线性回归：基于簇估计量的推断

Subhodeep Dey* Gopal K. Basak† Samarjit Das‡

印度统计研究所

arXiv:2602.04691v1 [stat.ME] 4 Feb 2026

摘要

本文针对聚类数据中的回归系数提出了一种新颖的估计量，该估计量明确考虑了簇内依赖。我们研究了所提估计量在簇大小有限和无限两种情况下的渐近性质。随后，我们将分析扩展到标准随机系数模型，推导出平均（共同）参数的渐近结果，并构建了一种针对一般线性假设的 Wald 型检验。我们还考察了传统混合普通最小二乘（POLS）估计量在随机系数框架中的表现，并表明在广泛的、具有实证意义的设定下，该估计量可能并不可靠。此外，我们在更高层级（超块；第 2 层、第 3 层……）引入了一种新的参数稳定性检验，假设同一层级内各簇之间的参数保持稳定。大量模拟研究证明了所提出检验的有效性，而一项实证应用则展示了其实际相关性。

关键词：最小二乘法，截面数据，簇依赖。

JEL 分类：C01、C13

*理论统计与数学部，印度统计研究所，加尔各答，印度
†理论统计与数学部，印度统计研究所，加尔各答，印度
‡经济研究部，印度统计研究所，加尔各答，印度

1 引言

经济学应用研究者经常使用由相互依赖的随机变量组成的独立簇数据。Moulton (1986) 和 Moulton (1990) 提供了若干典型示例，说明聚类如何自然地出现在实证研究中。当抽样设计涉及抽取随机群体（如学校、家庭或城镇），并对每个被抽中的群体内所有单位进行调查时，聚类便经常出现。数据收集中的成本削减策略也可能导致聚类，例如采访位于同一街区的多个家庭。此外，即使采用简单随机抽样方案，由于单位之间存在未观测到的共同影响，也可能产生依赖关系；例如，某个潜在因子可能为某一州内的所有家庭所共有。簇数据的基本特征是，同一簇内的个体通常相互依赖，而不同簇之间的个体则预期相互独立。

大量文献致力于通过适当校正标准误来促进稳健推断（参见 White（1984）、Arellano（1987）、Conley（1999）、Wooldridge（2003）、Cameron 和 Trivedi（2005）、Wooldridge（2006）、Cameron 等（2008）、Cameron 和 Miller（2015）、MacKinnon 和 Webb（2019）、Abadie 等（2020）、Canay 等（2021）、Hansen（2022）、Abadie 等（2023）、Chiang 等（2023）、MacKinnon 等（2023））。相关文献大多假设簇大小（N，记作）是有限的，而簇数（G，记作）发散至无穷大。这种框架限制了相关稳健标准误调整的适用范围。在实践中，可能出现三种不同的渐近情形：(1) G 是有限的，且 $N \to \infty$。(2) $N \to \infty$，且 $G \to \infty$；(3) N 是有限的，但 $G \to \infty$. 情形 (1) 通常会导致最小二乘估计量不具一致性，并使中心极限定理失效。[1] [2] [3] 当 $N \to \infty$ 和 $G \to \infty$均如此，且没有关于其相对大小的额外限制时，也会出现同样的推断难题。情形 (3) 中，簇大小固定而簇数增加，已在上述文献中得到广泛研究。

然而，针对情形（2）的文献仍然有限，只有两个值得注意的例外：Hansen and Lee(2019）和 Djogbenou 等人(2019）。这两项研究都假设簇大小具有一定的同质性，因此未涵盖少数簇规模远大于多数簇的情形。在这种情况下，这些研究所使用的合并普通最小二乘法（POLS）估计量可能表现不佳。例如，Djogbenou 等人(2019）发现，当一个簇占据主导地位、包含一半观测值时，会出现一些不理想的性质。这些情况并非纯属理论假设。例如，在美国研究州法律与公司治理时，常会遇到以州为单位的聚类，而仅特拉华州就有近半数的公司（Spamann (2019））。为进一步说明这一点，我们考察印度 2017—18年和 2019—20 年 Annual Survey of Industries (ASI) 的示例。来自特里普拉邦、马哈拉施特拉邦和西孟加拉邦等地区的 ASI 数据尤为相关，其中四位数国家工业分类代码（NIC4s）可作为自然的簇标识符。

示例 *1*：I　在特里普拉邦，总共50个簇中存在一个极其庞大的簇。对于t　he

[1] Andrews (2005) 证明，当样本量增长而簇数固定时，最小二乘估计量不相合。[2] Bester 等人 (2011年) 研究了固定 G 和较大 N 条件下的渐近性质，并采用“有限记忆”假设；在该假设下，依赖随距离增加而消失，从而使误差协方差矩阵变得稀疏。[3] Ibragimov 和 M¨uller(2016) 针对簇数较少的情形提出了推断策略，实质上假设每个簇内的模型误差服从高斯分布。

2017—18年，最大簇的尺寸为323，而其余簇的平均尺寸仅为6。特里普拉邦被认为是印度工业化程度最低的邦之一。

示例 *2*：在马哈拉施特拉邦，162 个簇中有 17 个簇各自包含 500 多个单位（依据 2017—18年数据，其中将 4 位数国家工业分类代码视为簇）。其余簇的平均尺寸为 95，另有 56 个簇仅包含 1 至 25 个单位。马哈拉施特拉邦被认为是印度工业化程度最高的邦之一。示例 *3*：在西孟加拉邦，148 个簇中有 3 个簇各自包含 500 多个单位（依据 2017—18 年数据，其中将 4 位数国家工业分类代码视为簇）。其余簇的平均尺寸为 55，另有 70 个簇的尺寸为 1 至 25 个单位。西孟加拉邦被认为是工业化程度中等的邦。

这些示例说明了这样一种情形：有限数量的簇相较于其他簇大得多。

本文表明，少数大型簇（主导集群）的存在可能导致合并普通最小二乘法不相合。我们提出一种新的估计量，并证明即使存在少数大型簇，该估计量仍然相合。随后，我们扩展基本集群模型，以适应基础参数的随机变异。随机系数模型已被广泛用于处理观测单位之间回归关系中的不可观测异质性。 Swamy(1970) 建立了线性随机系数模型的估计方法。正如他恰当地指出的那样：“.....横截面样本中观察到的个体间差异不太可能用仅包含少数自变量的简单回归方程来解释。在这种情况下，可以将回归模型的系数向量视为随机变量，以反映个体间异质性。”我们证明，所提估计量在随机系数模型下同样相合。我们研究了这一所提估计量的渐近性质，并展示了相应的推断方法。

Swamy的研究将分析扩展到面板数据，允许回归系数在个体和企业之间变化，同时保持共同均值结构。关于参数稳定性检验，可参阅一篇有价值的综述：Pe- saran和Yamagata(2008)。该论文通过标准化S wamy的离散统计量，提出了一种斜率同质性检验；在大 N 和 T 面板数据下，该检验适用于检验截面单位间系数同质的原假设。Breitung等人(2016)针对面板数据中的斜率同质性提出了拉格朗日乘数检验，将Breusch–Pagan框架加以推广，同时允许存在异方差性、序列相关和非正态性。

对于纯截面数据，似乎还没有参数稳定性检验。现有文献不允许底层参数在簇之间变化。本文提出一种新检验，用于考察参数向量在超块层级上是否跨簇保持稳定。这里需要指出的是，这种新的参数稳定性检验针对更高层级（超块层级）开发，并假设超块内各簇之间的参数保持稳定。从实践者的角度来看，这一假设或许并不苛刻。请看以下示例。

例*4*：假设我们要估计印度的恩格尔曲线。印度是一个极其多元化的国家，根据共同的语言、文化、价值体系和饮食习惯划分为若干相对同质的邦。接着，每个邦又主要依据一些共同特征（同质性）划分为若干地区。地区再划分为分区。考虑家庭消费

调查（HCES）2022–23年度印度数据，其中家庭作为个体观测，第一阶段抽样单位（FSU）作为簇。每个FSU包含约5至18个家庭。对于印度农村地区，FSU来自村庄；对于城市地区，FSU来自小型市镇。我们可以将地区或邦视为超块。自然可以假设，恩格尔曲线的参数向量会因邦而异；同一分区内甚至同一地区内的不同FSU之间，参数可能不会变化。

若要检验州层面的参数稳定性，我们假设每个州内所有簇（FSU）的参数均保持稳定。另一方面，若要检验地区层面的参数稳定性，我们假设每个地区内所有簇（FSU）的参数均保持稳定。

示例 *5*：在印度，学校（16—18 岁学生）的辍学率或辍学生比例极高，近年来尤其如此。辍学决策（二元）预计由家庭收入（消费）、户主的教育水平、家庭规模、宗教信仰状况等变量决定。教育体系主要由邦政府管理。观察发现，各邦或地区之间的辍学率存在显著差异。学校可视为基本聚类。县或邦可视为超块。

总之，本文提出了一种新估计量，在基本线性模型和随机系数模型下均具有相合性。无论簇大小的规模或不均衡程度如何，所提估计量均保持有效 $G \to \infty$。我们表明，与 Moulton(1986)、Moulton (1990)、Cameron 等人(2008)、Cameron 和 Miller (2015年)、Wooldridge (2003)、Wooldridge(2006)、Cameron 和 Trivedi (2005)，以及 Hansen andLee (2019)所研究的标准估计量相比，所提估计量具有若干优势。特别是，我们证明，当少数大型簇主导样本时，在各种形式的截面依赖下，传统混合普通最小二乘（POLS）估计量在随机系数模型中并不相合。相比之下，所提估计量在此类情形下仍具有相合性。我们进一步基于所提估计量构造了 Wald 型检验统计量，用于在基本模型和随机系数模型下检验一般线性假设。此外，我们还提出了一种有效检验，用于评估超块之间的参数恒定性；超块代表更高层级的聚类结构。最后，我们给出模拟结果和实证应用，以支持本文的理论发现。

本文其余部分安排如下。第 2节介绍参数恒定性框架下的基本模型及其相关假设。本节引入所提估计量，并确立其在各种依赖形式下的渐近性质。第 3 节将分析扩展至允许回归参数在不同簇之间变化的情形，并给出这一广义模型对应的渐近结果。我们还给出了变参数设定下传统混合普通最小二乘（POLS）估计量的结果。第4 节报告模拟研究，以及使用现实世界数据进行的实证分析。第5 节总结全文。所有正式证明均列于附录中。

2 参数恒定的模型

令 Y_{gi} 表示第 i^{th} 个簇中的第 g^{th} 个个体的观测值，并假设其由以下聚类数据线性模型生成：

$$Y_{gi} = X'_{gi}\beta + \epsilon_{gi}, \quad i = 1, 2, \ldots, N_g; \; g = 1, 2, \ldots, G, \tag{2.1}$$

其中， X_{gi} 是第 i^{th} 个簇中 g^{th} 个体的观测回归变量（包括分类变量）组成的 $k \times 1$ 向量，ϵ_{gi} 是相应的随机误差。我们仅考虑单向聚类。将一个簇内的观测值堆叠起来，可得以下模型。

$$Y_g = X_g\beta + \epsilon_g, \quad g = 1, 2, \ldots, G, \tag{2.2}$$

其中， Y_g是一个 $N_g \times 1$向量， X_g是一个 $N_g \times k$矩阵， β是一个 $k \times 1$向量。请注意，对每个 $g = 1$，2，...，G，ϵ_g 的方差-协方差矩阵定义为 Ωg，它是一个 $N_g \times N_g$ 正定矩阵，对角元素一致有界，且与 X相互独立。[4]

在给出主要结果之前，我们先提出一些基本假设并说明几个关键观察。

假设 A1。 我们假设 $X = (X'_1, X'_2, \ldots, X'_G)'$ 相对于ϵ_g, $\forall g$严格外生，即：

$$\mathbb{E}\left[\epsilon_g | X\right] = 0, \quad \forall g.$$

定义 2.1. (i) 设 $\{z_n\}$ 为实数序列， $\{a_n\}$ 为正实数序列。(a) 所谓 $z_n = O_e(a_n)$，是指对某个 $0 < c_1 \leq c_2 < \infty$,

$$c_1 \leq \left|\frac{z_n}{a_n}\right| \leq c_2, \quad \text{for all } n.$$

(b) 所谓 $z_n = O(a_n)$，是指对某个 $0 < c < \infty$,

$$\left|\frac{z_n}{a_n}\right| \leq c, \quad \text{for all } n.$$

(ii) 设 $\left\{z_{N_g}\right\}$为实数的双指标序列， $\left\{a_{N_g}\right\}$为正实数的双指标序列，其中 N_g 是第 g^{th} 个簇的簇大小，它可以增大到 $+\infty$.

(a) 所谓 $z_{N_g} = O_e\left(a_{N_g}\right)$ 关于 g一致地，是指对某个 $0 < c_1 \leq c_2 < \infty$,

$$c_1 \leq \inf_{g, N_g} \left|\frac{z_{N_g}}{a_{N_g}}\right| \leq \sup_{g, N_g} \left|\frac{z_{N_g}}{a_{N_g}}\right| \leq c_2.$$

(b) 所谓 $z_{N_g} = O\left(a_{N_g}\right)$ 关于 g一致地，是指对某个 $0 < c < \infty$,

$$\sup_{g, N_g} \left|\frac{z_{N_g}}{a_{N_g}}\right| \leq c.$$

定义 2.2. (i) 设 $\{Z_n\}$ 为一列随机变量， $\{a_n\}$ 为一列正实数。所谓 $Z_n = O_P(a_n)$，是指对每一个 $\varepsilon > 0$, 存在 $M_\varepsilon > 0$ 和 $L_\varepsilon \in \mathbb{N}$，使得

$$\mathbb{P}\left(\left|\frac{Z_n}{a_n}\right| > M_\varepsilon\right) < \varepsilon, \quad \text{for all } n \geq L_\varepsilon.$$

[4]在对 X.作出一些假设的情况下，可以允许条件异方差

(ii) 设 $\{Z_{N_g}\}$ 为随机变量的双重索引序列，$\{a_{N_g}\}$ 为正实数的双重索引序列，其中 N_g 是 g^{th} 簇的簇大小，该大小可能增加到 $+\infty$. 。根据$Z_{N_g} = O_P\left(a_{N_g}\right)$，在 g，上一致地，我们的意思是：对于每个 $\varepsilon > 0$，存在 $M_\varepsilon > 0$ 使得

$$\sup_{g,N_g} \mathbb{P}\left(\left|\frac{Z_{N_g}}{a_{N_g}}\right| > M_\varepsilon\right) < \varepsilon.$$

定义 2.2（i）本质上意味着，对于每一个 $\varepsilon > 0$，存在 $M_\varepsilon > 0$ 使得 $\sup_n \mathbb{P}\left(\left|\frac{Z_n}{a_n}\right| > M_\varepsilon\right) < \varepsilon$.我们现在刻画误差项的依赖结构，这对我们的分析至关重要。

定义 2.3。 我们定义三种可能的截面依赖。记 $\Omega_g = \mathbb{E}\left[\epsilon_g \epsilon_g' | X\right]$对应于固定簇 g，且第 g^{th} 个簇包含 N_g 个观测值。

1. **强依赖：** 称 N_g 个个体之间的依赖为强依赖，当 $\lambda_{\max}(\Omega_g) = O_e(N_g)$。

2. **半强依赖：** 当满足以下条件时，称 N_g 个个体之间的依赖为半强依赖或中等依赖： $\lambda_{\max}(\Omega_g) = O_e(h(N_g))$，其中 $h(N_g) \uparrow \infty$，随着 $N_g \uparrow \infty$，但 $\frac{h_{(N_g)}}{N_g} \to 0$，随着 $N_g \to \infty$.

3. **弱依赖：** 当 $\lambda_{\max}(\Omega_g) = O_e(1)$。时， N_g 个体之间的依赖被称为弱依赖

强依赖意味着几乎所有个体都彼此相关或相互连接。半强依赖意味着依赖对的数量随样本量增加而增加。弱依赖意味着 Ω 的所有特征值$_g$ 均为有限值。独立性被视为弱依赖。若观测值之间的距离按某种度量增大时，依赖衰减得足够快，也可能满足弱依赖。关于这几种截面依赖类型的更多细节和示例，请参见 Basak 和 Das (2018年）。考虑一种情形：我们有聚类数据，其中包含 G 个簇，且 g^{th} 簇包含 N_g 个观测值。如果 g^{th} 簇呈现强依赖或半强依赖，那么 N_g 必须发散至 $+\infty$，同时 $G \to \infty$. 否则，如果 N_g 为有限值，且 $G \to \infty$，那么总会剩下弱依赖的观测值。

注 1。 在本文中，强依赖是指这样一种情形： $O(G)$ 个簇表现出强依赖，而其他簇可能表现出半强依赖或弱依赖。需要注意的是，强依赖簇的数量可能是 $\frac{G}{2}$ 或 $\frac{G}{3}$，这些情况也包含在上述情形中。半强依赖是指 $O(G)$ 个簇表现出半强依赖，且不存在强依赖簇，但仍可能存在弱依赖簇。另一方面，弱依赖描述的是所有簇均表现出弱依赖，且不存在强依赖或半强依赖簇的情形。

注释 2. 为简化起见，我们假设各簇之间的半强依赖速率相同，即 $h_g(N_g) \simeq h(N_g)\ \forall g$，尽管这一假设可以推广。

假设 A2. 我们假设簇之间的误差相互独立，因此对于每个 g,

$$\mathbb{E}\left[\epsilon_g \epsilon_{g_1}' \middle| X\right] = 0, \ \forall g_1 \neq g \quad and$$
$$\mathbb{E}\left[\epsilon_g \epsilon_g' \middle| X\right] = \Omega_g,$$

这种结构允许同一簇中的个体之间存在任意的簇内异方差和相关性。根据定义2.3，进一步假设

$$\lambda_{\max}(\Omega_g) = \begin{cases} O_e(N_g), & for\ strong\ dependence \\ O_e(h(N_g)), & for\ semi\text{-}strong\ dependence \\ O_e(1), & for\ weak\ dependence \end{cases}$$

一致地对 g，如定义2.1所述。

我们使用精确的定义和阶数记号，旨在系统地追踪假设、定理及其相应证明中出现的众多固定常数。

定义2.4。我们使用两种范数，即最大特征值范数和迹范数。任意矩阵 A 的最大特征值范数定义为 $\|A\|_e = \max_{z\neq 0} \frac{\|Az\|}{\|z\|}$，其中对于一个 $n \times 1$ 向量 y，$\|y\| = (y'y)^{1/2}$. 特别地，对于任意半正定矩阵 A, 最大特征值范数变为$\|A\|_e = \max_{z:\ z'z=1} z'Az$. 迹范数定义为 $\|A\| = [\text{tr}(A'A)]^{\frac{1}{2}}$. 当 $F = \frac{1}{2}$ 时，它也称为弗罗贝尼乌斯范数。

定义 2.5. **(a) 近似平衡簇：** $\min_g\{N_g\}/\max_g\{N_g\}$ 与零之间的距离有正下界 o.
(b) 不平衡簇：$\min_g\{N_g\}/\max_g\{N_g\} = o(1)$。

2.1 拟议估计量

我们分别针对每个簇估计 β，然后通过取简单平均来合并所有这些 G估计值。g^{th} 簇的估计值为 $\hat{}$ $\beta_g = \left(X_g'X_g\right)^{-1} X_g'Y_g, \quad g = 1, 2 \ldots\ldots, G$. 因此，我们的所提估计量是所有这些估计值的平均值，定义如下:

$$\hat{\bar{\beta}} = \frac{1}{G}\sum_{g=1}^{G} \hat{\beta}_g. \tag{2.3}$$

$\hat{}$的方差-协方差矩阵在模型（2.2）下由下式给出:

$$V = \frac{1}{G^2}\sum_{g=1}^{G} \mathbb{E}\left[\left(X_g'X_g\right)^{-1} X_g'\Omega_g X_g \left(X_g'X_g\right)^{-1}\right]. \tag{2.4}$$

假设 A3。 对于某个固定的 $0 < M < \infty$，假设存在某个 $p > 1$,

$$\sup_g \mathbb{E}\left[\left(\text{tr}\left(\left(\frac{X_g'X_g}{N_g}\right)^{-1}\right)\right)^p\right] \leq M.$$

假设 A3 要求各簇中归一化设计矩阵的逆矩阵具有一致矩有界。基本上，这要求 $\left(X_g'X_g/N_g\right)^{-1}$ 的迹具有有限的 p 阶矩，且关于g一致。该假设意味着 $X_g'X_g/N_g$ 的特征值在期望意义下一致地远离零。

定理 2.1。 在假设 *A1*、*A2* 和*A3*成立时，对于强依赖、半强依赖或弱依赖，以及任意簇大小， β 是 β的相合估计量。

定理 2.1 的证明见附录。我们现在给出一个引理，以阐明后续结果。

引理 2.1. 设 $\{z_{N_g}\}$ 与 $\{a_{N_g}\}$ 为正实数的双重索引序列，使得 $z_{N_g} = O_e\left(a_{N_g}\right)$ 关于 g一致成立，其中 N_g 是第 g^{th} 个簇的簇大小。则对任意 $G \geq 1$， $\sum_{g=1}^{G} z_{N_g} = O_e\left(\sum_{g=1}^{G} a_{N_g}\right)$.

