A真实解聚：小区域估计视角 *

作者：Yunhan Wu[1], Finn Lindgren[2] 和 Heidi A. Hanson[1]

[1]橡树岭国家实验室，美国田纳西州橡树岭，邮编 37830[2]爱丁堡大学数学学院及麦克斯韦数学科学研究所，英国爱丁堡

arXiv:2603.04246v1 [stat.ME] 4 Mar 2026

摘要

在精细区域尺度上生成可靠的健康与人口指标估计值，对于考察异质性和支持地方健康政策至关重要。然而，许多调查仅按较粗的行政层级发布结果，从而限制了这些数据对决策的参考价值。我们提出一种用于区域层级解聚的完全贝叶斯单阶段空间建模框架，可直接利用粗粒度聚合调查数据生成指标的精细尺度估计值。该框架在目标分辨率上定义潜在空间过程，并通过聚合步骤将其与观测结果相连接；它采用小区域估计技术，同时纳入协变量并提供一致的不确定性量化。我们使用 inlabru 实现所提方法，以提高计算效率。我们通过一项针对肯尼亚总生育率的模拟研究评估模型表现，展示模型在不同数据生成情景下恢复精细尺度变异的能力。我们还将该框架应用于两项全国性调查，利用 2022年肯尼亚人口与健康调查生成地区层级的生育率估计值；更重要的是，利用 2021年肯尼亚时间利用调查生成地区层级的无偿照护和家务劳动以及大众媒体使用指标。

关键词： 小区域估计、贝叶斯平滑、空间模型、分解、调查统计

1 引言

小区域估计（SAE）在现代公共卫生和流行病学中发挥着重要作用，可为直接调查数据稀少或无法获取的地理亚群体提供可靠的估计值。随着人们日益重视本地化规划和解决健康不平等问题，对精细区域尺度上的疾病负担和健康指标估计值的需求显而易见。例如，在低收入和中等收入国家，这通常对应于地区（二级行政区）层级，该层级通常是管理决策和资源分配的执行单位。实证研究和国际倡议——尤其是可持续发展目标（SDG）相关工作——明确呼吁以区域层面的精细粒度报告健康指标（Oosterhof，2018；Janocha 等，2021；J¨onsson 和 Bexell，2021）。研究证实了针对特定指标的这些需求，包括但不限于疫苗接种覆盖率（Utazi等，2021）、教育程度（Delprato 等，2024；Wu 等，2025）、生育率（Abate 等，2024）和儿童死亡率（Wu 等，2021）。

满足这些需求，需要能够在直接调查数据不足时生成稳健估计值的统计方法。基于设计的小区域估计方法，例如 Horvitz-Thompson 估计量（Horvitz 和 Thompson，1952），在处理许多感兴趣区域的小样本量时存在局限。为解决这一问题，基于模型的小区域估计方法日益流行（Rao 和 Molina，2015）。经典的区域层级模型，例如 Fay–Herriot（FH）模型（Fay 和

*本文由 UT-Battelle有限责任公司根据与美国能源部（DOE）签订的合同 DE-AC05-00OR22725 撰写。美国政府保留相关权利；出版商接受本文发表，即承认美国政府保留一项非独占、已付费、不可撤销的全球许可，可为美国政府目的发表或复制本文的已发表版本，或允许他人这样做。DOE将根据 DOE公共访问计划（https://www.energy.gov/doe-public-access-plan）向公众提供这些由联邦资助的研究成果。

Herriot，1979），以及单位层级模型，例如由 Battese等人（1988）提出的模型，允许通过纳入通常来自人口普查或行政记录的辅助信息，在各区域之间借用统计强度（Pratesi，2016；Morales 等人，2021；Moretti 和 Whitworth，2021）。这些方法的贝叶斯形式自然带有不确定性量化，并便于进行纳入协变量以及空间或时间依赖结构的层次建模。

尽管取得了这些进展，小区域估计框架通常假设响应数据与目标推断层级具有相同的地理分辨率。例如，许多人口与健康调查（DHS）提供 GPS坐标，因此可以将个体划分到各区（一级行政区）。然而，在许多其他情况下，出于保密或数据收集限制，调查结果仅以较粗的空间尺度发布，而政策和研究应用却需要更精细尺度的估计值。

以2021年肯尼亚时间利用调查（KTUS）（肯尼亚国家统计局，2023年）为例。尽管许多关键指标理想情况下应按地区尺度绘制地图，以指导因地制宜的干预措施，但公开发布的数据仅具有县（一级行政区）层面的地理索引。这一局限对可持续发展目标5.4.1（联合国统计司，2020年）尤为重要，该指标追踪按性别、年龄和地点划分的无偿家务和照护工作时间占比。区级估计对于揭示异质性以及制定有针对性的家庭福利和劳工政策至关重要。然而，如果没有方法上的创新，这类粗尺度数据就无法用于精细尺度推断。

尽管小域估计方法旨在解决预先定义的地理单元内小样本量带来的挑战，分解方法处理的则是一个不同但相关的问题：在比观测（聚合）数据更精细的空间分辨率下生成预测值。这一挑战存在于地理学、遥感、人口学和流行病学等多个领域，并有多种称谓，包括降尺度、分解、人口密度分区制图或超分辨率。尽管表述方式各异，这些方法都旨在利用辅助数据和建模假设来提高空间细节。

一种广泛使用的方法是人口密度分区制图，它利用土地覆盖、建筑密度或遥感代理指标等辅助协变量，将区域数据重新分配到更小的区域。LandScan（Dobson 等，2000）和 WorldPop（Tatem，2017）等人口制图产品，以及环境健康领域的应用（Requia 等，2018），都是典型案例。然而，这些方法通常将汇总总量视为固定值，并忽略抽样误差或测量误差。虽然这些方法适用于人口数等稳定数量，但对于源自调查的结局则不太适用，因为这类结局的不确定性必须得到明确传递。在这种情况下，小区域估计模型更能胜任这项任务。

另一种策略是空间微观模拟，即构建受约束、与观测边际分布相匹配的合成微观数据。这种方法与后分层多层回归（多层回归与后分层）（Park 等，2004）类似：先按人口群体对结果变量建模，再通过合成人口构成进行重新分配。其应用涵盖健康、福利、交通和人口统计学（Tanton，2014；Smith 等，2021）。然而，微观模拟面临计算可扩展性、验证和不确定性量化方面的挑战。

另一类方法采用地统计模型来估计连续索引的空间表面，通常使用随机偏微分方程框架（Lindgren 等，2011）。这类模型将聚合数据与连续潜在空间场关联起来，广泛用于疾病制图，包括 COVID-19 研究（Python等，2022）。它们提供了一个连贯的贝叶斯框架，能够量化不确定性，并有 R 语言中的 disaggregation 软件包等工具提供支持（南迪 等，2023）。然而，这类模型的表现高度依赖高质量的空间协变量，而这些协变量并不总能有效预测健康结局。此外，由于缺乏真实值数据，在精细分辨率下进行验证仍然颇具挑战（Arambepola等，2020）。

重要的是，许多此类分解方法并未考虑调查设计，而调查设计是健康与人口指标调查数据的核心特征。小域估计方法能够通过结合基于设计的一致性与基于模型的灵活性，独特地弥合这一差距。

鉴于现有方法的局限性，以及生成可靠区域级估计值这一实际目标，我们提出了一种用于健康指标空间分解的新型区域级小域估计方法。我们将该方法构建为一个完全贝叶斯单阶段空间模型，在目标分辨率上定义潜在过程，并纳入区域层级协变量的信息。模型内置的聚合步骤将这一潜在场与较粗尺度上观测到的结果变量相连接，从而确保观测到的调查数据与更精细层级的推断保持一致。尽管我们的示例源自低收入和中等收入国家的情境，所提出的框架具有广泛的适用性

，可用于各种不同情境。这一建模方式使该框架能够纳入经典小区域估计方法，例如 FH 模型和单位层级模型；在人口统计微观数据可用时，也能纳入多层回归与后分层。计算方面，我们使用 inlabru 及相关工具实现这些模型，以提高计算效率（Bachl 等，2019年；Lindgren 等，2024；Suen等，2025）。该框架在构建时便实现了连贯的不确定性量化，并支持标准的贝叶斯推断。

本文其余部分安排如下。第 2 节介绍方法框架：首先回顾现有的小域估计方法，然后介绍我们提出的区域层级分解模型。第 3 节通过一项关于肯尼亚总和生育率（GFR）的模拟研究评估模型性能。第 4 节和第 5 节将该框架应用于两个真实案例——2022年肯尼亚人口与健康调查和 KTUS。最后，第 6 节讨论研究发现及其影响。

2 方法

2.1 小区域估计模型概述

本节介绍小区域估计的三类核心方法：直接估计、区域层级（FH）模型和单位层级模型（Rao 和 Molina，2015；Fay 和 Herriot，1979；Battese 等，1988）。这些方法假设调查数据在关注的层级上具有地理索引，这意味着结局是在更细尺度的子区域中针对抽样个体观测得到的，而非仅以聚合形式呈现。在本文后续部分，我们将拓展这些标准框架，介绍区域分解方法，以应对数据仅在较粗地理层级上可用这一更具挑战性的情形。

这些方法构成了一个自然的谱系：从完全依赖抽样设计的基于设计的估计量，到越来越多地采用基于模型的方法；后者以牺牲一定的设计一致性为代价，换取处理稀疏数据时更大的灵活性。除了这三种方法，我们还考虑多层回归与后分层（Downes 等，2018），它可以与区域层级模型或单位层级模型结合，以纳入额外的人口统计特征，并使估计值与人口结构相匹配。多层回归与后分层通过人口统计变量将总体划分为若干单元，对各单元内的结果变量建模，再利用已知人口数汇总预测值，是一种扩展方法。

考虑一个有限目标总体，其中有 N 个个体居住在 $|\mathcal{I}|$ 个粗尺度区域（例如一级行政区）内，这些区域的索引为 $i = 1,\ldots,|\mathcal{I}|$。每个区域 i 包含一组更细尺度的子区域（例如二级行政区），其索引为 $j = 1,\ldots,|\mathcal{J}|$。我们用 $i[j]$ 表示包含子区域 j的上级较粗区域。在每个子区域 j内，有 N_j 个个体，其索引为 $k = 1,\ldots,N_j$。

令 y_{jk} 表示子区域 j中个体 k 的结局。我们的研究方法框架聚焦于健康指标小区域估计中最受关注的两种结局类型：二元结局和泊松结果。二元情形（$y_{jk} \in \{0,1\}$）在疾病制图中很常见，对应于患病率估计。泊松情形（$y_{jk} \in \{0,1,2,\ldots\}$）适用于罕见事件或生育率、死亡率等比率型指标。

我们的主要目标是估计子区域层级的患病率或比率，定义为 a s

$$\mu_j = \frac{1}{N_j}\sum_{k=1}^{N_j} y_{jk},$$

其中，N_j 是子区域 j的人口规模。

虽然并非直接关注的目标，我们也将粗区域层级的患病率或比率定义为其各子区域的加权聚合，权重为人口规模。这一表述为后续章节介绍的细分方法提供了依据：

$$\mu_i = \sum_{j:i[j]=i} \frac{N_j}{N_i}\mu_j,$$

其中，$N_i = \sum_{j:\ i[j]=i} N_j$ 是区域 i的人口。

小区域估计的一个重要方面是调查抽样设计。个体通常通过涉及不等入选概率的复杂抽样设计进行抽取。对于每个子区域 j，按照调查设计抽取一个样本量为 n_j 的样本，我们将其记为 $S_j \subset \{y_{jk}:\ k = 1,\ \ldots,\ N_j\}$，其中 $|S_j| = n_j$。对于每个抽中的个体 $k \in S_j$，其对应的设计权重 w_{jk}定义为其纳入概率的倒数，使得 $w_{jk} = \frac{1}{\pi_{jk}}$，其中 π_{jk} =Pr($k \in S_j$)。

在此设定下，我们介绍三种标准的小区域估计方法，适用于数据的地理位置标注至子区域层级的情形。

2.1.1 直接估计

直接估计量仅使用每个子区域内的抽样数据，并通过权重考虑调查设计（Rao 和 Molina，2015）。子区域 j 中患病率的加权估计量由下式给出（H´ajek，1971）：

$$\hat{\mu}_j^w = \frac{\sum_{k\in S_j} w_{jk} y_{jk}}{\sum_{k\in S_j} w_{jk}}.$$

这些估计值在设计上是一致的，并且所需的建模假设极少，但当样本量较小时，它们往往具有较大的方差。这促使人们使用基于模型的方法，从邻近区域借用强度。

2.1.2 区域层级模型

区域层级模型，即FH模型（Fay 和 Herriot，1979），通过层级模型将各区域联系起来，从而扩展了直接估计。令$\hat{}\mu_j^w$表示子区域j 中患病率或比率的直接估计值，并将变换后的估计值定义为

$$\hat{\lambda}_j^w = g(\hat{\mu}_j^w), \qquad V_j = \widehat{\mathrm{var}}(\hat{\lambda}_j^w),$$

其中 $g(\cdot)$为连接函数，对患病率和比率分别采用逻辑斯蒂变换和对数变换。变换尺度上的方差 V_j 可通过德尔塔法由基于设计的方差得到。

FH模型假设

$$\begin{aligned} \hat{\lambda}_j^w \mid \lambda_j &\sim \mathcal{N}(\lambda_j, V_j), \\ \lambda_j &= \alpha + \boldsymbol{x}_j^\top \boldsymbol{\beta} + b_j, \end{aligned} \tag{1}$$

其中， $\boldsymbol{x}_j$ 为子区域层级协变量， $\boldsymbol{\beta}$ 为其系数， b_j 为随机效应。子区域 j 的患病率或比率为 $\mu_j = g^{-1}(\lambda_j)$。

为捕捉 b_j中的空间依赖，我们采用 BYM2 表述（Riebler 等，2016）。该表述是Besag-York-Mollie（BYM）（Besag等，1991）的一种重新参数化，并将随机效应分解为一个无结构的独立同分布（IID）项和一个具有空间结构的内在条件自回归成分。总方差以及归因于结构化成分的比例由超参数控制，我们为这些超参数指定惩罚复杂度先验（Simpson 等，2017）。详细信息（包括超先验设定）见补充材料的 S2 节。对于后文讨论的所有模型，我们都采用相同的空间结构。

平滑通过在子区域之间借力来降低方差，但 FH模型仍依赖稳定的直接估计值及其方差；当数据非常稀疏时，这可能会限制其表现。

2.1.3 单位层级模型

单位层级小区域估计模型直接使用个体结局，在采用整群抽样的调查中通常于聚类层级实施。对于包含 C 个抽样簇的调查，令 Y_c 与 n_c 分别表示聚类 c中的阳性结局数和抽样个体数。视结局类型而定，数据被建模为

$$Y_c \mid \mu_c \sim \begin{cases} \mathrm{Binomial}(n_c, \mu_c), & \text{binary outcomes}, \\ \mathrm{Poisson}(n_c \mu_c), & \text{count outcomes}, \end{cases}$$

其中 μ_c 表示聚类 c中的潜在患病率或比率。

均值结构通过链接函数加以设定：

$$g(\mu_c) = g(\mu_{c[j]}) = \alpha + \boldsymbol{x}_j^\top \boldsymbol{\beta} + b_j, \tag{2}$$

