }\ \left|\frac{n_{\mathbf{s}}^{-1}}{\pi_{\mathbf{s}}(1_{-i}^{\mathbf{s}})}-1\right|\leq\exp(2\sup_{i,j\in\mathbf{s}}|\theta_i-\theta_j|)-1$$

由引理 8，得：

$$|\operatorname{E}[\tilde{\psi}_{\mathbf{s}}(\eta) - \psi_{\mathbf{s}}(\eta) \mid \mathcal{G}_n]| \leq \tilde{w}_{\mathbf{s}} n_{\mathbf{s}} [\exp(2 \sup_{i,j \in \mathbf{s}} |\theta_i - \theta_j|) - 1] \operatorname{E}(|V^{(n,\mathbf{s})}| \mid \mathcal{G}_n). \tag{65}$$

由此得证 (30)。

第 (iii) 部分。 由于条件期望算子在L_p中是压缩映射，

$$\begin{aligned}\operatorname{E}\left(\frac{\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}} \operatorname{E}(|V^{(n,\mathbf{s})}| \mid \mathcal{G}_n)}{\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}}\right) &\leq \frac{\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}} \operatorname{E}|V^{(n,\mathbf{s})}|}{\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}} \\ &\leq \frac{\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}} \operatorname{E}|V|}{\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}} = \operatorname{E}|V|.\end{aligned}$$

由马尔可夫不等式可得 $\sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}}{}^{-1} \sum_{\mathbf{s}\in\mathcal{S}_n} \tilde{w}_{\mathbf{s}} n_{\mathbf{s}} \operatorname{E}(|V_{(n,\mathbf{s})}| \mid \mathcal{G}_n) = O_P(1)$。因此，$|\{\theta_i - \theta_j \colon i \sim j\}|_\infty = o_P(1)$ 与 (30) 合并可推出

$$\sup_{\eta} |\operatorname{E}[\tilde{\psi}_{\mathcal{S}_n}(\eta) - \psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n]| = o_P(1). \tag{66}$$

根据命题 9，$\eta \mapsto \operatorname{E}(\psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n)$ 有唯一的根 τ_n；根据该命题的第 (i) 部分，$\eta \mapsto \operatorname{E}(\psi_{\mathcal{S}_n}(\eta) \mid \mathcal{G}_n)$ 有一个由 (29) 给出的唯一根；由于这两个函数的斜率都远离零，(66) 表明这些根必然彼此收敛。

第 (iv) 部分。 该命题的第 (iv) 部分现由命题 9 的结论 iii 得证。

□