该引理的证明见附录。若将 $O_e(\cdot)$ 替换为 $O(\cdot)$，则类似结果成立。注意 $\hat{\beta} - \beta = \frac{1}{G}\sum_{g=1}^{G}\left(\frac{X_g'X_g}{N_g}\right)^{-1}\frac{X_g'\epsilon_g}{N_g}$. 显然，$\hat{\beta}$ 是 β. 的无偏估计量。由附录中给出的定理2.1 的证明，我们有

$$\mathbb{E}\left\|\hat{\beta} - \beta\right\|^2 \leq \frac{1}{G^2}\sum_{g=1}^{G}\mathbb{E}\left[\operatorname{tr}\left(\left(\frac{X_g'X_g}{N_g}\right)^{-1}\right)\right]\frac{\lambda_{max}\left(\Omega_g\right)}{N_g}.$$

然后，根据假设 A2、假设 A3和引理 2.1，我们得到

$$\mathbb{E}\left\|\hat{\beta} - \beta\right\|^2 = \begin{cases} O\left(\frac{1}{G}\right), & \text{for strong dependence} \\ O\left(\frac{1}{G^2}\sum_{g=1}^{G}\frac{h(N_g)}{N_g}\right), & \text{for semi-strong dependence} \\ O\left(\frac{1}{G^2}\sum_{g=1}^{G}\frac{1}{N_g}\right), & \text{for weak dependence} \end{cases} \tag{2.5}$$

推论 2.1。 上述讨论表明， $\hat{\beta}$ 在强依赖情形下是 $\sqrt{G}$-相合的。 $\hat{\beta}$ 在半强依赖情形下是 $\sqrt{G}\,\bar{h}_1^{-\frac{1}{2}}$-相合的，在弱依赖情形下是 $\sqrt{G}\cdot\bar{m}^{-\frac{1}{2}}$-相合的，其中 $\bar{h}_1 = \frac{1}{G}\sum_{g=1}^{G}\frac{h(N_g)}{N_g}$ 和 $\bar{m} = \frac{1}{G}\sum_{g=1}^{G}\frac{1}{N_g}$.

假设 A4。 假设存在某个 $0 < c < \infty$，且存在某个 $p > 1$，

$$\mathbb{E}\left[\left|\epsilon_{gi}\right|^{2p}\middle| X\right] \leq c, \quad \textit{for all } i, g \textit{ and uniformly in } X.$$

上述假设实质上确保了二阶以上矩的存在，从而控制误差项 ϵ_{gi} 的尾部行为。它还为推导估计量的极限分布建立了必要的正则条件。

假设 A5。假设

$$V_G = \frac{1}{G}\sum_{g=1}^{G}\mathbb{E}\left[\left(X_g'X_g\right)^{-1}X_g'\Omega_g X_g\left(X_g'X_g\right)^{-1}\right]$$

是有限的且为正定矩阵。此外，假设

$$O_e\left(\lambda_{min}\left(V_G\right)\right) = O_e\left(\lambda_{max}\left(V_G\right)\right) = \begin{cases} O_e\left(1\right), & \textit{for strong dependence} \\ O_e\left(\frac{1}{G}\sum_{g=1}^{G}\frac{h(N_g)}{N_g}\right), & \textit{for semi-strong dependence} \\ O_e\left(\frac{1}{G}\sum_{g=1}^{G}\frac{1}{N_g}\right), & \textit{for weak dependence} \end{cases}$$

如定义*2.1*所述。

对于弱依赖，上述结论可直接得出，因为最小和最大特征值均满足$\lambda_{min}(\Omega_g) = O_e(1) = \lambda_{max}(\Omega_g)$ 在 g. 上一致成立。此外，通过应用引理 2.1，我们可以严格刻画 $\lambda_{\min}(V_G)$ 和 $\lambda_{\max}(V_G)$在三种依赖类型下的阶。

定理 2.2。 在假设 *A1*、 *A2*、 *A3*、*A4* 和 *A5*下，对于强依赖

$$\sqrt{G}V_G^{-1/2}\left(\hat{\hat{\beta}}-\beta\right) \xrightarrow{d} N_k\left(0, I_k\right), \quad as\ G \to \infty.$$

事实上，对于半强依赖，如果对于某些 $p > 1$、 $\frac{G}{\left[\sum_{g=1}^{G} \frac{h(N_g)}{N_g}\right]^p} \to 0$，当 $G \to \infty$. 时，渐近正态性成立。

对于弱依赖，如果对于某些 $p > 1$、 $\frac{G}{\left[\sum_{g=1}^{G} \frac{1}{N_g}\right]^p} \to 0$，当 $G \to \infty$. 时，渐近正态性成立

定理 2.2 的证明见附录。

注 3。 注意，对于 $N_g \simeq N$ 的近似平衡簇情形，在强依赖下，李雅普诺夫条件显然成立，因为它化简为 $O\left(G^{1-p}\right)$，当 $G \to \infty$ 时，对于某些 $p > 1$.，其趋于 0。对于具有 $h(N) \to \infty$ 且 $N \to \infty$ 的半强依赖情形，该条件化简为 $O\left(\left(\frac{N}{h(N)}\right)^p G^{1-p}\right) \to 0$，当$G \to \infty$时，对于某些 $p > 1$.。对于具有 $N \to \infty$ 的弱依赖情形，该条件化简为 $O\left(N^p G^{1-p}\right) \to 0$，当$G \to \infty$时，对于某些 $p > 1$.。这些条件与合并普通最小二乘法估计量具有渐近正态性时所采用的条件有些相似。

注 4. 假设对某个 $p > 1$,

$$\frac{G}{\left[\lambda_{min}\left(\sum_{g=1}^{G} \mathbb{E}\left[\left(X_g' X_g\right)^{-1} X_g' \Omega_g X_g \left(X_g' X_g\right)^{-1}\right]\right)\right]^p} \to 0, \ as\ G \to \infty.$$

那么，$\hat{\beta}$ 的渐近正态性成立，无需假设*A5*。尽管该假设相对较弱，但假设 *A5*中还规定了更具体的条件。

假设 A6. 假设

$$\mathbb{E}\left\|X_g' \epsilon_g \epsilon_g' X_g - X_g' \Omega_g X_g\right\|^2 = \begin{cases} O\left(N_g^4\right), & for\ strong\ dependence \\ O\left(N_g^2 h^2\left(N_g\right)\right), & for\ semi\text{-}strong\ dependence \\ O\left(N_g^2\right), & for\ weak\ dependence \end{cases}$$

对 g.一致地

该假设控制簇级二次型 $X_g' \epsilon_g \epsilon_g' X_g$ 围绕其均值 $X_g' \Omega_g X_g$ 的二阶变异性，且随着簇大小增大，这种变异性也会受到控制。偏差允许随 N_g增大，但其增速取决于簇内依赖是强、半强还是弱。

引理 2.2. 设 $\{N_g\}_{g=1}^{\infty}$ 是一个正整数序列， h 是 N_g 的函数，满足 $h(N_g) \uparrow \infty$，当 $N_g \uparrow \infty$时成立，但 $h(N_g) = o(N_g)$。那么，

$$\frac{\sum_{g=1}^{G} \frac{h^2(N_g)}{N_g^2}}{\left(\sum_{g=1}^{G} \frac{h(N_g)}{N_g}\right)^2} \to 0 \quad and \quad \frac{\sum_{g=1}^{G} \frac{1}{N_g^2}}{\left(\sum_{g=1}^{G} \frac{1}{N_g}\right)^2} \to 0, \quad as\ G \to \infty$$

对于 *(i)* 近似平衡情形，以及 *(ii)* 含有 $O(G)$ 个簇且这些簇近似平衡的不平衡情形，均成立。

定理 2.3. 在假设 *A1*、*A2*、*A3*、*A4*、*A5*、*A6*以及引理 *2.2*成立的情况下，对于近似平衡的簇，ˆ在强依赖、半强依赖或弱依赖下， V_G 关于 V_G相合，其中

$$\hat{V}_G = \frac{1}{G}\sum_{g=1}^{G}\left(X_g'X_g\right)^{-1}X_g'e_ge_g'X_g\left(X_g'X_g\right)^{-1}$$

ˆ其中 $e_g = Y_g - X_g\beta$. 对于不平衡情形，一致性在以下情况下成立：*(i)* 通常的强依赖；*(ii)* 弱依赖和半强依赖，若 $O(G)$ 簇近乎平衡。

定理 2.3的证明见附录。

ˆ**注 5.** 对于不平衡簇，即使 $O(G)$ 簇（例如， $G/2$ 簇）极大，而 $O(G)$ 簇（例如， $G/2$ 簇）近乎平衡，在弱依赖或半强依赖下， V_G 仍具有一致性。

ˆ**注 6.** *1.* 在强依赖下，不平衡情形中 V_G 仍保持一致性。

2. 对于不平衡情形，若所有簇均呈半强依赖，则要求 $\sum_{g=1}^{G}$ $\frac{\frac{h^2(N_g)}{N_g^2}}{\left(\sum_{g=1}^{G}\frac{h\ N_g()}{\ N_g}\right)^2}\to$

ˆ0 对于 V_G. 的一致性

3. 对于不平衡情形，如果所有簇均弱依赖，则 $\frac{\sum_{g=1}^{G}\frac{1}{N_g^2}}{\left(\sum_{g=1}^{G}\frac{1}{N_g}\right)^2}\to 0$ 对于ˆ V_G. 的一致性 $\frac{\quad —}{\quad —}$

ˆ**注 7.** 请注意，在假设 *A6* 下，半强依赖条件下 β与 $p=2$ 渐近正态性的李雅普诺夫条件可简化为

$$\frac{\sum_{g=1}^{G}\frac{h^2(N_g)}{N_g^2}}{\left(\sum_{g=1}^{G}\frac{h(N_g)}{N_g}\right)^2}\to 0, \quad as\ G\to\infty.$$

此外，若将假设 *A6* 改为以下条件：对于某个 $p>1$,

$$\mathbb{E}\left\|X_g'\epsilon_g\epsilon_g'X_g - X_g'\Omega_gX_g\right\|^p = \begin{cases} O\left(N_g^{2p}\right), & for\ strong\ dependence \\ O\left(N_g^p h^p\left(N_g\right)\right), & for\ semi\text{-}strong\ dependence \\ O\left(N_g^p\right), & for\ weak\ dependence \end{cases}$$

对于 g一致地成立，则李雅普诺夫条件可简化为

$$\frac{\sum_{g=1}^{G}\frac{h^p(N_g)}{N_g^p}}{\left(\sum_{g=1}^{G}\frac{h(N_g)}{N_g}\right)^p}\to 0, \quad as\ G\to\infty.$$

根据引理*2.2*，当 $O(G)$ 簇近乎平衡时，上述条件均成立。

在弱截面依赖下也有类似结论，此时某些$p>1$ 的李雅普诺夫条件可简化为

$$\frac{\sum_{g=1}^{G}\frac{1}{N_g^p}}{\left(\sum_{g=1}^{G}\frac{1}{N_g}\right)^p}\to 0, \quad as\ G\to\infty.$$

定理2.2和定理2.3 为使用我们所提估计量构建一般线性假设问题的 Wald 型检验提供了一个框架。考虑检验一般线性假设的问题： H_0： $R\beta = r$ 对立于 H_1： $R\beta \neq r$，其中 R 是一个已知的 $q \times k$ 矩阵，而 r 是一个已知的 $q \times 1$ 向量，且 rank$(R) = q \leq k$. 。下述定理给出了我们所提的检验。

定理 2.4。 设 $\Pi_G^{q\times q} = RV_GR'$ 且 $\hat{} \ \Pi_G^{q\times q} = R\hat{}V_GR'$。那么，在 H_0： $R\beta = r$，并满足假设 *A1*、*A2*、*A3*、*A4*、*A5*、*A6*以及引理 *2.2*的条件下，对于近似平衡簇，

$$G\left(R\hat{\bar{\beta}} - r\right)' \hat{\Pi}_G^{-1}\left(R\hat{\bar{\beta}} - r\right) \xrightarrow{d} \chi_q^2, \ as \ G \to \infty.$$

对于不平衡情形，上述结论在以下情况下成立：*(i)* 一般的强依赖；*(ii)* 弱依赖和半强依赖，前提是 $O(G)$ 簇近似平衡。

定理 2.4 的证明见附录。该定理所述的检验可用于判断在存在簇依赖的截面数据中，模型的解释变量是否显著。

下一节将讨论带有变化参数的簇模型，并给出三种依赖类型下的一些渐近结果。

3 参数变化的线性模型

假设系数向量 β 在各簇之间同质，并没有充分依据。因此，允许 β 在各簇或高层级单位（例如州或县）之间变化是有意义的。本节将在 Swamy (1970) 的框架下，说明当 β 允许在各簇之间变化时，所提估计量仍然相合。我们还将为该估计量建立中心极限定理，并基于超块的概念介绍一种新颖的检验程序。

3.1 带有簇特定随机系数的模型

我们考虑一个适用于聚类数据的线性回归模型，其中回归系数允许因簇而异。具体而言，对于每个簇 $g = 1$, 2、.....、 G，结果变量满足

$$Y_g = X_g\beta_g + \epsilon_g, \tag{3.6}$$

其中 β_g 是一个 $k \times 1$ 簇特定回归系数组成的向量，而 X_g、 Y_g和 ϵ_g 的定义与模型（2.2）中的定义类似。为允许簇间存在系统性异质性，我们采用随机系数设定。具体而言，我们假设 $\beta_g = \beta + u_g$，其中 β 是一个共同的 $k \times 1$ 参数向量，而u_g 表示相对于共同均值的簇特定偏差。将其代入式（3.6）可得

$$Y_g = X_g\beta + X_gu_g + \epsilon_g, \quad g = 1, 2, \ldots, G. \tag{3.7}$$

假设 A7. 对于每个簇 $g = 1$, 2, ..., G，随机系数向量 u_g 独立于特质误差项 ϵ_g 和回归变量矩阵 X_g。此外，集合 $\{u_g\}_{g=1}^G$ 在各簇之间相互独立。

假设A7确保随机系数的外生性，其要求是 u_g 在每个簇内独立于回归变量和特质误差，并且在各簇之间相互独立。

假设 A8. 随机向量 u_g 在各簇之间独立分布，且

$$\mathbb{E}[u_g] = 0 \quad and \quad \mathrm{Var}(u_g) = \Delta,$$

其中 Δ 是有限的正定 $k \times k$ 矩阵。

假设A8 通过允许 u_g的各个元素之间具有不受限制的协方差，同时保持共同均值为零，从而形式化了随机系数结构。

对于模型（3.6），（2.3）中所述的拟议估计量为

$$\hat{\bar{\beta}} = \bar{\beta} + \frac{1}{G}\sum_{g=1}^{G} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g. \tag{3.8}$$

注意，对于随机系数模型（3.7），

$$\hat{\bar{\beta}} = \beta + \frac{1}{G}\sum_{g=1}^{G} u_g + \frac{1}{G}\sum_{g=1}^{G} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g. \tag{3.9}$$

显然，β 是 β 在模型（3.6）下的无偏估计量。在模型（3.7）的随机系数设定下，该估计量对于共同参数 β仍然无偏。定义模型（3.7）下 β的方差-协方差矩阵为

$$V^* = V + \frac{1}{G}\Delta\,, \tag{3.10}$$

其中 V是 $\frac{1}{G}\sum_{g=1}^{G} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g$ 在（2.4）中定义的方差-协方差矩阵。

定理 3.1。 对于模型 (3.7)，在假设 *A1*、*A2*、*A3*、*A7*和 *A8*成立的条件下，对于强、半强、或弱依赖，以及任意簇大小，β 是 β的相合估计量。

定理 3.1 的证明见附录。

注意，根据附录中给出的定理 3.1 的证明，我们可以得到

$$\mathbb{E}\left\|\hat{\bar{\beta}} - \beta\right\|^2 = \mathrm{tr}(V) + \frac{1}{G}\,\mathrm{tr}\left(\Delta\right) = O\left(\frac{1}{G}\right),$$

无论 ϵ_g 的依赖结构如何，且对于任意簇大小。因此，对于模型（3.7），$\hat{\bar{\beta}}$ 在 $\sqrt{G}$-一致估计 β，适用于任何类型的截面依赖。

注释 8。 由于 β 具有 (3.8)中给出的表示形式，因此它是各簇平均回归参数的相合估计量。因此，当关注参数为 β时，所提估计量能够相合地估计该量。

假设 A9. 对于某个 $p > 1$，每个分量 u_{gj} 的 $2p$ 阶绝对矩对所有 $j = 1, \ldots, k$ 以及 $g = 1, 2, \ldots, G$ 均存在且有限。

假设 A9 为存在随机系数时应用中心极限定理提供了矩条件。

定理 3.2。 在假设 *A1*、*A2*、*A3*、*A4*、*A5*、*A7*、*A8*和*A9*下，对于强依赖、半强依赖或弱依赖，以及任意簇大小， $V^{*-1/2}(\hat{\bar{\beta}} - \beta) \xrightarrow{d} N_k(0, I_k)$。

定理 3.2 的证明见附录。

注 9. 对于随机系数模型 (3.7) 当 $\Delta \neq 0$ 时，在弱依赖和半强依赖的情况下，所提估计量的中心极限定理不需要满足定理*2.2* 中陈述的限制条件。

定义 $V_G^* = GV^*$. 为了估计随机系数模型中的这一方差，我们采用定理 2.3 中的方差-协方差矩阵估计量：

$$\hat{V}_G^* = \frac{1}{G}\sum_{g=1}^{G} \left(X_g'X_g\right)^{-1} X_g' e_g e_g' X_g \left(X_g'X_g\right)^{-1}.$$

$\hat{}\,\hat{}$- -该模型的残差为 $e_g = Y_g - X_g\beta = X_g u_g + \epsilon_g - X_g(\beta - \beta)$。

定理 3.3. 在假设*A1*、*A2*、*A3*、*A4*、*A5*、*A7*、*A8* 和 *A9*成立时，对于强依赖、半强依赖或弱$\hat{}$依赖，且簇大小任意， V_G^* 是 V_G^* 在模型 (3.7)下的相合估计量。

定理 3.3 的证明见附录。

我们来考虑对一般线性假设进行检验的问题 H_0： $R\beta = r$，以检验 H_1： $R\beta \neq r$ 对于模型（3.7），其中 R 是已知的 $q \times k$ 矩阵， r 是已知的 $q \times 1$ 向量，且 rank$(R) = q \leq k$. 以下定理给出了所提出的检验统计量及其渐近性质。

定理 3.4。 设 $\Gamma_G = RV_G^*R'$ 和 $\hat{}\,\Gamma_G = R\hat{}V_G^*R'$。在原假设 H_0： $R\beta = r$ 下，并且假设*A1*、*A2*、*A3*、*A4*、*A5*、*A7*、*A8*以及*A9*，对于强依赖、半强依赖或弱依赖，以及任意簇大小，

$$G\left(R\hat{\bar{\beta}} - r\right)' \hat{\Gamma}_G^{-1} \left(R\hat{\bar{\beta}} - r\right) \xrightarrow{d} \chi_q^2, \quad as\ G \to \infty.$$

定理 3.4 的证明见附录。在下一小节中，我们提出一种新方法，用于检验各簇之间的参数是否恒定。

3.2 使用超块检验参数恒定性

假设我们有按超块结构组织的聚类数据，其中数据被分为 D 个超块。 l^{th} 个超块包含 P_l 个簇，使得

$$\sum_{l=1}^{D} P_l = G.$$

我们假设，对于 l^{th} 超块 $\mathcal{S}_l$， $\beta_g = \beta + u_l$， 对于 $g \in \mathcal{S}_l$，其中 β 是一个共同的 $k \times 1$ 参数向量，且 u_l 具有与假设 A7 和假设A8中所述类似的性质，即 u_l均值为 0， 且方差矩阵为 Δ. 将此表示代入式（3.6）可得

$$Y_g = X_g\beta + X_g u_l + \epsilon_g; \quad g \in \mathcal{S}_l,\ l = 1, 2, \ldots, D. \tag{3.11}$$

该设定允许每个超块内部的斜率系数保持一致，同时通过随机效应 u_l，允许不同超块之间存在系统性的异质性。因此，复合误差

项同时捕捉特异性冲击和超区块层面的参数变化。从实际角度来看，这一假设可能并不会限制模型。对于大多数应用而言，可以较为稳妥地假设，在低层级（村庄、学校.....），参数不会发生变化，但底层参数向量可能会在地区、州或国家层面发生变化，即在高层级（第 2 层、第 3 层.....）发生变化。

对于 l^{th} 超块 $\mathcal{S}_l$，定义 $\tilde{\beta}_l = \frac{1}{P_l}\sum_{g\in\mathcal{S}_l}\hat{\beta}_g$，它由属于该 l^{th} 超块的簇取平均构造而成。因此，式（2.3）中提到的所提估计量变为

$$\hat{\bar{\beta}} = \frac{1}{G}\sum_{l=1}^{D} P_l\tilde{\beta}_l.$$

按照定理2.1的证明，很容易注意到，对于每个 l，β_l 都是 β的相合估计量，当 $P_l \to \infty$. 对于 l第 $\mathcal{S}_l$ 个超块，的方差-协方差矩阵为

$$V_{G,l} = \frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\mathbb{E}\left[\left(X_g'X_g\right)^{-1}X_g'\Omega_g X_g\left(X_g'X_g\right)^{-1}\right].$$

假设 A10。假设$V_{G,\ l}$ 有限且正定。此外，根据假设*A 5*

$$O_e\left(\lambda_{min}\left(V_{G,l}\right)\right) = O_e\left(\lambda_{max}\left(V_{G,l}\right)\right) = \begin{cases} O_e\left(\frac{1}{P_l}\right), & for\ strong\ dependence \\ O_e\left(\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}\right), & for\ semi\text{-}strong\ dependence \\ O_e\left(\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}\right), & for\ weak\ dependence \end{cases}$$

对 l一致地。

对于弱依赖，上述结论可直接得出，因为最小特征值和最大特征值均满足$\lambda_{min}(\Omega_g) = O_e(1) = \lambda_{max}(\Omega_g)$，对 g. 一致成立。此外，应用引理2.1，我们可以严格刻画三种依赖类型下 $\lambda_{\min}(V_{G,\ l})$和 $\lambda_{\max}(V_{G,\ l})$的阶，

引理 3.1. 对于模型 (3.11)，定义 $\tilde{V}_l = \frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\left(X_g'X_g\right)^{-1}X_g'e_g^*e_g^{*\prime}X_g\left(X_g'X_g\right)^{-1}$，其中 $e_g^* = \tilde{Y}_g - X_g\beta_l$. 那么，在假设 *A1*、*A2*、*A3*、*A6*和*A10*下，对于每个 $l = 12$、...、D，V_l 关于 $V_{G,l}$ 是相合的，其中 $P_l \to \infty$.