其中， μ_c 是抽样簇 c中的潜在患病率或比率，我们通常假设同一区域的抽样簇具有相同的值。令 $c[j]$ 表示抽样簇 c 所在的子区域 j 。我们

于是， $\boldsymbol{x}_j$ 为协变量， b_j 为子区域 j中的子区域级随机效应（如FH模型中所示）。

单位层级模型的一个局限是未直接纳入调查权重，这意味着没有明确考虑抽样设计。在分层设计中，这一问题尤为突出：若忽略城乡状况等层，且分层变量与结局相关，同时对某些层过度或抽样不足，就可能产生有偏估计（Wu and Wakefield，2024）。一种可行的解决方案是将此类分层变量直接纳入均值结构，以调整其影响，然后将各层的估计值汇总为总体估计值。

2.1.4 多层回归与后分层

多层回归与后分层为结合基于模型的估计、调查设计信息和人口统计学协变量提供了一个有原则的框架。例如，假设我们要按人口统计群体进行分层 $a \in \mathcal{A}$ （例如，年龄、性别或城乡类别）。对于 FH 方法和单位层面方法，我们都可以针对群体特定均值，以群体特定似然的形式指定分层模型。我们在由城乡状况和人口统计类别等分层变量定义的后分层单元内对结果变量进行建模。

对于FH模型，我们使用

$$\begin{aligned}\hat{\lambda}^{w}_{j,a} \mid \lambda_j &\sim \mathcal{N}(\lambda_{j,a}, V_{j,a}),\\ \lambda_{j,a} &= \alpha + \boldsymbol{x}_j^{\top}\boldsymbol{\beta} + b_j + f(a),\end{aligned}$$

with 通过从建模尺度反变换得到的组特异性患病率或比率：

$$\mu_{j,a} = g^{-1}(\lambda_{j,a}).$$

对于单位层级模型，我们使用

$$Y_{c,a} \mid \mu_{c,a} \sim \begin{cases}\text{Binomial}(n_{c,a}, \mu_{c,a}), & \text{binary outcomes},\\ \text{Poisson}(n_{c,a}\mu_{c,a}), & \text{count outcomes},\end{cases}$$

$$g(\mu_{c,a}) = g(\mu_{c[j],a}) = \alpha + \boldsymbol{x}_j^{\top}\boldsymbol{\beta} + b_j + f(a).$$

分层效应记为 $f(a)$，与协变量和空间随机效应一并纳入。当分组变量较少且类别有限时，可以直接对固定效应（及其交互作用）建模。然而，随着组数增加，每个单元格内的数据会变得稀疏。在这种情况下，可以采用分层平滑来稳定估计值。

在后分层阶段，利用外部人口计数$N_{j,a}$（通常来自人口普查数据）将预测值汇总为区域总体估计值，从而使估计值反映真实的人口构成：

$$\mu_j^{\text{MRP}} = \frac{1}{N_j}\sum_a N_{j,a}\,\mu_{j,a}, \qquad N_j = \sum_a N_{j,a}. \tag{3}$$

因此，多层回归与后分层通过显式调整预测性人口统计变量，并确保估计值与人口结构一致，从而扩展了FH模型或单位层级模型。这一调整考虑了由人口统计异质性引起的子区域变异，并提高了小区域估计的精度。

2.2 区域分解

我们提出的区域分解方法超越了上文介绍的标准 SAE框架。例如，调查结果仅按较粗的层级进行地理索引，但分析和政策目标要求在更细的分辨率下获得估计值。

具体而言，假设调查记录的数据 $\{y_i, n_i\}$ 按较粗区域 i进行地理索引，而估计目标是 μ_j 子区域 j 中的患病率或比率 i。如果仅有汇总数据，由于无法识别每个粗粒度区域内结果变量的分布，因此在没有额外假设或信息的情况下，无法对 $\{\mu_j\}$ 进行估计。

我们所提出的框架通过系统地纳入能够反映区域内变异的辅助信息来源，应对这一挑战。这些来源大致可分为三类。第一，空间结构使我们能够在相邻

子区域之间借力。通过在更精细的区域层级构建邻近矩阵，我们可以将粗尺度区域中的空间模式延伸至更精细的空间层级。第二，来自外部来源（如卫星数据、人口普查数据）的协变量，涵盖从环境到人口统计指标的各种信息，有助于解释粗粒度区域内的结果变异。第三，多层回归与后分层通过使用真实的人口统计构成汇总群体特定估计值来改进估计值，使人口统计方面的异质性能够反映子区域中的变异——即使各子区域的群体特定估计值相同。

以下小节将介绍通过结合这些互补信息来源来进行区域分解的统计模型。

2.2.1 FH细分

我们首先扩展区域层级（FH）模型以用于分解。从粗粒度区域 i 的调查加权ˆ直接估计值入手。具体而言，令 λ_i^w 表示区域 i 的变换后直接估计值，其基于设计的方差为 V_i。按照 FH模型的设定，我们假设其服从高斯分布：

$$\hat{\lambda}_i^w \mid \lambda_i \sim \mathcal{N}(\lambda_i, V_i),$$

其中， λ_i 是区域 i 的潜在患病率或比率。

为将区域层级参数 λ_i 与定义在精细空间尺度上的潜在过程关联起来，我们构建子区域均值的人口加权聚合：

$$\lambda_i = g(\mu_i) = g\left(\sum_{j:i[j]=i} \frac{N_j}{N_i}\mu_j\right), \tag{4}$$

其中， N_j 是子区域 j 的人口， $N_i = \sum_{j:\ i[j]=i} N_j$ 是较粗区域的人口 i,

and $g(\cdot)$ 是链接函数。

潜在过程以子区域层级为索引。具体而言，对于每个子区域 j，我们有 ve

$$g(\mu_j) = \alpha + \boldsymbol{x}_j^\top \boldsymbol{\beta} + b_j,$$

其中，参数建模和平滑结构的设置与第2.1.2节讨论的FH模型一致。

贝叶斯层级模型的计算细节和实现方面将在第2.3节讨论。特别是，方程（4）中的非线性聚合需要特殊处理。

2.2.2 单元级模型分解

我们将此框架扩展至单位层级模型，其中观测记录在嵌套于粗尺度区域 i 内的抽样簇 c 中。令 Y_c 表示抽样簇 c 中观测到的结果，且该簇的集群规模为 n_c。目标是将地理索引位于较粗层级 $i[c]$ 的群组层级结果与定义在更细尺度子区域 j 上的潜在过程关联起来。

在这种设定下，同一聚类内的结果变量并非独立地从同一个共享均值中抽取，而是存在相关性。由于聚类 c 中的所有个体都属于同一个子区域，因此与从区域均值中独立抽样相比，他们的结果变量更为相似。这种集群内相似性会增加聚合结果的变异，因此标准的泊松分布或二项似然会低估不确定性。为解决这一问题，我们采用过度离散边际抽样模型：对计数结果使用负二项分布，对二元结果使用贝塔-二项分布：

$$Y_c \sim \begin{cases} \text{NegBin}(n_c\mu_c, \phi), & \text{count outcomes,} \\ \text{BetaBinomial}(n_c, \mu_c, \phi), & \text{binary outcomes,} \end{cases}$$

其中， ϕ是一个过度离散参数，用于描述集群内相关性所导致的额外变异。

继 FH模型 的扩展之后，我们将较粗区域 i 中观测到的抽样簇与细尺度子区域上的潜在结构联系起来。位于区域 i 中的抽样簇的均值模型表示为子区域均值的人口加权平均：

$$\mu_c = \mu_{i[c]} = \sum_{j:i[j]=i} \frac{N_j}{N_i}\,\mu_j,$$

其中， N_j 是子区域 j的人口数，而 $N_i = \sum_{j:\ i[j]=i} N_j$ 是较粗区域的人口数。

在子区域层级，潜在过程通过线性潜在场建模：

$$g(\mu_j) = \alpha + \boldsymbol{x}_j^\top \boldsymbol{\beta} + b_j, \tag{5}$$

其中， $\boldsymbol{x}_j$ 是协变量， b_j 是随机效应， $g(\cdot)$是链接函数。随机效应b_j 捕捉结构化和非结构化空间变异。

2.2.3 扩展以纳入多层回归与后分层

基于上述单位层级和 FH 分解模型的设定，我们可以自然地纳入多层回归与后分层。即使结果变量仅在较粗的地理层级上观测到，该框架也能利用人口统计异质性为子区域估计值提供信息。

设分层组为 $a \in \mathcal{A}$。与第 2.1.4 节中的多层回归与后分层类似，我们按组定义数据模型和均值结构。

对于 FH模型，我们将层级结构设定为：

$$\begin{aligned} \hat{\lambda}^{w}_{j,a} \mid \lambda_{j,a} &\sim \mathcal{N}(\lambda_{j,a}, V_{j,a}), \\ \lambda_{i,a} &= g\left(\sum_{j:i[j]=i} \frac{N_{j,a}}{N_{i,a}}\, \mu_{j,a} \right), \\ g(\mu_{j,a}) &= \alpha + \boldsymbol{x}_j^\top \boldsymbol{\beta} + b_j + f(a) \end{aligned} \tag{6}$$

其中，$N_{j,\ a}$ 是子区域 j中第 a 组的人口， $N_{i,\ a} = \sum_{j:\ i[j]=i} N_{j,\ a}$ 表示较粗区域 i的人口。

类似地，对于单位层级模型，我们有

$$\begin{aligned} Y_{c,a} \mid \mu_{c,a} &\sim \begin{cases} \text{Binomial}(n_{c,a}, \mu_{c,a}), & \text{binary outcomes}, \\ \text{Poisson}(n_{c,a}\mu_{c,a}), & \text{count outcomes}, \end{cases} \\ \mu_{c,a} = \mu_{c[i],a} &= \sum_{j:i[j]=i} \frac{N_{j,a}}{N_{i,a}}\, \mu_{j,a}, \\ g(\mu_{c,a}) &= \alpha + \boldsymbol{x}_{j[c]}^\top \boldsymbol{\beta} + b_{j[c]} + f(a). \end{aligned} \tag{7}$$

所有步骤均遵循第 2.1.4 节所述的标准多层回归与后分层。由于均值结构均定义在精细区域 j，我们可以使用公式（3）中所述的相同形式，将结果汇总回整体估计值。

2.3 模型实现与非线性

通过 R-INLA 软件包（版本 25.06.22）实现的集成嵌套拉普拉斯近似（INLA）方法广泛用于贝叶斯小区域估计，因为它能为潜在高斯模型（LGM）提供快速且准确的近似推断（Bakka 等，2018；Rue 等，2009；Lindgren 和 Rue，2015；Wakefield等，2025）。对于空间分层模型，R-INLA 利用高斯马尔可夫随机场的稀疏精度结构，大幅提升计算效率，并内置了 BYM2先验（Riebler 等，2016）等空间平滑先验；我们在分析中采用了该先验。这些特性使 R-INLA 成为实现我们建模框架的自然选择。

基准 INLA 方法假设线性预测子具有 $\boldsymbol{\eta} = \boldsymbol{A}\boldsymbol{u}$ 的形式，其中 $\boldsymbol{u}$ 汇集了潜在高斯分量，包括固定效应以及空间和非空间随机效应（例如，式（1）右侧）。在抽样模型中，每个观测 y_k 都通过可能为非线性的链接函数 $h(\cdot)$ 依赖于 η_k。在 FH模型中， $h(\cdot)$ 是恒等链接，因为结局的变换在抽样模型之前进行；而对于单位层级模型，$h(\cdot) = g^{-1}(\cdot)$。因此，唯一允许的非线性是从 η_k 到位置参数的链接；从 $\boldsymbol{u}$到 $\boldsymbol{\eta}$ 的映射必须严格线性（Rue 等，2009）。以细粒度子区域为地理索引的模型符合这一约束，并可按照式（1）和式（2）的规定，直接在 R-INLA 中实现。

相比之下，我们的分解模型违反了这一要求，因为聚合引入了非线性。与抽样模型的位置参数相关联的预测变量成为潜在高斯分量的非线性组合。

为了说明这一点，我们从式（5）中的潜在场出发，子区域均值为

$$\mu_j = g^{-1}\big(\alpha + \boldsymbol{x}_j^\top \boldsymbol{\beta} + b_j\big) = g^{-1}(\boldsymbol{u}_j).$$

对于 FH 模型和单位层级分解模型，预测变量的形式为

$$\tilde{\eta}_i(\boldsymbol{u}) \;=\; g\left(\sum_{j:i[j]=i} \frac{N_j}{N_i}\,\mu_j\right) \;=\; g\left(\sum_{j:i[j]=i} \frac{N_j}{N_i}\,g^{-1}(\boldsymbol{A}\boldsymbol{u}_j)\right), \tag{8}$$

其中，非线性预测变量 $\tilde{\eta}_i(\boldsymbol{u})$ 涉及对聚合预测变量 [1] 的变换（例如，加权指数和的对数、加权 expit 和的逻辑斯蒂变换），且无法表示为线性 ˜形式 $\boldsymbol{A}\boldsymbol{u}$。

为克服这一挑战，我们使用 inlabru 软件包（版本 2.13.0.9011）（Lindgren等，2024；Bachl等，2019；Lindgren等，2025）。该软件包扩展了 R-INLA，使潜在预测变量可以是潜在高斯场的非线性函数。在 inlabru 的表述中，我们将非线性预测变量的堆叠向量 $\tilde{\boldsymbol{\eta}}(\boldsymbol{u})$ 表示为 $\boldsymbol{u}$ 的确定性函数，如公式（8）所示。选取线性化点 $\boldsymbol{u}_0$ 后，我们进行一阶泰勒近似：

$$\bar{\boldsymbol{\eta}}(\boldsymbol{u}) \;=\; \tilde{\boldsymbol{\eta}}(\boldsymbol{u}_0) + \boldsymbol{B}\,(\boldsymbol{u} - \boldsymbol{u}_0) \;=\; \boldsymbol{B}\boldsymbol{u} + \big[\tilde{\boldsymbol{\eta}}(\boldsymbol{u}_0) - \boldsymbol{B}\boldsymbol{u}_0\big],$$

wher其中 e $\boldsymbol{B}$ 是非线性预测变量在 $\boldsymbol{u}_0$ 处的导数矩阵。这得到一个线性化 d 版 原始非线性预测变量的版本，其中偏移量取决于 $\boldsymbol{u}_0$。

精确的边缘抽样模型定义为

$$\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{\theta} \;\sim\; p(\boldsymbol{y} \mid \tilde{\boldsymbol{\eta}}(\boldsymbol{u}), \boldsymbol{\theta})\,,$$

其中 $\boldsymbol{\theta}$ 是超参数。我们通过将非线性预测变量替换为其线性化形式来近似该模型，得到

$$\bar{p}(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{\theta}) \;=\; p(\boldsymbol{y} \mid \bar{\boldsymbol{\eta}}(\boldsymbol{u}), \boldsymbol{\theta}) \;\approx\; p(\boldsymbol{y} \mid \tilde{\boldsymbol{\eta}}(\boldsymbol{u}), \boldsymbol{\theta}) \;=\; \tilde{p}(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{\theta}).$$

通过这种线性化，非线性观测模型可以近似为标准潜高斯模型，并可使用 INLA 进行拟合，从而能够使用 INLA 推断及后验摘要的全部功能。

inlabru 方法的一个关键步骤是选择合适的线性化点 $\boldsymbol{u}_0$。这通过不动点迭代过程实现，该过程不断更新线性化点，使其与线性化模型的条件后验众数一致。从一个初始点出发，算法交替执行 INLA 方法以更新超参数，并计算潜在场的条件后验众数，不断改进 $\boldsymbol{u}_0$，直至收敛。有关该迭代方案的完整细节，包括更新规则和收敛准则，见 Lindgren等 (2024)。

