arXiv:2611.29632v1 [econ.EM] 22 Nov 2026

概率可行的历史迁移

分布式前趋势设计的尖锐界限与证伪检验

Ulrich Hounyo*

2026年11月24日

摘要

分布差分中的差分研究通常假设，处理组与对照组在政策实施后的差距，是政策实施前各差距的混合分布。这一历史迁移假设可能以单一统计量检验无法发现的方式失效。合并多个政策事件时，平均反事实可能是一个有效分布，尽管没有任何单个事件的反事实是有效分布，而合并界限也可能排除真实平均效应。过去差距的混合分布可能会给结果赋予负概率，按截断值计算的界限也可能彼此不相容。我们刻画了尖锐界限，给出崩溃半径、无网格置信区域，以及有无实验时的检验。这些检验在全国支持性工作项目数据中拒绝迁移假设，并揭示了合并最低工资界限背后不可能成立的州级反事实。对于1993年劳动所得税抵免扩张，检验结果截至1995年均未拒绝该假设；届时，界限显示，单身母亲相对于无子女女性的就业增幅为5至7个百分点。剔除较早实施福利豁免的州后，就业增幅仍为正。

关键词：双重差分；部分识别；分布式处理效应；前趋势；崩溃分析；研究内比较。
JEL分类：C14, C21, C23, J23, J38.

*纽约州立大学奥尔巴尼大学经济学系。电子邮件：khounyo@albany.edu。

1 引言

双重差分研究日益关注政策如何改变整个结果变量分布：工资低于某一水平的劳动者比例、有任何收入的概率，或某个分位数。当平行趋势假设存疑时，政策实施前的记录自然成为寻找约束条件之处。Ban 等人(2026年)最近针对平均效应提出了一项方案，以一项更弱的要求取代平行趋势假设：政策实施后的处理组与对照组差距，必须与政策实施前观察到的差距相似，即属于这些差距的混合分布。本文探讨了当研究对象是一个分布时，这项要求意味着什么，以及它能带来什么。

这个问题并非技术细节。令 $\mu^0_{g,\,t}$ 表示组 $g \in \{0,1\}$ 在日期 t的未处理结果分布，其中 $g=1$ 是处理组，$t=1$ 是政策实施后的第一个日期，并令 S 为一组政策前日期。该要求的分布版本为

$$\mu^0_{1,1} - \mu^0_{0,1} = \sum_{s\in S} \lambda_s \big(\mu^0_{1,s} - \mu^0_{0,s}\big) \quad \text{for some weights } \lambda_s \geq 0 \text{ with } \textstyle\sum_s \lambda_s = 1. \tag{1}$$

分析者使用（1）时，可能出现两种问题。对照组当前分布加上过去差异的混合，是一种有符号测度，因此可能对某些结果赋予负概率；这样的“反事实”并不描述任何总体。此外，分析者在多个工资阈值处分别优化权重来界定效应，所报告的端点可能来自彼此不相容的混合；每个端点单独看都是尖锐的，但不存在一个反事实分布能同时产生这些端点。逐个考察单一标量时，这两种问题都无从察觉，但二者都会影响经济结论。例如，Cengiz 等人（2019年）的聚集估计量，将新最低工资以下缺失的就业岗位与其上方多出的就业岗位相抵；只有当这两项计数都来自同一个反事实工资分布时，对两者之差的界定才有意义。第三种问题出现在如今已成惯例的多个政策事件合并分析中：事件层面反事实的平均值看起来可能像一个分布，即便其中没有任何一个本身是分布；于是，合并界限可能无法涵盖其被解读为所衡量的平均效应。我们研究的数据中，这三种问题都存在。

论文的研究内容。 精确识别集对所有结果事件施加带有共同权重的（1），并与概率分布集合取交集（定理 1）；在包含 K 个单元的网格上，任何线性对比的界限都可表示为关于 $|S|$ 个历史权重的线性规划，其中包含 K 个约束。本文探讨这一限制对分布式前趋势设计的使用方式有何影响：

- 合并事件可能掩盖不可能的反事实。将事件堆叠起来，并且只检验平均值是否可行，是最容易导致误导的做法。当某个历史

权重在各事件间共享时，平均后的反事实可能是一个有效分布，但对于每个允许的权重，总有某个事件的反事实并非如此（定理4）。此时，叠加界限描述的是一个没有任何事件会产生的复合结果，而不是事件效应的平均值。这不仅影响解释：当各事件以不同权重利用各自的历史时，在我们的模拟设计中，叠加界限有 29% 的情况下会排除真实平均效应。我们提出一种精确检验，每个事件和单元对应一个线性规划；给出三种合并模型（共同权重、事件特定权重和合并约束）的尖锐集合；并构造一族模型，通过限制事件权重之间的差异程度，在这些模型之间进行插值。同样的结构也适用于协变量：在协变量单元内进行迁移，需要逐单元满足相容性，而无条件检验无法发现这一点。在最低工资数据中，无论采用共同权重还是州特定权重，堆叠反事实都是一个分布，但每个州的反事实都不是。

- 相容性如何改变结论。当且仅当所有达到无约束界限的历史混合分布均不可行时，概率约束才会收窄政策对比的界限（命题1）；它还可能使符号反转。当不存在可行混合分布时，某个单一的非负结果变量得分在每个历史混合分布下的期望均为负，从而推翻整个模型（定理3）。分别界定截点，以及将候选分布投影到单纯形，都会改变识别集。

- 网格。对于在两种网格上均有定义的对比量，较粗的网格只会扩大界限，却可能完全掩盖不可行性。通过一个线性规划对所有区间计算置信域，可以消除对网格的选择（命题3–5）。

- 失效分析与推断。 我们给出一个精确公式：只需一个线性规划，其中两个项共用一个权重向量，即可求出效应对结果占比的影响方向不再确定时的半径，并给出置信区间（命题 7）。这与使模型达到可行所需的距离不同：在 NSW 数据中，修复不足 1% 的概率质量即可恢复可行性，而效应方向在全变差接近 40% 的偏离下仍然保持不变。若改用收入必须变动多少来衡量（Wasserstein 距离），则平均变动约 $1,050 时，效应方向便不再确定，远低于实验揭示的 $2,630 偏离；距离的选择很重要，而这些程序可适用于两种距离。由于混合分布的误差绝不会超过任一历史日期上的最大误差，针对有限个历史候选分布构造的同时置信带可用于得到置信集、进行相容性检验；当实验识别出缺失分布时，还可检验迁移本身，并给出其误差的置信下界（定理6，命题 8）。若个人历程相互独立，且采用 Dvoretzky–Kiefer–Wolfowitz 半径，这些方法在有限样本下均有效；自助法和学生化版本则更尖锐，并在样本量增长时有效。对于没有

实验的研究者而言，学生化相干性检验是主力方法：模拟结果表明，它在原假设下是保守的，而且在未学生化置信带毫无功效的情况下仍有功效。它是一个带有单纯形干扰权重的矩不等式检验，通用的矩不等式程序也适用于相同的约束。对于按簇（例如州）抽取的数据，在固定网格上，只要置信带按簇进行校准，同样的结果在有限样本下成立，或在簇数增长时均匀成立；单侧校准有效且功效更高。相容性统计量的精确局部极限，即绑定单元上的小型线性规划的值，给出了其功效函数，并解释了为何基于所有顶点单元的临界值会过于保守。

应用所揭示的情况。 三个应用各有不同目的。第一个应用考察历史迁移能否还原已知真实值。在国家支持工作（NSW）示范项目中，随机化揭示了参与者在没有该项目的情况下本会获得的1978年收入分布。我们根据参与者和两个调查对照样本的1974—1975年收入构建历史迁移界限，并将其与实验对照分布进行比较。迁移结果相差甚远。以累积分布函数（CDF）之间最大的垂直间距衡量，真实值与最近的迁移后分布之间距离的95%置信下界，对于当前人口调查（CPS）对照样本为0.20，对于收入动态追踪调查（PSID）样本为0.23；由此推得的就业效应（0.38至0.49）是实验效应（0.11）的三至四倍。迁移失败有其经济原因，且在项目实施前就已显现：资格要求近期失业，因此参与者与调查受访者之间项目实施前的非就业差距（49至62个百分点）远远超过实验揭示的项目实施后差距（22至24个百分点）。当选择机制通过前期结果本身发挥作用时，历史差距无法涵盖政策实施后的差距。即使不使用实验，学生化相干性检验也已拒绝从PSID迁移（$p = 0.013$）。对于CPS，尽管相干性检验未能拒绝迁移假设，只有实验才能揭示迁移失败，这说明前期数据能够检测到的问题有限。当一半实验对照组替代调查样本、从而使迁移在研究设计上成立时，基准检验在200次随机划分中仅拒绝1次，相干性检验则一次也未拒绝。

第二项应用的情境值得检验这一假设：1993年劳动所得税抵免（EITC）扩张，提高了有子女家庭的税收抵免。处理组和对照组（有子女和无子女的单身女性）之间的差异来自一种稳定特征，而非近期结果。我们在估计之前以书面形式固定了研究设计。诊断检验均通过：相干性检验未拒绝，1993年安慰剂基准未拒绝（$p = 0.96$），且对照组的收入分布截至1995年始终处于其历史范围内。到1995年，界限表明，单身母亲的就业增幅

相对于无子女女性为4.7至7.0个百分点；这是税收抵免以及这些年份州级福利豁免的效应。到1996年，这一增幅为11.5至13.9个百分点，但在1996年，对照组的分布超出了其历史范围，因此方案将该年份标记出来。与这些界限相比，抽样不确定性很大：无论重新抽样的是女性还是完整州，95%外部区域在1996年均不包含零，但在1995年包含零。剔除截至1995年已获批福利豁免的州后，所有诊断检验均通过，包括1996年的检验；1996年的界限变为8.4至11.8个百分点，相干性检验会进一步缩窄这一范围。剔除截至1996年所有获批豁免的州后，界限变宽，1995年的界限则包含零。共同权重还能得出分别估计界限无法提供的信息：在1995年，任何就业增幅高于6.2个百分点的反事实，都意味着有收入母亲中处于税收抵免逐步退出区间的比例下降。

第三个应用展示了标准研究设计中的汇总问题，以及在现实网格上相容性能够带来多大增益。在Cengiz 等人（2019年）所使用的118个区间的州—季度工资分布中，针对五次2014年最低工资上调，四个历史候选分布都在某些工资区间上产生了负质量；而施加概率可行性约束后，就业人口占比变化的界限宽度缩小了一半，从 $[-0.52, -0.32]$个百分点。但尽管叠加反事实是一个分布，无论历史权重对所有州是否相同，还是为每个州分别选择，每个州的反事实都不是分布，因此该界限并不是各州效应的平均值。审计还提供了进一步的理由，说明不应将其作因果解读：2014年的劳动力市场超出了其历史范围；删除一个历史季度后——在该季度，两次上调已经签署或获批——相容集合变为空集；而按人口为事件加权，则会使该界限包含零。

由此得到两点启示。对于任何基于历史数据构建的分布反事实，概率一致性都是必要条件；施加这一条件可以显著收紧界限，而忽略它则会产生没有任何总体能够达到的端点。但相容性只是检查，并非验证。历史数据能否约束反事实，取决于政策实施后的环境是否处于历史环境的范围之内；在使用迁移方法之前，可以也应当对此进行审计。若选择反映的是组间稳定差异，而非暂时性项目前状态，那么这一假设便值得认真考虑；EITC 应用表明，在这种情况下，诊断指标能够通过检验，界限也能提供有用信息。

相关文献。 本文将 Ban 等人 （2026年）提出的历史凸包思想从平均效应扩展到分布。对每个有界结果评分使用一个权重来施加其标量限制，等价于（1）（命题S2.5）；若权重可以随评分变化，则这种等价性不成立。因此，我们的模型是其模型的分布闭包；对于单个预先指定的评分，其界限已经是尖锐的；新的问题包括可行性、联合可达性、跨事件汇总、失效分析和检验。其他分布型

双重差分模型施加了不同的限制：变化中的变化中的跨期秩不变性（Athey 和 Imbens，2006）、面板数据中的联结函数稳定性（Callaway and Li，2019）、变换不变性（Roth and Sant'Anna，2023）、分布回归（Fern´andez-Val 等人，2024）以及分布型合成控制（Gunsilius，2023）。第2.2节将它们的经济含义与我们的模型进行比较，并说明何时选择每种方法更为合适。平行趋势的有界偏离（Manski and Pepper，2018；Rambachan and Roth，2023）限制了标量趋势。相容性是一个由单纯形中的权重索引的矩不等式系统，因此相关检验与矩不等式推断相联系（Cher-nozhukov、Hong 和 Tamer，2007；Andrews 和 Soares，2010；Romano、Shaikh 和 Wolf，2014）。合并结果与错位实施文献中的结论相呼应：当效应在事件之间存在差异时，双向固定效应估计量和合并估计量未必会对事件层面效应取平均（de Chaisemartin and D' Haultfœuille，2020；Goodman-Bacon，2021；Sun and Abraham，2021；Callaway and Sant'Anna，2021）；在这里，即使是识别阶段也会出现这一问题，早于任何估计：当不存在任何有效的事件层面分布反事实时，合并分布反事实仍可能是一个有效分布。重新排列估计分布函数（Chernozhukov、Fern´andez-Val 和 Galichon，2010）可以在不施加共同权重约束的情况下确保单调性，因此回答的是另一个问题。这些界限采用 Mogstad、Santos 和 Torgovitsky（2018）以及 Kaido、Molinari 和 Stoye（2019）传统中的线性规划，而松弛则是 Masten 和 Poirier（2020）意义上的失效分析。NSW 实证沿用了 LaLonde（1986年）、Dehejia and Wahba（1999年）、Smith and Todd（2005）、Cal´onico 和 Smith（2017年）以及 Imbens and Xu（2025年）的研究；据我们所知，这是对分布前趋势方法的首次实验评估。劳动所得税抵免应用沿用了 Eissa and Liebman（1996年）以及 Meyer and Rosenbaum（2001年）对有子女和无子女单身女性的比较；Kleven（2019年）提出疑问：1993年之后的就业增幅有多大程度可以由该税收抵免解释。最低工资应用建立在 Cengiz 等人（2019）、Dube、Lester 和 Reich（2010年）以及 Callaway and Sant'Anna（2021年）的研究基础上。

第 2节介绍模型及其经济含义，第3节讨论相容性何时重要的结果，第4 节讨论分位数，第 5 节讨论松弛，第 6 节讨论推断与模拟。第7、8和9节介绍应用。正文中未列出的证明见附录。

2 模型

2.1 设置

有两组，$G \in \{0,1\}$，日期为 $t = -T_0, \ldots, 0, 1$；组 $G = 1$ 从 $t = 1$ 起接受处理。结果变量取值于标准博雷尔空间（$\mathcal{Y}$，$\mathcal{B}$），该空间涵盖实值

结果变量、有限类别，以及收入在零点有质量等混合结果。单元具有潜在结果 $Y_t(0)$ 和 $Y_t(1)$，观测结果为 Y_t。记

$$\mu^d_{g,t}(A) = \mathbb{P}\{Y_t(d) \in A \mid G = g\}, \qquad \mu_{g,t}(A) = \mathbb{P}\{Y_t \in A \mid G = g\}, \qquad A \in \mathcal{B},$$

