arXiv:2603.02611v1 [stat.ME] 3 Mar 2026

一种面向调查加权有界计数的贝叶斯分层障碍贝塔-二项模型及其在儿童保育入托中的应用

JoonHo Lee

摘要。有界离散比例——已知总数中的计数——在数据呈现结构性零值、过度离散和分层聚类时带来建模挑战。我们提出带有州变化系数的贝叶斯分层障碍贝塔-二项模型，以应对全部四项特征。该框架做出三项方法论贡献：（i）通过跨块协方差分量研究跨边际依赖，并阐明该参数何时以及如何经由层次层而非条件似然得到识别；（ii）针对调查权重下的伪后验推断，提出基于乔列斯基分解的夹心方差校准，并以参数特定的设计效应比诊断加以引导；（iii）为障碍模型引入对数尺度边际效应分解，将回归系数转化为政策相关量。将该模型应用于来自2019年全国早期照护与教育调查、覆盖51个州的6,785家托儿服务提供机构后，模型揭示了“贫困反转”：贫困降低入托参与，却提高参与者中的参与强度，其中广延边际占总效应的三分之二。设计校准模拟表明，夹心校正区间显著提高覆盖率，在固定效应的90%名义水平上达到82–88.5%。R 软件包 hurdlebb 实现了全部方法。

关键词：贝叶斯分层模型；贝塔-二项分布；有界离散比例；障碍模型；夹心方差校正；调查加权

日期：2026年3月2日。Lee：阿拉巴马大学，美国阿拉巴马州塔斯卡卢萨。jlee296@ua.edu。本研究由美国卫生与公众服务部儿童与家庭管理局规划、研究与评估办公室（OPRE）通过向内华达大学拉斯维加斯分校提供的90YE0346号资助项目资助（项目负责人：Gerilyn Slicker），并由该校向阿拉巴马大学提供分项资助。文中所表达的观点仅代表作者本人，并不代表儿童与家庭管理局或美国卫生与公众服务部的观点。

1. 引言

有界离散比例——已知总数中的计数——出现在有限容量被部分利用的任何情形：医院床位被占用、保险计划被参保、牙齿表面发生龋坏、调查条目得到肯定回答，或托育中心的入学名额被填满。当结果变量因结构性而非抽样原因可以取零值，并且方差远超二项分布时，标准广义线性模型便不再适用。美国的婴幼儿儿童保育入托同时体现了所有这些复杂情形的典型实例。来自2019年全国早期照护与教育调查（全国早期照护与教育调查项目团队，2022年）的数据记录了每一家中心式服务提供者在总容量中三岁以下儿童的入学人数。超过三分之一的中心报告婴幼儿入学为零——并非因为没有需求，而是因为严格的师幼比和较高的人均儿童成本促使许多服务提供者完全拒绝为该年龄组提供服务（Adams & Rohacek，2010年）。由此产生的服务可及性差距不成比例地落在高贫困社区（Dobbins等人，2016年；Malik等人，2018年），然而贫困与入学之间的关系实际上远比简单的短缺叙事更为复杂。

对数据的初步考察揭示了一种鲜明的模式，我们将其称为贫困反转：社区贫困会降低中心提供婴幼儿服务的概率，但在确实参与服务的中心中，位于贫困程度较高地区的中心会将更大份额的容量用于这一年龄组。这种反转带来了方法学挑战：任何将参与和服务强度合并到单一方程中的模型都无法检测到这一现象；同样，无法同时处理有界支持集、过度离散、结构性零值和调查加权的模型也无法检测到它。这四项特征共同促使本文构建了层级障碍贝塔-二项框架。

针对存在过多零值的计数数据，两部分模型已有相当长的发展历史。Cragg（1971）为连续支出引入了门槛框架，Mullahy（1986）则将其改用于计数数据，把二元参与阶段与截断计数阶段区分开来。贝叶斯扩展大大拓宽了这类模型的应用范围：采用泊松分布核和负二项核的层级零膨胀模型或门槛模型，如今能够处理空间、纵向和多层结构（Altinisik，2023年；B. Neelon，2019；B. Neelon 等人，2013、2016年；B. H. Neelon等人，2010年）。与本文最相关的是，Ghosal 等人（2020年）提出了一种用于时空计数数据的层级门槛模型，通过单一共享标量捕捉跨边际依赖。

然而，所有这些模型都假设采用泊松分布核或负二项核，因而将计数视为无界。当响应变量具有有界支持集时，贝塔-二项分布是自然的替代方案，它在保持 $\{0, 1, \ldots, n\}$ 支持集有界的同时引入过度离散（Wagner 等人，2015年）。Bandyopadhyay 等人（2011年）将贝塔-二项分布纳入空间框架，以处理聚类有界计数；Wen 等人（2025）则提出了带有空间变系数和跨组件协方差的零膨胀贝塔-二项分布模型。然而，他们采用的是零膨胀形式，即零值来自潜在混合；而在儿童保育场景中，报告婴儿入托人数为零的中心是经过审慎考虑后作出的运营决策（参见 Hall，2000；Lambert，1992），因此门槛模型设定才是自然的选择。

另一个独立的方法学分支关注调查数据中贝叶斯模型的基于设计的推断。Savitsky 和 Toth（2021年）提出了一个伪后验框架，将调查权重纳入似然指数；Williams 和 Sav- itsky（2021）则提供了经校准的三明治方差校正，从而恢复名义可信区间覆盖率（建立在 Binder，1983年；Pfeffermann，1993的基础上）。这些研究针对的是标准的单方程回归；而将其扩展到两部分层级设定尚未有人尝试。总而言之，现有文献分别提供了门槛、贝塔-二项分布、层级结构和调查校正这些组成部分，但尚未将它们综合起来。

本文构建了这一综合框架。该模型沿两个方向扩展了 Ghosal 等人（2020年）的框架，每个方向都针对现有方法的一项局限。

首先，我们提出一个具有随州变化的系数和跨边际协方差的贝叶斯分层障碍贝塔-二项模型。障碍分量将参与与强度分离；贝塔-二项核遵循有界支持集，并容纳所观测到的过度离散，其超出二项方差约十二倍；而对州级系数的层级先验识别出一个跨边际协方差结构，其有效样本量为$S = 51$ 个州，而非 $N = 6$、785 名服务提供者，且LKJ 先验为高维协方差提供有限样本正则化（$2q = 10$ 个参数）。该模型将Ghosal 等人（2020年）的标量跨边际参数作为特例纳入，并通过纳入针对结构性零值的门槛，扩展了Bandyopadhyay 等人（2011年）和Wen 等人（2025年）的贝塔-二项分布模型。应用于NSECE 数据时，该设定揭示了这一反转及其在51个州之间的地理异质性，并给出完整的不确定性量化。

其次，我们将Savitsky 和 Toth（2016年）以及Williams 和 Savitsky（2021）的伪后验框架扩展至两部分层级设定，实施基于 Cholesky 的夹心方差校正（Binder，1983年；Pfeffermann，1993），并将设计效应比作为参数特异性诊断引入。一项配套模拟研究（第4节）提供了坦率的诊断：在现实的设计效应下，夹心校正区间显著提高了固定效应的覆盖率（在90%名义水平上为82–88.5%），而方差分量则需要非加权估计——我们透明地报告这一发现，而不是将夹心估计量视为普遍适用的补救措施。

这两项创新共同实现了边际效应分解（命题5），将反转分解为广延边际（可及性）和集约边际（强度）两个部分。单方程设定会将这两种相互抵消的作用混为一谈，从而得到较弱的净效应；两部分分解则量化了各作用渠道及其相对大小，其中广延边际约占总效应的三分之二。配套的 R 软件包 hurdlebb （Lee，2026）实现了完整的方法。

本文其余部分安排如下。第2节介绍 NSECE 数据，并阐述促使我们构建该模型的经验模式。第3节构建层级障碍 Beta-二项设定，并介绍其解析性质及三明治校正。第4节报告模拟研究结果。第5节将模型应用于 2019年NSECE数据。第6节讨论相关启示与局限性。补充材料 A–E 提供证明、识别条件、扩展表格和计算细节。

2. 研究动机：数据与初步分析

本节介绍数据，并记录共同决定第3节所构建模型的实证特征。第3节我们的介绍围绕四种分布特征展开——结构性零值、有界支持集、过度离散和复杂抽样设计，它们分别直接对应层级障碍贝塔-二项模型设定中的一个组成部分。随后，初步分析证实了反转现象和州际异质性，这两者正是引入层级结构和跨边际协方差的动因。

2.1. NSECE 调查。

2019年全国早期照护与教育调查（NSECE；全国早期照护与教育调查项目团队，2022年）是一项具有全国代表性的早期照护与教育服务提供者调查，由

芝加哥大学 NORC 受规划、研究与评估办公室（OPRE）委托开展。该办公室隶属于儿童与家庭管理局。数据收集于2018年11月至2019年7月进行，共有6,917名中心式服务提供者完成调查。中心式服务提供者部分采用分层多阶段整群抽样设计：按州分组划分为30个层，包含415个初级抽样单位，以及6,809个有非缺失调查答复的中心。其中，24个中心报告的总入学人数为 $n_i = 0$ ，因贝塔-二项分布核要求分母为正数而予以排除，最终分析样本包括来自全部 $S = 51$ 个州（50个州和哥伦比亚特区）的$N = 6{,}785$名服务提供者。有关抽样框、权重构建和信息性检验的完整细节，见补充材料 D。

对于每个服务提供者 i，调查记录其总入学人数 n_i （0–5岁）以及婴幼儿（IT）入学人数 $Y_i \in \{0, 1, \ldots, n_i\}$ （0–2岁）。结果变量对(Y_i, n_i)是自然的分析单位。服务提供者层面的协变量包括服务提供者所在人口普查区的社区贫困率、城市化程度和种族／族裔构成（Dobbins等人，2016年；Malik等人，2018年）。州级政策变量涵盖儿童保育与发展基金（CCDF）补贴政策的三个维度——市场费率百分位、分级报销以及婴幼儿费率附加额，并作为跨层调节变量纳入第5节的模型中。表1总结了关键描述性统计数据。

2.2. **四项数据特征及其对建模的影响**。现在，我们考察NSECE 数据的四项分布特征，这些特征分别及共同决定了第3节中的模型设定。图1展示了婴幼儿入托比例的边际分布，该分布同时呈现了全部四项特征。**结构性零值与障碍模型设定**。在6,785家服务提供者中，2,393家（35.3%）报告婴幼儿入学人数为零——这一比例远高于抽样变异所能解释的程度，反映了服务提供者有意决定不接收三岁以下儿童（Adams & Rohacek，2010年）。障碍模型设定（Mullahy，1986）将参与和强度视为由不同参数支配的独立过程，详见第3.2节。

有界支持集与贝塔-二项核。在服务机构（$z_i = 1$）中，婴幼儿入园人数 Y_i 的上限为总入学人数 n_i，因此份额Y_i/n_i位于$(0, 1]$。贝塔-二项分布符合这一 $\{0, 1, \ldots, n_i\}$ 有界支持集，并通过集中参数κ (Prentice，1986；Wagner 等人，2015年)纳入超二项变异；参见第3.2节。

表1. 数据摘要：NSECE 2019 中心式服务提供者。A面板描述结果变量；B面板描述服务提供者层面的协变量（连续变量为标准化前的值，以均值、标准差和范围表示）；C面板描述调查设计特征，包括 Kish 设计效应 $\text{DEFF}_{\text{Kish}} = 1 + \text{CV}^2(w)$。

变量	数值
*A*面板：结果变量	
服务提供者总数（N）	6,785
IT服务提供者（$z_i = 1$）	4,392（64.7%）
非服务提供者（$z_i = 0$）	2,393（35.3%）
总入学人数（n_i）：均值 / 中位数 / 范围	60.5 / 48 / 1–378
IT入托（$Y_i \mid z_i = 1$）：均值 / 中位数	29.4 / 24
IT占比（$Y_i/n_i \mid z_i = 1$）：均值（标准差）	0.478 (0.214)
*B*面板：协变量 — 均值（标准差） [范围]	
社区贫困率（%）	17.3 (8.3) [1.9, 54.2]
城市（%）	93.3
社区黑人比例	17.2 (20.5) [0.0, 96.5]
社区西班牙裔比例	24.2 (22.6) [0.0, 98.4]
面板 *C*：调查设计	
州（S）	51（50 +哥伦比亚特区）
层 / 初级抽样单位	30 / 415
抽样权重范围	1–462
权重变异系数	1.66
Kish有效样本量	1,803
Kish设计效应	3.76

过度离散与集中参数 κ。参与者中Y_i/n_i 的观测方差在汇总层面约为二项分布预测的12倍（边际过度离散，即未考虑协变量之前），且所有51个州均各自表现出过度离散。在以模型协变量和层级结构为条件后，估计的集中参数$\hat{\kappa} \approx 7$ 意味着条件过度离散因子约为7。贝塔-二项分布以 $\kappa \to \infty$的形式包含二项分布，从结构上对这种过度离散进行建模，而非事后校正（Prentice，1986）。

复杂抽样设计与伪后验分布。 抽样权重介于 1 至 462 之间，Kish有效样本量为 $\text{ESS}_{\text{Kish}} = 1,803$，设计效应为 $\text{DEFF}_{\text{Kish}} = 3.76$；一项正式检验证实，该设计对集约边际具有信息性（Pfeffermann，1993）。忽略该设计会导致可信

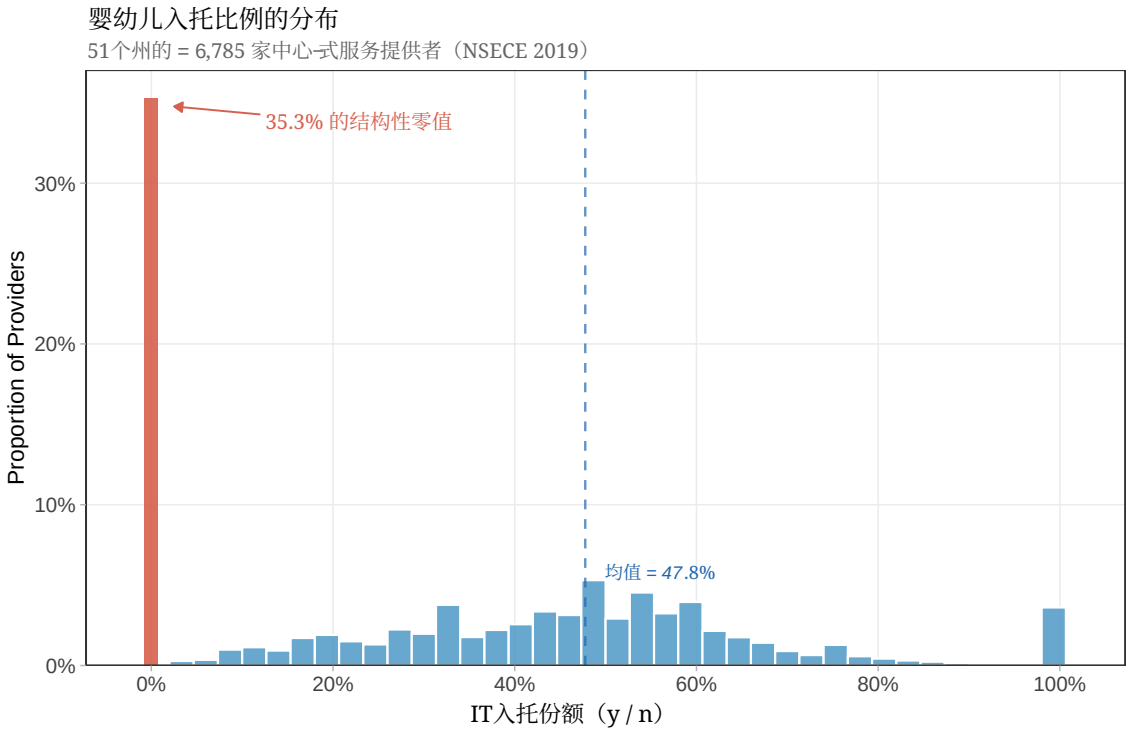

图1. 婴幼儿入托比例 Y_i/n_i在 $N = 6{,}785$ 家中心式服务提供者中的分布（NSECE 2019）。零值处的峰值（占中心的 35.3%）代表不接收任何 3 岁以下儿童的服务提供者。在提供 IT 服务的服务提供者中（$z_i = 1$, $N_{\text{pos}} = 4{,}392$ 家），分布大致呈单峰，均值为 0.478，且离散程度较高（标准差 $= 0.214$）；这与过度离散程度约为二项分布的 12 倍相一致。以上三个特征共同构成采用跨栏贝塔-二项模型的动机。

区间的经验覆盖率远低于名义覆盖率；第 3.5 节介绍了伪后验分布和三明治校正。

每种特征单独都排除了一类候选模型：结构性零值排除了单方程设定，有界支持集排除了泊松分布和负二项分布核，普遍过度离散排除了二项分布，而复杂抽样设计则排除了朴素的贝叶斯后验。四项约束的交集将可行的设定范围缩小到采用贝塔-二项核的障碍模型，并通过带有三明治校正的伪后验进行估计。

2.3. **贫困反转的初步证据**。这四种数据特征决定了模型的分布设定。模型的层级结构——州变化系数及其跨边际协方差——源于第五种经验模式：贫困反转在各州之间呈现系统性差异。

汇总估计。 以标准化社区贫困率为自变量，对婴幼儿参与（z_i）进行合并逻辑回归，得到广延边际系数 $\hat{}\alpha_{\text{pov}} = -0.020$ **(**$p < 10^{-10}$**)**，证实社区贫困率较高地区的服务提供者为婴儿提供服务的可能性显著较低。以贫困率为自变量，对入学份额（Y_i/n_i 服务机构中的份额）进行合并普通最小二乘回归，得到集约边际系数 $\hat{}\beta_{\text{pov}} = +0.003$**(**$p < 10^{-14}$**)**，说明在以提供服务为条件时，较贫困社区的服务提供者将更大比例的入学名额分配给婴儿。系数符号相反，构成了最简单形式的反转：贫困降低参与，但提高强度。

其内涵解释值得关注。贫困构成了进入壁垒：婴儿照护成本高昂，抑制了服务参与，而需求恰恰在这些地区最为旺盛（Dobbins等人，2016年；Malik等人，2018年）。然而，在跨过这一壁垒的中心中，弱势社区的中心会将其入学结构更多地倾向于最低年龄组，这或反映了替代选择较少地区的市场需求，或体现了使命驱动型服务供给。任何将参与和强度合并到单一方程中的模型，都会估计出较弱的净效应，从而掩盖这两种相互制衡的作用。

州际异质性。按州分别估计这些回归时，51个州中有23$\hat{}$个州（45%）呈现经典反转模式（$\hat{}\alpha_{\text{pov},\,s} < 0$ 和$\beta_{\text{pov},\,s} > 0$）。其余各州则呈现不同组合：有些州在两个边际上均表现出负效应，另一些州则在两个边际上均表现出正效应。两个边际特定贫困系数之间的州际相关性为 $r = 0.021$——接近于零——这表明广延机制和密集型机制在很大程度上独立运作。图2展示了原始数据中的总体反转；州级双变量散点图见图5。跨边际相关性接近于零，意味着知道某州存在较强的参与障碍，基本无法提供有关其强度响应的信息。这一发现促使我们在第3节中构建完整协方差结构（Sipple等人，2020年）。

政策相关性。 在州级政策变量中，分级报销与贫困—参与关系的关联最强：

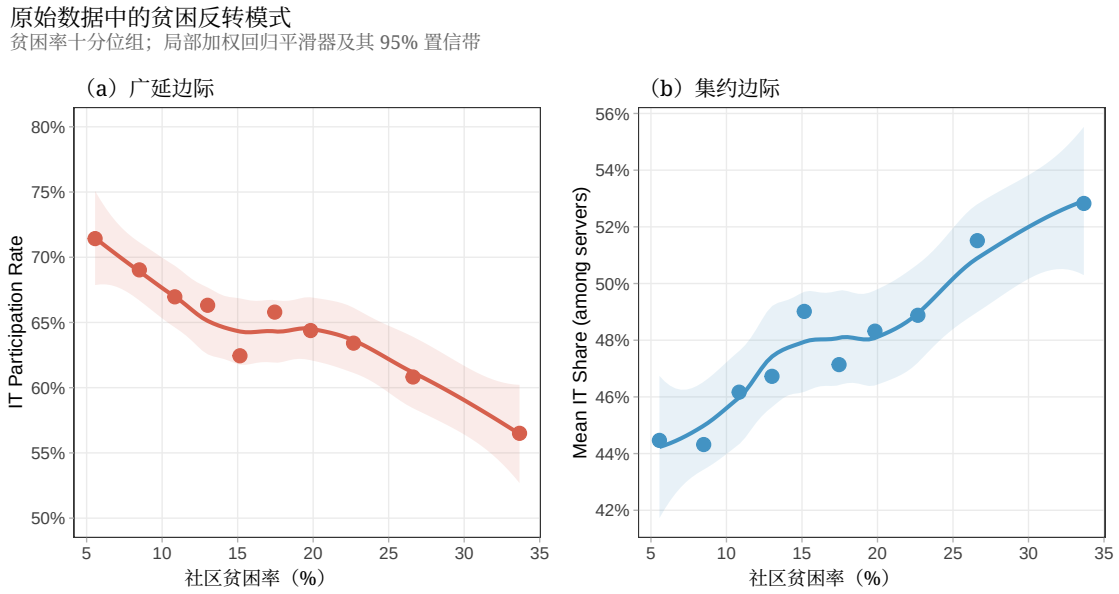

图2. 原始NSECE 2019年数据中贫困反转的初步证据。左图：按社区贫困率十分位划分的IT参与率（为任何婴儿提供服务的中心所占比例）；带95%置信带的局部加权回归平滑器证实了下降趋势。右图：按贫困率十分位划分的服务机构中IT入学份额（Y_i/n_i）的均值，显示出相反的正向趋势。两幅图共同以最简单的形式展示了这一反转：参与随贫困下降，但强度上升。

实行分级报销的州，其广延边际上的贫困系数往往较不负，这与以下假设一致：与质量挂钩的支付激励抵消了在弱势社区服务婴儿所面临的成本劣势。这些初步相关性为第5节中的

总之，这些数据需要一个能够同时做到以下几点的模型：(i) 通过门槛将结构性零值与正计数分离，(ii) 通过贝塔-二项分布核尊重有界支持集，(iii) 通过集中参数容纳过度离散，(iv) 通过带三明治校正的伪后验纳入调查权重，以及 (v) 通过具有完整协方差的层级先验刻画州际异质性与跨边际依赖。第3节建立该模型。

3. 层级门槛贝塔-二项模型

本节构建分层障碍贝塔-二项模型（HBB），其分布设定的依据是

第2.2节中记录的四项数据特征。该模型将三个组成部分——门槛、贝塔-二项核和层级先验——整合为一个统一框架，并为此类模型增添了两项新的方法学贡献：一种跨边际协方差，其识别完全通过层级先验而非条件似然实现；以及一种基于乔列斯基分解的夹心校正，将伪后验推断扩展至两部分层级设定。

通俗地说，该模型的工作方式如下。首先，门槛 将提供婴幼儿托育服务的提供者与不提供此类服务的提供者区分开来，并将参与情况建模为具有州特定截距和斜率的逻辑回归。对于参与者，贝塔-二项分布核函数用于建模入学人数中婴幼儿所占的份额，同时处理有界计数和州内过度离散。两个边际共享一个针对州级偏差的联合层级先验，因此某州的参与障碍与其条件入学强度可以呈正相关或负相关——这种跨边际协方差正是本文核心的结构性创新。抽样权重通过伪似然纳入模型，而三明治校正则恢复总体平均参数的设计一致覆盖率。

3.1. **记号与数据结构**。 我们观察到 $N = 6{,}785$ 家以中心为基础的托儿服务提供者，来自 $S = 51$ 个州（包括哥伦比亚特区）。每家提供者 i 均隶属于一个州 $s[i]$。结果变量对 (Y_i, n_i) 记录婴幼儿入园人数 $Y_i \in \{0, 1, \ldots, n_i\}$ 占总入学人数 n_i的情况；参与指示变量为 $z_i = \mathbf{1}(Y_i > 0)$。 表2 汇总了全文使用的符号。所有连续协变量均在分析前进行标准化。

3.2. **障碍贝塔-二项似然。贝塔-二项分布**。采用均值–精度参数化的贝塔-二项分布（Griffiths, 1973年; Prentice, 1986）$(\mu,\ \kappa)$ 设定 $a = \mu\kappa$ 和 $b = (1-\mu)\kappa$，其中 $\mu \in (0,\ 1)$ 是条件均值参数，而 $\kappa > 0$ 是集中参数。较大的 κ意味着过度离散程度较低，当二项分布$(n,\ \mu)$ 分布在 $\kappa \to \infty$时恢复。其概率质量函数（PMF）为

$$\mathbb{P}(Y = y \mid n, \mu, \kappa) = \binom{n}{y} \frac{B\big(y + \mu\kappa,\ n - y + (1-\mu)\kappa\big)}{B\big(\mu\kappa,\ (1-\mu)\kappa\big)}, \quad y = 0, 1, \ldots, n, \tag{3.1}$$

其中， $B(\cdot, \cdot)$ 是贝塔函数。前两个矩为

$$\mathbb{E}[Y] = n\mu, \qquad \mathrm{Var}(Y) = n\mu(1-\mu)\,\frac{n+\kappa}{1+\kappa}. \tag{3.2}$$

表2。符号摘要。维度以括号表示。

符号	维度	领域	描述
索引和维度			
i	—	$\{1,\dots,N\}$	服务提供商索引 ($N = 6,785$)
s	—	$\{1,\dots,S\}$	州索引（$S = 51$，含哥伦比亚特区）
$s[i]$	—	$\{1,\dots,S\}$	提供者 i 的州归属
N_s	—	$\mathbb{Z}_{>0}$	州 s 内服务提供商数量；$\sum_s N_s = N$
结果变量			
n_i	(1)	$\mathbb{Z}_{>0}$	总入学人数，0–5岁（视为固定）
Y_i	(1)	$\{0,1,\dots,n_i\}$	婴幼儿（IT）入学人数
z_i	(1)	$\{0,1\}$	$z_i = \mathbf{1}(Y_i > 0)$：婴幼儿参与指标
协变量			
$\mathbf{x}_i$	$(P \times 1)$	$\mathbb{R}^P$	服务提供者协变量，$P = 5$：截距、贫困、城市、黑人、西班牙裔
$\mathbf{x}_i^{(r)}$	$(q \times 1)$	$\mathbb{R}^q$	随状态变化的子向量 $\mathbf{x}_i$ ($q = P$ 在 M2/M3 中)
q_i	(1)	$(0,1)$	参与概率（广延边际）；参见 (3.10)
$\mathbf{v}_s$	$(Q \times 1)$	$\mathbb{R}^Q$	状态政策向量，$Q = 4$：截距 $+ Q - 1$ 政策工具
参数			
$\boldsymbol{\alpha}$	$(P \times 1)$	$\mathbb{R}^P$	总体平均广延边际系数
$\boldsymbol{\beta}$	$(P \times 1)$	$\mathbb{R}^P$	总体平均集约边际系数
κ	(1)	$\mathbb{R}_{>0}$	贝塔-二项分布集中度
$\boldsymbol{\delta}_{k,s}$	$(q \times 1)$	$\mathbb{R}^q$	州偏差，边际 $k \in \{1,2\}$，州 s
$\boldsymbol{\Gamma}_k$	$(q \times Q)$	$\mathbb{R}^{q \times Q}$	跨层级政策调节变量矩阵，边际 k
$\boldsymbol{\Sigma}_\delta$	$(2q \times 2q)$	$\mathbb{S}^{2q}_{++}$	交叉边际状态效应协方差
调查设计			
w_i	(1)	$\mathbb{R}_{>0}$	抽样权重
$\tilde{w}_i$	(1)	$\mathbb{R}_{>0}$	标准化权重：$\tilde{w}_i = w_i N / \sum_j w_j$

对于所有 $n \geq 2$，方差膨胀因子（$n + \kappa$）/（$1 + \kappa$）均大于 1，这反映了相较于二项分布的过度离散。在托育数据中，典型入托人数为 $n_i \approx 50$，估计的 $\kappa \approx 7$ 为 $57/8 \approx 7$，该因子约为 $57/8 \approx 7$，远超二项模型所能适应的范围。

计算说明。在实际计算中，当 n_i 超过 50 时，我们使用 lbeta() 函数计算 log f_{BetaBin}，以避免数值溢出。具体而言，对数概率质量函数的计算方式为 $\log\binom{n}{y} +$ lbeta$(y + a,\ n - y + b)$ $-$ lbeta$(a,\ b)$，即使 n_i 达到数百，也能保持稳定。

零概率。当 $y = 0$ 时，概率质量函数可简化为一种既易于解释又数值稳定的乘积形式：

$$p_0(n, \mu, \kappa) = \prod_{j=0}^{n-1} \frac{(1-\mu)\kappa + j}{\kappa + j}. \tag{3.3}$$

乘积中的每一因子，都是在此前所有“试验”均产生非IT结果的条件下，第 j次“试验”产生非IT结果的条件概率。乘积形式表明， p_0对 μ 和 κ 均严格递减（对于 $n \geq 2$）：更高的平均强度和更低的过度离散都会降低观察到零IT入托的概率。

障碍模型构造。 正如第2.2节所记载，托育数据中的零值是结构性的：它们反映的是有意不为婴儿提供服务的决定，而非从计数分布中的随机抽取。这促使采用障碍模型设定（Cragg，1971；Mullahy，1986），将参与决策与入托强度区分开来。

定义1 （跨栏贝塔-二项模型）。 对于招生容量为n_i的提供者 i，HBB 模型设定为

$$f(y_i \mid q_i, \mu_i, \kappa, n_i) = \begin{cases} 1 - q_i, & \text{if } y_i = 0, \\ q_i \cdot f_{\text{ZT-BetaBin}}(y_i \mid n_i, \mu_i, \kappa), & \text{if } y_i \in \{1, \ldots, n_i\}, \end{cases} \tag{3.4}$$

其中 $q_i \in$ （0，1）为参与概率（广延边际）， $\mu_i \in$ （0，1）为条件均值IT占比（集约边际），零截断贝塔-二项概率质量函数为

$$f_{\text{ZT-BetaBin}}(y \mid n, \mu, \kappa) = \frac{f_{\text{BetaBin}}(y \mid n, \mu, \kappa)}{1 - p_0(n, \mu, \kappa)}, \quad y = 1, \ldots, n. \tag{3.5}$$

门槛将托育中心面临的两项不同决策区分开来。广延边际刻画一家中心是否收托任何婴儿——即准入决策——由 q_i 支配。集约边际刻画一家中心相对于其总容量提供多少婴儿名额——即强度决策——由 μ_i支配。这种区分反映了一个真正的两阶段过程：中心必须首先决定承担婴儿照护的成本与监管负担（跨越门槛），然后才确定提供多少婴儿名额。

在条件独立下，对数似然可加地分解为广延成分与集约成分，

$$\ell(\boldsymbol{\theta}) = \underbrace{\sum_{i=1}^{N} \big[z_i \log q_i + (1 - z_i) \log(1 - q_i) \big]}_{\ell_{\mathrm{ext}}(\boldsymbol{\alpha})} + \underbrace{\sum_{i:\, z_i = 1} \big[\log f_{\mathrm{BetaBin}}(y_i \mid n_i, \mu_i, \kappa) - \log\big(1 - p_{0,i}\big) \big]}_{\ell_{\mathrm{int}}(\boldsymbol{\beta}, \kappa)}, \quad (3.6)$$

其中 $p_{0,i} = p_0(n_i, \mu_i, \kappa)$。这种分离是障碍模型的结构性质：二元指示变量 z_i 构成一个充分统计量划分，使广延参数 $\boldsymbol{\alpha}$ 与集约参数（$\boldsymbol{\beta}, \kappa$）在似然中变差独立。相比之下，零膨胀模型引入潜在风险指示变量，将两个分量耦合在一起，从而使估计和解释都更为复杂。

无条件期望与强度函数。无条件入学计数均值为

$$\mathbb{E}[Y_i] = q_i \cdot n_i \cdot h(\mu_i, n_i, \kappa), \qquad h(\mu, n, \kappa) \equiv \frac{\mu}{1 - p_0(n, \mu, \kappa)}, \quad (3.7)$$

其中 h是强度函数，将潜在平均比例 μ 映射为参与者中的期望比例。由于 $h(\mu) \geq \mu$（截断使条件均值上移），预计参与中心的IT占比高于仅由潜在参数 μ 所暗示的水平。函数 h 是反转解释的核心：关键问题在于，广延系数与集约系数的相反符号能否被 h的非线性响应“吸收”，抑或这种反转确实是系数层面的现象。下述定理对此作出解答。

定理 1（h的单调性）。对于所有 $n \geq 2$，$\kappa > 0$和 $\mu \in (0, 1)$，

$$\frac{\partial h}{\partial \mu} > 0. \quad (3.8)$$

对于 $n = 1$, $h(\mu) \equiv 1$ 对所有 $\mu \in (0, 1)$。

证明概要。定义 $\Phi(\mu) = (1 - p_0) - \mu\, p_0\, \Lambda$，其中 $\Lambda = \kappa \sum_{j=0}^{n-1} [(1-\mu)\kappa + j]^{-1} > 0$。由于 $\partial h / \partial \mu = \Phi(\mu)/(1 - p_0)^2$, 导数的符号等于 Φ 的符号。可以验证，$\Phi(0) = 0$ 且 $\Phi'(\mu) = \mu\, p_0(\Lambda^2 - \Lambda_2)$，其中 $\Lambda_2 = \kappa^2 \sum_{j=0}^{n-1} [(1-\mu)\kappa + j]^{-2}$。对于 $n \geq 2$，由柯西–施瓦茨不等式可得 $\Lambda^2 > \Lambda_2$，因此 $\Phi'(\mu) > 0$ 在 $(0, 1)$ 上。由于

Φ 从零开始且严格递增，$\Phi(\mu) > 0$ 对所有 $\mu \in (0, 1)$。完整证明见A 节。□

对贫困反转的解释。定理 1证明，强度函数 h 关于 μ严格递增：潜在 IT 强度 μ 较高的中心，其参与者的预期 IT 份额始终较高，无论过度离散程度如何。因此，贫困反转——即贫困加剧会降低参与率（$\alpha_{\text{pov}} < 0$），却会提高条件强度（$\beta_{\text{pov}} > 0$）——纯粹是系数层面的现象，并非非线性响应函数造成的假象。研究者可以直接考察 α_{pov} 和β_{pov} 的后验分布来评估这种反转，无需考虑非线性链接函数的影响。

Prop命题 1（h的弹性）。定义弹性 $\varepsilon_h = (\partial h/\partial \mu)\ (\mu/h)$。则

$$\varepsilon_h(\mu) = 1 - \omega(\mu), \qquad \omega(\mu) = \frac{\mu\, p_0\, \Lambda}{1 - p_0} \in (0, 1) \quad \textit{for } n \geq 2. \tag{3.9}$$

因此 $\varepsilon_h \in (0,1)$：强度函数对 μ 的变化呈非弹性响应。

这种非弹性响应的原因是，增加 μ 会同时提高这样一种概率：即使没有门槛，某个中心也会获得正值抽样（提高 $1 - p_0$，即 h 的分母），从而部分抵消分子效应。实际而言，中心潜在 IT 倾向 μ 每增加百分之一，其参与者预期 IT 份额的增幅就会小于百分之一。在我们的估计参数下（$\mu \approx 0.30$, $\kappa \approx 7$, $n \approx 50$），$\omega \approx 0.07$，因此弹性约为 0.93——在这一参数范围内，抑制效应相对温和。

3.3. **具有跨边际协方差的州变化系数（SVC）**。反转并非全国统一的模式。初步估计显示，呈现经典反转的州不到一半——即贫困的广延边际系数为负，而集约边际系数为正（第2节.3）。这种地理变异引出了一个根本问题：贫困与入学关系是否在各州之间呈现系统性差异？如果是，广延边际与集约边际是否彼此关联？

线性预测变量。沿用 Gelfand 等人（2003年）的空间变系数框架，广延边际和集约边际与服务提供者特征

通过 logit 线性预测变量关联：

$$\operatorname{logit}(q_i) = \mathbf{x}_i \intercal \boldsymbol{\alpha} + \mathbf{x}_i^{(r)\intercal} \boldsymbol{\delta}_{1,s[i]}, \tag{3.10}$$

$$\operatorname{logit}(\mu_i) = \mathbf{x}_i \intercal \boldsymbol{\beta} + \mathbf{x}_i^{(r)\intercal} \boldsymbol{\delta}_{2,s[i]}, \tag{3.11}$$

其中，$\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^P$ 是总体平均固定效应——即“全国性叙事”；$\boldsymbol{\delta}_{k,s} \in \mathbb{R}^q$（$k = 1$ 用于广延边际，$k = 2$ 用于密集型边际）是均值为零的州特定偏差，用于捕捉各州的贫困与入学关系相对于全国平均值的偏离。子向量 $\mathbf{x}_i^{(r)}$ 由$\mathbf{x}_i$ 的前 q 个分量组成，这些分量对应的系数允许随州而异。协变量 j 在边际 k 上的州特定总系数 为

$$\tilde{\alpha}_{j,s} = \alpha_j + \delta_{1,s,j}, \qquad \tilde{\beta}_{j,s} = \beta_j + \delta_{2,s,j}. \tag{3.12}$$

˜在儿童保育应用中，若某州的 $\tilde{\alpha}_{\text{pov},s} < 0$ 且 $\beta_{\text{pov},s} > 0$ ，则会出现反转：社区贫困程度越高，参与率越低，但条件强度越高。

跨边际协方差：关键创新。 标准障碍模型分别估计两个边际。我们打破这一惯例，将州偏差堆叠成一个 $2q$ 维向量：

$$\boldsymbol{\delta}_s = \begin{pmatrix} \boldsymbol{\delta}_{1,s} \\ \boldsymbol{\delta}_{2,s} \end{pmatrix} \sim \mathsf{N}_{2q}(\mathbf{0},\ \boldsymbol{\Sigma}_\delta)\,, \qquad s = 1, \ldots, S, \tag{3.13}$$

及其 $2q \times 2q$ 正定协方差

$$\boldsymbol{\Sigma}_\delta = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{3.14}$$

其中 $\boldsymbol{\Sigma}_{11}$ 控制广延边际系数的州间变异，$\boldsymbol{\Sigma}_{22}$ 控制集约边际，而非对角块 $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}\intercal$ 刻画跨边际相依性（在政策调节模型 M3b 中，$\boldsymbol{\Sigma}_{12}$ 指残差跨边际协方差，即去除政策效应之后的协方差；参见注释 B.19）。这一跨边际协方差是该模型的核心结构性创新：它允许参与障碍更强的州（更大的$|\delta_{1,\ \text{pov},\ s}|$）同时表现出更强或更弱的强度效应（$\delta_{2,\ \text{pov},\ s}$），其方向与幅度由数据估计。协变量 j 的跨边际相关性为

$$\varrho_j^{\text{cross}} = \frac{[\boldsymbol{\Sigma}_{12}]_{jj}}{\sqrt{[\boldsymbol{\Sigma}_{11}]_{jj}\,[\boldsymbol{\Sigma}_{22}]_{jj}}}. \tag{3.15}$$

命题 2（对 $\boldsymbol{\Sigma}_{12}$ 基于条件似然的不可识别性）。 条件于州级偏差 $\boldsymbol{\delta}_s$，*HBB* 模型的费 *Fisher* 信息矩阵在广延边际参数（$\boldsymbol{\alpha}$，$\boldsymbol{\delta}_{1,\,1}$，$\ldots$，$\boldsymbol{\delta}_{1,\,S}$）与集约边际参数（$\boldsymbol{\beta}$，$\boldsymbol{\delta}_{2,\,1}$，$\ldots$，$\boldsymbol{\delta}_{2,\,S}$，$\kappa$）之间呈分块对角。因此，$\boldsymbol{\Sigma}_{12}$ 从条件似然中获得的信息量为零，并经由作用于 $S = 51$ 个已估计州级偏差之上的关于 $\boldsymbol{\delta}_s$ 的分层先验得以识别。

证明。 对数似然可分解为 $\ell = \ell_{\text{ext}} + \ell_{\text{int}}$ (公式 (3.6))。由于 ℓ_{ext}仅取决于 $(\boldsymbol{\alpha}, \boldsymbol{\delta}_{1,\cdot})$，而 ℓ_{int} 仅取决于 $(\boldsymbol{\beta}, \boldsymbol{\delta}_{2,\cdot}, \kappa)$，因此对于任意广延边际参数 ϕ_1 和集约边际参数 ϕ_2，交叉导数$\partial^2 \ell / \partial \phi_1 \partial \phi_2$⊺ 均为零。因此，观测信息矩阵和期望信息矩阵在不同边际之间呈分块对角结构，非对角块 $\boldsymbol{\Sigma}_{12}$ 不会出现在似然中。□

$\boldsymbol{\Sigma}_{12}$ 作为层级估计目标。 命题 2阐明了 $\boldsymbol{\Sigma}_{12}$的信息结构：其后验分布完全由来自共同 $\mathrm{N}_{2q}(\mathbf{0}, \boldsymbol{\Sigma}_\delta)$ 分布的 $S = 51$ 组 $(\boldsymbol{\delta}_{1,s}, \boldsymbol{\delta}_{2,s})$ 决定，因此其有效样本量为 S，而非 N。原则上，频率学派可以通过对边际似然使用限制性最大似然来估计 $\boldsymbol{\Sigma}_{12}$；然而，相对于组数 $(S = 51)$，高维协方差结构（每个边际有 $2q = 10$ 个自由参数）使不受约束的频率学派估计较为脆弱。LKJ(η) 先验提供有限样本正则化，使后验以 $O(S^{-1/2})$ 的速率集中。相比之下，边际内区块 $\boldsymbol{\Sigma}_{11}$和 $\boldsymbol{\Sigma}_{22}$还通过州特定系数获得直接的似然信息，并以 $O(N^{-1/2})$ 的速率集中。

注释 1（条件似然与边际似然）。命题 2 中的分块对角性是关于条件似然 $p(\mathrm{y} \mid \boldsymbol{\delta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \kappa)$ 的陈述。对随机效应积分后得到的边际似然 $p(\mathrm{y} \mid \boldsymbol{\Sigma}_\delta, \boldsymbol{\alpha}, \boldsymbol{\beta}, \kappa) = \int p(\mathrm{y} \mid \boldsymbol{\delta})\, p(\boldsymbol{\delta} \mid \boldsymbol{\Sigma}_\delta)\, \mathrm{d}\boldsymbol{\delta}$，在 $\boldsymbol{\Sigma}_{12} \neq \mathbf{0}$ 时不会因子分解，因此原则上可以通过边际似然识别 $\boldsymbol{\Sigma}_{12}$。然而，条件形式适用于 MCMC 计算，其中将 $\boldsymbol{\delta}_s$ 作为潜变量进行抽样，并在给定已实现偏差的条件下更新 $\boldsymbol{\Sigma}_\delta$。在频率学派框架下，原则上可以通过边际似然对 $\boldsymbol{\Sigma}_{12}$ 进行限制最大似然估计；但当有 $S = 51$ 个组为一个 $2q = 10$ 维协方差矩阵提供信息时，数值计算会很不稳定：在这种情况下，边界解和奇异海森矩阵很常见。采用 LKJ 正则化的贝叶斯方法能够提供稳定的估计，但代价是集中速率较慢；命题 3 记载了 $O(S^{-1/2})$ 的这一集中速率。

命题 3（通过先验实现 $\boldsymbol{\Sigma}_{12}$ 的可识别性）。 假设 *(i)* $S > 2q$，*(ii)* 每个州 s 至少有 q 个协变量线性独立的服务提供者，且 *(iii)* $\boldsymbol{\Sigma}_\delta$ 的先验在 $\mathbb{S}^{2q}_{++}$ 上具有全支持。那么，$\boldsymbol{\Sigma}_\delta$（包括 $\boldsymbol{\Sigma}_{12}$）的后验分布集中良好，且 $\boldsymbol{\Sigma}_{12}$ 每个元素的后验标准差以 $O(S^{-1/2})$ 的速率递减。对于我们的数据，$S = 51 \gg 2q = 10$。

备注 2 (边界条件**)**。 当 $S < 2q$ 即使使用适当先验，跨边际协方差仍难以识别，因为 S 来自 $2q$维分布的抽样会产生奇异的样本协方差矩阵。在这种情况下，Ghosal 等人（2020年）提出的标量跨边际参数 a，其嵌套形式为 $\boldsymbol{\Sigma}_{12} = a\,\boldsymbol{\Sigma}_{22}$，能够简洁但有效地概括跨边际依赖。

计算说明。 在 Stan（Carpenter 等人，2017年）中，州偏差通过非中心化参数化（NCP）进行抽样：抽取$\mathbf{z}_s \overset{\text{i.i.d.}}{\sim} \mathrm{N}_{2q}(\mathbf{0},\ \mathbf{I})$，并令 $\boldsymbol{\delta}_s = \mathrm{diag}(\boldsymbol{\tau})\,\mathbf{L}_\varepsilon\,\mathbf{z}_s$，其中$\mathbf{L}_\varepsilon$是相关矩阵$\mathbf{R}_\varepsilon$的 Cholesky 因子。这消除了中心参数化下因小 τ_j 而产生的漏斗几何（Betancourt 和 Girolami，2015年）：当$\delta_{s,\,j}$ 接近零时，会导致哈密顿蒙特卡洛出现发散转移。对于 $S = 51$ 个州，且 N_s 的取值范围为 17 至 568，NCP 是恰当的默认选择。

3.4. 跨层级政策调节变量与边际效应分解。 为什么这种反转会出现在某些州，而在其他州却不会？州级补贴政策对此提供了一种自然的解释：它们直接影响中心在决定是否以及接收多少婴儿时所面临的成本结构。我们通过跨层级政策调节变量将这一假设形式化。

调节框架。

定义 2（政策调节变量设定）。 对于每个边际 $k \in \{1, 2\}$，州特定偏差分解为

$$\boldsymbol{\delta}_{k,s} = \boldsymbol{\Gamma}_k\,\mathbf{v}_s + \boldsymbol{\varepsilon}_{k,s}, \tag{3.16}$$

其中 $\boldsymbol{\Gamma}_k \in \mathbb{R}^{q \times Q}$ 是跨层级交互矩阵，$\boldsymbol{\varepsilon}_{k,s}$ 是州残差效应。将两个边际堆叠后，残差服从一个保留跨边际依赖的联合分布：

$$\boldsymbol{\varepsilon}_s \equiv \begin{pmatrix} \boldsymbol{\varepsilon}_{1,s} \\ \boldsymbol{\varepsilon}_{2,s} \end{pmatrix} \sim \mathrm{N}_{2q}\left(\mathbf{0},\ \boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} \boldsymbol{\Sigma}_{\varepsilon,1} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{\varepsilon,2} \end{pmatrix}\right), \tag{3.17}$$

从而使完整模型估计残差跨边际协方差 $\boldsymbol{\Sigma}_{12} = \mathrm{Cov}(\boldsymbol{\varepsilon}_{1,s}, \boldsymbol{\varepsilon}_{2,s})$ 在去除政策解释成分之后（见命题 2和补充材料 B）。

州级政策向量为$\mathbf{v}_s \in \mathbb{R}^Q$ 包含 $Q = 4$，由一个截距和$Q-1 = 3$ 直接影响中心面临的婴儿照护成本结构的儿童保育与发展基金（CCDF）政策工具组成。具体的政策变量及其实际作用详见第5节。

嵌套模型层级。HBB框架将一系列复杂度逐步增加的模型嵌套在一起，汇总见表3。

表3. 嵌套模型层级。每个模型都在前一个模型的基础上增加结构。

模型结构		所探讨的问题	近似参数数目
M0	仅固定效应	国家级贫困逆转?	11
M1	+ 随机截距†	入学情况是否因州而异?	116
M2	+ 状态变化系数	反转是否因州而异?	∼ 576
M3a	+ 跨边际协方差	两个边际是否相关?	∼ 576
M3b	+ 政策调节变量	政策能否解释州之间的差异—性?	∼ 626

†M通过将状态变化系数限制为仅包含截距（$q = 1$），可将 1 视为 M2 的一个特例 .