3 模拟

3.1 概述

我们设计了一项模拟研究，以评估空间分解方法在估计肯尼亚总生育率方面的表现。在数据生成过程中，我们首先模拟个体层级的生育率结果，并将其地理定位到二级行政区（选区），作为真实值。随后，我们复现人口与健康调查的抽样程序，从该合成人口中抽取调查样本。在模型拟合时，我们施加了数据限制：仅提供带有地理索引的一级行政区（县）级数据，并移除所有 Admin-2 标识符。接着，应用所提出的分解方法，根据一级行政区聚合数据推断二级行政区级总生育率估计值。最后，将这些估计值与已知的二级行政区真实值进行比较，以量化该分解方法的准确性和可靠性。

全文均采用总生育率（GFR）的传统时期定义，即育龄女性（15–44岁）每1,000女性人年暴露量中的活产数：

$$\mathrm{GFR} = 1000 \times \frac{\sum_{a \in A} B_a}{\sum_{a \in A} E_a/12},$$

[1]关于为何一种看似更简单的单位层级模型表述与 inlabru 框架不兼容，补充材料的 S1节中提供了讨论。

其中，A $=\{15$–19、20–24、.....、40–$44\}$ 表示年龄组，B_a 表示参考期内年龄组 a 的活产总数，E_a 表示年龄组 a中的女性暴露人月总数。除以12可将暴露月数转换为女性人年，乘以1、000则将比率换算为每1、000女性人年的活产数。在我们的模拟中，参考窗口为调查年份之前的60个月，我们的目标是每位女性 $T^\star = 48$ 个月的平均暴露量。

在分析中，根据肯尼亚的行政区划边界，我们将一级行政区视为47个县，将二级行政区视为290个选区。模拟样本量和分层方式仿照2022年肯尼亚人口与健康调查，该调查以2019年人口普查作为抽样框。

3.2 数据来源、主抽样框的构建与抽样

我们将行政区划边界矢量文件与2019年肯尼亚人口普查数据相结合，为每个 Admin-2 单元构建社会人口学边际分布。关键变量包括城镇人口比例、具有中等及以上教育程度的女性比例，以及其他区域层面协变量，例如平均家庭规模和手机使用条件。

人口数取自 WorldPop 数据（Tatem，2017）。我们将网格化的分年龄人口估计值汇总至 Admin-2 级，作为基准人口。其他栅格协变量，例如夜间灯光、医疗设施可及性和植被指数，则按 Admin-2 级计算人口加权平均值。

为构建主抽样框，我们采用肯尼亚2022年人口与健康调查的抽样设计。通过将 Admin-2 $\times$ 年龄 $\times$ 城乡属性 $\times$ 教育单元扩展为个体记录，创建合成人口。随后，按照女性人口的比例将普查区（即抽样簇）分配到各层。

最后，我们仿照人口与健康调查的两阶段抽样策略。在各层内，按人口规模概率抽取抽样簇，之后从每个抽样簇中随机抽取20名女性。由此得到的入样概率用于确定基础抽样权重，以确保模拟能高度还原人口与健康调查的设计。更多细节见补充材料的S3节。

3.3 结局生成与模拟情景

我们的目标是构建若干模拟情景，以体现次国家级结果估计中潜在的变异来源。我们希望从简单情景入手，再逐步纳入更多因素，以更好地近似真实世界的数据。具体而言，我们考虑了四种情景，如表1所示；这些情景在是否纳入观测协变量、未观测协变量、区域层面独立随机效应和模型设定错误方面有所不同。

在第一种情景中，观测协变量可以通过反向聚合模型自然地捕捉。在第二种情景中，我们模拟这样一种情况：未观测协变量通常具有空间依赖性，可通过空间平滑部分地加以考虑，而空间平滑能够捕捉 Admin-2 层级的空间相关随机效应。在第三种情景中，我们在 Admin-2 层级引入独立随机效应；这些效应无法从地理索引到 Admin-1 层级的数据中识别，因此也无法恢复。在最后一种情景中，我们为城乡分层指定不同的系数值，以此引入模型设定错误；其他模型则假设总体效应估计。尽管引入交互项在一定程度上反映了现实情况，但我们指出，实际情况可能比即使在最复杂的情景中所捕捉到的也更加复杂。

为形式化结局生成过程，令 $E_{c,\ k}$ 表示个体 k 在聚类 c中的参考期暴露量。个体层面的出生过程独立生成如下：

$$Y_k \sim \text{Poisson}(E_{c,k}\,\mu_{c,k}),$$

其中，$\mu_{c,k}$ 是个体生育率。

情景1：在第一种情景中，我们仅纳入观测协变量。位于二级行政区 j 的聚类 c 中个体 k 的对数尺度均值结构设定为

$$\begin{aligned}\log \mu_{c,k}^{s_1} = {} & \alpha_{\text{rural}}\, I(c \in \text{rural}) + \alpha_{\text{urban}}\, I(c \in \text{urban}) \\ & + \delta_{\text{age}(k)} + \delta_{\text{educ}(k)} \\ & + \boldsymbol{\beta}\boldsymbol{X}_{j[k]} + e_k + e_c.\end{aligned}$$

表1：四种情景下空间异质性的变异来源，以及分解模型对其进行捕捉的预期能力。

变异来源	情景1	情景2	情景3	情景 4	可恢复
观测协变量	✓	✓	✓	✓	✓ 显式捕捉 通过固定效应
未观测协变量 （空间相关的 二级行政区效应)	✗	✓	✓	✓	❍ 部分捕获 被空间效应捕获
IID Admin-2 随机效应	✗	✗	✓	✓	✗ 不可恢复 ， 完全非结构化
变 系数	✗	✗	✗	✓	✗ 导致模型 设定错误

该结构纳入聚居类型截距、个体层协变量（年龄和教育）、区域层面协变量（例如夜间灯光、到最近医疗机构的出行时间、家庭规模），以及个体层级和群组层级的随机效应。本情景及后续情景的完整参数设定见补充材料S3.4节。

情景2： 第二种情景通过在区域层级纳入额外的未观测协变量，对这一均值结构进行了扩展。具体来说，

$$\log(\mu_{c,k}^{s_2}) = \log(\mu_{c,k}^{s_1}) + \boldsymbol{\beta}_{\text{unobs}} \boldsymbol{X}_{j[k]}^{\text{unobs}},$$

其中， $\boldsymbol{X}_{j[k]}^{\text{unobs}}$通过引入两个未观测协变量进一步增加变异：植被指数和女性手机使用情况，两者均经过标准化。这些协变量未纳入建模过程，但预计分解模型能够通过空间相关的二级行政区效应部分捕捉这些变异。

情景3： 在第三种情景中，我们通过引入独立的二级行政区层级随机效应，允许存在额外的未解释异质性：

$$\log(\mu_{c,k}^{s_3}) = \log(\mu_{c,k}^{s_2}) + e_j, \qquad e_j \sim \mathcal{N}(0, \sigma_j^2).$$

这捕捉了分解模型无法解释的非结构化残差变异 in.

情景 4： 第四种情景通过允许同一 Admin-2 单元内城市地区和农村地区的系数不同，引入模型设定错误。这种设定允许城乡人口之间存在异质性：

$$\log(\mu_{c,k}^{s_4}) = \alpha^{(s)} + \delta_{\text{age}(k)}^{(s)} + \delta_{\text{educ}(k)}^{(s)} + \boldsymbol{\beta}^{(s)} \boldsymbol{X}_{j[k]} + e_k + e_c,$$

其中， $s \in$ {城市，农村}。

虽然情景 4 比此前的情景更接近现实，但仍可能无法充分体现实际数据生成过程的复杂性。

在情景 1 和 2 中，合成人口只生成一次，并在各次重复实验中重复使用。相比之下，在情景 3 和 4 中，由于独立的 Admin-2 级效应可能对结果产生显著影响，因此每次重复实验都会生成一个新的合成人口。对于每种情景，我们重复抽取样本 500 次，从而得到 500 个不同的数据集。随后，根据这 500 个数据集的平均结果进行模型拟合并计算性能指标。

表2：模拟研究中比较的模型汇总。

模型	配置	地理索引级别	规格
FH 分解	Fay–Herriot 解聚模型；no MRP.	一级行政区	§2.2.1
FH-MRP 分解	Fay–Herriot 分解模型；多层回归与后分层按年龄和教育（不含城乡）。	一级行政区	第2.2.3节，式(6)
单元分解	单位层级细分模型；分层按城乡。	一级行政区	§2.2.2
单位-MRP 分解	单位层级细分模型；多层回归与后分层按年龄、教育和城乡。	一级行政区	第2.2.3节，公式（7）
直接二级行政区（基准）	调查加权直接估计值 at 二级行政区。	二级行政区	补充材料第2.1.1节
FH二级行政区（基准）	Fay–Herriot 模型拟合直接 at 二级行政区。使用观测协变量。	二级行政区	§2.1.2

3.4 纳入考量的模型与比较指标

我们将四种定义在 Admin-1 层级的分解模型，与两种基于 Admin-2 数据的基准方法进行比较（表2）。分解模型分为 FH 模型和单位层级模型两类，每类又有两个变体：使用或不使用多层回归与后分层（MRP）。所有情况下，潜在结构均定义在 Admin-2 级。对于分解模型，我们使用 inlabru框架将潜在场聚合至 Admin-1 层级，以便将预测值与观测数据关联起来。

FH细分模型使用Admin-1层级的调查加权直接估计值，并在逻辑尺度上建模。FH-MRP变体进一步将多层回归与后分层应用于人口统计群体，但由于数据稀疏，且调查权重已纳入这一分层，因此不考虑城乡维度。单位层级模型以群组层级结果为拟合对象，但非MRP设定仍按城乡分层，以减轻前文所述的偏差。单位层级MRP模型也按年龄、教育和城乡进行多层回归与后分层。

作为基准，我们纳入两种标准方法：使用地理索引至Admin-2层级的数据进行直接调查加权估计，以及直接在Admin-2层级拟合的FH模型。

所有模型均纳入相同的三项观测协变量，且所有未观测协变量均不用于模型拟合。在非MRP变体中，教育作为区域层级协变量纳入，并以各Admin-2区域至少接受过中等教育的女性比例进行测量。

我们定义了一组指标，用于系统地评估模型在模拟情景中的表现。在空间下推模型的背景下，我们主要关注模型捕捉一级行政区区域内部变异的能力，而非仅关注全局或跨区域变异。具体而言，我们旨在确定模型能否有效利用协变量信息、空间结构和人口构成，捕捉同一二级行政区内各区域的相对排名和异质性。

因此，我们首先关注两项一级行政区内部指标：基于回归的 R^2 和基于相关性的度量。我们将二级行政区 j 的真实总生育率记为 μ_j，相应的模型预测值记为 $\hat{\mu}_j$。每个一级行政区组的真实总生育率为 $\bar{\mu}_i = \frac{1}{n_i} \sum_{j:\ i[j]=i} \mu_j$。

一级行政区内回归 R^2： 对于每个一级行政区区域，我们计算误差平方和（SSE）和组内总平方和（$\text{SST}_{\text{within}}$）：

$$\text{SSE} = \sum_{i \in \mathcal{I}} \sum_{j:i[j]=i} \left(\mu_j - \hat{\mu}_j\right)^2, \qquad \text{SST}_{\text{within}} = \sum_{i \in \mathcal{I}} \sum_{j:i[j]=i} \left(\mu_j - \bar{\mu}_i\right)^2.$$

然后计算 Admin-1 内回归 R^2 如下

$$R^2_{\text{within-reg}} = 1 - \frac{\text{SSE}}{\text{SST}_{\text{within}}}.$$

该指标量化了模型预测所解释的真实结果中 Admin-1 内方差的比例，而非所有二级行政区的总方差。

一级行政区内的皮尔逊相关系数：我们使用皮尔逊相关系数评估二级行政区预测值的排序。在每个一级行政区内，真实总生育率与预测总生育率之间的相关系数为

$$r_i \;=\; \mathrm{corr}\big(\mu_j, \hat{\mu}_j \mid i[j] = i\big).$$

总体指标通过对所有一级行政区的这些相关系数取平均，再将平均相关系数平方得到:

$$r_{\text{within-corr}} = \bar{r} \;=\; \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}} r_i.$$

This 基于相关性的度量强调了行政区内预测值的排序准确性 -1 区域ns.

不确定性指标：我们还评估模型所提供的不确定性量化的质量。考察三项指标：预测区间的平均宽度、这些区间的频率学派覆盖率，以及 Gneiting 和 Raftery（2007）定义的区间评分（IS）。

令 $(l_j,\ u_j)$ 表示来自 $\hat{}\mu_j$ 后验预测分布的不确定性区间。具体而言，下界和上界分别对应 $\alpha/2$ 和 $1-\alpha/2$ 分位数。根据 Gneiting 和 Raftery（2007）提出的区间评分（IS）评估区间质量，该评分会惩罚过宽的区间以及未能包含观测值的区间。对于一组区间$\{(l_{ij},\ u_{ij})\}$，其评分为

$$\mathrm{IS}_\alpha \;=\; \frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}}\Big[\,u_j - l_j + \tfrac{\alpha}{2}(l_j-\mu_j)\,\mathbf{I}\big(l_j>\mu_j\big) + \tfrac{\alpha}{2}(\mu_j-u_j)\,\mathbf{I}\big(u_j<\mu_j\big)\Big].$$

该指标会同时惩罚区间过宽以及未能包含观测值的区间 d valu。得分越低，模型性能越好。

3.5 模拟结果

我们展示了所提出模型以及基准模型在四种预先设定情景下的性能。除直接估计不纳入协变量外，所有模型均使用同一组观测协变量进行拟合。图1对这些模型进行了总结，更多细节请参见补充材料中的S4节。总之，我们提出的模型包括FH模型和单位层级分解模型，且每种模型都分别考虑了有无MRP扩展的情况。这些模型基于在一级行政区层级进行地理索引且移除了二级行政区标识符的数据。基准模型包括FH模型和直接估计，所使用的数据则在二级行政区层级进行地理索引。

每项实验重复进行500次，以减小调查过程中随机抽样变异所引入的噪声。随后对500次重复实验的指标取平均值，并使用不确定性误差线来展示这些指标的变异性。

比较分解模型与基准模型时，可以看出一个明显趋势：随着情景变得越来越复杂、越来越接近现实条件，分解模型的表现逐渐下降。在最后两个情景中，FH二级行政区模型优于细分方法。这一规律与情景设计相符。从情景1到情景4，我们依次引入了分解模型较难解释的变异来源：观测协变量、未观测协变量（代表空间相关效应）、IID Admin-2 级随机效应以及模型设定误差。最显著的转折点出现在引入 IID Admin-2 随机效应时；由于地理索引数据具有足够的分辨率，基准模型能够捕捉这些效应。相比之下，分解模型仅依赖一级行政区信息时，无法生成高保真估计值。