引理 3.1的证明见附录。

假设 A11. 假设当 $P_l \to \infty$ 和 $D \to \infty$ 时，以下收敛成立 *s:*

$$\frac{1}{\sqrt{D}}\sum_{l=1}^{D}\frac{P_l^2}{\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}}\cdot\frac{1}{G^2}\sum_{l=1}^{D}\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g} \to 0.$$

定理 3.5. 令 $T_{SB} = \sum_{l=1}^{D}\left(\tilde{\beta}_l - \hat{\bar{\beta}}\right)'\tilde{V}_l^{-1}\left(\tilde{\beta}_l - \hat{\bar{\beta}}\right)$. 那么，在假设 *A1*、*A2*、*A3*、*A6*、*A10*、*A11*以及引理 *3.1*下，对于近似平衡簇，无论存在何种依赖，

$$Z = \frac{T_{SB} - kD}{\sqrt{2kD}} \xrightarrow[H_0]{d} N(0,1), \quad as\ \min_l P_l \to \infty,\ D \to \infty,\ and\ \frac{D}{\min_l P_l} \to 0.$$

定理 3.5 的证明见附录。

该定理中构造的检验统计量旨在检验超块之间的参数恒定性。假设 $D/\min_l P_l \to 0$ as $D \to \infty$ and $\min_l P_l \to \infty$ 在经济学和统计学上都有意义，因为它确保超块数的增长速度低于每个超块内簇数最小值的增长速度。

在接下来的小节中，我们考察传统合并普通最小二乘（POLS）估计量在随机系数模型下的性质，并考虑任意簇大小下横截面误差依赖的不同形式。

3.3 混合普通最小二乘法

聚类框架中最常用的估计量是混合普通最小二乘（POLS）估计量。将模型（2.2）中的各簇观测值堆叠起来，可得

$$Y = X\beta + \epsilon,$$

其中， Y和 ϵ 是维度为 $\left(\sum_{g=1}^{G} N_g\right) \times 1$的向量， X 是一个由回归变量构成的 $\left(\sum_{g=1}^{G} N_g\right) \times k$ 矩阵。对于上述模型， β 的POLS估计量定义为

$$\hat{\beta}_{POLS} = (X'X)^{-1} X'Y, \tag{3.12}$$

当数据按簇合并时，该估计量与普通最小二乘估计量一致。在假设A1 和A2成立的条件下， β_{POLS} 的方差-协方差矩阵可表示为

$$\Sigma = \mathbb{E}\left[(X'X)^{-1} X'\Omega X (X'X)^{-1}\right], \tag{3.13}$$

其中，Ω =Var(ϵ)表示堆叠误差向量 ϵ的方差-协方差矩阵。在此设定下，传统簇稳健方差估计量（CRVE）的形式为:

$$\hat{\Sigma} = (X'X)^{-1} \left(\sum_{g=1}^{G} X_g'\hat{\epsilon}_g\hat{\epsilon}_g'X_g\right)(X'X)^{-1}, \tag{3.14}$$

ˆ其中，$\hat{\epsilon}_g = Y_g - X_g\beta_{POLS}$表示第 g个簇的残差向量，其中 $g = 1$，2、.....、 G。在存在任意簇内相关性的情况下，只要簇数 G 足够大且簇大小满足温和的齐性条件，该ˆ估计量就是真实方差的相合估计量。

在若干种设定下，可以证明所提估计量比混合普通最小二乘（POLS）估计量渐近更有效。在此背景下，我们考虑以下结果。

结果 3.1。如果$\Omega_g = a_g I_{N_g}$ 且 $X_g'X_g = a_g$，那么 $V < \Sigma$.

结果3.1 的证明见附录。上述结果清楚地表明，即使存在较弱的截面依赖性，所提估计量也更有效。

结果 3.2。假设$\Omega_g = (a-b)I_{N_g} + b\mathbb{1}\mathbb{1}'$ 且 $X_g = \mathbb{1}$, $\forall g$，其中 $0 < b < a < \infty$. 。此外，假设第一个簇渐近占主导，且 $N_1 \to \infty$，而其余簇的ˆ ˆ尺寸有限。则，所提估计量 β 渐近上比 β_{POLS}更有效。

上述结果的证明见附录。请注意，在此设定下，Ω_g 表现出较强的截面依赖。由该结果可知，在存在强误差依赖和不平衡簇大小的情况下，所提估计量渐近上更有效。

我们将模型（3.7）改写为

$$Y_g = X_g\beta + \eta_g, \quad \text{where } \eta_g = X_g u_g + \epsilon_g. \tag{3.15}$$

这里，复合误差项 η_g 同时包含特异性冲击和由随机系数变异引起的不可观测异质性。模型（β 3.15）的合并普通最小二乘估计为

$$\hat{\beta}_{POLS} = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}\sum_{g=1}^{G} X_g'\eta_g. \tag{3.16}$$

ˆ显然，$\mathbb{E}[\beta_{POLS}] = 0$，因为 $\mathbb{E}[\eta_g] = 0$，根据假设 A1 和 A8。此外，根据假设 A7、A8 以及ˆ方程（3.13），模型（β_{POLS} 3.15）下的方差-协方差矩阵为

$$\Sigma^* = \Sigma + \mathbb{E}\left[(X'X)^{-1}\sum_{g=1}^{G} X_g'X_g\Delta X_g'X_g(X'X)^{-1}\right]. \tag{3.17}$$

现在我们建立簇 β 变化时合并普通最小二乘法的渐近理论。

假设 A12。 *(i)* 假设 $C_G = \mathbb{E}\left[\frac{X'X}{\sum_{g=1}^{G} N_g}\right]$ 有限且在 G.上一致正定。此外，假设

$$\frac{X'X}{\sum_{g=1}^{G} N_g} - C_G = o_P(1).$$

(ii) 假设对于某个 $p > 1$,

$$\mathbb{E}\left[\left(\operatorname{tr}\left(X_g'X_g\right)\right)^p\right] = O_e\left(N_g^p\right) \quad \textit{uniformly in } g,$$

如定义*2.1*所述。

根据假设 A12（i），我们的含义是 $\left|\left(\frac{X'X}{\sum_{g=1}^{G} N_g}\right)_{j,\,j_1} - (D_G)_{j,\,j_1}\right| = o_P(1)$，对所有 j，$j_1 = 1, 2,$ ....., k. 这意味着样本矩阵逐元素依概率收敛于其期望。假设A12（ii）通过要求当 N_g 增加时，回归变量的平均平方幅度一致地保持在远离零的水平，来控制每个簇内回归变量矩阵的增长 g。它排除了 $X_g'X_g$ 增长速度超过 N_g阶的设计。

定理 3.6。在假设 *A1*、*A2*、*A7*、*A8*和*A12*下，在存在强、半强或ˆ弱截面依赖的情况下，β_{POLS} 对于模型（3.15）中的 β 是相合的，前提是：（*i*）各簇近似平衡，或（*ii*）各簇不平衡且满足限制条件$\max_g N_g / \sum_{g=1}^{G} N_g \to 0$，当 $G \to \infty$.时

定理 3.6 的证明见附录。

ˆ**定理 3.7**。 在满足假设 *A1*、*A2*、*A7*、*A8*和 *A12*时，*POLS*估计量 β_{POLS} 对于不平衡簇并非相合，除非 $\frac{\sum_{g=1}^{G} N_g^2}{\left(\sum_{g=1}^{G} N_g\right)^2} \to 0$.

定理 3.7 的证明见附录。

如果有限数量的簇极大，而其余簇有界，或以低于这些簇的˙速率无界，则 β_{POLS} 会变得不相合。然而，无需赘述，即使在此类˚情形下， β 仍然是相合的。簇规模异质性是企业层面数据集和行业层面数据集的常见特征。例如，印度的 Annual Survey of Industries (ASI)、美国的 Compustat 以及覆盖多个国家的 ORBIS 等企业层面数据集显示，不同行业的企业数量存在显著差异。在这些数据中，少数行业占据了企业总数中不成比例的较大份额，而许多其他行业仅包含相对较少的企业。这种不平衡簇大小在实践中自然产生，并且在大规模经济数据集中经常可以观察到。

下一节将介绍所提出的两个检验以及基于POLS的检验在模拟数据和实证数据集中的应用。

4 模拟与实证分析

˚本节考察拟议估计量 β 在各种数据生成过程下的有限样本表现。我们研究前文各节提出的所有检验程序和检验统计量的表现，重点关注其经验尺寸和检验功效。我们进一步比较所提出的检验程序与基于混合普通最小二乘法（POLS）的传统检验在检验一般线性假设方面的表现。此外，我们还通过蒙特卡罗模拟实验以及使用印度家庭消费支出调查数据开展的实证应用，对用于检验参数恒定性的超块估计方法进行评估。

4.1 模拟研究

为考察我们方法的有限样本性质，我们针对含截距和单个回归变量的线性模型进行了蒙特卡洛模拟。

4.1.1 一般线性假设检验

我们感兴趣的是检验 $H_0: \beta = \beta_0$ 对 $H_1: \beta \neq \beta_0$ 在模型（2.2）下。我们生成 10000 次重复，每次重复都会从数据生成过程中抽取一组新数据，并据此决定是否拒绝 H_0。每次重复中都有 G 个簇，每个簇中有 N_g 个体。模拟过程如下所述。模拟数据生成如下：

$$Y_{gi} = X'_{gi}\beta + \epsilon_{gi}.$$

误差项 ϵ_g 服从多元正态分布 $N_{N_g}(0, \Omega_g)$，其中 Ω_g 的构造通过以下设置确保强依赖：

$$\Omega_g = M_g M'_g, \tag{4.18}$$

其中， M_g 是一个 $N_g \times N_g$ 矩阵，其元素 m_{ij} 独立地取自 Unif $[-5, 10]$。根据构造，对于每个组 g，这一设定使 Ω_g 的最大特征值的阶为 $O(N_g)$，从而

在数据中生成所需的强依赖结构。在模拟设计中，第一个簇设为大型簇。对于 $g = 2$、3、.....、G，簇大小 N_g 独立地取自 $\{25, \ldots, 50\}$上的离散均匀分布。回归变量 X_{gi} 为二维，由截距和斜率组成。对于第一个簇，令 $p_1^{(\max)}$ 表示与最大特征值 $\lambda_{\max}(\Omega_1)$对应的特征向量。定义 $X_1^* = \operatorname{sgn}\left(p_1^{(\max)}\right)$，并构造回归变量矩阵 $X_1 = [\,\mathbb{1}_{N_1}, c_1 X_1^*\,]$，其中缩放系数 $c_{1j} \sim$独立地取自 Unif(2, 10)，对应 $j = 1$、2、.....、N_1。对于簇 $g = 2$、3、.....、G，回归变量矩阵 X_g 的生成方式为：第一列为 $\mathbb{1}_{N_g}$，第二列由 N_g个独立抽样构成，抽样来自正态分布 $N(\mu_g, \omega_g^2)$，其中 $\mu_g \sim$取自 Unif(10, 100)，$\omega_g^2 \sim$取自 Unif(200, 300)。生成 $\{X_{gi}, Y_{gi}\}$后，按照 (2.3) 计算所提估计量。固定回归变量和模型参数后，我们将整个数据生成过程重复 $R^* = 10{,}000$次。对于每次重复，我们计算 β及其方差估计量

$$\hat{V}_G = \frac{1}{G}\sum_{g=1}^{G} \left(X_g' X_g\right)^{-1} X_g' e_g e_g' X_g \left(X_g' X_g\right)^{-1},$$

其中 $e_g = Y_g - X_g\beta$，其定义见定理 2.3。相应的检验统计量为

$$T = G\left(R\hat{\beta} - r\right)' \hat{\Delta}_G^{-1}\left(R\hat{\beta} - r\right), \qquad \hat{\Delta}_G = R\hat{V}_G R',$$

在原假设 H_0 和备择假设 H_1下，分别对其进行评估。

在各次重复中，误差项均从相同的分布中生成，而协方差矩阵$\{\Omega_g\}_{g=1}^G$ 保持不变。随后计算该检验在 5% 显著性水平下的经验尺寸和经验功效。经验尺寸定义为在 H_0下，落入拒绝域的检验统计量 T 所占的比例；经验功效则在 H_1下以类似方式定义。为便于进行有意义的功效比较，我们进行了尺寸校正。具体而言，尺寸校正临界值取为：在 H_0下所有重复中得到的 T 经验分布的 95th 分位数。随后使用这一校正后的临界值计算尺寸校正后的经验功效。

作为比较，我们还考虑基于合并普通最小二乘法（POLS）的常用检验统计量，其定义为

$$T_{\text{POLS}} = \left(R\hat{\beta}_{\text{POLS}} - r\right)' \left(R\hat{V}_{\text{POLS}} R'\right)^{-1} \left(R\hat{\beta}_{\text{POLS}} - r\right),$$

其中

$$\hat{V}_{\text{POLS}} = (X'X)^{-1}\left(\sum_{g=1}^{G} X_g' \hat{\epsilon}_g \hat{\epsilon}_g' X_g\right)(X'X)^{-1},$$

与 $\hat{\epsilon}_g = Y_g - X_g\beta_{\text{POLS}}$. 在下表中，我们比较了在误差强依赖和簇大小不平衡的情况下，基于所提估计量的检验与基于POLS的检验，并报告了不同 G 和 N_1取值下各检验的经验尺寸和检验功效。

表 1：一个大型簇存在强依赖时，所提出的检验与基于POLS的检验的比较

$(G,\ N_1)$	Size		临界值		检验功效		尺寸校正后的检验功效	
	$\hat{\tilde{\beta}}$	$\hat{\beta}_{POLS}$	$\hat{\tilde{\beta}}$	$\hat{\beta}_{POLS}$	$\hat{\tilde{\beta}}$	$\hat{\beta}_{POLS}$	$\hat{\tilde{\beta}}$	$\hat{\beta}_{POLS}$
(25, 100)	0.064	0.138	4.27	6.75	0.996	0.995	0.995	0.993
(25, 500)	0.066	0.835	4.42	30.68	0.991	0.866	0.986	0.741
(50, 100)	0.058	0.137	4.21	6.52	1	0.999	1	0.998
(50, 500)	0.057	0.765	4.16	15.44	1	0.9164	1	0.814
(100, 100)	0.051	0.091	3.87	5.08	1	1	1	1
(100, 500)	0.052	0.219	3.89	6.41	1	0.928	1	0.898

重复次数为 10,000。经验尺寸在 $\beta_0 = (1，0。5)$ 处计算，经验功效在 $\beta_1 = (1，1。6)$。

模拟结果表明，所提出的检验程序具有令人满意的尺寸控制，且结果与名义显著性水平非常接近。相比之下，基于POLS的检验存在严重的尺寸扭曲。为解决这一差异并便于进行有效比较，我们使用尺寸校正临界值评估两种检验程序的检验功效。结果显示，与基于POLS的检验相比，所提出的检验始终具有更高的检验功效。此外，随着簇数增加，所提方法相对于POLS的表现会有所提升。

4.1.2 参数恒定性检验

我们感兴趣的是检验 $H_0：\ \Delta = 0$对 $H_1：\ \Delta \neq 0$ 在模型（3.11）下。我们进行 10000 次重复模拟，每次重复模拟都会从数据生成过程中重新抽取一组数据，由此得到对 H_0的拒绝或不拒绝结果。每次重复模拟中有 G 个簇，每个簇包含 N_g 个体。此外，我们假设存在超块结构，其中有 D 个超块，在第 l个超块中，有 P_l 个簇。为简便起见，我们假设采用平衡配置，其中 $P_l = P$，对所有 $l = 1，2，\ldots\ldots，D$。模拟数据根据模型（3.11）生成。

簇大小 N_g 独立地从 $\{25, \ldots, 50\}$ 上的离散均匀分布中抽取。误差项 ϵ_g 从多元正态分布 $N_{N_g}(0, \Omega_g)$ 中抽取，其中 Ω_g 按照（4.18）为每个组 g构造，以在数据中形成所需的强依赖结构。在模拟设计中，回归变量 X_{gi} 是二维的，由截距和斜率组成。回归变量矩阵X_g 的第一列为 $\mathbb{1}_{N_g}$，第二列由 N_g 个独立抽样值组成，这些值来自正态分布 $N(\mu_g, \omega_g^2)$，其中 $\mu_g \sim$服从 Unif(10, 100)，$\omega_g^2 \sim$服从 Unif(200, 300)。超块层面的随机系数偏差 $\{u_l\}$ 从均匀分布和正态分布中抽取，取值覆盖不同的参数范围，具体见表格。

固定回归变量和模型参数，重复整个数据生成过程

$R^* = 10,000$ 次。对于每次重复，我们计算

$$T_{SB} = \sum_{l=1}^{D} \left(\tilde{\beta}_l - \hat{\tilde{\beta}}\right)' \tilde{V}_l^{-1} \left(\tilde{\beta}_l - \hat{\tilde{\beta}}\right),$$

其中

$$\tilde{\beta}_l = \frac{1}{P_l} \sum_{g \in \mathcal{S}_l} \hat{\beta}_g, \qquad \hat{\tilde{\beta}} = \frac{1}{G} \sum_{l=1}^{D} P_l \tilde{\beta}_l,$$

and

$$\tilde{V}_l = \frac{1}{P_l^2} \sum_{g \in \mathcal{S}_l} \left(X_g' X_g\right)^{-1} X_g' e_g^* e_g^{*\prime} X_g \left(X_g' X_g\right)^{-1},$$

~其中 $e_g^* = Y_g - X_g \beta_l$. 相应的检验统计量为

$$Z = \frac{T_{SB} - kD}{\sqrt{2kD}},$$

在原假设 H_0 和备择假设 H_1下均对其进行评估。

协方差矩阵 $\{\Omega_g\}_{g=1}^G$ 在各次重复中保持固定。有限样本表现以名义显著性水平为 5% 时的经验尺寸和检验功效来评估，使用标准正态临界值 $z_{0.975}$。

表2：强依赖下参数恒定性检验的经验尺寸和检验功效

$(P,\ D)$	Size	检验功效		
		$u_l \sim$ 均匀分布 $[-0.1, 0.1]$	$u_l \sim$ 均匀分布 $[-0.2, 0.2]$	$u_l \sim N(0,\ 0.01)$
(25, 25)	0.152	0.441	0.916	0.9278
(25, 50)	0.239	0.7193	0.9988	0.9897
(25, 100)	0.396	0.892	1	0.9898
(50, 25)	0.075	0.6127	0.996	0.965
(100, 25)	0.051	0.8376	1	1
(100, 50)	0.056	0.939	1	1

重复次数为 10,000。经验尺寸在 $\beta_0 = (1,\ 2)$ 处计算。

模拟结果表明，随着 G和 P 增加，检验的经验尺寸趋近于名义水平，而经验功效则趋近于一。此外，模拟证据反映了条件 $D/P \to 0$ 的重要性，该条件似乎是所提程序获得理想有限样本表现所必需的。

4.2 实证示例

以下实证练习使用印度2022–23年家庭消费支出调查（HCES）的家庭层面数据。我们将该调查的抽样设计映射到超块框架，其中家庭为个体观测，第一阶段抽样单位（FSU）为簇，邦为超块。在实证分析中，由于中央直辖区包含的簇数较少，因此将其排除

此外，由于每个FSU最多包含18个家庭，我们剔除了家庭数少于5个的簇，以确保簇内存在有意义的变异。经过这些排除后，最终样本包括来自29个邦的14,491个FSU中的239,744个家庭。

分析重点关注三个关键变量。变量$food$表示参考期内家庭在食品上的支出，$total$表示家庭总消费支出，而$hhsize$表示家庭规模，定义为居住在该家庭中的成员总数。

我们考虑五种不同的恩格尔曲线设定，用于对家庭食品支出行为建模。令 w_i 表示家庭食品支出占家庭总支出的比例， v_i 表示家庭食品支出， M_i 表示家庭总消费支出。以下模型按实证分析中的顺序列出。

模型 *1* （线性食品支出份额模型）。

$$w_i = a_i + b_i M_i + \epsilon_i,$$

模型 *2* （线性模型）。

$$v_i = a_i + b_i M_i + \epsilon_i,$$

模型 *3* （双对数模型或柯布–道格拉斯模型）。

$$\ln v_i = \ln a_i + b_i \ln M_i + \epsilon_i.$$

模型 *4* （半对数模型）。

$$\ln v_i = a_i + b_i M_i + \epsilon_i.$$

模型 *5* （Working–Leser模型）。

$$w_i = a_i + b_i \ln M_i + \epsilon_i.$$

针对汇总后的全印度样本，分别估计这五个模型，其中既包括将家庭规模（家庭规模）作为附加控制变量的情形，也包括不将其纳入的情形。

下表报告了采用拟议方法和混合普通最小二乘法（POLS）估计量所得的估计值，以及基于拟议估计量计算的检验统计量值：

$$Z = \frac{T_{SB} - kD}{\sqrt{2kD}},$$

其中， T_{SB} 是上一小节中定义的基于超块的检验统计量。

表 3：用于检验州级变异的家庭消费支出表

	不含家庭规模			含家庭规模		
	$\hat{\tilde{\beta}}$	$\hat{\beta}_{POLS}$	Z	$\hat{\tilde{\beta}}$	$\hat{\beta}_{POLS}$	Z
模型 1	$\begin{pmatrix} 4.727393e-01 \\ -2.008456e-06 \end{pmatrix}$	$\begin{pmatrix} 4.926134e-01 \\ -2.499980e-06 \end{pmatrix}$	463.2962	$\begin{pmatrix} 4.542658e-01 \\ -5.726445e-06 \\ 2.017493e-02 \end{pmatrix}$	$\begin{pmatrix} 4.370622e-01 \\ -3.009854e-06 \\ 1.539651e-02 \end{pmatrix}$	348.6733
模型 2	$\begin{pmatrix} 1119.4529217 \\ 0.3744132 \end{pmatrix}$	$\begin{pmatrix} 3832.1177046 \\ 0.2303107 \end{pmatrix}$	453.2751	$\begin{pmatrix} 529.7656275 \\ 0.2946063 \\ 551.367824 \end{pmatrix}$	$\begin{pmatrix} 1254.8449853 \\ 0.2066553 \\ 714.3146155 \end{pmatrix}$	290.6762
模型 3	$\begin{pmatrix} 0.2202282 \\ 0.8882348 \end{pmatrix}$	$\begin{pmatrix} 1.2618063 \\ 0.7822229 \end{pmatrix}$	509.3487	$\begin{pmatrix} 1.7128509 \\ 0.7123225 \\ 0.0594726 \end{pmatrix}$	$\begin{pmatrix} 1.79900738 \\ 0.70199741 \\ 0.05776171 \end{pmatrix}$	379.0587
模型 4	$\begin{pmatrix} 8.066576e+00 \\ 4.906426e-05 \end{pmatrix}$	$\begin{pmatrix} 8.497149e+00 \\ 2.022707e-05 \end{pmatrix}$	821.6923	$\begin{pmatrix} 7.968004e+00 \\ 3.739323e-05 \\ 7.544114e-02 \end{pmatrix}$	$\begin{pmatrix} 8.13683e+00 \\ 1.69199e-05 \\ 9.98657e-02 \end{pmatrix}$	594.0962
模型 5	$\begin{pmatrix} 0.80060181 \\ -0.03727856 \end{pmatrix}$	$\begin{pmatrix} 1.23852479 \\ -0.08156762 \end{pmatrix}$	486.5201	$\begin{pmatrix} 1.35120594 \\ -0.10225263 \\ 0.02095447 \end{pmatrix}$	$\begin{pmatrix} 1.43733095 \\ -0.11125730 \\ 0.02137633 \end{pmatrix}$	329.247