以及 $\mathsf{P}(\mathcal{Y})$ 表示 $\mathcal{Y}$上概率测度的集合。数据为面板数据或重复横截面，且 P 表示观测变量的分布；在重复横截面的情况下，下文使用的每个组—日期单元都具有正概率。组成员身份固定；对于重复横截面，每个组在每个日期都视为同一总体；其构成变化则纳入下文定义的选择偏差。第 2.2 节说明了这些变化如何违反识别假设。

假设1。 *(i)* $0 < \mathbb{P}(G = 1) < 1$。*(ii)* 单元间不存在干扰； $Y_t = Y_t(0) = Y_t(1)$ 在两个组中均成立；并且在 $t = 1$ 时，若 $G = 0$ 则 $Y_1 = Y_1(0)$ ，若$G = 1$则 $Y_1 = Y_1(1)$ 。

第(ii)部分将一致性与无预期效应结合起来。在此条件下，除 $\mu^0_{1,1}$ 外，所有未处理分布都可观测到：$\mu^0_{g,t} = \mu_{g,t}$ 对于 $t \le 0$，以及 $\mu^0_{0,1} = \mu_{0,1}$。令 $\nu_t = \mu^0_{1,t} - \mu^0_{0,t}$ 表示日期 t 的未处理选择偏差，它是总质量为零的有符号测度。固定一个有限非空的历史日期集合 $S \subseteq \{-T_0, \dots, 0\}$，并令 $\Delta_S = \{\lambda \in \mathbb{R}^S \colon \lambda_s \ge 0,\ \sum_s \lambda_s = 1\}$。

假设2（历史迁移）。 存在 $\lambda \in \Delta_S$，使得 $\nu_1 = \sum_{s \in S} \lambda_s \nu_s$ 作为有符号测度成立。

只有一个历史日期时，假设2表示分布差分中的差分，且偏差固定为该日期对应的值。存在多个历史日期时，偏差可以漂移，但只能在历史所覆盖的值域内漂移。该等式是测度之间的等式，因此同一个 λ 适用于每一个结果事件。

2.2 经济含义：环境模型

一个简单的环境模型使假设2 具体化。假设经济在每个时期都处于潜在环境的混合分布中 $e \in \{1, \dots, m\}$，例如衰退与扩张，或低技能劳动力需求高与低的环境，各环境的份额为 π_t，位于 Δ_m的 $\mathbb{R}^m$单位单纯形内，且两组共有；并且在每种环境中 e，两组的未处理结果分布 F_{1e} 和 F_{0e}不随时间变化。那么 $\nu_t = \sum_e \pi_{te}(F_{1e} - F_{0e})$，只要政策实施后的份额 π_1 位于历史份额的凸包 $\mathrm{co}\{\pi_s \colon \ s \in S\}$之内，假设S1.1 就成立。附录 S1.1 说明，组别特定的环境占比会增加一个构成项，而该假设排除了这一项。环境模型足以保证假设2，但并非必要条件：不同的份额也可能产生相同的选择偏差。它

还具有自身可检验的含义。每组在日期 t 的未处理概率律 $\sum_e \pi_{te} F_{ge}$是 π_t的仿射函数，因此 $\pi_1 \in \mathrm{co}\{\pi_s : s \in S\}$ 意味着$\mu_{0,1} \in \mathrm{co}\{\mu_{0,s} : s \in S\}$：对照组的当前分布必须是其自身过去分布的混合分布，因此，任何关于该分布的线性统计量也必须如此，例如非就业率。对照组统计量若落在其历史范围之外，就能推翻环境模型，但不能推翻假设2 本身。该模型还指出了何时值得考虑迁移：当两组的差异源于教育或当地劳动力需求等稳定特征，而非触发处理的暂时状态时。这是假设2在经济学上的解读，也为下文的实证分析提供了框架。当两组的环境份额不同时，同一模型表明，单凭构成差异就可能导致假设2不成立；附录S1.1 给出了由此产生的构成项，并说明，对环境份额差异设定一个界限，就能得到第5节的松弛模型，其半径等于该界限（推论S1.1）。

协变量以与事件相同的方式进入。如果在离散的预先确定的协变量 X的各单元内实施迁移，并使用权重 $\lambda(x)$，则处理组的反事实为 $\sum_x p_1(x)\mu^0_{1,1}(\cdot \mid x)$，其中 p_1 是处理组在政策日期的协变量分布，且每个条件反事实都必须是一个概率律。于是，尖锐集合是条件尖锐集合的 p_1加权闵可夫斯基和（命题S2.11）。仅对无条件分布施加迁移约束，只能检验平均值本身的相容性，正如对事件进行堆叠时的情形一样。当条件迁移成立时，这种做法也可能失败：在协变量分布随时间保持稳定的情况下，两者之间的差距等于两组协变量分布之差乘以对照组条件分布偏离其自身过去混合分布的程度（命题S2.11）；而协变量分布随时间变化还会增加一项。因此，要为迁移假设辩护，就需要在协变量单元内实施条件迁移；这是一项更强的假设，并且要求每个单元都具有各自的历史支持集。

应用中的每项审计都会提出三个问题之一：根据刚刚推导出的蕴含关系，π_1 是否可以位于 $\mathrm{co}\{\pi_s : s \in S\}$ 中；历史日期是否不受政策影响；以及组构成是否稳定?

表 1 比较了历史迁移与其他分布假设。变化中的变化（Athey 和 Imbens，2006）将 $Y_t(0) = h(U, t)$ 表示为 h 关于标量严格递增的函数 U，且该标量在各组内的分布随时间保持不变。它允许政策实施后的分布移出此前观察到的所有范围，只要这种变化是共同单调变换；而历史迁移则不允许这种外推。反过来，环境模型通常不满足变化中的变化：当两个环境对个体的排序不同时，结果变量就不是时不变秩的单一单调函数，但只要政策实施后的环境份额被历史份额所界定，历史迁移便成立。

表1：处理组未处理分布的识别假设

假设	限制	外推超越历史	可检验的含义 data
分布的平行趋势	$\nu_1 = \nu_s$ 针对一个基准日期 s	No	候选分布是一个分布；ν_s 在 s之间相等 如果对所有基准日期都施加
变化中的变化	$Y_t(0) = h(U, t)$，h 递增 在 U中，U 的概率律在组内稳定	是，通过 h	日期变换 稳定 跨前期
有界趋势	组 $\left\lvert \int u\, d\nu_1 - \int u\, d\nu_0 \right\rvert$ 受……界限约束 过去变化的一个倍数	是的，最多达到多个	None
标量历史的值域	得分 $\int u\, d\nu_1$ 在 $\int u\, d\nu_s$的值域内，一个得分	No	无，一个得分
历史迁移	$\nu_1 \in \mathrm{co}\{\nu_s\}$，所有事件共用一个权重事件	No	相容性；基准 test

注：“可通过数据检验”的限制条件意味着它可能不符合观测分布。相容性是指至少有一个迁移后的候选分布为概率分布（第3节）。 ν_t 表示未处理选择偏差，即在日期 t处理组与对照组的未处理结果分布之差。变化中的变化由 Athey 和 Imbens (2006提出；有界趋势见 Manski and Pepper(2018年) 和 Rambachanand Roth(2023年)；标量历史范围见 Ban 等人(2026年)。

有界趋势（Manski andPepper，2018年；Rambachan and Roth，2023年）限制标量偏差相对于其过去变化幅度的变动，并允许按选定倍数进行外推；历史迁移不允许外推，但会约束整个分布。因此，如何选择取决于经济学考量。历史迁移适用于政策时期与历史记录相似、但条件可能导致个体重新排序的情形；变化中的变化适用于秩保持稳定、但政策时期可能超出历史范围的情形。在 NSW 应用中，项目后时期超出了历史范围，因此历史迁移在此不成立；变化中的变化也不成立（第 7节）。

2.3 未处理结果分布的尖锐集合

对于 $\lambda \in \Delta_S$，定义传输候选分布和历史顶点

$$\mu_\lambda = \mu_{0,1}^0 + \sum_{s \in S} \lambda_s \nu_s = \sum_{s \in S} \lambda_s v_s, \qquad v_s = \mu_{0,1}^0 + \nu_s. \tag{2}$$

每个 μ_λ 都是确定的，且总质量为 1，但不一定非负。相容权重集为

$$\Lambda_S(P) = \{\lambda \in \Delta_S : \ \mu_\lambda \in \mathsf{P}(\mathcal{Y})\},$$

的一个紧凸子集 Δ_S。对于实值结果，记 $H_\lambda(y) = \mu_\lambda((-\infty, y])$；那么 $\lambda \in \Lambda_S(P)$ 当且仅当 H_λ 是非递减的，因为在 $\pm\infty$ 处的极限和右连续性均继承自观测分布函数。令 $\mathfrak{Q}_S(P$ 表示能够重现 P 且满足假设1–2的潜在结果、组别标签和日期的联合分布集合，并记 $\mu^0_{1,1,Q}$ 为在 $Q \in \mathfrak{Q}_S(P)$ 下，处理组在 $t = 1$ 时的未处理分布。

定理 1（精确识别）。在假设*1*（*i*）下，

$$\left\{\mu^0_{1,1,Q} : Q \in \mathfrak{Q}_S(P)\right\} = \left[\mu^0_{0,1} + \mathrm{co}\{\nu_s : s \in S\}\right] \cap \mathsf{P}(\mathcal{Y}) = \{\mu_\lambda : \lambda \in \Lambda_S(P)\}.$$

特别地，当且仅当 $\Lambda_S(P) \neq \varnothing$ 时，模型与 P 相容。

必要性显而易见。为证明充分性，取一个相容的 μ_λ，并使用独立的均匀随机变量从 μ_λ 中抽取处理单元缺失的未处理结果，同时保持所有观测变量不变（附录S1.5）。该定理针对的是边际未处理分布。它不限制 $Y_1(0)$ 与 $Y_1(1)$ 之间，以及 $Y_1(0)$ 与单元历史之间的依赖关系，因此能够识别分位数处理效应，即两个分布之间的差异，却无法说明个体效应的分布。对个体收入动态施加限制的模型，其识别集可能更小。

历史凸包距离分布集有多远？对于带有约当分解 $\mu = \mu^+ - \mu^-$ 的有符号测度 μ，令 $b(\mu) = \mu^-(\mathcal{Y})$ 表示其负质量；对于单位质量的有符号测度 μ，Q，令 $d_{\mathrm{TV}}(\mu, Q) = \sup_{A\in\mathcal{B}} |\mu(A) - Q(A)|$ 表示它们的全变差（TV）距离。

定理2（到概率空间的距离）。对于每个满足 $\mu(\mathcal{Y}) = 1$ 的有限符号测度 μ，

$$\min_{Q\in\mathsf{P}(\mathcal{Y})} d_{\mathrm{TV}}(\mu, Q) = b(\mu) = \sup_{A\in\mathcal{B}} [-\mu(A)]_+,$$

其中 $[x]_+ = \max(x, 0)$，且最低工资在 $Q^\dagger = \mu^+/\{1 + b(\mu)\}$ 处取得。因此，最低工资修复 $d_S = \min_{\lambda\in\Delta_S} b(\mu_\lambda)$ 就是历史凸包与 $\mathsf{P}(\mathcal{Y})$ 之间的全变差距离，并且当且仅当 $\Lambda_S(P) \neq \varnothing$ 时成立。

证明。令 $b = b(\mu)$，并令 N 为哈恩负集，因此 $\mu(N) = -b$。对于任意概率律 Q，$d_{\mathrm{TV}}(\mu, Q) \geq |\mu(N) - Q(N)| = b + Q(N) \geq b$。此外，$\mu - Q^\dagger = \{b/(1+b)\}\mu^+ - \mu^-$ 是两个质量为 b 的互相奇异测度之差，因此 $d_{\mathrm{TV}}(\mu, Q^\dagger) = b$。在 λ 上取最低值可以达到，因为 $\lambda \mapsto b(\mu_\lambda)$ 是凸且利普希茨的（引理S1.2）。 □

修复后的规律 $Q^{\dagger}$ 是一种诊断工具。它通常位于历史凸包之外，因此并非该模型允许的反事实。

2.4 有限网格

应用报告有限划分 $\Pi = \{A_1, \dots, A_K\}$ of $\mathcal{Y}$上的结果。令 $p_{g,\ t} \in \mathbb{R}^K$收集 $\mu^0_{g,\ t}$的单元概率，令 $z_s = p_{0,\ 1} + p_{1,\ s} - p_{0,\ s}$ 为顶点 v_s的单元向量，并令 $z_\lambda = \sum_s \lambda_s z_s$。网格相容权重为

$$\Lambda_{S,\Pi} = \{\lambda \in \Delta_S :\ z_\lambda \geq 0\},$$

以及网格修复为 $d_{S,\ \Pi} =$ 最小化$_\lambda \sum_k [-z_{\lambda,k}]_+$。当$z_s$ 的某个坐标为负时，顶点无效。无效顶点并不意味着集合为空：对于两个单元，$z_1 = (-0.1,\ 1.1)$ 和 $z_2 = (1.1,\ -0.1)$ 均为无效顶点，但 $\lambda z_1 + (1-\lambda) z_2$ 是 $\lambda \in [1/12, 11/12]$的概率向量。因此，单独检验每个历史日期并不能检验假设 2。下面对线性对比量的每个界限，都是关于 $|S|$ 权重且带有 K非负性约束的线性规划；修复过程增加了 K 松弛变量。

3 相容性的重要性

3.1 政策界限

令 $\mathcal{H}_S = \mathrm{co}\{v_s :\ \ s \in S\}$ 为带符号凸包，且 $\mathcal{H}_S^+ = \mathcal{H}_S \cap \mathsf{P}(\mathcal{Y})$ 为其相容部分。如果每个顶点都是概率律，则二者重合（命题 S2.1）；否则 $\mathcal{H}_S^+$ 严格更小。集合严格变小并不一定会改变给定的界限，而下一个结果将准确说明其发生的条件。

命题 1 (相容性何时收紧线性界)。设 $\Lambda_S(P) \neq \varnothing$。令 ℓ为在各顶点上取有限值的线性泛函，令 $a_s = \ell(v_s)$，$a_{\min} =$ 最低工资$_s\, a_s$，且$S_{\min} = \{s :\ \ a_s = a_{\min}\}$。在 $\Lambda_S(P)$ 上，$\ell(\mu_\lambda)$ 的最低值等于 $a_{\min}$，当且仅当存在某个$\lambda \in \Lambda_S(P)$ 在 $S_{\min}$之外赋予零权重。对于最大值，使用 $S_{\max}$时同样成立。在有限网格上，这两个条件都是线性可行性问题。

证明。对于每个 $\lambda \in \Delta_S$, $\ell(\mu_\lambda) - a_{\min} = \sum_s \lambda_s (a_s - a_{\min}) \geq 0$，等号当且仅当$\lambda$ 的支持集包含于 $S_{\min}$时成立。在紧集 $\Lambda_S(P)$ 上，最低工资取到，因此它当且仅当面 $\{\lambda :\ \lambda_s = 0,\ s \notin S_{\min}\}$ 与 $\Lambda_S(P)$ 相交时等于$a_{\min}$。对最大值，将此论证应用于 $-\ell$。

□

增益可能使符号反转。设有三个单元，顶点为 $z_1 =$(。45，。30，。25) 和 $z_2 =$(。10，$-.35$，1。25)，且观测到的处理组在第一单元中的份额为 。20。若权重 t 施加于 z_2，