每次转换都在回答前一模型提出的问题：M0 到 M1，考察基准入学率是否因州而异；M1 到 M2，考察反转模式是否因州而异；M2 到 M3a，考察跨边际协方差能否改善拟合；M3a 到 M3b，则考察可观测政策能否解释州际异质性。

贫困系数展开。

命题4 (贫困系数展开)。 在 *M3b* 下，广延边际上的州特定贫困系数总量为

$$\tilde{\alpha}_{\mathrm{pov},s} = \underbrace{\alpha_{\mathrm{pov}}}_{\substack{\textit{national}\\ \textit{average}}} + \underbrace{\boldsymbol{\gamma}_{1,\mathrm{pov}}^{\intercal}\,\mathbf{v}_s}_{\substack{\textit{policy-explained}\\ \textit{deviation}}} + \underbrace{\varepsilon_{1,\mathrm{pov},s}}_{\substack{\textit{residual}\\ \textit{state effect}}}, \tag{3.18}$$

其中 $\boldsymbol{\gamma}_{1,\ \mathrm{pov}} = (\gamma_{1,\ 2,\ 1}, \ldots, \gamma_{1,\ 2,\ Q})^{\intercal}$ 是 $\boldsymbol{\Gamma}_1$ 的第二行。类似地，密集型-˜边际对应的项为 $\beta_{\mathrm{pov},\ s} = \beta_{\mathrm{pov}} + \boldsymbol{\gamma}_{2,\ \mathrm{pov}}^{\intercal}\mathbf{v}_s + \varepsilon_{2,\ \mathrm{pov},\ s}$。

命题4将州特定贫困效应分解为三个易于解释的组成部分：全国平均值（α_{pov}）、政策解释偏离（$\boldsymbol{\gamma}_{1,\ \mathrm{pov}}^{\intercal}\mathbf{v}_s$）以及残差特异成分（$\varepsilon_{1,\ \mathrm{pov},\ s}$）。州特定贫困系数的方差

相应地分解为 $\mathrm{Var}(\tilde{\alpha}_{\mathrm{pov},\,s}) = \gamma_{1,\,\mathrm{pov}}^{\top}\,\mathrm{Var}(\mathbf{v}_s)\,\gamma_{1,\,\mathrm{pov}} + (\Sigma_{\varepsilon,\,1})_{\mathrm{pov},\,\mathrm{pov}}$，将州间变异划分为政策解释部分和残差组成部分。

边际效应分解。为了将模型参数转化为与政策相关的量，我们将任意协变量 x_k 对预期入学人数的总效应分解为可及性和强度两个作用渠道。

命题 5（对数可及性与对数强度分解）。对于州s中的提供者 i，协变量 x_k 对对数期望入学的边际效应可分解为

$$\frac{\partial \log \mathbb{E}[Y_i]}{\partial x_{i,k}} = \underbrace{(1-q_i)\,\tilde{\alpha}_{k,s}}_{\mathrm{LAE}_k} + \underbrace{(1-\mu_i)\,\varepsilon_{h,i}\,\tilde{\beta}_{k,s}}_{\mathrm{LIE}_k}, \tag{3.19}$$

其中LAE_k 表示对数可及性效应，而LIE_k 表示对数强度效应，其中$\varepsilon_{h,\,i} = 1-\omega_i$ 来自命题 *1*。

证明。由于 $\log \mathbb{E}[Y_i] = \log\, q_i + \log\, n_i + \log\, h_i$ 且 n_i 为固定值，链式法则给出两项。对于可及性项，logit 链接函数得到 $\partial(\log q_i)/\partial x_{i,\,k} = (1-q_i)\,\tilde{\alpha}_{k,\,s}$。对于强度项，$\partial(\log\, h_i)/\partial x_{i,\,k} = \varepsilon_{h,\,i}\,(1-\mu_i)\beta_{k,\,s}$，其中使用了逻辑斯蒂导数$\partial\mu/\partial\eta = \mu(1-\mu)$和弹性定义 $\varepsilon_h = \mu\,\partial(\log\, h)/\partial\mu$。□

具体解释。在样本均值（$q \approx 0.64$, $\mu \approx 0.30$, $\varepsilon_h \approx 0.93$）处，使用总体平均系数 $\alpha_{\mathrm{pov}} = -0.324$ 和 $\beta_{\mathrm{pov}} = +0.090$：$\mathrm{LAE}^{\mathrm{pov}} = (1-0.64)\,(-0.324) = -0.117$，$\mathrm{LIE}^{\mathrm{pov}} = (1-0.30)\,(0.93)\,(0.090) = +0.058$。净效应为 -5.9%：准入障碍占主导，但强度响应抵消了近一半。该分解有三个值得强调的特点：它在对数尺度上严格可加；两个乘数$(1-q_i)$和$(1-\mu_i)\,\varepsilon_{h,\,i}$均为正且有界，因此每个分量的符号完全由系数的符号决定；反转模式也得到确认，属于与定理 1 一致的系数层面现象。

3.5.**调查加权伪后验与三明治校正**。NSECE 2019 采用分层多阶段整群抽样设计，包含 30 个层、415 个初级抽样单位（PSU），抽样权重范围为 1 至 462（Kish 有效样本量 $= 1,\,803$；设计效应 $= 3.76$）。在信息性抽样设计下，忽略这些权重会导致点估计产生偏差，并使可信区间校准不当（Pfeffermann 等，1998年）。遵循 Savitsky 和 Toth（2016年）以及 Williams 和 Savitsky（2021），我们采用伪后验方法，并在事后进行三明治校正（多层模型中加权的一般处理亦可参见 Rabe-Hesketh 和 Skrondal，2006）。

伪对数似然。定义加权对数似然

$$\ell^{(w)}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \tilde{w}_i \log f_{\mathrm{HBB}}(y_i \mid \boldsymbol{\theta}), \tag{3.20}$$

其中 $\tilde{w}_i = w_i N / \sum_j w_j$ 为归一化的调查权重。

定理 2 (伪后验适定性**)**。 在适当先验 $\pi(\boldsymbol{\theta})$下，伪后验分布

$$\pi^{(w)}(\boldsymbol{\theta} \mid \mathbf{y}) \propto \exp\!\big(\ell^{(w)}(\boldsymbol{\theta})\big)\, \pi(\boldsymbol{\theta}) \tag{3.21}$$

对于任意一组正且有限的权重 $\{w_i\}_{i=1}^{N}$均为适当的。

证明概要。 由于 $f_{\mathrm{HBB}}(y_i \mid \boldsymbol{\theta}) \in (0,1]$ 和 $\tilde{w}_i > 0$，每个加权因子均满足$f_i^{\tilde{w}_i} \le 1$。因此，利用$\pi(\boldsymbol{\theta})$的适当性可得 $\int \exp(\ell^{(w)})\ \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} \le \int \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} = 1 < \infty$。验证边际似然严格为正的完整证明见补充材料C。 □

命题 6 (加权可分离性**)**。 加权对数似然仍保留两部分分解：

$$\ell^{(w)}(\boldsymbol{\theta}) = \ell^{(w)}_{\mathrm{ext}}(\boldsymbol{\alpha}) + \ell^{(w)}_{\mathrm{int}}(\boldsymbol{\beta}, \kappa), \tag{3.22}$$

其中 $\ell^{(w)}{}_{\text{扩展}} = \sum_i \tilde{w}_i [z_i \log\ q_i + (1 - z_i) \log(1 - q_i)]$ 和 $\ell^{(w)}{}_{\text{截距}} = \sum_{i:\ z_i = 1} \tilde{w}_i [\log\ f_{\mathrm{BetaBin}}(y_i \mid n_i,\ \mu_i,\ \kappa) - \log(1 - p_{0,\ i})]$。

证明。 按 $\tilde{w}_i > 0$ 加权会将每个观测的对数密度乘以一个正标量，因此不会在原本属于不同加性分量的参数之间引入函数依赖关系。□

伯恩斯坦–冯·米塞斯（BvM）条件。 权重比 $w_{\max}/w_{\min} = 462$超过 $N^{1/2} \approx 82.5$。有限样本诊断 $\delta \approx \log(462)/\log(6{,}785) \approx 0.695 > 0.5$ 表明伪后验方差可能低估真实的抽样变异性，因此有必要采用下文的三明治校正。

三明治方差。 定义观测海森矩阵和聚类稳健外积：

$$\mathbf{H}_{\mathrm{obs}} = -\sum_{i=1}^{N} \tilde{w}_i \left. \frac{\partial^2 \log f_{\mathrm{HBB}}(y_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^{\mathsf{T}}} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}, \tag{3.23}$$

$$\mathbf{J}_{\mathrm{cluster}} = \sum_{c=1}^{C} \mathbf{s}_c\, \mathbf{s}_c^{\mathsf{T}}, \qquad \mathbf{s}_c = \sum_{i \in \mathrm{PSU}_c} \tilde{w}_i \left. \frac{\partial \log f_{\mathrm{HBB}}(y_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}, \tag{3.24}$$

ˆ 其中， $C = 415$ 为初级抽样单位的数量， $\boldsymbol{\theta}$ 为后验均值。障碍模型可分离性意味着$\mathbf{H}_{\mathrm{obs}}$ 在广延边际参数与

集约边际参数之间呈分块对角结构；然而，$\mathbf{J}_{\text{cluster}}$ 并非分块对角结构，因为单个初级抽样单位会同时对两个边际贡献得分。

注释 3（简化说明）。 公式 (3.24)为清晰说明，采用非分层形式。实际使用的估计量是命题 C.29中的分层聚类稳健形式，该形式在 $H = 30$ NSECE 层内对初级抽样单位得分进行中心化，并应用有限总体校正因子 $C_h/(C_h - 1)$。

设计一致的夹心方差为

$$\mathbf{V}_{\text{sand}} = \mathbf{H}_{\text{obs}}^{-1}\,\mathbf{J}_{\text{cluster}}\,\mathbf{H}_{\text{obs}}^{-1}. \tag{3.25}$$

定理 3（Cholesky 仿射变换）。 设 $\boldsymbol{\theta}$ 为后验均值，$\mathbf{L}_{\text{sand}}$ 为$\mathbf{V}_{\text{sand}}$的下三角 *Cholesky* 因子，$\mathbf{L}_{\text{MCMC}}$ 为 *MCMC* 后验协方差 $\boldsymbol{\Sigma}_{\text{MCMC}}$的下三角 *Cholesky* 因子。变换后的抽样值

$$\boldsymbol{\theta}^{*(m)} = \hat{\boldsymbol{\theta}} + \mathbf{L}_{\text{sand}}\,\mathbf{L}_{\text{MCMC}}^{-1}\big(\boldsymbol{\theta}^{(m)} - \hat{\boldsymbol{\theta}}\big), \qquad m = 1, \ldots, M, \tag{3.26}$$

满足 $\mathbb{E}[\boldsymbol{\theta}^*] = \boldsymbol{\theta}$（均值保持）以及$\mathrm{Cov}(\boldsymbol{\theta}^*)$ $=\mathbf{V}_{\text{sand}}$（设计一致的协方差）。该变换与目标协方差结构相匹配；基于边际方差构造的 *Wald* 置信区间不受变换矩阵选择的影响。

计算说明。 观测海森矩阵$\mathbf{H}_{\text{obs}}$ 由加权伪对数似然（3.23）的负二阶导数计算得出，不包括先验曲率，并在后验均值处求值；而非根据 MCMC 样本协方差计算。这一区别在层级模型中至关重要：固定效应的边际后验方差 $\boldsymbol{\Sigma}_{\text{MCMC}}$通过对随机效应分布积分得出，因此其置信区间远宽于条件精度$\mathbf{H}^{-1}{}_{\text{obs}}$。在我们的应用中，大多数固定效应的层级方差比（$\boldsymbol{\Sigma}_{\text{MCMC}}$）$_{pp}$/（$\mathbf{H}^{-1}{}_{\text{obs}}$）$_{pp}$ 超过 600——这是层级结构所致，而非先验主导——这证实标准替代 $\boldsymbol{\Sigma}_{\text{MCMC}} \approx \mathbf{H}^{-1}$ 在层级模型中并不成立。

设计效应比与分块校正策略。作为一种针对单个模型参数的诊断指标，它将经典设计效应（Pfeffermann 等人，1998年；Rao 与 Scott，1981年）扩展至各个模型参数。定义参数 p 的设计效应比为

$$\mathrm{DER}_p = \frac{[\mathbf{V}_{\text{sand}}]_{pp}}{[\mathbf{H}_{\text{obs}}^{-1}]_{pp}}. \tag{3.27}$$

该设计效应比$_p$ 量化了调查设计使 p阶参数的方差相较于仅由数据决定的（简单随机抽样）方差增加了多少，并据此提出以下三层校正策略。Cholesky变换（3.26）按块应用：

- 固定效应 ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, 对数 κ)：完整的夹心校正。在我们的数据中，设计效应比介于 1.14 到 4.18 之间（均值为 2.11，与 Kish设计效应 3.76 一致）。相应的置信区间膨胀因子为 $\sqrt{\text{设计效应比}} \in [1.07, 2.04]$。
- 随机效应 ($\boldsymbol{\delta}_{k,\,s}$)：根据设计效应比进行部分校正。有效样本量较小的州，其设计效应比接近 1（先验分布起主导作用）；而较大的州，其设计效应比最高可达 4。
- 超参数 ($\boldsymbol{\Sigma}_\delta$, $\boldsymbol{\Gamma}_k$)：不进行校正。夹心估计量不适用于根据州际变异估计的方差分量；设计效应比概念不适用于这些参数。因此，超参数的后验区间是基于模型的摘要，其有效性取决于层次分布的正确设定，而非设计一致性。当某个超参数在实质上至关重要时——例如 在本应用中（命题 2 和3）——建议进行先验敏感性分析。

备注4 （分块应用的注意事项）。 分块应用 定理 3 时忽略了固定效应与随机效应之间的跨块协方差；对于最大的三至五个州，这一协方差可能不可忽略，因为州内有效样本量已接近 N_s。补充材料 C 报告了比较分块应用与联合校正的敏感性分析。模拟研究中评估了三明治校正区间的有限样本覆盖率（第4节）。

3.6. **先验设定与计算。先验层级。** 每个先验均选为弱信息先验——提供足够的正则化以确保计算稳定性和正常后验，同时保持足够宽泛，使数据在识别良好的区域占主导地位。 表4 总结了完整的设定，并遵循 Gelman 等人（2008）关于正则化回归以及 Gelman（2006）关于方差分量的建议。

先验$\log \kappa$。我们将对数集中参数先验的中心设为 2（对应$\kappa \approx 7.4$），因为对 IT入学计数的初步分析表明，过度离散程度适中：参与中心之间的方差均值比约为二项基线的 7–15 倍，与 κ处于 5–20 的范围内一致。尺度设为 1.5，取值范围较宽，可涵盖小于 1（接近最大过度离散）到大于 140（接近二项分布）的 κ值。

表 4. 先验设定。NCP = 非中心化参数化。

参数	先验	依据
$\boldsymbol{\alpha}$, $\boldsymbol{\beta}$	$\mathrm{N}(\mathbf{0},\ 2^2\,\mathbf{I}_P)$	在 logit 尺度上为弱信息；± 4logit单位覆盖了大部分概率范围
$\mathrm{vec}(\boldsymbol{\Gamma}_k)$	$\mathrm{N}(0,\ 1^2)$	弥散；允许适度的跨层效应
$\log \kappa$	$\mathrm{N}(2,\ 1.5^2)$	中心设定为中等程度的过度离散（$\kappa \approx 7.4$）；95%先验区间 $\kappa \in [0.4, 140]$
τ（$\boldsymbol{\Sigma}_\delta$的标准差）	半正态分布$(0,\ 1)$	具有正则化作用；有利于收缩
$\mathbf{R}_\varepsilon$（相关系数）	LKJ(2)（Lewandowski 等人，2009）	向单位矩阵的轻度收缩
$\mathbf{z}_\delta$（非中心化参数化辅助变量）	$\mathrm{N}(\mathbf{0},\ \mathbf{I}_{2q})$	标准正态分布（非中心化参数化）

可识别性。

定理 4（可识别性）。假设以下条件成立：

(C1) 设计矩阵满秩：$\mathrm{rank}(\mathbf{X}) = P$ 且$\mathrm{rank}(\mathbf{V}) = Q$.(C2) 每个州 s至少有 q 个服务提供者，其协变量向量$\mathbf{x}_i^{(r)}$ 线性独立，且至少有一个零响应者和至少 q 个正响应者。(C3) 对所有满足 $z_i = 1$的 i， $n_i \geq 3$，并且每个州都存在服务提供者 i、j，使得 $s[i] = s[j]$ 且$n_i \neq n_j$。(C4) 先验 $\pi(\boldsymbol{\theta})$ 在参数空间上具有全支持。

于是，完整参数向量 $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \kappa, \{\boldsymbol{\delta}_s\}, \boldsymbol{\Sigma}_\delta, \{\boldsymbol{\Gamma}_k\})$ 在后验分布下可识别。

解释。门槛似然可识别州特定的总系数$\tilde{\alpha}_{j,s}$以及$\beta_{j,s}$ (方程(3.12))，前提是满足标准的秩和重复条件；随后，层级结构将它们分解为总体平均效应 $\boldsymbol{\alpha}^{(r)}$,$\boldsymbol{\beta}^{(r)}$、政策调节 $\boldsymbol{\Gamma}_k$ 和州级残差偏差 $\varepsilon_{k,s}$。完整证明见补充材料 B。

条件（C3）对于集约边际至关重要：当 $n_i = 1$ 时，强度函数 $h \equiv 1$ 和参数 μ_i、 κ 仅通过 $1-p_0 = \mu_i$ 进入似然，因此 κ 无法识别。当 $n_i = 2$ 时， p_0 同时取决于 μ 和 κ，但单一约束 $p_0(2, \mu, \kappa) = (1-\mu)[(1-\mu)\kappa + 1]/(\kappa + 1)$ 只能确定一个自由度。对于 $n_i \geq 3$，乘积公式（3.3）提供了足够的结构，将 μ 与 κ 区分开来。完整证明见补充材料 B。

估计流程。算法 1概述完整的估计和推断程序。

算法 1 HBB 模型的估计流程

1：数据准备：对权重 $\tilde{w}_i = w_i N / \sum_j w_j$ 进行归一化；对协变量进行标准化。2：编译：使用非中心化参数化和 lbeta() 编译 Stan 模型，以确保数值稳定性。3：MCMC 抽样：运行 4 条链（NUTS；Hoffman 与 Gelman，2014），并设置 adapt_delta $= 0.95$；迭代次数和树深度设置见补充材料E。ˆ 4：收敛诊断：确认所有参数均满足 $R < 1.01$、有效样本量$_{\text{bulk}} > 400$ （Vehtari 等人，2021），且不存在发散转移。5：得分提取：提取观测层级得分 $\partial \log f_{\text{HBB}} / \partial \boldsymbol{\theta}$，这些得分在 Stan 的生成量代码块中计算，并在每次 MCMC 抽样时进行评估。6：三明治计算：根据公式 (3.23) 和 (3.24) 求得 H_{obs} 和 $\mathrm{J}_{\text{cluster}}$；根据公式 (3.25) 计算 V_{sand}。7：Cholesky变换：按块应用公式 (3.26)，以获得经设计校正的后验抽样 $\{\boldsymbol{\theta}^{*(m)}\}$。8：模型比较：对 M0–M3b 计算留一交叉验证（Vehtari 等，2017年）；$\widehat{\ }$报告 elpd$_{\text{loo}}$ 和 Pareto-k 诊断。

模型从 M0 到 M3b 依次拟合，并将每个较简单模型的后验均值作为后续模型的起始点。与冷启动相比，这种热启动策略可将总计算时间缩短约 40%。完整的 MCMC 设置和收敛诊断见补充材料 E；得分计算细节以及夹心估计量各组成部分的闭式表达式见hurdlebb 软件包手册（Lee，2026）。

本估计流程的有限样本性质——尤其是三明治校正在现实调查设计下的表现——将在下文的模拟研究中评估。

4. 模拟研究

第 3.5 节提出的三明治校正伪后验 基于渐近论证，其有限样本表现必须先得到刻画，才能建议在实践中采用该方法。本节通过一项受控模拟，考察三种推断策略——未加权、朴素加权和三明治校正——在不同设计信息性条件下的表现范围；这些条件依据现实的调查参数进行校准。我们将此项研究视为诚实的诊断，而非单纯的验证：结果揭示了各策略在哪些情况下有效、在哪些情况下失效，以及原因何在，并暴露出三明治校正在层级方差分量方面的根本结构性局限。通过确立这些

表现特性，我们为第5节中采用的报告策略提供了证据依据。

4.1. 设计。数据生成过程。我们构建了一个包含 $M = 50{,}000$ 个服务提供者的有限超总体，分布于 $S = 51$ 个州，各州规模与观测到的调查计数成比例。对于州 s 中的每个服务提供者 i，从经验边际分布中抽取一个协变量向量 $\mathbf{x}_i = (1, x_{\text{pov}}, x_{\text{urb}}, x_{\text{blk}}, x_{\text{his}})^\intercal$。真实参数根据 M1（随机截距）未加权拟合进行校准：

$$\begin{aligned} \boldsymbol{\alpha}_0 &= (0.696,\ -0.119,\ 0.253,\ -0.070,\ -0.139)^\intercal, \\ \boldsymbol{\beta}_0 &= (-0.032,\ 0.057,\ -0.018,\ 0.080,\ 0.040)^\intercal, \quad \log \kappa_0 = 1.655, \end{aligned} \tag{4.1}$$

州随机效应标准差为 $\tau_{\text{ext}} = 0.577$，$\tau_{\text{int}} = 0.208$，跨边际相关性为 $\rho_{\text{cross}} = 0.285$。每次重复模拟中，州随机效应 $\boldsymbol{\delta}_s = (\delta_{1s}, \delta_{2s})^\intercal \sim \text{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_\delta)$ 抽取一次。参与指标生成为 $z_i \sim$ 伯努利分布(q_i)，其中 $\text{logit}(q_i) = \mathbf{x}_i^\intercal \boldsymbol{\alpha}_0 + \delta_{1,s[i]}$；正计数生成为$Y_i \mid z_i = 1 \sim$零截断 Beta-二项分布$(n_i, \mu_i, \kappa_0)$，其中 $\text{logit}(\mu_i) = \mathbf{x}_i^\intercal \boldsymbol{\beta}_0 + \delta_{2,s[i]}$，且 n_i 从总入学人数的经验分布中抽取。零截断贝塔-二项分布通过从无条件贝塔-二项分布(n_i, μ_i, κ_0)中拒绝抽样获得。

抽样设计与情景。 从每个有限总体中抽取一个分层整群样本，约包含 $N \approx 7{,}000$ 名服务提供者。通过规模偏倚纳入概率控制调查信息性：

$$\log \pi_i = c_0 + \rho_{\text{inc}} \cdot y_i^*, \tag{4.2}$$

其中，y_i^*表示标准化潜在结果，c_0 的取值用于达到目标样本量。信息性参数 ρ_{inc} 对应三种情景（表5）。情景 S3 的校准结果与 NSECE 的设计相匹配（设计效应 $_{\text{Kish}} = 3.76$；表1）。情景 S4 将信息性加倍，以在比实际调查更极端的条件下对三明治校正进行压力测试。

估计量。 每次重复模拟均采用三种估计量（表 6），分别对应 第 3.5 节中的三种推断策略。E-UW 使用未加权似然拟合模型，并使用朴素 MCMC 后验构造区间。E-WT 将似然替换为伪对数似然（公式 (3.20)；Savitsky 和 Toth，2016年），针对信息性校正点估计，但保留朴素后验方差。E-WS 则在 E-WT 的基础上，通过 Cholesky 仿射变换使用三明治校正方差（定理 3；Williams 和 Savitsky,

表 5. 模拟情景。信息性参数 ρ_{inc}控制纳入概率与结果之间的依赖关系（公式 (4.2)）； $\mathrm{CV}(w)$ 是所得抽样权重的变异系数；Kish 设计效应 $= 1 + \mathrm{CV}^2(w)$。

情景	ρ_{inc}	CV(w)	Kish设计效应	描述
S0	0.00	≈ 1.0	≈ 2.0	非信息性基线
S3	0.15	≈ 1.67	≈ 3.8	经NSECE校准
S4	0.50	≈ 2.0	≈ 5.0	压力测试

表6. 估计量定义。

估计量	似然	标准误	目的
E-UW	未加权	基于模型（朴素）	忽略设计
E-WT	伪（加权）	基于模型（朴素）	校正点估计
E-WS	伪（加权）	夹心估计量（3.25）	完全校正

2021），从而得到 Wald型置信区间。根据 设计效应比 在 第 3.5 节中的分类，按区块应用 Cholesky变换：对固定效应进行完全校正，不对超参数进行校正。

目标参数与计算。 我们跟踪五个涵盖固定效应和方差分量的参数：贫困斜率 α_{pov} 和 β_{pov}、过度离散对数 κ以及州效应标准差 τ_{ext} 和 τ_{int}。性能通过¯ˆ 90% 区间覆盖率、相对偏差 RB $= 100 \times(\theta - \theta_0)/|\theta_0|$ (%)、均方根误差（RMSE）和宽度比来衡量

$$\mathrm{WR} = \frac{\text{median CI width (E-WS)}}{\text{median CI width (E-WT)}},$$

这隔离了可归因于三明治校正的膨胀。每种情景—估计量组合使用 $R = 200$ 次重复模拟。每个覆盖率估计值的蒙特卡洛标准误（MCSE）为 $\mathrm{MCSE} = \sqrt{\hat{p}(1-\hat{p})/R}$，范围为 0.0 个百分点（边界覆盖率处）至最大 3.5 个百分点（覆盖率接近 50 % 时）；在 90% 名义水平下，$\mathrm{MCSE} \approx 2.1$ 个百分点。各单元格的 MCSE 报告在 表 7的括号中；偏离 90% 超过 $2 \times \mathrm{MCSE}$ 的覆盖率以 †标记。每次重复模拟均采用 M1设定（表3），运行 4 条链，每条链迭代 3,500 次（1,500 次预热，2,000 次抽样），总计获得 8,000 个预热后抽样。所有重复模拟

ˆ均达到 $R < 1.01$ (Vehtari 等人，2021)，且主体有效样本量最低超过 1,600。

4.2. **结果**。 表7 报告了三种情景和三种估计量下全部五个参数的覆盖率、相对偏差、均方根误差和宽度比。 图3以图形方式展示了带有蒙特卡洛标准误带的覆盖率。我们围绕五项发现组织结果，这些发现共同刻画了该估计框架的运行特性。

发现1：在正确设定下，未加权估计量对固定效应实现了接近名义水平的覆盖率。在 E-UW 下， α_{pov} 和β_{pov} 的覆盖率在所有三种情景中均处于 89–93% 的范围内，与 90% 的名义水平相差不到两个蒙特卡洛标准误。这一结论在 S4 下仍然成立，此时抽样设计具有很强的信息性。结果证实，当生成模型正确设定 且推断仅基于模型似然时，忽略调查权重并不会损害区间覆盖率。这一发现不应 被解读为反对调查加权：在实践中，生成模型绝不会完全正确，而发现2–3表明，结合三明治校正的加权方法能够有效抵御模型错设。专门的错设实验（E.5节）证实了这一点：在遗漏了随州变化的贫困斜率的 M2 数据生成过程中，E-UW 对贫困系数的覆盖率降至 0.5–30%，而 E-WS 仍保持 67.5–79.0% 的覆盖率。未加权估计量对 $\log \kappa$（90.5–92.0%）以及超参数 τ_{ext}和 τ_{int} （91.0–98.0%）也维持了名义覆盖率，唯一例外是 τ_{ext} 在 S4 下的覆盖率（77.0%）；此时极端的权重结构会在抽样随机效应中引发构成效应。**发现2：朴素加权推断会导致严重的覆盖不足**。 随着设计复杂度增加，加权估计量 E-WT 的覆盖不足会变得严重且持续加剧。对于 α_{pov}，覆盖率从 S0 下的 81.0% 降至 S3 下的 58.5%，再降至 S4 下的 53.0%——从轻度覆盖不足骤降至仅略多于一半的模拟抽样包含真实值。对于 β_{pov} ，情况类似（76.5%、62.5%、56.5%）。其机制已得到充分理解（Savitsky & Toth，2016年）：伪似然会夸大有效样本量，使后验分布集中于带有权重诱导偏差的点估计周围，从而产生同时过窄且中心偏移的区间。

发现3：三明治校正提供了有意义但不完美的恢复。经三明治校正的E-WS从E-WT的崩溃中大幅恢复，

两个贫困系数的覆盖率达到82.0–88.5%，在S3和S4中较E-WT提高约24–33个百分点。宽度比随设计效应单调变化：WR从S0下的1.22增至S4下的1.64–2.06（表7），与$\mathrm{WR} \approx \sqrt{\mathrm{DER}}$ 以及实证DER 范围1.14–4.18（见表10）相一致。剩余覆盖率差距反映了三明治方差估计量的有限样本偏差——这是簇数量适中时聚类稳健估计量的已知性质（Binder，1983年）。

为更精确地刻画该差距，我们将其分解为宽度与偏差成分，通过标准误校准比标准差/标准误，其中标准差为各次重复模拟中点估计的实际标准差，标准误为所报告夹心标准误的均值（表 E.8）。对于S3下的贫困系数，标准误比为1.13–1.14，证实夹心标准误将真实抽样变异性低估了约13%，而标准化平均偏差 $|b|$/标准差可忽略不计（0.06–0.22）。因此，该差距是由宽度驱动的：区间中心正确，但窄了13%。这一13%的低估与 $O(S^{-1})$ 有限样本偏差相一致，该偏差来自具有 $S = 51$ 个簇的聚类稳健方差估计量（Binder，1983年），并表明小样本校正（例如CR2；Pustejovsky与Tipton，2018年）或可进一步缩小该差距——此为未来工作的一个方向。将区间放宽至95%名义区间证实了这一解释：在S3下，E-WS对贫困系数的覆盖率升至87.5–88.5%——这一增幅与预期的 ≈ 5 个百分点（来自正态分布尾概率）相一致，但仍留下成比例的剩余差距（表 E.9）。

发现4：三明治校正对过度离散参数的效果较弱。对于 $\log \kappa$，E-WS 的覆盖率为57.0–76.5%，显著低于贫困系数的 82–88.5%。这一差异源于 κ以非线性方式进入贝塔-二项分布的方差函数：加权得分方差与经三明治校正的参数方差之间的关系，无法像三明治公式所依据的线性泰勒展开那样得到充分近似。$\log \kappa$ 的宽度比与线性系数的相当（1.31–2.04），这证实三明治校正确实适当地增大了区间宽度；覆盖率不足主要是由点估计偏差而非方差低估所致。

发现5：超参数无法进行三明治校正。一个值得注意的发现涉及 τ_{ext} 和 τ_{int}。在E-WT和E-WS下，所有情景的覆盖率均降至0.0–8.0%，相对偏差为 $+68$至 $+110$%。宽度比恰好为宽度比 $= 1.00$，从而证实就这些参数而言，E-WS $\equiv$ E-WT在构造上成立：

三明治校正仅应用于通过得分进入加权对数似然的参数，而层级方差分量不会产生个体层级得分贡献。该设计效应比无定义，其所针对的是 τ（公式（3.27））， 因此无法计算方差膨胀。相当大的正偏差之所以出现，是因为伪似然将每个观测视为携带 $\tilde{w}_i$ 个信息量单位，从而使表观州际异质性以与 $\mathbb{E}[\tilde{w}_i^2]$成比例的因子膨胀。未加权估计量E-UW保持91–98%的覆盖率，针对两个 τ 参数（S0和S3）；在S4中，τ_{ext} 降至77.0%——这是整张表中唯一的E-UW失效——反映了信息性抽样权重（被忽略）与广延边际随机效应方差之间日益加剧的张力。

跨边际相关性。我们还追踪跨边际相关性 $\rho_{\text{cross}} = \text{cor}(\delta_j^{\text{ext}}, \delta_j^{\text{int}})$的恢复情况，该量由来自$S = 51$ 后验均值随机效应的插入估计$\hat{\rho} = \text{cor}(\bar{\delta}_j^{\text{ext}}, \bar{\delta}_j^{\text{int}})$估计（表 E.5）。在经NSECE校准的设计（S3）下，未加权估计量E-UW高估了真实 $\rho_{\text{cross}} = 0.285$, 相对偏差为 $+104\%$（均方根误差 $= 0.364$），而加权估计量E- WT表现出中等程度的向下偏差 -42%（均方根误差 $= 0.226$）。与 τ_{ext} 和τ_{int}一样，对于该超参数，E-WS $\equiv$ E-WT：三明治校正仅调整固定效应协方差，无法修改随机效应相关结构。由于模拟存档中未保留Ω的完整后验抽样，无法获得逐次重复模拟的可信区间；表E.5改为报告各次重复模拟的频率学派汇总。

实践建议。模拟结果支持对调查加权贝叶斯推断采用由三部分组成的报告策略，并使用经三明治校正的伪后验。第一，对于总体平均固定效应，将经三明治校正的Wald 置信区间作为主要推断工具进行报告，同时指出，当 Kish设计效应超过 3 时，有限样本覆盖率可能比名义值低 2–8 个百分点。第二，对于层级方差分量，报告非加权后验区间，因为三明治校正在结构上不适用，而且加权伪后验会产生系统性偏差的点估计。第三，务必为每个关注参数计算并报告设计效应比(公式（3.27）)：这一标量汇总以可与经典设计效应直接比较的单位传达调查设计效应的大小，并使读者能够评估三明治校正将各区间扩大了多少。实证设计效应比数值见表10(范围为 1.14–4.18)，而表7中的模拟宽度比（范围为 1.22–2.06）与之相互吻合，并为设计效应大小提供了互补证据。

表7。模拟结果：90% 区间的经验覆盖率（%，括号内为蒙特卡洛标准误）、相对偏差（RB，%）、均方根误差（RMSE）以及宽度比（WR），按每种情景的 $R = 200$ 重复模拟次数。E-UW=未加权；E-WT =加权，采用朴素后验方差；E-WS = 加权，采用三明治校正（公式（3.25））。WR =置信区间宽度中位数（E-WS）/置信区间宽度中位数（E-WT）；WR $= 1.00$ 对于 τ参数表示，按构造，E-WS $\equiv$ E-WT。 † = 覆盖率与90%的偏差超过2个MCSE。MCSE $= \sqrt{\hat{p}(1-\hat{p})/R}$；范围为0.0–3.5个百分点。

	覆盖率（%，MCSE）			相对偏差（%）			RMSE			
参数	E-UW	E-WT	E-WS	E-UW	E-WT	E-WS	E-UW	E-WT	E-WS	宽度比
面板 *A*：*S0* — 无信息（设计效应 ≈ 2.0）										
α_{pov}	93.0 (1.8)	81.0 (2.8)†	88.5 (2.3)	+13.2	+4.9	+4.9	0.033	0.044	0.044	1.22
β_{pov}	93.0 (1.8)	76.5 (3.0) †	86.5 (2.4)	−9.0	−4.8	−4.8	0.016	0.022	0.022	1.22
log κ	91.5 (2.0)	64.0 (3.4)†	76.5 (3.0)†	−0.8	+1.6	+1.6	0.023	0.040	0.040	1.31
τ_{ext}	97.5 (1.1)†	3.0 (1.2)†	3.0 (1.2)†	+9.6	+68.8	+68.8	0.082	0.423	0.423	1.00
τ_{int}	91.0 (2.0)	8.0 (1.9)†	8.0 (1.9)†	−6.6	+70.6	+70.6	0.036	0.158	0.158	1.00
*B*面板：*S3* — 经*NSECE*校准（设计效应 ≈ 3.8）										
α_{pov}	89.5 (2.2)	58.5 (3.5)†	82.0 (2.7)†	+16.5	+12.9	+12.9	0.037	0.072	0.072	1.71
β_{pov}	89.0 (2.2)	62.5 (3.4) †	82.5 (2.7) †	−16.7	−3.2	−3.2	0.017	0.032	0.032	1.64
log κ	92.0 (1.9)	34.0 (3.3) †	60.5 (3.5)†	−0.5	+3.3	+3.3	0.020	0.068	0.068	1.75
τ_{ext}	98.0 (1.0)†	0.5 (0.5)†	0.5 (0.5) †	+8.0	+105.6	+105.6	0.091	0.639	0.639	1.00
τ_{int}	97.5 (1.1)†	0.0 (0.0)†	0.0 (0.0)†	−8.6	+97.6	+97.6	0.040	0.211	0.211	1.00
面板 *C*：*S4* — 压力测试（设计效应 ≈ 5.0）										
α_{pov}	92.5 (1.9)	53.0 (3.5)†	86.0 (2.5)	+20.5	+11.9	+11.9	0.036	0.083	0.083	2.06
β_{pov}	91.0 (2.0)	56.5 (3.5)†	84.0 (2.6)†	−21.7	−2.2	−2.2	0.017	0.035	0.035	1.91
log κ	90.5 (2.1)	23.0 (3.0)†	57.0 (3.5)†	+0.8	+4.1	+4.1	0.020	0.083	0.083	2.04
τ_{ext}	77.0 (3.0)†	0.0 (0.0)†	0.0 (0.0) †	+26.5	+110.2	+110.2	0.175	0.671	0.671	1.00
τ_{int}	97.0 (1.2)†	0.0 (0.0) †	0.0 (0.0) †	−0.7	+103.9	+103.9	0.043	0.225	0.225	1.00

这些建议广泛适用于任何使用调查加权伪似然估计的层级贝叶斯模型，而不仅限于障碍贝塔-二项模型设定：三明治校正对方差分量在结构上不适用，这是由$\mathbf{V}_{\text{sand}}$（公式（3.25））的基于得分的构造所决定的；对于任何超参数缺少个体层级得分贡献的层级模型，这一结论都成立。第5节采用的双轨报告策略——固定效应使用Wald（三明治）区间，超参数使用基于模型的后验区间——直接落实了这些建议；后一轨道只有在层级模型设定正确的情况下，才能提供有效推断。补充图展示了相对偏差分布、宽度比

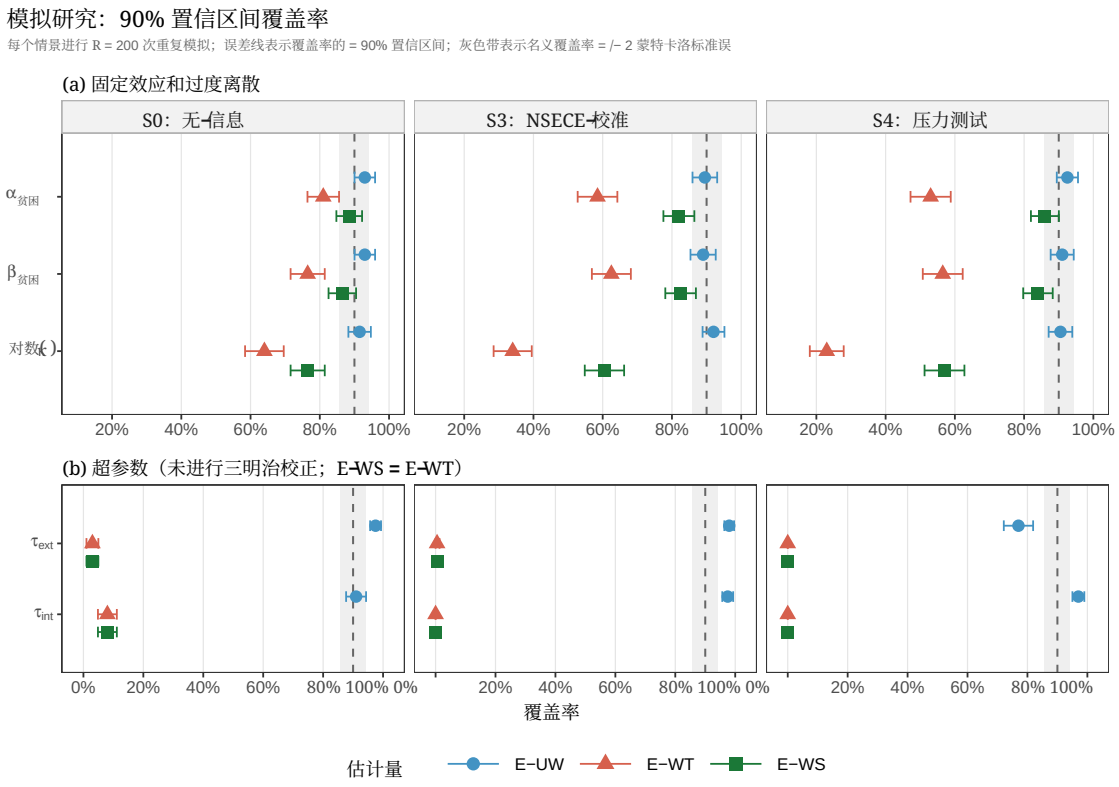

图 3. 在 $R = 200$ 次重复模拟中的 90% 区间经验覆盖率。每个面板对应一个模拟情景（S0、S3、S4，Kish设计效应依次增加）。点表示每个参数—估计量组合的覆盖率；水平虚线表示 90% 名义水平；阴影带表示 ± 2 蒙特卡洛标准误（≈ 4.2 个百分点）。E-UW（未加权）始终保持接近名义覆盖率。随着设计效应增加，E-WT（加权、朴素）估计量的覆盖率急剧下降。E-WS（三明治校正）可显著改善固定效应的覆盖率，但无法校正超参数 τ_{ext} 和 τ_{int}。