数据取自印度 2022—23年的家庭消费支出调查（HCES）。

表中报告的结果表明，对于模型 1，拟议估计量和合并普通最小二乘法（POLS）估计量的斜率系数均为负，这在经济学上是合理的。此外，在全部五种模型设定下，原假设 $H_0:\ \Delta = 0$ 均被拒绝，这为各邦之间存在显著的参数异质性提供了有力证据。估计结果还表明，纳入家庭规模（家庭规模）对结果的影响可以忽略不计。

5 结论

本文提出了一种适用于聚类数据的新型回归系数估计量，能够处理簇依赖问题。我们证明，在各种实际情形下，标准的合并普通最小二乘法（POLS）可能产生不一致的结果。相比之下，我们的所提估计量已被证明在这些条件下仍具有一致性。我们分析了该估计量的渐近性质，涵盖有限簇大小和无限簇大小两种情形，随后探讨了一种典型的经典随机系数模型。在此基础上，我们推导出了平均（共同）参数的渐近结果，并构造了用于一般线性假设检验的 Wald 型检验统计量。此外，我们还提出了一种新型的更高层次（超块）参数稳定性检验，并假设超块内各簇之间的参数保持稳定。我们还考虑了一个实际案例，以展示所提参数稳定性检验的广泛潜在适用范围。

未来的研究可以从多个方面扩展所提出的方法，例如纳入双向聚类、多层建模，并允许存在聚类特定的非随机均值。目前正在研究这些扩展。

6 附录

为便于阅读证明，我们简要说明备注中讨论的各种依赖类型的含义 1。强依赖指的是这样一种情形： $O(G)$ 个簇表现出强依赖，而其他簇可能具有半强依赖或弱依赖。半强依赖是指 $O(G)$ 个簇具有半强依赖，且不存在强依赖的簇，但仍可能存在弱依赖的簇。另一方面，弱依赖描述的是所有簇均具有弱依赖，且不存在强依赖或半强依赖簇的情形。

引理的证明2.1
根据定义2.1，对于双指标正序列 $\left\{z_{N_g}\right\}$ 与 $\left\{a_{N_g}\right\}$， $z_{N_g}=O_e\left(a_{N_g}\right)$ 关于 g 一致地意味着对某个 $0<c_1\leq c_2<\infty$，

$$c_1\leq \inf_{g,N_g}\frac{z_{N_g}}{a_{N_g}}\leq \sup_{g,N_g}\frac{z_{N_g}}{a_{N_g}}\leq c_2,$$

也就是说，对所有 g，存在正常数 c_1 与 c_2，使得 $c_1\cdot a_{N_g}\leq z_{N_g}\leq c_2\cdot a_{N_g}$. 对任意 $G\geq 1$，将这些不等式从 $g=1$ 到 G 求和，得到

$$c_1\cdot\sum_{g=1}^{G}a_{N_g}\leq\sum_{g=1}^{G}z_{N_g}\leq c_2\cdot\sum_{g=1}^{G}a_{N_g}.$$

这可化简为 $c_1\leq\frac{\sum_{g=1}^{G}z_{N_g}}{\sum_{g=1}^{G}a_{N_g}}\leq c_2$，从而意味着

$$c_1\leq\frac{\sum_{g=1}^{G}z_{N_g}}{\sum_{g=1}^{G}a_{N_g}}\leq c_2,\quad \text{for all } G.$$

因此，该引理直接由定义2.1得出。

■

为求完整起见，我们给出以下结果。

结果 6.1。 我们列出证明中用到的一些已知结果，这些结果可见于标准文献。这里，正半定矩阵 $A\in\mathbb{R}^{n\times n}$ 是指 A 为实对称矩阵，且二次型 $z'Az\geq 0$ 对所有 $z\in\mathbb{R}^n$.

1. $\|AB\|\leq\|A\|\,\|B\|$，对于任意矩阵 A 和 B *2.* $tr(AB)\leq tr(A)\,tr(B)$，对于半正定矩阵 A 和 B. *3.* $tr(A)\ \lambda_{\min}(B)\leq tr(AB)\leq tr(A)\ \lambda_{\max}(B)$，对于半正定矩阵 A 和 B *4.* $z'ABA'z\leq(z'AA'z)\ tr(B)$，对于任意 $z\in R^k$、任意矩阵 A 和半正定矩阵 B *5.* $\lambda_{\max}(ABA')\leq\lambda_{\max}(AA')\,\lambda_{\max}(B)$，对于任意矩阵 A 和半正定矩阵 B *6.* $\lambda_{\min}(ABA')\geq\lambda_{\min}(AA')\,\lambda_{\min}(B)$，对于任意矩阵 A 和半正定矩阵 B *7.* $\lambda_{\max}(A+B)\leq\lambda_{\max}(A)+\lambda_{\max}(B)$，对于半正定矩阵 A 和 B.

8. 对于两个半正定矩阵 A 和 B，若 $A \geq B$，则 $\|A\| \geq \|B\|$.

定理证明2.1

^为证明 β的一致性，，首先注意到

$$\hat{\hat{\beta}} - \beta = \frac{1}{G}\sum_{g=1}^{G}\left(\frac{X_g'X_g}{N_g}\right)^{-1}\frac{X_g'\epsilon_g}{N_g}.$$

现在，根据假设A1，注意到

$$\mathbb{E}\left[\hat{\hat{\beta}} - \beta\right] = \mathbb{E}\left[\frac{1}{G}\sum_{g=1}^{G}\left(\frac{X_g'X_g}{N_g}\right)^{-1}\mathbb{E}\left[\frac{X_g'\epsilon_g}{N_g}\middle| X\right]\right] = 0.$$

利用第三行中簇之间的独立性，以及第六行中的结果6.1 ，

$$\begin{aligned}
\mathbb{E}\left\|\hat{\hat{\beta}} - \beta\right\|^2 &= \mathbb{E}\left\|\frac{1}{G}\sum_{g=1}^{G}\left(\frac{X_g'X_g}{N_g}\right)^{-1}\frac{X_g'\epsilon_g}{N_g}\right\|^2 \\
&= \mathbb{E}\mathbb{E}\left[\frac{1}{G^2}\sum_{g=1}^{G}\frac{\epsilon_g'X_g}{N_g}\left(\frac{X_g'X_g}{N_g}\right)^{-1}\sum_{g_1=1}^{G}\left(\frac{X_{g_1}'X_{g_1}}{N_{g_1}}\right)^{-1}\frac{X_{g_1}'\epsilon_{g_1}}{N_{g_1}}\middle| X\right] \\
&= \mathbb{E}\mathbb{E}\left[\frac{1}{G^2}\sum_{g=1}^{G}\frac{\epsilon_g'X_g}{N_g}\left(\frac{X_g'X_g}{N_g}\right)^{-2}\frac{X_g'\epsilon_g}{N_g}\middle| X\right] \\
&= \mathbb{E}\left[\frac{1}{G^2}\sum_{g=1}^{G}\operatorname{tr}\left(\frac{1}{N_g^2}X_g\left(\frac{X_g'X_g}{N_g}\right)^{-2}X_g'\mathbb{E}\left[\epsilon_g\epsilon_g'|X\right]\right)\right] \\
&= \mathbb{E}\left[\frac{1}{G^2}\sum_{g=1}^{G}\operatorname{tr}\left(\frac{1}{N_g^2}X_g\left(\frac{X_g'X_g}{N_g}\right)^{-2}X_g'\Omega_g\right)\right] \\
&\leq \mathbb{E}\left[\frac{1}{G^2}\sum_{g=1}^{G}\operatorname{tr}\left(\frac{1}{N_g}X_g\left(\frac{X_g'X_g}{N_g}\right)^{-2}X_g'\right)\frac{\lambda_{max}\left(\Omega_g\right)}{N_g}\right] \\
&= \mathbb{E}\left[\frac{1}{G^2}\sum_{g=1}^{G}\operatorname{tr}\left(\left(\frac{X_g'X_g}{N_g}\right)^{-1}\right)\frac{\lambda_{max}\left(\Omega_g\right)}{N_g}\right] \\
&= \frac{1}{G^2}\sum_{g=1}^{G}\mathbb{E}\left[\operatorname{tr}\left(\left(\frac{X_g'X_g}{N_g}\right)^{-1}\right)\right]\frac{\lambda_{max}\left(\Omega_g\right)}{N_g}.
\end{aligned}$$

现在，利用以下事实： $\mathbb{E}\left[\operatorname{tr}\left(\left(\frac{X_g'X_g}{N_g}\right)^{-1}\right)\right] = O$在 g 上按假设A3一致地为 1， $\lambda_{max}(\Omega_{N_g}{}^{g} = O)$ 在 g 上对任意类型的依赖均一致地为 1（根据假设A2和引理2.1），可得

$$\mathbb{E}\left\|\hat{\hat{\beta}} - \beta\right\|^2 = O\left(\frac{1}{G}\right) \to 0, \text{ as } G \to \infty.$$

^因此， β 对于 β是一致的，，无论ϵ_g存在何种依赖关系，也无论簇大小如何。

■

定理的证明2.2
为了证明 $\sqrt{G}V_G^{-1/2}\left(\hat{\bar{\beta}} - \beta\right)$，的渐近正态性，首先注意到

$$\sqrt{G}V_G^{-1/2}\left(\hat{\hat{\beta}} - \beta\right) = \frac{1}{\sqrt{G}}V_G^{-1/2}\sum_{g=1}^{G}\left(X_g'X_g\right)^{-1}X_g'\epsilon_g.$$

要证明渐近正态性，只需证明中心极限定理的李雅普诺夫条件。为证明李雅普诺夫条件，需要证明：对于任意固定的 $z \in \mathbb{R}^k$，以及某个 $p > 1$,

$$\frac{\sum_{g=1}^{G} \mathbb{E}\left|z' \frac{1}{\sqrt{G}} V_G^{-1/2} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g\right|^{2p}}{\left(\sum_{g=1}^{G} \mathbb{E}\left[z' \frac{1}{\sqrt{G}} V_G^{-1/2} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g\right]^2\right)^p} \to 0, \text{ as } G \to \infty$$

$$\Longrightarrow \frac{\sum_{g=1}^{G} \mathbb{E}\left|z' V_G^{-1/2} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g\right|^{2p}}{\left(\sum_{g=1}^{G} \mathbb{E}\left[z' V_G^{-1/2} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g\right]^2\right)^p} \to 0, \text{ as } G \to \infty.$$

利用第二行和第三行中的迹不等式，可以观察到

$$\begin{aligned}
\sum_{g=1}^{G} \mathbb{E}\left|z' V_G^{-1/2} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g\right|^{2p} &= \sum_{g=1}^{G} \mathbb{E}\mathbb{E}\left[\left(z' V_G^{-1/2} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g \epsilon_g' X_g \left(X_g' X_g\right)^{-1} V_G^{-1/2} z\right)^p \middle| X\right] \\
&\leq \sum_{g=1}^{G} \mathbb{E}\mathbb{E}\left[\left(z' V_G^{-1/2} \left(X_g' X_g\right)^{-1} V_G^{-1/2} z\right)^p \left(\operatorname{tr}\left(\epsilon_g \epsilon_g'\right)\right)^p \middle| X\right] \\
&\leq \sum_{g=1}^{G} \mathbb{E}\left[\left(z' V_G^{-1} z\right)^p \left(\operatorname{tr}\left(\left(X_g' X_g\right)^{-1}\right)\right)^p \mathbb{E}\left[\left(\epsilon_g' \epsilon_g\right)^p \middle| X\right]\right].
\end{aligned}$$

现在，利用 H ¨older 的不等式和假设A4， p^{th} 的绝对矩 $\epsilon_g' \epsilon_g$ 在给定 X的条件下， i s

$$\mathbb{E}\left[\left(\epsilon_g' \epsilon_g\right)^p \middle| X\right] \leq \mathbb{E}\left[\left(\sum_{i=1}^{N_g} 1^{\frac{p}{p-1}}\right)^{p-1} \sum_{i=1}^{N_g} |\epsilon_{gi}|^{2p} \middle| X\right] = O\left(N_g^p\right) \quad \text{uniformly in } g. \tag{6.19}$$

那么，假设 $\mathbb{E}\left[\left(\epsilon_g' \epsilon_g\right)^p \middle| X\right] \leq c_0 N_g^p$， 对于某个 $0 < c_0 < \infty$， 我们有

$$\sum_{g=1}^{G} \mathbb{E}\left|z' V_G^{-1/2} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g\right|^{2p} \leq c_0 \left(z' V_G^{-1} z\right)^p \sum_{g=1}^{G} \mathbb{E}\left[\left(\operatorname{tr}\left(\left(\frac{X_g' X_g}{N_g}\right)^{-1}\right)\right)^p\right].$$

注意，根据假设A5，我们得到

$$\frac{z' V_G^{-1} z}{z' z} \leq \lambda_{\max}\left(V_G^{-1}\right) = \frac{1}{\lambda_{\min}\left(V_G\right)} = \begin{cases} O(1), & \text{for strong dependence} \\ O\left(\frac{G}{\sum_{g=1}^{G} \frac{h(N_g)}{N_g}}\right), & \text{for semi-strong dependence} \\ O\left(\frac{G}{\sum_{g=1}^{G} \frac{1}{N_g}}\right), & \text{for weak dependence} \end{cases}$$

此外，对于分母，我们有

$$\begin{aligned}
&\left(\sum_{g=1}^{G} \mathbb{E}\left[z' V_G^{-1/2} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g\right]^2\right)^p \\
=& \left(\sum_{g=1}^{G} \mathbb{E}\mathbb{E}\left[\left(z' V_G^{-1/2} \left(X_g' X_g\right)^{-1} X_g' \epsilon_g \epsilon_g' X_g \left(X_g' X_g\right)^{-1} V_G^{-1/2} z\right) \middle| X\right]\right)^p \\
=& \left(z' V_G^{-1/2} \mathbb{E}\left[\sum_{g=1}^{G} \left(X_g' X_g\right)^{-1} X_g' \Omega_g X_g \left(X_g' X_g\right)^{-1}\right] V_G^{-1/2} z\right)^p \\
=& \left(z' V_G^{-1/2} G \cdot V_G V_G^{-1/2} z\right)^p
\end{aligned}$$

$$= \quad G^p \left(z'z\right)^p .$$

因此，在第二行使用假设A3 和引理 2.1 后，李雅普诺夫条件可化简为

$$\begin{aligned}
\frac{\sum_{g=1}^{G} \mathbb{E}\left|z'V_G^{-1/2}\left(X_g'X_g\right)^{-1}X_g'\epsilon_g\right|^{2p}}{\left(\sum_{g=1}^{G}\mathbb{E}\left[z'V_G^{-1/2}\left(X_g'X_g\right)^{-1}X_g'\epsilon_g\right]^2\right)^p} &\leq \frac{c_0\left(z'V_G^{-1}z\right)^p\sum_{g=1}^{G}\mathbb{E}\left[\left(\operatorname{tr}\left(\left(\frac{X_g'X_g}{N_g}\right)^{-1}\right)\right)^p\right]}{G^p\left(z'z\right)^p} \\
&\leq \left(\frac{z'V_G^{-1}z}{z'z}\right)^p\frac{c_0MG}{G^p} \\
&= \begin{cases} O\left(G^{1-p}\right), & \text{for strong dependence} \\ O\left(\frac{G}{\left[\sum_{g=1}^{G}\frac{h(N_g)}{N_g}\right]^p}\right), & \text{for semi-strong dependence} \\ O\left(\frac{G}{\left[\sum_{g=1}^{G}\frac{1}{N_g}\right]^p}\right), & \text{for weak dependence} \end{cases}
\end{aligned}$$

因此，在强依赖下，渐近正态性显然成立。在半强依赖下，若存在 $p>1$, $O\left(\frac{G}{\left[\sum_{g=1}^{G}\frac{h(N_g)}{N_g}\right]^p}\right)\to 0$), 使得当 $G\to\infty$. 时，渐近正态性成立。在弱依赖下，若存在 $p>1$, $O\left(\frac{G}{\left[\sum_{g=1}^{G}\frac{1}{N_g}\right]^p}\right)\to 0$, 使得当 $G\to\infty$.时，渐近正态性成立。

■

引理2.2的证明

注意

$$\frac{\sum_{g=1}^{G}\frac{h^2(N_g)}{N_g^2}}{\left(\sum_{g=1}^{G}\frac{h(N_g)}{N_g}\right)^2} = \frac{1}{1+\frac{\sum_{g_1\neq g_2}\frac{h(N_{g_1})h(N_{g_2})}{N_{g_1}N_{g_2}}}{\sum_{g=1}^{G}\frac{h^2(N_g)}{N_g^2}}} = \frac{1}{1+a_G}, \text{ say.}$$

(i) 所有簇都近乎均衡，因此 $N_g\simeq N$, 对所有 g. 均成立。于是,

$$a_G = \frac{G(G-1)\frac{h^2(N)}{N^2}}{G\frac{h^2(N)}{N^2}} = G-1 = O(G).$$

(ii) 假设 L 个簇与其他簇相比极大，满足 $N_1,\ldots,N_L\to\infty$且 $N_g\simeq N\to\infty$, 对 $g=L+1,\ldots,G$, 成立，但 $\frac{N}{N_j}=o(1)$, 对 $j=1,2,\ldots,L$. 成立。令 $\mathcal{L}$ 为极大簇的集合，并且不失一般性地假设 $\frac{h(N_1)}{N_1}<\frac{h(N_2)}{N_2}<\cdots<\frac{h(N_L)}{N_L}$. 于是,

$$\begin{aligned}
a_G &= \frac{\sum_{g_1,g_2\in\mathcal{L}}\frac{h(N_{g_1})h(N_{g_2})}{N_{g_1}N_{g_2}}+\sum_{g_1\in\mathcal{L},g_2\in\mathcal{L}^c}\frac{h(N_{g_1})h(N_{g_2})}{N_{g_1}N_{g_2}}+(G-L)(G-L-1)\frac{h^2(N)}{N^2}}{\sum_{g\in\mathcal{L}}\frac{h^2(N_g)}{N_g^2}+(G-L)\frac{h^2(N)}{N^2}} \\
&\geq \frac{L(L-1)\frac{h(N_1)h(N_2)}{N_1N_2}+2L(G-L)\frac{h(N_1)h(N)}{N_1N}+(G-L)(G-L-1)\frac{h^2(N)}{N^2}}{L\frac{h^2(N_L)}{N_L^2}+(G-L)\frac{h^2(N)}{N^2}} \\
&= \frac{\frac{L(L-1)}{(G-L)}\frac{h(N_1)h(N_2)N^2}{N_1N_2h^2(N)}+2L\frac{h(N_1)N}{N_1h(N)}+(G-L-1)}{\frac{L}{G-L}\frac{h^2(N_L)N^2}{N_L^2h^2(N)}+1}, \text{ dividing by } (G-L)\frac{h^2(N)}{N^2}.
\end{aligned}$$

显然，分子大于 $G-L-1$，分母为 $o\binom{L}{G-L}+1$，这是因为 $\frac{h^2(N_L)N^2}{N_L^2h^2(N)}\to 0$. 因此，如果 L 是有限的，那么 $a_G=O(G)$。另外，请注意， L 也可能是 $O(G)$，此外还具有

$G-L=O(G)$，例如，$L=G/2$. 那么，$a_G=O(G)$。因此，对于情形 (i) 和情形 (ii)，$\frac{\sum_{g=1}^{G} h(N_g)\left(\sum_{g=1}^{G} h(N_g)\right)^2}{\left(\sum_{g=1}^{G} h(N_g) N_g^2\right)^2}=\frac{\sum_{g=1}^{G} h(N_g)}{\left(\sum_{g=1}^{G} h(N_g) N_g^2\right)^2}\left(\sum_{g=1}^{G} h(N_g)\right)^2$ 令 $h(N_g)=1$，可得 $\frac{\sum_{g=1}^{G} \frac{1}{N_g^2}}{\left(\sum_{g=1}^{G} \frac{1}{N_g}\right)^2} \to 0$，因为 $G \to \infty$ 与上文所述的类似情形相同。引理 2.2 的证明至此结束。

■

定理2.3的证明

要证明$\hat{}V_G$对于V_G是相合的，我们需要证明 $\frac{1}{\|V_G\|}\left\|\hat{}V_G - V_G\right\| = o_P(1)$。为此，注意到$e_g = Y_g - X_g\hat{}\bar{\beta} = \epsilon_g - X_g\left(\hat{}\bar{\beta} - \beta\right)$，这意味着 $e_g e_g' = \epsilon_g\epsilon_g' + X_g\left(\hat{}\bar{\beta} - \beta\right)\left(\hat{}\bar{\beta} - \beta\right)' X_g' - \epsilon_g\left(\hat{}\bar{\beta} - \beta\right)' X_g' - X_g\left(\hat{}\bar{\beta} - \beta\right)\epsilon_g'$. 现在，

$$\begin{aligned}\hat{V}_G - V_G &= \frac{1}{G}\sum_{g=1}^{G}\left(\left(X_g'X_g\right)^{-1}X_g'e_g e_g' X_g\left(X_g'X_g\right)^{-1} - \mathbb{E}\left[\left(X_g'X_g\right)^{-1}X_g'\Omega_g X_g\left(X_g'X_g\right)^{-1}\right]\right)\\ &= T_1 + T_2 + T_3 + T_4, \text{ say,}\end{aligned}$$

where $T_1 = \frac{1}{G}\sum_{g=1}^{G}\left(\left(X_g'X_g\right)^{-1}X_g'\epsilon_g\epsilon_g' X_g\left(X_g'X_g\right)^{-1} - \mathbb{E}\left[\left(X_g'X_g\right)^{-1}X_g'\Omega_g X_g\left(X_g'X_g\right)^{-1}\right]\right)$,
$T_2 = \frac{1}{G}\sum_{g=1}^{G}\left(X_g'X_g\right)^{-1}X_g'X_g\left(\hat{\bar{\beta}} - \beta\right)\left(\hat{\bar{\beta}} - \beta\right)' X_g'X_g\left(X_g'X_g\right)^{-1}$,
$T_3 = -\frac{1}{G}\sum_{g=1}^{G}\left(X_g'X_g\right)^{-1}X_g'\epsilon_g\left(\hat{\bar{\beta}} - \beta\right)' X_g'X_g\left(X_g'X_g\right)^{-1}$ and
$T_4 = -\frac{1}{G}\sum_{g=1}^{G}\left(X_g'X_g\right)^{-1}X_g'X_g\left(\hat{\bar{\beta}} - \beta\right)\epsilon_g' X_g\left(X_g'X_g\right)^{-1}$.