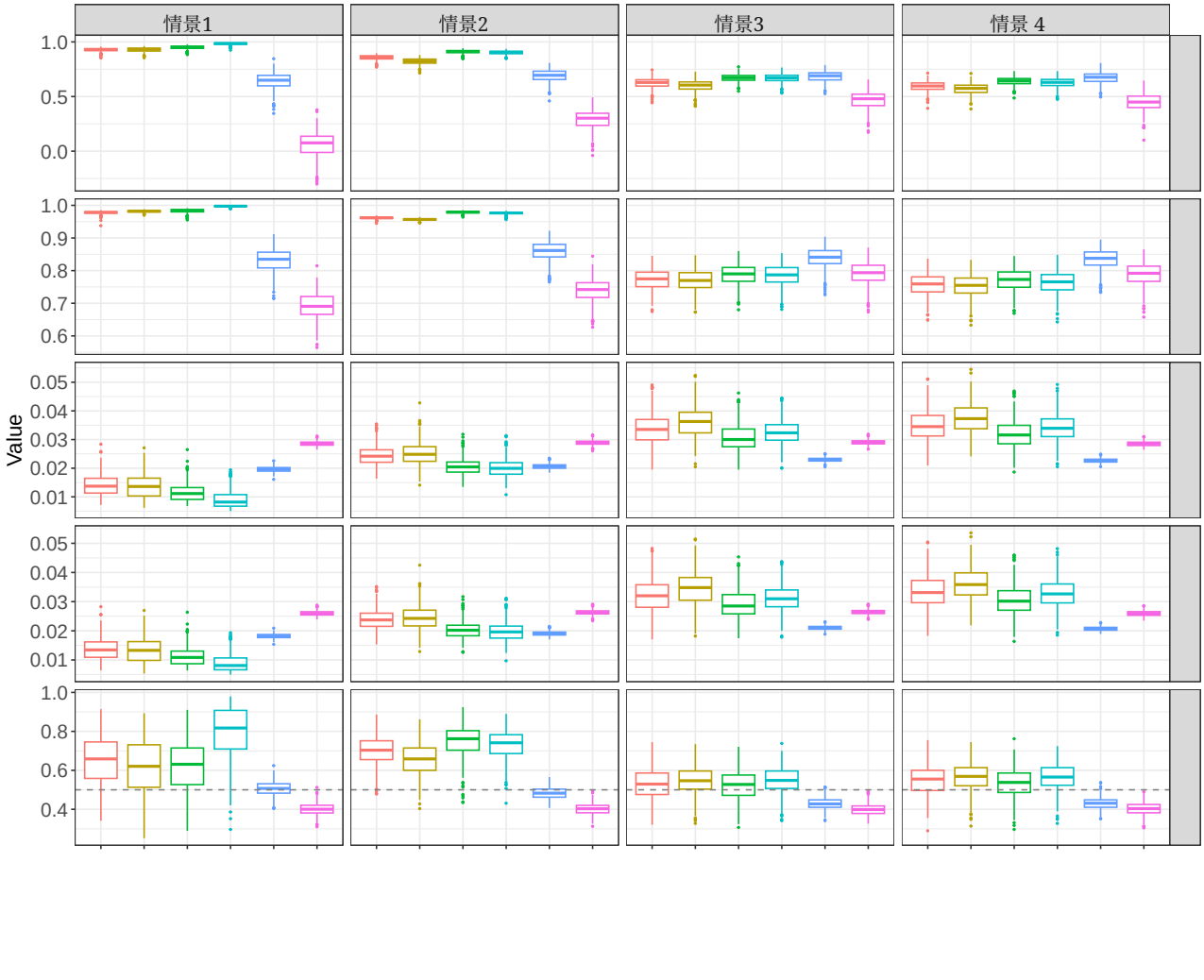


图1：四种模拟情景下分解模型与基准模型（FH 和 Admin-2 级直接估计值）的比较，并使用 Admin-1 内回归、 R^2；皮尔逊相关系数、 r；不确定性 IS；宽度；以及覆盖率进行评估。情景1包括个体层级和区域层级的观测协变量及随机效应。情景2在情景1的基础上增加了两项假定为未观测的区域层级指标。情景3在情景2的基础上增加了独立同分布的Admin-2级随机效应。情景4是最复杂的情景，通过允许同一Admin-2单元内城市和农村地区的系数不同，引入了模型设定错误。

基准模型在情景 1 和 2 中表现较弱，可以用其倾向于拟合二级行政区层级的独立随机效应来解释，尽管这些情景中并不存在此类效应。这一问题在直接估计中尤为明显，而 FH二级行政区模型则受益于协变量和平滑处理，能够在一定程度上稳定结果。另一方面，分解模型在情景 1 和 2 中表现异常出色，其一级行政区内部的 R^2 和 Pearson 的 r 接近于一，对此应谨慎解读。选择这些情景，是为了突出分解模型预期表现良好的情形，并为评估其表现提供清晰的基准。

来看不确定性度量，在情景 3 和 4 中，引入无法识别的二级行政区随机效应后，分解模型的不确定性有所增加。这是预期之中的，因为这些模型实际上是在从一级行政区外推到二级行政区。外推时略微倾向于过度覆盖是可取的，因为这体现了对不确定性作出适当保守的评估。与基准模型相比，分解模型在最后两种情景中的 IS 值更高。同样，这反映了外推所面临的挑战。总体而言，在 500 次重复实验中的变异性对于分解模型也更大，从而显示出它们对随机抽样变异的敏感性。这也是预期之中的，因为

这些模型中的预测值更加依赖观测数据。

在分解模型内部进行比较后发现，单位层面方法在各种情景下的表现都略优于 FH模型，但优势幅度很小。就 MRP扩展的影响而言，MRP变体在情景 1 和 2 中略占优势，但在情景 3 和 4 中效果较差。一个合理的解释是，MRP变体试图将独立同分布的二级行政区随机效应所产生的部分未解释变异归因于人口统计学协变量，从而引入额外的变异。对于 FH模型，由于模型是在汇总层面拟合的，而后分层所用的协变量处于个体层级，因此 MRP扩展也可能受到生态偏差的影响。S4节所示偏差指标的升高值体现了这种偏差。然而，对于单位层级模型，通过 MRP进行后分层也有助于抵消人口统计变量方面信息性抽样带来的潜在偏差，因为所提出的单位层级模型无法纳入抽样设计权重。

总而言之，结果表明，分解模型表现合理。它们能够解释一级行政区内部的变异，这一点反映在 R^2 和 Pearson's r中；其不确定性量化效果也不比基准模型差太多。同时，模拟结果凸显了分解模型在面对无法识别的二级行政区变异时的局限性。基于模拟研究的发现，我们在后续真实数据应用中重点考察两个模型：不含 MRP 扩展的 FH 模型，以及含 MRP 扩展的单位层级模型。

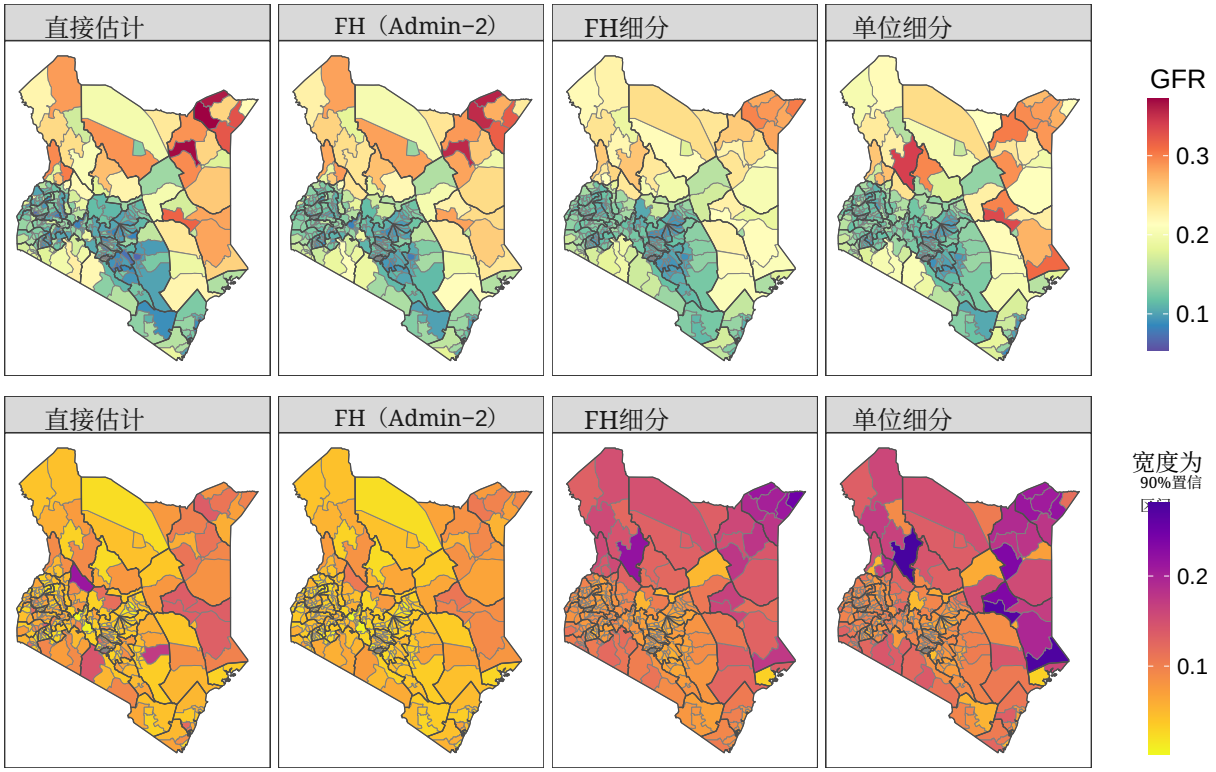


图2：基于2022年人口与健康调查数据，通过四种模型获得的肯尼亚二级行政区总生育率估计值：直接估计、二级行政区 FH、FH细分，以及结合 MRP 的单位层级分解。上方面板显示总生育率点估计，下方面板显示90%不确定性区间的宽度。

4 案例研究1：肯尼亚的 GFR指标（2022年人口与健康调查）

我们的模拟研究旨在模拟基于2022年人口与健康调查的肯尼亚总生育率数据生成与抽样过程。在本案例研究中，我们将分解模型应用于实际的2022年人口与健康调查数据。由于该调查提供了抽样簇的GPS坐标，因此数据具有二级行政区层面的地理索引。我们评估了四种模型：两种直接使用二级行政区索引的基准模型（直接估计，以及带协变量的FH模型），以及两种

假设仅提供一级行政区索引的分解模型（不采用多层回归与后分层的FH模型，以及采用多层回归与后分层的单位层级模型）。所用协变量包括夜间灯光强度、前往最近医疗机构的出行时间、平均家庭规模、手机使用情况，以及中学教育完成率。

目标是评估在掩盖Admin-2标识符的情况下，分解模型的表现，包括将其与基准模型进行比较，以及相互之间进行比较。

图2显示，分解模型生成的Admin-2 级总生育率估计值总体上与基准模型一致。在Admin-2 FH模型估计区域层面随机效应较大的区域，差异更为明显；这与模拟研究的发现一致，即分解模型在设计上无法完全恢复这些效应（补充材料的S5.1节提供了更多细节）。

以Admin-2 FH模型作为基准，一级行政区内部的分解估计值与基准估计值之间的皮尔逊相关系数r 平均而言，FH模型为0.52，单元级模型为0.55，表明两者具有较为合理的一致性（在全球层面，所有二级行政区单元之间的相关系数均超过0.9）。不过，基准模型估计值本身仍存在不可忽视的不确定性，因此应谨慎解读这些比较结果。

总体而言，分解模型的不确定性区间更宽，这与模拟研究的结果一致，因为分解必须从一级行政区外推至二级行政区。相比之下，基准模型的区间窄得多，因为它们直接利用了二级行政区的信息；不过，若仅采用一级行政区索引，就无法达到这种精度。

总体而言，该案例研究表明，当只能获得一级行政区地理索引时，分解模型仍能以合理的精度提供二级行政区级别的总生育率估计值。其代价是得到更宽、但也更为审慎的不确定性区间。

5 案例研究 2：来自 2021年KTUS 的指标

现在我们来看一个案例：由于数据仅按较粗的行政级别进行地理索引，而非按目标级别索引，因此分解方法必不可少。我们的应用使用了 2021年KTUS，该调查作为肯尼亚连续住户调查的一个模块实施。与人口与健康调查类似，KTUS采用了两阶段分层设计，并以肯尼亚2019年人口普查为主抽样框。与人口与健康调查不同，KTUS不提供GPS坐标，因此个体记录仅关联到一级行政区，无法开展标准的二级行政区分析。由于政策制定和可持续发展目标监测需要更精细的次国家级细节，尤其是对于我们下文讨论的此类指标，分解模型是唯一可行的选择。

5.1 无偿照护或家务劳动

所考虑的第一个指标是用于无偿照护或家务劳动的时间，这是可持续发展目标具体目标5.4.1的核心。受访者报告了前一天各项活动所花费的分钟数，我们将其四舍五入为半小时单位。我们采用以当天半小时总数为暴露量的负二项模型，并比较三种方法：（1）一级行政区的直接估计；（2）不含多层回归与后分层的FH细分模型；（3）使用年龄组和城乡状况进行后分层、并采用多层回归与后分层的单位层级细分模型。协变量包括夜间灯光强度、到最近医疗机构的出行时间、平均家庭规模和高中教育完成率。

图3显示了女性一天中用于无偿照护和家务劳动的估计比例。结果在部分区域约为15%（3.5–4小时），在其他区域则超过30%（超过7小时）。内罗毕和蒙巴萨的估计值最低，反映出更好的服务可及性和正规就业，而北部和东北部的估计值最高。除这些总体模式外，分解模型进一步表明一级行政区内部存在显著异质性。两种分解模型均突出显示了马萨比特和瓦吉尔境内靠近埃塞俄比亚边境的特定二级行政区，这些地区的女性将超过30%的时间用于无偿照护或家务劳动。两种分解模型得出的估计值一致，其中单位层级模型产生的空间模式更为平滑。分解模型的不确定性区间宽于一级行政区直接估计值，这反映了从粗地理层级向细地理层级的外推。男性的具体结果见补充材料中的图S5。总体而言，男性一天中用于无偿照护或家务劳动的时间不足6%（不到1.5小时）。