箱线图和 设计效应比所有情景和参数的汇总结果见补充材料 E，其中还提供了频率学派情境化（E.6 节），表明经三明治校正的伪后验固定效应具有设计一致性，与传统的 `svyglm` 估计值一致，同时还能涵盖频率学派方法无法处理的层级和分布特征。

在确定了适用范围后，我们现在将双轨报告策略应用于NSECE案例分析：对总体平均固定效应采用三明治校正Wald区间，对州级随机效应和方差分量采用非加权后验区间。

5. 在 NSECE 托育数据中的应用

现在，我们将第3节中建立的完整分析流程应用于第2节所述的2019年NSECE数据，在真实的调查情境中逐一展示各个方法组成部分。第4节中的模拟证据表明，三明治校正显著提高了频率学派覆盖率，DER诊断能够可靠地识别受信息性抽样影响的参数；本节案例分析将展示这些工具的实际应用。分析分为五个部分：通过LOO-CV进行模型比较（第5.1节）、采用双轨报告策略和DER诊断进行总体平均推断（第5.2节）、州特定异质性与跨边际独立性（第5.3节）、边际效应分解（第5.4节），以及简要考察政策调节（5.5节）。这些小节共同展示了三项方法学贡献：解决先验主导问题的双轨报告策略、量化参数特定设计效应的DER诊断，以及将障碍模型系数转换为政策相关量的边际效应分解。

5.1. **模型比较与选择**。 我们拟合了汇总于表3的五个嵌套模型——从汇总的 M0 到含政策调节变量的完整 M3b——采用算法 1所述的估计流程，并在 Stan 中实现（Carpenter 等人，2017年）。每个模型使用 4 条链、每条 3,500 次迭代进行估计（1,500 次预热 +2,000 次抽样），每个参数得到 8,000 次预热后抽样。

收敛。 五个模型均满足标准诊断（Vehtari 等人，2021）：$\hat{R} < 1.01$ 对每个监测参数均成立，最小主体有效样本量超过 400，且未记录到发散转移。 表 8 汇总了这些统计量；非中心化参数化（Betancourt 和 Girolami，2015年）成功消除了漏斗几何。逐参数的完整诊断见补充材料E。

留一交叉验证比较。我们使用帕累托平滑重要性抽样留一交叉验证（LOO-CV；Vehtari 等人，2017年）比较模型，其中“留一”指的是单个服务提供者，而非整个州。每个服务提供者的逐点对数似然 i 包括两个障碍分量：广延边际（伯努利分布）对数似然始终有贡献，而集约边际（零截断贝塔-二项分布）对数似然仅对参与者有贡献（$z_i = 1$）。 表9列出了各模型的期望逐点对数预测密度（$\mathrm{elpd}_{\mathrm{LOO}}$）。对于调查加权设定 M3b-W，LOO 计算采用

ˆ 表 8. 所有模型的收敛诊断。 $R_{\max}$ 是所有监测参数中的最大潜在尺度缩减因子。$\text{ESS}_{\min}$ 报告主体和尾部有效样本量的最小值。所有 ˆ 模型均达到建议的阈值（$R < 1.01$、ESS > 400、发散数为零）。

模型	$\hat{R}_{\max}$	$\text{ESS}_{\min}^{\text{bulk}}$	$\text{ESS}_{\min}^{\text{tail}}$	发散
M0：汇总	1.001	3,390	2,679	0
M1：随机截距	1.004	815	1,415	0
M2：块对角 SVC	1.004	1,078	1,244	0
M3a：跨边际	1.004	1,385	1,257	0
M3b：政策调节	1.004	1,245	1,463	0

未加权的名义对数似然，因为伪对数似然并非适当的预测密度；因此，比较依据模型的名义预测能力对其进行排序。由于 LOO-CV 在服务提供者层级进行，而许多推断目标是州级参数，因此 LOO 排名应被视为观测层级拟合优度的参考，而非对州级推断质量的直接评估。模型逐步扩展的效果呈现边际递减：加入州随机截距 $\widehat{\ }$（M0→M1）带来了显著改进（$\Delta\text{elpd} = +317.8$，标准误 $= 25.9$，超过 12 个标准误），允许州变化系数（M1→M2）则进一步带来了显著提升（$+23.6$ 增量上）。引入跨边际协方差 $\widehat{\ }$（M2→M3a）并未改善预测表现（$\Delta\text{elpd} < 2$），这与接近零的实证跨边际相关性以及命题 2一致：由于 $\boldsymbol{\Sigma}_{12}$未获得似然信息，其价值在于推断而非预测。加入政策调节变量（M3a→M3b）同样没有带来预测增益；这些$\boldsymbol{\Gamma}$ 矩阵重新分配了州际异质性的来源，却没有改变服务提供者层级的预测分布。

模型选择依据。尽管 M3b、M2 和 M3a 的预测性能相当，我们仍采用 M3b 作为报告模型。跨边际协方差$\boldsymbol{\Sigma}_{12}$ 使我们能够对反转作出联合后验陈述（例如，$\Pr(\tilde{\alpha}_{\text{pov},\,s} < 0$ 和$\beta_{\text{pov},\,s} > 0 \mid$ 数据)），而政策调节变量则将州际异质性分解为政策解释的部分和残差部分（命题4）。留一交叉验证的差异不显著，证实增加这一结构不会造成预测性能损失。后验预测检验证实，M3b 能够重现观测到的零值比例（35.3%）以及服务机构 IT 份额的分布；详情见补充材料 D。

表9。模型 M0–M3b 的留一交叉验证比较。$\widehat{\text{elpd}}_{\text{LOO}}$ 是通过帕累托平滑重要性抽样估计的期望逐点对数预测密度（Vehtari 等人，2017年）；$\widehat{\Delta\text{elpd}} = \widehat{\text{elpd}}_{\text{M0}} - \widehat{\text{elpd}}_{\text{Mj}}$ 衡量相对于 M0 的预测增益（数值越负表示预测效果越好）；SE(Δ) 是差异的标准误。当 $|\Delta\text{elpd}| > 2 \times \text{SE}(\Delta)$ 时，则将模型改进标记为显著。所有帕累托 k 诊断值均低于 0.7。

模型	参数	$\widehat{\text{elpd}}_{\text{LOO}}$	SE	$\widehat{\Delta\text{elpd}}$	SE(Δ)	显著性
M0：汇总	11	−20,945。0 140.3		—	—	—
M1：随机截距	116	−20,627。2 144.0		−317.8	25.9	Yes
M2：块对角 SVC	551	−20,603。5 144.4		−341.4	27.6	Yes
M3a：跨边际	576	−20, 601。6 144.4		−343.4	27.6	No
M3b：政策调节	616	−20,607.6 144.6		−337.3	28.5	No

5.2. **总体平均效应与夹心估计量校正。** 表10报告了全部11个固定效应参数的后验均值、朴素的95%可信区间（来自原始MCMC输出）、经夹心估计量校正的95% Wald置信区间，以及设计效应比（DER）。11个参数中，有9个在夹心估计量校正下达到95%水平的统计显著性；两个例外是 α_{Hisp}（广延边际的西班牙裔系数，Wald 置信区间 $[-0.095, 0.106]$）和 β_{Black}（集约边际的黑人系数，Wald 置信区间 $[-0.052, 0.019]$）。

调查加权的影响。 表11 将加权（M3b-W）与未加权（M3b）的后验均值并列呈现。加权会改变每一个固定效应估计值，中位偏移量为5.2个夹心标准误，但仍保留了所有具有实质意义的符号模式——最重要的是，贫困反转（$\alpha_{\text{pov}} < 0$，$\beta_{\text{pov}} > 0$）在两种设定中均存在。三个系数的符号发生变化（α_{Hispanic}，β_{Black}，β_{Hispanic}），但集约边际上的两个符号变化涉及的参数，其Wald置信区间均包含零，因此符号本身在统计上无法确定。总体而言，这一比较证实，NSECE抽样设计会实质性地影响点估计——这也说明采用伪后验方法是合理的——但核心实证发现仍然成立。

贫困反转。 这一反转是核心实证发现。广延边际贫困系数为 $\hat{\alpha}_{\text{pov}} = -0.324$（Wald 95%置信区间：$[-0.426, -0.221]$；设计效应比 =2.41）：社区贫困每增加一个标准差（约8.3个百分点），会使提供婴儿服务的对数优势降低0.324。密集型^边际贫困系数为 $\beta_{\text{pov}} = +0.090$（Wald 95%置信区间：$[+0.053, +0.127]$；设计效应比 =

表10. 来自加权 M3b的总体平均固定效应估计。“后验均值”是伪后验的后验均值。“朴素95%置信区间”是未经校正的MCMC后验的第2.5至第97.5百分位区间。“Wald 95%置信区间”使用经夹心估计量校正的方差（定理 3）。“设计效应比”是设计效应比（方程（3.27））。星号表示在Wald检验下5%水平的显著性。所有协变量均已标准化。

参数	后验均值	朴素95%置信区间	Wald 95%置信区间	DER
广延边际（$\boldsymbol{\alpha}$）				
截距	0.764	$[-1.109,\ 2.628]$	$[0.677,\ 0.851]^*$	2.04
贫困	−0.324	$[-2.205,\ 1.497]$	$[-0.426,\ -0.221]^*$	2.41
城市	0.442	$[-1.387,\ 2.224]$	$[0.375,\ 0.509]^*$	4.18
黑人	0.478	$[-1.459,\ 2.405]$	$[0.392,\ 0.565]^*$	1.54
西班牙裔	0.006	$[-1.902,\ 1.903]$	$[-0.095,\ 0.106]$	1.82
集约边际（$\boldsymbol{\beta}$）				
截距	−0.242	$[-2.058,\ 1.587]$	$[-0.272,\ -0.213]^*$	1.36
贫困	0.090	$[-1.713,\ 1.904]$	$[0.053,\ 0.127]^*$	1.61
城市	−0.047	$[-1.858,\ 1.768]$	$[-0.087,\ -0.007]^*$	3.18
黑人	−0.017	$[-1.851,\ 1.760]$	$[-0.052,\ 0.019]$	1.38
西班牙裔	−0.097	$[-1.974,\ 1.716]$	$[-0.130,\ -0.064]^*$	1.14
过度离散 $\log \kappa$	1.919	$[1.866,\ 1.972]$	$[1.861,\ 1.976]^*$	2.58

1.61）：在通过门槛的中心中，同样的贫困程度上升会使 IT 份额的 logit 增加 0.090。两个区间均不包含零，符号明显相反；根据夹心校正后验抽样值，反转在总体平均值处成立的后验概率为 Pr(α_{pov} <0 and $\beta_{\text{pov}} > 0 \mid$ 数据) $= 1.000$（起决定作用的边际 β_{pov} 距离零 4.8 个夹心标准差；二元正态近似给出的互补概率约为 10^{-6}）。

朴素区间与Wald区间的对比：解决先验主导问题。 表10中朴素区间与Wald区间之间的对比 表10 清楚地实证展示了先验主导现象。对于 $\boldsymbol{\alpha}$和 $\boldsymbol{\beta}$，朴素可信区间在logit尺度上的跨度约为3.7个单位——宽到每个参数的区间都同时包含正值和负值，导致无法进行显著性检验。相比之下，Wald区间的宽度为0.06–0.21个单位，因此可以得出明确的推断。

出现这种差异，是因为总体平均系数的朴素后验分布主要受先验而非数据支配。在层级模型中，

表11。未加权（M3b）与调查加权（M3b-W）固定效应后验均值的比较。“变化量”为 M3b-W 减去 M3b。“变化量/标准误”将变化量除以夹心标准误进行标准化，提供衡量加权影响的无尺度指标。“Wald 95%置信区间”是 M3b-W 的三明治校正区间（定理 3）。星号表示在 5% 水平上具有显著性。所有协变量均已标准化。

参数	M3b	M3b-W	变化量	变化量/标准误	Wald 95%置信区间
广延边际（$\boldsymbol{\alpha}$）					
截距	$+0.558$	$+0.764$	$+0.206$	$+4.64$	$[0.677,\ 0.851]^*$
贫困	-0.220	-0.324	-0.103	-1.97	$[-0.426,\ -0.221]^*$
城市	$+0.264$	$+0.442$	$+0.178$	$+5.19$	$[0.375,\ 0.509]^*$
黑人	$+0.248$	$+0.478$	$+0.230$	$+5.20$	$[0.392,\ 0.565]^*$
西班牙裔	-0.146	$+0.006$	$+0.151$	$+2.94$	$[-0.095,\ 0.106]$
集约边际（$\boldsymbol{\beta}$）					
截距	-0.096	-0.242	-0.146	-9.65	$[-0.272,\ -0.213]^*$
贫困	$+0.040$	$+0.090$	$+0.051$	$+2.68$	$[0.053,\ 0.127]^*$
城市	-0.055	-0.047	$+0.008$	$+0.39$	$[-0.087,\ -0.007]^*$
黑人	$+0.102$	-0.017	-0.119	-6.51	$[-0.052,\ 0.019]$
西班牙裔	$+0.060$	-0.097	-0.157	-9.35	$[-0.130,\ -0.064]^*$
过度离散 对数 κ	$+1.661$	$+1.919$	$+0.257$	$+8.80$	$[1.861,\ 1.976]^*$

州随机效应$\boldsymbol{\delta}_{k,\ s}$ 几乎吸收了有关州特定系数的全部数据信息$\tilde{\alpha}_{k,\ s} = \alpha_k + \delta_{k,\ s}$（参见 Rabe-Hesketh 和 Skrondal，2006）；总体平均 α_k 仅通过随机效应分布得到识别，其有效样本量为 $S = 51$（州的数量），而不是 $N = 6,\ 785$（服务提供者的数量）。因此，边际后验方差 $[\boldsymbol{\Sigma}_{\text{MCMC}}]_{pp}$ 仍接近先验方差，而不是收敛于数据信息精度 $[\mathbf{H}^{-1}{}_{\text{obs}}]_{pp}$。层级方差比——即 MCMC 后验方差与观测信息方差之比——对于所有固定效应均超过 600（参见第 3.5 节中的计算说明；完整的诊断表见表 D.4）。

三明治校正Wald区间完全绕过了这种分层膨胀。Wald标准误 $\sqrt{[\mathbf{V}_{\text{sand}}]_{pp}}$ 是一个频率学派量，由观测信息和聚类稳健外积推导得出（公式（3.25））（Binder，ˆ1983年；Williams & Savitsky，2021）：它衡量的是在对调查进行假设性重复模拟时 $\boldsymbol{\theta}$ 的抽样变异性，而非贝叶斯后验分布的宽度。对于分层调查模型中的固定效应，这种频率派精度是进行总体层面推断的恰当依据。朴素后验区间仍适用于

超参数（此处不适用夹心估计量）以及州随机效应（此处先验能够恰当地正则化小样本估计）。

设计效应比。 表表10中的设计效应比列揭示了一种具有实质意义且可解释的模式。最大设计效应比为 4.18，对应 α_{Urban}：由于 NSECE 的分层方案（国家早期保育与教育调查项目团队，2022年）明确与城市化程度相关——城市和农村地区的抽样率不同，因此该系数受调查设计的影响最大。第二大的设计效应比为 3.18，对应 β_{Urban}，这证实设计效应具有特征特异性，而非边际特异性。相反，最小的设计效应比为 1.14，对应 β_{Hisp}，表明调查设计对该系数的影响极小。总体平均设计效应比为 2.11，与 3.76 的 Kish设计效应大致一致。 图 4并列展示了全部 11 个参数的朴素区间和三明治区间，以及设计效应比柱状图。

过度离散。 估计的集中参数为 $\hat{\kappa} = \exp(1.919) \approx 6.81$ (Wald 置信区间（κ对数 $[1.861, 1.976]$）为，这证实了第2节中预示的显著过度离散。入学人数通常为 $n_i \approx 50$ 时，方差膨胀因子 $(n+\kappa)/(1+\kappa) \approx 57/8 \approx 7$约为二项基线的七倍。值得注意的是，对数 κ 是唯一一个朴素区间与 Wald 区间宽度相近的参数（0.106 对 0.115），因为 κ没有对应的州随机效应——它是一个由所有 $N_{\text{pos}} = 4$个正观测值直接提供信息的单一全局参数，因此标准的伯恩斯坦–冯·米塞斯近似成立，也不会出现先验主导。

5.3.州特定异质性与跨边际独立性。总体平均系数描述了全国层面效应；州变化系数则揭示了这些效应在各辖区的变化位置和方式。

跨边际散点图与独立性。 图5绘制后验均值 $(\tilde{\alpha}_{\text{pov},\,s},\ \tilde{\beta}_{\text{pov},\,s})$ 涵盖全部51个州。在层级M2设定下，51个州中有48个（94%）落在左上象限（$\alpha < 0,\ \beta > 0$），而初步OLS分析中为23个（45%）（第2.2节）——这一显著增加是由层级收缩将小州估计值拉向负的总体平均广延边际系数所致。贫困系数的后验跨边际相关性为 $\hat{\varrho}^{\text{cross}}_{\text{pov}} \approx 0.021$，与零无法区分，表明贫困效应的广延边际与集约边际主要通过在很大程度上相互独立的

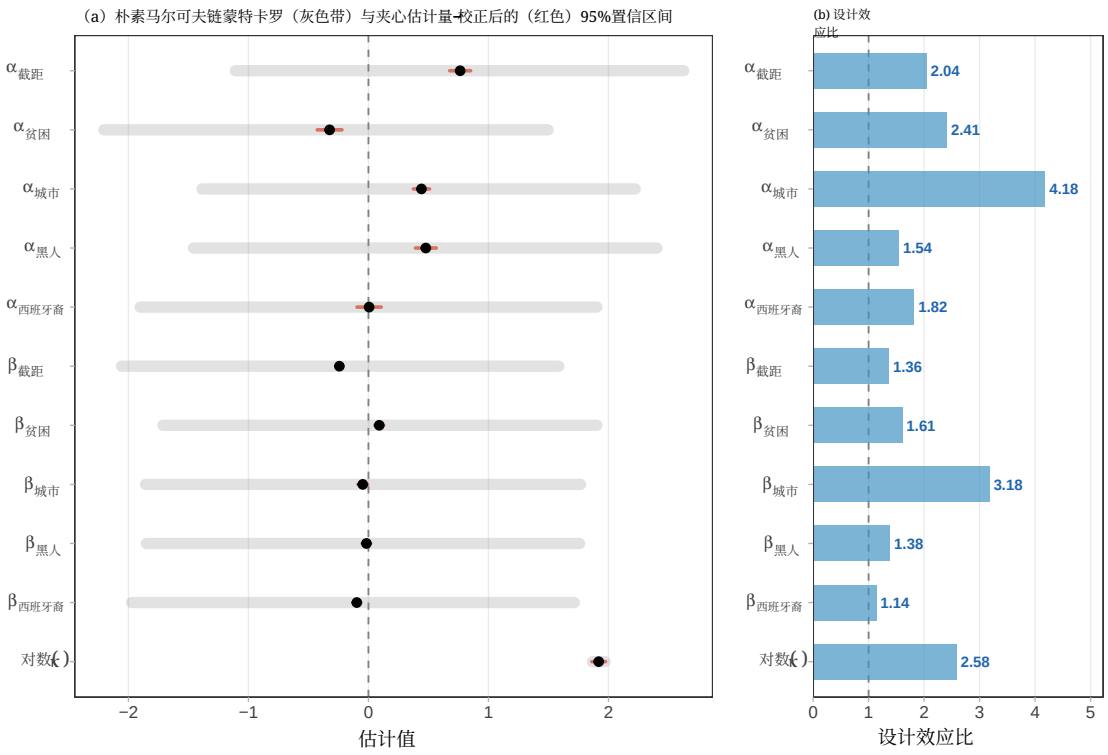

图 4. 三明治方差校正对固定效应推断的影响。左图：所有 11 个固定效应的朴素 MCMC 95% 可信区间（灰色）与经三明治校正的 Wald 95% 置信区间（彩色）。朴素区间反映先验宽度；Wald 区间反映根据调查设计调整后的数据信息量。右图：各参数的设计效应比（DER）。DER 为 1 表示不存在调查设计效应；观测范围为 1.14–4.18（均值为 2.11）。

机制发挥作用——这一结构性发现与分块对角可识别性结果一致，该结果出自命题 2。完整的州特定系数估计值见补充材料 D （ D.3 节）；反转概率图见 D.6 节。

5.4. **边际效应分解**。 本小节展示边际效应分解——这一由两部分结构支持的分析框架——通过将 logit 尺度系数转换为概率尺度量，并以命题 5中的对数尺度分解为基础。该分解将每个协变量的总效应分解为广延（可及性）和密集型（份额）部分，这种划分只有在两部分设定下才可实现。 表 12报告了结果。

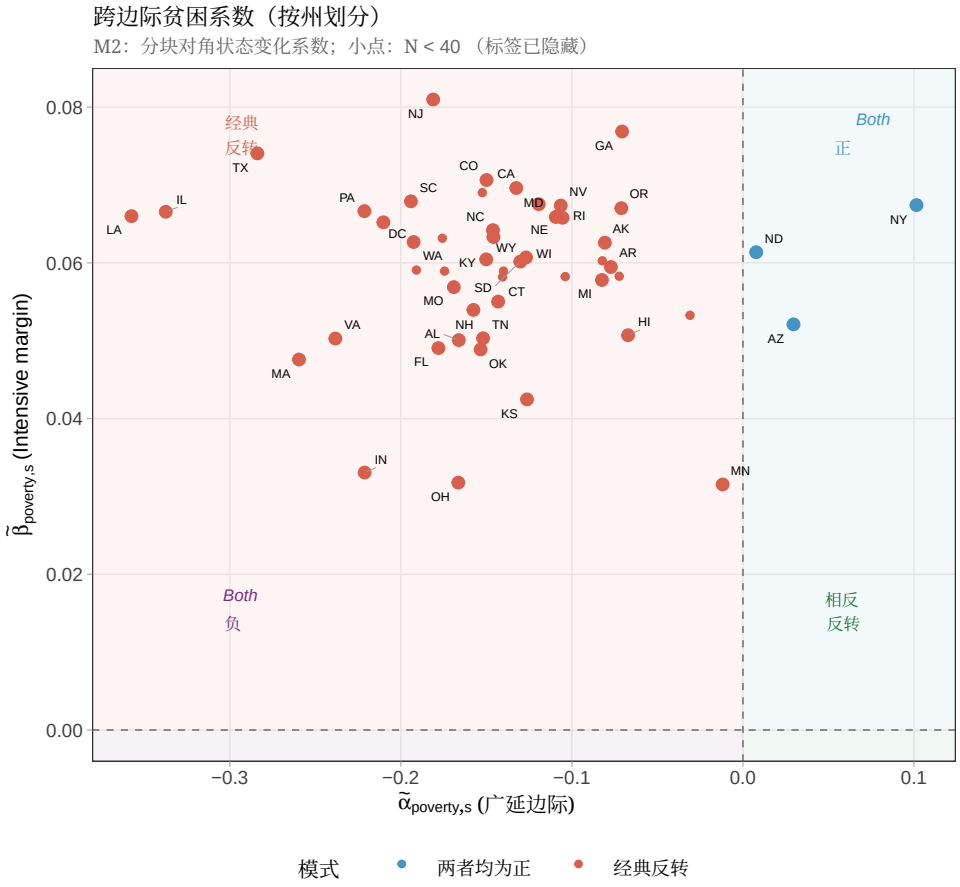

图5. 未加权 M2 模型（块对角 SVC）中各州特定贫困系数的跨边际散点图。每个点代表一个州的后验均值（$\tilde{\alpha}_{\text{pov},\ s}$，$\tilde{\beta}_{\text{pov},\ s}$）。没有标签的小点表示具有 $N < 40$ 的州。四个象限对应不同的贫困—入学模式：左上象限（阴影部分）包含 51 个州中的 48 个（94%），表明在层级收缩下，逆转模式几乎普遍存在。虚线表示总体平均系数（α_{pov}，β_{pov}）。跨边际相关性接近于零（$\hat{\varrho} \approx 0.021$），这意味着两个边际对贫困的响应主要通过彼此独立的机制产生。

从对数尺度到概率尺度。 命题 5 对对数预期入学率尺度上的边际效应进行了分解，其中两个作用渠道完全具有加性。要转换到概率尺度，请对 $E[Y_i/n_i] = q_i\, h_i$应用乘积法则：

$$\frac{\partial\,(q_i\,h_i)}{\partial x_{i,k}} = \underbrace{h_i\,q_i(1-q_i)\,\tilde{\alpha}_{k,s}}_{\text{extensive}} + \underbrace{q_i\,h_i\,\varepsilon_{h,i}\,(1-\mu_i)\,\tilde{\beta}_{k,s}}_{\text{intensive}} = q_i\,h_i\big(\text{LAE}_k + \text{LIE}_k\big). \quad (5.1)$$

平均边际效应条目在表 12中对每个分量按服务提供者与经 Cholesky 校准的后验抽样取平均 $\boldsymbol{\theta}^{*(m)}$ （定理 3），并将命题 5 的对数尺度分解映射到百分点尺度。

贫困。在广延边际上，贫困的平均边际效应为 -0.029 （95%置信区间：$[-0.038, -0.020]$）：社区贫困一个标准差的增加（≈ 8.3 百分点）会使服务任何婴儿的概率降低2.9个百分点。在集约边际上，平均边际效应 $= +0.015$ （95%置信区间：$[+0.009, +0.021]$）：在确实服务婴儿的中心中，同样的贫困增加会使IT占比提高1.5个百分点。总平均边际效应为 -0.014 （95%置信区间：$[-0.027, -0.001]$），其中广延边际占总绝对效应的66.0%：准入障碍占主导，但强度响应抵消了其中大约一半。这一分解——在任何单方程设定下都无法得到——体现了两部分结构（命题 5）的分析价值：门槛将准入障碍识别为主导作用渠道；若采用单方程模型，这一发现会被归并为一个减弱的净效应，即 -1.4 个百分点。

其他协变量。 一个值得注意的模式是，表 12 显示，四个协变量均呈现相反的广延效应和集约效应——反转模式并非贫困所独有。城市化程度提高可及性（$+3.9$百分点），但降低强度（-0.8 百分点），其中广延边际占总效应的 83.6%。黑人占比提高可及性（$+4.3$百分点），而强度效应可以忽略不计（广延份额 $= 93.9\%$）。西裔人口构成尤为突出：广延平均边际效应实际上为零（$+0.001$），因此全部效应都通过集约边际发挥作用；西裔人口构成越高，IT占比越低（-1.6 百分点）。这种以集约边际为主导的模式，与其他协变量以广延边际为主导的模式形成鲜明对比。在每种情况下，单方程模型都只会报告净总效应，从而掩盖相互抵消的作用渠道——而这正是门槛框架以及命题 5 中的平均边际效应公式旨在揭示的分解。

5.5.**政策调节**。 第5.3节 中记录的反转率和地理聚集 引出了一个自然的问题：可观测的 CCDF 政策能否解释为何反转在某些州表现强烈、而在另一些州却不存在? $\boldsymbol{\Gamma}_k$ M3b 中的矩阵将协变量效应的州级变异分解为政策解释部分和残差部分（命题4）。完整的40元素 $\boldsymbol{\Gamma}$矩阵见补充材料 D（第D.4节）；此处我们概述三个关键模式。

表 12. 模型M3b的平均边际效应（AME）分解。效应以概率尺度表示（协变量每增加一个标准差所引起的变化）。“广延份额”是归因于广延（获得）边际的总绝对效应百分比，计算如下：|广延|/(|广延| + |集约型|)。四个协变量的广延效应和集约效应方向均相反。95%区间为在 Cholesky校准抽取处计算的各 AME 分量后验百分位数的第 2.5 和第 97.5 百分位 $\boldsymbol{\theta}^{*(m)}$ (定理 3)。

协变量	广延	集约型	总计	广延份额
贫困	-0.029 $[-0.038, -0.020]$	$+0.015$ $[+0.009, +0.021]$	-0.014 $[-0.027, -0.001]$	66.0%
城市	$+0.039$ $[+0.034, +0.045]$	-0.008 $[-0.014, -0.001]$	$+0.032$ $[+0.022, +0.041]$	83.6%
黑人	$+0.043$ $[+0.035, +0.051]$	-0.003 $[-0.009, +0.003]$	$+0.040$ $[+0.029, +0.052]$	93.9%
西班牙裔	$+0.001$ $[-0.009, +0.010]$	-0.016 $[-0.021, -0.010]$	-0.016 $[-0.027, -0.004]$	2.8%

首先，分级报销减轻了广延边际上的贫困障碍（$\hat{\gamma}_{1,\,\mathrm{pov},\,\mathrm{TR}} = +0.495$, $\Pr(>0) = 0.962$），并提高了集约边际上的基线入学水平（$\hat{\gamma}_{2,\,\mathrm{int},\,\mathrm{TR}} = +0.271$, $\Pr(>0) = 0.985$）。其次，IT 费率附加项提高了基线参与率（$\hat{\gamma}_{1,\,\mathrm{int},\,\mathrm{IT}} = +0.728$, $\Pr(>0) = 0.996$），但出人意料地强化了广延边际上的贫困梯度（$\hat{\gamma}_{1,\,\mathrm{pov},\,\mathrm{IT}} = -0.488$, $\Pr(<0) = 0.982$），这表明附加项主要惠及中度贫困地区。第三，IT附加项在缩小婴儿入托人数的人口统计差距方面最为有效，尤其是在集约边际上惠及黑人社区的中心（$\hat{\gamma}_{2,\,\mathrm{Black},\,\mathrm{IT}} = +0.256$, $\Pr(>0) = 1.000$）。

这三个可观测政策变量解释了贫困系数州际差异的 4% 至 16%。这些关联是横截面的，不能支持因果主张；尽管如此， Γ分解展示了第3节中的分层结构如何将州际异质性划分为政策可归因部分和残差组成部分。

6. 讨论

摘要。本文针对含有结构性零值的调查加权有界计数，构建了一种分层障碍贝塔-二项模型（HBB）框架，将 Williams 和 Savitsky（2021）的伪后验方法扩展至两部分分层模

els. The t概要 i摘要 thodological contributi际边际 i协变量iance identi

—识别（命题 2）和DER引导的三明治校正（定理 3）—将在下文与可推广性、实践建议和局限性一并讨论。可推广性。第3节提出的HBB框架适用于结果变量为有界离散比例的情形——即具有已知分母的计数 $Y_i \in \{0, 1, \ldots, n_i\}$ ——且该结果嵌入具有结构性零值的层级结构中。以下五个领域展示了该框架在儿童保育入托之外的广泛适用性。(i) 在牙科流行病学中， Y_i 表示检查的 n_i 个牙面中龋坏牙面的数量；对于无龋个体，其零计数反映的是没有疾病，而非风险较低，因此存在结构性零值（Bandyopadhyay 等人，2011）。(ii) 在物种占域建模中， Y_i 表示某地点 n_i 次重复调查中的检测事件数量；对于物种确实不存在的地点，则存在结构性零值（MacKenzie等，2002）。(iii) 在医院床位管理中， Y_i 表示病区已占用床位数，分母为 n_i 张配置人员的床位；因人员短缺或感染控制措施而暂时停止接收入院的病区则存在结构性零值。(iv) 在保险登记中， Y_i 表示某计划—地区单元内 n_i 名符合资格的个人中的参保人数；对于未提供该计划的计划—地区组合，则存在结构性零值。(v) 在调查项目认可中， Y_i 表示某分量表中受访者认可的项目数，分母为 n_i 个施测项目；对于不适用该构念的受访者，则存在结构性零值（例如，向终身戒除者询问物质使用项目）。每个领域都具备共同促成HBB设定的三个特征：具有已知分母的有界支持集、活跃人群中的超二项变异，以及由不同于计数过程的定性机制产生的大量结构性零值。在所有情形下，门槛机制都将结构性参与决策与参与条件下的强度区分开来，而贝塔-二项核则能够处理相对于二项基线的过度离散。

推断贡献。 除建模框架外，本文还对推断作出了两项具体贡献。首先， 命题 2 证明观测层条件似然在各边际之间呈分块对角结构，因此跨边际协方差 $\boldsymbol{\Sigma}_{12}$ (第3节. 3) 是一个层级估计目标，其有效样本量为 S，而非 N。尽管频率学派方法原则上可以通过对 $\boldsymbol{\Sigma}_{12}$ 进行限制性最大似然估计来进行目标推断

边际似然（注释 1），当 $q \times q$ 组的交叉协方差来自 $S = 51$ 时，无约束估计并不稳健； 命题 3 表明，LKJ先验可提供有限样本正则化，使后验集中速度达到O（$S^{-1/2}$）。其次，采用 设计效应比引导的分块应用（定理 3）三明治校正流程，将 Williams 和 Savit- sky（2021）的伪后验框架扩展至两部分分层模型；当所有参数的 设计效应比恒定时，该流程将 Ghosal 等人（2020年）提出的标量设计效应乘数作为一种特例（公式（3.27））。

实践建议。 来自应用（第5节）与模拟（第4节）的综合证据表明，可为将贝叶斯层级模型拟合于复杂调查数据的应用研究者提供三条指导原则。首先，在决定是否加权之前检验信息性。遵循Pfeffermann（1993），在完整模型协变量向量条件下将抽样权重对结果变量进行回归，并检验权重系数是否显著；对完整协变量集进行条件控制至关重要，因为单变量检验会将信息性与混杂相混淆（Pfeffermann 等人，1998年）。对于两部分模型，应分别检验每个边际，因为设计可能对一个边际无信息、对另一个边际有信息（第2节）。

第二，同时报告朴素置信区间和三明治校正置信区间以及DER。本研究的实证DER范围为1.14–4.18（表10），而1.22–2.06的模拟宽度比（表7）提供了独立验证。线性插值启发式方法对DER 在中间收缩水平下的高估幅度最高可达50%；补充材料C（C节）中的精确共轭公式提供了更准确的基准。

第三，对于层级方差分量，应报告基于模型的后验区间，并进行先验敏感性分析。三明治校正从结构上不适用于那些不会产生个体层级得分贡献的参数，例如州效应标准差 τ 和跨边际相关性 ρ（第4.2节）；模拟也证实，这些参数的宽度比恒为1.00。我们在D.7节进行敏感性分析，在LKJ(η) 先验下 $\eta \in \{1, \ldots, 8\}$重新拟合：跨边际相关性仍接近于零（$\hat{\rho} \in [0.10,\ 0.19]$），且所有固定效应估计值均保持稳定，表明后验分布主要由数据主导。对于最大的3至5个簇，还建议进行敏感性分析，比较分块 Cholesky 校正与联合 Cholesky 校正（备注4；补充材料 C）。

局限性与未来方向。有几项局限性值得讨论。权重比 $w_{\max}/w_{\min} = 462$ 超过 $N^{1/2} \approx 82.4$ （第 3.5 节），因此，在信息性抽样条件下，伯恩斯坦–冯·米塞斯定理的充分条件 C3 并未得到充分满足；这使得三明治校正更适合作为一种有限样本稳健性度量，而非渐近保证，也凸显了第4节中模拟研究作为覆盖性质独立检验的重要性。贝塔-二项分量将 n_i 视为已知上界，但总入学人数本身可能会随社区特征而变化，这属于一种结果内生性。联合建模(Y_i, n_i) 可以解决这一问题，但需要同时设定总入学人数过程和 IT 占比过程；这一扩展留待未来研究。模型假设所有州共享一个集中参数κ ；因州而异的 κ_s 或可捕捉异质性过度离散，但要识别每个 κ_s ，需要对分母取值进行充分的州内重复观测 n_i （命题 B.24）。最后，三明治校正能为固定效应提供设计一致推断，但无法推广至层级方差分量，例如 $\boldsymbol{\Sigma}_{12}$；如何将设计一致推断推广到这些参数仍是一个开放问题，重复权重方法虽是一个可行方向，但计算成本较高。

替代调查方法。 本文采用的伪后验框架并非进行调查加权贝叶斯推断的唯一途径。多层伪最大似然（MPML；Rabe-Hesketh 和 Skrondal，2006）提供了一种频率学派对应方法，可将抽样权重直接纳入边际似然，从而避免伪似然构造。完全基于模型的方法将抽样机制纳入概率模型（Si 等人，2020年），在设计具有信息性时具有理论优势，但需要对选择过程进行额外设定。最后，平衡重复复制（BRR）或刀切重复权重方法（Rao 和 Wu，1988）无需分布假设即可提供基于设计的方差估计，不过，目前在每组重复权重下重新拟合贝叶斯分层模型的计算成本高得难以承受。将这些替代方法与伪后验三明治方法进行比较，是未来研究的一个重要方向。

结论。 本文提出一种适用于具有结构性零值的调查加权有界计数的贝叶斯层级障碍贝塔-二项框架。应用于2019年 NSECE，该模型揭示出一种贫困反转——社区贫困会降低

婴幼儿照护的可及性，但会提高参与中心的入托强度——在层级收缩下，这一现象在各州近乎普遍（51个州中的48个），且其广延边际约占总效应的三分之二。三明治校正与DER诊断为拟合于复杂调查数据的贝叶斯层级模型提供了一套可用于总体层面推断的实用工具。配套的R 软件包 hurdlebb （Lee，2026）及复现代码仓库使该方法论可立即供应用研究者使用。

软件与可复现性。 配套的 R 软件包 hurdlebb (Lee，2026) 实现了本文提出的完整方法，包括调查加权伪后验推断、通过 Cholesky 变换进行夹心方差校正，以及边际效应分解。该软件包可从 https://github.com/joonho112/hurdlebb 获取，文档见 https://joonho112.github.io/hurdlebb/。另一个复现包位于 https://github.com/joonho112/hurdlebb-replication 其中包含复现本文报告的所有结果所需的全部分析脚本，但须能够访问 NSECE 限制使用数据。完整的估计流程详见 算法 1；补充材料E（第 E 节）提供了计算细节，包括 Stan 配置、收敛诊断和运行时间基准。

数据可用性。 本文的实证结果基于 2019年全国早期照护与教育调查（NSECE）的数据。分析中使用的所有关键变量——包括州标识符、社区贫困率、城市化程度、种族／族裔构成、婴幼儿入托人数和总入学人数——仅可通过 NSECE限制使用文件（ICPSR研究38445，https://www.icpsr.umich.edu/web/ICPSR/studies/38445）获取。对于有意了解该方法但无法访问受限数据的研究人员，hurdlebb 软件包包含一个合成数据集（`nsece_synth`），该数据集保留了原始数据的基本分布特征。

资助。 本研究由美国卫生与公众服务部儿童与家庭管理局规划、研究与评估办公室（OPRE）通过90YE0346号资助向内华达大学拉斯维加斯分校（项目负责人：Gerilyn Slicker）提供支持，并向阿拉巴马大学提供子资助。所表达的观点仅为作者本人的观点，并不代表儿童与家庭管理局或美国卫生与公众服务部的观点。

参考文献

Adams, G.与Rohacek, M.（2010年）。在情境中理解质量：托儿中心、社区、市场与公共政策（技术报告）（ERIC 文献 ED511767）。Urban Institute。华盛顿特区。 [2, 5]。Altinisik, Y.（2023年）。通过新的多层异质障碍模型处理聚类计数数据的过度离散与零膨胀。《应用统计学杂志》，50（2），408–433。https : / / doi . org / 10 . 1080 / 02664763 . 2022 . 2096875[2]。Bandyopadhyay, D.、Reich, B. J.与Slate, E. H.（2011年）。用于龋齿聚类计数数据的空间贝塔-二项模型。《医学研究统计方法》，20（2），85–102。https://doi.org/10.1177/0962280210372453 [3, 42,61]。Betancourt, M.与Girolami, M.（2015年）。层级模型的哈密顿蒙特卡洛。载于S. K. Upadhyay、U. Singh、D. K. Dey与A. Loganathan（编），《贝叶斯方法论及其应用的当前趋势》（第79–101页）。Chapman & Hall/CRC。https://doi.org/10.1201/b18502 [17, 32]。Binder, D. A.（1983年）。论复杂抽样调查中渐近正态估计量的方差。《国际统计评论》，51（3），279–292。https : / / doi.org/10.2307/1402588 [3, 4, 28, 36, 67, 83, 99, 117]。Carpenter, B.、Gelman, A.、Hoffman, M. D.、Lee, D.、Goodrich, B.、Betancourt, M.、Brubaker, M.、Guo, J.、Li, P.与Riddell, A.（2017年）。Stan：一种概率编程语言。《统计软件杂志》，76（1），1–32。https : // doi.org/10.18637/jss.v076.i01 [17, 32, 67, 104]。Cragg, J. G.（1971）。有限因变量的若干统计模型及其在耐用品需求中的应用。《计量经济学》，39（5），829–844。https://doi.org/10.2307/1909582 [2, 12, 52, 71]。Dobbins, D.、Tercha, J.、McCready, M.与Liu, A.（2016年）。托儿服务荒漠：制定托儿服务供需解决方案（技术报告）。Child Care Aware of America。弗吉尼亚州阿灵顿。 [2, 5, 8]。Gelfand, A. E.、Kim, H.-J.、Sirmans, C. F.与Banerjee, S.（2003年）。基于空间变系数过程的空间建模。《美国统计学会杂志》，98（462），387–396。https://doi.org/10.1198/016214503000170 [14]。Gelman, A.（2006）。层级模型中方差参数的先验分布（对Browne与Draper文章的评论）。《贝叶斯分析》，1（3），515–534。https://doi.org/10.1214/06-BA117A [22]。

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). 逻辑回归及其他回归模型的弱信息默认先验分布。《应用统计年鉴》，2(4), 1360–1383。https://doi.org/10.1214/08-AOAS191[22, 79, 81]. Ghosal, S., Lau, T. S., Gaskins, J., & Kong, M. (2020). 时空计数数据的分层混合效应障碍模型及其在识别影响卫生专业人员短缺的因素中的应用。《皇家统计学会会刊：C辑（应用统计）》，69 (5), 1121–1144。https://doi.org/10. 1111/rssc.12434 [2, 3, 17, 43, 75]. Griffiths, D. A. (1973). 贝塔-二项分布的最大似然估计及其在疾病病例总数家庭分布中的应用。《生物统计学》，29 (4), 637–648。https://doi.org/10.2307/2529131 [10, 52, 54]. Hall, D. B. (2000). 带随机效应的零膨胀泊松回归和二项回归：案例研究。《生物统计学》，56 (4), 1030–1039。https://doi.org/10.1111/j.0006- 341X.2000.01030.x [3]. Hoffman, M. D., & Gelman, A. (2014). 无转向采样器：在哈密顿蒙特卡洛中自适应设定路径长度。《机器学习研究杂志》，15(47), 1593–1623 [24, 104]. Kleijn, B. J. K., & van der Vaart, A. W. (2012). 模型设定错误条件下的伯恩斯坦-冯·米泽斯定理。《电子统计学杂志》，6, 354–381。https:// doi.org/10.1214/12-EJS675 [77, 80]. Lambert, D. (1992). 零膨胀泊松回归及其在制造缺陷中的应用。《计量技术学》，34 (1), 1–14。https : / / doi . org / 10 . 1080 / 00401706.1992.10485228 [3, 71]. Lee, J. (2026). hurdlebb：用于有界离散比例的障碍贝塔-二项模型[R 软件包版本 0.1.0]. https://github.com/joonho112/hurdlebb [4, 24, 45]. Lewandowski, D., Kurowicka, D., & Joe, H. (2009). 基于藤 copula 和扩展洋葱法生成随机相关矩阵。《多元分析杂志》，100 (9), 1989–2001。https://doi.org/10.1016/j.jmva.2009.04.008 [23,76, 79, 81]. Lumley, T. (2004). 复杂抽样调查样本分析。《统计软件杂志》，9 (8), 1–19。https://doi.org/10.18637/jss.v009.i08 [115]. Lumley, T. (2010). 复杂抽样调查：使用 R 进行分析指南。约翰·威立父子出版公司。https://doi.org/10.1002/9780470580066 [115]。