根据假设 A2，在给定 X 的条件下，注意到

$$\begin{aligned}\mathbb{E}[T_1] &= \mathbb{E}\mathbb{E}\left[T_1 \middle| X\right]\\ &= \frac{1}{G}\sum_{g=1}^{G}\mathbb{E}\left[\left(X_g'X_g\right)^{-1}X_g'\mathbb{E}\left[\epsilon_g\epsilon_g'\middle|X\right]X_g\left(X_g'X_g\right)^{-1} - \mathbb{E}\left[\left(X_g'X_g\right)^{-1}X_g'\Omega_g X_g\left(X_g'X_g\right)^{-1}\right]\right]\\ &= 0,\end{aligned}$$

并利用第二行中簇之间的独立性以及第三行中的范数不等式，可得

$$\begin{aligned}\mathbb{E}\|T_1\|^2 &= \mathbb{E}\left\|\frac{1}{G}\sum_{g=1}^{G}\left(X_g'X_g\right)^{-1}\left(X_g'\epsilon_g\epsilon_g' X_g - X_g'\Omega_g X_g\right)\left(X_g'X_g\right)^{-1}\right\|^2\\ &= \frac{1}{G^2}\sum_{g=1}^{G}\mathbb{E}\left[\left\|\left(X_g'X_g\right)^{-1}X_g'\left(\epsilon_g\epsilon_g' - \Omega_g\right)X_g\left(X_g'X_g\right)^{-1}\right\|^2\middle|X\right]\\ &\le \frac{1}{G^2}\sum_{g=1}^{G}\mathbb{E}\left[\left\|\left(X_g'X_g\right)^{-1}\right\|^4\mathbb{E}\left[\left\|X_g'\left(\epsilon_g\epsilon_g' - \Omega_g\right)X_g\right\|^2\middle|X\right]\right]\end{aligned}$$

利用以下事实：$\mathbb{E}\left\|X_g'\epsilon_g\epsilon_g' X_g - X_g'\Omega_g X_g\right\|^2 = \begin{cases} O\left(N_g^4\right) \\ O\left(N_g^2 h^2\text{，对于强依赖（}N_g\text{）}\right)\text{，对于半强依赖，对于弱依赖} \\ O\left(N_g^2\right)\end{cases}$

关于 g，根据假设 A6，且 $\mathbb{E}\left[\left\|\left(X_g'X_g\right)^{-1}\right\|^4\right]=O\left(N_g^{-4}\right)$ 关于 g，根据假设 A3，由引理 2.1

$$\mathbb{E}\left\|T_1\right\|^2=\begin{cases}\frac{1}{G^2}\sum_{g=1}^{G}O\left(N_g^{-4}\right)O\left(N_g^4\right)=O\left(\frac{1}{G}\right), & \text{for strong dependence}\\ \frac{1}{G^2}\sum_{g=1}^{G}O\left(N_g^{-4}\right)O\left(N_g^2h^2\left(N_g\right)\right)=O\left(\frac{1}{G^2}\sum_{g=1}^{G}\frac{h^2(N_g)}{N_g^2}\right), & \text{for semi-strong dependence}\\ \frac{1}{G^2}\sum_{g=1}^{G}O\left(N_g^{-4}\right)O\left(N_g^2\right)=O\left(\frac{1}{G^2}\sum_{g=1}^{G}\frac{1}{N_g^2}\right), & \text{for weak dependence}\end{cases}$$

由此可得

$$\left\|T_1\right\|=\begin{cases}O_P\left(\frac{1}{\sqrt{G}}\right), & \text{for strong dependence}\\ O_P\left(\frac{1}{G}\sqrt{\sum_{g=1}^{G}\frac{h^2(N_g)}{N_g^2}}\right), & \text{for semi-strong dependence}\\ O_P\left(\frac{1}{G}\sqrt{\sum_{g=1}^{G}\frac{1}{N_g^2}}\right), & \text{for weak dependence}\end{cases}$$

现在，对于第二项，我们有

$$\mathbb{E}\left\|T_2\right\|=\mathbb{E}\left\|\left(\hat{\bar{\beta}}-\beta\right)\left(\hat{\bar{\beta}}-\beta\right)'\right\|=\frac{1}{G}\mathbb{E}\left[\operatorname{tr}\left(V_G\right)\right].$$

根据假设A5，

$$O_e\left(\lambda_{min}\left(V_G\right)\right)=O_e\left(\lambda_{max}\left(V_G\right)\right)=\begin{cases}O_e\left(1\right), & \text{for strong dependence}\\ O_e\left(\frac{1}{G}\sum_{g=1}^{G}\frac{h(N_g)}{N_g}\right), & \text{for semi-strong dependence}\\ O_e\left(\frac{1}{G}\sum_{g=1}^{G}\frac{1}{N_g}\right), & \text{for weak dependence}\end{cases}$$

由此得到

$$\left\|T_2\right\|=\begin{cases}O_P\left(\frac{1}{G}\right), & \text{for strong dependence}\\ O_P\left(\frac{1}{G^2}\sum_{g=1}^{G}\frac{h(N_g)}{N_g}\right), & \text{for semi-strong dependence}\\ O_P\left(\frac{1}{G^2}\sum_{g=1}^{G}\frac{1}{N_g}\right), & \text{for weak dependence}\end{cases}$$

注意

$$T_3=-\frac{1}{G}\sum_{g=1}^{G}\left(X_g'X_g\right)^{-1}X_g'\epsilon_g\left(\hat{\bar{\beta}}-\beta\right)'=-\left(\hat{\bar{\beta}}-\beta\right)\left(\hat{\bar{\beta}}-\beta\right)'=-T_2,$$

且 $T_4=T_3$. 因此，$\left\|T_3\right\|=\left\|T_2\right\|=\left\|T_4\right\|$. 根据范数的三角不等式

$$\frac{\left\|\hat{V}_G-V_G\right\|}{\left\|V_G\right\|}\leq\frac{1}{\left\|V_G\right\|}\left[\left\|T_1\right\|+\left\|T_2\right\|+\left\|T_3\right\|+\left\|T_4\right\|\right].$$

由引理2.2，我们有$\sum_{g=1}^{G}\frac{h^2(N_g)}{\left(\sum_{g=1}^{G}\frac{h\left(N_g\right)}{N_g}\right)^2}\frac{N_g^2}{N_g}\to 0$ 以及 $\frac{\sum_{g=1}^{G}\frac{1}{N_g^2}}{\left(\sum_{g=1}^{G}\frac{1}{N_g}\right)^2}\to 0$，当 $G\to\infty$，对于一般的近似平衡情形，以及 $O(G)$个簇近似平衡的不平衡情形。因此，

$$\frac{\left\|T_1\right\|}{\left\|V_G\right\|}=\begin{cases}O_P\left(\frac{1}{\sqrt{G}}\right), & \text{for strong dependence}\\ O_P\left(\frac{\sqrt{\sum_{g=1}^{G}\frac{h^2(N_g)}{N_g^2}}}{\sum_{g=1}^{G}\frac{h(N_g)}{N_g}}\right), & \text{for semi-strong dependence}\\ O_P\left(\frac{\sqrt{\sum_{g=1}^{G}\frac{1}{N_g^2}}}{\sum_{g=1}^{G}\frac{1}{N_g}}\right), & \text{for weak dependence}\end{cases}$$

$= o_P(1)$，对于任何类型的依赖。

此外，利用引理2.2，

$$\frac{\|T_2\|}{\|V_G\|} = O_P\left(\frac{1}{G}\right) = o_P(1), \quad \text{for any kind of dependence.}$$

因此，我们已经证明

$$\frac{\left\|\hat{V}_G - V_G\right\|}{\|V_G\|} = o_P(1),$$

ˆ 由此可知，对于近似平衡情形，在任何类型的依赖下， V_G 对 V_G相合。对于不平衡情形，上述结论适用于：(i) 通常的强依赖；(ii) 弱依赖和半强依赖，前提是 $O(G)$簇近似平衡。

■

为求完整起见，我们给出以下引理。

引理 6.1. 假设 A 是一个满行秩为 $m(\leq n)$的 $m \times n$ 矩阵，其中 m 有限。那么，对于任意正定矩阵 B $n \times n$，*(i)* $\lambda_{\min}(ABA') \geq \lambda_{\min}(AA')\lambda_{\min}(B)$ 以及 *(ii)* $\lambda_{\max}(ABA') \leq \lambda_{\max}(AA')\lambda_{\max}(B)$。

引理6.1 的证明

注意

$$\lambda_{\min}(ABA') = \min_{z\in\mathbb{R}^k} \frac{z'ABA'z}{z'z} \geq \min_{z\in\mathbb{R}^k} \frac{z'AA'z}{z'z} \lambda_{\min}(B) \geq \lambda_{\min}(AA')\lambda_{\min}(B),$$

and

$$\lambda_{\max}(ABA') = \max_{z\in\mathbb{R}^k} \frac{z'ABA'z}{z'z} \leq \max_{z\in\mathbb{R}^k} \frac{z'AA'z}{z'z} \lambda_{\max}(B) \leq \lambda_{\max}(AA')\lambda_{\max}(B).$$

■

定理的证明2.4

ˆ我们分两个步骤展示基于 β 的 Wald 型检验统计量的渐近分布。步骤 *1*：在此步骤中，我们证明在 H_0，条件下

$$\Pi_G^{-1/2}\sqrt{G}\left(R\hat{\beta} - r\right) \xrightarrow{d} N_q(0, I_q), \quad \text{as } G \to \infty$$

(i) 对于近似平衡情形；(ii) 对于不平衡情形且具有强依赖；(iii) 对于不平衡情形且具有弱依赖或半强依赖，如果 $O(G)$ 个簇近似平衡。

首先注意，在 H_0，条件下

$$\Pi_G^{-1/2}\sqrt{G}\left(R\hat{\beta} - r\right) = \Pi_G^{-1/2}\sqrt{G}R\left(\hat{\beta} - \beta\right) = \Pi_G^{-1/2}RV_G^{1/2}V_G^{-1/2}\sqrt{G}\left(\hat{\beta} - \beta\right).$$

注意，定理2.2 给出了 $V_G^{-1/2}\sqrt{G}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N_k(0, I_k)$，(i) 对于近似平衡情形，(ii) 对于不平衡情形和强依赖，(iii) 对于不平衡情形以及弱依赖或半强依赖，前提是 $O(G)$ 个簇近似平衡。由于 R 是一个维数为 $q \times k$ 且满行秩为 q的有限矩阵，并且V_G 是有限且正定的，根据White (1984) 的推论 4.24 可知

$$\Pi_G^{-1/2}RV_G^{1/2}V_G^{-1/2}\sqrt{G}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N_q(0, I_q), \quad \text{as } G \to \infty$$

对于前面提到的情形。

步骤 2：在这一步中，我们证明在 H_0,

$$\hat{\Pi}_G^{-1/2}\sqrt{G}\left(R\hat{\beta}-r\right)\xrightarrow{d}N_q(0,I_q),\quad \text{as } G\to\infty.$$

为证明这一点，首先注意到，利用范数不等式，

$$\left\|\hat{\Pi}_G^{-1/2}\sqrt{G}\left(R\hat{\beta}-r\right)-\Pi_G^{-1/2}\sqrt{G}\left(R\hat{\beta}-r\right)\right\|\leq\left\|\hat{\Pi}_G^{-1/2}\Pi_G^{1/2}-I_l\right\|\left\|\Pi_G^{-1/2}\sqrt{G}\left(R\hat{\beta}-r\right)\right\|.$$

步骤 1，我们已经证明，在 H_0, $\Pi_G^{-1/2}\sqrt{G}\left(R\hat{\beta}-r\right)\xrightarrow{d}N_q(0,I_q)$, 由此可推得根据

$$\left\|\Pi_G^{-1/2}\sqrt{G}\left(R\hat{\beta}-r\right)\right\|=O_P(1),\quad \text{under } H_0.$$

因此，只需证明 $\left\|\hat{\Pi}_G^{-1/2}\Pi_G^{1/2}-I_l\right\|=o_P(1)$.

由于 $\Pi_G=RV_GR'$, 由引理 6.1 可知

$$\lambda_{\min}(\Pi_G)\geq\lambda_{\min}(RR')\lambda_{\min}(V_G)=O_e(1)\lambda_{\min}(V_G)$$
$$\lambda_{\max}(\Pi_G)\leq\lambda_{\max}(RR')\lambda_{\max}(V_G)=O_e(1)\lambda_{\max}(V_G).$$

由于根据假设 A5，$O_e(\lambda_{\min}(V_G))=O_e(\lambda_{\max}(V_G))$，因此我们有 $\|\Pi_G\|=O_e(\|V_G\|)$。在定理 2.3 的证明中，我们已经证明 $\frac{1}{\|V_G\|}\left\|\hat{V}_G-V_G\right\|=o_P(1)$，这意味着 $O_e\left(\frac{1}{\|V_G\|}\right)\left\|R\hat{V}_GR'-RV_GR'\right\|=o_P(1)$，进而意味着 $\frac{1}{\|\Pi_G\|}\left\|\hat{\Pi}_G-\Pi_G\right\|=o_P(1)$。根据假设 A5，我们推得 $\frac{V_G}{\|V_G\|}$ 一致正定，因此 $\frac{\Pi_G}{\|\Pi_G\|}$ 也一致正定。于是，根据概率收敛连续性定理，将 $g(Z)=Z^{-1/2}\left(\frac{\Pi_G}{\|\Pi_G\|}\right)^{1/2}$ 视为 Z 的连续函数，并令 $Z=\hat{\frac{\Pi_G}{\|\Pi_G\|}}$，可得 $\left\|\hat{\Pi}_G^{-1/2}\Pi_G^{1/2}-I_l\right\|=o_P(1)$。

因此，我们已经证明，在 H_0下，

$$\hat{\Pi}_G^{-1/2}\sqrt{G}\left(R\hat{\beta}-r\right)\xrightarrow{d}N_q(0,I_q),\quad \text{as } G\to\infty.$$

因此，由连续映射定理可得

$$G\left(R\hat{\beta}-r\right)'\hat{\Pi}_G^{-1}\left(R\hat{\beta}-r\right)\xrightarrow{d}\chi_q^2,\quad \text{as } G\to\infty$$

（i）对于近似平衡情形，（ii）对于不平衡情形和强依赖，（iii）对于不平衡情形以及弱依赖和半强依赖，如果 $O(G)$个簇近似平衡。这就证明了该定理。

■

定理的证明3.1

为证明 β 对模型（3.7）的一致性，首先注意到，由（3.9）

$$\hat{\beta}-\beta=\frac{1}{G}\sum_{g=1}^{G}u_g+\frac{1}{G}\sum_{g=1}^{G}\left(X_g'X_g\right)^{-1}X_g'\epsilon_g.$$

因此，$\hat{\beta}$ 根据假设 A1 和 A8是无偏的。因此，为证明 β的一致性，只需证明$\mathbb{E}\left\|\hat{\beta}-\beta\right\|^2\to0$，当 $G\to\infty$. 利用 ϵ_g与 u_g之间由假设 A7所给出的独立性，

$$\mathbb{E}\left\|\hat{\beta}-\beta\right\|^2=\operatorname{tr}\left(\mathbb{E}\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'\right]\right)=\operatorname{tr}(V^*),$$

其中 $V^* = V + \frac{1}{G}\Delta$ 如（3.10）中所定义。由定理 2.1的证明，我们已经表明，对于任何形式的依赖，在假设 A1、A2和 A3下，

$$\operatorname{tr}(V) = O\left(\frac{1}{G}\right),$$

无论簇大小如何。此外，在假设A8下，

$$\operatorname{tr}(\Delta) = O(1).$$

因此，

$$\mathbb{E}\left\|\hat{\bar{\beta}} - \beta\right\|^2 = O\left(\frac{1}{G}\right),$$

这意味着，对于ϵ_g 的任意依赖形式， $\hat{\bar{\beta}}$ 都是 β 的相合估计量，且与簇大小无关。

■

定理3.2的证明 为证明 $V^{*-1/2}\left(\hat{\bar{\beta}} - \beta\right)$ 的渐近正态性，首先注意，$\hat{\bar{\beta}} - \beta = \frac{1}{G}\sum_{g=1}^{G}\left(X_g'X_g\right)^{-1}X_g'\epsilon_g + \frac{1}{G}\sum_{g=1}^{G}u_g$ 如（3.9）所示。假设 $w_g = \left(X_g'X_g\right)^{-1}X_g'\epsilon_g$，我们得到

$$V^{*-1/2}\left(\hat{\bar{\beta}} - \beta\right) = V^{*-1/2}\left[\frac{1}{G}\sum_{g=1}^{G}w_g + \frac{1}{G}\sum_{g=1}^{G}u_g\right].$$

要证明渐近正态性，只需证明中心极限定理的李雅普诺夫条件。为证明李雅普诺夫条件，需要证明对于任意固定的 $z \in \mathbb{R}^k$，以及某个 $p > 1$,

$$\begin{aligned}
&\frac{\sum_{g=1}^{G}\mathbb{E}\left|z'V^{*-1/2}\frac{1}{G}(w_g + u_g)\right|^{2p}}{\left(\sum_{g=1}^{G}\mathbb{E}\left[z'V^{*-1/2}\frac{1}{G}(w_g + u_g)\right]^2\right)^p} \to 0, \quad \text{as } G \to \infty \\
\Longrightarrow\ &\frac{\sum_{g=1}^{G}\mathbb{E}\left|z'V^{*-1/2}(w_g + u_g)\right|^{2p}}{\left(\sum_{g=1}^{G}\mathbb{E}\left[z'V^{*-1/2}(w_g + u_g)\right]^2\right)^p} \to 0, \quad \text{as } G \to \infty.
\end{aligned}$$

在第二行使用迹不等式、在第四行使用不等式 $|a+b+c|^p \le 3^{p-1}(|a|^p + |b|^p + |c|^p)$，注意到

$$\begin{aligned}
\sum_{g=1}^{G}\mathbb{E}\left|z'V^{*-1/2}(w_g + u_g)\right|^{2p} &= \sum_{g=1}^{G}\mathbb{E}\left[\left|z'V^{*-1/2}(w_g + u_g)(w_g + u_g)'V^{*-1/2}z\right|^p\right] \\
&\le \sum_{g=1}^{G}\mathbb{E}\left[\left|z'V^{*-1}z\right|^p\left|\operatorname{tr}\left((w_g + u_g)(w_g + u_g)'\right)\right|^p\right] \\
&= \left|z'V^{*-1}z\right|^p\sum_{g=1}^{G}\mathbb{E}\left[\left|w_g'w_g + u_g'u_g + 2w_g'u_g\right|^p\right] \\
&\le \left|z'V^{*-1}z\right|^p\sum_{g=1}^{G}3^{p-1}\left(\mathbb{E}\left|w_g'w_g\right|^p + \mathbb{E}\left|u_g'u_g\right|^p + \mathbb{E}\left|2w_g'u_g\right|^p\right).
\end{aligned}$$

根据迹不等式，

$$\begin{aligned}
\mathbb{E}\left|w_g'w_g\right|^p &\le \mathbb{E}\left[\left|\operatorname{tr}\left(X_g\left(X_g'X_g\right)^{-2}X_g'\right)\right|^p\left|\epsilon_g'\epsilon_g\right|^p\right] \\
&= \mathbb{E}\left[\left|\operatorname{tr}\left(\left(X_g'X_g\right)^{-1}\right)\right|^p\mathbb{E}\left[\left|\epsilon_g'\epsilon_g\right|^p|X\right]\right].
\end{aligned}$$

由于 $\mathbb{E}\left[\left|\epsilon_g'\epsilon_g\right|^p|X\right]=O\left(N_g^p\right)$ 在 g 上一致地成立，这可由定理 2.2证明中的（6.19）结合假设 A4推出，因此由假设 A3

$$\mathbb{E}\left|w_g'w_g\right|^p=O(1),\quad \text{uniformly in } g.$$

在第二行使用 H¨older 不等式，并在第四行使用假设A9

$$\begin{aligned}\mathbb{E}\left|u_g'u_g\right|^p&=\mathbb{E}\left|\sum_{j=1}^k u_{gj}^2\right|^p\\&\leq\left(\sum_{j=1}^k 1^{p^*}\right)^{p/p^*}\mathbb{E}\left[\sum_{j=1}^k\left|u_{gj}\right|^{2p}\right],\quad \text{where } \frac{1}{p}+\frac{1}{p^*}=1\\&=k^{p/p^*}\sum_{j=1}^k\mathbb{E}\left[\left|u_{gj}\right|^{2p}\right]\\&=O\left(k^p\right)=O(1),\quad \text{uniformly in } g.\end{aligned}$$