为考察性别差距，我们计算女性与男性在无偿家务和照护工作时长上的差值，并绘制女性比男性多花四小时以上的概率分布图

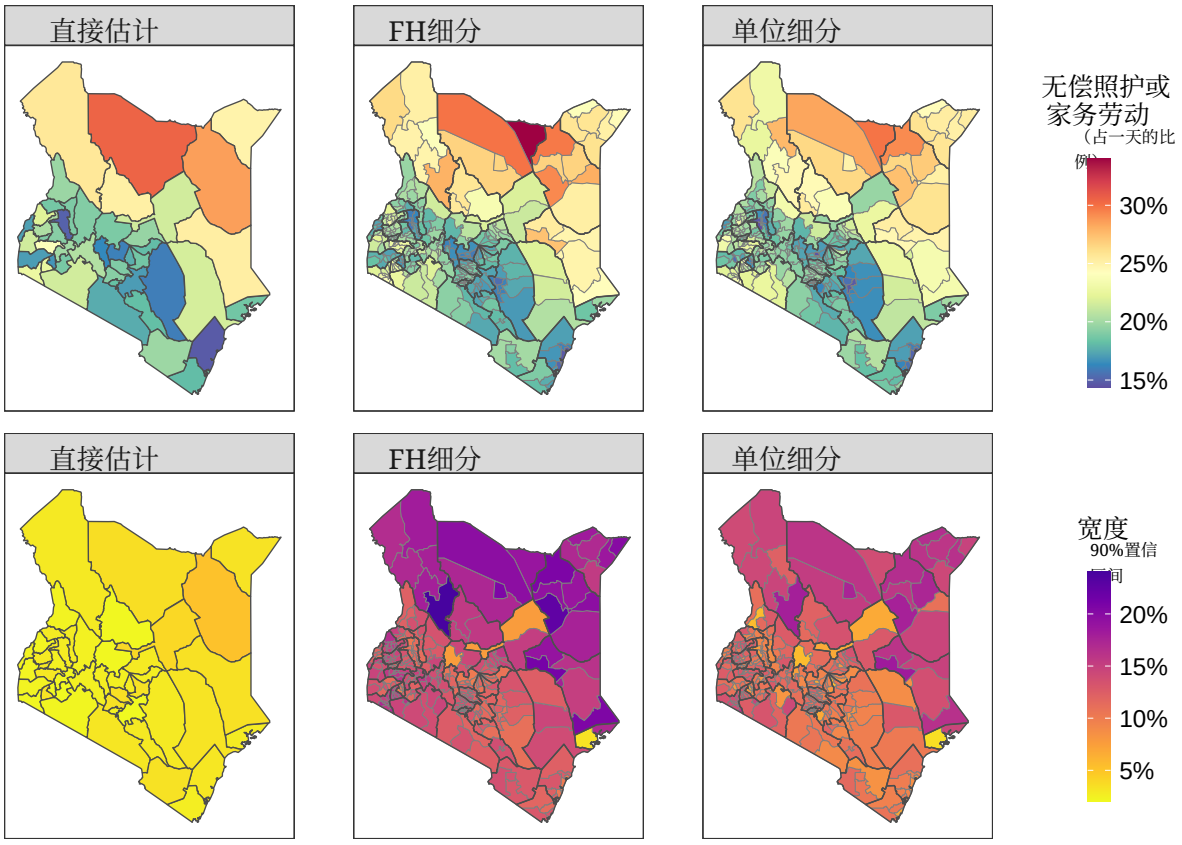


图3：通过一级行政区直接估计、FH细分和单位层级分解模型，估算女性用于无偿照护和家务劳动的一天中的比例。地图展示了点估计及相应90%不确定性区间的宽度。

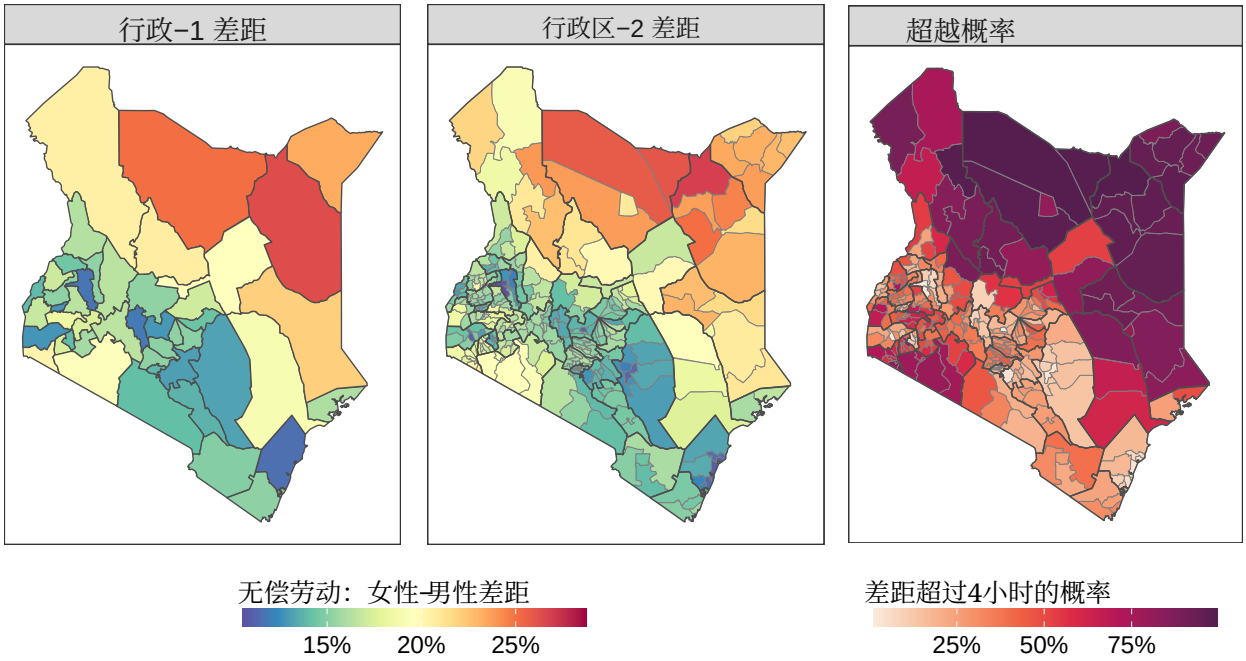


图4：无偿照护和家务劳动中女性与男性差距的地图，以及每日差距超过四小时的相应概率。一级行政区估计值基于直接估计，二级行政区估计值则来自结合多层回归与后分层的单位层级分解模型。

女性每天用于此类活动的时间比男性多（相当于一天的16.7%）。这些超越概率根据单位层级细分模型的后验抽样得出，并展示于图4中。结果呈现出清晰的空间模式：北部和东北部区域的概率始终高于75%，表明有充分证据显示这些地区存在较大的性别差距。相比之下，中部高地和沿海地区（如蒙巴萨和基利菲）的概率要低得多，部分地区甚至低于25%。分解揭示了更精细的空间模式，其中包括聚合估计值无法显示的局部热点区域。这些地区的男女差距尤为显著，凸显了为推动可持续发展目标具体目标5.4.1的进展而采取针对性空间干预措施的必要性。

5.2 大众媒体使用

现在，我们考察一个二元指标，以说明该建模框架同样适用于二元结果，而不仅仅适用于比率结果（例如总生育率）。在本例中，我们关注大众媒体使用情况，其定义为受访者前一天是否花过时间接触大众媒体。大众媒体是指触达大量受众的传播渠道，例如广播、电视、报纸和在线新闻平台。大众媒体使用是衡量信息获取和社会参与的重要指标，与公共卫生宣传、公民参与以及若干可持续发展目标具体目标的进展密切相关。

结局采用第2.2.2节所述的贝塔-二项分布设定进行建模。与前一种情况一样，我们比较了三种方法：Admin-1层级的直接估计、FH 分解，以及 Admin-2层级的单位层级分解。所用协变量包括夜间灯光强度、高中入学率、手机使用和医疗机构可及性。

图5展示了三种建模方法得出的大众媒体使用概率的男性和女性特定估计值。在肯尼亚，男性报告的大众媒体使用率普遍高于女性。从空间分布来看，内罗毕、蒙巴萨及其他城市中心的二级行政区中，男女的大众媒体使用率都较高；而北部地区的比率则明显较低。值得注意的是，分解模型能够还原二级行政区内部的异质性，而一级行政区直接估计值无法捕捉这种差异。由于该指标为二元变量，且调查样本的变异性较大，因此不确定性区间比此前的分析更宽。如补充材料中的图 S6所示，分解模型的可信区间反映了增加的不确定性，也凸显了在对细尺度差异作出推断时保持谨慎的必要性。

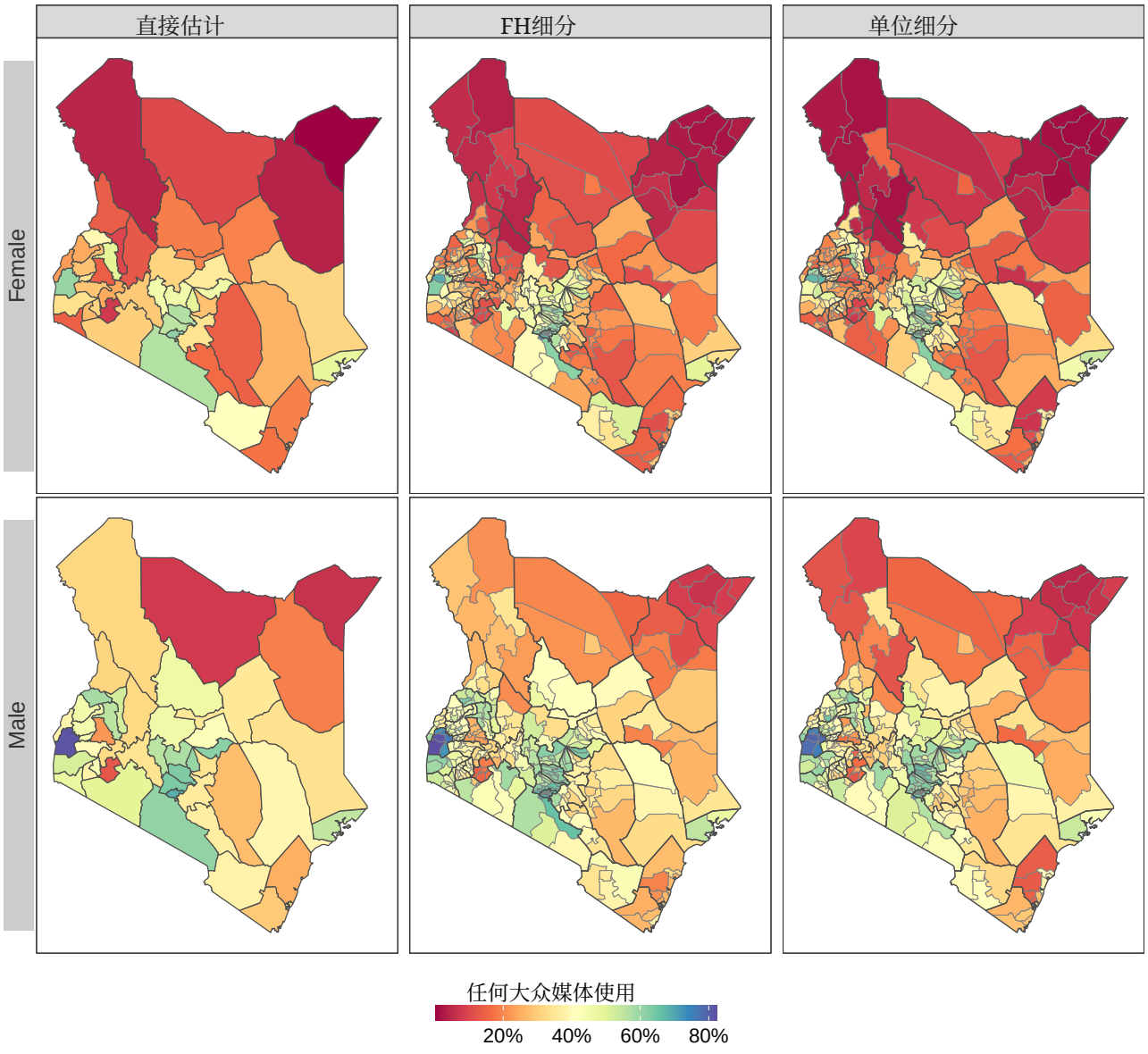


图5：大众媒体使用的女性和男性特定估计值地图。

6 讨论

本研究在小区域估计框架下提出了一系列分解模型，用于在调查数据仅能以较粗的行政层级获取时，获得精细尺度的区域估计值。这些方法基于两种标准的小区域估计方法——FH模型和单位层级模型，并纳入了多层回归与后分层扩展。该框架旨在利用三种关键信息来源：区域层级协变量、空间平滑和人口构成，将粗尺度的地理参照调查数据与在更精细目标分辨率上定义的潜在场联系起来。计算通过 inlabru 中的近似贝叶斯推断高效完成。

针对肯尼亚总生育率的模拟研究模拟了2022年DHS，并将以一级行政区为地理索引的分解模型与以二级行政区为地理索引的基准模型进行了比较。结果清楚地表明了分解何时有效，以及何时效用较为有限。当结局变异由观测协变量驱动，或由观测协变量与空间相关的未观测协变量共同驱动时，这些模型能够以较高准确度恢复二级行政区层级的Admin-1内部异质性。然而，引入非结构化Admin-2随机效应，或系数以模型未考虑的方式随层变化时，模型表现会大幅下降。分解模型的不确定性区间通常更宽；考虑到这项任务具有外推性质，这种保守性是可取的。多层回归与后分层扩展对预测能力的提升作用不大，但在单位层级情境中，通过纳入调查设计发挥了有益作用。总体而言，在无法获得精细尺度地理索引时，分解模型的表现与基准模型相当。根据模拟研究，以下两种模型适用于实际应用：不含多层回归与后分层的FH细分模型，以及含多层回归与后分层的单位层级细分模型。

这些案例研究展示了所提出的方法应用于真实调查数据时的实际价值。在DHS 2022年案例研究中，由于可获得用于验证的Admin-2标识符，分解模型得出的估计值与基于Admin-2索引的基准方法大体一致。在2021年KTUS中，分解模型是唯一可行的选择，因为发布的数据仅关联到一级行政区。在这种情况下，这些模型为无偿家务和照护工作、大众媒体使用等关键指标生成了有意义的二级行政区估计值。这些估计值揭示了县内的异质性，并凸显了局部性别差距，而仅进行直接的一级行政区分析则无法发现这些差距。总体而言，这些案例研究表明，分解能够为局部政策监测和可持续发展目标报告提供有价值的次国家层面洞见，尤其是在对精细尺度估计值的需求不断增长的情况下。与此同时，分解结果的不确定性区间较宽，这突显出从此类估计值得出推断时保持谨慎的重要性。

最后，我们总结本研究的局限性。首先，分解从根本上受限于可用信息。以区域层面IID随机效应形式纳入模型的精细尺度非结构化异质性，无法从粗粒度地理参照数据中识别。当结局主要受此类非结构化变异影响时，分解模型预计表现不佳。尽管模拟研究考察了一系列合理情景，但现实世界的数据生成过程可能更加复杂。此外，在分解最具相关性的情境中，即只能获得粗粒度地理索引（例如KTUS）时，无法通过更高分辨率的数据进行直接验证。因此，对分解结果作出的推断始终应谨慎解读。

其次，与应用于地理参考数据充足的情形下的传统小区域估计方法相比，分解模型更加依赖协变量。因此，协变量的选择和质量都更加重要。在本研究中，我们将协变量视为固定输入，未明确考虑其不确定性。在这一空间细分框架中，协变量主要用于预测，而非推断其与结果变量之间的因果关系。正如引言中所述，最有用的人口和健康协变量通常来自人口普查，但人口普查数据并不常见，而且其可获取性存在不确定性，在低收入和中等收入国家尤其如此。

最后，模型设定仍有改进空间。对于单位层级模型，从细粒度潜在场到较粗区域的聚合过程，最准确的表示方式应是混合似然，即根据每个单位属于不同子区域的概率，对其贡献进行聚合。然而，在这一设定下实施完整的贝叶斯推断对计算资源的要求很高。我们的方法采用过度离散近似，能够提供正确的均值表示，而均值正是主要关注的量。尽管在子区域之间存在极端异质性的情况下，这种近似的准确性可能较低，但借助inlabru框架可以显著提高计算效率。此外，我们的模拟结果表明，没有证据显示这种近似会导致系统性偏差或不稳定性。

致谢

本材料基于美国能源部科学办公室高级科学计算研究计划在奖项编号 DE-SC-ERKJ422 下支持的工作。

参考文献

Abate, B. B., Sendekie, A. K., Ayele, M., Lake, E. S., Wodaynew, T., Tilahun, B. D., Azmeraw, M., Habtie, T. E., Kassa, M., Munie, M. A., 等（2024年）。2000年至2019年埃塞俄比亚全国、次国家级和地方层面的生育率制图。《公共卫生前沿》,**12**。