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Lang- timm, C. A. (2002). 当检测概率小于 1 时估计地点占用率。《生态学》，83 (8), 2248–2255。https://doi.org/10.1890/ 0012-9658(2002)083[2248:ESORWD]2.0.CO;2 [42]. Malik, R., Hamm, K., Schochet, L., Novoa, C., Workman, S., & Jessen-Howard, S. (2018). 2018年美国的儿童保育荒漠（技术报告）。美国进步中心。华盛顿特区 [2, 5, 8]. Mullahy, J. (1986). 若干修正计数数据模型的设定与检验。《计量经济学杂志》，33 (3), 341–365。https : / / doi . org / 10 . 1016 / 0304 - 4076(86)90002-3 [2, 5, 12, 52, 56, 71, 74]. 全国早期照护与教育调查项目团队。(2022). 2019年全国早期照护与教育调查数据收集与抽样方法报告（技术报告）（OPRE报告2022-118）。规划、研究与评估办公室、儿童与家庭管理局、美国卫生与公众服务部。华盛顿特区。 [2, 4, 37, 85, 92]. Neelon, B. (2019). 基于波利亚-伽马混合的贝叶斯零膨胀负二项回归。《贝叶斯分析》，14 (3), 829–855。https : / / doi . org / 10 . 1214/18-BA1132 [2]. Neelon, B., Ghosh, P., & Loebs, P. F. (2013). 用于探索急诊就诊地理变异的空间泊松障碍模型。《英国皇家统计学会会刊：A辑（社会统计）》，176 (2), 389–413。https://doi.org/10.1111/j.1467-985X.2012.01039.x [2]. Neelon, B., O'Malley, A. J., & Smith, V. A. (2016). 卫生服务研究中零修正计数数据和半连续数据的建模第 1 部分：背景与概述。《医学统计学》，35 (27), 5070–5093。https://doi.org/10.1002/ sim.7050 [2, 71]. Neelon, B. H., O'Malley, A. J., & Normand, S.-L. T. (2010). 重复测量零膨胀计数数据的贝叶斯模型及其在门诊精神科服务使用中的应用。《统计建模》，10 (4), 421–439。https://doi.org/10.1177/1471082X0901000404 [2]. Pfeffermann, D. (1993). 抽样权重在调查数据建模中的作用。《国际统计评论》，61 (2), 317–337。https : / /doi . org /10 . 2307 /1403631 [3, 4, 6, 43, 92–94]。

Pfeffermann, D., Krieger, A. M., & Rinott, Y. (1998). 信息性概率抽样下复杂调查数据的参数分布。Statistica Sinica, 8 (4), 1087–1114 [19, 21, 43]. Prentice, R. L. (1986). 使用扩展贝塔-二项分布进行二元回归，并讨论协变量测量误差引起的相关性。Journal of the American Statistical Association, 81 (394), 321–327. https: //doi.org/10.1080/01621459.1986.10478275 [5, 6, 10, 52, 58, 68, 106]. Pustejovsky, J. E., & Tipton, E. (2018). 固定效应模型中聚类稳健方差估计和假设检验的小样本方法。Journal of Business & Economic Statistics, 36 (4), 672–683. https://doi.org/10.1080/ 07350015.2016.1247004 [28]. Rabe-Hesketh, S., & Skrondal, A. (2006). 复杂调查数据的多层建模。Journal of the Royal Statistical Society: Series A (Statistics in Society), 169 (4), 805–827. https://doi.org/10.1111/j.1467-985X.2006.00426.x [19, 36,44, 78]. Rao, J. N. K., & Scott, A. J. (1981). 复杂样本调查分类数据分析：列联表中拟合优度和独立性的卡方检验。Journal of the American Statistical Association, 76 (374), 221–230. https://doi.org/10.2307/2287815 [21]. Rao, J. N. K., & Wu, C. F. J. (1988). 复杂调查数据的重抽样推断。Journal of the American Statistical Association, 83 (401), 231–241. https: //doi.org/10.1080/01621459.1988.10478591 [44]. Savitsky, T. D., & Toth, D. (2016). 信息性抽样下的贝叶斯估计。Electronic Journal of Statistics, 10 (1), 1677–1708. https://doi.org/10.1214/16- EJS 1153 [3, 4, 19, 25, 27, 78, 80, 92, 107]. Si, Y., Pillai, N. S., & Gelman, A. (2020). 贝叶斯非参数加权抽样推断。Bayesian Analysis, 15 (2), 603–625. https://doi.org/10.1214/19- BA1164 [44]. Sipple, J. W., McCabe, L. A., & Casto, H. G. (2020). 纽约州的托儿服务荒漠：学前班实施情况及与婴幼儿照护容量相关的社区因素。Early Childhood Research Quarterly, 51, 167–177. https://doi.org/10.1016/j.ecresq.2019.10.007 [8]. Stan Development Team. (2024). CmdStanR：CmdStan 的 R 接口 [R 软件包版本 0.8.1]. https://mc-stan.org/cmdstanr/ [104]。

Vehtari, A., Gelman, A., & Gabry, J. (2017). 使用留一交叉验证和 WAIC 进行实用的贝叶斯模型评估。Statistics and Computing, 27(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4 [24, 32, 34]. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). 秩标准化、折叠与局部化：一种用于评估马尔可夫链蒙特卡罗收敛性的改进 R 方法（附讨论）。Bayesian Analysis, 16(2), 667–718. https: //doi.org/10.1214/20-BA 1221 [24, 27, 32, 105]. Wagner, B., Riggs, P., & Mikulich-Gilbertson, S. (2015). 分布选择在物质使用数据建模中的重要性：负二项分布、贝塔-二项分布和零膨胀分布的比较。The American Journal of Drug and Alcohol Abuse, 41 (6), 489–497. https://doi.org/10.3109/ 00952990.2015.1056447 [3, 5]. Wen, C.-C., Paul, R., Hunt, K. J., O'Malley, A. J., Li, H., Hill, E., Malek, A. M., & Neelon, B. (2025). 一种用于过度离散二项数据的贝叶斯零膨胀空间变系数模型 [在线优先发表。美国国立卫生研究院作者手稿]。Journal of the Royal Statistical Society: Series A (Statistics in Society). https://doi.org/10.1093/jrsssa/qnaf056 [3]. Williams, M. R., & Savitsky, T. D. (2021). 复杂抽样下伪贝叶斯推断的不确定性估计。International Statistical Review, 89(1), 72–107. https://doi.org/10.1111/insr. 12376 [3, 4, 19, 25, 36, 42, 43, 78,80, 84, 87]. Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). 纵向数据模型：广义估计方程方法。Biometrics, 44(4), 1049–1060. https://doi.org/10.2307/2531734 [116]。

在线补充材料

用于调查加权有界计数的贝叶斯分层障碍贝塔-二项模型及其在儿童保育入学中的应用

JoonHo Lee

目录

在线补充材料 52 附录 A. 跨栏贝塔-二项模型的数学性质 52 A.1. 定理1的证明 52 A.2. 截断贝塔-二项分布的高阶矩 54 A.3. 无条件方差分解 56 A.4. 强度函数的解析性质 59 A.5. 对数尺度边际效应分解 64 附录 B. 可识别性与费舍尔信息 66 B.1. 得分函数 67 B.2. 费舍尔信息与分块对角性 70 B.3. 识别 71 附录 C. 调查加权推断 77 C.1. 伪似然与可分性 77 C.2. 伪后验适定性 78 C.3. Bernstein–von Mises近似 80 C.4. 夹心方差估计量 83 C.5. Cholesky仿射变换 87 C.6. 设计效应比率分类与诊断 89 附录 D. 扩展实证结果 92 D.1. 调查设计特征 92 D.2. 信息性评估 92 D.3. 州级贫困系数 96 D.4. 完整政策调节参数 97 D.5. 设计效应诊断 97 D.6. 反转概率图 99

D.7. LKJ 先验敏感性 101附录 E. 模拟与计算细节 102E.1. 模型层级与架构 103E.2. MCMC 配置 104E.3. 数据生成过程与校准 106E.4. 补充诊断图与设计效应分析 107E.5. 模型设定错误稳健性 111E.6. 频率学派情境化 115E.7. 覆盖率差距分解 117E.8. 95% 名义水平下的覆盖率 118

附录A. 跨栏贝塔-二项模型的数学性质

本附录汇集了关于第3节中讨论的跨栏贝塔-二项模型数学性质的完整证明及支持性结果。采用均值–精度参数化的贝塔-二项分布遵循 Griffiths (1973年) 和 Prentice (1986) 的方法；障碍模型构造遵循 Mullahy (1986) 和 Cragg (1971) 的方法。全文沿用第 3.1 节中确立的记号。

A.1. **定理 1的证明。** 下面给出强度函数$h(\mu) = \mu/(1-p_0)$ 对每个 μ 都严格递增的完整证明，其中 $n \geq 2$， $\kappa > 0$。回顾一下，$p_0(n,\ \mu,\ \kappa) = \prod_{j=0}^{n-1}[(1-\mu)\kappa + j]/(\kappa + j)$ 是贝塔-二项分布的零概率，定义见 (3.3)。

辅助量。 令 $b = (1-\mu)\kappa$ 并定义

$$\Lambda = \kappa \sum_{j=0}^{n-1} \frac{1}{(1-\mu)\kappa + j} > 0, \qquad \Lambda_2 = \kappa^2 \sum_{j=0}^{n-1} \frac{1}{\big[(1-\mu)\kappa + j\big]^2} > 0. \tag{A.1}$$

注意， $\partial p_0/\partial \mu = -p_0\,\Lambda$ （通过对 $\ln p_0 = \sum_{j=0}^{n-1} \ln[(1-\mu)\kappa + j] - \ln[\kappa + j]$ 求导得到），而 $\partial \Lambda/\partial \mu = \Lambda_2$ （因为 $\partial b/\partial \mu = -\kappa$）。

引理 A.1 (倒数和不等式)。 设 $b > 0$、 $n \geq 1$，并令 $S_1 = \sum_{j=0}^{n-1} (b+j)^{-1}$和 $S_2 = \sum_{j=0}^{n-1} (b+j)^{-2}$，使得 $\Lambda = \kappa\, S_1$ 且 $\Lambda_2 = \kappa^2\, S_2$。

(i) 对于 $n = 1$: $S_1^2 = S_2$，因此 $\Lambda^2 = \Lambda_2$。

(ii) 对于 $n \geq 2$: $S_1^2 > S_2$，因此 $\Lambda^2 > \Lambda_2$。

证明。写作 $a_j = (b+j)^{-1}$。展开和的平方，

$$S_1^2 = \Big(\sum_{j=0}^{n-1} a_j\Big)^2 = \sum_{j=0}^{n-1} a_j^2 + 2\sum_{0\le i<j\le n-1} a_i\,a_j = S_2 + 2\sum_{0\le i<j\le n-1} \frac{1}{(b+i)(b+j)}.$$

当 $n=1$ 时，交叉乘积和为空，因此得到 $S_1^2 = S_2$。当 $n\ge 2$ 时，每一项 $1/[(b+i)(b+j)]$ 都严格为正，因此 $S_1^2 > S_2$。$\square$

定理 *1* 的证明[定理 *1*](#)。定义分子函数

$$\Phi(\mu) = \big(1 - p_0(\mu)\big) - \mu\, p_0(\mu)\,\Lambda(\mu). \tag{A.2}$$

对 $h = \mu(1-p_0)^{-1}$ 应用商法则，并结合 $\partial p_0/\partial\mu = -p_0\,\Lambda$，可得

$$\frac{\partial h}{\partial \mu} = \frac{\Phi(\mu)}{(1-p_0)^2}. \tag{A.3}$$

由于对于 $\mu\in$，$(1-p_0)^2 > 0$ （0，1），因此 $\partial h/\partial\mu$ 的符号与 $\Phi(\mu)$ 的符号相同。证明分三步进行。

步骤 *1*：边界值。当 $\mu = 0$时，(3.3) 中的每个因子都等于 1，因此 $p_0(0) = 1$。因此 $\Phi(0) = (1-1) - 0\cdot 1\cdot\Lambda(0) = 0$。

步骤 *2*：求导 Φ。对式（A.2）应用乘积法则求导，得 s

$$\Phi'(\mu) = p_0\,\Lambda - \big[p_0\,\Lambda + \mu\,\tfrac{d}{d\mu}(p_0\,\Lambda)\big] = -\mu\,\frac{d}{d\mu}\big(p_0\,\Lambda\big).$$

现在 $d(p_0\,\Lambda)/d\mu = (\partial p_0/\partial\mu)\,\Lambda + p_0\,(\partial\Lambda/\partial\mu) = -p_0\,\Lambda^2 + p_0\,\Lambda_2 = p_0(\Lambda_2 - \Lambda^2)$，因此

$$\Phi'(\mu) = -\mu\,p_0\big(\Lambda_2 - \Lambda^2\big) = \mu\,p_0\big(\Lambda^2 - \Lambda_2\big). \tag{A.4}$$

步骤 *3*：应用[引理 *A.1*](#)。

- 情形 $n=1$。由引理 A.1(i) 得 $\Lambda^2 = \Lambda_2$，因此 $\Phi'(\mu) = 0$ 恒成立。结合 $\Phi(0) = 0$，可得对于所有 $\mu\in(0,1)$，$\Phi(\mu) = 0$，从而 $\partial h/\partial\mu = 0$。等价地，当 $n=1$ 时，$p_0 = 1-\mu$，因此 $h(\mu) = \mu/(1-(1-\mu)) = 1$ 恒成立。
- 情形 $n\ge 2$。[引理 A.1](#)（ii）给出 $\Lambda^2 > \Lambda_2$。由于 $\mu>0$ 且 $p_0>0$ 在 $(0,1)$上，方程（A.4）得出 $\Phi'(\mu) > 0$ 对所有 $\mu\in(0,1)$均成立。

由于 $\Phi(0) = 0$ 且 Φ 在 (0，1) 上严格递增，因此我们得出 $\Phi(\mu) > 0$ 对所有 $\mu\in$ (0，1)都成立，故

$$\frac{\partial h}{\partial\mu} = \frac{\Phi(\mu)}{(1-p_0)^2} > 0. \qquad \square$$

A.2. **截断贝塔-二项分布的高阶矩。记号。**在本小节中，$Y \sim \text{BetaBin}(n,\ \mu,\ \kappa)$ 表示标准（未截断）贝塔-二项分布，其中 $a=\mu\kappa,\ b=(1-\mu)\kappa$，而 $Y^{+} \sim \text{ZT-BetaBin}(n,\ \mu,\ \kappa)$ 表示其零截断版本。我们记 $p_0=\mathbb{P}(Y=0)$、$V_{\mathrm{BB}}=\mathrm{Var}(Y)=n\mu(1-\mu)(n+\kappa)/(1+\kappa)$ 和 $h=\mu/(1-p_0)$。

引理 A.2（二阶阶乘矩）。对于 $Y \sim$ 贝塔-二项分布$(n,\ \mu,\ \kappa)$，

$$\boxed{\mathbb{E}\left[Y(Y-1)\right]=n(n-1)\cdot\frac{\mu(\mu\kappa+1)}{\kappa+1}.} \tag{A.5}$$

证明。贝塔-二项分布可表示为贝塔混合（Griffiths，1973年）：令 $p \sim$ 贝塔分布$(a,\ b)$，其中 $a=\mu\kappa,\ b=(1-\mu)\kappa$，且 $Y \mid p \sim$ 二项分布$(n,\ p)$。由于 $\mathbb{E}[Y(Y-1) \mid p]=n(n-1)p^2$，由塔式性质可得

$$\mathbb{E}\left[Y(Y-1)\right]=n(n-1)\,\mathbb{E}[p^2].$$

$p \sim$ 贝塔分布$(a,\ b)$ 的二阶矩为 $\mathbb{E}[p^2]=a(a+1)/[(a+b)(a+b+1)]$,，利用上升阶乘恒等式 $\mathbb{E}[p^k]=a^{(k)}/(a+b)^{(k)}$（其中 $k=2$）。代入 $a=\mu\kappa$、$a+b=\kappa$：

$$\mathbb{E}\left[Y(Y-1)\right]=n(n-1)\cdot\frac{\mu\kappa(\mu\kappa+1)}{\kappa(\kappa+1)}=n(n-1)\cdot\frac{\mu(\mu\kappa+1)}{\kappa+1}. \qquad \square$$

一致性检验。由（A.5），二阶矩为 $\mathbb{E}[Y^2]=\mathbb{E}[Y(Y-1)]+\mathbb{E}[Y]=n(n-1)\mu(\mu\kappa+1)/(\kappa+1)+n\mu$。我们验证 $\mathrm{Var}(Y)=\mathbb{E}[Y^2]-(n\mu)^2$：

$$\mathbb{E}[Y^2]-n^2\mu^2=\frac{n(n-1)\mu(\mu\kappa+1)+n\mu(\kappa+1)-n^2\mu^2(\kappa+1)}{\kappa+1}.$$

展开分子因子：$n\mu\left[(n-1)(\mu\kappa+1)+(\kappa+1)-n\mu(\kappa+1)\right]$。括号内化简为

$$(n-1)\mu\kappa+(n-1)+(\kappa+1)-n\mu\kappa-n\mu=(1-\mu)(n+\kappa).$$

因此，$\mathrm{Var}(Y)=n\mu(1-\mu)(n+\kappa)/(\kappa+1)=V_{\mathrm{BB}}$。$\square$

引理 A.3（一般 r 阶乘阶矩）。对于 $Y \sim$ 贝塔-二项分布(n,μ,κ)，

$$\mathbb{E}\left[\prod_{j=0}^{r-1}(Y-j)\right]=\frac{n^{(r)}\cdot(\mu\kappa)^{(r)}}{\kappa^{(r)}}, \tag{A.6}$$

其中 $c^{(r)}=c(c+1)\cdots(c+r-1)=\Gamma(c+r)/\Gamma(c)$ 是上升阶乘。

证明。由二项分布的 $\mathbb{E}[\prod_{j=0}^{r-1}(Y-j) \mid p]=n^{(r)}p^r$ 和贝塔分布的 $\mathbb{E}[p^r]=(\mu\kappa)^{(r)}/\kappa^{(r)}$ 可得。$\square$

命题 A.1（在零点消失的函数的截断恒等式）。 对于任意函数 g 满足 $g(0) = 0$,

$$\mathbb{E}\left[g(Y) \mid Y > 0\right] = \frac{\mathbb{E}\left[g(Y)\right]}{1 - p_0}. \tag{A.7}$$

特别地， $\mathbb{E}[Y^k \mid Y > 0] = \mathbb{E}[Y^k]/(1 - p_0)$ 对所有 $k \geq 1$ 均成立。

证明。 由于 $g(Y) \cdot \mathbf{1}(Y = 0) = g(0) \cdot \mathbf{1}(Y = 0) = 0$, 因此有 $\mathbb{E}[g(Y)] = \mathbb{E}[g(Y) \cdot \mathbf{1}(Y > 0)]$。除以 $\mathbb{P}(Y > 0) = 1 - p_0$ 即可得出结论。□

推论 A.1（条件二阶阶乘矩）。

$$\boxed{\mathbb{E}\left[Y(Y-1) \mid Y > 0\right] = \frac{n(n-1)\mu(\mu\kappa + 1)}{(\kappa + 1)(1 - p_0)}.} \tag{A.8}$$

P命题 A.2（截断BB方差）。 对于 $Y^+ \sim$ ZT-贝塔-二项分布$(n, \ \mu, \ \ \kappa)$,

$$\boxed{\mathrm{Var}(Y \mid Y > 0) = \frac{V_{\mathrm{BB}}(1 - p_0) - n^2\mu^2\, p_0}{(1 - p_0)^2}.} \tag{A.9}$$

证明。 根据截断恒等式（命题 A.1）,

$$\mathbb{E}[Y^2 \mid Y > 0] = \frac{\mathbb{E}[Y^2]}{1 - p_0}, \qquad \mathbb{E}[Y \mid Y > 0] = \frac{n\mu}{1 - p_0}.$$

Then

$$\mathrm{Var}(Y \mid Y > 0) = \frac{\mathbb{E}[Y^2]}{1 - p_0} - \frac{n^2\mu^2}{(1 - p_0)^2} = \frac{V_{\mathrm{BB}} + n^2\mu^2}{1 - p_0} - \frac{n^2\mu^2}{(1 - p_0)^2}.$$

将分子中含 $n^2\mu^2$ 的项通分至公分母 $(1 - p_0)^2$:

$$= \frac{V_{\mathrm{BB}}}{1 - p_0} + n^2\mu^2 \cdot \frac{(1 - p_0) - 1}{(1 - p_0)^2} = \frac{V_{\mathrm{BB}}}{1 - p_0} - \frac{n^2\mu^2\, p_0}{(1 - p_0)^2} = \frac{V_{\mathrm{BB}}(1 - p_0) - n^2\mu^2\, p_0}{(1 - p_0)^2}. \quad \square$$

备注 A.1. 公式（A.9）揭示了截断的两种相互竞争的影响。第一项 $V_{\mathrm{BB}}/\left(1 - p_0\right)$ 通过重新缩放概率使方差增大。第二项 $n^2\mu^2 p_0/\left(1 - p_0\right)^2$ 则会使方差减小，因为截断后的分布在零点处损失了概率质量，从而使均值远离零，并使相对于该偏移均值的有效离散程度降低。

替代形式。

形式*A*（因式分解）：

$$\mathrm{Var}(Y \mid Y > 0) = \frac{n\mu}{1 - p_0}\left[\frac{(1 - \mu)(n + \kappa)}{1 + \kappa} - \frac{n\mu\, p_0}{1 - p_0}\right]. \tag{A.10}$$

这表示方差为 $\mathbb{E}[Y \mid Y > 0]$ 倍的“离散因子”。

形式 *B*（通过阶乘矩）：

$$\mathrm{Var}(Y \mid Y>0) = \frac{\mathbb{E}[Y(Y-1)]}{1-p_0} + \frac{n\mu}{1-p_0} - \frac{n^2\mu^2}{(1-p_0)^2}. \tag{A.11}$$

极限情形的验证。

情形 *1*： $\kappa \to \infty$ （截断二项分布极限）。当 $\kappa \to \infty$ 时， $V_{\mathrm{BB}} \to n\mu(1-\mu)$ 且 $p_0 \to (1-\mu)^n$。该公式化为已知的零截断二项分布方差。对于 $n=1$: $Y \mid Y>0 = 1$ 是确定的，因此 $\mathrm{Var}=0$。在公式中：分子 $=\mu(1-\mu)\cdot\mu - \mu^2(1-\mu) = 0$。 □

其余极限情形可用类似方法验证：当 $p_0 \to 0$ 时，方差化简为 V_{BB}；而对于 $n=1$ ，两边均为零，因为 $Y \mid Y>0 = 1$ 是确定的。 □

数值交叉检验。考虑 $n=5$、 $\mu=0.3$、 $\kappa=2$ （因此 $a=0.6$、 $b=1.4$）。直接计算得 $p_0 = \prod_{j=0}^{4}(1.4+j)/(2+j) = 0.376992$、 $\mathbb{E}[Y]=1.5$、 V_{BB} =2.45 和 $\mathbb{E}[Y^2]=4.70$。截断矩为 $\mathbb{E}[Y \mid Y>0] = 1.5/0.623008$ =2.40770、 $\mathbb{E}[Y^2 \mid Y>0] = 4.70/0.623008 = 7.54409$，且 $\mathrm{Var}(Y \mid Y>0) = 7.54409 - 2.40770^2 = 1.74707$。形式 A 得到 $(2.45\times 0.623008 - 2.25\times 0.376992)/0.623008^2 =$ 1.74709，与结果在五位有效数字上吻合。 □

A.3. **无条件方差分解**。我们推导 Y_i 在跨栏贝塔-二项模型下的无条件方差、其方差均值比、相对于二项分布的过度离散指数，以及变异系数。全文中，我们记$V_{\mathrm{BB},\,i} = n_i\mu_i(1-\mu_i)(n_i+\kappa)/(1+\kappa)$ 为贝塔-二项方差（3.2）， $h_i = \mu_i/(1-p_{0,\,i})$ 为强度函数（3.7），以及$m_i^{+} = n_i h_i = n_i\mu_i/(1-p_{0,\,i})$ 为给定参与时的条件均值。

全方差定律。

命题 A.3 （障碍模型方差分解）。 设 Y_i 服从HBB模型（参见 *Mullahy*, *1986*, 关于泊松障碍模型类比），其参与概率为 q_i ，强度参数为$(n_i,\ \mu_i,\ \kappa)$。定义 $Z_i = \mathbf{1}\{Y_i>0\}$。则

$$\mathrm{Var}(Y_i) = q_i\cdot\mathrm{Var}(Y_i \mid Y_i>0) + q_i(1-q_i)\big[\mathbb{E}(Y_i \mid Y_i>0)\big]^2. \tag{A.12}$$

证明。 根据伯努利变量 Z_i 应用全方差定律：

$$\mathrm{Var}(Y_i) = \mathbb{E}\big[\mathrm{Var}(Y_i \mid Z_i)\big] + \mathrm{Var}\big(\mathbb{E}[Y_i \mid Z_i]\big).$$

第一项。当 $Z_i = 0$、$Y_i = 0$ 几乎必然时，因此 $\mathrm{Var}(Y_i \mid Z_i = 0) = 0$。当 $Z_i = 1$时，$\mathrm{Var}(Y_i \mid Z_i = 1) = \mathrm{Var}(Y_i \mid Y_i > 0)$。因此

$$\mathbb{E}\big[\mathrm{Var}(Y_i \mid Z_i)\big] = q_i \ \mathrm{Var}(Y_i \mid Y_i > 0).$$

第二项。条件期望 $\mathbb{E}[Y_i \mid Z_i]$ 以概率$1 - q_i$ 取值 0，以概率 q_i取值 $m_i^+ = n_i\mu_i/(1 - p_{0,\,i})$。因此，其方差为 $q_i(1 - q_i)(m_i^+)^2$。□

推论A.2 (展开形式**)**。将 命题*A.2* 中的截断方差代入 (A.12)：

$$\mathrm{Var}(Y_i) = \frac{q_i\, V_{\mathrm{BB},i}}{1 - p_{0,i}} - \frac{q_i\, n_i^2\mu_i^2\, p_{0,i}}{(1 - p_{0,i})^2} + \frac{q_i(1 - q_i)\, n_i^2\mu_i^2}{(1 - p_{0,i})^2}. \tag{A.13}$$

合并涉及 $n_i^2\mu_i^2/(1 - p_{0,\,i})^2$ 的各项，得到紧凑形式

$$\mathrm{Var}(Y_i) = \frac{q_i\, V_{\mathrm{BB},i}}{1 - p_{0,i}} + \frac{q_i\, n_i^2\mu_i^2\,(1 - q_i - p_{0,i})}{(1 - p_{0,i})^2}. \tag{A.14}$$

注记A.2。量 q_i 和 $p_{0,\,i}$ 彼此不同：q_i 是障碍参与概率，而 $p_{0,\,i}$ 是贝塔-二项分量的零概率。一般而言，$q_i \neq 1 - p_{0,\,i}$；量 $(1 - q_i - p_{0,\,i})$ 可能为正、负或零，因此（A.14）中的第二项可能使无条件方差增大或减小。

推论A.3（以无条件均值表示）。回顾（3.7）中的结论 $\mathbb{E}[Y_i] = q_i\, n_i\, h_i$。则

$$\mathrm{Var}(Y_i) = q_i \ \mathrm{Var}(Y_i \mid Y_i > 0) + \frac{1 - q_i}{q_i}\,\big(\mathbb{E}[Y_i]\big)^2. \tag{A.15}$$

分量间项与均值平方成正比，并由不参与优势进行缩放（$1 - q_i$）$/q_i$。

方差均值比。

命题A.4（方差比）。定义方差比$_i$ =Var（Y_i）$/\mathbb{E}[Y_i]$。则

$$\boxed{\mathrm{VR}_i = \frac{(1 - \mu_i)(n_i + \kappa)}{1 + \kappa} - \frac{n_i\,\mu_i\,p_{0,i}}{1 - p_{0,i}} + (1 - q_i)\, n_i\, h_i.} \tag{A.16}$$

证明。利用 $\mathbb{E}[Y_i] = q_i\, n_i\, h_i$ 和 推论A.3：

$$\mathrm{VR}_i = \frac{\mathrm{Var}(Y_i \mid Y_i > 0)}{n_i\, h_i} + (1 - q_i)\, n_i\, h_i.$$

对于第一项，代入截断方差（命题 A.2），并利用 $h_i = \mu_i/(1 - p_{0,\ i})$ 化简：

$$\frac{V_{\mathrm{BB},i}}{(1-p_{0,i})\,n_i\,h_i} = \frac{(1-\mu_i)(n_i+\kappa)}{1+\kappa}, \qquad \frac{n_i\,\mu_i^2\,p_{0,i}}{(1-p_{0,i})^2\,h_i} = \frac{n_i\,\mu_i\,p_{0,i}}{1-p_{0,i}}. \qquad \square$$

注释 A.3. (A.16) 中的三项有明确的解释：

(i) $(1-\mu_i)(n_i+\kappa)/(1+\kappa)$：贝塔-二项方差膨胀因子（Prentice，1986），对所有 $n_i > 1$ 都大于 1；(ii) $-n_i\,\mu_i\,p_{0,\ i}/(1 - p_{0,\ i})$：零截断带来的校正，它会降低方差比，反映了在 $Y = 0$ 处被移除的概率质量；(iii) $(1 - q_i)\,n_i\,h_i$：门槛机制产生的组分间贡献，只要 $q_i < 1$，就会增大方差比。

对于泊松数据，VR $= 1$（等离散）。在 NSECE 数据中，$\mathrm{VR}_i > 1$ 始终如此，反映了过度离散和门槛机制的共同作用。

相对于二项分布的过度离散指数。

定义 A.1（过度离散指数）。对于比例 Y_i/n_i，定义

$$\mathrm{OD}_i = \frac{\mathrm{Var}(Y_i)}{n_i\,\pi_i(1-\pi_i)}, \qquad \text{where } \pi_i = q_i\,h_i = \frac{\mathbb{E}[Y_i]}{n_i}. \tag{A.17}$$

分母 $n_i\,\pi_i(1-\pi_i)$ 是一个具有相同无条件均值比例的 $\mathrm{Bin}(n_i,\ \pi_i)$ 随机变量的方差，因此 OD_i 衡量的是相对于二项分布基准的过度离散程度。

极限情形。

- 无门槛，$p_0 \to 0$。于是 $q_i = 1$，$h_i \to \mu_i$，且 $\mathrm{Var}(Y_i) \to V_{\text{贝塔-二项分布 } i}$，因此 $\mathrm{OD}_i \to (n_i+\kappa)/(1+\kappa)$，即标准贝塔-二项过度离散因子（Prentice，1986）。
- 二项极限（$\kappa \to \infty$，$q_i = 1$）。$V_{\text{贝塔-二项分布 } i} \to n_i\mu_i(1-\mu_i)$，且门槛不起作用，因此 $\mathrm{OD}_i \to 1$。

变异系数。

命题 A.5（条件变异系数）。给定 Y_i 参与时，其变异系数的平方为

$$\boxed{\mathrm{CV}^2(Y_i \mid Y_i > 0) = \frac{(1-\mu_i)(n_i+\kappa)}{n_i\,\mu_i(1+\kappa)}\,(1-p_{0,i}) - p_{0,i}.} \tag{A.18}$$

证明。根据定义，$\mathrm{CV}^2 = \mathrm{Var}(Y_i \mid Y_i > 0)/[\mathbb{E}(Y_i \mid Y_i > 0)]^2$。代入截断方差（命题 A.2）和条件均值 $m_i^+ = n_i\mu_i/(1-p_{0,i})$：

$$\begin{aligned}\mathrm{CV}^2 &= \frac{V_{\mathrm{BB},i}(1-p_{0,i}) - n_i^2\mu_i^2\,p_{0,i}}{n_i^2\mu_i^2} = \frac{V_{\mathrm{BB},i}}{n_i^2\mu_i^2}\,(1-p_{0,i}) - p_{0,i} \\ &= \frac{(1-\mu_i)(n_i+\kappa)}{n_i\,\mu_i(1+\kappa)}\,(1-p_{0,i}) - p_{0,i}.\end{aligned}$$

□

命题 A.6（无条件变异系数）。 Y_i（无条件）的变异系数平方满足

$$\boxed{\mathrm{CV}^2(Y_i) = \frac{\mathrm{CV}^2(Y_i \mid Y_i > 0) + 1 - q_i}{q_i}.} \tag{A.19}$$

证明。 由 $\mathrm{Var}(Y_i) = q_i\,\mathrm{Var}(Y_i \mid Y_i > 0) + q_i(1-q_i)(n_ih_i)^2$ (命题 A.3) 和$[\mathbb{E}(Y_i)]^2 = q_i^2(n_ih_i)^2$:

$$\mathrm{CV}^2(Y_i) = \frac{\mathrm{CV}^2(Y_i \mid Y_i > 0)}{q_i} + \frac{1-q_i}{q_i} = \frac{\mathrm{CV}^2(Y_i \mid Y_i > 0) + 1 - q_i}{q_i}.$$

□

备注 A.4。 变异系数对于后验预测检验很有用，因为：**(i)** 它不受尺度影响，因此可以比较具有不同 n_i 的服务提供者；**(ii)** $(1-q_i)/q_i$ 项（见 A.19）揭示了参与率较低时门槛如何放大变异系数；**(iii)** 可以将观测到的样本变异系数与理论变异系数的后验分布进行比较，从而得到一种诊断方法，同时检验模型的均值、方差和零膨胀结构。

A.4. **强度函数的解析性质**。 本小节汇集了强度函数 $h(\mu) = \mu/(1-p_0)$ 的解析结果，这些结果支撑弹性公式（命题 1）、备注 备注 2中讨论的边界行为，以及边际效应的对数尺度分解。我们继续将 $b = (1-\mu)\kappa$ 记作Λ， Λ_2，其定义见（A.1）。

p_0**的二阶导数**。

引理 A.4（p_0的二阶导数）。 对所有 $n \geq 1$， $\mu \in (0,\ 1)$，且 $\kappa > 0$，

$$\frac{\partial^2 p_0}{\partial \mu^2} = p_0\big(\Lambda^2 - \Lambda_2\big). \tag{A.20}$$

证明。 从 $\partial p_0/\partial\mu = -p_0\,\Lambda$ 开始，应用乘积法则求导：

$$\frac{\partial^2 p_0}{\partial \mu^2} = -\frac{\partial p_0}{\partial \mu}\cdot\Lambda - p_0\cdot\frac{\partial \Lambda}{\partial \mu} = p_0\,\Lambda^2 - p_0\,\Lambda_2 = p_0\big(\Lambda^2 - \Lambda_2\big).$$

这里我们使用了 $\partial\Lambda/\partial\mu = \Lambda_2$（因为 $\partial b/\partial\mu = -\kappa$）。 □

推论 A.4 （ p_0 在 μ中的凸性）。 对于 $n \geq 2$: $\partial^2 p_0/\partial\mu^2 > 0$, 因此 p_0 关于 μ 严格凸。对于 $n = 1$: $p_0 = 1 - \mu$ 关于 μ是线性的。

证明。 由 引理 A.4 和 引理 A.1可直接得出: 当 $n \geq 2$时, $\Lambda^2 > \Lambda_2$ 且 $p_0 > 0$, 因此二阶导数严格为正。当 $n = 1$时, $\Lambda^2 = \Lambda_2$, 从而有 $\partial^2 p_0/\partial\mu^2 = 0$。 □

h的二阶导数。

命题 A.7 （ h的二阶导数）。 设 $D = 1 - p_0$。则

$$\frac{\partial^2 h}{\partial \mu^2} = \frac{p_0}{D^3}\Big[-2\Lambda D + \mu\Lambda^2(1 + p_0) - \mu\Lambda_2 D\Big]. \tag{A.21}$$

证明。 由 $\partial h/\partial\mu = N/D^2$ 其中 $N = \Phi(\mu) = D - \mu\, p_0\, \Lambda$ （见（A.2））, 应用商法则:

$$\frac{\partial^2 h}{\partial \mu^2} = \frac{N'D - 2ND'}{D^3},$$

其中 $D' = -\partial p_0/\partial\mu = p_0\, \Lambda > 0$。

计算 N'。 由于 $N = (1 - p_0) - \mu\, p_0\, \Lambda$, 我们有

$$N' = p_0\, \Lambda - \big[p_0\, \Lambda + \mu \cdot \tfrac{d}{d\mu}(p_0\, \Lambda)\big] = -\mu \cdot \frac{d}{d\mu}(p_0\, \Lambda) = \mu\, p_0\big(\Lambda^2 - \Lambda_2\big),$$

使用 $d(p_0\Lambda)/d\mu = p_0(\Lambda_2 - \Lambda^2)$ 由 (A.4)。

综合。

$$\begin{aligned} N'D - 2ND' &= \mu\, p_0(\Lambda^2 - \Lambda_2)D - 2\big(D - \mu\, p_0\, \Lambda\big)(p_0\, \Lambda) \\ &= \mu\, p_0(\Lambda^2 - \Lambda_2)D - 2D\, p_0\, \Lambda + 2\mu\, p_0^2\Lambda^2 \\ &= p_0\big[-2\Lambda D + \mu(\Lambda^2 - \Lambda_2)D + 2\mu\, p_0\, \Lambda^2\big]. \end{aligned}$$

注意到 $(\Lambda^2 - \Lambda_2)D + 2p_0\Lambda^2 = \Lambda^2(D + 2p_0) - \Lambda_2 D = \Lambda^2(1 + p_0) - \Lambda_2 D$ 即得所述形式。□

验证 ($n = 1$)。当 $n = 1, p_0 = 1-\mu, D = \mu, \Lambda = 1/(1-\mu)$, $\Lambda_2 = \Lambda^2$, 且$h \equiv 1$, 故 $\partial^2 h/\partial\mu^2 = 0$。在公式中: $N = D - \mu\, p_0\, \Lambda = \mu - \mu(1 - \mu)/(1 - \mu) = 0$恒成立, 故 $N' = 0$ 且分子为零。 □

p_0 关于所有参数的单调性。

命题 A.8 （ p_0 关于 μ的单调性）。 $\partial p_0/\partial\mu = -p_0\, \Lambda < 0$。增大 μ会严格降低零概率。

证明。 已在第 A.1 节中确立（定理 1的证明之前的辅助计算）。 □

命题 A.9 （ κ中 p_0 的单调性）。 对于 $n \geq 2$ 和 $\mu \in (0,\ 1)$:

$$\frac{\partial p_0}{\partial \kappa} = -p_0\,\mu \sum_{j=1}^{n-1} \frac{j}{(b+j)(\kappa+j)} < 0. \tag{A.22}$$

对于 $n = 1$: $\partial p_0/\partial\kappa = 0$ （因为 $p_0 = 1-\mu$ 与 κ相互独立）。

证明。 对 log $p_0 = \sum_{j=0}^{n-1}[\log(b+j) - \log(\kappa+j)]$ 关于 $\partial b/\partial\kappa = 1-\mu$求导:

$$\frac{\partial \log p_0}{\partial \kappa} = \sum_{j=0}^{n-1}\left[\frac{1-\mu}{b+j} - \frac{1}{\kappa+j}\right].$$

对于每一项，分子为 $(1-\mu)(\kappa+j)-(b+j) = (1-\mu)(\kappa+j)-(1-\mu)\kappa-j = -\mu j$。因此

$$\frac{\partial \log p_0}{\partial \kappa} = -\mu \sum_{j=1}^{n-1} \frac{j}{(b+j)(\kappa+j)} < 0,$$

其中 $j = 0$ 项消失。两边乘以 $p_0 > 0$，即可得到所述公式。 □

增大 κ （降低过度离散）会使概率质量集中在 $n\mu > 0$附近，从而降低零值的概率。

命题 A.10 （ n中 p_0 的单调性）。 对所有 $n \geq 1$， $\mu \in (0,\ 1)$，且 $\kappa > 0$:

$$\frac{p_0(n+1)}{p_0(n)} = \frac{b+n}{\kappa+n} < 1. \tag{A.23}$$

因此， p_0 关于 n严格递减。

证明。由乘积形式 (3.3)， $p_0(n+1) = p_0(n)\cdot(b+n)/(\kappa+n)$。由于对 $\mu \in (0,\ 1)$， $b = (1-\mu)\kappa < \kappa$， 比率 $(b+n)/(\kappa+n) < 1$。 □

注 A.5（ n中 p_0 的多项式衰减）。当 $n \to \infty$时，相邻比 $(b+n)/(\kappa+n) \to 1$，因此 p_0 关于 n呈多项式衰减，而非指数衰减。具体而言，根据上升阶乘比率的斯特林近似， $p_0(n) \sim C(\mu,\kappa)\,n^{-\mu\kappa}$， 当 $n \to \infty$时。相比之下，二项零概率 $(1-\mu)^n$ 呈指数衰减。这种较慢的衰减意味着，即使组规模适中且较大（$n = 50$–100）， 当 κ 较小时，结构性零概率 p_0 仍不可忽视，这进一步说明门槛分量的重要性（另见 Bandyopadhyay等人，2011年，其中空间贝塔-二项的应用体现了这一现象的实际意义）。

对数导数与 h的严格单调性。

命题 A.11 （ h的对数导数）。

$$\frac{\partial(\log h)}{\partial \mu} = \frac{1}{\mu} - \frac{p_0 \Lambda}{1 - p_0}. \tag{A.24}$$

证明。 由于 $\log h = \log \mu - \log(1 - p_0)$，我们对每一项求导： $d(\log \mu)/d\mu = 1/\mu$，以及 $d\log(1 - p_0)/d\mu = (-dp_0/d\mu)/(1 - p_0) = p_0\Lambda/(1 - p_0)$。相减即得结果。□

定理 A.1 （$\log h$ 严格递增）。 对于 $n \geq 2$:

$$\frac{\partial(\log h)}{\partial \mu} > 0 \qquad \text{for all } \mu \in (0, 1). \tag{A.25}$$

证明。 条件 $\partial(\log h)/\partial\mu > 0$ 等价于 $1/\mu > p_0\Lambda/(1 - p_0)$，整理后得 $(1 - p_0) > \mu\, p_0\, \Lambda$，即 $\Phi(\mu) > 0$ （见（A.2））。这已在
定理 1。 □

注释 A.6 （对数导数的分解）。 对数导数可分解为

$$\frac{\partial(\log h)}{\partial \mu} = \underbrace{\frac{1}{\mu}}_{\text{direct effect}} - \underbrace{\frac{p_0 \Lambda}{1 - p_0}}_{\text{truncation attenuation}}.$$

直接效应表示分子 μ 中的比例增加。截断衰减表示分母 $1 - p_0$ 中的比例增加。两项均为正，但直接效应始终占主导，因为 $\Phi(\mu) > 0$（定理 1），而这又源于 p_0 的严格凸性（推论 A.4）。

弹性。

定义 A.2 （ h 对 μ的弹性）。

$$\varepsilon_h(\mu) = \frac{\partial h}{\partial \mu} \cdot \frac{\mu}{h} = \mu \cdot \frac{\partial(\log h)}{\partial \mu}. \tag{A.26}$$

命题 A.12 （弹性公式与界）。

$$\boxed{\varepsilon_h(\mu) = 1 - \omega(\mu), \qquad \omega(\mu) = \frac{\mu\, p_0\, \Lambda}{1 - p_0}.} \tag{A.27}$$

此外:

(i) 对于 $n = 1$: $\varepsilon_h = 0$ （因为 $h \equiv 1$）。(ii) 对于 $n \geq 2$: $\varepsilon_h \in (0,1)$ 对所有 $\mu \in (0, 1)$。强度函数 h 对 μ缺乏弹性。

这是命题 *1* 在正文中的完整证明。

证明。 由命题 A.11,

$$\varepsilon_h = \mu\left[\frac{1}{\mu} - \frac{p_0\,\Lambda}{1-p_0}\right] = 1 - \frac{\mu\,p_0\,\Lambda}{1-p_0} = 1-\omega.$$

根据定理 A.1， $\Phi(\mu) > 0$ 对于 $n \geq 2$，这给出 $\omega < 1$ ，因此 $\varepsilon_h > 0$。由于 $\omega > 0$ （所有因子均为正）， $\varepsilon_h < 1$。□

注 A.7 （非弹性解释）。 μ 增加百分之一，会使 h增加不到百分之一，因为增加 μ 的同时也会增加 $1-p_0$（ h的分母），从而部分抵消分子效应。

命题 A.13 （ ε_h的极限行为）。

(a) 当 $\mu \to 0^+$ 趋于极限时（对于 $n \geq 2$）： $\varepsilon_h \to 0^+$。截断衰减为 $\omega \to 1^-$。
(b) 当 $\mu \to 1^-$趋于极限时： $\varepsilon_h \to 1 - C_\kappa$，其中 $C_\kappa = \prod_{j=1}^{n-1} j/(\kappa + j)$是命题 *A.15* 中的常数。由于 $p_0 \sim (1-\mu)\,C_\kappa$ 且 $\Lambda \sim 1/(1-\mu) + O(1)$，因此得到 $p_0\,\Lambda \to C_\kappa$ ，进而有 $\omega \to C_\kappa$。当 n 或 κ较大时， $C_\kappa \approx 0$ ，且极限趋近于 *1*。
(c) 当 $\kappa \to \infty$ 趋于极限时（二项极限）： $\varepsilon_h \to 1 - n\mu(1-\mu)^{n-1}\big/\big[1-(1-\mu)^n\big]$。
(d) 当 $\kappa \to 0$ 趋于极限时（最大过度离散）： $p_0 \to 1-\mu$， $\Lambda \to 1/(1-\mu)$，因此 $\omega \to 1$且 $\varepsilon_h \to 0$。在此极限下， $h \to 1$；强度函数为平坦的。*(e)* 当 $n \to \infty$ 趋于极限时： $p_0 \to 0$，因此 $\omega \to 0$ 且 $\varepsilon_h \to 1$。截断校正变得可以忽略不计。