在第一行应用柯西-施瓦茨不等式，根据假设 w_g 与 u_g 之间的独立性，在第二行使用假设A7，并整理前面各项的阶数，可得

$$\begin{aligned}\mathbb{E}\left|w_g'u_g\right|^p&\leq\mathbb{E}\left[\left|w_g'w_g\right|^{1/2}\left|u_g'u_g\right|^{1/2}\right]^p\\&=\mathbb{E}\left[\left|w_g'w_g\right|^{p/2}\right]\mathbb{E}\left[\left|u_g'u_g\right|^{p/2}\right]\\&=O(1),\quad \text{uniformly in } g.\end{aligned}$$

因此，对于某个 $0<c_1<\infty$，根据引理2.1

$$\sum_{g=1}^G\mathbb{E}\left|z'V^{*-1/2}\left(w_g+u_g\right)\right|^{2p}\leq c_1G\left|z'V^{*-1}z\right|^p.$$

注意，对于任意 $z\in\mathbb{R}^k$,

$$\frac{z'V^{*-1}z}{z'z}\leq\lambda_{max}\left(V^{*-1}\right)=\frac{1}{\lambda_{min}\left(V^*\right)}.$$

由于 $\lambda_{min}(V^*)\geq\lambda_{min}(V)+\lambda_{min}\left(\frac{1}{G}\underline{\Delta}\right)\geq\lambda_{min}\left(\frac{1}{G}\underline{\Delta}\right)\geq\frac{c_2}{G}$，根据假设A8 $0<c_2<\infty$，对于某个，可得

$$\frac{z'V^{*-1}z}{z'z}\leq\frac{G}{c_2}.$$

对于分母，将 V^* 的表达式（见3.10）代入第二行，得到

$$\begin{aligned}\left(\sum_{g=1}^G\mathbb{E}\left[z'V^{*-1/2}\left(w_g+u_g\right)\right]^2\right)^p&=\left(z'V^{*-1/2}\sum_{g=1}^G\mathbb{E}\left[\left(w_g+u_g\right)\left(w_g+u_g\right)'\right]V^{*-1/2}z\right)^p\\&=\left(z'V^{*-1/2}G^2V^*V^{*-1/2}z\right)^p\\&=G^{2p}\left(z'z\right)^p.\end{aligned}$$

因此，李雅普诺夫条件可简化为

$$\frac{\sum_{g=1}^G\mathbb{E}\left|z'V^{*-1/2}\frac{1}{G}\left(w_g+u_g\right)\right|^{2p}}{\left(\sum_{g=1}^G\mathbb{E}\left[z'V^{*-1/2}\frac{1}{G}\left(w_g+u_g\right)\right]^2\right)^p}\leq c_1G^{1-2p}\left|\frac{z'V^{*-1}z}{z'z}\right|^p\leq\frac{c_1}{c_2}G^{2(1-p)},$$

该值趋于 0，当 $G \to \infty$，对于某个 $p > 1$. 因此，对于强依赖、半强依赖和弱依赖，以及任意簇大小，$V^{*-1/2}\left(\hat{}\bar{\beta} - \beta\right)$ 满足中心极限定理，。

■

定理证明3.3

ˆ要证明 V_G^* 关于 V_G^*，是相合的，只需证明

$$\mathbb{E}\left[\frac{\|\hat{V}_G^* - V_G^*\|}{\|V_G^*\|}\right] \to 0, \quad \text{as } G \to \infty.$$

ˆˆ– –为此，首先注意到 $e_g = Y_g - X_g\beta = \epsilon_g + X_g u_g - X_g(\beta - \beta)$，这意味着

$$\begin{aligned} e_g e_g' =& (\epsilon_g + X_g u_g)(\epsilon_g + X_g u_g)' + X_g(\hat{\beta} - \beta)(\hat{\beta} - \beta)' X_g' \\ & - X_g(\hat{\beta} - \beta)(\epsilon_g + X_g u_g)' - (\epsilon_g + X_g u_g)(\hat{\beta} - \beta)' X_g'. \end{aligned}$$

ˆ然后， $V_G^* - V_G^*$ 可以表示为:

$$\begin{aligned} \hat{V}_G^* - V_G^* &= \frac{1}{G}\sum_{g=1}^{G} \left(X_g' X_g\right)^{-1} X_g' e_g e_g' X_g \left(X_g' X_g\right)^{-1} - V_G^* \\ &= T_1 + T_2 + T_3 + T_4, \quad \text{say,} \end{aligned}$$

其中

$$T_1 = \frac{1}{G}\sum_{g=1}^{G} (X_g' X_g)^{-1} X_g' \left[(\epsilon_g + X_g u_g)(\epsilon_g + X_g u_g)'\right] X_g (X_g' X_g)^{-1} - V_G^*,$$

$$T_2 = \frac{1}{G}\sum_{g=1}^{G} (\hat{\beta} - \beta)(\hat{\beta} - \beta)',$$

$$T_3 = -\frac{1}{G}\sum_{g=1}^{G} (\hat{\beta} - \beta)(\epsilon_g + X_g u_g)' X_g (X_g' X_g)^{-1}, \quad \text{and}$$

$$T_4 = -\frac{1}{G}\sum_{g=1}^{G} (X_g' X_g)^{-1} X_g' (\epsilon_g + X_g u_g)(\hat{\beta} - \beta)'.$$

首先，注意到根据假设A1、假设A7 和假设A8

$$\mathbb{E}\mathbb{E}\left[T_1\right] = \mathbb{E}\left[\frac{1}{G}\sum_{g=1}^{G} (X_g' X_g)^{-1} X_g' \left[(\epsilon_g + X_g u_g)(\epsilon_g + X_g u_g)'\right] X_g (X_g' X_g)^{-1} \middle| X\right] - V_G^* = 0.$$

现在，条件于 X,

$$T_1 = \frac{1}{G}\sum_{g=1}^{G} (X_g' X_g)^{-1} X_g' \left[(\epsilon_g + X_g u_g)(\epsilon_g + X_g u_g)' - (\Omega_g + X_g \Delta X_g')\right] X_g (X_g' X_g)^{-1}$$

假设 $A_g = (\epsilon_g + X_g u_g)(\epsilon_g + X_g u_g)' - (\Omega_g + X_g \Delta X_g')$，并利用假设A2下簇之间的独立性， $\mathbb{E}\,\mathrm{tr}[T_1 T_1']$ 可写为

$$\mathbb{E}\,\mathrm{tr}[T_1 T_1'] = \frac{1}{G^2}\sum_{g=1}^{G} \mathbb{E}\left[(X_g' X_g)^{-1} X_g' A_g X_g (X_g' X_g)^{-2} X_g' A_g' X_g (X_g' X_g)^{-1}\right]$$

$$\leq \frac{1}{G^2}\sum_{g=1}^{G}\mathbb{E}\left[\operatorname{tr}\left(X_g(X_g'X_g)^{-2}X_g'\right)\cdot\mathbb{E}\operatorname{tr}[(X_g'X_g)^{-1}X_g'A_gA_g'X_g(X_g'X_g)^{-1}\big|X]\right], \quad \text{by trace inequality}$$

$$\leq \frac{1}{G^2}\sum_{g=1}^{G}\mathbb{E}\left[\operatorname{tr}^2\left((X_g'X_g)^{-1}\right)\cdot\mathbb{E}\operatorname{tr}[A_gA_g'\big|X]\right], \quad \text{by trace inequality.}$$

Now,

$$\mathbb{E}\operatorname{tr}[A_gA_g'] = \mathbb{E}\operatorname{tr}\left[((\epsilon_g + X_gu_g)(\epsilon_g + X_gu_g)' - (\Omega_g + X_g\Delta X_g'))^2\right]$$

展开平方并取涉及四阶矩的期望后,

$$\begin{aligned}
&\mathbb{E}\operatorname{tr}[A_gA_g'] \\
&= \mathbb{E}\left[\{(\epsilon_g + X_gu_g)'(\epsilon_g + X_gu_g)\}^2\right] - \operatorname{tr}[(\Omega_g + X_g\Delta X_g')^2] \\
&= \mathbb{E}\left[(\epsilon_g'\epsilon_g)^2 + (u_g'X_g'X_gu_g)^2 + 4(\epsilon_g'X_gu_g)^2 + 2\epsilon_g'\epsilon_gu_g'X_g'X_gu_g + 2u_g'X_g'X_gu_g\epsilon_g'X_gu_g + 2\epsilon_g'\epsilon_g\epsilon_g'X_gu_g\right] \\
&\quad - \operatorname{tr}[(\Omega_g + X_g\Delta X_g')^2].
\end{aligned}$$

注意，根据假设 A4和假设 A9，上述所有项均为 $O(N_g^2$。应用不等式$\mathbb{E}\|T_1\| \leq (\mathbb{E}\operatorname{tr}[T_1T_1'])^{1/2}$，利用假设 A3 和引理 2.1，可得

$$\begin{aligned}
\mathbb{E}\|T_1\| &\leq \frac{1}{G}\left(\sum_{g=1}^{G}O(N_g^{-2})\cdot O(N_g^2)\right)^{1/2} \\
&= O\left(\frac{1}{\sqrt{G}}\right), \quad \text{for any kind of dependence.}
\end{aligned}$$

ˆ ˆ– –注意， $T_2 = (\beta-\beta)(\beta-\beta)'$。由于 $V_G^* = V_G + \Delta$，因此根据假设 A5 和假设 A8,

$$\mathbb{E}\|T_2\| = \mathbb{E}\left|(\hat{\bar{\beta}}-\beta)'(\hat{\bar{\beta}}-\beta)\right| = \operatorname{tr}(V^*) = \frac{1}{G}\operatorname{tr}(V_G^*) = O\left(\frac{1}{G}\right).$$

注意,

$$T_3 = -\frac{1}{G}\sum_{g=1}^{G}(\hat{\bar{\beta}}-\beta)(\epsilon_g + X_gu_g)'X_g(X_g'X_g)^{-1} = -\left(\hat{\bar{\beta}}-\beta\right)\left(\hat{\bar{\beta}}-\beta\right)' = -T_2,$$

同理， $T_4 = T_3$. 因此， $\|T_3\| = \|T_2\| = \|T_4\|$. 根据范数的三角不等式

$$\frac{\left\|\hat{V}_G^* - V_G^*\right\|}{\|V_G^*\|} \leq \frac{1}{\|V_G^*\|}\left[\|T_1\| + \|T_2\| + \|T_3\| + \|T_4\|\right].$$

此外， $\|V_G^*\| = \|V_G + \Delta\| \geq \|\Delta\| = O_e(1)$，根据假设 A8。因此,

$$\mathbb{E}\left[\frac{\|\hat{V}_G^* - V_G^*\|}{\|V_G^*\|}\right] = O\left(\frac{1}{\sqrt{G}}\right) \to 0, \quad \text{as } G \to \infty.$$

因此，ˆ V_G^* 是 V_G^* 的相合估计量，在模型（3.7）下。

■

定理的证明3.4

ˆ我们分两个步骤说明基于 β 的模型（3.7）的 Wald 型检验统计量的渐近分布。步骤 *1*：在此步骤中，我们说明在 H_0,

$$\Gamma_G^{-1/2}\sqrt{G}\left(R\hat{\bar{\beta}} - r\right) \xrightarrow{d} N_q(0, I_q), \quad \text{as } G \to \infty.$$

首先注意，在 H_0,

$$\Gamma_G^{-1/2}\sqrt{G}\left(R\hat{\bar{\beta}}-r\right)=\Gamma_G^{-1/2}\sqrt{G}R\left(\hat{\bar{\beta}}-\beta\right)=\Gamma_G^{-1/2}RV_G^{*1/2}V_G^{*-1/2}\sqrt{G}\left(\hat{\bar{\beta}}-\beta\right).$$

注意，定理 3.2 给出 $V_G^{*-1/2}\sqrt{G}\left(\hat{\bar{\beta}}-\beta\right)\xrightarrow{d}N_k(0,\ I_k)$。此外，$V_G^*=V_G+\Delta$ 是有限且正定的，这是因为 V_G 和 Δ 根据假设 A5 和假设A8，分别都是有限且正定的。由于 R 是一个维度为 $q\times k$ 且满行秩为 q 的有限矩阵，并且 V_G^* 有限且正定，因此根据 White 的推论 4.24（White (1984)）可得

$$\Gamma_G^{-1/2}RV_G^{*1/2}V_G^{*-1/2}\sqrt{G}\left(\hat{\bar{\beta}}-\beta\right)\xrightarrow{d}N_q(0,I_q),\quad \text{as } G\to\infty.$$

步骤 *2*：在这一步中，我们证明在 H_0,

$$\hat{\Gamma}_G^{-1/2}\sqrt{G}\left(R\hat{\bar{\beta}}-r\right)\xrightarrow{d}N_q(0,I_q),\quad \text{as } G\to\infty.$$

为证明这一点，首先注意，利用范数不等式,

$$\left\|\hat{\Gamma}_G^{-1/2}\sqrt{G}\left(R\hat{\bar{\beta}}-r\right)-\Gamma_G^{-1/2}\sqrt{G}\left(R\hat{\bar{\beta}}-r\right)\right\|\leq\left\|\hat{\Gamma}_G^{-1/2}\Gamma_G^{1/2}-I_l\right\|\left\|\Gamma_G^{-1/2}\sqrt{G}\left(R\hat{\bar{\beta}}-r\right)\right\|.$$

由步骤 1 可知，在 H_0，$\Gamma_G^{-1/2}\sqrt{G}\left(R\hat{\bar{\beta}}-r\right)\xrightarrow{d}N_q(0,\ I_q)$，从而可得

$$\left\|\Gamma_G^{-1/2}\sqrt{G}\left(R\hat{\bar{\beta}}-r\right)\right\|=O_P(1),\quad \text{under } H_0.$$

因此，只需证明 $\left\|\hat{\Gamma}_G^{-1/2}\Gamma_G^{1/2}-I_l\right\|=o_P(1)$。

由于 $\Gamma_G=RV_G^*R'$，由引理6.1 可得

$$\lambda_{\min}\left(\Gamma_G\right)\geq\lambda_{\min}\left(RR'\right)\lambda_{\min}\left(V_G^*\right)=O_e(1)\lambda_{\min}\left(V_G^*\right)$$
$$\lambda_{\max}\left(\Gamma_G\right)\leq\lambda_{\max}\left(RR'\right)\lambda_{\max}\left(V_G^*\right)=O_e(1)\lambda_{\max}\left(V_G^*\right).$$

在假设A5 和A8下，以下结论对任意类型的依赖均成立：

$$\lambda_{\max}(V_G^*)\leq\lambda_{\max}(V_G)+\lambda_{\max}(\Delta)=O_e(1),\quad \text{and}$$
$$\lambda_{\min}(V_G^*)\geq\lambda_{\min}(V_G)+\lambda_{\min}(\Delta)\geq\lambda_{\min}(\Delta)=O_e(1).$$

因此，$O_e\left(\lambda_{\min}\left(V_G^*\right)\right)=O_e\left(\lambda_{\max}\left(V_G^*\right)\right)$ 对任意类型的依赖均成立，因此，$\|\Gamma_G\|=O_e\left(\|V_G^*\|\right)$。在定理3.3的证明中，我们已证明 $\frac{1}{\|V_G^*\|}\left\|\hat{V}_G^*-V_G^*\right\|=o_P(1)$，这意味着

$$\frac{1}{O_e\left(\|V_G^*\|\right)}\left\|R\hat{V}_G^*R'-RV_G^*R'\right\|=o_P(1),$$

这进一步意味着 $\frac{1}{\|\Gamma_G\|}\left\|\hat{\Gamma}_G-\Gamma_G\right\|=o_P(1)$。根据假设A5和A8，我们推得 $\frac{V_G^*}{\|V_G^*\|}$ 是一致正定的，因此 $\frac{\Gamma_G}{\|\Gamma_G\|}$ 也一致正定。于是，由概率收敛连续性定理，将 $g(Z)=Z^{-1/2}\left(\frac{\Gamma_G}{\|\Gamma_G\|}\right)^{1/2}$ 视为 Z 的连续函数，并令 $Z=\hat{\frac{\Gamma_G}{\|\Gamma_G\|}}$，可得 $\left\|\hat{\Gamma}_G^{-1/2}\Gamma_G^{1/2}-I_l\right\|=o_P(1)$。因此

，我们已经证明，在 H_0 条件下 ，

$$\hat{\Gamma}_G^{-1/2}\sqrt{G}\left(R\hat{\bar{\beta}}-r\right)\xrightarrow{d}N_q(0,I_q),\quad \text{as } G\to\infty.$$

因此，根据连续映射定理可得

$$G\left(R\hat{\tilde{\beta}}-r\right)'\hat{\Gamma}_G^{-1}\left(R\hat{\tilde{\beta}}-r\right)\xrightarrow{d}\chi_q^2,\quad \text{as } G\to\infty.$$

■

引理的证明3.1

˜要证明 V_l 对于 $V_{G,\ l}$，是相合的，只需证明

$$\mathbb{E}\left[\frac{\left\|\tilde{V}_l-V_{G,l}\right\|}{\|V_{G,l}\|}\right]\to 0,\quad \text{as } G\to\infty.$$

为证明这一点，首先注意到 $e_g^*=Y_g-X_g\tilde{\beta}_l=\epsilon_g-X_g\left(\tilde{\beta}_l-\beta\right)$，这意味着

$$e_g^*e_g^{*\prime}=\epsilon_g\epsilon_g'+X_g\left(\tilde{\beta}_l-\beta\right)\left(\tilde{\beta}_l-\beta\right)'X_g'-\epsilon_g\left(\tilde{\beta}_l-\beta\right)'X_g'-X_g\left(\tilde{\beta}_l-\beta\right)\epsilon_g'.$$

˜于是，$V_l-V_{G,\ l}$ 可以表示为:

$$\tilde{V}_l-V_{G,l}=T_1+T_2+T_3+T_4,\quad \text{say,}$$

其中

$$T_1=\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\left(\left(X_g'X_g\right)^{-1}X_g'\epsilon_g\epsilon_g'X_g\left(X_g'X_g\right)^{-1}-\mathbb{E}\left[\left(X_g'X_g\right)^{-1}X_g'\Omega_gX_g\left(X_g'X_g\right)^{-1}\right]\right),$$

$$T_2=\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\left(\tilde{\beta}_l-\beta\right)\left(\tilde{\beta}_l-\beta\right)',$$

$$T_3=-\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\left(X_g'X_g\right)^{-1}X_g'\epsilon_g\left(\tilde{\beta}_l-\beta\right)',\quad \text{and}$$

$$T_4=-\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\left(\tilde{\beta}_l-\beta\right)\epsilon_g'X_g\left(X_g'X_g\right)^{-1}.$$

现在，在给定 X，的条件下，注意到

$$\begin{aligned}\mathbb{E}\left[T_1\right]&=\mathbb{E}\left[T_1\middle|X\right]\\&=\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\mathbb{E}\left[\left(X_g'X_g\right)^{-1}X_g'\mathbb{E}\left[\epsilon_g\epsilon_g'\middle|X\right]X_g\left(X_g'X_g\right)^{-1}-\mathbb{E}\left[\left(X_g'X_g\right)^{-1}X_g'\Omega_gX_g\left(X_g'X_g\right)^{-1}\right]\right]\\&=0,\end{aligned}$$

并利用假设A2中的簇间独立性以及第三行中的范数不等式，我们得到

$$\begin{aligned}\mathbb{E}\left\|T_1\right\|^2&=\mathbb{E}\left\|\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\left(X_g'X_g\right)^{-1}\left(X_g'\epsilon_g\epsilon_g'X_g-X_g'\Omega_gX_g\right)\left(X_g'X_g\right)^{-1}\right\|^2\\&=\frac{1}{P_l^4}\sum_{g\in\mathcal{S}_l}\mathbb{E}\left[\left\|\left(X_g'X_g\right)^{-1}X_g'\left(\epsilon_g\epsilon_g'-\Omega_g\right)X_g\left(X_g'X_g\right)^{-1}\right\|^2\middle|X\right]\\&\le\frac{1}{P_l^4}\sum_{g\in\mathcal{S}_l}\mathbb{E}\left[\left\|\left(X_g'X_g\right)^{-1}\right\|^4\mathbb{E}\left[\left\|X_g'\left(\epsilon_g\epsilon_g'-\Omega_g\right)X_g\right\|^2\middle|X\right]\right]\end{aligned}$$

现在，利用 $\mathbb{E}\left[\left\|\left(X_g'X_g\right)^{-1}\right\|^4\right]=O\left(N_g^{-4}\right)$ 在 g 上一致成立（根据假设A3），并应用假设A6和引理2.1，我们得到

$$\mathbb{E}\left\|T_1\right\|^2=\begin{cases}\frac{1}{P_l^4}\sum_{g\in\mathcal{S}_l}O\left(N_g^{-4}\right)O\left(N_g^4\right)=O\left(\frac{1}{P_l^3}\right), & \text{for strong dependence}\\ \frac{1}{P_l^4}\sum_{g\in\mathcal{S}_l}O\left(N_g^{-4}\right)O\left(N_g^2h^2\left(N_g\right)\right)=O\left(\frac{1}{P_l^4}\sum_{g\in\mathcal{S}_l}\frac{h^2(N_g)}{N_g^2}\right), & \text{for semi-strong dependence}\\ \frac{1}{P_l^4}\sum_{g\in\mathcal{S}_l}O\left(N_g^{-4}\right)O\left(N_g^2\right)=O\left(\frac{1}{P_l^4}\sum_{g\in\mathcal{S}_l}\frac{1}{N_g^2}\right), & \text{for weak dependence}\end{cases}$$

因此，

$$\left\|T_1\right\|=\begin{cases}O_P\left(\frac{1}{P_l^{3/2}}\right), & \text{for strong dependence}\\ O_P\left(\frac{1}{P_l^2}\sqrt{\sum_{g\in\mathcal{S}_l}\frac{h^2(N_g)}{N_g^2}}\right), & \text{for semi-strong dependence}\\ O_P\left(\frac{1}{P_l^2}\sqrt{\sum_{g\in\mathcal{S}_l}\frac{1}{N_g^2}}\right), & \text{for weak dependence}\end{cases}$$

对 l.一致地。现在，对于第二项，我们有

$$\mathbb{E}\left\|T_2\right\|=\mathbb{E}\left\|\frac{1}{P_l}\left(\tilde{\beta}_l-\beta\right)\left(\tilde{\beta}_l-\beta\right)'\right\|=\frac{1}{P_l}\mathbb{E}\left[\operatorname{tr}\left(V_{G,l}\right)\right].$$

根据假设假设A2，A3以及引理2.1，

$$\mathbb{E}\left[\operatorname{tr}\left(V_{G,l}\right)\right]=\begin{cases}O\left(\frac{1}{P_l}\right), & \text{for strong dependence}\\ O\left(\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}\right), & \text{for semi-strong dependence}\\ O\left(\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}\right), & \text{for weak dependence}\end{cases}$$

对 l.一致地。由此可得

$$\left\|T_2\right\|=\begin{cases}O_P\left(\frac{1}{P_l^2}\right), & \text{for strong dependence}\\ O_P\left(\frac{1}{P_l^3}\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}\right), & \text{for semi-strong dependence}\\ O_P\left(\frac{1}{P_l^3}\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}\right), & \text{for weak dependence}\end{cases}$$

注意到

$$T_3=-\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\left(X_g'X_g\right)^{-1}X_g'\epsilon_g\left(\hat{\tilde{\beta}}-\beta\right)'=-T_2.$$

以及 $T_4=T_3$. 因此，$\left\|T_3\right\|=\left\|T_2\right\|=\left\|T_4\right\|$. 根据范数的三角不等式，

$$\frac{\left\|\tilde{V}_l-V_{G,l}\right\|}{\left\|V_{G,l}\right\|}\leq\frac{1}{\left\|V_{G,l}\right\|}\left[\left\|T_1\right\|+\left\|T_2\right\|+\left\|T_3\right\|+\left\|T_4\right\|\right].$$

根据假设 A10，

$$\left\|V_{G,l}\right\|=\begin{cases}O_e\left(\frac{1}{P_l}\right), & \text{for strong dependence}\\ O_e\left(\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}\right), & \text{for semi-strong dependence}\\ O_e\left(\frac{1}{P_l^2}\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}\right), & \text{for weak dependence}\end{cases}$$

因此，利用引理2.2

$$\frac{\|T_1\|}{\|V_{G,l}\|} = \begin{cases} O_P\left(\frac{1}{\sqrt{P_l}}\right), & \text{for strong dependence} \\ O_P\left(\frac{\sqrt{\sum_{g\in\mathcal{S}_l}\frac{h^2(N_g)}{N_g^2}}}{\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}}\right), & \text{for semi-strong dependence} \\ O_P\left(\frac{\sqrt{\sum_{g\in\mathcal{S}_l}\frac{1}{N_g^2}}}{\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}}\right), & \text{for weak dependence} \end{cases}$$

$$= o_P(1), \quad \text{for any kind of dependence, since } P_l \to \infty.$$

此外，对于任何类型的依赖，

$$\frac{\|T_2\|}{\|V_{G,l}\|} = O_P\left(\frac{1}{P_l}\right) = o_P(1), \quad \text{as } P_l \to \infty.$$

因此，对于任何类型的误差依赖，

$$\frac{\left\|\tilde{V}_l - V_{G,l}\right\|}{\|V_{G,l}\|} = o_P(1), \quad \text{as } P_l \to \infty.$$

˜这意味着 V_l 对于 V_G是相合的, l，对于每个 $l = 1$，2、.....、D.