第二单元的质量为 $.30-.65t$，因此相容性要求 $t \le 6/13$。反事实减观测的第一单元份额的相容下界为 $.25-.35\cdot 6/13 = 23/260 \approx .088 > 0$，而带符号凸包给出的结果是 $-.10$；只有在第二单元存在负质量的“分布”中才能达到该值。使用阈值特定权重时，负端点可以达到（附录 S2.2），因此这种增益来自共同权重。

3.2 一个概率律，多个结论

研究者通常会报告多个阈值或对比量。根据假设2 这些都必须来自同一个 λ。下一个定理给出一个对偶证书，证明没有任何 λ 可行。

定理 3（不可行性的共同见证）。 $\Lambda_S(P) = \varnothing$ 当且仅当存在事件A_1、…、A_m、一个权重向量 $q \in \Delta_m$ 和 $\gamma > 0$，使得 $f = \sum_j q_j \mathbf{1}_{A_j}$ 满足 $0 \le f \le 1$且

$$\max_{\lambda \in \Delta_S} \int f \, d\mu_\lambda \le -\gamma.$$

在 K 单元网格上，这种边际的最大值等于网格修复：

$$\min_{\lambda \in \Delta_S} \sum_{k=1}^{K} [-z_{\lambda,k}]_+ = \max_{f \in [0,1]^K} \min_{s \in S} \bigl(-f^{\mathsf{T}} z_s\bigr). \tag{3}$$

证明。若 $\Lambda_S(P) = \varnothing$，则每个 λ 都会给出某个事件的负质量。开集$\{\lambda : \mu_\lambda(A) < 0\}$覆盖紧单纯形，因此有限多个事件 A_1、…、A_m 就足够，且 $\max_\lambda \min_j \mu_\lambda(A_j) = -\gamma < 0$。收益函数 $\sum_j q_j \mu_\lambda(A_j)$ 在 $\Delta_m \times \Delta_S$ 上是双线性的，极小极大定理给出 q，使得 $\max_\lambda \sum_j q_j \mu_\lambda(A_j) = -\gamma$。反过来，非负得分在任意概率律下的期望都非负。在网格上，$\sum_k [-z_{\lambda,k}]_+ = \max_{f \in [0,1]^K} (-f^{\mathsf{T}} z_\lambda)$，而极小极大定理结合关于 λ 的线性性给出（3）。 □

见证 f 是一个结果评分，例如低工资区间的加权计数，所有历史混合分布都预测其均值为负。它否定的是整个模型，而非任何单一前趋势。它是一个总体对象；第6 节将其转化为检验。

分别设定阈值。分别界定每个结果事件，比假设2的限制更弱。类别编码为0、1、2时，有效顶点为 $v_1 = (.2, .6, .2)$和 $v_2 = (.4, .1, .5)$，且平均类别为 $m(p) = p_1 + 2p_2$，共同权重给出 $m \in [1.0, 1.1]$。对每个事件分别施加历史值域限制，会纳入(.2, .3, .5)和(.4, .4, .2)，它们都不是v_1 和 v_2的混合分布，且 $m \in [.8, 1.3]$。对于单个预先指定的得分，将标量限制应用于该得分便已得到 $[1.0, 1.1]$；当多个报告量必须描述同一个总体时，共同权重才至关重要。

3.3 堆叠事件与复合反事实

许多应用会汇总多个策略事件，例如不同州的最低工资上调，并报告平均效应。设事件 $e = 1$、……、E的聚合权重为$\omega \in \Delta_E$，未处理分布为μ^0_e、g、t，顶点为v_e、$s = \mu^0_e$、0、$1 + \mu^0_e$、1、$s - \mu^0_e$、0、s，并令μ_e、$\lambda = \sum_s \lambda_s v_e$、$s$、$\bar{\mu}_\lambda = \sum_e \omega_e \mu_e$、$\lambda$和$z_e$、$\lambda$分别表示网格上$\mu_e$、$\lambda$的单元向量。目前有两种模型在使用中，但往往未加区分。在逐事件传输下，假设2对每个事件都成立，并且使用同一个 λ，相容集合为 $\Lambda^{\text{ev}} = \{\lambda：$ 对所有 $e\}$，均有μ_e、$\lambda \in \mathrm{P}(\mathcal{Y})$。在堆叠传输下，该假设仅施加于平均分布，相容集合为$\Lambda^{\text{st}} = \{\lambda:$ $\bar{\mu}_\lambda \in \mathrm{P}(\mathcal{Y})\}$。

定理 4（堆叠可能掩盖不可行性）。 (*i*) $\Lambda^{\text{ev}} \subseteq \Lambda^{\text{st}}$，因此对于平均反事实，逐事件尖锐集合 $\{\bar{\mu}_\lambda：\ \lambda \in \Lambda^{\text{ev}}\}$ 包含于堆叠集合中。 (*ii*) 该包含关系可能是严格的，并且当 Λ^{ev} 为空时，Λ^{st} 可能非空。 (*iii*) 在有限网格上，对于每个事件 e 和单元 k，一组 EK 线性规划中，$\Lambda^{\text{ev}} = \Lambda^{\text{st}}$当且仅当 $\min_{\lambda \in \Lambda^{\text{st}}} z_{e,\ \lambda,\ k} \geq 0$ 。

证明。 (i) 概率分布的凸组合仍是概率分布。 (ii) 对于一个历史日期、两个事件和两个单元，$z_1 =$ (−0.1，1。1) 和 $z_2 =$ (0。3，0。7) 以相等权重取平均，得到 (0。1，0。9)，因此 $\Lambda^{\text{st}} = \Delta_S$ 而 $\Lambda^{\text{ev}} = \varnothing$。 (iii) 如果某个 $\min_{\Lambda^{\text{st}}} z_{e,\ \lambda,\ k} < 0$，则其最小化点属于 $\Lambda^{\text{st}} \setminus \Lambda^{\text{ev}}$；如果所有最小值均为非负，则每个堆叠相容的 λ 逐事件相容。 □

该定理很简单，但其对实践的影响却并非如此。堆叠是概括多个政策事件的标准方法（Cengiz 等人，2019年；Callaway andSant'Anna，2021年），而堆叠分布界通常被解读为这些事件效应平均值的界限。当 $\Lambda^{\text{st}} \neq \varnothing = \Lambda^{\text{ev}}$时，在逐事件模型下，这种解读并不成立：堆叠模型允许的每一个历史权重，都会使至少一个事件对应的反事实包含负概率，因此，平均后的反事实是一种复合结果，任何逐事件应用该模型的方法都无法产生这一结果。这是关于模型的论断，而非逻辑上的不可能性。当研究对象是处理组总体，并且分析者愿意仅对该总体作出迁移假设时，合并才是正确的目标；例如，州层面的历史可能过于嘈杂，无法约束州层面的反事实。此时，该界限仅涉及合并总体，对任何单一事件都不作说明。我们的建议是，在解释堆叠分布界之前，先计算 EK 第 (iii) 部分中的项目，并在二者不同时报告逐事件界限。在最低工资应用中，堆叠集合非空，但不存在任何共同权重能使全部五个事件层面反事实均有效；而在接下来描述的第三个模型下，即使针对每个州分别选择权重，也同样无法做到这一点。

事件之间的权重各不相同，构成第三种模型；当我们认为不同事件对其历史赋予的权重不同时，应使用这种模型：每个事件都有自己的 $\lambda_e \in \Delta_S$，其 $\mu_{e,\ \lambda_e} \in$

P(Y)，而平均律的尖锐集合是闵可夫斯基平均 $\sum_e \omega_e\{\mu_{e,\lambda_e} : \mu_{e,\ \lambda_e} \in \mathsf{P}(\mathcal{Y})\}$，因此，平均对比量的界限是事件级界的平均值。当且仅当每个事件自身的相容集合非空时，该集合才非空；它包含逐事件集合，但未必包含或被堆叠集合包含（命题 S2.9）。介于前两者之间的是一个由 $\delta \geq 0$索引的族：事件权重 λ_e 满足 $\|\lambda_e - \lambda\|_1 \leq \delta$，且具有共同的 $\lambda \in \Delta_S$。其尖锐界限由线性规划给出，随 δ增大；当 $\delta = 0$ 时，界限等于逐事件界；当 $\delta = 2$时，界限等于事件特定界；而使每个事件都相容的最小 δ 本身也是一个线性规划（命题 S2.10）。它衡量数据所要求的权重异质性程度。

在这些模型之间作出选择会改变结论，而不仅仅是改变解释。在一项总体层面模拟中，三个事件各自都有真实权重和有效的反事实（附录 S10.1），事件特定界始终包含真实平均效应。叠加界限在 29% 的研究设计中将其排除；而共同权重的逐事件模型在 72% 的情况下为空集或将其排除。因此，只有在认为这些事件在其合并历史数据上共享同一个权重时，堆叠才是合理的假设；它并非对事件级迁移无害的概括。

3.4 投影

一种看似诱人的补救方法是将每个无效顶点投影到单纯形上，再混合这些投影。这会以另一种识别限制取而代之。

命题 2 (投影与相容迁移不可交换)。 在三个类别上，存在概率向量 a、b_s、c_s $(s = 1, 2)$，使得顶点 $v_s = a + b_s - c_s$ 构成的相容凸包非空，且与这些顶点在单纯形上的欧几里得投影所构成的凸包不相交。对整个带符号凸包进行投影，可能会严格扩大相容凸包。

证明。取 $a = (1\,5, 2\,5, 2\,5)$、$b_1 = (0, 3\,10, 7\,10)$、$c_1 = (\frac{1}{2}, 0, \frac{1}{2})$、$b_2 = (1\,2, 3\,10, 1\,5)$、$e_2 = (0, 0, 1)$。那么 $v_1 = (-\frac{3}{10}, 7\,10, 3\,5)$，且 $v_2 = (7\,10, 7\,10, -\frac{2}{5})$。以权重 t 对 v_1 进行混合，得到 $(\frac{7}{10} - t, 7\,10, t - \frac{2}{5})$，它对 $t \in [\frac{2}{5}, \frac{7}{10}]$而言是相容的，中间质量始终为 $7\,10$。投影分别为 $(0, 11\,20, 9\,20)$ 和 $(1\,2, 1\,2, 0)$，其混合分布的中间质量属于 $[\frac{1}{2}, \frac{11}{20}]$。投影保持相容混合分布不变，但会将 v_1 映射到相容凸包之外。 □

目前有三种修正方法，应加以区分：将每个候选分布欧几里得投影到单纯形上，对估计分布函数进行单调重排（Chernozhukov, Fern´andez-Val, and Galichon，2010年），以及对整个包络进行最小距离投影。前两种方法在混合前分别修正每个候选分布，而无效顶点的修正取决于所选距离；第三种方法则是第5 节中半径为 d_S的全变差松弛。投影仍可作为一种明确声明的敏感性分析。

第5节的松弛改为修正混合分布，所需的全变差最小，并且其界限对于一个明确的模型是尖锐的（命题7）。若希望得到修正后的模型，我们建议采用这种方法，而不是投影。

3.5 网格

假设 2 对分布施加限制，而任何网格只能观察到其中一部分。令 Π_n 为嵌套有限划分，其网格相容集合为$\Lambda_{S,\,n}$，网格修复为 $d_{S,\,n}$。

命题 3（细化）。 $\Lambda_S(P) \subseteq \Lambda_{S,\,n+1} \subseteq \Lambda_{S,\,n}$ 以及 $d_{S,\,n} \le d_{S,\,n+1} \le d_S$。若这些划分生成 $\mathcal{B}$，则 $\Lambda_S(P) = \bigcap_n \Lambda_{S,\,n}$ 以及 $d_{S,\,n} \uparrow d_S$。

任何网格上的不相干性都会驳斥该模型；可在网格上测量的对比量的界限在网格细化时会弱收紧；而一个网格上的相容性并不能说明更精细网格上的情况，也不能说明 $\mathcal{Y}$ 上的概率律。粗网格可能掩盖很多信息。

命题4（粗化）。 (*i*) 若 Π 的某个单元包含两个点，则对每个观测分布都有 $d_S \le 1$，且存在一个观测分布仅有一个历史日期，使该界限可达，并且满足 $d_{S,\,\Pi} = 0$。 (*ii*) 设 $\mathcal{Y} = [a, b]$，每个顶点的密度均为利普希茨连续，常数为 L，且 Π 的每个单元都是长度不超过 h 的区间。那么

$$0 \le d_S - d_{S,\Pi} \le \frac{L(b-a)h}{8},$$

且常数 $1/8$无法进一步改进。

在第 (i) 部分中，若在同一单元的两个点 δ_x，δ_y 处有点质量，$\mu^0_{0,\,1} = \mu^0_{1,\,s} = \delta_x$ 以及 $\mu^0_{0,\,s} = \delta_y$则得到候选分布 $2\delta_x - \delta_y$，它在网格上看似相容，但负质量为 1。第 (ii) 部分说明光滑性带来的好处。光滑性并不意味着相容性无关紧要：对于 $[0, 1]$,有效的历史密度，$1 \pm \sin(2\pi x)$ 会产生顶点密度 $1 + 2\sin(2\pi x)$，与均匀分布的混合分布只有在权重不超过 $1/2$时才相容，并且这会使每个足够精细的网格上 $\int \sin(2\pi x)\, d\mu$ 的上界减半（附录 S2.11）。工资分布可能在整数处出现堆积现象，并在最低工资处有尖峰，因此未必具有光滑密度；所以第 (ii) 部分对实际应用具有指导意义，但并非保证。

网格由分析者选择，因此人们可能担心，分析者会通过选择网格来得到想要的答案。命题 3 限制了这种操作的空间。粗化绝不会缩小界限，只会掩盖不可行性。因此，风险在于错误接受模型；而同时对每个区间网格检验相容性，就能消除这一风险。

命题 5（网格选择）。 令 $\mathcal{Y} \subseteq \mathbb{R}$，令 $H_s(y) = v_s$（$(-\infty, y]$）为顶点分布函数，且 $H_\lambda = \sum_s \lambda_s H_s$，令 H_s 用 $H_\lambda = \sum_s \lambda_s H_s$估计 H_s，并假设

最大值$_s \left\|\widehat{H}_s - H_s\right\|_\infty \le \varepsilon$，其中 $\|\cdot\|_\infty$ 为上确界范数。那么对于将 $\mathcal{Y}$划分为区间的每个划分 Π，包括在观察到数据之后选择的划分，

$$\Lambda_{S,\Pi} \subseteq \{\lambda \in \Delta_S : \ \widehat{z}_\lambda^{\Pi} \ge -2\varepsilon\},$$

其中 $\widehat{z}_\lambda^{\Pi}$ 是 $\widehat{H}_\lambda$ 的估计单元质量。特别地，在上确界范数事件上校准的相干性检验，无论使用何种网格，都具有相同的检验尺寸。

证明。对于一个区间单元 $(u,\ w]$，$|\widehat{H}_\lambda(w) - \widehat{H}_\lambda(u) - H_\lambda(w) + H_\lambda(u)| \le 2\left\|\widehat{H}_\lambda - H_\lambda\right\|_\infty \le 2\varepsilon$，最后一步使用下文的引理 1；开区间单元和半开区间单元则通过左极限处理。 □