数值验证。 对于 $n = 10$、 $\kappa = 10$、 $\mu = 0.3$：解析公式给出 $\varepsilon_h = 0.734979$，中心有限差分得到 0.734979。两者精确到小数点后六位一致。 □

边界分析。

命题 A.14 （ h 当 $\mu \to 0^+$时的行为）。对于 $n \geq 2$ 和 $\kappa > 0$:

$$h(\mu) \to \frac{1}{\Lambda_0} \quad \text{as } \mu \to 0^+, \tag{A.28}$$

其中 $\Lambda_0 = \kappa \sum_{j=0}^{n-1} 1/(\kappa + j)$ 。

证明。 泰勒展开： $p_0(\mu) = 1 - \mu\Lambda_0 + O(\mu^2)$（由 $p_0(0) = 1$和 $p_0'(0) = -\Lambda_0$得出）。于是 $1 - p_0 = \mu\Lambda_0 + O(\mu^2)$，且 $h = \mu/\big(\mu\Lambda_0 + O(\mu^2)\big) \to 1/\Lambda_0$。□

例如，当 $n = 2$ 且 $\kappa = 1$： $\Lambda_0 = 1/1 + 1/2 = 3/2$，因此 $h(0^+) = 2/3$。对于$n = 10$ 且 $\kappa = 1$： $\Lambda_0 = H_{10} \approx 2.929$，因此 $h(0^+) \approx 0.341$。正如备注 2所指出的，只要 $n \geq 2$，该极限就严格小于 1，这证实了即使在 $\mu \to 0^+$时，截断调整仍有不可忽略的影响。

命题 A.15 (h 在 $\mu \to 1^-$时的行为)。对于 $n \geq 1$ 和 $\kappa > 0$:

$$h(\mu) \to 1 \quad \text{as } \mu \to 1^-, \qquad 1 - h(\mu) \sim (1-\mu)(1-C_\kappa), \tag{A.29}$$

其中 $C_\kappa = \prod_{j=1}^{n-1} j/(\kappa + j) \in (0,1)$。

证明。由于 $\mu \to 1$, $b = (1-\mu)\kappa \to 0$, 且乘积形式（3.3）给出 $p_0 \sim (1-\mu)\, C_\kappa$, 其中 $C_\kappa = \prod_{j=1}^{n-1} j/(\kappa + j)$。于是

$$h = \frac{\mu}{1-p_0} \sim \frac{1-(1-\mu)}{1-(1-\mu)C_\kappa},$$

and

$$1 - h \sim \frac{(1-\mu)(1-C_\kappa)}{1-(1-\mu)C_\kappa} \sim (1-\mu)(1-C_\kappa)$$

因为 $\mu \to 1^-$。 □

κ **中的弹性单调性。**

猜想 A.1 （ ε_h 关于 κ的单调性）。 对于固定的 $\mu \in (0,1)$ 和 $n \geq 2$, 弹性 ε_h 随着 κ 增大而增大。等价地，衰减比 ω 随着 κ 增大而减小。

以下证明概要提供了支持性证据，但尚未构成严格证明。

证明概要。随着 κ 增加， p_0 减少（命题 A.9），而对 $\omega = \mu\, p_0\, \Lambda/$（$1 - p_0$）起主导作用的是分子因子 p_0的减少。边界行为证实了单调性：当 $\kappa \to 0$ 时， $\varepsilon_h \to 0$（命题 A.13(d)）；当 $\kappa \to \infty$ 时， $\varepsilon_h \to 1 - n\mu$（$1-\mu$）$^{n-1}/[1-$（$1-\mu$）$^{n}] > 0$（命题 A.13(c)）。闭式证明需要证明，在(0, 1)上， p_0 的减少始终大于 Λ同时发生的减少。

注释 A.8 （猜想的数值验证）。 我们在每个 $n \in \{2, 5, 10, 20, 50, 100\}$上，对（$\mu$, κ）取 $\in$ (0。01，0。99)，对 $\times$(0。01，100)中的 100×100 个值构成的网格，使用中心有限差分计算了 $\partial\varepsilon_h/\partial\kappa$ 。所有情况下，导数都严格为正，支持了这一猜想。目前尚未发现反例。

A.5. **对数尺度边际效应分解。** 命题 5已证明，协变量 x_k 对对数期望入学的边际效应可精确分解为 $\mathrm{LAE}_k + \mathrm{LIE}_k$，其乘数分别为 $(1-q_i)$ 和 $(1-\mu_i)\varepsilon_{h,\,i}$。本小节给出三项补充结果：关于 LIE 符号结构的修正说明、分解比相对于 κ 的敏感性分析，以及贫困逆转区间的阈值刻画。

注 A.9 （LIE 中截断衰减的符号）。 正确的 LIE 在截断衰减项前应为负 号：

$$\mathrm{LIE}_k = \left[\frac{1}{\mu_i} - \frac{p_{0,i}\,\Lambda_i}{1-p_{0,i}}\right]\mu_i(1-\mu_i)\,\tilde{\beta}_{k,s} = (1-\mu_i)(1-\omega_i)\,\tilde{\beta}_{k,s}, \tag{A.30}$$

其中 $\omega_i = \mu_i\, p_{0,\ i}\,\Lambda_i/(1-p_{0,\ i})$ 是命题 A.12 中的命题 A.12所给出的衰减比。衰减效应是相减而非相加，因为增大 μ 会同时增大$1-p_0$ (即 h的分母)，从而削弱而非增强 h对 μ的响应。

P命题 A.16 (分解比)。 定义分解比

$$\mathrm{DR}(\kappa) = \frac{|\,\mathrm{LAE}_k\,|}{|\,\mathrm{LAE}_k\,| + |\,\mathrm{LIE}_k\,|} = \frac{(1-q)\,|\tilde{\alpha}_k|}{(1-q)\,|\tilde{\alpha}_k| + (1-\mu)\,\varepsilon_h\,|\tilde{\beta}_k|}. \tag{A.31}$$

于是：

(a) 当 $\kappa \to 0$ 趋于 $\varepsilon_h \to 0$ 时，*LIE* $\to 0$，*DR* $\to 1$。最大过度离散会使强度渠道失效。*(b)* 当 $\kappa \to \infty$ 趋于 $\varepsilon_h \to \varepsilon_h^{\mathrm{binom}} \in$ 时，*(0, 1)* 中的 *DR* $\to DR^{\mathrm{binom}} \in$ 。*(c)* 条件于猜想 *A.1*，*DR* 随 κ 增大而减小（因为 ε_h 随 κ 增大而增大）。

证明。 性质 (a) 和 (b) 分别直接由命题 A.13(d) 和 (c) 得出。性质 (c) 成立是因为式 (A.31) 的分子不随 κ变化，而分母则通过因子 κ随 ε_h增大。□

表 A.1 展示了分解比对 κ的敏感性，评估时采用˜代表性参数值（$n = 10$、$\mu = 0.30$、 $q = 0.70$、 $|\tilde{\alpha}| = 0.5$、 $|\beta| = 0.8$）。

表 A.1. 分解比对 κ的敏感性。参数：˜$n = 10$、 $\mu = 0.30$、 $q = 0.70$、$|\tilde{\alpha}| = 0.5$、 $|\beta| = 0.8$。

κ	ε_h	ω	|LIE |	DR
0.01	0.008	0.992	0.005	0.970
0.1	0.071	0.929	0.040	0.790
1	0.343	0.657	0.192	0.439
10	0.735	0.265	0.412	0.267
100	0.859	0.141	0.481	0.238
1000	0.874	0.126	0.489	0.235

备注 A.10。 DR 对 κ 高度敏感，其取值范围为 (0.1，10)。在我们的数据情境中，$\kappa \approx 5$–15，可及性和强度作用渠道均对总效应作出了实质性贡献。

命题 A.17 (可及性效应主导的阈值)。 在贫困反转˜区域 ($\tilde{\alpha}_{\text{pov},\ s} < 0$, $\beta_{\text{pov},\ s} > 0$) 中，当且仅当以下条件成立时，对数期望入学的总边际效应为负（可及性效应占主导）：

$$\boxed{\frac{|\tilde{\alpha}_{\text{pov},s}|}{|\tilde{\beta}_{\text{pov},s}|} > \frac{(1-\mu)\,\varepsilon_h}{1-q}.} \tag{A.32}$$

右侧随 q 增大而增大，并且随 κ 增大而增大（通过 ε_h）。当参与率较高（q 接近 1）时，即使集约边际系数很小，也可能占主导；当 q 较低时，广延边际更容易占主导。

˜证明。 根据 命题 5，总效应为负当且仅当 $(1-q)|\tilde{\alpha}| > (1-\mu)\varepsilon_h|\beta|$；˜两边除以 $|\beta|(1-q)$ 得（A.32）。□

表 A.2 报告了 $(1-\mu)\varepsilon_h/(1-q)$ 这一阈值在涵盖 NSECE 服务提供者范围的参数组合下的取值。弹性 ε_h 根据 命题 A.12中的精确公式计算。

注记 A.11。在我们的后验估计（$\kappa \approx 7$、 $\mu \approx 0.30$、 $q \approx 0.64$、 $n \approx 50$）下，阈值约为 1.87（表 A.2，第 $q = 0.64$、 $\mu = 0.30$、 $\kappa = 7$ 行）。由于经验系数比 $|\alpha_{\text{pov}}|/|\beta_{\text{pov}}| = 0.324/0.090 \approx 3.6 \gg 1.87$，可及性效应具有决定性的主导地位。对于接近样本均值的服务提供者，系数比需要降至 1.87 以下——也就是说，广延边际系数需要小于集约边际系数的两倍——强度才可能占主导。在 $q = 0.80$ （所考虑的最极端参与率）处，阈值升至 2.87–3.77，接近经验比率。这证实，可及性主导在低 μ 且 κ较大的高参与中心中最为脆弱，此时强度弹性接近 1，而广延乘数 $(1-q)$ 较小。

附录B. 可识别性与费舍尔信息

本附录给出跨栏贝塔-二项模型的得分函数、费舍尔信息结构及可识别性证明，这些内容在 第3节中暂未展开。我们沿用 第 3.1 节中的记号，其中 $a = \mu\kappa$， $b = (1-\mu)\kappa$，双伽玛函数 $\psi(x) = \Gamma'(x)/\Gamma(x)$。

表 A.2. 代表性参数组合的访问支配阈值 $(1-\mu)\,\varepsilon_h/(1-q)$。低于经验系数比 $|\alpha_{\mathrm{pov}}|/|\beta_{\mathrm{pov}}| = 3.6$ 的数值表示访问支配。所有条目均使用 $n = 50$ 计算。

		κ		
q	μ	3	7	15
0.50	0.15	0.90	1.26	1.51
	0.30	1.15	1.34	1.39
	0.45	1.04	1.10	1.10
0.64	0.15	1.25	1.75	2.09
	0.30	1.60	1.87	1.93
	0.45	1.44	1.52	1.53
0.80	0.15	2.25	3.16	3.77
	0.30	2.87	3.36	3.48
	0.45	2.60	2.74	2.75

注。阈值由精确弹性公式（A.27）计算，其中$p_0 = \prod_{j=0}^{n-1}[(1-\mu)\,\kappa + j]/(\kappa + j)$ $\kappa + j]$/以及$\Lambda = \kappa \sum_{j=0}^{n-1}[(1-\mu)\,\kappa + j]^{-1}$。数值四舍五入至小数点后两位。

B.1. **得分函数**。我们推导了障碍型贝塔-二项对数似然各组成部分的得分函数。这些表达式有三种用途：(i) 验证 Stan 中使用的自动微分梯度（Carpenter 等人，2017年）；(ii) 通过得分幅度监测进行诊断；以及 (iii) 为 1983年的夹心方差估计量（Binder）提供理论依据，详见 C节。回顾参数化$a = \mu\kappa$，$b = (1-\mu)\kappa$、零概率 p_0（见3.3），以及辅助量Λ，Λ_2，其定义见 公式 (A.1)。

命题 B.18（广延边际得分）。关于线性预测量 $\eta_i^{(1)} = \mathbf{x}_i^{\intercal}\boldsymbol{\alpha} + \mathbf{x}_i^{(r)\intercal}\boldsymbol{\delta}_{1,\,s}$，广延边际对数似然 $\ell_i^{(1)} = z_i \log q_i + (1 - z_i)\log(1 - q_i)$ 的得分为

$$\frac{\partial \ell_i^{(1)}}{\partial \eta_i^{(1)}} = z_i - q_i, \tag{B.1}$$

其中 q_i =逆逻辑函数($\eta_i^{(1)}$)。海森矩阵是非随机的：

$$\frac{\partial^2 \ell_i^{(1)}}{\partial (\eta_i^{(1)})^2} = -\,q_i(1 - q_i). \tag{B.2}$$

证明。 标准逻辑回归微积分。由于 $\partial q_i/\partial \eta_i^{(1)} = q_i(1-q_i)$，对 $\ell_i^{(1)}$ 求导得到 $z_i/q_i \cdot q_i(1-q_i) - (1-z_i)/(1-q_i)\cdot q_i(1-q_i) = z_i(1-q_i) - (1-z_i)q_i = z_i - q_i$。二阶导数为 $-q_i(1-q_i)$，它仅取决于参数，因此是非随机的。 □

注 B.12. (B.2) 的非随机性意味着广延边际的观测费舍尔信息与期望费舍尔信息相同：$\mathcal{I}_{\text{ext},i} = q_i(1-q_i)$。这简化了夹心估计量，因为第 1 部分的海森面包矩阵不含随机项。

命题 B.19 (集约边际得分，针对 μ)。 定义（未截断的）贝塔-二项分布得分函数（遵循 *Prentice* 的参数化方式，*1986*）

$$S_{\text{BB}}(y;\,\mu,\kappa,n) = \kappa\big[\psi(y+a) - \psi(n-y+b) - \psi(a) + \psi(b)\big], \tag{B.3}$$

其中 $\psi(\cdot)$ 是双伽玛函数，且 $a = \mu\kappa$, $b = (1-\mu)\kappa$。零截断对数似然 $\ell_i^{(2+)} = \log f_{\text{BB}}(y_i) - \log(1-p_{0,i})$ 关于 μ 的得分为

$$\boxed{\frac{\partial \ell_i^{(2+)}}{\partial \mu} = S_{\text{BB}}(y_i) - \frac{p_{0,i}\,\Lambda_i}{1-p_{0,i}}}. \tag{B.4}$$

证明。 分解式 $\ell_i^{(2+)} = \log f_{\text{BB}}(y_i) - \log(1-p_{0,i})$ 给出两项。

第 *1* 项（二项分布得分）。 利用 $\partial a/\partial\mu = \kappa$ 和 $\partial b/\partial\mu = -\kappa$，对贝塔-二项分布对数概率质量函数求导可得

$$\frac{\partial \log f_{\text{BB}}}{\partial \mu} = \kappa\big[\psi(y+a) - \psi(n-y+b) - \psi(a) + \psi(b)\big] = S_{\text{BB}}(y).$$

第 *2* 项（截断校正）。由于 $\ln p_0 = \sum_{j=0}^{n-1} \ln[(1-\mu)\kappa + j] - \ln[\kappa + j]$，我们有 $\partial p_0/\partial\mu = -p_0\Lambda$ （见公式（A.1）），因此

$$\frac{\partial \log(1-p_0)}{\partial \mu} = \frac{p_0\,\Lambda}{1-p_0} > 0.$$

合并，$\partial \ell^{(2+)}/\partial\mu = S_{\text{BB}}(y) - p_0\Lambda/(1-p_0)$。 □

命题 B.20（关于 $\log\kappa$ 的得分）。零截断对数似然关于对数离散参数的得分为

$$\boxed{\begin{aligned}\frac{\partial \ell_i^{(2+)}}{\partial \log\kappa} = \kappa\Big[&\mu\big[\psi(y_i+a)-\psi(a)\big]+(1-\mu)\big[\psi(n_i-y_i+b)-\psi(b)\big]\\ &+\psi(\kappa)-\psi(n_i+\kappa)\Big]-\frac{p_{0,i}\,\mu\kappa}{1-p_{0,i}}\sum_{j=1}^{n_i-1}\frac{j}{(b+j)(\kappa+j)}.\end{aligned}} \tag{B.5}$$

证明。根据 $\phi=$对数 κ和链式法则 $\partial/\partial\phi=\kappa\,\partial/\partial\kappa$，未截断的 BB 对数概率质量函数通过 $\partial a/\partial\kappa=\mu$ 和 $\partial b/\partial\kappa=1-\mu$得到第一行。截断校正需要 ∂对数$(1-p_0)/\partial$对数 κ。令 $\ln p_0=\sum_{j=0}^{n-1}\ln(b+j)-\ln(\kappa+j)$，并对 κ 求导（注意 $\partial b/\partial\kappa=1-\mu$），重新整理望远镜求和后得到第二行。□

命题 B.21（关于 μ的期望信息量）。相对于 μ，零截断对数似然的负期望二阶导数为

$$\begin{aligned}-\mathbb{E}_{Y|Y>0}\left[\frac{\partial^2\ell^{(2+)}}{\partial\mu^2}\right] = \frac{\kappa^2}{1-p_0}\,\mathbb{E}_{Y\sim\mathrm{BetaBin}}\big[\psi_1(Y+a)+\psi_1(n-Y+b)\big]\\ -\kappa^2\big[\psi_1(a)+\psi_1(b)\big]+T'(\mu),\end{aligned} \tag{B.6}$$

其中 $\psi_1(\cdot)=d\psi/dx$ 为三伽马函数，并且

$$T'(\mu)=\frac{p_0\,\Lambda_2}{1-p_0}-\frac{p_0\,\Lambda^2}{(1-p_0)^2},\qquad \Lambda_2=\kappa^2\sum_{j=0}^{n-1}\frac{1}{(b+j)^2}. \tag{B.7}$$

证明。对得分（B.4）关于 μ求导会产生两项。

该 S_{BB} 项。直接求导得到 $\partial S_{\mathrm{BB}}/\partial\mu=\kappa^2[\psi_1(y+a)+\psi_1(n-y+b)-\psi_1(a)-\psi_1(b)]$。

截断项。令 $T(\mu)=p_0\Lambda/(1-p_0)$。利用 $\partial p_0/\partial\mu=-p_0\Lambda$ 的商法则可得

$$\begin{aligned}T'(\mu)&=\frac{p_0(\Lambda_2-\Lambda^2)(1-p_0)-p_0^2\Lambda^2}{(1-p_0)^2}=\frac{p_0\Lambda_2(1-p_0)-p_0\Lambda^2}{(1-p_0)^2}\\ &=\frac{p_0\,\Lambda_2}{1-p_0}-\frac{p_0\,\Lambda^2}{(1-p_0)^2}.\end{aligned}$$

截断条件下的期望。对 $Y\mid Y>0$求期望，并通过单位矩阵恒等式进行转换 $\mathbb{E}_{Y|Y>0}[g(Y)]=\mathbb{E}_{Y\sim\text{贝塔-二项分布}}[g(Y)]/(1-p_0)$ 得到（B.6）。

非随机项 $-\kappa^2[\psi_1(a) + \psi_1(b)]$ 和 $T'(\mu)$ 在求期望时保持不变。□

注记 B.13（交叉信息）。 μ 与 κ 均进入 $\ell_i^{(2+)}$，因此交叉信息$\mathcal{I}_{\mu,\,\kappa}$ 不为零。集约边际信息块为

$$\mathcal{I}_{\text{int}+\kappa} = \begin{pmatrix} \mathcal{I}_{\text{int}} & \mathcal{I}_{\text{int},\kappa} \\ \mathcal{I}_{\kappa,\text{int}} & \mathcal{I}_{\kappa} \end{pmatrix}, \tag{B.8}$$

其并非进一步分块对角。集约边际回归参数与κ 必须联合估计，但它们借助Hurdle结构与广延边际参数完全分离（第3节）。

B.2. **费舍尔信息与分块对角性**。 个体得分函数与信息条目在第 B.1 节 中导出，涉及单个线性预测量或标量参数。本小节确立全局 费舍尔信息矩阵的结构——特别是广延边际与集约边际参数块之间的分块对角分离——并汇集若干关于信息量速率与夹心约定的辅助结果。

命题 B.22（分块对角费舍尔信息）。 设 $\boldsymbol{\theta} = (\boldsymbol{\phi}_1^\intercal\,,\,\boldsymbol{\phi}_2^\intercal\,)^\intercal$，其中$\boldsymbol{\phi}_1 = (\boldsymbol{\alpha}^\intercal, \boldsymbol{\delta}_{1\,,s}^\intercal)^\intercal$ 汇集广延边际参数，而 $\boldsymbol{\phi}_2 = (\boldsymbol{\beta}^\intercal, \boldsymbol{\delta}_{2\,,s}^\intercal, \log\,\kappa)^\intercal$汇集包含离散参数在内的集约边际参数。则单个观测的期望费舍尔信息分解为

$$\boxed{\mathcal{I}_i(\boldsymbol{\theta}) = \operatorname{diag}\big(\mathcal{I}_{\text{ext},i},\ \mathcal{I}_{\text{int}+\kappa,i}\big).} \tag{B.9}$$

也就是说， $\boldsymbol{\phi}_1$ 与 $\boldsymbol{\phi}_2$ 之间的交叉信息恒为零。

证明。 对数似然可分解为 $\ell_i = \ell_i^{(1)} + \ell_i^{(2)}$，其中

$$\ell_i^{(1)} = z_i \log q_i + (1 - z_i) \log(1 - q_i)$$

仅通过 $q_i = \operatorname{expit}(\mathbf{x}_i^\intercal \boldsymbol{\alpha} + \mathbf{x}_i^{(r)\intercal} \boldsymbol{\delta}_{1,\,s})$依赖于 $\boldsymbol{\phi}_1$，并且

$$\ell_i^{(2)} = z_i \big[\log f_{\text{BB}}(y_i \mid n_i, \mu_i, \kappa) - \log(1 - p_{0,i})\big]$$

仅通过 μ_i 和 κ依赖于 $\boldsymbol{\phi}_2$。

交叉导数 $\partial \ell^{(1)} / \partial \boldsymbol{\phi}_2$。 由于 $\ell_i^{(1)}$ 不涉及 $\boldsymbol{\phi}_2$ 的任何元素，因此该导数恒为零。

交叉导数 $\partial \ell^{(2)}/\partial \boldsymbol{\phi}_1$。 写作 $\ell_i^{(2)} = z_i \cdot g(\mu_i, \kappa, y_i, n_i)$。指示变量 $z_i = \mathbf{1}(y_i > 0)$ 是观测数据的函数，而不是参数的函数，因此 $\partial z_i/\partial \boldsymbol{\phi}_1 = \mathbf{0}$ ，进而 $\partial \ell_i^{(2)}/\partial \boldsymbol{\phi}_1 = \mathbf{0}$。

合并后， $\partial^2 \ell_i/\partial \boldsymbol{\phi}_1 \, \partial \boldsymbol{\phi}_2^{\intercal} = \mathbf{0}$，取期望后得到（B.9）。 □

注 B.14 （门槛模型与零膨胀模型）。命题 B.22 中的分块对角性 是门槛（两部分）模型的一般结构性质（Cragg，1971；Mullahy，1986）：参与指示变量 $z_i = \mathbf{1}$（$y_i > 0$）是可观测的，因此似然的每个组成部分都依赖于互不重叠的参数子集。相比之下，在零膨胀模型中（例如 Lambert，1992），风险指示变量是潜在的，膨胀参数和计数参数都会进入每个观测的边际似然，从而使费舍尔信息中的两个组成部分相互关联（全面比较参见 B. Neelon 等人，2016年）。

命题 B.23 （参数空间中的广延边际信息）。提供者的广延边际费舍尔信息块为 i

$$\mathcal{I}_{\text{ext},i} = q_i(1 - q_i)\, \mathbf{d}_{1,i}\, \mathbf{d}_{1,i}^{\intercal}, \tag{B.10}$$

其中，$\mathbf{d}_{1,\,i} = \partial \eta_i^{(1)}/\partial$（$\boldsymbol{\alpha}^{\intercal}$、$\boldsymbol{\delta}_{1,\,s}^{\intercal}$）$^{\intercal}$ 是相对于所有广延边际参数的广延边际线性预测量的梯度。

证明。根据注 B.12，线性预测量的标量信息量为$\mathcal{I}_{\text{ext},\,i} = q_i$（$1 - q_i$）。链式法则 $\partial \ell_i^{(1)}/\partial \boldsymbol{\phi}_1 =$ （$\partial \ell_i^{(1)}/\partial \eta_i^{(1)}$）$\mathbf{d}_{1,\,i}$ 给出了秩一外积形式（B.10）。□

注 B.15 （费舍尔信息率）。广延边际信息 q_i（$1 - q_i$）为 O（1）且是非随机的（注 B.12）。集约边际信息涉及三伽马函数和 ψ_1（a） $+ \psi_1$（b），其在 $\kappa \to 0$时发散；对于实证相关的范围 $\kappa \in [2.5,\, 26.5]$，信息量表现良好，但这些发散项是第 C 节和4中观察到的集约边际参数较大设计效应比的根源。

B.3. **可识别性**。 本小节给出第3节中所述可识别性定理的完整证明 第3节 (定理 4)。我们首先重述正则条件，然后展开五步证明。接下来给出三个辅助命题：截断贝塔-二项分布中 μ 与 κ 的分离（采用修正后的 $n \geq 3$阈值）、跨边际协方差 $\boldsymbol{\Sigma}_{12}$的似然不可识别性，以及后验分布充分集中的充分条件。

正则条件。 可识别性需要满足以下条件。这些条件与 定理 4 中的表述完全一致；我们还补充说明了各条件的作用。

(C1) 满秩 秩。 设计矩阵满秩：$\mathrm{rank}(\mathbf{X}) = P$ 且 $\mathrm{rank}(\mathbf{V}) = Q$。

(C2)州级重复。 每个州 s 至少有 q 个服务提供者，其协变量向量 $\mathbf{x}_i^{(r)}$ 线性无关，并且至少有一个零响应者，且至少有 q 个正响应者。(C3) 试验规模变异。 $n_i \geq 3$ 对所有 i 且 $z_i = 1$均成立，并且每个州都存在服务提供者 i, j满足 $s[i] = s[j]$ 和 $n_i \neq n_j$。(C4) 先验支持。 先验 $\pi(\boldsymbol{\theta})$ 在参数空间上具有满支持。

备注 B.16 (C2 中的不对称性)。 (C2) 中的要求是不对称的。广延边际要求每个州至少有一个零响应者（以识别 $q_s < 1$），而集约边际要求每个州至少有 q 个正响应者（以识别 q 维的 $\boldsymbol{\delta}_{2,s}$）。这种不对称性反映了单变量二元变量与 q 维回归参数在信息需求上的差异。

定理 B.2 （HBB 模型的可识别性）。 在条件 （C1）–（C4）下，完整参数向量 $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \kappa, \{\boldsymbol{\delta}_s\}, \boldsymbol{\Sigma}_\delta, \{\boldsymbol{\Gamma}_k\})$ 在后验分布下可识别。

证明。 论证分为两个层次的可识别性： （i）州内似然可识别州特定总系数和 κ；（ii）层级先验结构将这些总量加性分解为总体平均效应、政策调节变量和州级残差偏差。证明分为六个步骤。

步骤 *1*：广延边际的州级总系数。$\mathbf{1}(Y_i > 0)$ 的边际分布为 伯努利分布（q_i），其中

$$q_i = \mathrm{expit}\big(\mathbf{x}_i \intercal \boldsymbol{\alpha} + \mathbf{x}_i^{(r)\intercal} \boldsymbol{\delta}_{1,s[i]}\big).$$

划分 $\mathbf{x}_i = (\mathbf{x}_i^{(r)\intercal}, \mathbf{x}_i^{(f)\intercal})\intercal$ 并相应地 $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^{(r)\intercal}, \boldsymbol{\alpha}^{(f)\intercal})\intercal$，其中 $\mathbf{x}_i^{(r)} \in \mathbb{R}^q$ 包含具有州变化系数的协变量，而 $\mathbf{x}_i^{(f)} \in \mathbb{R}^{P-q}$ 包含仅固定效应协变量。[1] 定义 广延边际总州系数 (参见（3.12）)

$$\tilde{\boldsymbol{\alpha}}_s \;:=\; \boldsymbol{\alpha}^{(r)} + \boldsymbol{\delta}_{1,s} \;\in\; \mathbb{R}^q, \tag{B.11}$$

[1]当 $q = P$ （如模型 M2/M3 中）时，仅含固定效应的划分为空，论证适用于 $\boldsymbol{\alpha}^{(f)}$ 且$\mathbf{x}_i^{(f)}$ 为平凡情形。

因此，线性预测量变为 $\text{logit}(q_i) = \mathbf{x}_i^{(f)\top}\boldsymbol{\alpha}^{(f)} + \mathbf{x}_i^{(r)\top}\tilde{\boldsymbol{\alpha}}_{s[i]}$。参数 $\boldsymbol{\phi} := (\boldsymbol{\alpha}^{(f)}, \tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_S) \in \mathbb{R}^{(P-q)+Sq}$ 通过有效设计矩阵进入模型

$$\mathbf{Z} = \left[\mathbf{X}^{(f)}, \mathbf{D}_1\mathbf{X}_1^{(r)}, \ldots, \mathbf{D}_S\mathbf{X}_S^{(r)}\right] \in \mathbb{R}^{N\times[(P-q)+Sq]},$$

其中，$\mathbf{D}_s$ 用于选取州 s 中的服务提供者，$\mathbf{X}_s^{(r)}$ 是 $\mathbf{x}_i^{(r)}$ 值的州内子矩阵。条件 (C2) 确保对于每个 s，$\text{rank}(\mathbf{X}_s^{(r)}) = q$，因此州特定区块贡献了 Sq 个线性无关列。条件 (C1) 确保 $\text{rank}(\mathbf{X}) = P$，因此 $(P-q)$ 个仅含固定效应的列不在州特定区块的张成空间内。因此，$\text{rank}(\mathbf{Z}) = (P-q) + Sq$；根据 logit 的单射性，$\boldsymbol{\alpha}^{(f)}$ 和 $\{\tilde{\boldsymbol{\alpha}}_s\}_{s=1}^S$ 可由广延边际似然识别。

然而，$\boldsymbol{\alpha}^{(r)}$ 和 $\boldsymbol{\delta}_{1,s}$ 无法仅凭似然分别识别：对于任意 $\mathbf{c} \in \mathbb{R}^q$，作如下平移：$\boldsymbol{\alpha}^{(r)} \mapsto \boldsymbol{\alpha}^{(r)} + \mathbf{c}$，$\boldsymbol{\delta}_{1,s} \mapsto \boldsymbol{\delta}_{1,s} - \mathbf{c}$，对所有 s 均成立；这会保留每个 $\tilde{\boldsymbol{\alpha}}_s$，因此也保留整个似然。解决这一 q 维加性不确定性是步骤 2 的任务。

步骤 *2*：广延边际上的层级分解。层级模型规定 $\boldsymbol{\delta}_{1,s} = \boldsymbol{\Gamma}_1\mathbf{v}_s + \boldsymbol{\varepsilon}_{1,s}$ 其中 $\boldsymbol{\varepsilon}_{1,s} \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon,1})$。代入（B.11）得，

$$\tilde{\boldsymbol{\alpha}}_s = \boldsymbol{\alpha}^{(r)} + \boldsymbol{\Gamma}_1\mathbf{v}_s + \boldsymbol{\varepsilon}_{1,s}.$$

由于 $\{\tilde{\boldsymbol{\alpha}}_s\}$ 由步骤 1 识别，且观测到了 $\{\mathbf{v}_s\}$，因此这是一个以 $\tilde{\boldsymbol{\alpha}}_s$ 为因变量、以 $\mathbf{v}_s$ 为自变量且截距为 $\boldsymbol{\alpha}^{(r)}$ 的多元回归。在秩 $(\mathbf{V}) = Q$ ((C1)) 且 $S > Q$ 的条件下，普通最小二乘解可唯一确定斜率矩阵 $\boldsymbol{\Gamma}_1$ 和截距 $\boldsymbol{\alpha}^{(r)}$。重要的是，$\boldsymbol{\alpha}^{(r)}$ 的可识别性依赖于层级中心化约定 $\mathbb{E}[\boldsymbol{\varepsilon}_{1,s}] = \mathbf{0}$：这一结构性假设确定了州际回归的截距，从而消除了州内似然留下的平移歧义。识别出 $\boldsymbol{\alpha}^{(r)}$ 后，州偏差可写为 $\boldsymbol{\delta}_{1,s} = \tilde{\boldsymbol{\alpha}}_s - \boldsymbol{\alpha}^{(r)}$，残差则为 $\boldsymbol{\varepsilon}_{1,s} = \boldsymbol{\delta}_{1,s} - \boldsymbol{\Gamma}_1\mathbf{v}_s$。

步骤 3：密集边际州级总系数和 κ。条件于 $z_i = 1$（即 $Y_i > 0$），Y_i 的分布为 ZT-BetaBin(n_i, μ_i, κ)，其中 $\mu_i = \text{expit}(\mathbf{x}_i^\top\boldsymbol{\beta} + \mathbf{x}_i^{(r)\top}\boldsymbol{\delta}_{2,s[i]})$。根据下文的命题 B.24，只要满足 $n_i \geq 3$（(C3)），就能从截断似然中识别出参数对 (μ_i, κ)。一旦为所有正响应者确定了 $\{\mu_i\}$，将 $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(r)\top}, \boldsymbol{\beta}^{(f)\top})^\top$ 划分，并定义密集边际州系数总和

$$\tilde{\boldsymbol{\beta}}_s := \boldsymbol{\beta}^{(r)} + \boldsymbol{\delta}_{2,s} \in \mathbb{R}^q, \tag{B.12}$$

使得 $\operatorname{logit}(\mu_i) = \mathbf{x}_i^{(f)\top}\boldsymbol{\beta}^{(f)} + \mathbf{x}_i^{(r)\top}\tilde{\boldsymbol{\beta}}_{s[i]}$。与步骤 1 相同的论证——利用 logit 的单射性、有效设计矩阵和秩条件(C1)–(C2)——可从密集边际似然中识别出 $\boldsymbol{\beta}^{(f)}$ 和 $\{\tilde{\boldsymbol{\beta}}_s\}_{s=1}^{S}$。与广延边际一样，$\boldsymbol{\beta}^{(r)}$ 和 $\boldsymbol{\delta}_{2,\,s}$ 只能通过层次层单独识别。

步骤 *4*：对集约边际进行层级分解。与步骤 2 类似，˜层级设定 $\boldsymbol{\delta}_{2,s} = \boldsymbol{\Gamma}_2\mathbf{v}_s + \boldsymbol{\varepsilon}_{2,s}$ 得出州际回归 $\boldsymbol{\beta}_s = \boldsymbol{\beta}^{(r)} + \boldsymbol{\Gamma}_2\mathbf{v}_s + \boldsymbol{\varepsilon}_{2,s}$。在相同的秩条件下，$\boldsymbol{\Gamma}_2$ 和 $\boldsymbol{\beta}^{(r)}$ 可被唯一确定，因此$\boldsymbol{\delta}_{2,s} = \tilde{\boldsymbol{\beta}}_s - \boldsymbol{\beta}^{(r)}$ 且$\boldsymbol{\varepsilon}_{2,s} = \boldsymbol{\delta}_{2,s} - \boldsymbol{\Gamma}_2\mathbf{v}_s$。

步骤 *5*：对 $\boldsymbol{\Sigma}_\varepsilon$ 进行可识别性分析。</*style*}州残差效应 $(,_s\top,\boldsymbol{\varepsilon}_{2,s}\top)\top$ 可由步骤 1–4 对 $s = 1,\ldots,S$识别。这些来自 $\mathrm{N}_{2q}(\mathbf{0},\boldsymbol{\Sigma}_\varepsilon)$ 的 $S = 51$ 次独立抽样，在 $S > 2q$ 的条件下可以识别协方差矩阵 $\boldsymbol{\Sigma}_\varepsilon$（从而样本协方差满秩）。在我们的应用中，$q = 5$，因此 $S = 51 > 2q = 10$。随后利用单位矩阵关系 $\operatorname{Var}(\boldsymbol{\delta}_{k,s}) = \boldsymbol{\Gamma}_k \operatorname{Var}(\mathbf{v}_s)\boldsymbol{\Gamma}_k\top + \boldsymbol{\Sigma}_{\varepsilon,k}$，根据 $\boldsymbol{\Gamma}_k$、$\boldsymbol{\Sigma}_\varepsilon$ 以及$\mathbf{v}_s$ 的方差重构完整协方差矩阵 $\boldsymbol{\Sigma}_\delta$。

步骤 *6*：不存在标签切换歧义。与成分标签可交换的混合模型不同，障碍模型（Mullahy，1986）具有一个可观测的划分：$z_i = \mathbf{1}(Y_i > 0)$ 是数据的确定性函数。第 1 部分（广延）和第 2 部分（密集型）作用于不同的可观测子集。任何成分标签的置换都无法保持似然不变，因此不存在标签切换不变性。□

命题 B.24（μ 和 κ在零截断-贝塔-二项分布中的可识别性）。对于 $n \geq 3$，参数(μ, κ) 可由零截断贝塔-二项分布的完整似然识别：如果

$$f_{\text{ZT-BetaBin}}(y \mid n, \mu_1, \kappa_1) = f_{\text{ZT-BetaBin}}(y \mid n, \mu_2, \kappa_2) \quad \text{for all } y \in \{1,\ldots,n\}, \tag{B.13}$$

那么$(\mu_1, \kappa_1) = (\mu_2, \kappa_2)$。

证明。我们采用一种基于连续概率比的构造性论证。记$a = \mu\kappa$ 和 $b = (1-\mu)\kappa$为（未截断）贝塔-二项分布的形状参数。

步骤 *1*：截断在相邻比中会抵消。由于 $f_{\text{ZT-BetaBin}}(y \mid n, \mu, \kappa) = f_{\text{BetaBin}}(y \mid n, \mu, \kappa)/(1 - p_0)$ 对于 $y \geq 1$，相邻比

$$R_y := \frac{f_{\text{ZT-BetaBin}}(y)}{f_{\text{ZT-BetaBin}}(y-1)} = \frac{f_{\text{BetaBin}}(y)}{f_{\text{BetaBin}}(y-1)} = \frac{n-y+1}{y}\cdot\frac{a+y-1}{b+n-y}, \qquad y = 2,\ldots,n,$$

仅取决于（a，b，n），而不取决于 p_0。

步骤 *2*：两个比率确定（a，b）当 $n \geq 3$时。对于 $n \geq 3$，比率 R_2 和 R_3给出两个独立方程。定义

$$r_2 = \frac{2}{n-1}\,R_2 = \frac{a+1}{b+n-2}, \qquad r_3 = \frac{3}{n-2}\,R_3 = \frac{a+2}{b+n-3}.$$

交叉相乘得到线性方程组

$$a + 1 = r_2\,(b + n - 2),$$
$$a + 2 = r_3\,(b + n - 3).$$

将第一个方程从第二个方程中减去：$1 = (r_3 - r_2)\,b + r_3(3-n) - r_2(2-n)$，因此

$$b = \frac{r_2(n-2) - r_3(n-3) + 1}{r_3 - r_2}, \tag{B.14}$$

然后 $a = r_2(b+n-2) - 1$。分母 $r_3 - r_2 \neq 0$ 对所有 $a, b > 0$均成立，因为

$$r_3 - r_2 = \frac{(a+2)(b+n-2) - (a+1)(b+n-3)}{(b+n-2)(b+n-3)} = \frac{a+b+n-1}{(b+n-2)(b+n-3)} > 0.$$

因此，(a, b) 由观测截断分布唯一确定，并且 $\mu = a/(a+b)$，$\kappa = a + b$。

步骤 *3*：$n = 2$ 边界。对于 $n = 2$，只能得到单个比率 R_2，从而得到一个含两个未知数的方程：可识别性失败。有关等价参数所构成的显式一维流形，请参见 注释 B.17 。□

注释 B.17（$n = 2$ 边界情形）。对于 $n = 2$，截断支持集$\{1, 2\}$ 仅为两个未知量（μ，κ）提供一个自由概率，因此无法实现可识别性。连续比率证明明确说明了这一点：只有 R_2 可用，因而留下一个等价参数的一维流形。这说明了 (C3) 中要求$n_i \geq 3$ 的原因；NSECE 中全部 4,392 名正响应者（最小值 $n_i = 4$）均满足该要求。

注释 B.18（试验规模变异的作用）。(C3) 中要求每个州内 $n_i \neq n_j$，这在最小充分性之外增强了可识别性：具有相同 μ_i 但 n_i 不同的观测值，可通过 p_0 对 n 的非线性依赖，为 κ 提供不同的信息量。

命题 B.25（似然不可识别性 $\boldsymbol{\Sigma}_{12}$）。非对角块$\boldsymbol{\Sigma}_{12}$ = 协方差（$\varepsilon_{1,\,s}$，$\varepsilon_{2,\,s}$）无法仅由似然识别（参见 *Ghosal* 等人*2020*年年，第 *3.1* 节中的类似讨论）。对数似然 $\ell(\boldsymbol{\theta})$ 在给定已实现值 $\boldsymbol{\Sigma}_{12}$ 的条件下，对 $\{\varepsilon_{1,s}, \varepsilon_{2,s}\}_{s=1}^{S}$完全不变。

证明。 根据分块对角费舍尔信息（命题 B.22），对数似然可分解为 $\ell = \ell_{\text{ext}} + \ell_{\text{int}}$，其中 ℓ_{ext} 取决于（$\boldsymbol{\alpha}$，$\boldsymbol{\delta}_{1,\cdot}$），而 ℓ_{int} 取决于（$\boldsymbol{\beta}$，$\boldsymbol{\delta}_{2,\cdot}$，$\kappa$）。这两个部分都不涉及 $\boldsymbol{\Sigma}_{\varepsilon}$：协方差仅通过层级先验进入模型

$$\boldsymbol{\varepsilon}_s = \begin{pmatrix} \boldsymbol{\varepsilon}_{1,s} \\ \boldsymbol{\varepsilon}_{2,s} \end{pmatrix} \sim \mathsf{N}_{2q}\left(\mathbf{0},\ \begin{pmatrix} \boldsymbol{\Sigma}_{\varepsilon,1} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{\varepsilon,2} \end{pmatrix}\right).$$

给定实现的 $\{\boldsymbol{\varepsilon}_s\}$ 后，似然是仅关于 $(\boldsymbol{\alpha}、\ \boldsymbol{\beta}、\ \kappa)$ 的固定函数，而 $\boldsymbol{\Sigma}_{12}$ 则完全不起作用。□

注释 B.19 ($\boldsymbol{\Sigma}_{12}$ 作为残差协方差)。 跨边际协方差 $\boldsymbol{\Sigma}_{12}$ =Cov($\boldsymbol{\varepsilon}_{1、\ s}$、 $\boldsymbol{\varepsilon}_{2、\ s}$) 表示通过 $\boldsymbol{\Gamma}_1\mathbf{v}_s$ 和 $\boldsymbol{\Gamma}_2\mathbf{v}_s$剔除观测到的政策效应后的残差跨边际协方差。参与和强度之间的边际跨边际关系既包含这一残差作用渠道，也包含经由 $\boldsymbol{\Gamma}_k$的直接政策调节通道。具体而言，

$$\mathrm{Cov}(\boldsymbol{\delta}_{1,s},\ \boldsymbol{\delta}_{2,s}) = \boldsymbol{\Gamma}_1\ \mathrm{Var}(\mathbf{v}_s)\,\boldsymbol{\Gamma}_2{}^{\intercal} + \boldsymbol{\Sigma}_{12},$$

因此，发现 $\boldsymbol{\Sigma}_{12} \approx \mathbf{0}$ 并不意味着这两个边际之间毫无关联——这仅表明政策完全解释了它们的跨边际协变。

命题 B.26 ($\boldsymbol{\Sigma}_{12}$的集中性)。
当满足以下条件时， $\boldsymbol{\Sigma}_{12}$ 的后验分布会高度集中：

(i) $S > 2q + 1$，因此 $\{\boldsymbol{\varepsilon}_s\}$ 的样本协方差具有正定的后验众数；(ii) 各州的个体状态效应 $\boldsymbol{\varepsilon}_{1、\ s}$ 和 $\boldsymbol{\varepsilon}_{2、\ s}$ 本身都能得到良好估计，这要求满足状态层级重复条件 (C2)；以及(iii) 相关矩阵 $\mathrm{LKJ}(\eta)$ 先验（*Lewandowski* 等人，*2009*） $\mathbf{R}_{\varepsilon}$ 在 $\eta \geq 1$的条件下提供正则化；当数据对跨边际依赖所含信息较弱时，它会使后验分布向独立性收缩。