■

定理3.5的证明

我们分两个步骤证明检验统计量的渐近正态性。

步骤 *1*：令 $T_{SB}^* = \sum_{l=1}^{D}(\tilde{\beta}_l - \hat{\bar{\beta}})' V_{G\,l}^{-1}$，$(\tilde{\beta}_l - \hat{\bar{\beta}})$。我们的目标是证明

$$Z^* = \frac{T_{SB}^* - kD}{\sqrt{2kD}} \xrightarrow[H_0]{d} N(0,1), \quad \text{as } \min_l P_l \to \infty,\ D \to \infty.$$

展开二次型：

$$\begin{aligned} T_{SB}^* &= \sum_{l=1}^{D}(\tilde{\beta}_l - \beta)' V_{G,l}^{-1}(\tilde{\beta}_l - \beta) + \sum_{l=1}^{D}(\hat{\bar{\beta}} - \beta)' V_{G,l}^{-1}(\hat{\bar{\beta}} - \beta) - 2\sum_{l=1}^{D}(\tilde{\beta}_l - \beta)' V_{G,l}^{-1}(\hat{\bar{\beta}} - \beta) \\ &= A_D + B_D + C_D, \quad \text{say.} \end{aligned}$$

我们首先分析这一项 A_D. 。按照定理 2.2的证明，可以很容易地看出，在假设A1、A2、A3、A4 和 A10下，

$$V_{G,l}^{-1/2}(\tilde{\beta}_l - \beta) \xrightarrow{d} N(0, I_k), \quad \text{as } P_l \to \infty, \tag{6.20}$$

对于：(i) 一般的强依赖；(ii) 半强依赖，若对于某个 $p > 1$，$\frac{P_l}{\left[\sum_{g\in\mathcal{S}_l}\frac{h(N_g}{N_g}\right]^p} \to 0$)，~~当$P_l \to \infty$.~~ (iii) 弱依赖，若对于某个 $p > 1$，$\frac{P_l}{\left[\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}\right]^p} \to 0$，当 P_l ~~$\to \infty$.~~

因此，根据连续映射定理，对于固定的 D, $A_D \xrightarrow{d} \chi^2_{kD}$。所以，利用独立卡方分布之和的中心极限定理，我们得到

$$\frac{A_D - kD}{\sqrt{2kD}} \xrightarrow{d} N(0,1), \quad \text{as } \min_l P_l \to \infty \text{ and } D \to \infty.$$

对于项 B_D，注意，根据迹不等式

$$\mathbb{E}\left|B_D\right| = \mathbb{E}\left[\sum_{l=1}^{D}(\hat{\tilde{\beta}}-\beta)'V_{G,l}^{-1}(\hat{\tilde{\beta}}-\beta)\right] = \sum_{l=1}^{D}\operatorname{tr}\left(V_{G,l}^{-1}\,\mathbb{E}\left[(\hat{\tilde{\beta}}-\beta)(\hat{\tilde{\beta}}-\beta)'\right]\right) \le \sum_{l=1}^{D}\operatorname{tr}(V_{G,l}^{-1})\ \operatorname{tr}(V).$$

根据假设A5、假设A10和引理2.1，

$$\begin{aligned}
\mathbb{E}\left|B_D\right| &\le \sum_{l=1}^{D}\operatorname{tr}(V_{G,l}^{-1})\ \operatorname{tr}(V)\\
&= \begin{cases}
\sum_{l=1}^{D} O(P_l)\cdot O(\frac{1}{G}) & \text{, for strong dependence}\\
\sum_{l=1}^{D} O\left(\frac{P_l^2}{\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}}\right)\cdot O\left(\frac{1}{G^2}\sum_{g=1}^{G}\frac{h(N_g)}{N_g}\right) & \text{, for semi-strong dependence}\\
\sum_{l=1}^{D} O\left(\frac{P_l^2}{\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}}\right)\cdot O\left(\frac{1}{G^2}\sum_{g=1}^{G}\frac{1}{N_g}\right) & \text{, for weak dependence}
\end{cases}\\
&= \begin{cases}
O\left(1\right) & \text{, for strong dependence}\\
O\left(\sum_{l=1}^{D}\frac{P_l^2}{\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}}\cdot\frac{1}{G^2}\sum_{l=1}^{D}\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}\right) & \text{, for semi-strong dependence}\\
O\left(\sum_{l=1}^{D}\frac{P_l^2}{\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}}\cdot\frac{1}{G^2}\sum_{l=1}^{D}\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}\right) & \text{, for weak dependence}
\end{cases}
\end{aligned}$$

So,

$$\frac{B_D}{\sqrt{2kD}} = \begin{cases}
O_P\left(\frac{1}{\sqrt{D}}\right) & \text{, for strong dependence}\\
O_P\left(\frac{1}{\sqrt{D}}\sum_{l=1}^{D}\frac{P_l^2}{\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}}\cdot\frac{1}{G^2}\sum_{l=1}^{D}\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}\right) & \text{, for semi-strong dependence}\\
O_P\left(\frac{1}{\sqrt{D}}\sum_{l=1}^{D}\frac{P_l^2}{\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}}\cdot\frac{1}{G^2}\sum_{l=1}^{D}\sum_{g\in\mathcal{S}_l}\frac{1}{N_g}\right) & \text{, for weak dependence}
\end{cases}$$

根据假设A11，有：

$$\frac{1}{\sqrt{D}}\sum_{l=1}^{D}\frac{P_l^2}{\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}}\cdot\frac{1}{G^2}\sum_{l=1}^{D}\sum_{g\in\mathcal{S}_l}\frac{h(N_g)}{N_g}\to 0,\quad \text{as } D\to\infty.$$

在弱依赖情形下，通过令 $h(N_g)=1.$，该表达式同样成立。因此，对于任意类型的依赖，$\frac{B_D}{\sqrt{2kD}}=o_p(1)$ 均成立。

对于项 C_D，第一行应用三角不等式，第二行应用柯西-施瓦茨不等式

$$\begin{aligned}
\mathbb{E}\left|C_D\right| &\le 2\mathbb{E}\left[\sum_{l=1}^{D}\left|(\tilde{\beta}_l-\beta)'V_{G,l}^{-1}(\hat{\tilde{\beta}}-\beta)\right|\right]\\
&\le 2\sqrt{A_D^2}\sqrt{B_D^2}.
\end{aligned}$$

由于 $\frac{A_D-kD}{\sqrt{2kD}}\xrightarrow{d}N(0,\ 1)$，且 $\frac{B_D}{\sqrt{2kD}}=o_P(1)$，我们得到 $\frac{C_D}{\sqrt{2kD}}=o_P(1)$。因此，

$$\frac{T_{SB}^*-kD}{\sqrt{2kD}} = \underbrace{\frac{A_D-kD}{\sqrt{2kD}}}_{\xrightarrow{d}N(0,1)} + \underbrace{\frac{B_D}{\sqrt{2kD}}}_{\xrightarrow{p}0} + \underbrace{\frac{C_D}{\sqrt{2kD}}}_{\xrightarrow{p}0}$$

这意味着 $Z^*\xrightarrow{d}N(0,\ 1)$，因为 $P_l\to\infty$，$D\to\infty$.

步骤 2：在这一步中，我们证明当 $D\to\infty$ 时，$Z-Z^*\xrightarrow{P}0$ 与 D 最小值 $P_l\to 0.$。注意

$$|Z-Z^*| = \left|\frac{T_{SB}-T_{SB}^*}{\sqrt{2kD}}\right|$$

$$
\begin{aligned}
&\leq \frac{1}{\sqrt{2kD}} \sum_{l=1}^{D} \left| (\tilde{\beta}_l - \hat{\tilde{\beta}})' (\tilde{V}_l^{-1} - V_{G,l}^{-1}) (\tilde{\beta}_l - \hat{\tilde{\beta}}) \right| \\
&= \frac{1}{\sqrt{2kD}} \sum_{l=1}^{D} \left| (\tilde{\beta}_l - \hat{\tilde{\beta}})' V_{G,l}^{-1/2} \left[V_{G,l}^{1/2} \tilde{V}_l^{-1} V_{G,l}^{1/2} - I_k \right] V_{G,l}^{-1/2} (\tilde{\beta}_l - \hat{\tilde{\beta}}) \right| \\
&\leq \frac{1}{\sqrt{2kD}} \sum_{l=1}^{D} (\tilde{\beta}_l - \hat{\tilde{\beta}})' V_{G,l}^{-1} (\tilde{\beta}_l - \hat{\tilde{\beta}}) \cdot \left\| V_{G,l}^{1/2} \tilde{V}_l^{-1} V_{G,l}^{1/2} - I_k \right\|
\end{aligned}
$$

根据引理 3.1 的证明，我们有

$$
\frac{\left\| \tilde{V}_l - V_{G,l} \right\|}{\|V_{G,l}\|} = O_P\left(\frac{1}{\sqrt{P_l}} \right), \quad \text{uniformly in } l.
$$

此外，由于 $\frac{V_{G,l}}{\|V_{G,l}\|}$ 是一致正定的，根据概率收敛连续性定理，取 $g(W) = \left(\frac{V_{G,l}}{\|V_{G,l}\|} \right)^{1/2} W^{-1} \left(\frac{V_{G,l}}{\|V_{G,l}\|} \right)^{1/2}$ 并令 $W = \frac{\tilde{V}_l}{\|V_{G,l}\|}$，我们有

$$
\|V_{G,l}^{1/2} \tilde{V}_l^{-1} V_{G,l}^{1/2} - I_k\| = O_P\left(\frac{1}{\sqrt{P_l}} \right), \quad \text{uniformly in } l.
$$

因此，利用第二行中的范数不等式

$$
\begin{aligned}
|Z - Z^*| &\leq \frac{1}{\sqrt{2kD}} \sum_{l=1}^{D} (\tilde{\beta}_l - \hat{\tilde{\beta}})' V_{G,l}^{-1} (\tilde{\beta}_l - \hat{\tilde{\beta}}) \cdot \left\| V_{G,l}^{1/2} \tilde{V}_l^{-1} V_{G,l}^{1/2} - I_k \right\| \\
&\leq \max_l \left\| V_{G,l}^{1/2} \tilde{V}_l^{-1} V_{G,l}^{1/2} - I_k \right\| \cdot \frac{T_{SB}^*}{\sqrt{2kD}} \\
&= O_P\left(\frac{1}{\min_l P_l} \right) \cdot O_P\left(\sqrt{D} \right) \\
&= O_P\left(\frac{D}{\min_l P_l} \right).
\end{aligned}
$$

因此，若 $\frac{D}{\min_l P_l} \to 0$，$Z$ 和 Z^* 具有相同的渐近分布，且

$$
Z \xrightarrow[H_0]{d} N(0,1).
$$

■

结果3.1的证明

这里，在给定 X，的条件下，

$$
V = \frac{1}{G^2} \sum_{g=1}^{G} (X_g' X_g)^{-1} X_g' \Omega_g X_g (X_g' X_g)^{-1} = \frac{1}{G^2} \sum_{g=1}^{G} \frac{a_g X_g' X_g}{a_g^2} = \frac{1}{G},
$$

and

$$
\Sigma = \left(\sum_{g=1}^{G} X_g' X_g \right)^{-1} \sum_{g=1}^{G} X_g' \Omega_g X_g \left(\sum_{g=1}^{G} X_g' X_g \right)^{-1} = \frac{\sum_{g=1}^{G} a_g^2}{\left(\sum_{g=1}^{G} a_g \right)^2}.
$$

由于 $\left(\sum_{g=1}^{G} a_g \right)^2 \leq G \sum_{g=1}^{G} a_g^2$ 根据柯西-施瓦茨不等式，我们有

$$
\frac{1}{G} \leq \frac{\sum_{g=1}^{G} a_g^2}{\left(\sum_{g=1}^{G} a_g \right)^2}.
$$

因此得证。

■

结果的证明3.2

给定 Ω_g，在给定 X的条件下，POLS估计量的方差为

$$\begin{aligned}\Sigma &= \left(\sum_{g=1}^{G} X_g' X_g\right)^{-1} \left((a-b)\sum_{g=1}^{G} X_g' X_g + b\sum_{g=1}^{G} X_g' \mathbb{1}\mathbb{1}' X_g\right) \left(\sum_{g=1}^{G} X_g' X_g\right)^{-1} \\ &= \frac{(a-b)\sum_{g=1}^{G} N_g + b\sum_{g=1}^{G} N_g^2}{\left(\sum_{g=1}^{G} N_g\right)^2} \\ &= \frac{a-b}{\sum_{g=1}^{G} N_g} + \frac{b\sum_{g=1}^{G} N_g^2}{\left(\sum_{g=1}^{G} N_g\right)^2}.\end{aligned}$$

另一方面，β 的方差为

$$V = \frac{1}{G^2}\sum_{g=1}^{G} (X_g' X_g)^{-1} X_g' \Omega_g X_g (X_g' X_g)^{-1} = \frac{1}{G^2}\sum_{g=1}^{G} \frac{aN_g + bN_g(N_g-1)}{N_g^2} = \frac{(a-b)}{G^2}\sum_{g=1}^{G} \frac{1}{N_g} + \frac{b}{G}.$$

因此，我们有 $\mathrm{Var}(\hat{\beta}_{POLS}) = \Sigma = O\left(\frac{N_1^2+G}{N_1+G}()^2\right)$，以及 $\mathrm{Var}(\hat{\bar{\beta}}) = V = O(\frac{1}{G})$。于是，当 N_1 和 G充分大时，$\hat{\bar{\beta}}$相对于$\hat{\beta}_{POLS}$的效率为

$$\begin{aligned}\frac{\Sigma}{V} &= O\left(\frac{GN_1^2+G^2}{(N_1+G)^2}\right) \\ &= \begin{cases} O\left(\frac{G+\frac{G^2}{N_1^2}}{\left(1+\frac{G}{N_1}\right)^2}\right) & = O(G), \quad \text{if } \frac{N_1}{G} \to \infty \text{ or } N_1 = O(G) \\ O\left(\frac{\frac{N_1^2}{G}+1}{\left(\frac{N_1}{G}+1\right)^2}\right) & = \begin{cases} O\left(N_1^{1-\delta}\right), & \text{if } \frac{N_1}{G} = o(1) \text{ and } G = O\left(N_1^{1+\delta}\right), \text{ for } 0<\delta<1 \\ \to c, & \text{if } \frac{N_1}{G} = o(1) \text{ and } G = O_e\left(N_1^2\right), \text{ for } 1<c<\infty \\ \to 1, & \text{if } \frac{N_1}{G} = o(1) \text{ and } G = O\left(N_1^{2+\delta_1}\right), \text{ for } \delta_1 > 0 \end{cases} \end{cases}\end{aligned}$$

因此，除非 $\frac{N_1^2}{G} \to 0$，否则$\hat{\bar{\beta}}$ 渐近更有效于$\hat{\beta}_{POLS}$。 —

■

引理 6.2。设 $\{A_n\}$ 是一列有限维矩阵，满足 $A_n = B_n + o_P(1)$，其中 B_n在 n. 上有限且一致正定。则，$A_n = O_P(1)$。

引理的证明6.2

由于 B_n 是有限的，存在常数 $M > 0$ 使得 $\|B_n\| \le M$，对所有 n. B_n 是一致正定意味着存在常数 $c > 0$，使得对所有非零向量 v，$v^\top B_n v \ge c\|v\|^2$，关于 n一致地成立。根据随机阶的加法性质，$O_P(1) + o_P(1) = O_P(1)$，见Van der Vaart（2000）的第 2.2 节。为证明 $A_n = O_P(1)$，只需证明 $B_n = O_P(1)$，即我们证明 B_n 的每个元素$(B_n)_{ij}$ 都是 $O_P(1)$。

根据定义 2.2，随机变量序列 Z_n 是 $O_P(1)$，如果对于每个 $\varepsilon > 0$，都存在 $M_\varepsilon > 0$ 和$L_\varepsilon \in \mathbb{N}$ 使得

$$\mathbb{P}(|Z_n| > M_\varepsilon) < \varepsilon, \quad \text{for all } n \ge L_\varepsilon.$$

在这种情况下， B_n 是确定性的，因此概率陈述就变成了确定性界。由于对于所有 n， $\|B_n\| \leq M$，因此对于任意条目（B_n）$_{ij}$，

$$|(B_n)_{ij}| \leq \|B_n\| \leq M.$$

这是因为矩阵中任意元素的绝对值都不超过其范数。现在，给定 $\varepsilon > 0$，并选取 $M_\varepsilon = M$， $L_\varepsilon = 1$. 于是，对所有 $n \geq L_\varepsilon$，以及所有 i， j，

$$\mathbb{P}\left(|(B_n)_{ij}| > M_\varepsilon\right) = \mathbb{P}\left(|(B_n)_{ij}| > M\right) = 0 < \varepsilon.$$

这满足 B_n 中每个元素的 $O_P(1)$ 定义。由于 B_n 的每个元素都被 M 所界（且与 n 无关），我们得出 $B_n = O_P(1)$。因此，我们有

$$A_n = O_P(1).$$

■

定理3.6的证明

ˆ为了证明 β_{POLS} 的一致性，首先注意，在模型（3.15）下

$$\hat{\beta}_{POLS} - \beta = (X'X)^{-1} \sum_{g=1}^{G} X_g' \eta_g = (X'X)^{-1} \left[X'\epsilon + \sum_{g=1}^{G} X_g' X_g u_g \right].$$

注意， $\left(\frac{X'X}{\sum_{g=1}^{G} N_g}\right)^{-1} - C_G^{-1} = o_P(1)$，这由假设 A12 得出，其中 C_G^{-1} 在 G. 上有限且一致正定。然后，根据引理 6.2，

$$\left(\frac{X'X}{\sum_{g=1}^{G} N_g}\right)^{-1} = O_P(1).$$

现在，根据假设A1，注意到

$$\mathbb{E}\left[\frac{X'\epsilon}{\sum_{g=1}^{G} N_g}\right] = \mathbb{E}\mathbb{E}\left[\frac{X'\epsilon}{\sum_{g=1}^{G} N_g}\middle| X\right] = \frac{1}{\sum_{g=1}^{G} N_g}\mathbb{E}\left[X'\mathbb{E}\left[\epsilon|X\right]\right] = 0.$$

此外，根据假设A2所述的聚类间独立性，向量 $\frac{X'\epsilon}{\sum_{g=1}^{G} N_g}$ 的方差-协方差矩阵为

$$\mathbb{E}\left[\mathrm{Var}\left(\frac{X'\epsilon}{\sum_{g=1}^{G} N_g}\middle| X\right)\right] = \frac{\sum_{g=1}^{G} \mathbb{E}\left[X_g'\Omega_g X_g\right]}{\left(\sum_{g=1}^{G} N_g\right)^2}.$$

由于这是一个半正定矩阵，

$$\left\| \frac{\sum_{g=1}^{G} \mathbb{E}\left[X_g'\Omega_g X_g\right]}{\left(\sum_{g=1}^{G} N_g\right)^2} \right\|_e = \lambda_{\max}\left(\frac{\sum_{g=1}^{G} \mathbb{E}\left[X_g'\Omega_g X_g\right]}{\left(\sum_{g=1}^{G} N_g\right)^2} \right).$$