Arambepola, R., Gething, P., 和 Cameron, E.（2020）。用于绘制马达加斯加疟疾发病时空风险图的非参数因果特征选择。*arXiv*预印本 *arXiv:2001.07745* 。

Bachl, F. E., Lindgren, F., Borchers, D. L., 和 Illian, J. B.（2019年）。inlabru：一款基于生态调查数据进行贝叶斯空间建模的 R 包。《生态学与进化方法》,**10**(6), 760–766。

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D. 和 Lindgren, F.（2018）。使用 R-INLA 进行空间建模：综述。《*Wiley Interdisciplinary Reviews: Computational Statistics*》，**10**（6），e1443。Battese, G. E., Harter, R. T. 和 Fuller, W. A.（1988）。利用调查数据和卫星数据预测县级作物面积的误差成分模型。《美国统计学会杂志》，**83**（401），28–36。Besag, J., York, J. 和 Molli´e, A.（1991）。贝叶斯图像复原及其在空间统计学中的两项应用。《数理统计学年刊》，**43**，1–20。Delprato, M., Chudgar, A. 和 Frola, A.（2024）。撒哈拉以南非洲成就指标的空间教育不平等及其溢出效应。《世界发展》，**176**，106522。Didan, K., Munoz, A. B., Solano, R., Huete, A., 等（2015）。MODIS植被指数用户指南（MOD13系列）。亚利桑那大学：植被指数与物候实验室，**35**，2–33。Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C. 和 Worley, B. A.（2000）。LandScan：用于估计风险人口的全球人口数据库。《摄影测量工程与遥感》，**66**（7），849–857。Downes, M., Gurrin, L. C., English, D. R., Pirkis, J., Currier, D. 和 Spittal, M. J.（2018）。使用后分层多层回归进行健康指标小区域估计：慢性病案例研究。《医学研究中的统计方法》，**27**（11），3291–3305。Fay, R. E. 和 Herriot, R. A.（1979）。小地区收入估计：詹姆斯–斯坦方法在人口普查数据中的应用。《美国统计学会杂志》，**74**（366a），269–277。Gneiting, T. 和 Raftery, A. E.（2007）。严格适当的评分规则、预测与估计。《美国统计学会杂志》，**102**，359–378。H´ajek, J.（1971）。对 D. Basu 所著《抽样调查逻辑基础论述，第一部分》的讨论。载于 V. Godambe 和 D. Sprott（编），《统计推断基础》。多伦多：Holt, Rinehart and Winston。Horvitz, D. G. 和 Thompson, D. J.（1952）。有限总体不放回抽样的一种推广。《美国统计学会杂志》，**47**（260），663–685。Janocha, B., Donohue, R., Fish, T., Mayala, B. 和 Croft, T.（2021）。次国家级二级行政区指标估计值使用指南与建议。技术报告，ICF International。DHS空间分析报告第20号。J¨onsson, K. 和 Bexell, M.（2021）。本地化可持续发展目标：坦桑尼亚案例。《发展政策评论》，**39**（2），181–196。

肯尼亚国家统计局（2023）。《肯尼亚时间利用报告》（基于2021年肯尼亚连续住户调查）。https://www.knbs.or.ke/wp-content/uploads/2023/10/ Kenya-Time-Use-Survey-2021_1.pdf。访问日期：2025-10-02。Lindgren, F. 和 Rue, H.（2015）。使用 R-INLA 进行贝叶斯空间建模。《统计软件杂志》，**63**，1–25。Lindgren, F., Rue, H. 和 Lindstr¨om, J.（2011）。高斯场与高斯马尔可夫随机场之间的显式联系：随机偏微分方程方法。《皇家统计学会会刊：*B*辑（统计方法论）》，**73**（4），423–498。Lindgren, F., Bachl, F., Illian, J., Suen, M. H., Rue, H. 和 Seaton, A. E.（2024）。inlabru：用于拟合带非线性预测变量的潜在高斯模型的软件。*arXiv*预印本 *arXiv:2407.00791*。

Lindgren, F., Bachl, F. E., Borchers, D. L., Simpson, D., Scott-Howard, L., Seaton, A., Suen, M. H., Roudier, P., Meehan, T., Niharika, P., 和 Perepolkin, D. (2025). *inlabru*：使用 *INLA* 及其扩展进行贝叶斯潜在高斯建模。R 软件包版本 2.13.0.9011。

Morales, D., Santamar´ıa, L., 和 Molina, I. (2021). 小区域患病率的基于模型与基于设计的估计量：西班牙吸烟案例。《国际统计评论》，**89**(1), 28–50。

Moretti, A. 和 Whitworth, A.(2021). 使用调查数据进行患病率制图的小区域估计模型比较。《医学研究中的统计方法》，**30**(9), 2125–2141。

Nandi, A. K., Lucas, T.C., Arambepola, R., Gething, P., and Weiss, D. J. (2023年). 分解：用于贝叶斯空间分解建模的 R 包。《统计软件杂志》，**106**，1–19。

Oosterhof, P. D. (2018). 将可持续发展目标本地化，以加快《2030年可持续发展议程》的实施：亚太地区可持续发展目标本地化的现状。

Park, D. K., Gelman, A., Bafumi, J., and Kaplan, A. (2004). 结合事后分层的贝叶斯多层估计：基于全国民意调查的州级估计。《政治分析》，**12**(4)，375–385。

Pratesi,M. (2016). 基于小区域估计的贫困数据分析。John Wiley & Sons。

Python, A., Marshall, J. D., Clark, S. J., and Mosser, J. F. (2022年). 使用分解回归绘制 COVID-19 负担的高分辨率地图。《自然·通讯》，**13**(1)，1234。

Rao, J. N. K. 和 Molina, I.（2015）。小区域估计。John Wiley & Sons。

Requia, W. J.、Koutrakis, P. 和 Arain, A.（2018）。环境流行病学研究中人口空间分布的建模：比较采用分级设色地图与人口密度分区制图得出的暴露估计。《国际环境》，**119**，152–164。

Riebler, A., Sørbye, S., Simpson, D., 和 Rue, H.（2016）。一种直观的贝叶斯空间模型，用于考虑尺度调整的疾病制图。《医学研究中的统计方法》，**25**，1145–1165。Rom´an, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., Schultz, L., Bell, J., Stokes, E. C., Pandey, B., 等（2018）。美国国家航空航天局黑色大理石夜间灯光产品系列。《环境遥感》，**210**，113–143。Rue, H., Martino, S., 和 Chopin, N.（2009）。利用积分嵌套拉普拉斯近似对潜在高斯模型进行近似贝叶斯推断。《英国皇家统计学会杂志：*B*辑》，**71**(2)，319–392。Simpson, D., Rue, H., Riebler, A., Martins, T., 和 Sørbye, S.（2017）。惩罚模型组件复杂度：构建先验分布的一种有原则且实用的方法。《统计科学》，**32**，1–28。Smith, D. M., Heppenstall, A., 和 Campbell, M.（2021）。跨时空估计健康状况：空间微观模拟在公共卫生中的应用综述。*J*，**4**(2)，182–192。

Suen, M. H., Naylor, M., Mudd, S., 和 Lindgren, F.（2025）。河流下切对尼泊尔廓尔喀地震触发滑坡的影响：基于 inlabru 的像元易发性模型结果。*arXiv* 预印本 *arXiv:2507.08742*。Tanton, R.（2014）。空间微观模拟方法综述。《国际微观模拟期刊》，**7**(1)，4–25。Tatem, A. J.（2017）。WorldPop：用于空间人口学的开放数据。《科学数据》，**4**(1)，1–4。联合国统计司（2020）。可持续发展目标指标元数据：5.4.1 按性别、年龄和地点划分的无偿家务和照护工作时间占比。https://unstats.un.org/sdgs/ metadata/files/Metadata-05-04-01.pdf。访问日期：2025-10-02。Utazi, C. E., Nilsen, K., Pannell, O., Dotse-Gborgbortsi, W., 和 Tatem, A. J.（2021）。疫苗接种覆盖率的地区级估计：离散空间模型与连续空间模型的比较。《医学统计学》，**40**，2197–2211。Wakefield, J., Gao, P., Fuglstad, G.-A., 和 Li, Z. R.（2025）。患病率制图的两种范式：小区域估计与空间统计学。*arXiv* 预印本 *arXiv:2110.09576*。Weiss, D., Nelson, A., Vargas-Ruiz, C., Gligori´c, K., Bavadekar, S., Gabrilovich, E., Bertozzi-Villa, A., Rozier, J., Gibson, H., Shekel, T., 等（2020）。医疗设施出行时间全球地图。《自然·医学》，**26**，1835–1838。Wu, Y. 和 Wakefield, J.（2024）。中低收入国家城乡比例建模。《皇家统计学会会刊 *A*辑》，**187**，811–830。Wu, Y., Li, Z. R., Mayala, B., Wang, H., Gao, P., Paige, J., Fuglstad, G.-A., Moe, C., Godwin, J., Donohue, R., Croft, T., 和 Wakefield, J.（2021）。用于 Admin-2 小区域估计的时空建模。技术报告，ICF International。《DHS空间分析报告第21号》。Wu, Y., Dharamshi, A., 和 Wakefield, J.（2025）。中低收入国家教育水平的小区域估计。*arXiv* 预印本 *arXiv:2502.07946*。

《区域分解：小区域估计视角》的补充材料

S1 关于 inlabru线性化下恒等链接的局限

或许有人会考虑单位层级细分模型的另一种表述。由于数据模型中的位置参数 μ_i 本身就是分布的均值，因此看起来可以将其直接等同于预测变量 η_i，从而省去基于广义线性模型的设定中通常使用的链接函数。按照这种思路，预测变量中使用的外部变换（例如 log(·) 在 log(∑exp(·)) 中）似乎可以与抽样模型中的链接函数相互抵消。要说明这种做法为何不正确，可以考察泊松情形；同样的论证也适用于负二项和（二项）二项似然。

	有问题	**正确**
抽样模型	$Y_i \sim$ 泊松分布$(n_i\,\mu_i)$	$Y_i \sim$ 泊松分布$(n_i\,\mu_i)$
Link	$\mu_i = \eta_i$ (恒等链接)	对数 $\mu_i = \eta_i$ (对数链接)
预测变量	$\eta_i = \sum_j \exp(\boldsymbol{A}\boldsymbol{u}_j)$	$\eta_i =$ 对数$\Big(\sum_j \exp(\boldsymbol{A}\boldsymbol{u}_j)\Big)$

左侧（上方）的表达式体现了聚合预测变量中的对数与对数链接中的指数之间表面上存在、但实际上具有误导性的抵消关系。尽管这种设置在代数上颇具吸引力，但它丧失了链接函数的一项重要功能：它允许 η_i 在 $\mathbb{R}$ 上不受约束，同时确保 $\mu_i > 0$ 为有效均值。

在 inlabru框架中，这一作用尤为关键，因为诸如

$$\eta_i \;=\; \log\Big(\sum_j \exp(\boldsymbol{A}\boldsymbol{u}_j)\Big)$$

的预测变量会被线性化，以便进行推断。如果改为将

$$\sum_j \exp(\boldsymbol{A}\boldsymbol{u}_j) \;\approx\; \tilde{\boldsymbol{A}}\boldsymbol{u} + \delta(\boldsymbol{u}_0)$$

并使用恒等链接，那么无法保证该近似在潜在高斯场 $\boldsymbol{u}$ 的整个分布支撑上始终严格为非负值。尽管函数值在该线性化点应保持为正，但近似在分布支撑的尾部可能产生负值，这将与泊松似然（以及二项模型的约束）不相容。

相比之下，保留对数链接可确保即使在线性化之后，预测变量 η_i 仍不受约束，同时变换 exp(η_i) 保持了 μ_i所需的正性。因此，不能采用代数上更简单的表述，因为它会破坏 inlabru 所采用的线性化策略的有效性。

S2 SpatialModel的模型设定

区域的空间效应：术语 b_j 表示区域 j的空间随机效应，聚类 c 位于该区域。我们采用 BYM2模型（Riebler 等，2016），它是 Besag-York-Mollie（BYM）模型（Besag 等，1991）的一种重新参数化形式，并将 b_j 分解为独立同分布（IID）分量 e_j 和空间结构分量 S_j:

$$b_j = \sigma_b(\sqrt{1-\kappa}e_j + \sqrt{\kappa}S_j),$$

其中 $e_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$，且 $S = (S_1,\ldots,S_n)^\top$ 服从由邻接矩阵 W（基于各区域的地理构型）构建的缩放内蕴条件自回归先验。记$j \sim q$ 为空间邻接、 m_j 为邻居数量，则区域 j 的条件分布为

$$S_j \,\big|\, \{S_q : q \sim j\} \sim \mathcal{N}\left(\frac{1}{m_j}\sum_{q\sim j} S_q,\ \frac{\sigma_s^2}{m_j}\right),$$

并带有和为零约束 $\sum_j S_j = 0$。我们对超参数 σ_b （总标准差）和 κ （空间变异所占比例）使用惩罚复杂度先验（Simpson 等，2017），使得$\Pr(\sigma_b > 1) = 0.01$ 且$\Pr(\kappa > 0.5) = 0.5$。我们的先验分布设定主要遵循R-INLA和inlabru中的默认设置。

后验推断使用 inlabru 软件包（Lindgren 等，2024；Bachl 等，2019）进行，详情见主文稿第2.3节。

S3 模拟设置的补充详情

S3.1 数据来源与输入构建

模拟利用公开的地理空间协变量和行政边界数据构建真实的主抽样框。首先，我们从肯尼亚独立选举与边界委员会获取一级行政区和二级行政区级别的分析用形状文件，并将其名称和代码与人口普查及 WorldPop 图层统一，以确保地理索引一致。

其次，我们从2019年肯尼亚人口与住房普查中提取二级行政区层级的社会人口学边际分布，包括城镇人口比例、接受过中等或更高教育的女性比例，以及选取的其他区域层面协变量（例如平均家庭规模、手机使用条件）。由于人口普查统计单元与我们分析中使用的标准二级行政区边界并不完全匹配，我们使用将普查单元映射到分析单元的对应表，使两者保持一致。此映射主要用于转换比例或比率指标。人口数取自WorldPop（下文将介绍），并使用人口普查中的比例进行分配。对应表的构建详情及注意事项将在下一节中完整说明。

第三，我们使用WorldPop的2019年栅格人口估算，获取1 km × 1km网格上按5岁年龄组划分的女性人数。我们利用分析用形状文件将这些人数汇总到二级行政区层级，并生成15–19岁、……、40–44岁各年龄组的分年龄区域总数。这些总数作为基准人口。

第四，我们整理了其他 1 km 栅格协变量，以刻画生育率的空间异质性：夜间灯光（Rom´an*et al.*，2018年）、到最近医疗机构的出行时间（Weiss*et al.*，2020）以及植被指数（NDVI）（Didan *et al.*，2015）。随后，我们计算了这些协变量的按人口加权的二级行政区平均值。

第五，我们从 2022年肯尼亚人口与健康调查（人口与健康调查）中提取关键设计参数：分层方案（二级行政区—城市/农村 [值得注意的是，内罗毕和蒙巴萨完全属于城市地区]）、按层划分的抽样框中普查区（EAs）和初级抽样单位（PSUs）数量，以及每个普查区和每个普查区内女性的目标抽样数量。这些输入用于复现人口与健康调查的两阶段抽样。