由于 $\Lambda_S(P) \subseteq \Lambda_{S,\ \Pi}$ 对每个划分都成立，并且命题 5 中的包含关系对同一事件上的所有区间划分都成立，因此在该事件上，$\Lambda_S(P)$ 属于右端项的交集，

$$\widehat{\Lambda}_\infty(\varepsilon) = \left\{\lambda \in \Delta_S : \ \widehat{H}_\lambda(w) - \widehat{H}_\lambda(u) \ge -2\varepsilon \text{ for all } u < w, \ \ -\varepsilon \le \widehat{H}_\lambda \le 1 + \varepsilon\right\},$$

估计分布函数下降幅度始终不超过 2ε，且与 $[0,1]$ 的距离始终不超过 ε 的权重集合；范围条件成立，因为每个相容的 H_λ 都取值于 $[0,1]$。根据到分布函数集合的上确界范数距离的精确公式（定理 S6.1），这些恰好是使 H_λ 与某个分布函数的距离不超过 ε 的权重，因此 $\Lambda_\infty(\varepsilon)$ 是定理 6 的相干性检验所接受的权重集合。按照第 6 节的方法，在 $\Lambda_\infty(\varepsilon)$ 上对比量进行优化，并加上对比量本身的置信带，所得对比量的置信域因而有效，且至少与用同样方法在任一单一网格上构建的置信域一样窄。这提供了一套具体的程序。(1) 预先固定一个报告网格，使其包含所有待报告的截断值及结果变量的自然单元，例如零收入，并据此报告代入界限。(2) 在该网格上运行第 6 节的学生化相干性检验。这是一项证伪检验：任何网格上的不相干性都会推翻模型，因此，在预先指定的网格上拒绝模型，就意味着拒绝模型，无论更粗的网格显示什么。其临界值仅针对该网格校准。(3) 报告基于 $\Lambda_\infty(\varepsilon)$ 构建的置信域，其中 ε 按上确界范数进行校准；该方法不涉及网格选择。由于 H_λ 是一个阶梯函数，其跳跃点位于观测到的结果值处，$\Lambda_\infty(\varepsilon)$ 可通过一个线性规划精确计算：为每个支撑点添加一个变量，用于表示 H_λ 的运行最大值，便可将针对所有数对 $u < w$ 的条件转化为数量与支撑点数量成正比的线性不等式（附录 S11.4）。对于 n 个支撑点，该程序含有 $|S| + n$ 个变量和 $5n - 1$ 个稀疏不等式；在 NSW 和 CPS 样本中，22,339 个不同的收入值对应的每个程序使用标准求解器约需七秒。对于更大的 n，只在部分支撑点施加条件会扩大集合，因此置信域仍然有效。我们不建议选择

在未被检验拒绝的网格中给出最窄界限的那个网格，因为这种规则会根据检验结果来选择网格。

4 分位数与政策排序

对于实值结果，令 $\mathcal{F}_S(P) = \{H_\lambda : \lambda \in \Lambda_S(P)\}$ 非空，并定义包络 $F_S(y) = \min_{\lambda\in\Lambda_S(P)} H_\lambda(y)$ 和 $\overline{F}_S(y) = \max_{\lambda\in\Lambda_S(P)} H_\lambda(y)$，以及分位数函数$Q_F(\tau) = \inf\{y : F(y) \geq \tau\}$。

定理 5 (分布与分位数尖锐性)。 *(i)* $\underline{F}_S$和 $\overline{F}_S$是分布函数。 *(ii)* $\overline{F}_S \in \mathcal{F}_S(P)$ 当且仅当 $\bigcap_y \arg\min_{\lambda\in\Lambda_S(P)} H_\lambda(y) \neq \varnothing$，$\underline{F}_S$亦然。*(iii)* 对于 $\tau \in (0,1)$，

$$\min_{\lambda\in\Lambda_S(P)} Q_{H_\lambda}(\tau) = Q_{\overline{F}_S}(\tau), \qquad \sup_{\lambda\in\Lambda_S(P)} Q_{H_\lambda}(\tau) = Q_{\underline{F}_S}(\tau),$$

并且当 $\Lambda_S(P)$ 是多面体时，上确界可以取到；特别是在任何有限网格上。

（ii）区分了逐点尖锐性与函数尖锐性：一种混合分布可能在某些阈值处达到下包络，另一种则在其他阈值处达到，因此下包络未必是可容许分布。（iii）给出了分位数极值，但集合 $\mathcal{Q}_S(\tau) = \{Q_{H_\lambda}(\tau) : \lambda \in \Lambda_S(P)\}$未必是一个区间。若 $H_p = p\delta_0 + (1-p)\delta_1$且满足 $p \in [0.4, 0.6]$，则对于 $p < 1/2$，中位数为 1；对于 $p \geq 1/2$，中位数为 0，因此 $\mathcal{Q}_S(1/2) = \{0, 1\}$；区间 $[0, 1]$ 是有效但保守的报告。命题 S3.4 给出了使 $\mathcal{Q}_S(\tau)$ 成为区间的条件。这些是反事实分布分位数的界限；分位数之差并非个体处理效应的分位数，而个体处理效应无法由该模型识别。

许多政策问题对于反事实而言是线性的。对于有界得分函数 u 和观测到的处理组分布 $F_T = \mu_{1,\,1}$，令 $W_u(\lambda) = \int u\, dF_T - \int u\, dH_\lambda$。

命题 6 （政策对比与支配关系）。 W_u 的尖锐集合是一个区间，其端点为 $\int u\, dF_T -$最大值$_{\Lambda_S(P)} \int u\, dH_\lambda$和 $\int u\, dF_T -$最小值$_{\Lambda_S(P)} \int u\, dH_\lambda$，且两者均可达到。当且仅当 $F_T \leq F_S$ 逐点成立时，处理组分布才以一阶随机支配每一个可接受的反事实。

—

该证明可由定理 1和线性直接得出。不同得分对应的最优混合分布可能不同，因此多个得分的区间是联合集合的投影；若要同时对多个得分作出陈述，就需要联合优化 $\Lambda_S(P)$。

5 松弛与失效分析

经验界是否令人信服，取决于它对偏离假设 2的稳健性。我们采用邻域

$$\mathcal{M}_S(\tau) = \{Q \in \mathsf{P}(\mathcal{Y}) : \; d_{\mathrm{TV}}(Q, \mu_\lambda) \le \tau \text{ for some } \lambda \in \Delta_S\}, \qquad \tau \ge 0.$$

根据定理2， $\mathcal{M}_S(\tau)$ 当且仅当 $\tau \ge d_S$ 时非空，且 $\mathcal{M}_S(0)$ 是尖锐集合。当参考对象 μ_λ 是概率律时， τ 表示总体中未处理结果可能不同于历史迁移结果的最大占比。当 μ_λ 是带符号的， τ 的一部分会用于移除负质量，因此应将 τ 理解为距离，而不是人群占比。

全变差只是几种选择之一：历史凸包周围的 Kolmogorov 球或 Wasserstein 球在网格上也具有多面体结构，因此下文的项目同样适用。应将两个半径区分开来。**最小修复** d_S 是松弛模型首次变为非空时对应的半径；下文定义的**符号分解半径** 则是结论失效时对应的半径。考虑结果占比上的效应 $\theta = g - Q(A)$，其中 A 是一个非空真事件， $g = \mu_{1,1}(A)$ 是处理组占比，而 Q 是反事实。符号分解半径τ^{sign}是使得 $\mathcal{M}_S(\tau)$ 包含一个满足 $\theta \ge 0$（若结论相反，则满足 $\theta \le 0$）的概率律的最小 τ。对于网格上的带符号向量 z 和单元集合 A，记 $z^+(A) = \sum_{k \in A}[z_k]_+$以及 $b(z) = \sum_k [-z_k]_+$。

命题 7（精确崩溃半径）。 对于有限网格上质量为 *1* 的有符号 z 、非空真单元集合 A 以及 $g \in [0,1]$,

$$\min\{d_{\mathrm{TV}}(q, z) : \; q \text{ a probability vector}, \; q(A) \le g\} = \max\{b(z), \; z^+(A) - g\}.$$

因此，非负效应变得可接受时的半径为

$$\tau^{\mathrm{sign}} = \min_{\lambda \in \Delta_S} \max\big\{b(z_\lambda), \; z_\lambda^+(A) - g\big\},$$

这是一个线性规划，其中两项均在同一个 λ处计算。对于估计值 $\widehat{z}_s$ 和 $\widehat{g}$，若最大值$_s\, d_{\mathrm{TV}}(\widehat{z}_s, z_s) \le \varepsilon$且 $|\widehat{g} - g| \le \delta$，则代入式修复量与 d_S 相差不超过，且 $|\widehat{\tau}_{\text{符号}} - \tau_{\text{符号}}| \le \varepsilon + \delta$。因此，若该事件的概率至少为 $1-\alpha$，则 $[\widehat{\tau}_{\text{符号}} - \varepsilon - \delta, \; \widehat{\tau}_{\text{符号}} + \varepsilon + \delta]$是 τ^{sign}的一个 $1-\alpha$ 置信区间。

证明。下界：根据定理2， $d_{\mathrm{TV}}(q, z) \ge b(z)$，并且对于 $B = A \cap \{z > 0\}$， $d_{\mathrm{TV}}(q, z) \ge z(B) - q(B) \ge z^+(A) - g$。上界：从正值单元中移除质量 $b(z)$，先处理位于 A 中的单元，并将负值单元设为零；所得概率律的距离为 $b(z)$，其

$q(A) \leq g$ 当 $z^+(A) - b(z) \leq g$时成立。否则，从 A 中移除 $z^+(A) - g$，并将 $z^+(A) - g - b(z)$ 添加到 A^c，这是可行的，因为 A 是适当的；距离为 $z^+(A) - g$。稳定性： b 和 $z \mapsto z^+(A)$ 关于全变差是 1-利普希茨的，混合分布的全变差误差至多为最大的顶点误差，而最小化保持利普希茨界限。 □

分别最小化这两项会低估半径（附录S4）。置信区间继承置信带的覆盖率：对于单元质量，使用有限样本置信带时是精确的，使用自助法置信带时则是渐近的。图1 展示了放宽后的界限如何随 τ 在 NSW 和最低工资应用中变宽。半径必须根据人们认为合理的偏离程度来判断，而 NSW 应用提供了一个校准：在那里，实验揭示了约占概率质量三分之一的偏离。半径大于实际观测到的偏离，仅意味着符号仍然保持不变；这并不能验证未放宽的界限，因为其检验尺寸可能严重错误。也不应将差异 $\tau^{\text{sign}} - d_S$ 理解为模型修正后所需的偏离程度：这两个项目分别选择各自的权重。

全变差将一名工人从零收入移至\$30,000 与将她移至\$5,000 同等对待。当此类移动不合理时，Wasserstein 球会根据移动的距离收费，因此是自然的替代方案；在网格上，它仍保持线性规划形式。在结合 CPS 的 NSW 数据中，它反转了上面的比较：当收入可能平均移动\$1,053 时，就业效应失去其符号，而实验揭示的偏离为\$2,630（附录S11.7）。对哪些移动可被接受施加限制，例如单调处理反应——它要求反事实分布函数位于处理组分布函数之上——同样是线性的，并且可以添加到本文的每个项目中。

排除限制。 经济理论有时会指出某项政策在哪些方面不会产生作用。接近 \$9 的最低工资，不应改变时薪为 \$18 的就业人口占比。如果 $\mu^1_{1,1}(B) = \mu^0_{1,1}(B)$ 对属于某个已知类的事件 B 成立，则尖锐集合变为$\{\mu_\lambda : \lambda \in \Lambda_S(P) \ \ \mu_\lambda(B) = \mu_{1,1}(B)$ 对所有这类 $B\}$均成立，这会在网格上增加线性等式（命题 S4.5）。该限制是一项额外的识别假设，而非既定事实：工资溢出效应和构成变化都可能使上尾部发生移动。它既能收紧界限，也可以检验，因为空集意味着迁移假设与排除限制不可能同时成立；此外，还可以在安慰剂时期进行检验。

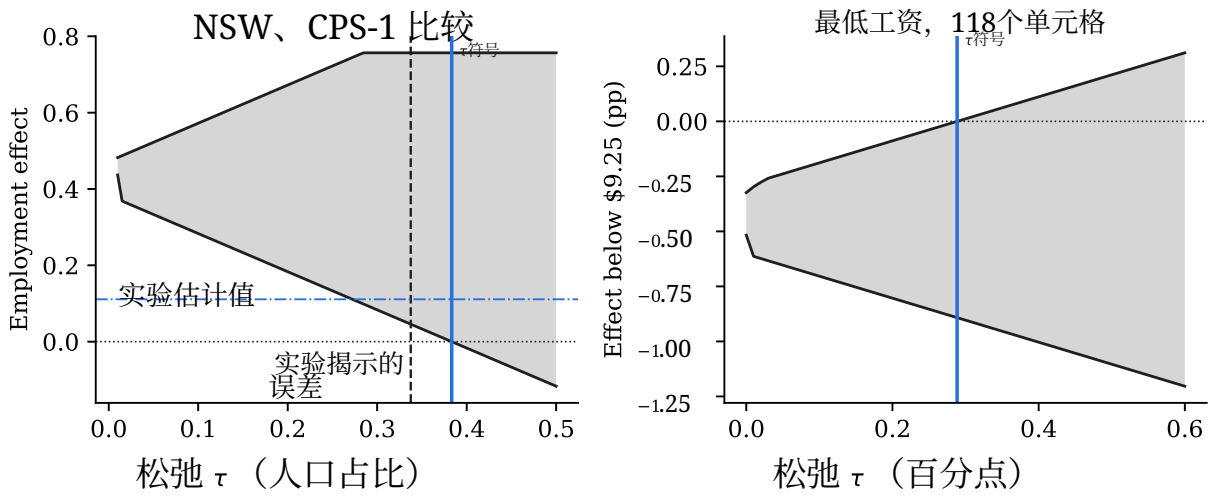


图1：随着与历史迁移的偏离程度增大，放宽后的界限注：阴影部分：在 $\mathcal{M}_S(\tau)$ 下效应的尖锐区间。左图：NSW、CPS-1 对照组、七个单元；当偏离程度低于最小修复 $\tau = 0.009$ 时，区间为空。虚线：实验对照分布与历史凸包之间的全变差距离。右图：最低工资应用，118个单元格；在 $\tau = 0$ 处相容集合非空。

6 推断

6.1 抽样及其理论涵盖的内容

我们考虑三种情形，并分别讨论。第一种情形中，处理组、比较组以及（如果有）基准样本相互独立；每个样本中的单元都是独立抽取的，且具有完整的结果历史；同一单元的结果在不同日期之间可以存在依赖关系。在这种情形下，只要临界值来自德沃雷茨基–基弗–沃尔福维茨不等式，或来自附录S7的样本拆分构造，以下程序在有限样本中即有效。第二种情形中，结构相同，但采用自助法或乘子临界值；对于固定的数据生成过程，随着样本量增加，这些临界值具有有效性；附录 S8 在明确的耦合条件下给出一致有效性。第三种情形中，结果变量是估计分布的单元，且联合抽样分布未知，例如基于 CPS 构建的州—季度工资分箱。本理论不涵盖这种情形，因此我们不报告其置信水平。当单元以相互独立的簇为单位抽样（例如州或村庄），且簇内存在任意形式的依赖关系时，只要置信带是按簇而非单元进行校准，以下覆盖定理便适用于固定网格：附录S8.2给出了有限样本半径和一致有效的高斯校准。加权调查估计量以及只有少数处理簇的研究设计仍不在理论涵盖范围内；若处理组州只有五个，有限样本簇半径约为 1，与单元的尺度相当。表 2总结了各程序所能保证的内容。