于是，根据结果6.1 和詹森不等式，可得

$$\lambda_{\max}\left(\frac{\sum_{g=1}^{G} \mathbb{E}\left[X_g'\Omega_g X_g\right]}{\left(\sum_{g=1}^{G} N_g\right)^2} \right) \leq \frac{\sum_{g=1}^{G} \lambda_{\max}\left(\mathbb{E}\left[X_g'\Omega_g X_g\right]\right)}{\left(\sum_{g=1}^{G} N_g\right)^2}$$

$$\leq \frac{\sum_{g=1}^{G}\mathbb{E}\left[\lambda_{\max}\left(X_g'\Omega_g X_g\right)\right]}{\left(\sum_{g=1}^{G}N_g\right)^2}$$
$$\leq \frac{\sum_{g=1}^{G}\mathbb{E}\left[\lambda_{\max}\left(X_g'X_g\right)\lambda_{\max}\left(\Omega_g\right)\right]}{\left(\sum_{g=1}^{G}N_g\right)^2}.$$

利用$\mathbb{E}\left[\lambda_{\max}\left(X_g'X_g\right)\right]=O\left(N_g\right)$在 g 上一致地成立这一事实，根据假设A12，由假设A2 和引理2.1 可得

$$\left\|\mathrm{Var}\left(\frac{X'\epsilon}{\sum_{g=1}^{G}N_g}\right)\right\| = \begin{cases} O\left(\frac{\sum_{g=1}^{G}N_g^2}{\left(\sum_{g=1}^{G}N_g\right)^2}\right), & \text{for strong dependence} \\ O\left(\frac{\sum_{g=1}^{G}N_g h(N_g)}{\left(\sum_{g=1}^{G}N_g\right)^2}\right), & \text{for semi-strong dependence} \\ O\left(\frac{1}{\sum_{g=1}^{G}N_g}\right), & \text{for weak dependence} \end{cases}$$

现在，对于涉及 u_g 的项，注意根据假设 A7 和 A8

$$\mathbb{E}\left[\frac{1}{\sum_{g=1}^{G}N_g}\sum_{g=1}^{G}X_g'X_g u_g\right]=0,$$

$\frac{1}{\sum_{g=1}^{G}N_g}\sum_{g=1}^{G}X_g'X_g u_g$ 的方差-协方差矩阵为

$$\frac{1}{\left(\sum_{g=1}^{G}N_g\right)^2}\sum_{g=1}^{G}\mathbb{E}\left[X_g'X_g\Delta X_g'X_g\right].$$

由于这是一个半正定矩阵，

$$\left\|\frac{1}{\left(\sum_{g=1}^{G}N_g\right)^2}\sum_{g=1}^{G}\mathbb{E}\left[X_g'X_g\Delta X_g'X_g\right]\right\|_e=\lambda_{\max}\left(\frac{1}{\left(\sum_{g=1}^{G}N_g\right)^2}\sum_{g=1}^{G}\mathbb{E}\left[X_g'X_g\Delta X_g'X_g\right]\right).$$

那么，根据结果6.1 和詹森不等式，我们有

$$\lambda_{\max}\left(\frac{1}{\left(\sum_{g=1}^{G}N_g\right)^2}\sum_{g=1}^{G}\mathbb{E}\left[X_g'X_g\Delta X_g'X_g\right]\right) \leq \frac{\sum_{g=1}^{G}\lambda_{\max}\left(\mathbb{E}\left[X_g'X_g\Delta X_g'X_g\right]\right)}{\left(\sum_{g=1}^{G}N_g\right)^2}$$
$$\leq \frac{\sum_{g=1}^{G}\mathbb{E}\left[\lambda_{\max}\left(X_g'X_g\Delta X_g'X_g\right)\right]}{\left(\sum_{g=1}^{G}N_g\right)^2}$$
$$\leq \frac{\sum_{g=1}^{G}\mathbb{E}\left[\lambda_{\max}\left((X_g'X_g)^2\right)\lambda_{\max}\left(\Delta\right)\right]}{\left(\sum_{g=1}^{G}N_g\right)^2}.$$

利用这一事实：$\mathbb{E}\left[\lambda_{\max}\left(\left(X_g'X_g\right)^2\right)\right]=O\left(N_g^2\right)$在 g上一致地成立。根据假设 A12，由假设A8 和引理2.1 可知:

$$\left\|\mathrm{Var}\left(\frac{1}{\sum_{g=1}^{G}N_g}\sum_{g=1}^{G}X_g'X_g u_g\right)\right\|=O\left(\frac{\sum_{g=1}^{G}N_g^2}{\left(\sum_{g=1}^{G}N_g\right)^2}\right).$$

由于 u_g 和 ϵ_g 根据假设 A7相互独立分布，因此可得:

$$\left\| \operatorname{Var}\left(\frac{X'\epsilon + \sum_{g=1}^{G} X_g' X_g u_g}{\sum_{g=1}^{G} N_g} \right) \right\| = O\left(\frac{\sum_{g=1}^{G} N_g^2}{\left(\sum_{g=1}^{G} N_g\right)^2} \right),$$

ϵ_g的任意类型的依赖
注意，$\frac{\sum_{g=1}^{G} N_g^2}{(\sum_{g=1}^{G} N_g)_2} \to 0$，若 $\frac{\max_{g:g=1(1)G} N_g}{\sum_{g=1}^{G} N_g} \to 0.$。特别地，对于近似平衡情形，假设 $N_g \simeq N$，

$\frac{\sum_{g=1}^{G} N_g^2}{(\sum_{g=1}^{G} N_g)_2} \simeq \frac{GN^2}{G^2N^2} = \frac{1}{G} \to 0$，当 $G \to \infty.$ 时。因此，

$$\frac{X'\epsilon + \sum_{g=1}^{G} X_g' X_g u_g}{\sum_{g=1}^{G} N_g} = o_P(1),$$

若 (i) 簇近似平衡，或 (ii) 簇不平衡且 $\frac{\max_g N_g}{\sum_{g=1}^{G} N_g} \to 0$，当 $G \to \infty$。由于 $\left(\frac{X'X}{\sum_{g=1}^{G} N_g}\right)^{-1} = O_P(1)$，我们得到

$$\hat{\beta}_{POLS} - \beta = \left(\frac{X'X}{\sum_{g=1}^{G} N_g} \right)^{-1} \left[\frac{X'\epsilon + \sum_{g=1}^{G} X_g' X_g u_g}{\sum_{g=1}^{G} N_g} \right] = O_P(1) \cdot o_P(1) = o_P(1),$$

(i) 近似平衡簇或 (ii) 满足 $\frac{\max_g N_g}{\sum_{g=1}^{G} N_g} \to 0$ 的不平衡簇，当 $G \to \infty$。这就证明了该定理。

■

定理的证明3.7

首先注意，在模型（3.15）下

$$\hat{\beta}_{POLS} - \beta = (X'X)^{-1} \sum_{g=1}^{G} X_g' \eta_g = (X'X)^{-1} \left[X'\epsilon + \sum_{g=1}^{G} X_g' X_g u_g \right].$$

由于 $\left(\frac{X'X}{\sum_{g=1}^{G} N_g}\right)^{-1} - C_G^{-1} = o_P$（1）由假设 A12 得出，其中 C_G^{-1} 在 G上有限且一致正定，由引理 6.2,

$$\left(\frac{X'X}{\sum_{g=1}^{G} N_g} \right)^{-1} = O_P(1).$$

假设 $Z_G = \sum_{g=1}^{G} X_g' \eta_g$，且 $Z_G^* = \frac{Z_G}{\sum_{g=1}^{G} N_g}$. 由于 ϵ_g 和 u_g 根据假设相互独立A7,

$$\operatorname{Var}\left(X'\epsilon + \sum_{g=1}^{G} X_g' X_g u_g \right) = \operatorname{Var}(X'\epsilon) + \operatorname{Var}\left(\sum_{g=1}^{G} X_g' X_g u_g \right) \geq \operatorname{Var}\left(\sum_{g=1}^{G} X_g' X_g u_g \right).$$

现在，根据假设A8

$$\operatorname{Var}\left(\sum_{g=1}^{G} X_g' X_g u_g \right) = \sum_{g=1}^{G} X_g' X_g \Delta X_g' X_g .$$

请注意

$$\lambda_{\max}\left(\sum_{g=1}^{G} X_g' X_g \Delta X_g' X_g \right) \leq \lambda_{\max}\left(\sum_{g=1}^{G} (X_g' X_g)^2 \right) \lambda_{\max}(\Delta) \quad \text{and}$$

$$\lambda_{\min}\left(\sum_{g=1}^{G} X_g' X_g \Delta X_g' X_g\right) \geq \lambda_{\min}\left(\sum_{g=1}^{G} (X_g' X_g)^2\right) \lambda_{\min}(\Delta).$$

使用假设 A8 和 A12，我们得到

$$\|\mathrm{Var}\left[Z_G^*\right]\| = O_e\left(\frac{\sum_{g=1}^{G} N_g^2}{\left(\sum_{g=1}^{G} N_g\right)^2}\right).$$

因此， Z_G^* 的范数的二阶矩有一个严格大于零的下界，除非 $\frac{\sum_{g=1}^{G} N_g^2}{(\sum_{g=1}^{G} N_g)_2} \to 0$。 Z_G 的范数的 4th 阶矩为

$$\begin{aligned}
\mathbb{E}\left[\|Z_G\|^4\right] &= \mathbb{E}\left[\left(\sum_{g=1}^{G} \eta_g' X_g \sum_{g_1=1}^{G} X_{g_1}' \eta_{g_1}\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{g,g_1=1}^{G} (X_g u_g + \epsilon_g)' X_g X_{g_1}' (X_{g_1} u_{g_1} + \epsilon_{g_1})\right)^2\right] \\
&= \mathbb{E}\left[(T_1 + T_2 + 2\,T_3)^2\right] \\
&\leq 3\,\mathbb{E}\left[T_1^2 + T_2^2 + 4\,T_3^2\right], \text{ by the C-S inequality,}
\end{aligned}$$

其中 $T_1 = \sum_{g,\, g_1=1}^{G} \epsilon_g' X_g X_{g_1}' \epsilon_{g_1}$， $T_2 = \sum_{g,\, g_1=1}^{G} u_g' X_g' X_g X_{g_1}' X_{g_1} u_{g_1}$，以及 $T_3 = \sum_{g,\, g_1=1}^{G} \epsilon_g' X_g X_{g_1}' X_{g_1} u_{g_1}$.考虑第一项，注意到

$$\mathbb{E}\left[T_1^2\right] = \mathbb{E}\left[\left(\sum_{g,g_1=1}^{G} \epsilon_g' X_g X_{g_1}' \epsilon_{g_1}\right)^2\right] = \sum_{g,g_1,h,h_1=1}^{G} \mathbb{E}\left[\epsilon_g' X_g X_{g_1}' \epsilon_{g_1}\ \epsilon_h' X_h X_{h_1}' \epsilon_{h_1}\right].$$

根据假设A2，误差在各簇之间相互独立，因此，乘积的期望为零，除非（g， h）=（g_1， h_1）或（g， h）=（h_1， g_1）。因此，对于某个 $c > 0$,

$$\begin{aligned}
\mathbb{E}\left[T_1^2\right] &\leq c \sum_{g=1}^{G}\sum_{g_1=1}^{G} \mathbb{E}\left[\left(\epsilon_g' X_g X_{g_1}' \epsilon_{g_1}\right)^2\right] \\
&\leq c\left(\sum_{g=1}^{G} \mathbb{E}\left[\epsilon_g' X_g X_g' \epsilon_g\right]\right)^2, \text{ by the C-S inequality} \\
&= c\left(\sum_{g=1}^{G} \mathbb{E}\,\mathrm{tr}\left[X_g' \Omega_g X_g\right]\right)^2
\end{aligned}$$

由于 $\mathbb{E}\mathrm{tr}\left[X_g' \Omega_g X_g\right] \leq \mathbb{E}\mathrm{tr}\left[X_g' X_g\right] \lambda_{\max}(\Omega_g)$ 根据结果6.1成立，因此由假设A2、A12和引理2.1，

$$\mathbb{E}\left[T_1^2\right] = \begin{cases} O\left(\left(\sum_{g=1}^{G} N_g^2\right)^2\right), & \text{for strong dependence} \\ O\left(\left(\sum_{g=1}^{G} N_g h(N_g)\right)^2\right), & \text{for semi-strong dependence} \\ O\left(\left(\sum_{g=1}^{G} N_g\right)^2\right), & \text{for weak dependence} \end{cases}$$

对于项 T_2，请注意

$$\mathbb{E}\left[T_2^2\right] = \sum_{g,g_1,h,h_1=1}^{G} \mathbb{E}\left[u_g' X_g' X_g X_{g_1}' X_{g_1} u_{g_1}\ u_h' X_h' X_h X_{h_1}' X_{h_1} u_{h_1}\right].$$

由于 u_g 根据假设 A7 在不同簇之间相互独立，乘积的期望值只有在 $(g,\ h) = (g_1,\ h_1)$ 或 $(g,\ h) = (h_1,\ g_1)$ 时才不为零。因此，对于某个 $c_1 > 0$,

$$\begin{aligned}
\mathbb{E}\left[T_2^2\right] &\leq c_1 \sum_{g=1}^{G}\sum_{g_1=1}^{G} \mathbb{E}\left[\left(u_g' X_g' X_g X_{g_1}' X_{g_1} u_{g_1}\right)^2\right] \\
&\leq c_1 \left(\sum_{g=1}^{G} \mathbb{E}\left[u_g' X_g' X_g X_g' X_g u_g\right]\right)^2, \quad \text{by the C-S inequality} \\
&= c_1 \left(\sum_{g=1}^{G} \mathbb{E}\,\mathrm{tr}\left[X_g' X_g \Delta X_g' X_g\right]\right)^2 \\
&\leq c_1 \left(\sum_{g=1}^{G} \mathbb{E}\,\mathrm{tr}\left[(X_g' X_g)^2\right] \lambda_{\max}(\Delta)\right)^2, \quad \text{by Result 6.1}
\end{aligned}$$

利用假设 A8，A12 和引理 2.1，我们得到

$$\mathbb{E}\left[T_2^2\right] = O\left(\left(\sum_{g=1}^{G} N_g^2\right)^2\right).$$

对于 T_3，注意到

$$\mathbb{E}\left[T_3^2\right] = \sum_{g,g_1,h,h_1=1}^{G} \mathbb{E}\left[\epsilon_g' X_g X_{g_1}' X_{g_1} u_{g_1}\ \epsilon_h' X_h X_{h_1}' X_{h_1} u_{h_1}\right].$$

由于 ϵ_g 和 u_g在假设 A2 和 A7下均在不同簇之间相互独立，乘积的期望不为零，除非 $(g,\ h) = (g_1,\ h_1)$ 或 $(g,\ h) = (h_1,\ g_1)$。因此，对于某个 $c_2 > 0$,

$$\begin{aligned}
\mathbb{E}\left[T_3^2\right] &\leq c_2 \sum_{g=1}^{G}\sum_{g_1=1}^{G} \mathbb{E}\left[\left(\epsilon_g' X_g X_{g_1}' X_{g_1} u_{g_1}\right)^2\right] \\
&\leq c_2 \sum_{g=1}^{G}\sum_{g_1=1}^{G} \mathbb{E}\left[\epsilon_g' X_g X_g' \epsilon_g u_{g_1}' X_{g_1}' X_{g_1} X_{g_1}' X_{g_1} u_{g_1}\right], \quad \text{by the C-S inequality} \\
&= c_2 \sum_{g=1}^{G}\sum_{g_1=1}^{G} \mathbb{E}\left[\epsilon_g' X_g X_g' \epsilon_g\right] \mathbb{E}\left[u_{g_1}' X_{g_1}' X_{g_1} X_{g_1}' X_{g_1} u_{g_1}\right], \quad \text{by Assumption A7} \\
&= c_2 \sum_{g=1}^{G} \mathbb{E}\,\mathrm{tr}\left[X_g' \Omega_g X_g\right] \sum_{g_1=1}^{G} \mathbb{E}\,\mathrm{tr}\left[X_{g_1}' X_{g_1} \Delta X_{g_1}' X_{g_1}\right].
\end{aligned}$$

由于 $\mathbb{E}\,\mathrm{tr}\left[X_g' \Omega_g X_g\right] \leq \mathbb{E}\mathrm{tr}\left[X_g' X_g\right] \lambda_{\max}(\Omega g)$ 根据结果 6.1成立，由假设 A2、A8、A12和引理 2.1,

$$\mathbb{E}\left[T_3^2\right] = O\left(\left(\sum_{g=1}^{G} N_g^2\right)^2\right).$$

综合以上各项，我们得到

$$\mathbb{E}\left[\|Z_G\|^4\right] = O\left(\left(\sum_{g=1}^{G} N_g^2\right)^2\right).$$

如果我们假设 $Z_G^* = \frac{Z_G}{\sum_{g=1}^{G} N_g}$，那么

$$\mathbb{E}\left[\|Z_G^*\|^4\right] = O\left(\frac{\left(\sum_{g=1}^{G} N_g^2\right)^2}{\left(\sum_{g=1}^{G} N_g\right)^4}\right) = O(1).$$

因此， Z_G^* 的范数的四阶矩有界。我们还证明了，除非 $\frac{\sum_{g=1}^G N_g^2}{(\sum_{g=1}^G N_g)_2} \to 0$，否则 Z_G^* 的范数的二阶矩有正的下界；也就是说， Z_G^* 不在均方意义下收敛到 0。由于 $(2+\delta)^{th}$ 阶矩有界意味着 $\left\{\|Z_G^*\|^2 \text{ 一致可积}: G \in \mathbb{N}\right\}$，因此该族是一致可积的。所以，我们可以断言 Z_G^* 不依概率收敛到 0，因为在 $\mathscr{L}^p$ 中，依概率收敛和一致可积性意味着在 p^{th}阶矩意义下收敛。因此，

$$Z_G^* = \frac{\sum_{g=1}^G X_g' \eta_g}{\sum_{g=1}^G N_g} \neq o_P(1),$$

ˆ 由此可推得，除非 $\frac{\sum_{g=1}^G N_g^2}{(\sum_{g=1}^G N_g)^2} \to 0.$，否则在簇不平衡且存在强依赖时， β_{POLS} 对 β并不相合。

———————

■

参考文献

Abadie, A., Athey, S., Imbens, G. W. 和 Wooldridge, J. M.（2020）。回归分析中基于抽样与基于设计的不确定性。《计量经济学》，88(1):265–296。

Abadie, A., Athey, S., Imbens, G. W. 和 Wooldridge, J. M.（2023）。何时应针对聚类调整标准误？《经济学季刊》，138(1):1–35。

Andrews, D. W.（2005）。存在共同冲击时的截面回归。《计量经济学》，73(5):1551–1585。

Arellano, M.（1987年）。如何为组内估计量计算稳健标准误。《牛津经济学与统计学通报》，49(4):431–434。

Basak, G. K. 和 Das, S.（2018）。理解面板数据中的截面依赖。*arXiv* 预印本 *arXiv:1804.08326*。

Bester, C. A.、Conley, T. G. 和 Hansen, C. B.（2011）。使用簇协方差估计量对依赖数据进行推断。《计量经济学杂志》，165(2)：137–151。

Breitung, J.、Roling, C. 和 Salish, N.（2016）。面板数据模型中斜率同质性的拉格朗日乘数型检验。《计量经济学期刊》，19(2)：166–202。

Cameron, A. C.、Gelbach, J. B. 和 Miller, D. L.（2008）。针对聚类误差推断的基于自助法的改进。《经济学与统计学评论》，90(3)：414–427。

Cameron, A. C. 和 Miller, D. L.（2015）。聚类稳健推断实践者指南。《人力资源杂志》，50(2)：317–372。

Cameron, A. C. 和 Trivedi, P. K.（2005）。微观计量经济学：方法与应用。剑桥大学出版社。

Canay, I. A.、Santos、和 Shaikh, A. M.（2021）。少数“大型”簇下的野生自助法。《经济学与统计学评论》，103(2):346–363。

Chiang, H. D.、Sasaki, Y.、和 Wang, Y.（2023）。针对聚类数据的真正稳健推断。*arXiv* 预印本 *arXiv:2308.10138*。Conley, T. G.（1999）。具有截面依赖的广义矩估计。《计量经济学杂志》，92(1):1–45。Djogbenou, A. A.、MacKinnon, J. G.、和 Nielsen, M. Ø.（2019）。具有聚类误差的渐近理论与野生自助法推断。《计量经济学杂志》，212(2):393–412。Hansen, B.（2022）。计量经济学。普林斯顿大学出版社。Hansen, B. E. 和 Lee, S.（2019）。聚类样本的渐近理论。《计量经济学杂志》，210(2):268–290。Ibragimov, R. 和 M¨uller, U. K.（2016）。具有少数异质聚类的推断。《经济学与统计学评论》，98(1):83–96。MacKinnon, J. G.、Nielsen, M. Ø.、和 Webb, M. D.（2023）。聚类稳健推断：实证实践指南。《计量经济学杂志》，232(2):272–299。MacKinnon, J. G. 和 Webb, M.（2019）。何时以及如何处理回归模型中的聚类误差。技术报告，Queen’s Economics Department 工作论文。Moulton, B. R.（1986）。随机组效应与回归估计值的精确性。《计量经济学杂志》，32(3):385–397。Moulton, B. R.（1990）。关于估计聚合变量对微观单位影响时一个陷阱的示例。《经济学与统计学评论》，第334–338页。Pesaran, M. H. 和 Yamagata, T.（2008）。大型面板中的斜率同质性检验。《计量经济学杂志》，142(1):50–93。Spamann, H.（2019）。使用州公司法进行识别时的推断。哈佛法学院。Swamy, P. A.（1970）。随机系数回归模型中的有效推断。《计量经济学：计量经济学学会杂志》，第311–323页。Van der Vaart, A. W.（2000）。渐近统计学，第3卷。剑桥大学出版社。White, H.（1984）。面向计量经济学家的渐近理论。学术出版社。Wooldridge, J. M.（2003）。应用计量经济学中的聚类抽样方法。《美国经济评论》，93(2):133–138。Wooldridge, J. M.（2006）。应用计量经济学中的聚类抽样方法：扩展分析。密歇根州立大学油印稿。