S3.2 人口普查—二级行政区对应表与数据协调 n

2019年肯尼亚人口与住房普查在剔除保护区及其他非居住单元后，报告了 333 个郡以下行政区层级的人口社会学指标。这些统计单元与我们分析中使用的 290 个二级行政区（选区）边界并不完全一致，而且 333 个普查郡以下行政区的官方边界文件也未公开 उपलब्ध。为协调这些定义，我们构建了一个 333×290 对应矩阵，将普查郡以下行政区与模拟框架中使用的二级行政区单元相连接。

交叉对应表是通过一个循序渐进的过程编制的。我们首先将人口普查表与二级行政区 Shapefile 中的单元名称进行匹配，并考虑已知的历史名称变更、其他拼写形式以及格式差异。当无法直接匹配时，我们人工查阅区域记录、行政区划地图和辅助地理元数据，以确定最合理的对应关系。对于人口普查次级县区与多个二级行政区单元相互重叠（或相反）的情况，我们在交叉对应表中允许多对多映射。由于此类关联可能扭曲绝对人口数，我们将交叉对应表的使用限制为传递比例和比率指标，例如城市比例或接受中等及以上教育的女性比例。

绝对人数则来自联合国调整后的 WorldPop 2019年栅格人口估算，数据按女性五岁年龄组划分，分辨率为1公里。这些人数汇总至分析中使用的二级行政区边界，随后利用人口普查得出的比例，按城乡和教育分层进行划分。此方法在保留人口普查所推导的社会人口结构的同时，减轻了潜在人口普查漏计的影响。尽管人口普查总人数漏计对我们推导的比例指标影响甚微，但若直接使用这些总人数，则可能导致人数估计出现偏差；这进一步印证了我们采用 WorldPop 数据获取绝对人口总量的决定。

S3.3 生成主抽样框与样本

步骤 1：个体名录。我们将二级行政区-2 × 年龄组 × 城乡属性 ×教育单元扩展为由15–19岁、20–24岁，……，40–44岁年龄组女性个体构成的合成人口。我们在 S3.4节讨论暴露量和结局的生成方法。

步骤2：普查区分配。 人口与健康调查报告包含抽样主框架中每个层 h （Admin-1 × 城乡）内普查区数量的数据。我们按各Admin-2 ×城乡单元在每个层内的女性人口占比，向其分配相应数量的普查区。为引入符合实际的聚类规模异质性，我们从对数为 $S \sim \mathcal{N}$（0，0.2₂）的对数正态分布中抽取普查区相对规模S。随后，我们依据这些相对规模（比例分配），将每个层中的女性个体划分到各普查区，以保持二级行政区归属。

步骤3：两阶段样本选取（模拟人口与健康调查）。 我们采用分层两阶段抽样来模拟人口与健康调查的设计。在第一阶段，于每个层 h内，我们以与规模成比例的概率（PPS）抽样选取 n_h 个抽样簇（普查区），其中规模测度为该聚类中的女性人数。令 N_{hc} 表示层 h 中聚类 c 的这一规模测度，并令$N_{h\cdot} = \sum_c N_{hc}$。第一阶段纳入概率为

$$\pi_{hc}^{(1)} = \frac{n_h \, N_{hc}}{N_{h\cdot}}$$

在第二阶段，我们从每个抽中的聚类 c 中，通过不放回简单随机抽样选取 m_{hc} 名女性。我们将 $m_{hc} = 20$设定为接近2022年肯尼亚人口与健康调查的样本量。随后，在聚类 c 中选中女性 k 的条件概率为

$$\pi_{k|hc}^{(2)} = \frac{m_{hc}}{N_{hc}}.$$

总体纳入概率为

$$\pi_k = \pi_{hc}^{(1)} \times \pi_{k|hc}^{(2)} = \frac{n_h \, N_{hc}}{N_{h\cdot}} \times \frac{m_{hc}}{N_{hc}} = \frac{n_h \, m_{hc}}{N_{h\cdot}},$$

且基础抽样权重为 $w_k = 1/\pi_k$。我们不模拟或调整无应答。

S3.4 结局生成与模拟情景

表 S1：四种情景下空间异质性的来源，以及区域细分模型捕捉这些异质性的预期能力。

变异来源	**情景1**	**情景2**	**情景3**	**情景 4**	**可恢复**
观测协变量	✓	✓	✓	✓	✓ 明确捕获 由固定效应
未观测协变量 （空间相关的 二级行政区效应）	✗	✓	✓	✓	❍ 部分捕获 由空间效应
IID Admin-2 随机效应	✗	✗	✓	✓	✗ 不可恢复 ， 完全非结构化
变化的 系数	✗	✗	✗	✓	✗ 导致模型 设定错误

情景1： 在第一个情景中，仅纳入观测协变量。位于二级行政区 j 中的聚类 c 内的个体 k 的对数尺度均值结构定义为

$$\begin{aligned}\log \mu_{c,k}^{s_1} = \alpha_{\text{rural}}\, I(c \in \text{rural}) + \alpha_{\text{urban}}\, I(c \in \text{urban}) \\ + \delta_{\text{age}(k)} + \delta_{\text{educ}(k)} \\ + \boldsymbol{\beta} \boldsymbol{X}_{j[k]} + e_k + e_c.\end{aligned}$$

截距固定为 $\alpha_{\text{urban}} = -2.1$ 和 $\alpha_{\text{rural}} = -1.8$。术语 $\delta_{\text{age}(k)}$ 和 $\delta_{\text{educ}(k)}$表示预计可通过多层回归与后分层捕获的个体层协变量。向量$\boldsymbol{X}_{j[k]}$ 表示区域层面协变量，其中包括夜间灯光、到最近医疗机构的出行时间和平均家庭规模，所有这些变量均已标准化。协变量效应定义为

$$\beta_{\text{ntl}} = -0.08, \quad \beta_{\text{health}} = 0.08, \quad \beta_{\text{edu}} = -0.6, \quad \beta_{\text{hh}} = 0.08.$$

教育建模为高中就读的二元指标，其效应为 $\delta_{\text{educ}} = -0.6$。年龄组效应构造如下：

$$\delta_{\text{age}} = \begin{cases} -0.3 & \text{if age 15–19,} \\ 0.6 & \text{if age 20–24,} \\ 0.5 & \text{if age 25–29,} \\ 0.3 & \text{if age 30–34,} \\ -0.2 & \text{if age 35–39,} \\ -0.9 & \text{if age 40–44.} \end{cases}$$

选择这些协变量效应值是为了引入亚组之间的变异，并在一定程度上近似现实模式。然而，它们并非旨在作为真实潜在效应的直接或准确估计值。个体层面随机效应分别从标准差为 0.05 的正态分布中抽取，适用于 e_k 和 e_c。

情景2： 在第二种情景中，除了已指定的协变量外，我们还将未观测协变量纳入模型。扩展后的结构为

$$\log(\mu_{c,k}^{s_2}) = \log(\mu_{c,k}^{s_1}) + \boldsymbol{\beta}_{\text{unobs}} \boldsymbol{X}_{j[k]}^{\text{unobs}},$$

其中， $\boldsymbol{X}_{j[k]}^{\text{unobs}}$包含两项指标：植被指数（NDVI）和女性手机使用情况，两者均经过标准化处理。这些协变量的效应设定如下：

$$\beta_{\text{ndvi}} = 0.08, \quad \beta_{\text{mobile}} = -0.08.$$

情景3： 第三种情景在第二种情景的基础上，引入了一个 Admin-2 级独立随机效应：

$$\log(\mu_{c,k}^{s_3}) = \log(\mu_{c,k}^{s_2}) + e_j, \qquad e_j \sim \mathcal{N}(0, 0.12^2).$$

新增的 Admin-2 级独立同分布随机效应代表非结构化残差变异，这些变异预计无法由分解模型解释。

情景 4： 在最后一种情景中，我们通过允许同一区域内城市和农村部分的系数有所不同，引入模型设定错误。与前几种情景中假设系数统一不同，这种设定反映了不同居住地类型之间的异质性。两个层的系数如下：

城市： $\alpha = -2.10$, $\beta_{\text{ntl}} = -0.10$, $\beta_{\text{health}} = 0.10$, $\beta_{\text{edu}} = -0.50$, $\beta_{\text{hh}} = 0.06$, $\beta_{\text{ndvi}} = 0.06$, $\beta_{\text{mobile}} = -0.06$, $\delta_k^{\text{educ}} = -0.5$

农村： $\alpha = -1.80$, $\beta_{\text{ntl}} = -0.06$, $\beta_{\text{health}} = 0.06$, $\beta_{\text{edu}} = -0.70$, $\beta_{\text{hh}} = 0.10$, $\beta_{\text{ndvi}} = 0.10$, $\beta_{\text{mobile}} = -0.10$, $\delta_k^{\text{educ}} = -0.7$

$$\delta_{\text{age}}^{(\text{Urban})} = \begin{cases} -0.4 & \text{if age group is 15–19,} \\ 0.5 & \text{if age group is 20–24,} \\ 0.5 & \text{if age group is 25–29,} \\ 0.3 & \text{if age group is 30–34,} \\ -0.1 & \text{if age group is 35–39,} \\ -0.8 & \text{if age group is 40–44.} \end{cases} \qquad \delta_{\text{age}}^{(\text{Rural})} = \begin{cases} -0.2 & \text{if age group is 15–19,} \\ 0.6 & \text{if age group is 20–24,} \\ 0.4 & \text{if age group is 25–29,} \\ 0.3 & \text{if age group is 30–34,} \\ -0.2 & \text{if age group is 35–39,} \\ -0.9 & \text{if age group is 40–44.} \end{cases}$$

S4 附加模拟结果

表 S2：各方法与情景的性能指标。数值显示每个选定指标在 500 次模拟重复中的均值。

方法	回归 R^2	皮尔逊r区间得分		Bias	绝对相对偏差
情景1					
FH 分解	0.927	0.978	0.014	0.000	3.6%
FH-MRP 分解	0.929	0.982	0.014	0.000	3.7%
单元分解	0.949	0.983	0.011	0.000	3.1%
单位-MRP 分解	0.983	0.997	0.009	0.000	1.7%
FH（Adm2）	0.642	0.832	0.020	0.000	7.6%
直接（Adm2）	0.062	0.690	0.029	0.000	13.3%
情景2					
FH 分解	0.857	0.962	0.024	0.000	5.5%
FH-MRP 分解	0.821	0.957	0.025	-0.003	6.3%
单元分解	0.911	0.979	0.021	0.000	4.2%
单位-MRP 分解	0.902	0.976	0.020	-0.001	4.3%
FH（Adm2）	0.691	0.859	0.021	0.001	8.1%
直接（Adm2）	0.291	0.738	0.029	0.001	13.4%
情景3					
FH 分解	0.622	0.773	0.034	0.000	10.8%
FH-MRP 分解	0.598	0.770	0.036	-0.002	11.2%
单元分解	0.670	0.787	0.031	0.000	10.0%
单位-MRP 分解	0.668	0.786	0.032	0.000	10.0%
FH（Adm2）	0.685	0.839	0.023	0.001	9.9%
直接（Adm2）	0.469	0.792	0.029	0.001	13.4%
情景 4					
FH 分解	0.593	0.756	0.035	0.000	10.9%
FH-MRP 分解	0.570	0.753	0.037	-0.002	11.4%
单元分解	0.641	0.771	0.032	0.000	10.3%
单位-MRP 分解	0.628	0.764	0.034	0.000	10.4%
FH（Adm2）	0.670	0.834	0.023	0.001	9.9%
直接（Adm2）	0.449	0.789	0.028	0.001	13.4%

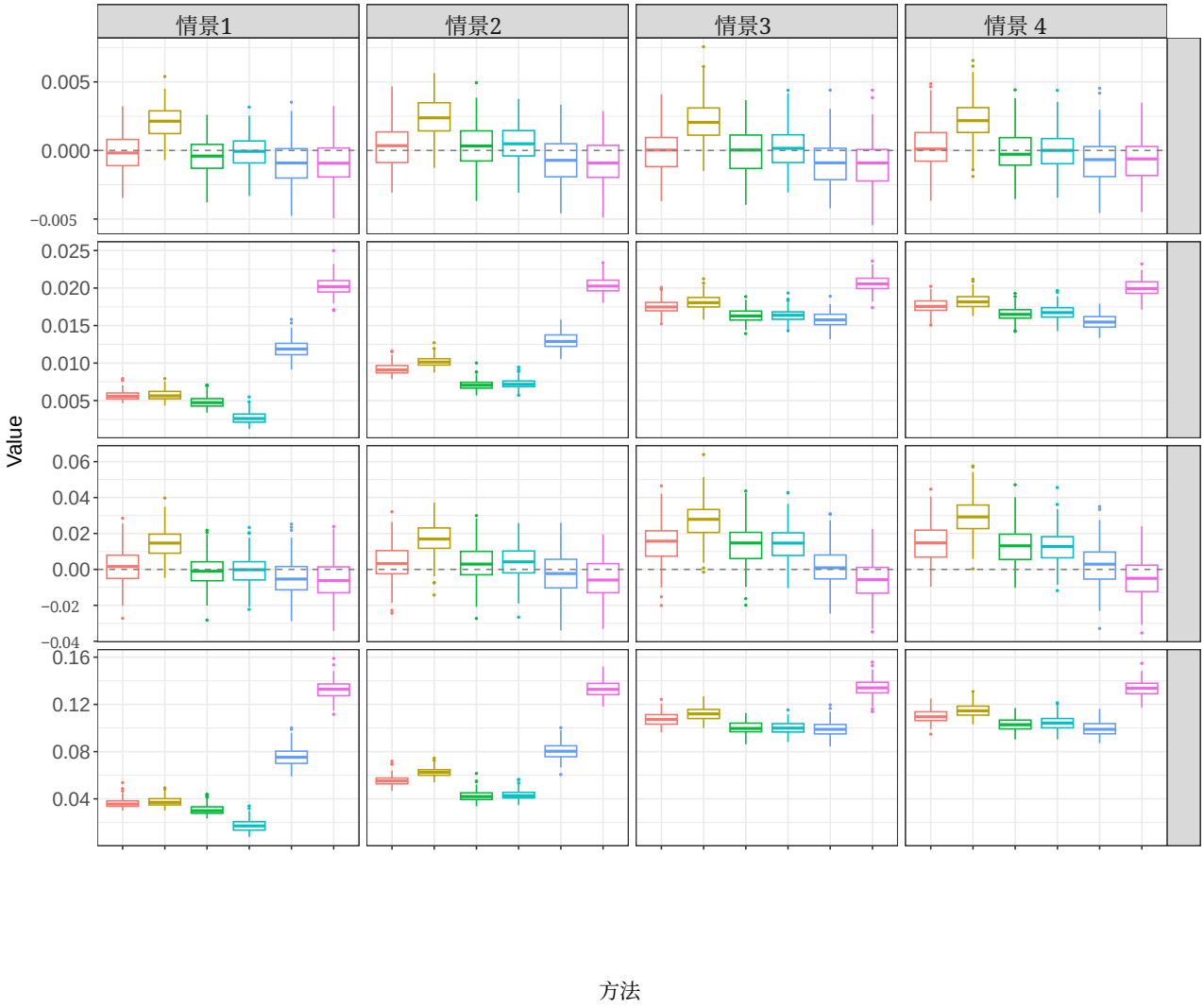


图 S1：在四种模拟情景下，使用偏差指标评估分解模型与基准模型的比较。

S5 附加可视化

S5.1 总生育率估计值的进一步比较（案例研究1）

图 S2 展示了各分解模型的二级行政区估计值与 Admin-2 Fay–Herriot（FH）模型估计值之差，以及由 Admin-2 FH 模型估计的区域层面随机效应。二者之间的强关联表明，分解模型的偏差出现在区域层面效应较强的区域。这与我们的模拟结果一致：按照设计，分解模型识别区域层面随机效应的能力有限，尤其是非结构化分量。