表2：推断程序的保证

程序	临界值	保证
置信集，相容性检验，基准检验	DKW–Massart 半径	有限样本水平，独立单元和样本
单个分数的置信集	样本拆分	有限样本水平、有界分数、相依的先导样本和估计样本
置信集、无网格的区域、崩溃区间	自助法历史	完整的固定分布的渐近；在……上一致耦合条件下的附录 S8
学生化相干性检验	自助法，估计标准误	渐近；在模拟中较为保守
置信集与一致性-ence 检验，聚类数据，固定网格	聚类上的霍夫丁	有限样本水平，独立的簇-簇，相等的聚类权重
same	高斯，簇协方差-ance	一致渐近水平，有界的重新报告，聚类方差有下界
估计分布的单元（最低工资）	None	无；结果仅作描述

对于日期 t 和组 g，令 $F_{g,\,t}$ 为给定 $G = g$ 的 Y_t 分布函数，$\widehat{F}_{g,\,t}$ 为基于 N_g 个单元的经验对应量，$H_s = F_{0,\,1} + F_{1,\,s} - F_{0,\,s}$ 为代入估计，且 $H_\lambda = \widehat{\sum_s} \lambda_s H_s$。

对代入估计施加精确相容性是错误的。对于连续结果，每个 $F_{0,\,s}$ 几乎必然会在没有正符号经验分布函数跳跃的地方发生跳跃，因此 $\widehat{H}_\lambda$ 对每个 λ 都会在某处递减，而经验相容集为空（命题 S5.1）。即使在固定网格上，当总体相容集为单点时，代入估计集合也约有一半的时间为空（第 6.4 节）。点估计不相容并非反对该模型的证据。

6.2 顶点约化与置信集

引理 1 (顶点约化)。对于任意半范数 ρ 以及有符号测度 e_s，$s \in S$，

$$\sup_{\lambda \in \Delta_S} \rho\Big(\sum_{s \in S} \lambda_s e_s\Big) = \max_{s \in S} \rho(e_s).$$

事实上，$\rho(\sum_s \lambda_s e_s) \le \sum_s \lambda_s \rho(e_s) \le$ 最大值$_s\, \rho(e_s)$，且在某个顶点处取等号。该引理意味着，无需估计任何最优历史权重；这一点很重要，因为在并列情形下，估计最优值并不规则。令 $\mathbb{D}$ 为 $\mathbb{R}$ 上的分布函数集合，并

定义相容性管

$$\mathfrak{C}_N(\varepsilon) = \bigcup_{\lambda \in \Delta_S} \{F \in \mathbb{D} : \left\| F - \widehat{H}_\lambda \right\|_\infty \le \varepsilon\}.$$

定理 6（相干推断）。若 $\mathbb{P}\big(\text{最大值}_{s\in S} \left\| \widehat{H}_s - H_s \right\|_\infty \le \varepsilon_N\big) \ge 1-\alpha$ 对一个可测的 ε_N 成立，则 $\mathbb{P}\{\mathcal{F}_S(P) \subseteq \mathfrak{C}_N(\varepsilon_N)\} \ge 1-\alpha$。当最小值$_\lambda$下确界$_{F\in\mathbb{D}} \left\| \widehat{H}_\lambda - F \right\|_\infty > \varepsilon_N$ 时予以拒绝的相干性检验，其检验尺寸至多为 α，只要 $\Lambda_S(P) \neq \varnothing$。

证明。在该事件上，每个 $H_\lambda \in \mathcal{F}_S(P)$ 均满足 $\left\| H_\lambda - \widehat{H}_\lambda \right\|_\infty \le$ 最大值$_s \left\| H_s - \widehat{H}_s \right\|_\infty \le \varepsilon_N$ ^根据引理 1。若 $\lambda_0 \in \Lambda_S(P)$，则 $H_{\lambda_0} \in \mathbb{D}$ 位于 ε_N 范围内，且与 H_{λ_0} 在同一事件上。可测性见附录S6。 □

半径 ε_N 可以通过两种方式进行校准。对于 $M = 1 + 2|S|$ 原始分布函数，结合采用 Massart 常数的德沃雷茨基–基弗–沃尔福维茨不等式和并集界，可得有限样本半径

$$\varepsilon_N^{\mathrm{DKW}} = \max_{s\in S}\{r_{0,1} + r_{1,s} + r_{0,s}\}, \qquad r_{g,t} = \left\{\frac{\log(2M/\alpha)}{2N_{g,t}}\right\}^{1/2},$$

^其中 $N_{g,t}$ 是 $F_{g,t}$ 背后的观测值数量（组大小，包含完整历史），在给定样本量的条件下有效，且不对跨日期依赖关系作任何限制（Dvoretzky、Kiefer 和 Wolfowitz，1956年；Massart，1990年）。更尖锐的半径是对 max$_s \left\| \widehat{H}_s^* - \widehat{H}_s \right\|_\infty$ 的自助法分位数，在每个样本中对完整历史进行重抽样（附录 S6）。在网格上，同样的逻辑适用于每个单元。学生化有助于推断：若 $\sigma_{s,k}$是 $\widehat{z}_{s,k}$ 的标准误，且 c 以概率 $1-\alpha$界定 max$_{s,k} |\widehat{z}_{s,k} - z_{s,k}|/\sigma_{s,k}$，那么对所有 k，每个 $\lambda \in \Lambda_{S,\Pi}$ 都满足 $\widehat{z}_{\lambda,k} \ge -c\sum_s \lambda_s \sigma_{s,k}$，这是关于 λ的线性约束。使用自助法临界值和估计标准误时，这一保证是渐近的。当不存在满足该约束的 λ 时，学生化相干性检验便拒绝原假设；其 p值是最大学生化误差超过最小 c 的自助法概率，而对于该最小值，存在某个 λ满足约束。稀有单元（通常会出现不可行性的单元）因此会获得较窄的置信带。只要没有可用的基准，我们就建议将此检验作为默认诊断工具。相容性意味着存在某个 $\lambda \in \Delta_S$ 满足 K 矩不等式 $z_{\lambda,k} \ge 0$，因此，针对带有干扰参数的矩不等式的一般程序（Chernozhukov、Hong 和 Tamer，2007年；Andrews 和 Soares，2010年；Romano、Shaikh 和 Wolf，2014年）可应用于相同约束，并且在边界附近可能不那么保守。这里的结构所带来的优势是顶点约化：它使针对 $|S|$ 候选分布的置信带成为适用于所有混合分布的置信带，并提供有限样本临界值。由于单元数量有限，或者在无网格情形下，半直线的指示函数类为有限类，所涉及的经验过程都是 Donsker 的，而自助法

在独立抽样下，有效性可由标准论证得出（附录S6）。关于功效，有两个事实无需额外假设（命题S8.3）。首先，检验可以使用 $\max_{s,k}[-(\widehat{z}_{s,k} - z_{s,k})]_+$的单侧临界值，该值从不大于双侧临界值。其次，统计量具有精确局部极限（定理 S8.3）。亏损是方向可微的；沿着局部备择 $Z_0 + HN^{-1/2}$，缩放后的统计量收敛于值 $D(H + W)$，它是一个小型线性规划的解，该规划针对相容权重下绑定的单元构建，且权重可在单纯形内移动。由此可得三个事实。零假设极限在随机意义上小于所有顶点单元上的最大值，这正是顶点临界值具有保守性的原因。当相容集合包含一个没有绑定单元的权重时，每个局部扰动仍然相容，极限为零：这类点位于原假设内部，检验在此处是保守的；要有局部功效，每个相容权重都必须有一个绑定单元。根据局部规划估计的临界值（如广义矩选择中的做法）具有逐点有效性且功效更高，但其一致有效性需要我们尚未确立的调节条件；在我们的聚类模拟中，一个粗略版本会导致过度拒绝（附录 S8.2）。对比量的置信区域可通过在同一组 λ上对其进行优化，并加上该对比量本身的置信带来构造；对于分位数效应，处理组分布必须联合覆盖（推论S5.1）。

6.3 针对基准进行检验

相容性是假设 2 的必要推论。当独立来源（例如随机对照组）识别出缺失的$\widehat{}$分布时，可以进行更严格的检验。令 F_R 为该基准分布函数，令 F_R 为其根据独立样本得到的估计。

命题 8（基准检验和迁移误差的置信界）。令 $D = \min_{\lambda\in\Delta_S} \|F_R - H_\lambda\|_\infty$ 和 $\widehat{D} = \min_{\lambda\in\Delta_S} \left\|\widehat{F}_R - \widehat{H}_\lambda\right\|_\infty$。若

$$\mathbb{P}\Big(\max_{s\in S}\big\|(\widehat{F}_R - \widehat{H}_s) - (F_R - H_s)\big\|_\infty \le c\Big) \ge 1 - \alpha,$$

$\widehat{}$则 $\mathbb{P}(|D - D| \le c) \ge 1 - \alpha$。在假设 2下，$D = 0$；当 $\widehat{}$$D > c$ 时拒绝的检验，其检验尺寸至多为 α，且 $D - c$ 是 $1 - \alpha$ 上的置信下界 D。选择$c = r_R + \varepsilon_N^{\mathrm{DKW}}$，其中 r_R 是基准样本量为 N_R 时的半径，并且每个半径均使用 $M + 1$ 替代 M进行计算，在有限样本下是有效的。

证明。对于每个 λ，$\left|\left\|\widehat{F}_R - \widehat{H}_\lambda\right\|_\infty - \|F_R - H_\lambda\|_\infty\right| \le \left\|\sum_s \lambda_s\{(\widehat{F}_R - \widehat{H}_s) - (F_R - H_s)\}\right\|_\infty \le c$在该事件上，根据引理1。对 λ 两边取最小值可保持不等式。 □

$\widehat{D}$ 是线性的 项目在合并样本点上的表现。该检验对于任何

固定 $D > 0$均具有一致性，而置信下界则将拒绝转化为关于幅度的陈述。基准必须识别处理组的未处理分布：在处理组总体的抽样框内进行随机化即可做到这一点；而经选择的实验子样本只能识别抽取该子样本的总体的分布。若没有实验，则可以用政策尚未产生作用的时期作为自然基准：处理组在该时期的观测分布就是其未处理分布，因此，将该时期作为结果日期，并采用该时期自身的历史日期，即可应用该命题。这样的基准检验的是不同日期和时间跨度上的迁移，而不是政策日期上的迁移。拒绝结果表明历史偏差并不稳定，因此有力地说明不应依赖这一假设；未拒绝则只能提供较弱的支持。

6.4 模拟

我们在六个结果单元的三种研究设计中考察检验尺寸、功效和覆盖率；这些设计根据含有零点质量的收入数据进行校准（附录 S10），样本量从（N_1，N_0）=（200，2，000），与 NSW 类似，扩展至（5，000，20，000）。在第一种设计中，历史迁移成立，三个历史顶点之一无效，而相容性使零收入占比的总体界限缩小了 39%。所有检验的拒绝率均低于 1%（1,000 次重复模拟的蒙特卡洛标准误至多为 1.6 个百分点），因此在此设计下检验较为保守；置信区域在至少 99.9% 的重复模拟中覆盖识别区间；随着样本量增大，相容性增益也体现在置信区域中（最大样本量下区域缩短了 19%）。在第二种设计中，基准与凸包的累积分布函数距离为 0.03、0.07 或 0.12；当距离为 0.07 且有 1,000 个处理单元时，自助法基准检验的功效为 0.85；当距离为 0.12 且有 200 个处理单元时，功效为 0.62；而 DKW 检验虽有效，但保守得多。在第三种设计中，当总体修正介于 0.005 和 0.04 之间时，相容集合是一个点或空集。朴素规则“若插件相容集合为空则拒绝”在模型成立时有一半概率拒绝；在边界处，学生化检验在不到 1% 的样本中拒绝，而对于 0.02 的修正，在 1,000 个处理单元下功效为 0.69；未学生化检验在此情形下没有功效。对于聚类数据，我们尝试的所有配置中，单侧高斯校准都能控制检验尺寸，且比双侧校准的功效高 4 至 8 个百分点；接触集规则虽能提高功效，但在 5% 的显著性水平下拒绝率为 6.7%（附录 S8.2）。由此得出两项实践结论：相容性检验应采用学生化方法；而在处理单元仅数百个的样本中，相容性主要是一种诊断工具，只有样本量较大时才能使置信区域更精确。

6.5 处理前数据无法做到什么

处理前结果揭示历史偏差如何变化，并可反驳其隐含集合为空的模型。但它们无法揭示政策后会出现何种偏差。如果两个嵌套模型生成相同的观测分布，却有不同的缺失分布，那么任何检验都无法区分它们；对于较大的模型，任何置信集都无法同时保持有效，并收敛至较小模型的识别集（附录 S9）。增加历史日期会削弱假设 2，并使识别集至少不会缩小。因此，选择能产生最窄区间的时间窗口，实际上是在可信度与表面上的精确性之间权衡；选择 S 是一项识别决策，必须以经济依据加以论证。

7 应用 I：实验基准

7.1 目的与研究设计

本应用探讨历史迁移能否恢复一个已知的分布，以及第6 节的诊断指标能否提醒一位对此并不知情的分析者。事后看来，失败是意料之中的，这也正是这一情境有用之处：问题在于检验能否发现这一失败，以及能否找对原因。新南威尔士州示范项目将弱势劳动者随机分配到补贴性工作中。分配在1975年至1977年间进行，而LaLonde (1986年)将男性样本限定为1976年1月及之后接受分配的人，因此对所有人而言，1975年的收入都早于分配（Dehejia and Wahba，1999年）。我们使用了Dehejia and Wahba (1999年)所选取的男性子样本，其中记录了1974年、1975年和1978年的收入：185名参与者和260名实验对照组成员。对照组为LaLonde (1986年)的CPS-1（15,992名男性）和PSID-1（2,490名男性）样本。收入以1982 年美元计。所有文件均来自Imbens and Xu (2025年)的复现档案。

我们扮演一名没有实验数据的分析者。处理组由185名参与者组成，对照组为CPS-1或PSID-1，历史日期为1974年和1975年，结果日期为1978年。顶点为 $H_s = F_{78}^C + F_s^T - F_s^C$，对应 $s \in \{74, 75\}$，其中F_s^T和F_s^C分别是参与者和对照组在 s年的收入分布。随机分配意味着实验对照组1978年的分布等于参与者的未处理分布，因此它是基准 F_R；它从不用于构建界限。目标是实验子样本中男性在1978年的未处理收入分布，而非所有NSW申请者或调查总体的收入分布。参与者和实验对照组在分配前是平衡的：在下方网格中，他们的全变差距离在1974年为0.06，在1975年为0.09。推断将每个样本视为来自其总体的随机样本。CPS和PSID采用复杂抽样设计；与该领域的其他研究一样，我们的结果未考虑这些设计。