对于带有 $q = 5$ 的模型：要求 $S > 2q + 1 = 11$ 由 $S = 51$ 充分满足。 $\boldsymbol{\Sigma}_{12}$ 中每个元素的后验标准差以速率 $O\left(S^{-1/2}\right) \approx 0.14$ 递减。

注释 B.20 (频率学派与贝叶斯识别)。 定理 B.2 中的可识别性结果针对的是后验分布，而非频率学派的最大似然估计量。在贝叶斯框架下，可识别性意味着当 $N \to \infty$（参见伯恩斯坦–冯·米塞斯定理

在模型设定错误的情况下；Kleijn & van der Vaart，2012），这要求同时满足似然识别（抽样模型能够确定 $\boldsymbol{\theta}$）和先验识别（先验分布不会在不可识别子空间上赋予概率质量）。步骤 1 和 3 确立了州级总系数 $\{\tilde{\boldsymbol{\alpha}}_s\}$、$\{\boldsymbol{\beta}_s\}$ 和 κ 的似然识别；步骤 2 和 4 表明，层级中心化约定 $\mathbb{E}[\boldsymbol{\varepsilon}_{k,s}] = \mathbf{0}$ 消除了加性平移歧义，并识别出个体分量（$\boldsymbol{\alpha}$，$\boldsymbol{\beta}$，$\boldsymbol{\delta}_s$，$\boldsymbol{\Gamma}_k$）；步骤 5 确立了边际内区块 $\boldsymbol{\Sigma}_{\varepsilon,\,1}$ 和 $\boldsymbol{\Sigma}_{\varepsilon,\,2}$的似然识别。

跨边际区块 $\boldsymbol{\Sigma}_{12}$ 是个例外：命题 B.25表明，它无法从条件似然中获得任何信息量。正如注释 1所述，对 $\boldsymbol{\delta}_s$ 积分后得到的边际似然确实会使各边际相互关联，并且原则上可以识别 $\boldsymbol{\Sigma}_{12}$，但此处采用的条件形式更适合进行 MCMC 计算。无论采用哪种表述，$\mathrm{LKJ}(\eta)$先验都提供了必要的正则化：对于由 $2q \times 2q$ 组构成的 $S = 51$ 协方差，频率学派 REML 估计在数值上较为脆弱（容易出现边界解和奇异海森矩阵）。这并非模型的缺陷，而是障碍因子分解在结构上的必然结果（命题 B.22 ）。 $\boldsymbol{\Sigma}_{12}$ 的后验分布以 $O(S^{-1/2})$ 的速率集中 —— 该速率由 $S = 51$ 州级观测决定——而非以其余参数所具有的参数收敛率$(N^{-1/2})$集中。较慢的集中速率是学习跨边际依赖结构所付出的代价；我们认为，鉴于理解限制可及性的州是否也会加剧条件入学这一问题具有实质意义，这一代价是值得的。

附录C. 调查加权推断

本附录阐述夹心方差估计量和 Cholesky 仿射变换所依据的调查设计理论，这些内容已在第3节中介绍。我们首先给出正式的伪似然框架，并证明伪后验分布是适定的（定理 2）；随后确立 NSECE 权重的伯恩斯坦–冯·米塞斯条件及其有限样本诊断，推导聚类稳健夹心估计量及其区块结构，给出 Cholesky 仿射变换的完整证明（定理 3），最后介绍设计效应比分类与诊断。全文将使用第 B 节 （补充材料 B）中的得分函数和费舍尔信息结果，以及第3节中确立的记号。

C.1. 伪似然与可分性。 我们总结第 3.5 节中的伪似然结构。根据命题 6，分解 $\ell^{(w)} = \ell^{(w)}_{\text{扩展}} + \ell^{(w)}_{\text{截距}}$在权重 $\tilde{w}_i = w_i N / \sum_j w_j$ 下成立。我们列出三条说明性注释。

注 C.21（权重归一化）。 调查文献中存在两种归一化约定。全文采用总和为N的约定， $\tilde{w}_i = w_i N / \sum_j w_j$使得 $\sum_i \tilde{w}_i = N$，这在贝叶斯伪后验文献中是标准做法（Sav-ˆitsky & Toth，2016年；Williams & Savitsky，2021）。另一种总和为M的约定保留未标准化权重 $\tilde{w}_i = w_i$，对于 NSECE，这给出 $\sum_i w_i \approx 119,\ 593$。在该约定下，伪似然大致相当于119，593个观测值，而不是 $N = 6,\ 809$个观测值，从而导致伪后验出现极度过度集中，可信区间也窄得多。总和为N 的归一化确保伪似然在尺度上与基于N 个观测值的普通似然相当。

注记 C.22（加权得分函数）。 广延边际和集约边际的观测层级得分函数，以及相应的费舍尔信息量和符号验证，详见第 B.1 节 （补充材料 B）。此处使用的加权版本，只需将每个观测的得分乘以 $\tilde{w}_i$；命题 6 的可分性确保加权广延边际得分仅取决于 $(\boldsymbol{\alpha}, \boldsymbol{\delta}_{1,\cdot})$，而加权集约边际得分仅取决于 $(\boldsymbol{\beta}, \boldsymbol{\delta}_{2,\cdot}, \kappa)$。

注记 C.23（加权下可分离性的重要性）。 命题 6 中的可分离性结论并非对所有调查调整都成立。改变似然结构的方法——例如多层权重平滑（Rabe-Hesketh 和 Skrondal，2006）或条件似然方法——可能会破坏加性结构。观测层级的指数化 $[f_i]^{\tilde{w}_i}$ 之所以能保留该结构，恰恰是因为 $\tilde{w}_i$ 在对数尺度上是一个标量乘数，不会将两个参数块耦合起来。夹心估计量（第 3.5 节）利用了这种可分性，其中 $\mathbf{H}_{\text{obs}}$ 的块对角性可直接由加性分解得出。

C.2. **伪后验适定性**。 现在，我们给出定理 2的完整证明；该定理已在第 3.5 节中陈述并给出了证明概要。

定理 *2* 的证明定理 2。 我们验证伪后验分布

$$\pi^{(w)}(\boldsymbol{\theta} \mid \mathbf{y}) \propto \exp\left[\ell^{(w)}(\boldsymbol{\theta})\right] \pi(\boldsymbol{\theta})$$

在参数空间上可积。

步骤 1（伪似然的有界性）。对于每个观测 i，HBB概率满足 $f_{\text{HBB}}(y_i \mid \boldsymbol{\theta}) \in (0, 1]$ 对所有位于参数空间内部的 $\boldsymbol{\theta}$

˜。由于 $\tilde{w}_i > 0$，当 $f_i \le 1$时，我们有 $f_i^{w_i} \le 1$。因此

$$\exp\big[\ell^{(w)}(\boldsymbol{\theta})\big] = \prod_{i=1}^{N} f_i^{\tilde{w}_i} \le 1 \qquad \text{for all } \boldsymbol{\theta}.$$

步骤 2（先验的适当性）。先验 $\pi(\boldsymbol{\theta})$ 的每个组成部分都是适当分布：

- $\boldsymbol{\alpha}$，$\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, 2_2 \mathbf{I}_P)$（Gelman 等人，2008）：具有有限方差的适当分布；
- $\boldsymbol{\tau} \sim$ 逐分量 Half-Normal(0，1)：在 $\mathbb{R}_{>0}$ 上是适当的；
- $\mathbf{R}_\varepsilon \sim \mathrm{LKJ}(2)$（Lewandowski 等人，2009）：在 $2q \times 2q$相关矩阵的空间上是适当的；
- $\log \kappa \sim \mathrm{N}(2, 1.5_2)$：适当。

联合先验是适当分布的乘积，因此是适当的：$\int \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} = 1$。

步骤 3（可积性）。综合步骤 1 与步骤 2：

$$\int \exp\big[\ell^{(w)}(\boldsymbol{\theta})\big]\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} \;\le\; \int 1 \cdot \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} = 1 < \infty.$$

此外，$\exp[\ell^{(w)}(\boldsymbol{\theta})] > 0$ 在参数空间内部（因为该处所有概率项 f_i 均严格为正），因此伪后验是非退化的。□

注 C.24（适当性的权重不变性）。上述证明并不依赖于权重的具体取值或归一化方式——仅依赖于其正性与有限性。因此，无论权重是否被归一化ˆ为总和等于 N、总和等于总体规模 M，还是采用任何其他正的缩放，适当性均成立。然而，伪后验的形状 与集中程度 都关键取决于ˆ归一化约定，正如注 C.21所讨论的：总和等于M 的约定会产生极端过度集中，而总和等于N 的约定则保留实际样本的信息量。

注 C.25（与非正规先验对比）。若回归系数采用非正规（平坦）先验，则要保证后验分布适当，必须要求伪似然本身可积——也就是说，$\int \exp[\ell^{(w)}(\boldsymbol{\theta})]\, d\boldsymbol{\theta} < \infty$。在这种情况下，可积性由有效样本量（而非名义样本量）决定。若 ESS $\approx 1,$ 797 且 $\dim(\boldsymbol{\theta}) \le 606$，可积性很可能成立，但正式验证较为复杂：必须确认对于每个参数块，伪最大似然估计处的加权海森矩阵均为满秩。采用适当先验可完全避开这一问题，从而无需对数据作任何条件，即可确保任意正权重配置下后验分布均适当。

C.3.**伯恩斯坦–冯·米塞斯近似**。第 3.5 节中的三明治校正基于伪后验的伯恩斯坦–冯·米塞斯（BvM）定理：当$N \to \infty$时，伪后验收敛于一个正态分布，其中心为伪最大似然估计，方差由观测海森矩阵的逆给出，即 $\mathbf{H}^{-1}$ 观测。我们陈述该精确结果，验证其在 HBB 模型中的正则条件，并针对 NSECE 权重配置提供有限样本诊断，以此说明三明治校正作为稳健性度量的依据。

定理 C.3（伪后验的设计一致的伯恩斯坦–冯·米塞斯定理）。 假设满足以下正则性条件：

（C1）光滑性。 对参数空间 Θ 内部的 $\boldsymbol{\theta}$，对数密度 $\log f_{\mathrm{HBB}}(y \mid \boldsymbol{\theta})$ 二阶连续可微。（C2）参数位于内部。 真实的超总体参数 $\boldsymbol{\theta}_0$ 位于 Θ 的内部。（C3）权重正则性。 最大标准化权重与最小标准化权重的比率 $\max_i \tilde{w}_i / \min_i \tilde{w}_i = O(N^{\delta})$ 满足条件，其中某个 $\delta < 1/2$，且 $N \to \infty$。（C4）信息正则性。 加权海森矩阵 $\mathbf{H}_{\mathrm{obs}}(\boldsymbol{\theta}_0)$ 为正定矩阵，其特征值与零和无穷远均保持一定距离。（C5）先验正则性。 先验密度 $\pi(\boldsymbol{\theta})$ 在 $\boldsymbol{\theta}_0$的某个邻域内连续且严格为正。

于是，当 $N \to \infty$,

$$\sup_{B \in \mathcal{B}} \left| \pi^{(w)}(\boldsymbol{\theta} \in B \mid \mathbf{y}) - \Phi_{\hat{\boldsymbol{\theta}}^{(w)}, \mathbf{H}_{\mathrm{obs}}^{-1}}(B) \right| \xrightarrow{P_{\boldsymbol{\theta}_0}} 0, \tag{C.1}$$

其中， $\mathcal{B}$ 是博雷尔可测凸集类，$\hat{}\boldsymbol{\theta}^{(w)}$ 是伪最大似然估计，$\Phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$表示 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的累积分布函数，$\mathbf{H}_{\mathrm{obs}}$ 是伪对数似然在$\hat{}\boldsymbol{\theta}^{(w)}$处取值时的观测海森矩阵。

这一结果是经典贝叶斯—冯·米塞斯定理的伪后验类比（Kleijn & van der Vaart, 2012），并由 Savitsky 和 Toth（2016年）以及 Williams 和 Savitsky（2021）将其扩展到调查加权情形。与标准贝叶斯—冯·米塞斯定理的关键区别在于条件(C3)，该条件限制了权重比随样本量增长的速度。值得注意的是，伪后验方差 $\mathbf{H}^{-1}$ 观测 不同于 $\hat{}\boldsymbol{\theta}^{(w)}$ 的设计一致的抽样方差 ,

$$\hat{\boldsymbol{\theta}}^{(w)} \dot{\sim} \mathsf{N}\big(\boldsymbol{\theta}_0,\ \mathbf{V}_{\mathrm{sand}}\big), \qquad \mathbf{V}_{\mathrm{sand}} = \mathbf{H}_{\mathrm{obs}}^{-1}\, \mathbf{J}_{\mathrm{cluster}}\, \mathbf{H}_{\mathrm{obs}}^{-1},$$

如式（3.25）所定义。在信息性抽样或模型错设的情况下，$\mathbf{J}_{\mathrm{cluster}} \neq \mathbf{H}_{\mathrm{obs}}$，因此，基于$\mathbf{H}^{-1}$ obs 的朴素可信区间作为一种衡量.....

频率学派不确定性。这种不匹配正是C.4节中所发展的三明治校正的动因：下文的定理 3量化了由此产生的覆盖率失真，并给出了操作性校准。

验证 。我们针对 HBB 模型与 NSECE 数据逐一验证各项条件 .

(C1)：光滑性。HBB 对数密度涉及 $\log\Gamma$ 函数（经由贝塔–二项分布概率质量函数）、逻辑斯蒂链接 $\sigma(\eta) = 1/(1+e^{-\eta})$，以及零概率 p_0（由有限个 $\log\Gamma$ 项求和）。每个分量在参数空间内部均为 C^∞，因此复合 $\log f_{\text{HBB}}$ 关于 $\boldsymbol{\theta}$ 光滑。 ✓

(C2)：内部参数。参数空间为 $\Theta = \mathbb{R}^P \times \mathbb{R}^P \times \mathbb{R}^{2QR} \times \mathbb{S}^{2Q}_{++} \times \mathbb{R}_{>0}$，且任意满足 $\kappa > 0$ 与 $\boldsymbol{\Sigma}_\varepsilon \succ 0$ 的有限参数向量均位于内部。 ✓

(C3)：权重正则性。该条件需要专门分析；参见命题 C.27 （见下文）。

(C4)：信息正则性。由命题 B.22，费舍尔信息在广延边际与集约边际之间呈块对角。对于广延块，正定性源于 秩$(\mathbf{X}) = P$ 以及 $q_i \in$ （0，1） 对所有 i 成立（标准逻辑回归理论）。对于集约块，μ 与 κ 的分离要求分母 n_i 存在变异（命题 B.24），且在实证上得到满足（n_i 在正响应者中的取值范围为 1 至 378）。 ✓

(C5)：先验正则性。第3.6节中指定的先验在内部均连续且严格为正：回归系数服从 N(0, $2^2\,\mathbf{I}_P$)（Gelman 等人，2008），尺度参数服从半正态分布(0，1)，相关矩阵服从 LKJ(2)（Lewandowski 等人，2009），$\log\kappa$ 服从 N(2，1.5$_2$)。 ✓

命题 C.27（有限样本权重诊断）。对于 $N = 6,785$ 和 $w_{\max}/w_{\min} = 462$，*NSECE 2019*年数据的权重比超过 $N^{1/2} \approx 82.4$，这表明，对于此设计，充分条件 (C3) 并未得到充分满足。计算

$$\delta = \frac{\log 462}{\log 6{,}785} = \frac{6.136}{8.822} \approx 0.695 > 0.5 \tag{C.2}$$

是一项有限样本诊断，而非对条件 O 的大样本渐近检验(C3)。这促使我们采用夹心方差校正作为稳健性度量。

注释 C.26（诊断结果的解释）。四项观察有助于明确命题 C.27的适用范围。

（a）条件（C3）是充分而非必要条件。即使不满足这一界限，只要高权重观测值没有主导伪似然，贝叶斯—冯·米塞斯近似仍可能足够准确。（b）极端权重观测值所占比例很小：权重排名前1%（权重 $w_i > 100$）的观测值不足6,785个观测值中的70个。诊断结果是由少数离群权重驱动的，并非系统性模式。（c）权重截尾可以恢复（C3）。将权重上限设为第 99 百分位数可使$w_{\max}$ 降至约100，从而得到 $\delta \approx 0.522$；将权重上限设为第95百分位数（$w_{\max} \approx 50$）则得到 $\delta \approx 0.443 < 0.5$，恢复（C3）。D 节中将权重截尾作为敏感性分析进行考察D 节。（d）在实践中，无论（3.25）是否完全满足（C3），三明治校正都能提供适当的方差调整：它会自动考虑极端权重导致的方差膨胀。

推论 C.5（朴素可信区间的覆盖率失准）。 在 定理 *C.3* 的条件下，定义：

- 置信区间$_{\text{朴素}}(\alpha)$： 伪后验边际分布给出的 α水平可信区间，方差为 $[\mathbf{H}_{\text{obs}}]^{-1}$;

- 置信区间及$(\alpha) = \hat{\theta}_{(w)} \pm \overline{z_{\alpha/2}}\sqrt{V}$三明治校正区间。于是：*(a)* 朴素区间的渐近覆盖率为 $\Phi(z_{\alpha/2}\, r) - \Phi(-z_{\alpha/2}\, r)$，其中 $r = \sqrt{\lambda H/\lambda}$是朴素标准差与三明治标准差的比率，$\Phi$ 表示标准正态累积分布函数。*(b)* 当抽样为无信息抽样且模型正确设定时，$\overline{Jcluster} \approx Hobs$，因此 $r \approx 1$，朴素区间的覆盖率近似正确。*(c)* 在信息性抽样或模型错设的情况下，$r \neq 1$。若 $r < 1$（朴素方差过小），区间是反保守的（覆盖不足）——这是伪后验过度集中的典型情况。若 $r > 1$，区间则是保守的。

证明。 贝叶斯—冯·米塞斯定理（定理 C.3）表明，伪后验分布收敛于 $\mathrm{N}(\hat{\boldsymbol{\theta}}^{(w)}, \mathbf{H}^{-1}{}_{\text{obs}})$，而 $\hat{\boldsymbol{\theta}}^{(w)}$ 的正确抽样分布近似为 $\mathrm{N}(\boldsymbol{\theta}_0, \mathbf{V}_{\text{sand}})$。边际标准差之比决定覆盖率失真。□

注记 C.27（通过 Cholesky 变换进行校准）。 定理 3 给出了实际校准策略：仿射变换（3.26）对每个 MCMC 抽样进行重新缩放，使调整后的后验分布具有协方差 $\mathbf{V}_{\text{sand}}$ ，同时保留后验均值。该变换将朴素方差$\mathbf{H}^{-1}{}_{\text{obs}}$ 替换为设计一致的夹心方差，从而校正推论 C.5所量化的覆盖率失真。分块实现——分别对固定效应、随机效应和超参数应用该变换——详见 第 3.5 节；完整证明见下文第 C.5 节 。

C.4. **三明治方差估计量**。 正文（第 3.5 节）定义了面包矩阵 $\mathbf{H}_{\text{obs}}$ （3.23）、聚类稳健肉矩阵 $\mathbf{J}_{\text{cluster}}$ （3.24）以及三明治方差$\mathbf{V}_{\text{sand}}$ （3.25）。本小节将详细推导：矩阵$\mathbf{H}_{\text{obs}}$的分块对角结构、矩阵 $\mathbf{J}_{\text{cluster}}$的非分块对角结构、考虑分层的聚类稳健汇总公式（Binder，1983年），以及州变化系数对州内聚类的部分吸收。全文中，我们使用第 B.1 节中的得分函数和下文定义的夹心约定。

面包矩阵。 面包矩阵是后验均值 $\boldsymbol{\theta}$ 处观测海森矩阵的负加权值，

$$\mathbf{H}_{\text{obs}} = -\sum_{i=1}^{N} \tilde{w}_i \, \frac{\partial^2 \log f_{\text{HBB}}(y_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \, \partial \boldsymbol{\theta}^{\mathsf{T}}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{C.3}$$

由于门槛对数似然（注记 C.23）的可分性以及分块对角费舍尔信息（命题 B.22），面包矩阵具有双区块结构：

命题 C.28（分块对角面包矩阵）。 观测海森矩阵分解为

$$\mathbf{H}_{\text{obs}} = \begin{pmatrix} \mathbf{H}_{\text{ext}} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{\text{int}} \end{pmatrix}, \tag{C.4}$$

其中：

(i) 广延块是加权逻辑信息

$$\mathbf{H}_{\text{ext}} = \sum_{i=1}^{N} \tilde{w}_i \, q_i(1-q_i) \, \mathbf{d}_{1,i} \, \mathbf{d}_{1,i}{}^{\mathsf{T}}, \tag{C.5}$$

其中 $\mathbf{d}_{1,\,i} = \partial \eta_i^{(1)} / \partial \left(\boldsymbol{\alpha}^{\mathsf{T}}, \boldsymbol{\delta}_{1,\,s[i]}{}^{\mathsf{T}}\right)^{\mathsf{T}}$。该矩阵是非随机的：它取决于$\hat{}\; q_i = q_i(\boldsymbol{\theta})$ ，但不取决于观测数据 z_i。

（ii）密集型部分是截断贝塔分布-二项分布对数似然的加权负海森矩阵，

$$\mathbf{H}_{\text{int}} = -\sum_{i\in\mathcal{I}_1} \tilde{w}_i \, \frac{\partial^2 \ell_i^{(2+)}}{\partial \boldsymbol{\phi}_2 \, \partial \boldsymbol{\phi}_2 \mathsf{T}}\bigg|_{\hat{\boldsymbol{\theta}}}, \tag{C.6}$$

其中 $\boldsymbol{\phi}_2 = (\boldsymbol{\beta}\mathsf{T}, \log \kappa, \boldsymbol{\delta}_{2,s[i]}\mathsf{T})\mathsf{T}$，$\mathcal{I}_1 = \{i : y_i > 0\}$，以及 $\ell_i^{(2+)} = \log f_{\text{BB}}(y_i \mid n_i, \mu_i, \kappa) - \log(1 - p_{0,i})$。该块依赖于观测计数 y_i。

证明。根据 命题 B.22，交叉导数 $\partial^2 \ell_i / \partial \boldsymbol{\phi}_1 \, \partial \boldsymbol{\phi}_2 \mathsf{T} = \mathbf{0}$ 对每个观测 i。乘以 $\tilde{w}_i > 0$ 并求和可保持该零块，从而得到（C.4）。广延边际形式（C.5）由 $-\partial^2 \ell_i^{(1)} / \partial (\eta_i^{(1)})^2 = q_i(1 - q_i)$（命题 B.23）以及链式法则得出。□

肉矩　　阵与跨边际耦合。个体层面肉矩阵 is

$$\mathbf{J}_{\text{ind}} = \sum_{i=1}^{N} \tilde{w}_i^2 \, \mathbf{s}_i \, \mathbf{s}_i \mathsf{T}, \tag{C.7}$$

其中，$\mathbf{s}_i = \partial \log f_{\text{HBB}}(y_i \mid \boldsymbol{\theta}) / \partial \boldsymbol{\theta}\big|_{\hat{\boldsymbol{\theta}}}$ 是观测层面的得分。

注 C.28（肉矩阵非分块对角）。与面包矩阵不同，广延边际与集约边际之间的肉矩阵并非分块对角。其交叉块为

$$\mathbf{J}^{\text{ext,int}} = \sum_{i=1}^{N} \tilde{w}_i^2 \, \mathbf{s}_i^{\text{ext}} \, (\mathbf{s}_i^{\text{int}}) \, \mathsf{T} \, . \tag{C.8}$$

对于 $y_i = 0$，集约得分$\mathbf{s}_i^{\text{int}} = \mathbf{0}$，因此交叉乘积为零。但对于$y_i > 0$，$\mathbf{s}_i^{\text{ext}} = (1 - q_i)\mathbf{d}_{1,i}$ 和 $\mathbf{s}_i^{\text{int}} \neq \mathbf{0}$均如此，因此每个正观测都会对交叉块产生非零贡献。因此，夹心方差$\mathbf{V}_{\text{sand}} = \mathbf{H}^{-1}{}_{\text{obs}} \mathbf{J} \mathbf{H}^{-1}{}_{\text{obs}}$ 并非分块对角，尽管$\mathbf{H}_{\text{obs}}$ 是分块对角的。这种跨边际耦合是障碍模型中夹心估计量的普遍特征：单个初级抽样单位会同时为两个边际贡献加权得分。

注 C.29（等权重下的简化）。在均匀权重下 $\tilde{w}_i = 1$且模型设定正确时，信息恒等式给出 $\mathbb{E}[\mathbf{J}_{\text{ind}}] = \mathbf{H}_{\text{obs}}$（参见 Williams & Savitsky，2021，关于伪后验类比），因此$\mathbf{V}_{\text{sand}} \to \mathbf{H}^{-1}{}_{\text{obs}}$ 且夹心估计量简化为基于模型的方差。对于广延边际，$\mathbb{E}[\mathbf{J}_{\text{ext}}] = \mathbf{H}_{\text{ext}}$ 精确成立，因为逻辑斯蒂海森矩阵是非随机的（见注 B.12；在不等调查权重下，该恒等式不再成立，因为$\mathbf{H}$ 涉及 $\tilde{w}_i$ 而 $\mathbb{E}[\mathbf{J}]$

涉及 $\tilde{w}_i^2$）。对于集约边际，该恒等式仅在期望意义下成立，且$\mathbf{J}_{\text{int}}$ 与$\mathbf{H}_{\text{int}}$ 之间的有限样本差异不为零。

聚类稳健聚合。 个体层面的 $\mathbf{J}_{\text{ind}}$ 将观测值视为相互独立。在 NSECE 分层整群设计中（全国早期照护与教育调查项目团队，2022年），同一初级抽样单位（PSU）内的服务提供者共享社区特征，从而产生 PSU 内相关。聚类稳健肉矩阵以 PSU 层级聚合量取代个体得分，并在层内进行中心化。

定义 C.3（PSU 层级得分总和）。 对于层 $h = 1, \dots, H$ 以及层 h 内的初级抽样单位 $c =1, \dots, C_h$，定义

$$\bar{\mathbf{s}}_{hc} = \sum_{i \in \text{PSU}(h,c)} \tilde{w}_i \, \mathbf{s}_i, \tag{C.9}$$

其中 $\text{PSU}(h, c) = \{i : \text{vstratum}[i] = h,\ \text{vpsu}[i] = c\}$。层均值为

$$\bar{\mathbf{s}}_h = \frac{1}{C_h} \sum_{c=1}^{C_h} \bar{\mathbf{s}}_{hc}. \tag{C.10}$$

命题 C.29（聚类稳健肉矩阵）。 在具有 H 个层且第 h 层包含 C_h 个 *PSU* 的分层整群设计下，聚类稳健肉矩阵为

$$\boxed{\mathbf{J}_{\text{cluster}} = \sum_{h=1}^{H} \frac{C_h}{C_h - 1} \sum_{c=1}^{C_h} (\bar{\mathbf{s}}_{hc} - \bar{\mathbf{s}}_h)\,(\bar{\mathbf{s}}_{hc} - \bar{\mathbf{s}}_h)^{\intercal}.} \tag{C.11}$$

该估计量考虑了：(i) *PSU* 内相关性，通过将得分聚合为 *PSU* 总和；(ii) 分层，通过在层内对 *PSU* 得分进行中心化，使得只有 *PSU* 间的层内变异产生贡献；(iii) 不等整群规模，通过加权聚合的 *PSU* 总和 $\bar{\mathbf{s}}_{hc}$；(iv) 有限样本自由度，通过因子 $C_h/(C_h - 1)$。

证明。 分层抽样下加权得分总和的基于设计的方差可分解为

$$\operatorname*{Var}_{\text{design}}\Big(\sum_i \tilde{w}_i \, \mathbf{s}_i\Big) = \sum_{h=1}^{H} \operatorname*{Var}_h\Big(\sum_{c=1}^{C_h} \bar{\mathbf{s}}_{hc}\Big),$$

因为各层之间的抽样相互独立。在层 h 内，从由 C_h^* 个初级抽样单位构成的总体中进行不放回抽样。根据“终极整群”假设，即 $C_h/C_h^* \approx 0$（初级抽样单位的抽样比例可忽略不计），有限总体

校正项被省略，层内方差估计量为

$$\widehat{\mathrm{Var}}_h = \frac{C_h}{C_h - 1} \sum_{c=1}^{C_h} (\bar{\mathbf{s}}_{hc} - \bar{\mathbf{s}}_h)\,(\bar{\mathbf{s}}_{hc} - \bar{\mathbf{s}}_h)^{\intercal},$$

来自 C_h 个单位的总量方差的标准无偏估计量。对各层求和可得 $\mathbf{J}_{\text{cluster}}$。□

备注 C.30（自由度）。聚类稳健估计量有 $\sum_{h=1}^{H}(C_h - 1) = 415 - 30 = 385$ 个有效自由度。C_h 的范围为 5 至 36（中位数为 ≈ 14），每层至少贡献 4 个自由度。这远高于稳定进行夹心方差估计的经验阈值 $\sum_h (C_h - 1) > 30$。

备注 C.31（个体层级与聚类层级的 J）。当协变量条件下不存在 PSU 内相关时，$\mathbf{J}_{\text{cluster}}$ 和 $\mathbf{J}_{\text{ind}}$ 收敛于相同的总体量。实际中，PSU 内相关会使 $\mathbf{J}_{\text{cluster}}$ 相对于 $\mathbf{J}_{\text{ind}}$ 增大；参数特定比率 $\mathrm{diag}(\mathbf{V}_{\text{sand}})/\mathrm{diag}(\mathbf{H}^{-1}{}_{\text{obs}}\mathbf{J}_{\text{ind}}\mathbf{H}^{-1}{}_{\text{obs}})$ 可用于估计聚类设计效应。设计效应比（DER）的分类与诊断见下文 C.6 节。

命题 C.30 (SVC 对聚类的部分吸收)。州变化系数 $\boldsymbol{\varepsilon}_s$ 会部分吸收州内聚类。对于同一州 s 且属于同一 *PSU* 的任意两个服务提供者 i, j，

$$\mathbb{E}_{\boldsymbol{\varepsilon}_s}\big[\mathrm{Cov}(s_i,\, s_j \mid \boldsymbol{\varepsilon}_s)\big] \;\le\; \mathrm{Cov}(s_i,\, s_j), \tag{C.12}$$

当且仅当 $\mathrm{Cov}(\mathbb{E}[s_i \mid \boldsymbol{\varepsilon}_s],\, \mathbb{E}[s_j \mid \boldsymbol{\varepsilon}_s]) = 0$时，等号成立。

证明。根据全协方差公式，

$$\mathrm{Cov}(s_i,\, s_j) = \mathbb{E}\big[\mathrm{Cov}(s_i,\, s_j \mid \boldsymbol{\varepsilon}_s)\big] + \mathrm{Cov}\big(\mathbb{E}[s_i \mid \boldsymbol{\varepsilon}_s],\, \mathbb{E}[s_j \mid \boldsymbol{\varepsilon}_s]\big).$$

第二项非负，因此 $\mathrm{Cov}(s_i,\ s_j) \ge \mathbb{E}[\mathrm{Cov}(s_i,\ s_j \mid \boldsymbol{\varepsilon}_s)]$。在后验均值 $\hat{}\,\boldsymbol{\theta}$（其中包含州特定的 $\hat{}\,\boldsymbol{\varepsilon}_s$）处，代入式条件协方差是 $\mathrm{Cov}(s_i,\ s_j \mid \boldsymbol{\varepsilon}_s)$ 的一个 $\hat{}$ 估计值；因此，在 $\boldsymbol{\theta}$ 处计算的 $\mathbf{J}_{\text{cluster}}$ 将小于不含 SVC 的模型下的值。□

备注 C.32 (残余聚类)。不等式（C.12）表明，州变协变量虽能降低但无法消除 PSU 内相关。残余相关反映了五个州变协变量（贫困、

种族/族裔、城市化程度）未能捕捉的邻里层级效应：当地劳动力市场条件、市政分区规划、位于同一地区的服务提供者，以及其他具有空间结构的不可观测因素。采用聚类稳健$\mathbf{J}_{\text{cluster}}$ 的三明治校正仍有必要，以处理这一残余设计效应。

C.5. **Cholesky 仿射变换。** 三明治方差$\mathbf{V}_{\text{sand}}$ (3.25) 提供了设计一致协方差，但来自伪后验的 MCMC 抽样 $\{\boldsymbol{\theta}^{(m)}\}_{m=1}^{M}$ 的经验协方差为 $\boldsymbol{\Sigma}_{\text{MCMC}} \approx \mathbf{H}^{-1}_{\text{obs}}$，而非$\mathbf{V}_{\text{sand}}$。在频率学派框架下，只需将$\mathbf{V}_{\text{sand}}$ 报告为方差估计量；在贝叶斯框架下，我们要求后验抽样本身具有正确的协方差，以便对可信区间进行适当校准（Williams & Savitsky, 2021）。正文中的定理 3 给出了实现这一目标的仿射变换。下面给出完整证明。

定理 3 的证明。 定义中心化抽样$\mathbf{u}^{(m)} = \boldsymbol{\theta}^{(m)} - \hat{\boldsymbol{\theta}}$，使得 $\boldsymbol{\Sigma}_{\text{MCMC}} = M^{-1}\sum_{m=1}^{M} \mathbf{u}^{(m)}(\mathbf{u}^{(m)})^{\intercal}$。令

$$\mathbf{A} = \mathbf{L}_{\text{sand}}\, \mathbf{L}_{\text{MCMC}}^{-1}, \tag{C.13}$$

其中$\mathbf{L}_{\text{MCMC}} = \text{chol}(\boldsymbol{\Sigma}_{\text{MCMC}})$ 和$\mathbf{L}_{\text{sand}} = \text{chol}(\mathbf{V}_{\text{sand}})$ 是下三角 Cholesky 因子。变换（3.26）则为 $\boldsymbol{\theta}^{*(m)} = \hat{\boldsymbol{\theta}} + \mathbf{A}\mathbf{u}^{(m)}$。

(a) 均值保持。

$$\frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\theta}^{*(m)} = \hat{\boldsymbol{\theta}} + \mathbf{A}\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{u}^{(m)}\right) = \hat{\boldsymbol{\theta}} + \mathbf{A}\cdot\mathbf{0} = \hat{\boldsymbol{\theta}}.$$

(b) 协方差恢复。

$$\begin{aligned}\frac{1}{M}\sum_{m=1}^{M}(\boldsymbol{\theta}^{*(m)} - \hat{\boldsymbol{\theta}})\,(\boldsymbol{\theta}^{*(m)} - \hat{\boldsymbol{\theta}})^{\intercal} &= \mathbf{A}\,\boldsymbol{\Sigma}_{\text{MCMC}}\,\mathbf{A}^{\intercal} \\ &= \mathbf{L}_{\text{sand}}\,\mathbf{L}_{\text{MCMC}}^{-1}\,\mathbf{L}_{\text{MCMC}}\,\mathbf{L}_{\text{MCMC}}^{\intercal}\,\mathbf{L}_{\text{MCMC}}^{-\intercal}\,\mathbf{L}_{\text{sand}}^{\intercal} \\ &= \mathbf{L}_{\text{sand}}\,\mathbf{I}\,\mathbf{L}_{\text{sand}}^{\intercal} = \mathbf{V}_{\text{sand}}. \end{aligned}\tag{C.14}$$

(c) 仿射等变性。标准化抽样 $\mathbf{z}^{(m)} = \mathbf{L}^{-1}$ 马尔可夫链蒙特卡罗$\mathbf{u}^{(m)}$ sati sfy

$$M^{-1}\sum_{m}\mathbf{z}^{(m)}(\mathbf{z}^{(m)})^{\intercal} = \mathbf{I}.$$

该变换映射

$$\boldsymbol{\theta}^{*(m)} - \hat{\boldsymbol{\theta}} = \mathbf{L}_{\text{sand}}\,\mathbf{z}^{(m)},$$

其为标准化抽样的线性重缩放。Wald 置信区间——计算为 $\hat{\theta}_p \pm z_{\alpha/2}\sqrt{[\mathbf{V}_{\text{sand}}]_{pp}}$——仅取决于 $\boldsymbol{\theta}^*$ 的边际方差，因此与 $\mathbf{A}$ 的选择无关。当 $\mathbf{A}$ 为非对角矩阵时，变换后抽样的边际分位数可能与原始抽样的边际分位数不同。□

注 C.28 (分块应用)。 实践中，变换 (3.26) 是按块应用的，而非同时应用于完整的 $\approx$616维向量。以下列出三类参数；数值 Cholesky 分解程序仅对块 (i) 和 (ii) 进行运算，因为块 (iii) 不进行校正（正文中的 备注4）。其依据来自 命题 C.28 和 注 C.28：

(i) 固定效应 $\boldsymbol{\xi} = (\boldsymbol{\alpha}^\intercal, \boldsymbol{\beta}^\intercal, \log \kappa)^\intercal \in \mathbb{R}^{11}$：完整的 Cholesky 校正。设计效应比的范围为 1.14 至 4.18（均值为 2.11），相应的置信区间膨胀因子为 $\sqrt{\text{设计效应比}} \in [1.07, 2.04]$。(ii) 州随机效应 $\boldsymbol{\varepsilon}_s \in \mathbb{R}^{10}$, $s = 1, \ldots, S$：部分校正。$\boldsymbol{\varepsilon}_s$ 的后验方差混合了来自州 s服务提供者的似然信息（受调查权重影响）与向先验均值 $\boldsymbol{\Gamma}_k \mathbf{v}_s$的收缩（不受影响）。先验占主导时（小州），设计效应比 $_s \approx 1$，无需校正；似然占主导时（大州），设计效应比$_s$ 可能超过 3，校正幅度较大。(iii) 超参数 $(\boldsymbol{\Gamma}_k, \boldsymbol{\Sigma}_\varepsilon)$：不进行校正。夹心估计量不适用于根据州间变异估计的方差分量；设计效应比概念不适用于这些参数。

注释 C.34 (分块校正的近似质量)。 分块应用忽略了固定效应 $\boldsymbol{\xi}$ 与随机效应 $\boldsymbol{\varepsilon}_s$ 在 $\mathbf{V}_{\text{sand}}$中的跨块协方差。当先验占主导时（$\text{ESS}_s \ll q$），近似结果基本精确，因为 $\boldsymbol{\varepsilon}_s$ 此时几乎独立于数据。当数据占主导时（$\text{ESS}_s \gg q$），跨块协方差可能不可忽略；D 节中的敏感性比较讨论了这种情况。

注释 C.35 （MCMC 与设计一致收缩）。收缩因子$B_s^{(w)} = \boldsymbol{\Sigma}_\varepsilon (\boldsymbol{\Sigma}_\varepsilon + \mathbf{V}_s^{-1})^{-1}$ 描述的是设计一致（三明治校正后）的后验行为，而非原始 MCMC 伪后验。MCMC 采样器所感知的精度与 $W_s = \sum_{i \in s} \tilde{w}_i$成正比；当州内权重存在异质性时，这一数值会超过有效样本量 $\text{ESS}_s = W_s^2 / \sum_{i \in s} \tilde{w}_i^2$。Cholesky 仿射变换（3.26）将 MCMC 输出映射到设计一致目标，并将方差扩大DER_s （C.6 节）所示的倍数。差距 $W_s - \text{ESS}_s$ 量化了

原始伪后验相对于考虑设计的目标而言过度集中：较大的差距（在只有少数权重很大的服务提供者的州中很常见）表明，朴素 MCMC 可信区间过窄，且三明治校正尤为重要。

备注 C.36 （非高斯后验形状）。 Cholesky变换是一种仿射运算，只能保留一阶和二阶矩。如果伪后验明显呈非高斯形状（偏斜或多峰），变换后的抽样值虽具有正确的协方差，却可能无法准确反映后验形状。对于 HBB 模型，根据 Bernstein–von Mises结果（定理 C.3），给定 $N = 6,\ 809$，固定效应 $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ 预计近似服从高斯分布；对数离散度 $\log \kappa$也可由正态分布很好地近似。对于方差分量（$\boldsymbol{\Sigma}_{\varepsilon}$），后验分布可能呈右偏，但这些参数不进行 Cholesky 校正（见上文第 3 块）。

备注 C.37 （正定性要求）。该变换要求 $\boldsymbol{\Sigma}_{\mathrm{MCMC}}$ 和 $\mathbf{V}_{\mathrm{sand}}$ 均为正定矩阵，才能存在相应的 Cholesky 因子。对于 $\boldsymbol{\Sigma}_{\mathrm{MCMC}}$，当 MCMC 抽样值的数量满足 $M \gg \dim(\boldsymbol{\xi}) = 11$时，正定性几乎必然成立；在我们的应用中，$M = 4,\ 000$ 个预热后抽样值。对于 $\mathbf{V}_{\mathrm{sand}}$，正定性要求有效自由度大于参数维度：$\sum_{h=1}^{H}(C_h - 1) = 386 > 11$ (备注 C.30)，这一条件很容易满足。

C.6. **设计效应比分类与诊断**。 三明治方差$\mathbf{V}_{\mathrm{sand}}$ （3.25）和 Cholesky变换（3.26）不必应用于每个参数。正文（3.27）中定义的设计效应比（DER）为设计效应大小提供了参数特异性诊断：

$$\mathrm{DER}_p = \frac{[\mathbf{V}_{\mathrm{sand}}]_{pp}}{[\mathbf{H}_{\mathrm{obs}}^{-1}]_{pp}}. \tag{C.15}$$

当 $\mathrm{DER}_p \approx 1$时，基于模型的方差已经近似于与设计一致的方差，无需进行校正；当 DER_p 显著大于 1 时，Cholesky校正对于校准推断至关重要。

命题 C.31 (参数的设计效应比分类)。 *HBB* 模型的参数在 *Cholesky* 校正方面分为三类：

(i) 固定效应 $\boldsymbol{\xi} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top,$ 对数 $\kappa)^\top$：需要完全校正。设计效应比$_p$反映了不等权重和初级抽样单元内聚类对伪似然信息量的联合影响。实证结果中，设计效应比$_p \in [1.14, 4.18]$（均值为2.11），与 *Kish*设计效应 $= 3.79$一致。(ii) 州随机效应 $\boldsymbol{\varepsilon}_s$：部分校正。 $\boldsymbol{\varepsilon}_s$的后验方差综合了基于似然的信息量（受设计影响）以及向先验均值 $\boldsymbol{\Gamma}_k \mathbf{v}_s$ 收缩的影响（不受设计影响）。先验占主导的州（有效样本量$_s \ll q$）具有设计效应比$_s \approx 1$；数据占主导的州（有效样本量$_s \gg q$）则可能具有设计效应比$_s > 3$。(iii) 超参数 $(\boldsymbol{\Gamma}_k, \boldsymbol{\Sigma}_\varepsilon)$：无需校正。夹心估计量不适用于根据州间变异估计的方差分量；设计效应比概念不适用于这些参数。因此，超参数的后验区间是基于模型的摘要，其校准取决于层次分布是否正确设定。

证明。第 (i) 部分成立，因为$\mathbf{H}_{\text{obs}}$ 和$\mathbf{J}_{\text{cluster}}$ 的固定效应块直接取决于观测级权重 $\tilde{w}_i$ 和初级抽样单元聚合，因此 DER$_p$ 由设计效应决定。第 (ii) 部分可由后验精度分解得出：对于正态–正态共轭近似， $\boldsymbol{\varepsilon}_s$ 的后验精度是数据精度（受设计影响）与先验精度（不受设计影响）之和。当先验精度占主导时，后验方差对设计不敏感，且 DER$_s \approx 1$。第 (iii) 部分成立，因为 $\boldsymbol{\Gamma}_k$ 和 $\boldsymbol{\Sigma}_\varepsilon$ 是根据 $S = 51$ 州级随机效应的州际变异识别的。调查设计会影响州内估计，但不会改变州际方差结构；因此，夹心校正对于这些参数没有自然对应项。□

启发式方法 C.1（随机效应设计效应比的线性插值）。当收缩因子为$\lambda_{s,\ pp} = \sigma_\varepsilon^2/$（$\sigma_\varepsilon^2 + \sigma_e^2/$有效样本量$_s$）时，$\varepsilon_{s,\ p}$的设计效应比近似为

$$\text{DER}_{s,p} \ \approx \ 1 + \lambda_{s,pp}\,(\text{DEFF}_s - 1), \tag{C.16}$$

在 DER$_{s,p} \ = \ 1$（强收缩）与 DER$_{s,p} \ = \ $ DEFF$_s$（无收缩）之间进行插值。

注释 C.38（线性启发式方法的高估）。**（修正 T2-2。）**线性插值（C.16）是一种启发式方法，而非精确结果。在一维

正态—正态共轭模型中，精确的设计效应比为

$$\mathrm{DER}_{s,p}^{\mathrm{exact}} = \frac{1 + \lambda_{s,pp}\,\mathrm{DEFF}_s}{(1+\lambda_{s,pp})^2}, \tag{C.17}$$