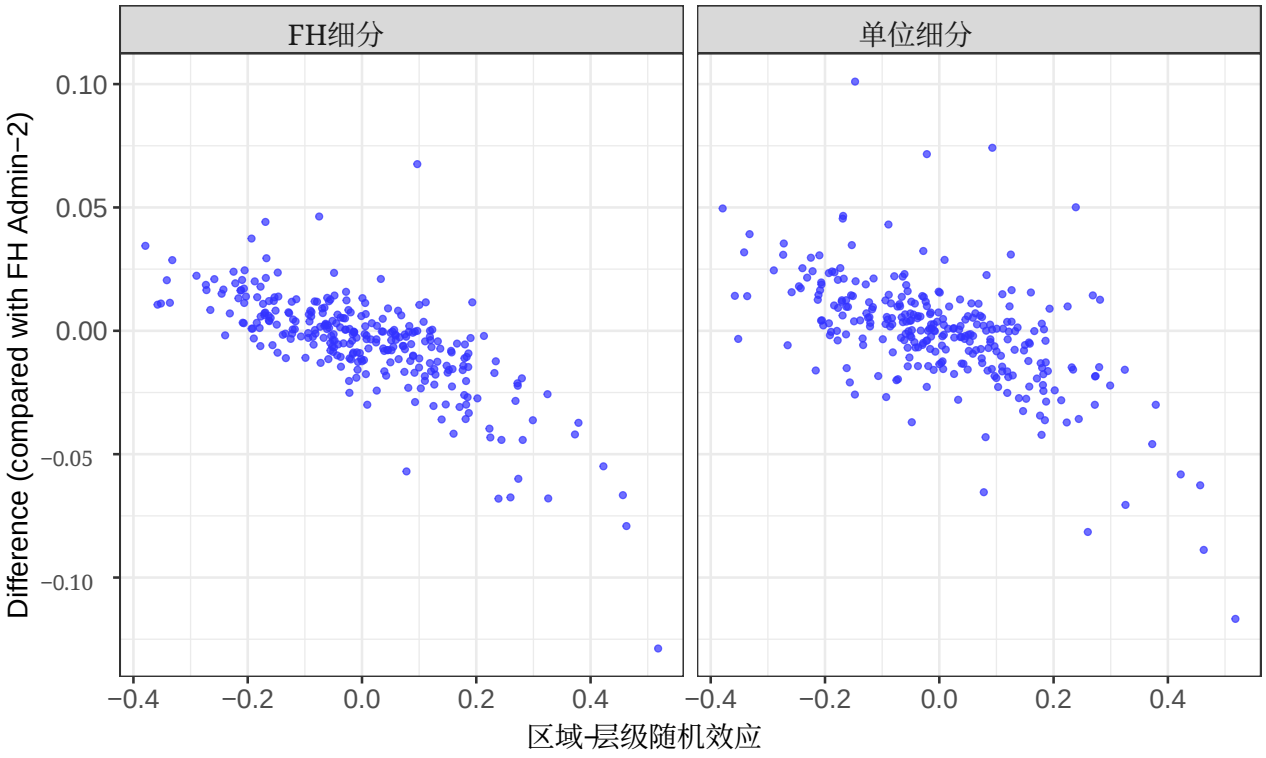
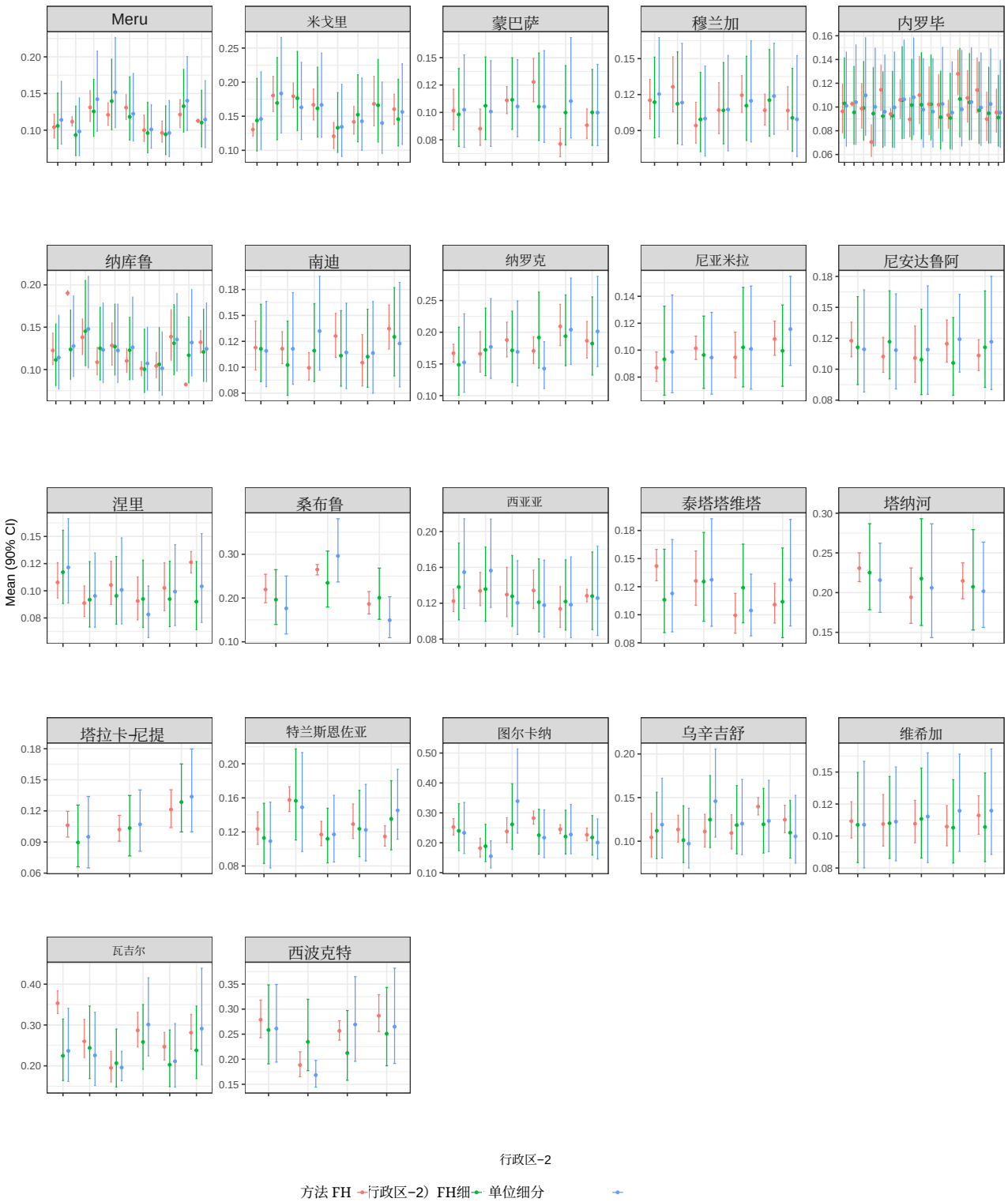


图 S2：区域层面随机效应与分解模型总生育率估计值相较于 FH二级行政区基准的差异之间关系的散点图。

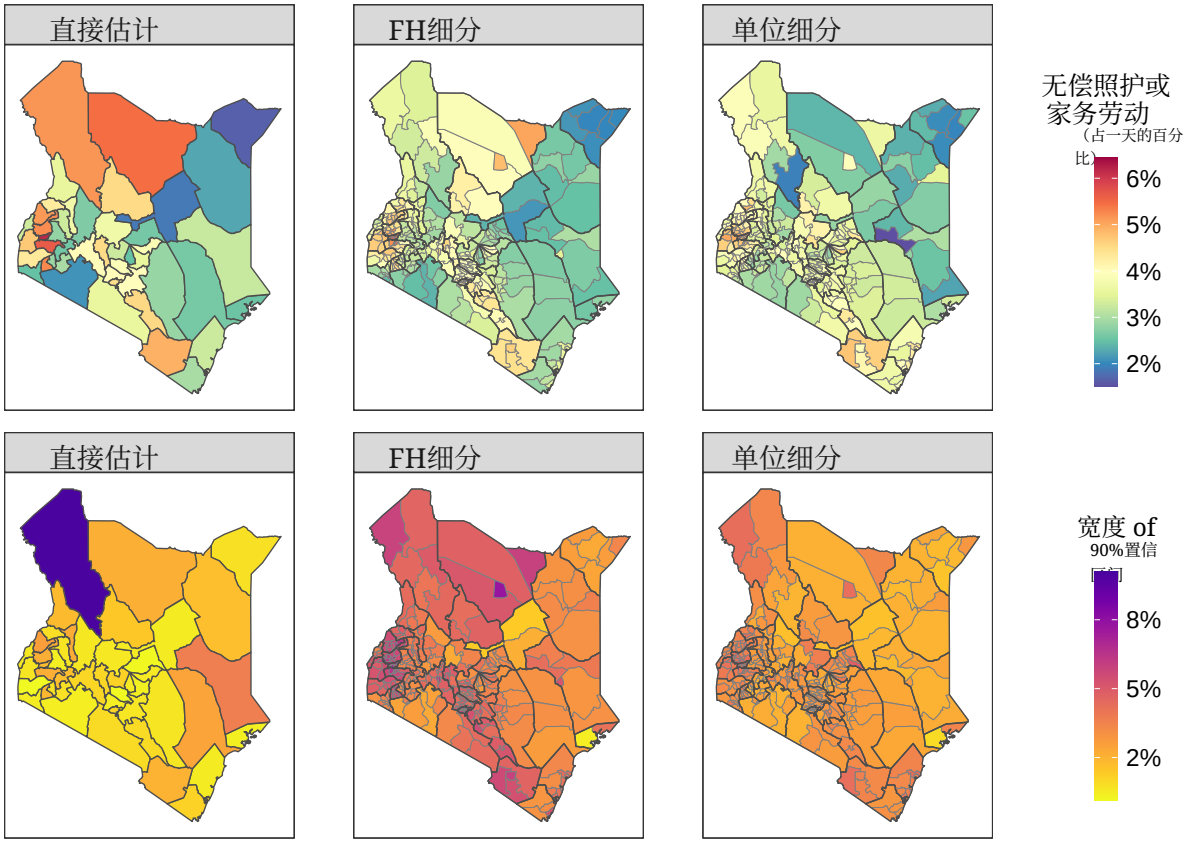


图 S3：基准 FH 二级行政区模型、FH 细分模型和单位层级分解模型中，一级行政区分组内显示的总生育率二级行政区点估计和 90% 可信区间比较。


Meru
米戈里
蒙巴萨
穆兰加
内罗毕
纳库鲁
南迪
纳罗克
尼亚米拉
尼安达鲁阿
涅里
桑布鲁
西亚亚
泰塔塔维塔
塔纳河
塔拉卡-尼提
特兰斯恩佐亚
图尔卡纳
乌辛吉舒
维希加
瓦吉尔
西波克特
Mean (90% CI)
行政区−2
方法 FH
行政区−2）FH细
单位细分

图S4：区间图（续）。

S5.2 男性无偿照护和家务劳动时间



图S5：基于一级行政区直接估计、FH细分和单位层级分解模型，估计男性用于无偿照护和家务劳动的一天中的比例。地图同时显示点估计及其对应的90%不确定性区间宽度。

S5.3 男性和女性大众媒体使用估计的不确定性

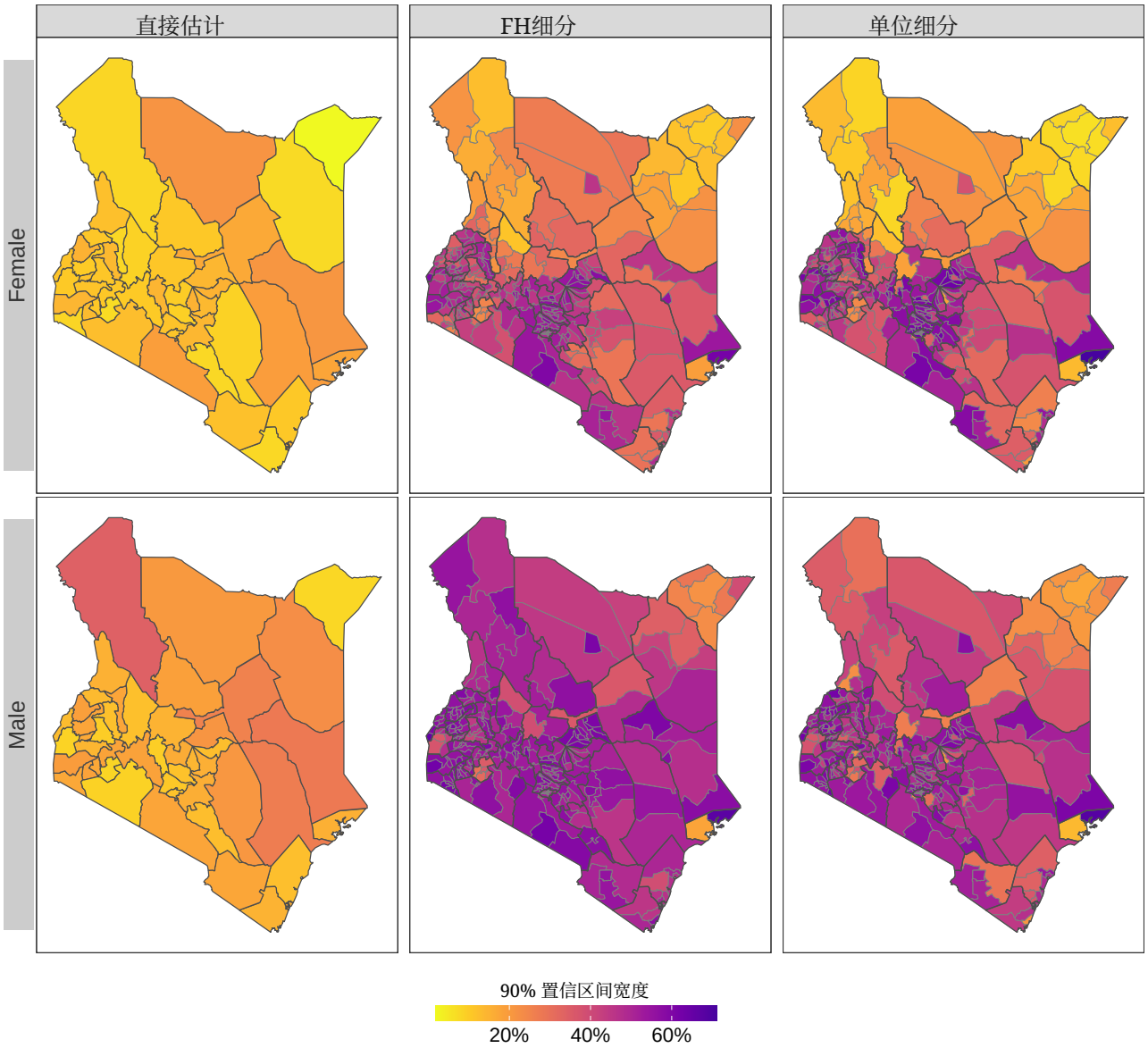


图 S6：男性和女性大众媒体使用估计值的 90% 置信区间宽度地图 s.