表3：相对于NSW实验基准的历史迁移

	CPS-1	CPS-1 重新加权	PSID-1	PSID-1 重新加权
相容性（七个单元）				
无效顶点（共 2 个）	2	1	2	2
代入估计最低工资修复（百分点）	0.88	0	3.73	0
学生化相干性检验， p 值	0.84	–	0.013	–
对正收益概率的效应				
有符号凸包区间	$[0.38, 0.48]$	$[0.32, 0.41]$	$[0.37, 0.49]$	$[0.38, 0.51]$
相干区间，代入估计	空	$[0.35, 0.41]$	空	$[0.40, 0.40]$
95%外部区域	$[0.23, 0.64]$	$[0.15, 0.58]$	$[0.21, 0.65]$	$[0.16, 0.74]$
变化中的变化（以1975年为基期）	0.41		0.51	
实验基准		0.111 $[0.026, 0.195]$		
基准与历史凸包之间的距离				
全变差 , 七个单元	0.34	0.29	0.35	0.32
累积分布函数距离 $\widehat{D}$	0.32	0.25	0.35	0.38
自助法 / DKW 5% 临界值	0.12 / 0.25	0.14 / –	0.12 / 0.29	0.18 / –
95% 下界（自助法/ DKW） D 0.20 / 0.07		0.11 / –	0.23 / 0.06	0.20 / –
自助法 p-值	< 0.001	0.001	< 0.001	0.001

注： 185名参与者、260名实验对照组成员、15,992名CPS-1男性和2,490名PSID-1男性；历史日期为1974年和1975；结果变量为1978年。“重新加权”根据男性比较组在年龄、教育及其平方项，以及黑人、西班牙裔、已婚和无学位示性函数上的逻辑回归倾向得分赔率，对其权重进行调整；自助法在999次抽样中分别重新估计权重。由于权重是估计得到的，因此未报告DKW值。外部区域在历史权重范围内对效应进行优化，该范围由单元质量和处理组占比的同时自助法置信带所限定。 D 是实验对照组1978年分布函数与最近的迁移后分布函数之间的上确界距离。变化中变化法采用Athey 和 Imbens (2006)的界限，这里两者一致。

预先设定的网格有七个单元：零收入、 $(0, 5]$， $(5, 10]$， $(10, 15]$， $(15, 20]$，$(20, 30]$ 及以上30,000美元。主要结果变量是1978年收入为正的概率。其实验效应为0.111，95%的置信区间为 $[0.026, 0.195]$。所有自助法均在每个样本内对完整的个人历程进行重抽样，共抽取2,000次。

7.2 结果

表 3 和图 2总结了结果。

迁移失败了，基准检验也证实了这一点。实验对照分布与最近的迁移后分布之间的距离为 0.32（CPS）和 0.35（PSID），自助法检验和有限样本检验均拒绝原假设。置信下界表明，在某个收入水平上，每个迁移后分布函数与真实值的差距至少为 20 个百分点。在七个单元上计算全变差，需要移动约三分之一的概率质量。对于政策问题，历史迁移得出的就业效应介于 0.38 和 0.49 之间；即使是 95%外部区域也排除了 0.11 这一实验估计值

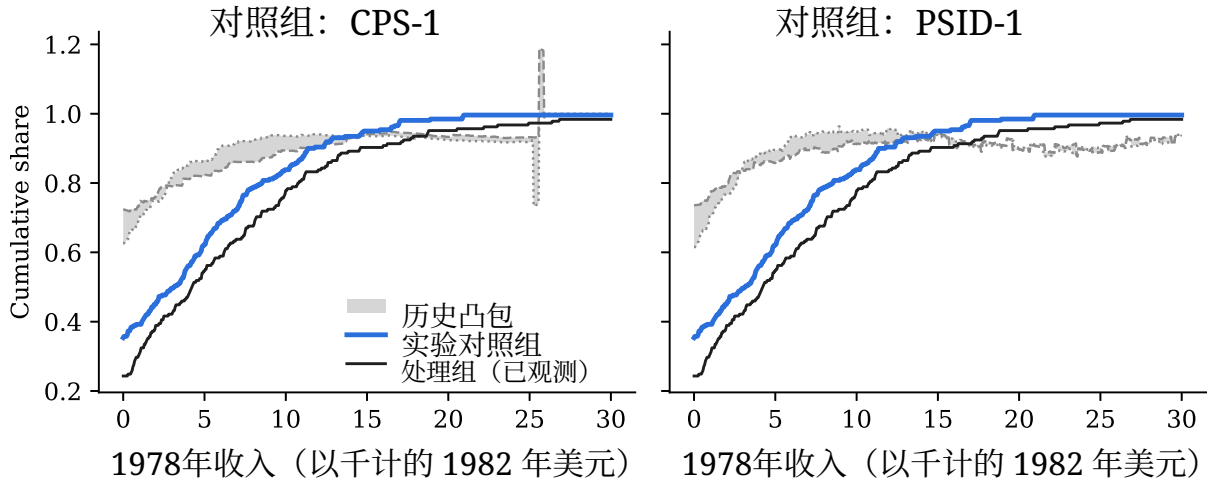


图2：1978年收入：实验对照组与历史凸包注：阴影部分：混合分布的分布函数 $\lambda H_{74} + (1-\lambda)H_{75}$（虚线：$H_{74}$；点线：$H_{75}$）。超过 1 的取值以及 CPS 顶端编码附近的向下跳跃表明，迁移后的候选分布并非分布。

。采用单一基准年的分布差分中的差分给出了有符号区间的端点（1974年的结果为 0.48，1975年的结果为 0.38），而变化中的变化得出的结果为 0.41（CPS）和 0.51（PSID）。所有利用项目实施前差距的方法都将效应高估了至少三倍。

即使没有实验，相容性也能揭示部分问题。对于两个对照组，两个历史顶点在某些收入单元格上都赋予了负质量，且任何混合分布都无法修复这一问题；CPS 的代入式修复量为 0.9 个百分点，PSID 为 3.7 个百分点，位置在\$10,000 至 \$20,000 之间，此处调查分布在 1974—1975年与 1978年之间发生变化。正如第 6节所提醒的，仅凭代入式不相容性并不能证明任何结论。学生化检验仅使用项目实施前数据和对照组数据，对 PSID 拒绝相容性（$p = 0.013$），但对 CPS 不拒绝（$p = 0.84$）。因此，没有实验数据的分析者可能仅凭数据就拒绝基于 PSID 的迁移。网格选择不会左右这些结论：第 3.5节的无网格区域基于联合上确界范数置信带校准，并经精确计算，CPS 为 $[0.21, 0.62]$，PSID 为 $[0.19, 0.67]$；其范围不宽于使用同一置信带、基于 7、12 或 23 个单元得到的区域（附录 S11.4），且不包含实验估计值。

当迁移成立时，检验会拒绝吗？将调查样本替换为实验对照组中的随机一半，并将另一半用作基准，便可通过研究设计使所有历史偏差归零。在200次随机划分中，基准检验拒绝了一次，而学生化相干性检验从未拒绝；累积分布函数距离的中位数则从0.32–0.35降至0.10（附录 S11.5）。上述拒绝并非小样本或稀疏单元格造成的假象。

迁移为何失败。NSW接纳了当时没有工作、且前几个月工作很少的人。资格取决于一个暂时状态，即 Ashenfelter (1978年)所说的项目实施前收入低谷。在1974年和1975年，参与者中无收入者的比例分别为71%和60%，而CPS-1中分别为12%和11%，因此历史非就业差距分别为0.59和0.49。1978年，实验对照组（35%无收入）与CPS-1（14%）之间的差距为0.22。政策后偏差远超历史偏差的值域；用第2.2节的术语来说，政策后环境并不是政策前环境的混合分布。PSID-1呈现出相同的图景（0.62和0.50，相比之下为0.24）。这种失败并非方法本身的偶然结果：当对处理的选择通过前期结果发挥作用时，历史无法界定政策后偏差的范围。

构成与第二个目标组。 将调查样本重新加权，使其符合参与者的人口统计特征后，七单元凸包彼此相容，但迁移仍遭到拒绝，推得的效应仍在0.35至0.40左右。对于示范项目中领取受抚养子女家庭援助（Cal´onico and Smith，2017年）的女性，她们是根据长期福利领取情况选出的；迁移后的就业效应为0.27，而实验结果为0.01，基准检验拒绝了该假设；由于只有一个历史日期，这检验的是单期分布差分中的差分（附录S11）。

失效分析。 对于CPS-1，松弛模型从 $\tau = 0.009$ 起非空，并在 $\tau^{\text{sign}} = 0.38$时首次允许就业效应为零，95%的置信区间为 $[0.26, 0.51]$。实验揭示的迁移偏离程度，即基准到凸包的全变差距离，为0.34。效应的符号在这种偏离下仍然成立，但其大小则不然：足以涵盖真实值的松弛程度允许效应取0.05至0.76之间的任意值（图1，左图）。

该应用表明，在一个典型设定中，历史迁移可能会出现很大的偏差；其原因在项目实施前的数据中既具有经济意义又可被检测出来；第 6节的检验能够检测出这一点，无论是否进行实验，同时在迁移按研究设计成立时不会拒绝迁移。

8 应用二：1993年劳动所得税抵免的扩张

8.1 目的与研究设计

本应用探讨的是：当这一假设值得予以审视时，该方法能够带来什么结果。1993年综合预算协调法案于1993年2月提出、同年8月颁布，在1994–1996年纳税年度提高了有子女家庭的劳动所得税抵免，其中有两个或更多子女的家庭受益最多。标准研究设计比较有子女和无子女的单身女性（EissaandLiebman，1996年；MeyerandRosenbaum，2001年）。

表4：1993年劳动所得税抵免扩展：界限与诊断指标，有子女女性

	1994	1995	1996
就业效应，界限	$[-0.010, 0.014]$	$[0.047, 0.070]$	$[0.115, 0.139]$
95%外部区域	$[-0.07, 0.08]$	$[-0.02, 0.13]$	$[0.05, 0.20]$
无网格95%区域	$[-0.16, 0.16]$	$[-0.10, 0.22]$	$[-0.04, 0.29]$
逐步退出份额效应，界限	$[-0.031, -0.008]$	$[-0.009, 0.014]$	$[0.053, 0.076]$
符号分解半径，就业	0	0.047	0.115
比较分布在其自身凸包内：距离	0.032	0.044	0.077
自助法 p-值	0.76	0.38	0.01

相容性：代入式修复量 0；学生化 $p = 1.00$ （5 个单元，1996年；所有年份共 21 个单元） 1993年安慰剂基准：累积分布函数距离 0.032，自助法临界值 0.081， $p = 0.96$1994–96年间的共同权重：在 5 个单元上可行

注：效应为母亲的观测份额减去基于1991年和1992年迁移所得的反事实份额，采用五单元格网格；在此网格上，无论是否要求相容性，界限都相同。外区域：通过顶点约化（引理 1），结合针对两个顶点效应的联合自助法带。无网格区域：第3.5节。比较法检验：无子女女性的收入分布与其1991年和1992年分布的凸包之间的累积分布函数上确界范数距离。有两个或更多子女的女性的结果见附录 S12。

两组的差异在于是否有子女，而这一特征不会对近期的劳动力市场冲击作出反应，因此，在NSW中导致迁移失效的选择机制并不存在。这里的迁移意味着：如果没有扩张，1994—1996年母亲与无子女女性之间的收入差距，本应类似于1991年和1992年收入差距的混合分布。

我们使用一份公开的3月当前人口调查记录摘录，涵盖20—54岁、未婚且没有高中学历的女性，纳税年度为1991—1996年（共13,746名女性；1996年有1,138名母亲和867名无子女女性）。结果变量是纳税年度的年度收入，以1997年美元计，并包含零点原子；我们将正收入称为就业。（该摘录将年份变量标记为纳税年度；3月调查中的收入指前一个日历年的收入。）摘录还记录了每位女性的居住州。在计算任何按组别和年份划分的结果之前，我们已书面确定研究设计：处理组（有子女女性；次要分组为有两个或更多子女）、对照组（无子女女性）、历史年份（1991年和1992年，这两年在提案之前结束；1993年作为敏感性检验）、网格（零收入，以及大致对应于1996年有两个子女的家庭税收抵免逐步增加、平台和逐步退出阶段及其以外范围的区间，以1997年美元计：$(0, 9{,}000], (9{,}000, 12{,}000], (12{,}000, 28{,}000]$, 以上）、对比量、诊断指标和三项成功标准：相干性检验不拒绝、安慰剂基准不拒绝，以及1996年的就业效应界限排除零。该摘录没有调查权重或设计变量。表4将记录视为相互独立；第8.3 节则改为对完整州进行重抽样。

8.2 结果

诊断检验通过（表4）。每个迁移后的候选分布都是定义在预先指定网格上的分布；无论采用该网格还是21 单元网格，学生化检验均不拒绝。安慰剂基准将1991—1992年差距迁移至1993年；该年法案并未改变税收抵免。结果显示，母亲1993年的实际收入分布与预测相差0.032个累积分布函数距离，远低于自助法临界值。三项成功标准均已满足。

随着税收抵免逐步实施，界限较窄且不断变化：1994年未检测到变化，1995年的就业增幅为4.7至7.0个百分点，1996年为11.5至13.9个百分点，而项目实施前就业率为53%。对于获益最多的有两个或两个以上子女的母亲，1996年的界限为13.4至17.1个百分点。相对于这些区间宽度而言，抽样不确定性较大：95%外部区域在1996年排除了零值，但在1995年没有；而无网格区域虽然避免了使用网格，却没有任何年份排除零值。变化中的变化在1996年给出0.15至0.16，高于传输界限。

第10 节的方案标记了1996年。1994年和1995年，无子女女性的收入分布是其1991年和1992年分布的混合分布，抽样误差除外；而1996年则不是（$p = 0.01$）；她们的就业率从67.7%和69.5%降至64.6%。在环境模型中，1996年是对历史的外推，同时也是福利改革和强劲的劳动力市场改变单身母亲激励机制的一年（Kleven，2019年）。因此，我们将1996年的界限解读为：相对于无子女女性，所有改变单身母亲处境的因素所产生的效应，而非税收抵免单独产生的效应。1995年的界限通过了诊断指标，但只有在1992年至1995年间发生的其他变化没有对母亲产生相对于无子女女性而言的影响时，它才同样是税收抵免的单独效应。这些年份间逐渐扩大的州级福利豁免确实产生了影响；Meyerand Rosenbaum (2001年)认为，尽管并非全部，但1984—1996年单身母亲就业率的上升有很大一部分应归因于劳动所得税抵免和其他税制变化。如果这些豁免提高了母亲的就业率，那么在传输框架下，1995年的界限就高估了税收抵免单独产生的影响。第8.3节排除了豁免州。

8.3 福利豁免与州聚类

州标识符还允许进行两项书面方案中未包含的进一步检验。首先，重抽样完整州而非女性——这允许一个州内跨年份和组存在任意依赖关系——不会使表4外区域的任何端点变化超过0.01： $[-0.01, 0.13]$ 在1995年为 $[0.05, 0.20]$ 在1996年为 。其次，我们剔除有福利豁免的州。每个州首次获得全州性AFDC豁免批准的年份取自Schanzenbach 和 Strain (2021年，表A3)，然后再考察各州的任何结果变量。一种研究设计剔除截至1995年已获批准的27个州，这可能影响1994年和1995年的结果变量；另一种则剔除截至1996年已获批准的全部36个州，即稳健性样本

表5：剔除福利豁免州后，EITC扩展的效应：有子女女性

	所有州	截至1995年未获批准	截至1996年未获批准
州	51	24	15
1995年就业效应，界限	$[0.047, 0.070]$	$[0.055, 0.102]$	$[-0.010, 0.079]$
95%外部区域	$[-0.01, 0.13]$	$[-0.04, 0.20]$	$[-0.16, 0.22]$
1996年就业效应界限	$[0.115, 0.139]$	$[0.084, 0.118]$	$[0.031, 0.120]$
95%外部区域	$[0.05, 0.20]$	$[0.01, 0.20]$	$[-0.06, 0.21]$
1996年比较法检验，p-值	0.02	0.20	0.56