这始终是 $\leq$ (C.16)。差异在中等收缩程度下最大（$\lambda_{s,\ pp} \approx 0.5$），此时线性启发式方法可能高估多达 50%。在实际应用中，该启发式方法仍可作为保守上界，用于快速识别需要校正的参数；精确公式（C.17）也可同时计算，作为更准确的基准。

命题 C.32 (DER诊断阈值)。 作为估计后诊断，对DER_p 每个参数进行计算，并按以下方式分类：

- $\mathrm{DER}_p \in [0.8,\ 1.2]$：设计不敏感。朴素可信区间的校准已足够；三明治校正可选。
- $\mathrm{DER}_p \in (1.2,\ 1.5]$：中度膨胀。置信区间宽度增加 $\sqrt{1.5} \approx 22\%$；建议进行校正。
- $\mathrm{DER}_p > 1.5$：设计敏感。应将三明治校正方差用作主要推断摘要。

在 *NSECE* 应用中，全部 *11* 个固定效应参数的 $\mathrm{DER}_p > 1.14$ 均处于中等至敏感范围内。

备注 C.39 （设计效应比校正下的截距可识别性）。**（校正 T2- 4。）**由于政策调节变量向量$\mathbf{v}_s$ 包含截距列（见 3.4节），固定截距 α_1 与 $\mathbf{\Gamma}_k$ 的第一列具有部分重叠的作用：州特定截距为 $\alpha_1 + \gamma_{k,\ 1,\ 1} + \varepsilon_{k,\ 1,\ s}$。$\alpha_1$ 的 DER 包含可归因于似然中的设计效应以及与$\gamma_{k,\ 1,\ 1}$的部分混杂的方差。形式上，分解

$$[\mathbf{V}_{\mathrm{sand}}]_{\alpha_1,\alpha_1} = [\mathbf{V}_{\mathrm{sand}}^{\mathrm{design}}]_{\alpha_1,\alpha_1} + [\mathbf{V}^{\mathrm{id}}]_{\alpha_1,\alpha_1} \tag{C.18}$$

并不能清晰地分离；对于截距参数，应谨慎解释 DER。对 $\mathbf{\Gamma}_k$ 截距列施加更紧的先验 $\mathsf{N}(0,\ 0。5_2)$ 所提供的（见 3.6节）弱识别通过将跨州均值归因于 α_1 并将偏差视为随机效应，缓解了这一问题。

备注 C.40 (DER 分类带来的计算节省)。 这种三分法意味着，只需对固定效应块（3.26）和最大州的随机效应块应用 Cholesky变换。这样便将 Cholesky 分解的维度从完整参数

维度（≈616）降至最多 11–20，从而大幅节省计算开销。实证 DER 值见正文（表 10）。

附录D. 扩展实证结果

本附录展示了 NSECE 2019 应用（第5节）的扩展实证结果。我们提供完整的调查权重分布（D.1 节）、遵循 Pfeffermann（1993）进行的正式信息性评估（D.2 节）、完整的州级贫困系数集（D.3 节）、完整的 40 项政策调节矩阵 $\boldsymbol{\Gamma}$（第D.4节）、详细的设计效应诊断（D.5 节）、反转概率的分级设色地图（D.6 节），以及 LKJ 先验敏感性分析（D.7节）。收敛诊断和后验预测检验可在复现包中获取，网址为 https://github.com/joonho112/hurdlebb-replication。

D.1. **调查设计特征。** NSECE 2019（全国早期照护与教育调查项目团队，2022年）采用分层整群设计，包含 $H = 30$个层和 $C = 415$个初级抽样单位（PSU），最终得到 $N = 6{,}785$名中心式服务提供者（剔除了协变量缺失的观测值后）。调查权重$w_i = 1/\pi_i$ 范围为 1 至 462 ，均值为 17。5 ，变异系数为 1。66。 图 D.1 展示了标准化权重的分布 $\tilde{w}_i = w_i/\bar{w}$，其呈显著右偏：第 99 百分位数（$\tilde{w} = 8.8$）接近中位数（$\tilde{w} = 0.47$）的十九倍，表明少数服务提供者具有不成比例的影响力。

Kish 设计效应为 $\text{DEFF}_{\text{Kish}} = 1 + \text{变异系数}_w^2 = 3.76$，使有效样本量从 $N = 6{,}785$ 降至有效样本量$_{\text{Kish}} \approx 1{,}803$。这一显著的损失说明了采用夹心方差校正和 Cholesky 校准的必要性，相关内容见第 3.5 节和C节。

D.2. **信息性评估。** 如果即使在对协变量进行条件化后，抽样单位中结果变量的分布仍与超总体分布不同，则该抽样设计具有信息性（Pfeffermann，1993）。在信息性抽样下，基于未加权似然的推断针对的是特定于样本而非总体的量，因此，伪似然加权对于一致估计至关重要（Savitsky & Toth，2016年）。我们通过三种互补的诊断方法评估 NSECE 2019 抽样设计的信息性。

（更正 T3-1。）原始的信息性检验以单个协变量（贫困率）为条件。由于完整模型为每个边际纳入 $P = 5$ 个协变量，

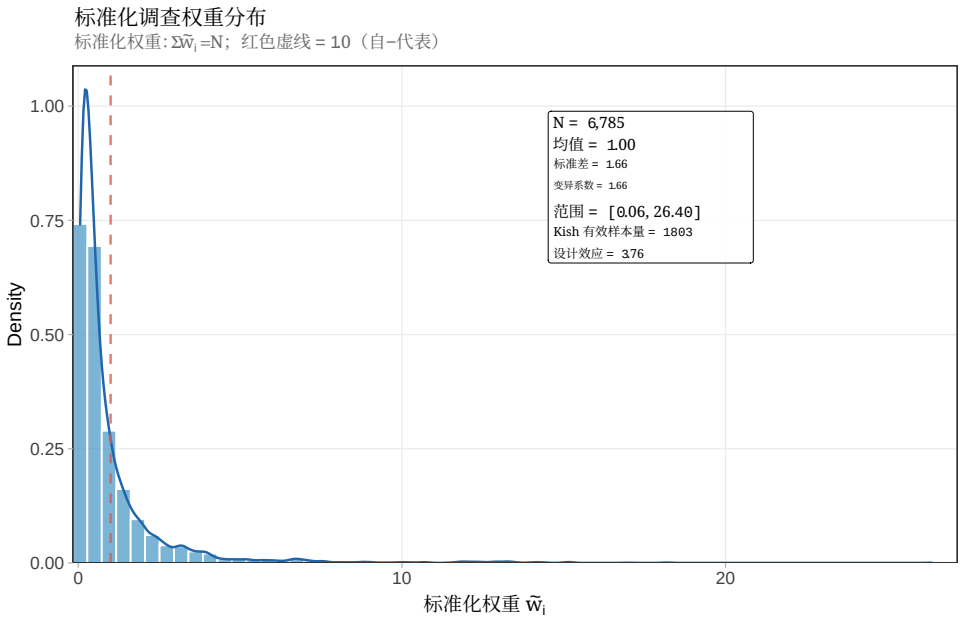

图 D.1。 $\tilde{w}_i = w_i/\bar{w}$ 至$N = 6{,}785$ 家中心式服务提供者的标准化调查权重分布。红色虚线标示 $\tilde{w} = 1.0$（自代表单位）。该分布呈明显右偏（变异系数 $= 1.66$），Kish设计效应为 $= 3.76$，有效样本量 ESS 为 $\approx 1{,}803$。

遗漏变量偏误可能会掩盖或夸大表面上的信息性。现在我们同时报告原始的单变量设定和更正后的全协变量设定。

（修正 T3-2。）原始 Hausman比较使用 lm(weights = w) 作为加权估计量，该方法仅提供异方差稳健标准误。我们现在使用 survey::svyglm() 并采用完整的分层整群设计，从而得到能够考虑 PSU 内相关性的聚类稳健标准误。

检验1：边际权重—结果相关性。 z_i（服务婴儿）与调查权重 w_i 之间的相关系数为$\mathrm{Corr}(z_i,\ w_i) = -0.041$；IT占比 Y_i/n_i（在服务机构中）与 w_i 之间的相关系数为$\mathrm{Corr}(Y_i/n_i,\ w_i) = -0.003$。二者都小得可以忽略，表明无条件信息性有限。然而，如果抽样设计对与结果变量相关的协变量进行了过度抽样，边际相关可能会掩盖条件依赖，因此有必要进行以下基于回归的检验。**检验2：基于回归的信息性检验。**根据 Pfeffermann（1993），我们在增广模型中检验 H_0: $\gamma = 0$

$$g(y_i) = \mathbf{x}_i^{\intercal}\boldsymbol{\beta} + \gamma \log(w_i) + \varepsilon_i, \tag{D.1}$$

其中，$g(\cdot)$ 是广延边际的 logit 链接函数，集约边际的单位链接函数。如果设计是无信息的，那么 log(w_i) 对 y_i 的预测能力不会超出$\mathbf{x}_i$，因此 $\gamma = 0$。

表 D.1 报告估计的$\hat{\gamma}$及其 p值，基于两种协变量设定：(i) 单变量设定，仅包含标准化贫困率；(ii) 全协变量设定，包含全部四个社区层面预测变量（贫困、城市化程度、黑人比例、西班牙裔比例）。

表 D.1。基于回归的信息性检验（Pfeffermann，1993）。增广模型（D.1）将$\log(w_i)$ 加入协变量向量；$\hat{\gamma}$ 是$\log(w_i)$ 的系数，其双侧 p值为。 **（更正 T3-1：）**全协变量设定将广延边际的结论从不显著反转为显著。

设定	边际	$\hat{\gamma}$	p值
单变量（仅贫困）	广延	0.020	0.369
	密集型	0.009	0.002
完整协变量（T3-1）	广延	0.086	< 0.001
	密集型	0.006	0.050

表 表 D.1 值得评论的两个特点。首先，在单变量设定下，广延边际似乎没有信息（$p = 0.369$），而集约边际则明显具有信息（$p = 0.002$）。其次，加入其余三个协变量后，广延边际的结论发生反γ 转，$\hat{}$增加了四倍（从0.020到0.086），且变得高度显著（$p < 0.001$）。这一反转源于单变量模型将权重与结果变量的直接关系，与通过被遗漏协变量（城市化程度、种族构成）产生的间接路径混杂在一起。纳入这些混杂因素后，log(w_i) 与参与之间的剩余关联变得可检测。两种设定下，集约边际均具有信息性，不过随着完整模型吸收了部分权重与结果变量之间的关联，系数略有减弱（从0.009到 0.006）。

在完整设定下，两个边际均具有信息性，这一发现进一步支持在贝叶斯分层模型中采用伪似然加权，以及采用C节中提出的夹心方差校正。

检验3：Hausman型比较。作为一项补充诊断，我们比较每个边际的未加权与加权系数估计值。在非信息性抽样下，两种估计量对同一总体参数均具有一致性，但加权估计量效率较低；较大的差异表明该设计扭曲了有效的协变量—结果关系。

(修正 T3-2：) 加权估计量使用 survey::svyglm()，并采用完整的分层整群设计（$H = 30$ 层，$C = 415$ 初级抽样单位），从而得到聚类稳健标准误。每个参数的 Hausman 检验统计量为

$$Z_k = \frac{\hat{\beta}_k^{\mathrm{WT}} - \hat{\beta}_k^{\mathrm{UW}}}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_k^{\mathrm{WT}}) - \widehat{\mathrm{Var}}(\hat{\beta}_k^{\mathrm{UW}})}}, \tag{D.2}$$

在 H_0（无信息性）下，该统计量渐近服从标准正态分布；分母中的方差差是有效的，因为在无信息性条件下，加权估计量的方差不小于未加权估计量的方差。

表 D.2。未加权估计值与设计加权系数估计值的 Hausman 型比较。**(修正 T3-2：)** 加权估计量使用 `survey::svyglm`，并采用聚类稳健标准误。 Z是逐参数 Hausman 统计量（D.2）。星号： $*$ 表示 $p < 0.10$， $**$ 表示 $p < 0.05$。

边际	参数	$\hat{\beta}^{\mathrm{UW}}$	$\hat{\beta}^{\mathrm{WT}}$	Z	p-值	
广延	截距	0.614	0.662	0.82	0.41	
	贫困	−0.158	−0.199	−0.64	0.52	
	城市	0.176	0.234	2.02	0.044	**
	黑人	−0.044	0.086	1.89	0.059	*
	西班牙裔	−0.043	0.059	1.51	0.13	
密集型	全部 $p > 0.24$；均不显著					

表 D.2 揭示了一种细致的模式。在广延边际上，贫困系数本身在不同估计量之间没有显著差异（$Z = -0.64$， $p = 0.52$），但城市化程度系数出现了显著变化（$\hat{}\,\beta^{\mathrm{UW}} = 0.176$ 与$\hat{}\,\beta^{\mathrm{WT}} = 0.234$；$Z = 2.02$， $p = 0.044$），而黑人构成系数则仅达到边际显著（$Z = 1.89$， $p = 0.059$）。这表明，NSECE 对高需求地区的过度抽样主要扭曲了城市化程度对 IT 参与的估计效应——这与抽样框明确针对城市和高贫困初级抽样单位这一事实相符。在集约边际上，没有任何单个参数显示出显著的

Hausman差异（全部 $p > 0.24$），这表明检验2检测到的信息性是分散的，而非集中于某一个协变量。
概要。这三项诊断得出了相互一致的结论。边际相关（检验1）可以忽略，但基于回归的检验（检验2）表明，在控制完整协变量向量后，两个边际都存在统计显著的信息性——这一发现被原始单变量设定所掩盖。Hausman比较（检验3）显示，广延边际的信息性来自城市化程度和种族构成协变量，而非贫困本身。综合来看，这些结果支持主分析中的两项设计选择（第5节）：（i）通过伪似然纳入调查权重；（ii）应用夹心方差校正（第 3.5 节）和 Cholesky仿射校准（C节），以确保后验可信区间反映真实的设计调整后不确定性。

D.3. **州级贫困系数**。本小节总结了所有 $S = 51$个州的模型M2（不含政策调节变量的块对角 SVC）州级贫困系数。广延边际系数$\alpha_{\text{pov},\ s}$ 以及集约边际系数$\beta_{\text{pov},\ s}$ 分别表示社区贫困增加一个标准差对 IT 参与对数优势和 IT 强度 logit 的州特定影响。图5 中的跨边际散点图展示了这些估计值的联合分布。

在51个州中，48个（94%）的后验均值符号模式与ˆ经典反转一致（ˆ$\alpha_{\text{pov},\ s} < 0$ 和$\beta_{\text{pov},s} > 0$）；这种近乎普遍的模式反映了层次收缩将小州估计值拉向负的总体平均广延边际系数。其中，15个（29%）的后验反转概率超过0.80，表明除点估计符号之外还有强有力的证据，另有21个州的反转概率介于0.60与0.80之间。

反转证据最强的州。反转概率最高的五个大样本州分别是德克萨斯州（$N = 578$, Pr $= 0.989$）、伊利诺伊州（N =471，Pr $= 0.970$）、加利福尼亚州（$N = 1,$ 110，Pr $= 0.937$）、路易斯安那州（$N = 77$，Pr =0.923）和新泽西州（$N = 201$，Pr $= 0.908$）。在这些州中，广延边际系数（$\alpha_{\text{pov},s} \in [-0.34, -0.13]$）明确为负，其90%可信区间不含零，而集约边际系数（$\beta_{\text{pov},s} \in [0.07, 0.08]$）为正，区间勉强包含零。

值得注意的例外：纽约州。纽约州（$N = 558$）是最突出的例外，其广延边际系数为正（$\alpha_{\text{pov}} = 0.101$；90% 置信区间 $[-0.051, 0.261]$，Pr(反转) $= 0.143$），这表明较高的贫困程度与

服务婴幼儿的可能性更高相关——这可能反映了纽约州对儿童早期项目的大量公共投资。亚利桑那州（Pr $= 0.40$）、明尼苏达州（Pr $= 0.44$）和北达科他州（Pr $= 0.48$）也呈现出微弱或不存在的反转模式。

小州收缩。小州（例如缅因州、$N = 18$；密西西比州、$N = 17$；蒙大拿州、$N = 22$）的可信区间更宽，反转概率也更接近全国平均值（≈ 0.73），这反映了预期中的向层级均值收缩，这一点已在C节中说明。完整的 51 州系数表可在重复模拟资料包中获取，网址为https://github.com/joonho112/hurdlebb-replication。

D.4. **完整的政策调节参数**。表 D.3列出了加权 M3b 模型中跨层调节矩阵的全部 40 个元素，其中包括 $\mathbf{\Gamma}_1$（广延边际）和 $\mathbf{\Gamma}_2$（集约边际）。正文（5.5节）总结了三种主要模式；此处我们给出全部5 个协变量 $\times 4$ 个政策维度 $\times$ 2 边际的完整矩阵。

四个政策维度分别为：(i) 市场费率百分位数（MR_百分位数_标准化），用于衡量各州的报销费率相对于服务提供者市场费率的慷慨程度；(ii) 分层报销，即针对采用分层报销质量体系的州的指示变量；(iii) 信息技术附加项，即针对为婴幼儿照护提供额外补贴的州的指示变量；以及 (iv) 截距列，用于捕捉与所测量的三项政策无关的州际剩余变异（ $\mathbf{\Gamma}_k$的截距列由其先验进行弱识别；参见3.6节）。

除正文中强调的模式外，还有两种模式值得关注。首先，IT附加补贴变量在两个边际上都表现出最一致的影响：它与广延边际截距呈正相关（$\gamma = 0.728$，Pr$(> 0) = 0.996$），与广延边际西班牙裔系数也呈正相关（$\gamma = 0.750$，Pr$(> 0) = 0.999$），而与广延边际贫困系数则呈强负相关（$\gamma = -0.488$，Pr$(> 0) = 0.018$）。在集约边际上，IT附加补贴变量与显著较高的黑人系数相关（$\gamma = 0.256$，Pr$(> 0) = 1.000$）。

两个边际的截距列后验分布都很宽（SD ≈ 0.93），这进一步证实了弱识别：层级先验（截距列采用 N(0，0。5_2)，其约束比政策列采用的 N$(0, 1^2)$ 更强）对这些参数形成约束，但不会将它们推至零。

D.5. **设计效应诊断**。设计效应比（DER）在（3.27）中定义，并在命题 C.31中进行分类，它量化了调查设计使

边际	协变量	政策	Mean	SD	Q025	Q975	Pr(>0)
广延边际 ($\mathbf{\Gamma}_1$)							
ext	截距	截距	0.127	0.937	−1.684	1.960	0.558
ext	截距	MR_百分位数	0.018	0.142	−0.268	0.297	0.543
ext	截距	分层报销	−0.413	0.359	−1.121	0.278	0.128
ext	截距	信息技术附加项	0.728	0.285	0.182	1.297	0.996
ext	贫困	截距	−0.044	0.933	−1.845	1.789	0.483
ext	贫困	MR_百分位数	0.064	0.107	−0.152	0.271	0.732
ext	贫困	分层报销	0.495	0.272	−0.064	1.021	0.962
ext	贫困	信息技术附加项	−0.488	0.225	−0.917	−0.035	0.018
ext	城市	截距	0.073	0.945	−1.766	1.886	0.531
ext	城市	MR_百分位数	0.011	0.062	−0.112	0.129	0.574
ext	城市	分层报销	−0.248	0.175	−0.601	0.094	0.076
ext	城市	信息技术附加项	−0.205	0.146	−0.488	0.087	0.072
ext	黑人	截距	0.085	0.929	−1.740	1.927	0.530
ext	黑人	MR_百分位数	−0.283	0.144	−0.565	0.010	0.026
ext	黑人	分层报销	−0.652	0.339	−1.346	0.014	0.029
ext	黑人	信息技术附加项	0.365	0.276	−0.193	0.890	0.910
ext	西班牙裔	截距	0.012	0.925	−1.799	1.850	0.509
ext	西班牙裔	MR_百分位数	0.072	0.134	−0.198	0.331	0.712
ext	西班牙裔	分层报销	−0.351	0.350	−1.052	0.318	0.157
ext	西班牙裔	信息技术附加项	0.750	0.259	0.260	1.290	0.999
集约边际 ($\mathbf{\Gamma}_2$)							
int	截距	截距	−0.074	0.922	−1.899	1.735	0.468
int	截距	MR_百分位数	−0.024	0.039	−0.104	0.052	0.265
int	截距	分层报销	0.271	0.120	0.026	0.497	0.985
int	截距	信息技术附加项	0.001	0.084	−0.158	0.173	0.488
int	贫困	截距	0.019	0.935	−1.783	1.831	0.505
int	贫困	MR_百分位数	0.028	0.026	−0.024	0.079	0.865
int	贫困	分层报销	−0.064	0.064	−0.193	0.060	0.156
int	贫困	信息技术附加项	−0.087	0.051	−0.187	0.014	0.041
int	城市	截距	−0.015	0.932	−1.821	1.783	0.497
int	城市	MR_百分位数	0.027	0.017	−0.007	0.060	0.941
int	城市	分层报销	0.015	0.054	−0.090	0.122	0.612
int	城市	信息技术附加项	0.039	0.040	−0.035	0.120	0.837
int	黑人	截距	0.029	0.912	−1.731	1.848	0.503
int	黑人	MR_百分位数	0.015	0.036	−0.056	0.082	0.662
int	黑人	分层报销	0.047	0.088	−0.126	0.224	0.706
int	黑人	信息技术附加项	0.256	0.072	0.117	0.397	1.000
int	西班牙裔	截距	−0.045	0.936	−1.853	1.811	0.482
int	西班牙裔	MR_百分位数	−0.115	0.054	−0.224	−0.010	0.016
int	西班牙裔	分层报销	0.227	0.164	−0.088	0.554	0.920
int	西班牙裔	信息技术附加项	−0.135	0.108	−0.349	0.077	0.104

表 D.3. 完整的政策调节参数（加权 M3b，所有 40 个 $\mathbf{\Gamma}_1$ 和 $\mathbf{\Gamma}_2$）。每行报告后验均值、标准差、95% 可信区间以及效应为正的后验概率。正文（5.5节）总结了三种关键模式；加粗结果对应于 $\Pr(>0) > 0.95$ 或 $\Pr(>0) < 0.05$。

将每个参数的方差膨胀至超出基于模型的（SRS）方差。表 D.4呈现了全部 11 个固定效应参数的完整方差分解，而图 D.2提供了由三个面板组成的诊断可视化。

方差分解。表 D.4中的三个方差量值得仔细解释。$[\mathbf{H}^{-1}{}_{\text{obs}}]_{pp}$ 是从观测信息矩阵中获得的基于模型的方差——在简单随机抽样下的适当基准。 $[\mathbf{V}_{\text{sand}}]_{pp}$ 是三明治校正方差（Binder，1983年），该方差考虑了复杂抽样设计，并构成主要的推断摘要。 $[\mathbf{\Sigma}_{\text{MCMC}}]_{pp}$是原始 MCMC 后验方差；相对于$\mathbf{H}^{-1}{}_{\text{obs}}$，它大幅膨胀（膨胀倍数为 600–6,500），这是因为层级先验对固定效应的影响超过了伪似然贡献。

DER 的范围为 1.14（β_{Hisp}）至 4.18（α_{Urban}），均值为 2.11。这与 Kish设计效应 $= 3.76$大体一致：DER 系统性地较小，因为州变化系数部分吸收了州内聚类（命题 C.30）。相应的置信区间膨胀因子为$\sqrt{\text{DER}} \in [1.07, 2.04]$，因此最宽的Wald区间约为假设的 SRS 设计下区间宽度的两倍。

诊断图。图 D.2将三明治校正分解为三个组成部分。面板 (a) 显示层级方差比——朴素（基于 MCMC）置信区间宽度与 Wald（基于三明治估计量）置信区间宽度之比——并以对数尺度呈现；所有参数的比率都大于 10，其中若干超过 100，这证实原始 MCMC 后验分布无法提供有关设计一致的不确定性的信息。面板 (b) 比较基于 $\mathbf{H}^{-1}{}_{\text{obs}}$ 的设计效应比（DER）与基于 $_{\text{MCMC}}$ 的“DER $\mathbf{\Sigma}_{\text{MCMC}}$”；除 $\log \kappa$ 外，后者对所有参数都接近于零，进一步说明了分层膨胀。面板 (c) 报告了固定效应块的 Cholesky 缩放矩阵$\mathbf{A}$ (定理 3) 的对角元素。对于 α 和 β 系数，所有对角元素都低于 $0。06$，表明 MCMC 抽样值明显收缩，这是分层膨胀的后验分布远宽于设计一致目标的直接结果。

D.6. **反转概率图**。 图 D.3 展示了各州发生贫困反转的后验概率：$\Pr(\tilde{\alpha}_{\text{pov},\ s} < 0$ 和 $\beta_{\text{pov},\ s} > 0 \mid$ 数据), 该概率由未加权 M2 模型（块对角 SVC ）计算得出。反转概率较高（$\Pr > 0.9$）的州集中在美国南部、美国西南部和大型城市州；概率较低（$\Pr < 0.1$）的州主要位于中西部北部和山区

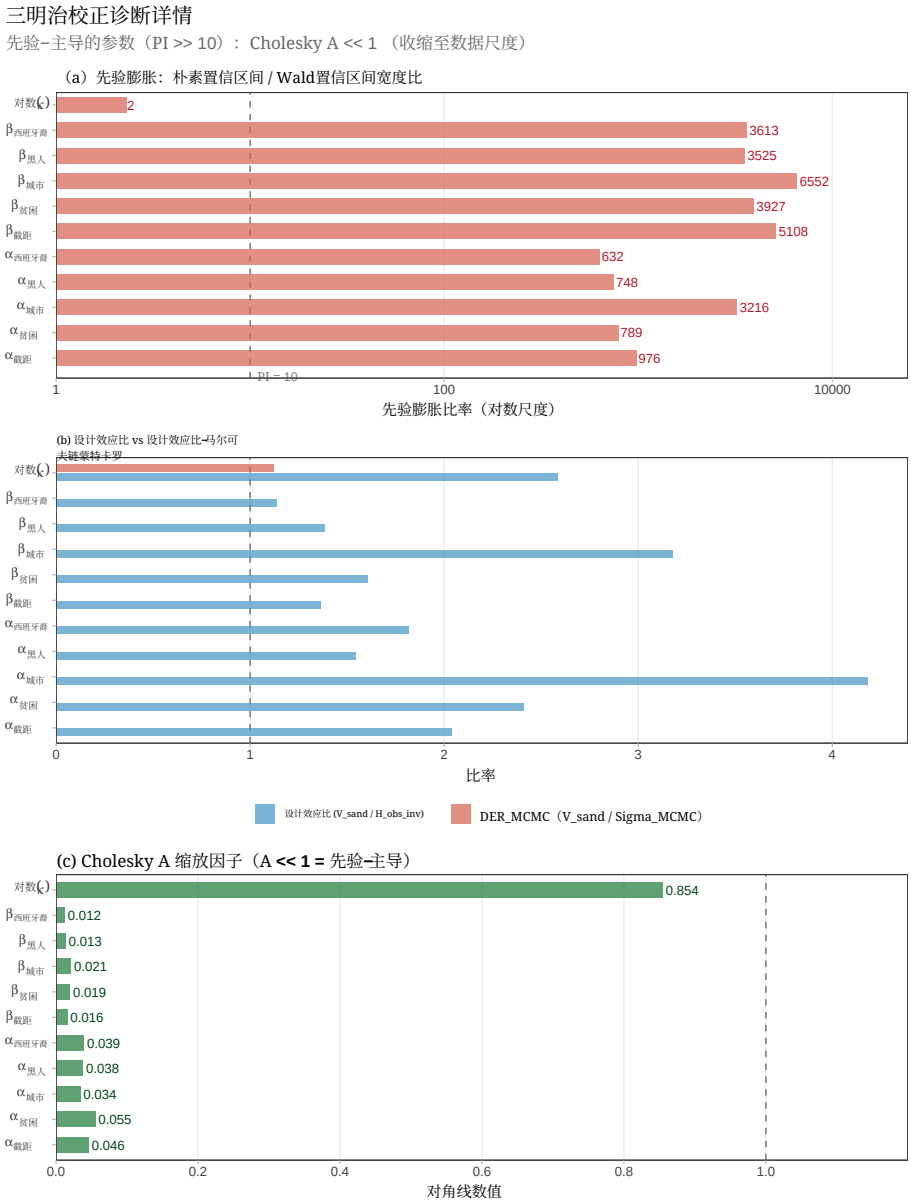

图 D.2。 三明治校正诊断详情。（a）分层方差比（对数尺度）：朴素方法（基于 MCMC）与 Wald 方法（基于夹心估计量）的置信区间宽度之比。高于虚线 10 的数值表示显著的分层膨胀。（b）DER 比较：使用 $\mathbf{H}^{-1}{}_{\text{obs}}$ （主要）与 $\boldsymbol{\Sigma}_{\text{MCMC}}$ （仅用于诊断）计算得到的设计效应比。（c）Cholesky A 对角线：仿射变换（定理 3）中的缩放因子。数值 $\ll 1$ 表明，该变换会将 MCMC 抽样值显著收缩至与设计一致的目标。

参数	H_{obs}^{-1}	V_{sand}	Σ_{MCMC}	设计效应比 设计效应比$_{\text{mc}}$		标准差(H) 标准差(三明治估计量) 标准差(MCMC)		
α_{Int}	0.00097	0.00198	0.946	2.04	0.002	0.031	0.044	0.972
α_{Pov}	0.00114	0.00275	0.900	2.41	0.003	0.034	0.052	0.949
α_{Urb}	0.00028	0.00118	0.907	4.18	0.001	0.017	0.034	0.952
α_{Blk}	0.00127	0.00195	0.947	1.54	0.002	0.036	0.044	0.973
α_{Hisp}	0.00146	0.00265	0.920	1.82	0.003	0.038	0.051	0.959
β_{Int}	0.00017	0.00023	0.860	1.36	0.000	0.013	0.015	0.927
β_{Pov}	0.00022	0.00036	0.875	1.61	0.000	0.015	0.019	0.936
β_{Urb}	0.00013	0.00042	0.873	3.18	0.000	0.012	0.021	0.934
β_{Blk}	0.00024	0.00033	0.845	1.38	0.000	0.016	0.018	0.919
β_{Hisp}	0.00025	0.00028	0.892	1.14	0.000	0.016	0.017	0.945
log κ	0.00033	0.00086	0.001	2.58	1.119	0.018	0.029	0.028

表 D.4。全部 11 个固定效应参数的设计效应比诊断。各列依次为：H^{-1} 观测 = 基于模型的方差（观测信息逆矩阵）、V_{sand} = 三明治方差、Σ_{MCMC} = 原始 MCMC 后验方差、观测设计效应比= V_{sand}/H^{-1} 观测、设计效应比$_{\text{mc}}$ = $V_{\text{sand}}/\Sigma_{\text{MCMC}}$以及相应的标准差。设计效应比范围为 1.14–4.18（均值为 2.11），与 Kish设计效应 = 3.76一致。

西部。其余州（$0.1 \leq$ Pr ≤ 0.9）往往是较小的州，由于样本量有限，后验区间较宽。

D.7. **LKJ 先验敏感性**。为评估对 $10 \times$10 随机效应相关矩阵 Ω 的 LKJ 先验的敏感性，我们采用七种集中参数 $\eta \in \{1.0, 1.5, 2.0, 3.0, 4.0, 6.0, 8.0\}$重新拟合 M3b-W。基准设定使用 η =2，该设定采用弱正则化分布，倾向于中等相关性；$\eta = 1$ 则在相关矩阵上给出均匀分布，而较大的取值会使概率质量集中在单位矩阵附近。

表 D.5 报告了跨边际相关性 ρ_{cross} =Ω$_{1,6,}$、贫困系数 (α_{pov}, β_{pov}) 和截距随机效应标准差 (τ_1^{ext}, τ_{int1}) 的后验摘要。在 η 的七倍范围内，$\hat{\rho}_{\text{cross}}$介于 0.097 和 0.191 之间，始终接近零，且完全处于任一单独设定的后验不确定性范围内。所有情况下，95% 可信区间均包含零。贫困系数同样稳定：$\alpha_{\text{pov}} \in [-0.359, -0.310]$ 和 $\beta_{\text{pov}} \in [0.074, 0.132]$，证实符号反转对 LKJ 先验具有稳健性。

各设定下的留一交叉验证模型拟合几乎完全相同（ELPD 跨度为 < 5），所有 MCMC 诊断结果均令人满意（唯一一个边际 R-hat 超标值$\eta = 6$ 为 1.011，完全处于实际容许范围内）。

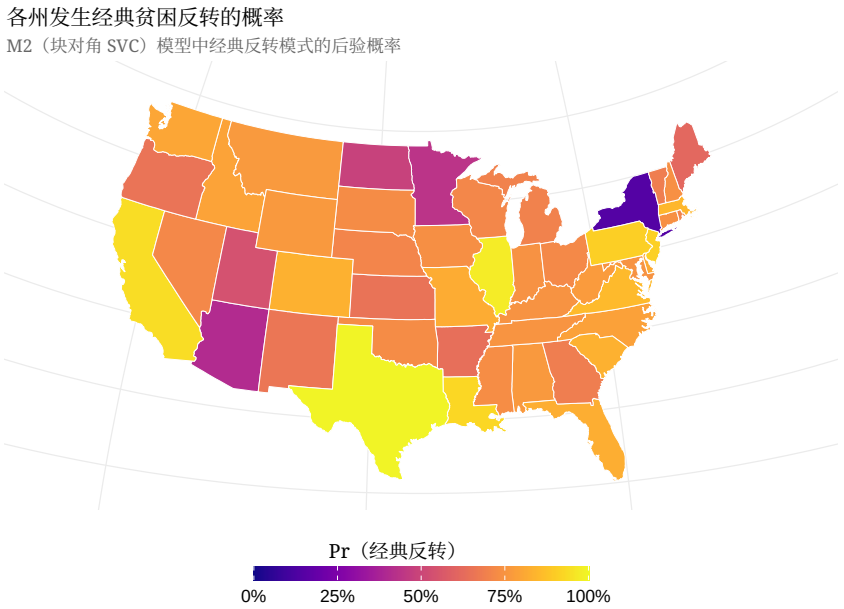

图 D.3. 未加权 M2 模型（分块对角 SVC）的后验反转概率分级统计地图：$\Pr(\tilde{\alpha}_{\text{pov},\,s} < 0 \;\cap\; \beta_{\text{pov},\,s} > 0 \mid \text{数据})$。颜色越深表示反转概率越高。总体平均反转概率为 1.000。

图 D.4叠加展示了所有七种设定下 ρ_{cross} 的后验密度。这些密度几乎完全重合，证实尽管簇大小适中，数据对于该参数的影响仍超过先验（$S = 51$）。

这些结果证实，接近零的跨边际相关性和贫困反转模式是由数据驱动的发现，而非 LKJ(2) 先验设定的产物。

附录E. 模拟与计算细节

本附录提供计算与模拟细节，用以补充第3 和4节。E.1节给出五模型层级结构及其参数数量。E.2节描述MCMC配置与运行时间基准。E.3 节规定模拟研究的数据生成过程，E.4节给出补充模拟诊断图，E.5节评估模型错设下的稳健性，E.6节提供频率学派情境化，第E.7节将覆盖率差距分解为偏差与宽度分量，并且E.8节报告95%名义水平下的覆盖率。实现细节（非中心化参数化、数值稳定性、得分

表 D.5. LKJ 先验敏感性分析。每行报告使用 LKJ(η)先验拟合 M3b-W 后得到的后验汇总，先验作用于 10×10随机效应相关矩阵 Ω。基准设定为 $\eta = 2$（相关系数服从均匀分布）。 $\rho_{\text{cross}} = \Omega_{1,6}$ 表示广延截距与集约边际截距之间的跨边际相关系数。 α_{pov} 和 β_{pov} 分别表示广延边际和集约边际的贫困系数。 τ_1^{ext}和τ_1^{int}表示截距随机效应标准差。所有协变量均已标准化。

η	$\hat{\rho}_{\text{cross}}$ [95% 置信区间]	α_{pov}	β_{pov}	τ_1^{ext}	τ_1^{int}	$\widehat{\text{elpd}}_{\text{loo}}$	对角线。
1.0	0.191 $[-0.261,\ 0.604]$	-0.322	0.088	0.693	0.149	-20913.0	✓
1.5	0.172 $[-0.261,\ 0.572]$	-0.312	0.074	0.683	0.146	-20915.0	✓
2.0⋆	0.170 $[-0.259,\ 0.573]$	-0.320	0.098	0.682	0.144	-20915.7	✓
3.0	0.145 $[-0.269,\ 0.527]$	-0.310	0.132	0.666	0.143	-20917.2	✓
4.0	0.134 $[-0.274,\ 0.516]$	-0.330	0.093	0.653	0.143	-20916.2	✓
6.0	0.111 $[-0.247,\ 0.460]$	-0.359	0.113	0.646	0.142	-20914.8	!
8.0	0.097 $[-0.256,\ 0.429]$	-0.340	0.110	0.639	0.140	-20915.4	✓

⋆基准设定e 设定（$\eta = 2$ 对应于相关矩阵上的均匀分布）

计算）记录于hurdlebb 软件包指南；完整模拟结果与复现说明可在复现包中获取，网址为https://github.com/joonho112/hurdlebb-replication。

E.1. **模型层级与架构。** 表 E.1 总结了在应用中拟合的五个嵌套模型（第5节）。每个模型均在前一模型的基础上增加一项结构特征：随机截距（M1）、具有块对角协方差的州变化系数（M2）、跨边际协方差（M3a）以及政策调节变量（M3b）。最终模型M3b-W将调查权重纳入伪后验。参数计数包括所有固定效应、随机效应和超参数（Cholesky因子、尺度和相关系数）。

这些模型采用热启动初始化依次拟合：模型 k 的后验均值为模型 $k+1$提供初始值。较简单模型中不存在的随机效应参数初始化为零（NCP 变量）或单位矩阵（Cholesky 因子）。与冷启动相比，这种渐进式策略可将总计算时间缩短约 40%，并确保每个模型都针对前一个模型提出的问题进行分析：M0 到 M1 探讨基准入学率是否因州而异；M1 到 M2 探讨贫困反转是否因州而异；M2 到 M3a 探讨跨边际协方差是否能改善拟合；M3a 到 M3b 探讨可观测政策是否能够解释州际异质性。

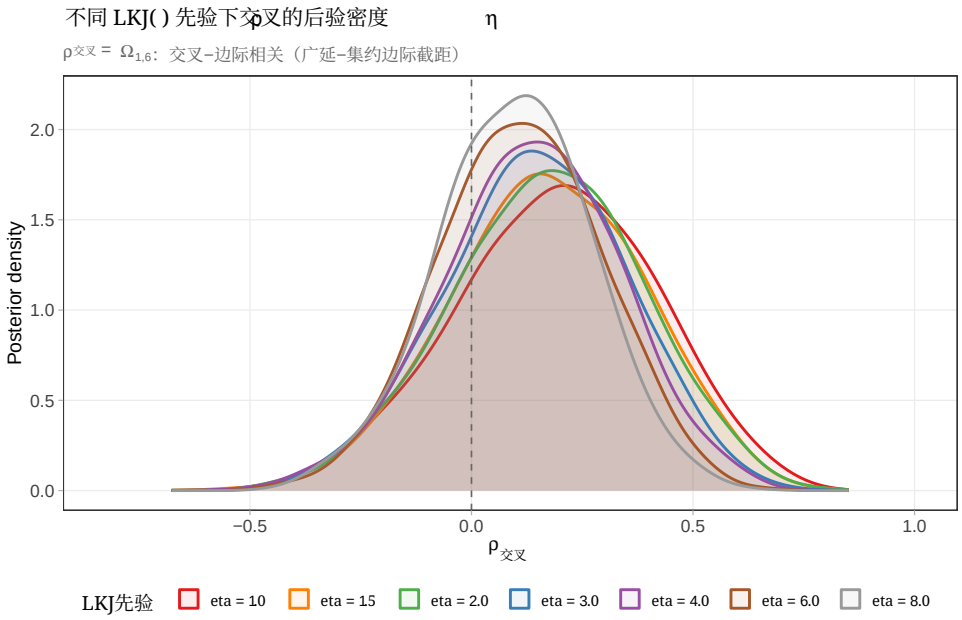

图 D.4. 在七种 LKJ 浓度参数 $\eta \in \{1, 1.5, 2, 3, 4, 6, 8\}$ 下，跨边际相关性 $\rho_{\text{cross}} = \Omega_{1,6}$ 的后验密度。这些密度几乎完全重合，表明后验分布主要由数据决定，而非先验驱动。垂直虚线表示零。

模型结构		参数	协方差	$\widehat{\text{elpd}}_{\text{loo}}$
M0	合并 HBB	11	—	−20,945
M1	随机截距	116	2×2	−20, 627
M2	块对角 SVC	551	$(5 \times 5) \oplus (5 \times 5)$	−20, 604
M3a	完整 SVC	576	10×10	−20,602
M3b	政策调节变量	616	10×10	−20, 608

表 E.1. 模型层级。参数：自由参数总数（固定效应 + 随机效应 + 超参数）。协方差：随机效应偏差的状态层级协方差矩阵维度 $\boldsymbol{\delta}_s$。LOO：预期对数预测密度（第5节）。从 M0 到 M3a 的逐步演进显著提升了预测准确度；M3b 加入了政策调节变量，但未进一步提升预测能力。

E.2. **MCMC 配置。** 所有模型均使用 Stan（Carpenter 等人，2017年）通过无转向采样器（NUTS；Hoffman & Gelman，2014）并经由 Cmd-StanR 接口进行拟合（Stan 开发团队，2024）。 表 E.2报告了采样器配置。

设置	实证（M0–M3b）模拟	
链	4（并行）	4（并行）
预热迭代	每条链 1,000 次	每条链 1,500 次
采样迭代	每条链 1,000 次	每条链 2,000 次
预热后抽样总数	4,000	8,000
adapt_delta	0.95	0.95
max_treedepth	12	14

表 E.2. MCMC 配置。实证案例研究（M0–M3b）使用默认的 hurdlebb 设置；模拟研究采用更长的运行和更高的 max_treedepth，以确保在 200 次重复模拟中进行充分探索。在这两种情况下，adapt_delta $= 0.95$ 会减小步长，以避免状态变化系数模型中的发散转移。

收敛目标。 如算法 1所述，我们要求所有参数满足以下诊断ˆ阈值（Vehtari 等人，2021）：(i) split-$R < 1.01$；(ii) 最小主体有效样本量 > 400；(iii) 最小尾部有效样本量 > 400；以及 (iv) 零发散转移。全部五个模型均满足这些阈值；详细诊断可在重复模拟包中获得。

热启动初始化。 五模型序列 M0 $\rightarrow$ M1 $\rightarrow$ M2 $\rightarrow$ M3a$\rightarrow$ M3b 逐步拟合：模型 k 的后验均值用于初始化模型 $k+1$。具体而言：

- 固定效应系数（$\boldsymbol{\alpha}$、 $\boldsymbol{\beta}$、$\log \kappa$）初始化为前一模型的后验均值。

- 当维度匹配时，尺度参数（$\boldsymbol{\tau}$）初始化为前一模型的后验均值；引入新维度时，则初始化为 0.1。
- Cholesky因子（$\mathbf{L}_{\varepsilon}$）初始化为单位矩阵，这对应于零相关——一种保守的起点，使采样器能够在预热期间发现相关性。

- NCP 变量（$\mathbf{z}_s$）初始化为零，使所有州均位于总体均值处。
- 政策调节系数（$\boldsymbol{\Gamma}$，仅限 M3b）初始化为前一模型的后验均值；对于新引入的参数，则初始化为零。

该策略避免了复杂模型因初始化值远离后验众数而可能出现的缓慢探索。

运行时间基准。 在一台 16 核工作站上，估算的实际运行时间为：M0（$\sim$5 分钟）、M1（$\sim$15 分钟）、M2（$\sim$30分钟）、M3a（$\sim$50 分钟）、

M3b（∼60 分钟）和 M3b-W（∼70 分钟，其中包括生成量代码块中的得分提取）。采用热启动初始化时，总顺序拟合时间约为 4 小时，而采用冷启动时约为 7 小时。

E.3. **数据生成过程与校准**。 本节提供第4节模拟研究的完整数据生成过程（DGP）。第4节。该DGP旨在匹配NSECE 2019年数据的结构，同时允许对调查信息性进行参数化控制。

总体生成。 一个包含 $M = 50{,}000$个服务提供者的有限超总体分布在 $S = 51$个州，各州规模与观测到的NSECE各州计数成比例。对于州 s中的每个服务提供者 i：