注：界限如表 4所示，基于居住在保留州的女性计算。外部区域和p-值通过对完整州进行重抽样得到。豁免批准年份来自Schanzenbach 和 Strain (2021年，表A3）。1994年以及有两个或两个以上子女的母亲的结果见附录 S12。

Schanzenbach 和 Strain (2021年)的稳健性样本。批准并不等同于实施，而且剔除州会改变总体，因此，这些研究设计检验的是界限对豁免的敏感性，而不是将税收抵免与福利改革区分开来。

表 5 报告了结果。在1995年之前未获批准的24个州中，所有诊断工具均通过，包括1996年的比较法检验；1995年的就业界限为5.5至10.2个百分点，1996年为8.4至11.8个百分点。此处相容性构成约束：1996年的带符号界限达到13.1个百分点，而凸包中没有任何概率分布支持超过11.8个百分点的增益。在截至1996年均未获批准的15个州中，两个历史差距相差更大，界限也扩大为 -1.0 1995年为至7.9个百分点，1996年为3.1至12.0个百分点。对于15个或24个州，抽样不确定性很大，只有24州设计中的1996年区域排除了零。由此得出三点结论。所有研究设计中，1996年的界限符号都保持不变；1995年的界限也保持不变，除非剔除截至1996年曾获任何批准的所有州。豁免州以外未能检出1996年比较法检验的失败，不过在较小样本中，该检验的功效较低。豁免并不能解释迁移下的就业增幅，但州数太少，数据无法将这些增幅分摊到税收抵免和豁免之间。Schanzenbach 和 Strain (2021年)同样发现，剔除豁免州后，其回归估计几乎没有变化。

8.4 什么是共同权重

在预先指定的网格上，相容性约束并未起作用：每个候选分布本身已经是一个分布，因此概率约束在该网格上没有带来任何收窄；而在 21 单元网格上，它将 1996年 的就业效应界限收窄至 $[0.128, 0.131]$ （在该网格上，其他年份以及有两个或两个以上子女的母亲对应的插件相干集合为空，但学生化检验在这些情形下均未拒绝）。即便如此，结果事件之间的共同权重仍然很重要。逐一对比量计算的标量界限表明，1995年，就业效应在 $[0.047, 0.070]$ ，而处于逐步退出区间的母亲比例效应在 $[-0.009, 0.014]$ ，

因此允许任意组合。在迁移约束下，两者沿同一线段共同变化：每个就业增幅超过 6.2 个百分点的反事实都意味着逐步退出比例下降；而每个逐步退出比例上升的反事实都意味着就业增幅低于 6.2 个百分点。税收抵免究竟是以低收入水平吸引母亲就业，还是也使收入向逐步退出区间移动，无法从各自独立的界限中看出；一旦就业增幅确定，答案便随之明确。

9 应用三：2014年各州最低工资

9.1 目的与研究设计

本应用展示了第 3.3 节所述的汇总问题在标准堆叠设计中的表现，以及在现实网格上，概率可行性会在多大程度上改变界限。这些界限是描述性的。我们并未声称最低工资具有因果效应：下文的审计表明，此处的传输假设并不可信，而且这些单元没有可用的抽样分布。

我们使用 Cengiz 等人 (2019年Roth andSant'Anna (2023年提供的州—季度面板数据，该数据由 Roth and Sant'Anna 分发（）。数据涵盖 51 个辖区在 2007—2015 年每个季度的情况，给出按 117 个以 2016年美元计、每 0.25 美元划分的工资区间统计的人均劳动者数量，并设有一个残差单元，其数值等于 1 减去就业人口比。哥伦比亚特区按工作地点统计就业岗位、按居住地统计人口，因此残差为负，故予以剔除。这些向量表示每位居民对应的就业岗位数，而不是个人结果变量的分布：多重就业和通勤使其无法作个人层面的解释。一致性约束并不依赖这种解释。在任何解释下，某工资区间内的反事实就业人数都不可能为负；而在以下每个事件层面的反事实中，负质量都落在工资区间内，绝不会落在残差单元中（附录 S13）。五个州在 2014 年上调了最低工资，此前自 2010年起未曾上调：新泽西州（1 月 1 日生效）、特拉华州（6 月 1 日）、加利福尼亚州（7 月 1 日）、明尼苏达州（8 月 1 日）和密歇根州（9 月 1 日）。每个州均以首次完全受到政策影响的季度作为评估时点，并与从 2010年到该季度期间未上调最低工资的对照州进行比较。历史日期分别提前 4、8、12 和 16 个季度，以保持季节一致。结果变量是低于 $9.25 的就业人口占比变化；当就业岗位消失或工资升至截断值以上时，该比例都会下降。

事件以相同权重堆叠，并对平均分布施加迁移约束。根据定理4 这比逐事件迁移的约束更弱，因此所得界限仅适用于这五个事件合并后的总体。

表6：低于截断值的就业人口占比效应（百分点）

	带符号凸包	相干凸包
低于 $9.25，七个单元	$[-0.614, -0.228]$	$[-0.614, -0.228]$
低于 $9.25，118个单元格	$[-0.614, -0.228]$	$[-0.517, -0.324]$
其他预先指定的阈值，*118*个单元格		
低于 $7.25		$[-0.416, -0.284]$
低于 $11.25		$[-0.442,\ \ 0.450]$
低于 $14.25		$[-0.680,\ \ 0.170]$
低于 $18.25		$[-0.366,\ \ 0.716]$
最低工资修复：堆叠 / 逐事件	0 / 0.303	
符号分解半径 τ^{sign}，低于 $9.25	0.288	
提前一年安慰剂检验：与观测分布的全变差距离	4.92	

注：低于 $9.25 的处理组占比为 7.136%。七个单元包括残差和由五个截点划分的区间。修正值、半径和距离均以总体的百分点表示。不同截点对应的区间不一定能由同一组历史权重达到。

9.2 合并与相容性

在七单元格网格上，每个顶点都是概率向量，相容性不起作用（表6）。在完整网格上，四个顶点都包含负值区间，不过堆叠凸包仍包含相容混合分布，下面的界限$9.25 从 $[-0.61, -0.23]$ 缩小至$[-0.52, -0.32]$，宽度减半；中点几乎没有变化。这正是命题1 所揭示的：达到无约束端点的权重，会在报告结果变量从未提及的区间上赋予负质量。给出这一界限的堆叠反事实是一个复合结果。不存在一组共同权重，能使全部五个事件层面反事实都有效；至少必须迁移总体中 0.3% 的个体才能实现这一点。针对各州分别选择权重也无济于事：每个州自身的相容集合都是空集，而州层面反事实的最小负质量介于 0.03 个百分点（密歇根州）和 0.52 个百分点（加利福尼亚州）之间。堆叠集合之所以非空，只是因为不同州的负值区间在平均时相互抵消。约三分之二的负值区间低于$8.00，此处单元数据稀疏，抽样噪声最大；没有单元的抽样分布，我们无法判断不可行性中有多少源于噪声，又有多少源于迁移失效。这个界限或许仍可作为关于合并总体的结论来辩护，但不能将其视为州效应的平均值。表格的其余部分还提出两点警示：在 $\tau = 0.29$ 个百分点处，符号会反转；而在早一年、任何政策实施之前，最近的迁移分布与观测分布之间的全变差为 4.9 个百分点。

9.3 对传输假设的审计

第2.2节的环境模型中，只有当2014年环境π_1 位于$\{\pi_s$ 内时，界限才具有因果性： $s \in S\}$；历史季度不受该政策影响，且构成保持稳定。每项检查均未通过；详情及完整表格见附录S13 （表S8）。这些检查针对堆叠模型；事件层面模型仅凭相容性不足便已失败。

括定。根据第2.2节，环境模型要求对照州的2014年分布是其自身历史分布的混合分布。事实并非如此：在四个历史日期，对照州的非就业率为43.8%至44.4%，而在政策日期为43.4%，比历史最低值低0.41个百分点（95%区间 $[0.35, 0.47]$，对照州重抽样，忽略了各州之间的共同冲击）；在完整的118单元网格上，2014年对照分布与其自身历史分布凸包之间的全变差相距12个百分点（一年前为13个百分点，部分原因是名义工资在以实际美元定义的分箱之间发生漂移）。在环境模型中，2014年的劳动力市场是对历史的外推，而不是历史的混合分布。

预期与聚合权重。 加利福尼亚州的增幅于2013年9月签署，新泽西州的修正案于2012—2013年通过州议会，二者都发生在最近的历史季度之前或之内。删除该季度后，相容集合变为空集。按总体对事件加权，使加利福尼亚州获得60%的权重，会将相容界限移至 $[-0.01, 0.08]$。等权重描述平均州级事件，而人口权重描述接受处理的州的平均居民；数据无法在二者之间作出选择。

构成与比较组。 假定这些增幅未改变时薪18.25美元或以上人群的就业比例，与政策日期的数据相容，并将界限缩小至 $[-0.51, -0.37]$；在一年前、尚未发生政策变化时，观测到的上尾份额位于凸包值域之外，正如构成变化所会导致的结果。从未接受处理的州或非毗连比较州会使界限向相反方向移动约0.1个百分点。

安慰剂分配。 将实际事件日期赋予随机抽取的五个从未接受处理的州，在 55% 的抽样中会得到一个相容集合；在这些抽样中，有 25% 的上端点不高于实际值（若从样本池中较大的一半抽取安慰剂州，则这一比例为 15%）。无条件来看，所有抽样中有 13% 得到这样的界限。安慰剂分布很宽，因此这项检验对真实效应的检验力较低：它无法证明该界限是虚假的。但它确实表明，在没有任何政策的情况下，相当一部分抽样也会产生类似实际界限的结果。

这些单元格是调查汇总数据，未公布联合抽样分布；而且，处理组州有五个，事件窗口彼此重叠，因此附录 S8.2中的聚类校准无法提供这样的分布：有限样本半径超过 1；对于 118个单元格，高斯近似的误差界为 $\{\log^7(JKC)/C\}^{1/6}$ 量级，对于任何现实的州数都不小。因此，我们在此不报告置信集；崩溃半径 0.29

个百分点，安慰剂距离为 4.9 个百分点，这些都是描述性比较。

9.4 启示

在现实的网格上，概率可行性可以通过远离报告结果的分箱，将分布界的宽度缩小一半。但在这里，该界限经不起审计：政策日期位于历史范围之外，相容集合取决于一个受到前瞻性影响的季度以及事件的加权方式，而事件层面反事实并不可行。可信的分析应在最早的公告或法案颁布之前终止历史，使用既未提前预期政策也不邻近处理州的对照州，利用工人层面数据固定构成，逐个事件施加迁移约束，并为单元获得抽样分布（附录 S13）。

10 结论

使用政策前差距来界定政策后反事实分布，是用一个由历史提供信息的值域替代一个无法检验的等式。该值域必须由分布构成，而报告的结论必须来自其中之一。尖锐集合是历史凸包与概率单纯形的交集。我们刻画了这一限制何时会改变结论、如何证明其遭到违反、为何堆叠事件和粗网格可能掩盖违反情况、要推翻一个符号需要将松弛扩大到何种程度，以及如何将抽样误差纳入检验和置信集，包括针对实验基准的检验，以及一种无需网格选择的置信域。

这些应用展示了正反两面。相容性是必要条件，而且可以显著收紧界限，但它并不能证明方法有效。在环境模型中，当政策实施后环境处于历史环境范围内时，历史迁移成立。在 NSW 中并非如此，因为资格取决于暂时性项目前状态；在最低工资数据中，2014年劳动力市场超出了近期历史范围，而合并界限依赖于不可能的状态反事实。对于 EITC，由于各组之间的差异源于稳定特征，诊断检验通过了1995年，且界限具有信息价值。剔除早期实施福利豁免的州后，增益仍为正；但由于州数较少，无法将增益拆分为税收抵免和豁免各自的贡献。同一组诊断指标表明，1996年是对照组超出其历史值域的一年，而在豁免州以外并未检测到这一偏离。对于应用研究，这提示了一套简短的方案：选择那些由组间稳定差异而非近期结果所决定的设置。在查看政策实施后的数据之前，确定历史日期、对照组、网格和结果变量；并在政策公布之前而非仅在政策开始之前结束历史窗口。

进行学生化相干性检验；若事件较多，还要进行定理4的事件层面检验。检查对照组的当前分布是否为其自身过去分布的混合分布，并确认构成稳定。对于聚类数据，应以簇而非单元为单位校准置信带。报告无网格置信区域和崩溃半径；凡存在基准之处，还应报告迁移检验。

EITC 应用没有实验基准。自然的验证方法是采用一项实验：资格取决于稳定特征，而非近期结果变量；同时，在项目实施前的多个日期，按照与处理组相同的测量方式观测未接受处理的观察性对照组。对于有条件现金转移项目的随机分阶段实施，例如 PROGRESA（Schultz，2004年；ToddandWolpin，2006），如果至少在两个基线轮次中始终如一地测量结果变量，且资格依据家庭贫困指数确定，那么这类实验就符合上述描述。应将实验对照组留作基准，并在查看基准之前固定网格、日期、结果变量和对照组。

数据可用性

NSW 和调查对照样本来自 Imbensand Xu （2025年的公开复现档案，该档案重新分发了 Dehejia and Wahba （1999年）以及 Cal´onicoand Smith （2017年）的文件。州—季度工资面板是随 Cengiz 等人 （2019年）数据一同发布的 Roth and Sant'Anna （2023年）版本；原始数据和程序位于 Cengiz 等人 （2019年）在哈佛 Dataverse 上存放的数据集（doi:10.7910/DVN/TJCTC7）中。复现包利用这些文件复现了论文中的每一个数字。

附录

附录中的结果编号为 S$x.y$；不带前缀 S 的引用指向正文。

S1 模型：解释与尖锐性

S1.1 环境解释

假设2 可能源自环境的聚合混合分布，但环境限制具有实质性。令 R 取值于历史日期的有限集 S。在处理后日期，假设处理组中给定 $R = s$时，假设未处理结果的条件分布为 $M_{g,s}$，并假设对于两组，$\Pr(R = s \mid g) = \lambda_s$。设

$$M_{1,s} - M_{0,s} = \mu^0_{1,s} - \mu^0_{0,s}, \qquad s \in S,$$

其中右侧的测度是历史上观察到的未处理分布。于是，根据全概率定律，处理后未处理组间差异为

$$\mu^0_{1,1} - \mu^0_{0,1} = \sum_{s \in S} \lambda_s (M_{1,s} - M_{0,s}) = \sum_{s \in S} \lambda_s (\mu^0_{1,s} - \mu^0_{0,s}).$$

这只是一种充分的经济学解释。它既不能由概率一致性推出，也无法仅凭处理前拟合进行检验；尤其是，它要求两组具有相同的环境权重。为了解相同制度份额排除了什么，写作 $\lambda_{g,s} = \Pr(R = s \mid g)$，并保留条件分布 $M_{g,s}$。若份额不相同，未处理差异为

$$\sum_s \lambda_{1,s}(M_{1,s} - M_{0,s}) + \sum_s (\lambda_{1,s} - \lambda_{0,s}) M_{0,s}.$$