（1）通过在州 s内从NSECE经验边际分布进行自助重抽样，抽取协变量向量 $\mathbf{x}_i = (1,\ x_{\text{pov}},\ x_{\text{urb}},\ x_{\text{blk}},\ x_{\text{his}})^\intercal$ 。（2）从NSECE 数据中0至5岁入学总人数的经验分布中抽取试验规模 n_i 。

真实参数值。 真实参数根据M1（随机截距）未加权NSECE拟合进行校准（附录E.1）。表E.3报告了DGP中使用的所有参数值。这些值经过选择，以产生符合实际情况的结构零比例（约36%）和与经验数据一致的过度离散水平。

结果变量生成。 对于每次重复模拟 $r = 1, \ldots, R$:

(1) 抽取州随机效应 $\boldsymbol{\delta}_s = (\delta_{1s},\ \delta_{2s})^\intercal \sim \mathrm{N}_2(\mathbf{0},\ \boldsymbol{\Sigma}_\delta)$，其中

$$\boldsymbol{\Sigma}_\delta = \operatorname{diag}(\boldsymbol{\tau})\, \mathbf{R}_\delta\, \operatorname{diag}(\boldsymbol{\tau}), \quad \boldsymbol{\tau} = (\tau_{\text{ext}},\ \tau_{\text{int}})^\intercal, \quad [\mathbf{R}_\delta]_{12} = \rho_{\text{cross}}.$$

(2) 计算线性预测量： $\eta_i^{\text{ext}} = \mathbf{x}_i^\intercal \boldsymbol{\alpha}_0 + \delta_{1,s[i]}$, $\eta_i^{\text{int}} = \mathbf{x}_i^\intercal \boldsymbol{\beta}_0 + \delta_{2,s[i]}$。

(3) 生成参与变量： $z_i \sim$伯努利分布(q_i)，其中 $q_i =$逆逻辑函数(η_i^{ext})。

(4) 对于具有 $z_i = 1$ 的服务提供者，通过 **拒绝抽样**从无条件贝塔-二项分布（Prentice，1986）中生成正计数：

抽取 $Y_i \sim$ 贝塔-二项分布$(n_i,\ \mu_i,\ \kappa_0)$，其中 $\mu_i =$ 逆逻辑函数(η_i^{int})；重复，直到 $Y_i > 0$；设定 $y_i = Y_i$。

这确保了对所有具有 $z_i = 1$ 的服务提供者均有 $y_i > 0$ ，与障碍模型中的零截断贝塔-二项分量相一致。接受概率为 $1 - p_0(n_i, \mu_i, \kappa_0)$，在我们的校准中，该概率对绝大多数观测值超过 0.80。

（5）对于具有 $z_i = 0$ 的服务提供者，设定 $y_i = 0$。

参数	数值	描述
α_{int}	$+0.696$	广延边际截距
α_{pov}	-0.119	广延边际贫困
α_{urb}	$+0.253$	广延边际城市化
α_{blk}	-0.070	广延边际黑人
α_{his}	-0.139	广延边际西班牙裔
β_{int}	-0.032	集约边际截距
β_{pov}	$+0.057$	集约边际贫困
β_{urb}	-0.018	集约边际城市化
β_{blk}	$+0.080$	集约边际黑人
β_{his}	$+0.040$	集约边际西班牙裔
对数 κ_0	$+1.655$	对数集中参数
κ_0	5.232	集中参数
τ_{ext}	0.577	状态标准差（广延）
τ_{int}	0.208	状态标准差（集约）
ρ_{cross}	0.285	跨边际相关性

表 E.3。模拟数据生成过程的真实参数值，根据 M1非加权拟合进行校准。固定效应采用 logit尺度； κ_0 =exp(log κ_0)。

注释 E.41 （T3-3：零截断贝塔-二项分布数据生成过程）。 从无条件贝塔-二项分布中抽样并拒绝零值，即可从零截断贝塔-二项分布 ZT-BetaBin(n, μ, κ) 中精确抽样。另一种方法——将无条件概率质量函数除以 $(1 - p_0)$——需要自定义抽样程序；而拒绝抽样可重复使用标准贝塔-二项分布抽样器，并且只要 p_0 不太接近 1，就能保持高效。Stan 的生成量代码块也采用相同方法进行后验预测复制。

抽样设计。 从每个有限总体中，利用规模偏倚入样概率抽取一个约 $N \approx 7{,}000$ 名服务提供者的分层整群样本$\log \pi_i = c_0 + \rho_{\text{inc}} \cdot y_i^*$(方程（4.2）)，其中 y_i^*是标准化潜在结果，且 c_0 根据目标样本量进行校准（参见 Savitsky & Toth，2016年）。三种情景（S0、S3、S4）定义于表 5；完整讨论见第4节 。调查权重计算为 $w_i = 1/\pi_i$ ，并进行归一化，使其总和为 N ，用于伪后验分布。

E.4. **补充诊断图和设计效应分析**。本节展示了三幅补充图，直观呈现所有模拟情景中的偏差、区间宽度和

设计效应，以及涵盖全部 11 个固定效应参数的设计效应比（DER）完整表格。

相对偏差分布。图 E.1 展示了各参数和情景下的相对偏差分布。E-UW 的绝对偏差最大（尤其是在 S4 下的 τ_{ext} 和 τ_{int}），而 E-WT/E-WS 通过纳入调查权重，改善了固定效应的点估计。对于贫困斜率 α_{pov} 和 β_{pov}，E-WS 在所有情景下的相对偏差均低于 15%。

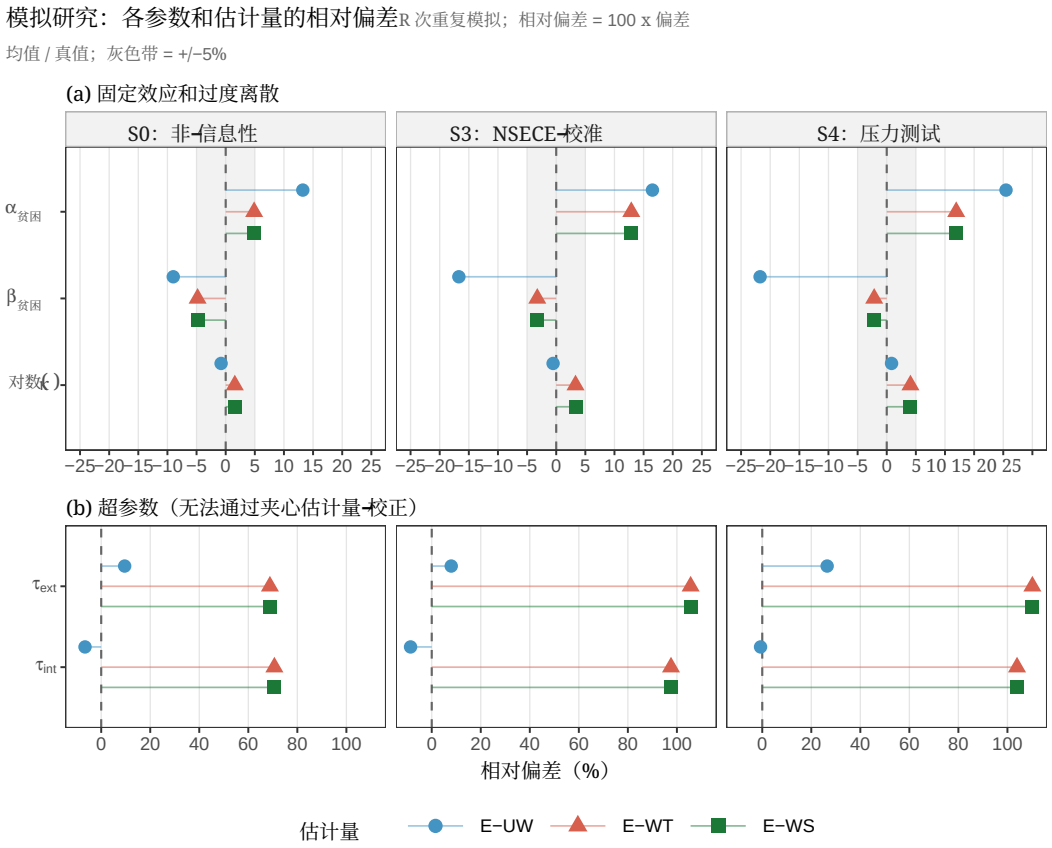

图 E.1. 按参数、估计量和情景划分的相对偏差（RB，%）。每个面板对应五个目标参数之一。点的位置表示相对偏差；颜色区分不同情景（S0、S3、S4）。竖直灰线表示零偏差。

宽度比值分布。 图 E.2 展示了宽度比值（WR = 重复模拟中 E-WS 下的置信区间宽度中位数 / E-WT 下的置信区间宽度中位数）的分布。对于所有固定效应参数，从 S0 到 S4 的渐进膨胀都清晰可见。对于 $\log \kappa$ 的宽度比与贫困斜率的宽度比相近或略大，

这反映了非线性离散度得分（附录 B）。超参数宽度比均为 1。000（未显示），因为未应用三明治校正。

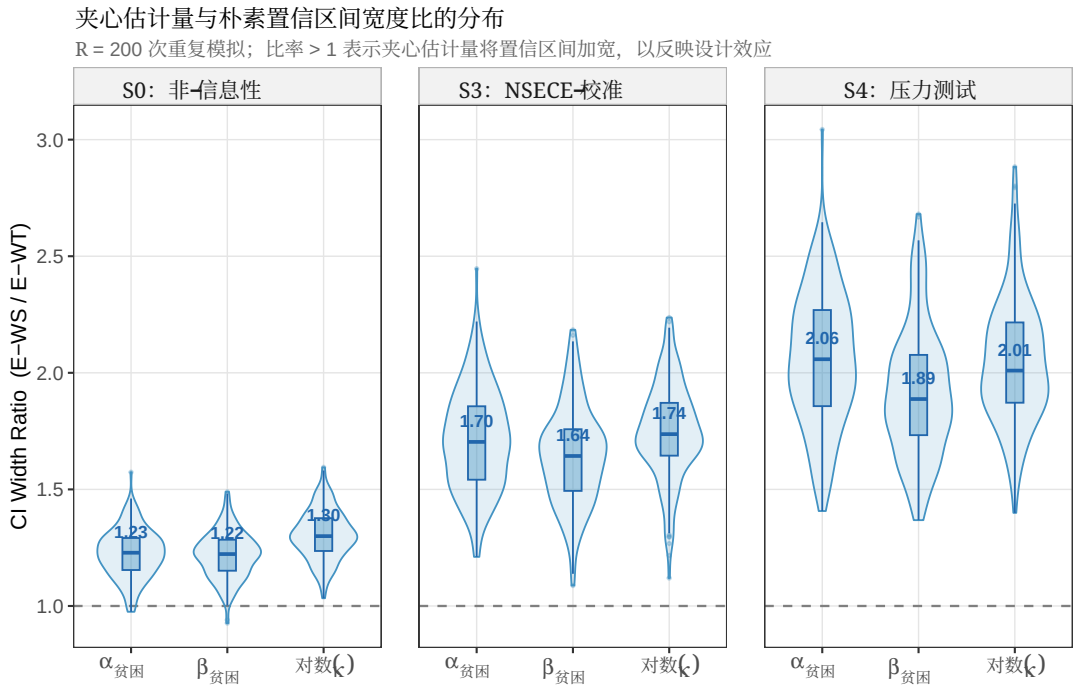

图 E.2. 按参数和情景展示跨 $R = 200$ 次重复模拟的宽度比分布。箱线图上叠加了小提琴图，以展示完整的分布形态。WR$=$ 1.0 处的水平虚线表示无校正基准。

设计效应比。 图 E.3 汇总了所有 11 个固定效应参数（5 个广延参数、5 个密集型参数和 $\log \kappa$）在三个情景下的模拟设计效应比。设计效应比随信息性参数 ρ_{inc}单调增加，这与命题 C.31中的理论预测一致。在所有情景中，城市密度参数均表现出最大的设计效应比（从 S0 中的均值 3.25 增至 S4 中的 6.43），这与 α_{urb} 在 NSECE 案例研究中的实证设计效应比 4.18 相呼应（表 D.4）。

完整设计效应比表。下表报告了所有 11 个固定效应参数在三个情景下，设计效应比（DER）在各次重复模拟中的完整分布。均值、中位数、标准差和范围（最小值–最大值）反映了重复模拟之间的变异性。值得注意的是两种模式：（i）设计效应比的标准差随情景严重程度增加而增大，表明在设计效应更大的情况下，抽样变异性也更高；（ii）城市参数的

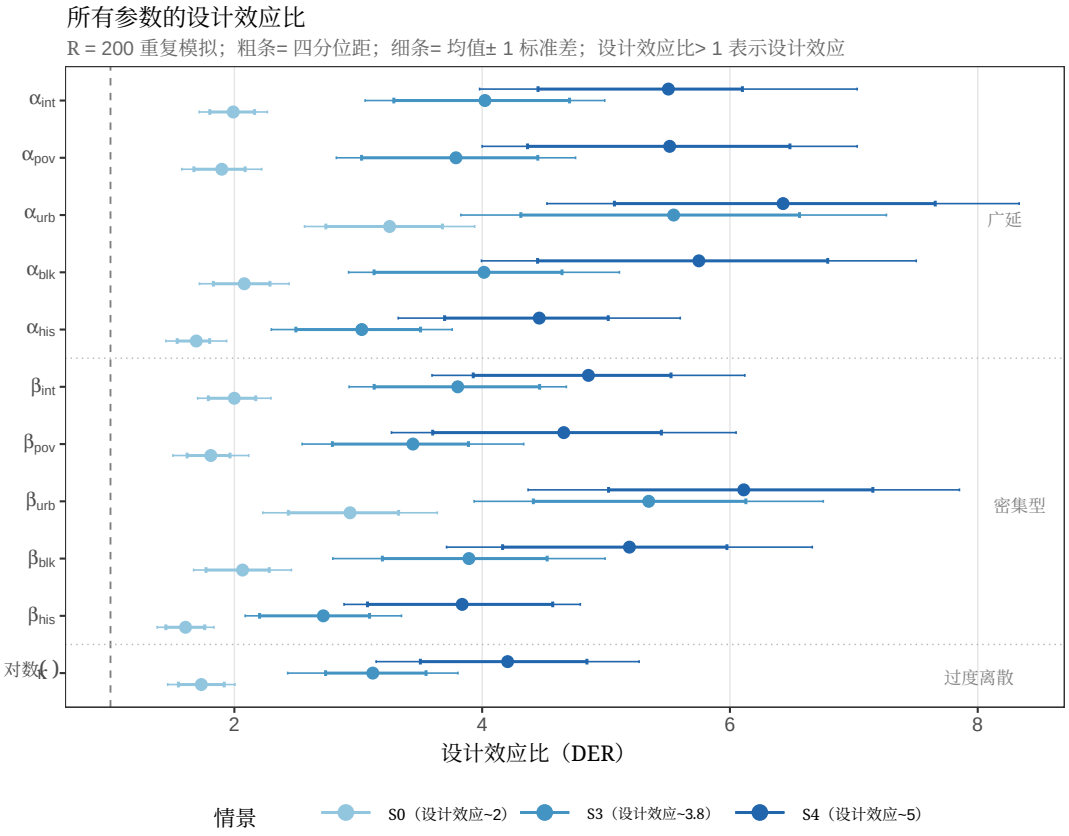

图 E.3。按参数和情景划分的设计效应比（设计效应比）。每个点表示200次重复模拟中设计效应比的均值；颜色用于区分情景。参数按边际（广延与密集型）分组，并将$\log \kappa$ 作为单独类别。

设计效应比范围始终最宽，这反映了城市服务提供者的空间集中会放大设计效应。

跨边际相关性恢复。表 E.5 报告了使用基于各 $\rho_{\text{cross}} = \text{cor}(\delta_j^{\text{ext}}, \delta_j^{\text{int}})$ 州后验均值随机效应的插件估计量所恢复的跨边际相关性。未加权估计量 E-UW 系统性地高估 ρ_{cross}，相对偏差从 S0 的 $+84\%$ 增至 $+127$S4 中的%。加权估计量 E-WT 显示出中等程度的向下偏差（-26%至-45%），而 E-WS 与 E-WT 完全相同，因为 Cholesky 三明治校正（定理 3）仅调整固定效应协方差。由于模拟存档中未保留相关矩阵 Ω 的完整后验抽样，因此无法获得逐次重复模拟的可信区间；这里只报告各次重复模拟的频率学派汇总结果。

情景	参数	R	均值	中位数	标准差	最小值	第25百分位数	第75百分位数	最大值
S0：无信息	截距（扩展）	200	1.99	1.95	0.28	1.29	1.80	2.16	2.92
S0：无信息	贫困（扩展）	200	1.90	1.89	0.32	1.23	1.67	2.09	3.07
S0：无信息	城市（扩展）	200	3.25	3.18	0.69	1.59	2.74	3.68	5.28
S0：无信息	黑人（扩展）	200	2.08	2.05	0.36	1.14	1.83	2.29	3.46
S0：无信息	西班牙裔（扩展）	200	1.69	1.67	0.24	1.05	1.54	1.80	2.68
S0：无信息	截距（截距）	200	2.00	1.97	0.30	1.38	1.79	2.17	2.92
S0：无信息	贫困（截距）	200	1.81	1.80	0.31	1.06	1.62	1.96	3.11
S0：无信息	城市（截距）	200	2.93	2.86	0.70	1.41	2.44	3.33	5.35
S0：无信息	黑人（截距）	200	2.07	2.02	0.39	1.32	1.77	2.28	3.55
S0：无信息	西班牙裔（截距）	200	1.61	1.59	0.23	1.11	1.45	1.76	2.34
S0：无信息	对数 κ	200	1.73	1.71	0.27	1.12	1.55	1.92	2.36
S3：经NSECE校准	截距（扩展）	200	4.02	3.88	0.97	1.97	3.29	4.71	7.47
S3：经NSECE校准	贫困（扩展）	200	3.79	3.72	0.97	1.90	3.03	4.45	7.53
S3：经NSECE校准	城市（扩展）	200	5.55	5.43	1.72	1.62	4.31	6.56	11.21
S3：经NSECE校准	黑人（扩展）	200	4.01	3.81	1.09	2.07	3.13	4.64	7.23
S3：经NSECE校准	西班牙裔（扩展）	200	3.03	2.94	0.73	1.58	2.50	3.50	5.26
S3：经NSECE校准	截距（截距）	200	3.80	3.70	0.88	1.68	3.13	4.46	6.40
S3：经NSECE校准	贫困（截距）	200	3.44	3.42	0.89	1.53	2.79	3.89	7.32
S3：经NSECE校准	城市（截距）	200	5.34	5.26	1.41	2.34	4.41	6.13	11.40
S3：经NSECE校准	黑人（截距）	200	3.89	3.67	1.10	1.96	3.19	4.52	7.45
S3：经NSECE校准	西班牙裔（截距）	200	2.72	2.65	0.63	1.67	2.20	3.09	5.16
S3：经NSECE校准 logkappa		200	3.12	3.07	0.69	1.38	2.74	3.55	5.44
S4：压力测试	截距（扩展）	200	5.50	5.23	1.52	2.79	4.45	6.10	13.23
S4：压力测试	贫困（扩展）	200	5.51	5.31	1.51	2.72	4.37	6.48	11.59
S4：压力测试	城市（扩展）	200	6.43	6.22	1.91	2.97	5.07	7.66	17.11
S4：压力测试	黑人（扩展）	200	5.75	5.41	1.75	2.30	4.45	6.79	11.38
S4：压力测试	西班牙裔（扩展）	200	4.46	4.36	1.14	1.91	3.70	5.02	10.86
S4：压力测试	截距（截距）	200	4.86	4.71	1.26	2.38	3.93	5.52	9.53
S4：压力测试	贫困（截距）	200	4.66	4.44	1.39	2.31	3.60	5.45	9.04
S4：压力测试	城市（截距）	200	6.11	5.96	1.74	2.24	5.02	7.15	13.79
S4：压力测试	黑人（截距）	200	5.19	5.01	1.48	2.13	4.16	5.98	10.13
S4：压力测试	西班牙裔（截距）	200	3.84	3.74	0.95	1.78	3.07	4.57	6.44
S4：压力测试	对数 κ	200	4.21	4.10	1.06	2.11	3.50	4.85	7.69

表 E.4. 各模拟重复中的设计效应比（DER）。DER $=$ diag(V_{sand}) / diag(H^{-1} 观测值)。显示全部 11 个固定效应参数。

E.5. **模型设定错误稳健性**。第4节中的模拟研究评估了正确设定情形：数据由随机截距模型 M1 生成，并使用 M1 拟合。一个自然的问题是，三种估计量（E-UW、

表 E.5. 跨边际相关性 ρ_{cross} =cor(δ_j^{ext}, δ_j^{int}) 在 $R = 200$ 次模拟重复中的恢复情况。插件估计量 $\hat{\rho} = \text{cor}(\bar{\delta}_j^{\text{ext}}, \bar{\delta}_j^{\text{int}})$ 使用后验均值随机效应（$S = 51$ 个州）。真实值：$\rho_{\text{cross}} = 0.2853$。$\text{E}_{\text{WS}}$ 与 E_{WT} 相同，因为三明治校正仅适用于固定效应。

情景	估计量	$\bar{\hat{\rho}}$	Bias	相对偏差（%）	RMSE	经验标准误
S0（基准）	E_{UW}	0.5238	+0.2386	+83.6	0.2990	0.1807
	E_{WT}	0.2104	−0.0749	−26.3	0.2294	0.2173
	$\text{E}_{\text{WS}}{}^{\dagger}$	0.2104	−0.0749	−26.3	0.2294	0.2173
S3（NSECE）	E_{UW}	0.5814	+0.2961	+103.8	0.3643	0.2128
	E_{WT}	0.1658	−0.1195	−41.9	0.2256	0.1918
	$\text{E}_{\text{WS}}{}^{\dagger}$	0.1658	−0.1195	−41.9	0.2256	0.1918
S4（重尾）	E_{UW}	0.6468	+0.3615	+126.7	0.4075	0.1886
	E_{WT}	0.1579	−0.1274	−44.7	0.2372	0.2005
	$\text{E}_{\text{WS}}{}^{\dagger}$	0.1579	−0.1274	−44.7	0.2372	0.2005

$^{\dagger}\text{E}_{\text{WS}}$ 与 E_{WT}使用相同的 Stan 拟合结果；Cholesky 三明治校正仅调整固定效应协方差，不调整超参数后验分布。

当拟合模型遗漏了相关的异质性来源时，E-WT、E-WS) 的表现如何。因此，我们在模拟中加入一个模型设定错误情景，以检验模型错误时三明治校正是否仍然有益。

M2 数据生成过程。 通过在 M1 线性预测量中加入随机贫困斜率，定义州别系数数据生成过程（M2）：

$$\eta_i^{\text{ext}} = \mathbf{x}_i^{\intercal}\boldsymbol{\alpha} + \delta_{s[i]}^{\text{ext}} + \gamma_{s[i]}^{\text{ext}}\, x_{\text{pov},i}, \tag{E.1}$$

$$\eta_i^{\text{int}} = \mathbf{x}_i^{\intercal}\boldsymbol{\beta} + \delta_{s[i]}^{\text{int}} + \gamma_{s[i]}^{\text{int}}\, x_{\text{pov},i}, \tag{E.2}$$

其中，$\gamma_s^{\text{ext}} \sim \text{N}(0,\, 0.152)$，且在各个 $s = 1$（.....）51 个州之间相互独立。所有其他参数（$\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、κ、$\boldsymbol{\tau}$、ρ）以及总体结构（$M = 50$，000 个服务提供者）均与 M1 数据生成过程（E.3 节）相同。标准差 0.15（广延）和 0.10（密集型）所产生的州别贫困效应，相对于固定贫困系数（$\alpha_{\text{pov}} = -0.119$、$\beta_{\text{pov}} = +0.057$）而言处于中等水平。

设计。 我们在 S3（NSECE 校准的）信息性情景下进行 $R = 200$ 次独立重复模拟，设定 DEFF 为 ≈ 3.79，并使用全部三种估计量拟合 M1 随机截距模型。M1 拟合忽略了随州变化的贫困斜率 γ_s，从而导致一种遗漏变量模型设定错误，并通过贫困系数估计值产生影响。

结果。 表 E.6 比较了正确设定的 S3 情景（M1 真值、M1 拟合）与错误设定的 B5 情景（M2 真值、M1 拟合）下的覆盖率、相对偏差、均方根误差和宽度比。

表 E.6。模型设定错误稳健性：经正确设定的 S3 与模型设定错误的 B5（M2 真值，M1 拟合）在经 NSECE 校准的抽样下（设计效应 ≈ 3.79）。 $R = 200$ 次重复模拟；名义覆盖率为 $= 90\%$； † 表示显著偏离（$> 2\times$ 蒙特卡洛标准误）。蒙特卡洛标准误 $= \sqrt{\hat{p}(1-\hat{p})/200}$；范围为 0.0–3.5 个百分点。

		覆盖率（%）		相对偏差（%）			RMSE			WR	
参数估计		S3	B5	S3	B5	Δ	S3	B5	Δ	S3	B5
固定效应											
α_{pov}	E-UW	89.5	0.5†	+16.5	−117.8	−134.3	0.037	0.143	+0.106	—	—
	E-WT	58.5†	52.0†	+12.9	−37.6	−50.5	0.072	0.079	+0.007	—	—
	E-WS	82.0†	79.0†	+12.9	−37.6	−50.5	0.072	0.079	+0.007	1.71	1.67
β_{pov}	E-UW	89.0	30.0†	−16.7	+63.1	+79.8	0.017	0.039	+0.022	—	—
	E-WT	62.5†	42.0†	−3.2	+52.7	+55.9	0.032	0.043	+0.011	—	—
	E-WS	82.5†	67.5†	−3.2	+52.7	+55.9	0.032	0.043	+0.011	1.64	1.60
对数 κ	E-UW	92.0	82.5†	−0.5	−1.3	−0.8	0.020	0.027	+0.007	—	—
	E-WT	34.0†	36.5†	+3.3	+2.9	−0.4	0.068	0.064	−0.004	—	—
	E-WS	60.5†	64.0†	+3.3	+2.9	−0.4	0.068	0.064	−0.004	1.75	1.74
方差分量											
τ_{ext}	E-UW	98.0†	96.0†	+8.0	−2.5	−10.5	0.091	0.117	+0.026	—	—
	E-WT	0.5†	0.0†	+105.6	+102.4	−3.2	0.639	0.625	−0.014	—	—
	E-WS	0.5†	0.0†	+105.6	+102.4	−3.2	0.639	0.625	−0.014	1.00	1.00
τ_{int}	E-UW	97.5†	57.0†	−8.6	−34.8	−26.2	0.040	0.079	+0.039	—	—
	E-WT	0.0†	1.0†	+97.6	+97.0	−0.6	0.211	0.213	+0.002	—	—
	E-WS	0.0†	1.0†	+97.6	+97.0	−0.6	0.211	0.213	+0.002	1.00	1.00

图 E.4 以图形方式展示了相应的覆盖率。

通过比较得出了三项发现。

发现 B5-1：未加权估计量对贫困系数设定错误极为敏感。在正确设定（S3）下，E-UW 对两个贫困斜率的覆盖率均接近名义水平（89.5% 和 89.0%）。在 M2数据生成过程下，E-UW 对 α_{pov} 的覆盖率骤降至 0.5%，对 β_{pov}的覆盖率降至 30.0%——分别下降 89 和 59 个百分点。其机制是遗漏变量偏差：由于 M1 未纳入随州变化的贫困斜率 γ_s，固定系数 α_{pov}和 β_{pov} 吸收了被遗漏的异质性，导致相对偏差分别达到 -117.8%和$+63.1\%$。这证实，E-UW 的基于模型的区间只有在

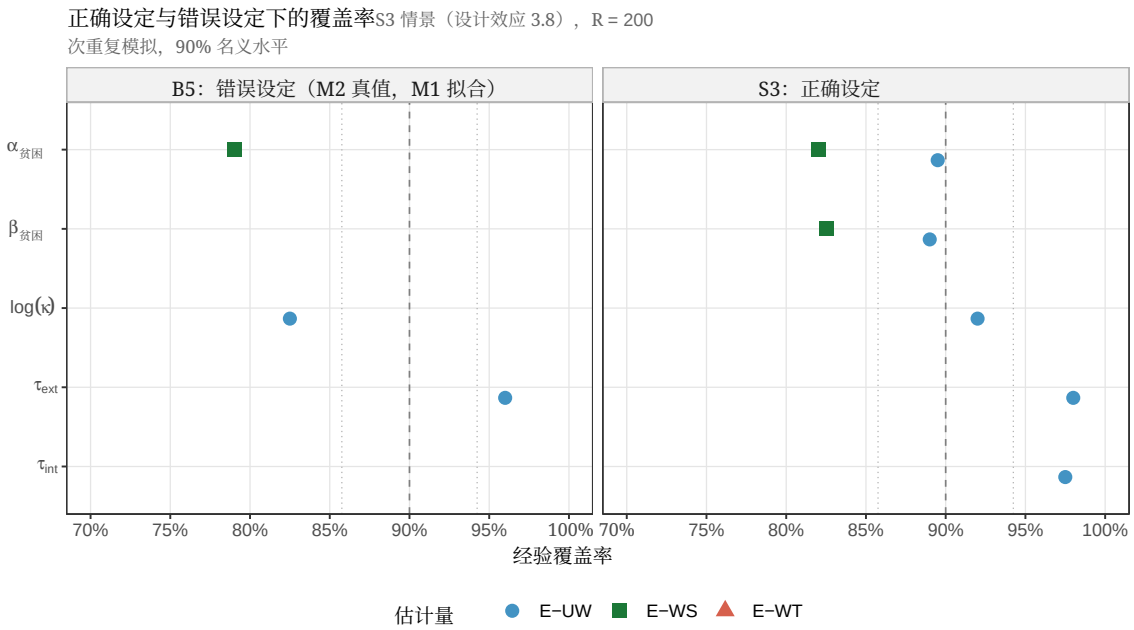

图 E.4. 正确设定（S3）与模型设定错误（B5）下的覆盖率。垂直虚线：90% 名义水平；点线： ±2蒙特卡洛标准误区间。

正确设定时才有效——这有力地说明了调查加权所提供的基于设计的保护为何重要。

发现 B5-2：三明治校正可在模型设定错误时保持覆盖率稳定性。对于 α_{pov}，E-WS 的覆盖率仅从 82.0%（S3）降至 79.0%（B5）——下降 3 个百分点；而 E-WT 则从 58.5% 降至 52.0%。对于 β_{pov}，E-WS 从 82.5% 降至 67.5%（-15 个百分点），相比之下，E-WT 从 62.5% 降至 42.0%（-20.5个百分点）。在这两种情况下，E-WS 都明显优于 E-WT（B5 情景下分别高出 27 和 25.5 个百分点），这表明三明治方差校正吸收了模型设定错误所导致的方差增加中的相当一部分。不同情景下的宽度比保持稳定（B5 下 WR $= 1.67$–1.74，而 S3 下为 1.64–1.75），说明无论模型是否正确设定，设计效应大小都相近。

发现 B5-3：方差分量表现出不同的敏感性。过度离散对数 κ 在 E-UW 下相对稳健（覆盖率为 92.0% $\rightarrow$ 82.5%；相对偏差为 -0.5% $\rightarrow$ -1.3%），这符合预期，因为被遗漏的随机斜率不会直接影响离散结构。在方差分量中，τ_{ext} 在 E-UW 下仍具有良好的覆盖率（98.0% $\rightarrow$ 96.0%），但 τ_{int} 则大幅下降（97.5% $\rightarrow$ 57.0%）。

这种不对称性源于：与 $\tau_{\text{int}} = 0.208$相比，被遗漏的集约边际斜率（$\sigma_{\gamma}^{\text{int}} = 0.10$ ）带来的未建模州际方差比例更高，而广延边际斜率（相对于 $\tau_{\text{ext}} = 0.577$的 $\sigma_{\gamma}^{\text{ext}} = 0.15$ ）则不然。与正确设定情形一样，E-WT 和 E-WS 对这两个τ 参数的覆盖率都接近于零（WR $= 1.00$），这再次说明三明治校正从结构上并不适用于方差分量。

综合来看，这些结果进一步支持了第4节所建议的双轨报告策略：对于总体平均固定效应，经夹心估计量校正的估计量 E-WS 对模型误设具有实质性稳健性；对于层级方差分量，未加权估计量 E-UW 仍是唯一可行的选择，但需注意其有效性取决于随机效应结构的正确设定。

E.6.**频率学派情境化**。 一个自然的问题是，更简单的基于设计的分析能否取代完整的贝叶斯分层模型。我们将 HBB 伪后验估计与从 `survey::svyglm` （Lumley, 2004年，2010年）得到的常规调查加权频率学派估计进行比较，以（i）验证伪后验固定效应的设计一致性，并（ii）阐明哪些模型特征无法由频率学派方法加以容纳。

频率学派设定。 我们采用与 HBB 模型相同的五协变量设定，分别拟合边际模型。对于广延边际，我们使用 svyglm(..., family = 拟二项()) 拟合调查加权逻辑回归，共$N = 6{,}785$ 个观测值。对于集约边际，我们对 IT入学份额 y_i/n_i 关于相同协变量拟合调查加权线性回归，并限定于$N^{+} = 4{,}392$ 名参与者。两个模型均采用 NSECE 复杂抽样设计（分层、聚类和抽样权重），并给出基于线性化的聚类稳健标准误。随机效应、beta-二项结构或跨边际依赖均无法纳入。

结果。 表E.7 比较了调查加权 svyglm 估计值与 M3b-W 三明治校正伪后验摘要。

对于广延边际，两种方法都采用 logit 链接函数，因此可以在共同尺度上进行有意义的比较。五个预测变量的系数符号和相对大小均一致。HBB 的点估计绝对值系统性地大于 svyglm 的估计值，这符合预期：HBB 模型以州级随机效应为条件，而混合模型中的条件效应已知会

表E.7. 频率学派情境化：调查加权 svyglm 与贝叶斯 HBB（M3b-W，三明治校正）的点估计和标准误。两种方法的广延边际均采用逻辑斯蒂链接函数；集约边际则分别采用恒等连接函数（svyglm）和 logit 链接函数（HBB），因此尺度不同。标准误比 =HBB 夹心标准误 / svyglm 标准误。

		svyglm（基于设计的）		HBB（三明治校正）			
边际	参数	Est.	SE	Est.	SE	Δ估计值	标准误比
广延	截距	0.662	0.063	0.764	0.044	+0.103	0.70
	贫困	-0.199	0.071	-0.324	0.052	-0.125	0.74
	城市	0.234	0.039	0.442	0.034	+0.208	0.89
	黑人	0.086	0.075	0.478	0.044	+0.392	0.59
	西班牙裔	0.059	0.074	0.006	0.051	-0.054	0.70
密集型	截距	0.483	0.005	-0.242	0.015	-0.725	3.02
	贫困	0.015	0.007	0.090	0.019	+0.075	2.77
	城市	-0.009	0.004	-0.047	0.021	-0.038	4.63
	黑人	0.034	0.008	-0.017	0.018	-0.050	2.30
	西班牙裔	0.010	0.006	-0.097	0.017	-0.107	2.92

大于其边际对应项（Zeger等人，1988）。标准误比（HBB 三明治估计 / svyglm）范围为 0。59 至0。89，这反映了通过随机效应对州际异质性建模，而非将其纳入残差所实现的方差缩减。

对于集约边际，两种方法的比较不那么直接，因为它们采用的尺度不同：svyglm 在单位矩阵连接函数下对入托比例 y_i/n_i 进行回归，而 HBB 模型在 logit 链接函数下对贝塔-二项分布的均值参数 μ_i 进行建模。截距不同（svyglm：$+0.483$；HBB：-0.242）反映的是尺度差异，而非实质性分歧，因为 $\text{logit}^{-1}(-0.242) \approx 0.44$，处于观测到的平均份额范围内。集约边际较大的标准误比（2.30–4.63）也源于 logit 尺度参数化，它相较于概率尺度会放大方差。

结构性局限。 这一比较凸显了 svyglm 无法处理的 HBB框架的若干特点：

(1)贝塔-二项分布： svyglm 通过高斯误差对连续份额 y/n 进行建模，忽略了结果变量的有界计数性质以及超二项过度离散（$\kappa = 6.81$）。(2) *Hurdle*结构：两个边际必须分别进行拟合，从而无法对零值生成过程进行联合推断。

(3) 州级随机效应： svyglm 是一个固定效应估计量；若要纳入$S = 51$ 州截距（以及 M3b 中的斜率），则需要使用固定效应虚拟变量——这会增大 SE 并牺牲收缩性——或者采用无法传播不确定性的两步法。(4)跨边际相关性： 二元随机效应结构 $\boldsymbol{\delta}_s \sim \mathrm{N}_{2q}(\mathbf{0}, \boldsymbol{\Sigma}_\delta)$ 无法在分边际回归中得到复现。(5) 政策调节： 随州变化的政策系数 $\boldsymbol{\Gamma}$以一种没有自然频率学派对应方法的方式，将层级成分与基于设计的成分进行交互。

总的来说，svyglm 比较可作为固定效应层面的健全性检验：贝叶斯伪后验固定效应与频率学派目标设计一致，而完整的 HBB框架还提供了层级结构、分布灵活性和联合推断，这些都是频率学派替代方法无法实现的。

E.7. **覆盖率差距分解**。 表 E.8对 S3（经 NSECE 校准）情景下每个参数—估计量组合的覆盖率差距进行了分解。以下两个诊断量可区分由宽度驱动和由偏差驱动的不足：

- 标准误校准比 $= \mathrm{SD}/\widehat{\mathrm{SE}}$，其中 SD 是点估计在 $R = 200$ 次重复模拟中的经验标准差，SE 是报告的标准误的均值。比率大于 1 表明报告的标准误低估了实际抽样变异性（宽度驱动缺口）。

- 标准化平均偏差 $= |b|/\mathrm{SD}$，其中 b 是重复模拟中的平均偏差。数值超过 ~ 0.5 表明存在不可忽略的中心化误差（偏差驱动缺口）。

在 E-WS 下，贫困系数的标准误比为 1.13–1.14，标准化偏差可忽略（0.06–0.22），这证实了 8 个百分点的覆盖率差距完全由区间宽度所致：夹心方差估计量将真实抽样方差低估了约 13%。这与已有充分文献记载的 $O(S^{-1})$ 聚类稳健方差估计量在 $S = 51$ (Binder, 1983年)下的有限样本向下偏差一致。对于$\log \kappa$，情况则相反：标准误比为 1.00（区间宽度校准完美），但标准化偏差为 1.37，表明伪似然点估计存在系统性偏移，导致未覆盖完全发生在单侧（低于目标值的比例为 39.5%，高于目标值的比例为 0%）。对于 τ_{ext} 和 τ_{int}，巨大的标准化偏差（> 3）压倒了任何区间宽度方面的考量，进一步证实三明治校正从结构上不适用于方差分量。

表 E.8. S3（NSECE 校准）情景下的覆盖率差距分解。“低于”和“高于”分别表示真实值低于或高于置信区间的重复模拟百分比。标准误比 > 1 表示区间过窄； $|b|$/标准差 > 0.5 表示点估计偏差不可忽略。 $\star$ =夹心估计量在结构上不适用（E-WS $\equiv$ E-WT）。

参数	Est.	覆盖率（%）	未覆盖率（%）		标准误比	$\lvert\bar{b}\rvert$/标准差	主要来源
			低于	高于			
α_{pov}	E-UW	89.5	1.0	9.5	0.86	0.63	—
	E-WT	58.5	15.5	26.0	1.93	0.22	宽度
	E-WS	82.0	5.0	13.0	1.13	0.22	宽度
β_{pov}	E-UW	89.0	0.5	10.5	0.85	0.69	—
	E-WT	62.5	15.5	22.0	1.88	0.06	宽度
	E-WS	82.5	8.5	9.0	1.14	0.06	宽度
对数 κ	E-UW	92.0	1.0	7.0	0.85	0.48	—
	E-WT	34.0	65.0	1.0	1.75	1.37	Both
	E-WS	60.5	39.5	0.0	1.00	1.37	Bias
τ_{ext}	E-UW	98.0	1.5	0.5	0.64	0.59	—
	E-WT/WS	0.5	99.5	0.0	1.22	3.18	偏差*
τ_{int}	E-UW	97.5	1.0	1.5	0.73	0.51	—
	E-WT/WS	0.0	100.0	0.0	1.11	3.55	偏差*

E.8. **95% 名义水平下的覆盖率**。表 E.9报告了三个情景ˆ下五个目标参数在 90% 和 95% 名义水平下的覆盖率，以及三种估计量的结果。对于所有估计量，95%区间均按 $\theta \pm 1.96 \times$标准误计算，其中标准误为后验标准差（E-UW、E-WT）或夹心标准误（E-WS）。验证结果表明，这种 Wald 构造与原始的基于分位数的 90%区间相差不超过 0.1 个百分点，从而验证了正态近似。

这些结果进一步印证了第E.7节所指出的 E-WS 覆盖率差距是由区间宽度驱动的解释。对于 S3 下的贫困系数，从 90% 扩大到 95% 时，覆盖率仅增加 5.5–6.0 个百分点（从 82.0–82.5% 升至 87.5–88.5%），相对于 95% 名义水平仍有 6.5–7.5 个百分点的差距。如果夹心标准误经过了完美校准，预期增幅应为 ≈ 5.0 个百分点（来自正态分布尾部在 $z_{0.05}$ 与 $z_{0.025}$ 之间增加的概率质量）；观察到的增幅与这一预测一致，证实覆盖率不足会随区间宽度成比例变化，而非反映一种依赖水平的现象。

表 E.9. 90% 和 95% 名义水平下的区间覆盖率（%）。括号内数值为蒙特卡洛标准误（$= \sqrt{\hat{p}(1-\hat{p})/R}$；$R = 200$）。在 95% 名义水平下，蒙特卡洛标准误 ≈ 1.5 百分点。

参数估计		S0（非信息）		S3（NSECE）		S4（压力）	
		90%	95%	90%	95%	90%	95%
α_{pov}	E-UW	93.0 (1.8)	97.0 (1.2)	89.5 (2.2)	96.0 (1.4)	92.5 (1.9)	97.5 (1.1)
	E-WT	81.0 (2.8)	89.0 (2.2)	58.5 (3.5)	67.5 (3.3)	53.0 (3.5)	61.0 (3.4)
	E-WS	88.5 (2.3)	96.0 (1.4)	82.0 (2.7)	87.5 (2.3)	86.0 (2.5)	89.5 (2.2)
β_{pov}	E-UW	93.0 (1.8)	96.5 (1.3)	89.0 (2.2)	95.0 (1.5)	91.0 (2.0)	96.0 (1.4)
	E-WT	76.5 (3.0)	86.0 (2.5)	62.5 (3.4)	70.5 (3.2)	56.5 (3.5)	64.0 (3.4)
	E-WS	86.5 (2.4)	91.5 (2.0)	82.5 (2.7)	88.5 (2.3)	84.0 (2.6)	89.5 (2.2)
$\log \kappa$	E-UW	91.5 (2.0)	96.0 (1.4)	92.0 (1.9)	96.5 (1.3)	90.5 (2.1)	93.5 (1.7)
	E-WT	64.0 (3.4)	73.0 (3.1)	34.0 (3.3)	38.5 (3.4)	23.0 (3.0)	28.0 (3.2)
	E-WS	76.5 (3.0)	84.5 (2.6)	60.5 (3.5)	68.5 (3.3)	57.0 (3.5)	68.0 (3.3)
τ_{ext}	E-UW	97.5 (1.1)	100.0 (0.0)	98.0 (1.0)	99.5 (0.5)	77.0 (3.0)	93.0 (1.8)
	E-WT	3.0 (1.2)	7.0 (1.8)	0.5 (0.5)	0.5 (0.5)	0.0 (0.0)	0.0 (0.0)
	E-WS	3.0 (1.2)	7.0 (1.8)	0.5 (0.5)	0.5 (0.5)	0.0 (0.0)	0.0 (0.0)
τ_{int}	E-UW	91.0 (2.0)	95.0 (1.5)	97.5 (1.1)	99.5 (0.5)	97.0 (1.2)	98.5 (0.9)
	E-WT	8.0 (1.9)	12.0 (2.3)	0.0 (0.0)	0.5 (0.5)	0.0 (0.0)	0.0 (0.0)
	E-WS	8.0 (1.9)	12.0 (2.3)	0.0 (0.0)	0.5 (0.5)	0.0 (0.0)	0.0 (0.0)

相比之下，对于所有固定效应，E-UW 在两个水平上均实现了接近名义覆盖率（90% 时为 89.0–93.0%；95% 时为 95.0–97.5%），这进一步表明，在模型设定正确时，非加权后验经过了良好校准。对于$\log \kappa$，E-WS 下的 95% 覆盖率（68.5%）仍远低于名义水平，这与表 E.8中指出的偏差驱动的不足一致。