第二项是带符号的构成差异。组份额相等时，该项为零；而共同聚合冲击通常无法消除这一项。对于每个结果事件，其绝对值不超过 $\frac{1}{2}\sum_s |\lambda_{1,\ s} - \lambda_{0,\ s}|$，因为混合权重差异之和为零，且每个 $M_{0,\ s}(A)$ 都介于零和一之间。除非能基于经济依据对组别特定的环境份额差异加以界定，否则这一界限没有经验内容。如果各组的 R 分布不同，即使每个条件分布都有效，该界限也可能不成立。它也不意味着观测到的历史日期本身就是未观测的随机处理后制度。要作出这种解释，必须有实质性论证将这些环境与历史日期联系起来。

推论 S1.1 (有界组成差异)。 在该模型中，令 $\lambda_1 = (\lambda_{1,\ s})_s$ 并假设 $\frac{1}{2}\sum_s |\lambda_{1,\ s} - \lambda_{0,\ s}| \le \eta$。则 $\mu^0_{1,\ 1} \in \mathcal{M}_S(\eta)$，即第 5 节的松弛模型。因此，根据 $\mathcal{M}_S(\eta)$计算出的每个界限，都是有效的外界；该外界适用于组别特定的环境份额差异至多为 η 的情形，而该节中的半径 τ 可以理解为环境份额差异的界限。

证明。 未处理差异等于 $\sum_s \lambda_{1,s}\nu_s + \kappa$ 其中 $\kappa = \sum_s(\lambda_{1,s} - \lambda_{0,s})M_{0,s}$。由于$\kappa(\mathcal{Y}) = 0$ 且 $|\kappa(A)| \le \eta$对每个 A都成立，如上所示，因此根据引理S1.1， $d_{\mathrm{TV}}(\mu^0_{1,1}, \mu_{\lambda_1}) = \sup_A |\kappa(A)| \le \eta$，且$\mu^0_{1,1}$ 是一个概率律。 □

该推论并未声称其结果是尖锐的：组成模型将 κ 限制为未观测分布 $M_{0,s}$的组合，而 $\mathcal{M}_S(\eta)$并未施加这一限制。

设 $(\mathcal{Y}, \mathcal{B})$ 为一个可测空间。对于有限符号测度 η，记 $|\eta|$ 为其全变差测度，记 $\eta = \eta^+ - \eta^-$ 为其约当分解，并且 $b(\mu) = \mu^-(\mathcal{Y})$。

S1.2 等质量全变差

引理 S1.1 （等质量表示）。 假设 $\eta(\mathcal{Y}) = 0$。那么

$$\sup_{A\in\mathcal{B}} |\eta(A)| = \eta^+(\mathcal{Y}) = \eta^-(\mathcal{Y}) = \frac{1}{2}|\eta|(\mathcal{Y}).$$

证明。 由于 $\eta(\mathcal{Y}) = 0$，正质量与负质量相等。若 P 是一个哈恩正集， $N = P^c$是其负补集，则 $\eta(P) = \eta^+(\mathcal{Y})$，且 $\eta(N) = -\eta^-(\mathcal{Y})$。对于任意 A， $-\eta^-(\mathcal{Y}) \le \eta(A) \le \eta^+(\mathcal{Y})$。结论得证。 □

本引理确定了主论文中采用的约定：

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_A |\mu(A) - \nu(A)|$$

当且仅当 μ和 ν 具有相同的总质量时。

S1.3 相干性缺口的连续性

对于有限个零质量符号测度 $\nu_1, \ldots, \nu_J$以及一个单位质量基准测度 μ_0，定义 $\mu_\lambda = \mu_0 + \sum_j \lambda_j \nu_j$。

引理S1.2 (利普希茨连续性)。 映射 $\lambda \mapsto b(\mu_\lambda)$ 在 $\mathbb{R}^J$ 上是凸且连续的。更准确地说，

$$|b(\mu_\lambda) - b(\mu_{\lambda'})| \le d_{\mathrm{TV}}(\mu_\lambda, \mu_{\lambda'}) \le \sum_{j=1}^{J} |\lambda_j - \lambda'_j| \sup_A |\nu_j(A)|.$$

证明。 度量空间中到固定集合的距离是 1-Lipschitz。根据概率相容性定理， $b(\mu)$ 正是 μ 到概率空间在单位质量测度上的 d_{TV} 距离。第二个不等式是三角不等式。凸性既可由有符号测度范数下到凸集 $\mathsf{P}(\mathcal{Y})$ 的距离的凸性推出，也可直接由逐事件表示得到。

□

因此 有限维单纯形上的最低相干性亏损可以取到 .

S1.4 负质量与最小替代

记 $b(\mu) = \mu^{-}(\mathcal{Y})$ 为单位质量有符号测度的负质量，并令 $\Delta_S = \{\lambda \in \mathbb{R}^S : \lambda_s \geq 0, \sum_s \lambda_s = 1\}$。不受限制的替代质量的最低工资比例衡量不相干性的严重程度；它不是一项新的识别假设。对于总质量为一的有符号测度 μ，定义

$$r(\mu) = \inf\bigl\{\varepsilon \in [0,1] : (1-\varepsilon)\mu + \varepsilon R \in \mathsf{P}(\mathcal{Y}) \text{ for some } R \in \mathsf{P}(\mathcal{Y})\bigr\}.$$

这里， ε 是替代比例， R 是概率律；两者都不是可识别的政策量。

命题 S1.1 (精确的最低替代比例)。对于任意总质量为一的有限符号测度 μ，

$$r(\mu) = \frac{b(\mu)}{1 + b(\mu)},$$

其中 $b(\mu)$ 是其负质量。若 $b(\mu) > 0$，则最低值在 $R = \mu^{-}/b(\mu)$处取得，所得概率律为 $\mu^{+}/(1 + b(\mu))$。对于由单位质量有符号测度构成的有限历史族 $\{\mu_s : s \in S\}$，令 $\mu_\lambda = \sum_s \lambda_s \mu_s$。在所有历史权重中，最小替代比例为

$$r_S = \min_{\lambda \in \Delta_S} r(\mu_\lambda) = \frac{d_S^{\mathrm{TV}}}{1 + d_S^{\mathrm{TV}}}, \qquad d_S^{\mathrm{TV}} = \min_{\lambda \in \Delta_S} b(\mu_\lambda).$$

对于固定的有限结果划分 Π，定义 $d_S^{\mathrm{TV}}{}_{,\Pi} = \min_{\lambda \in \Delta_S} \sum_{A \in \Pi} [-\mu_\lambda(A)]_+$；相应的比例为 $r_{S,\Pi} = d_S^{\mathrm{TV}}{}_{,\Pi}/\bigl(1 + d_S^{\mathrm{TV}}{}_{,\Pi}\bigr)$。对于一个有限划分的嵌套序列，若其生成的$\sigma$-代数在 $\nu = \sum_{s \in S} |\mu_s|$, r_{S,Π_n} 下与结果σ-代数相差一个零测集，则 r_S 递增至。

证明。 令 $b = b(\mu)$，并选取哈恩负集 N，使得 $\mu(N) = -b$。若 $(1-\varepsilon)\mu + \varepsilon R$是概率测度，则在 N上的非负性意味着 $\varepsilon R(N) \geq (1-\varepsilon)b$。由于 $R(N) \leq 1$，必有 $\varepsilon \geq b/(1+b)$。取等号时，对于 $b > 0$，选取 $R = \mu^{-}/b$。有符号组合变为 $\mu^{+}/(1+b)$，这是一个概率测度。对于 $b = 0$， μ 已经是一个概率

律，且最低工资为零。映射 $b \mapsto b/(1+b)$ 在 $[0,\infty)$ 上连续且严格递增；因此，在历史单纯形上的最小化与该变换可交换。对于划分极限，令 $f_n(\lambda) = \sum_{A\in\Pi_n}[-\mu_\lambda(A)]_+$，并令 $f(\lambda) = b(\mu_\lambda)$。细化给出$f_n \le f_{n+1} \le f$。对于每个固定的 λ，$f_n(\lambda) \uparrow f(\lambda)$：使用陈述中的有限支配测度 ν，并写成 $h_\lambda = d\mu_\lambda/d\nu$。由于这些划分是嵌套的，且模去 ν-零测集后生成$\mathcal{B}$，条件期望 $\mathbb{E}_\nu[h_\lambda \mid \sigma(\Pi_n)]$ 在 $L^1(\nu)$ 中收敛到 h_λ（必要时对 ν 归一化）。因此，它们的负部积分收敛到 $\int[-h_\lambda]_+\,d\nu = f(\lambda)$。每个 f_n 在紧单纯形 Δ_S 上连续；此外，f 在此处连续，因为 $|b(\mu_\lambda) - b(\mu_{\lambda'})| \le \|\mu_\lambda - \mu_{\lambda'}\|_{\mathrm{TV}}$、$_{\mathrm{var}} \le \sum_s |\lambda_s - \lambda'_s|\,|\mu_s|(\mathcal{Y})$，其中 $\|\eta\|_{\mathrm{TV}}$、$_{\mathrm{var}} := |\eta|(\mathcal{Y})$。因此，由 Dini定理可知，$f_n \to f$ 在 Δ_S 上一致收敛，进而 $\min_\lambda f_n(\lambda) \uparrow \min_\lambda f(\lambda)$。$x \mapsto x/(1+x)$ 的连续性给出了r_S、$_{\Pi_n}$的所述收敛性。

□

尽管每个历史顶点都可能无效，但它们的混合分布可能是一个概率律。因此，必须在识别限制所使用的同一组共同权重下，将替代比例最小化。它衡量的是负符号质量的数量，而不是结果发生变化的劳动者比例。替换历史凸包之外的质量是一种诊断操作，并不能识别出新的反事实律。

S1.5 精确识别

命题 S1.2（固定历史约束与相容混合分布）。 假设处理前日期的未处理组—日期律以及处理后对照律均可观测。记 $\mu^0_{g,t}$ 为组 $g \in \{0,1\}$ 在日期 t 的未处理潜在结果分布，其中 $g = 1$ 表示处理单元，$t = 1$ 表示处理后日期。固定一个有限且非空的处理前日期集合 S。对于 $s \in S$，令 $v_s = \mu^0_{0,1} + \mu^0_{1,s} - \mu^0_{0,s}$ 表示将处理后未处理偏差固定为时期 s 的取值所隐含的反事实。定义 $\mathcal{H}^+_S = \mathrm{co}\{v_s :\ s \in S\} \cap \mathsf{P}(\mathcal{Y})$，其中 $\mathsf{P}(\mathcal{Y})$ 表示概率律集合。于是：

(i) 在 s处的固定历史限制与未处理反事实概率律相容，当且仅当 $v_s \in \mathsf{P}(\mathcal{Y})$。相容时，其识别出的概率律 v_s 属于 $\mathcal{H}^+_S$。

(ii) 可能存在 $\mathcal{H}^+_S \neq \varnothing$，尽管对于每个 $s \in S$，$v_s \notin \mathsf{P}(\mathcal{Y})$。

(iii) 如果每个 v_s 都是概率律，那么 $\mathcal{H}^+_S = \mathrm{co}\{v_s :\ s \in S\}$。

这些陈述比较历史约束下的候选集；它们并不对这些约束的可信度进行排序。

证明。 (i) 部分成立，因为将偏差固定为 s 后，得到唯一的候选分布 v_s，而它必须是一个概率律。对于 (ii)，考虑两个结果类别 $S = \{s_1, s_2\}$、一个后处理期

控制律（$1/2,\ 1/2$），以及在 s_1 时的历史处理组和对照组控制律（$1/5,\ 4/5$）和（$4/5,\ 1/5$），并在 s_2 时互换其角色。两个迁移后的候选分布为 $v_{s_1} = (-1/10,\ 11/10)$ 和 $v_{s_2} = (11/10,\ -1/10)$。它们都不是概率律，但二者等权混合后得到（$1/2,\ 1/2$）。(iii) 部分由概率律的凸性得出。 □

标准博雷尔实现。

引理 S1.3（任意边际律的实现）。设 $\mathcal{Y}$ 为标准博雷尔空间，且 $\mu \in \mathsf{P}(\mathcal{Y})$。存在一个博雷尔映射 $q\colon [0,1] \to \mathcal{Y}$，使得 $q(U) \sim \mu$，对 $U \sim U[0,1]$ 均成立。

证明。若 $\mathcal{Y}$ 为可数集，将其枚举为 $\{y_j : j \geq 1\}$（有限情形作显而易见的修改），把 $[0,1]$ 划分为连续的半开区间，使其长度为 $\mu(\{y_j\})$，并在第 j 个区间上令 $q(u) = y_j$。如有剩余的原假设零集，可将其映射到 $\mathcal{Y}$ 中任意一点。

若 $\mathcal{Y}$ 为不可数集，则博雷尔同构定理给出一个双射 $\phi : \mathcal{Y} \to [0,1]$，使得 ϕ 和 ϕ^{-1} 均为博雷尔可测。令 $\widetilde{\mu} = \mu \circ \phi^{-1}$，并令

$$r(u) = \inf\{x \in [0,1] : \widetilde{\mu}([0,x]) \geq u\}, \qquad u \in (0,1).$$

映射 r 是博雷尔可测的，且广义概率积分变换给出$r(U) \sim \widetilde{\mu}$。在 $(0,\ 1)$ 上定义 $q = \phi^{-1} \circ r$，并在端点处任意赋值，即可得到$q(U) \sim \mu$。这适用于所有标准博雷尔结果空间，且不要求 $\mathcal{Y}$具有序结构或向量空间结构。

□

尖锐性构造。设 P 为面板数据 $O^P = (G, (Y_t)_t)$或重复横截面 $O^R = (G, T, Y)$的观测数据分布。对于一个有限且非空的历史组集合S，令 $\mathfrak{Q}_S(P)$为诱导出 P 且满足假设1–2的完整数据分布类。

定理 S1.1 (尖锐相干识别)。假设结果空间是标准博雷尔空间，且 P 满足假设*1(i)*。对于重复横截面，假设每个进入组–日期单元均具有正概率。则

$$\left\{\mu^0_{1,1,Q} : Q \in \mathfrak{Q}_S(P)\right\} = \left[\mu^0_{0,1} + \mathrm{co}\{\mu^0_{1,s} - \mu^0_{0,s} : s \in S\}\right] \cap \mathsf{P}(\mathcal{Y}).$$

证明。必要性由假设2以及潜在结果边际分布必须为一个概率律这一要求得出。

为证明充分性，固定右端项上的概率律 $\mu^\star$，并选择能够表示该概率律的 $\lambda^\star \in \Delta_S$。先考虑面板数据。令 $(\Omega_P,\ \mathcal{A}_P,\ P)$ 承载完整的观测数据向量，并定义乘积扩展

$$(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P}) = (\Omega_P \times [0,1], \mathcal{A}_P \otimes \mathcal{B}[0,1], P \otimes \mathrm{Leb}).$$

将 U 记为第二个坐标。引理 S1.3 给出一个可测的 q，满足 $q(U) \sim \mu^\star$。保留 G 以及每个观测结果作为坐标投影。对于每个 $t \le 0$，定义$Y_t(0) = Y_t(1) = Y_t$。在日期 1 定义

$$Y_1(0) = \begin{cases} Y_1, & G = 0, \\ q(U), & G = 1, \end{cases} \qquad Y_1(1) = \begin{cases} q(U), & G = 0, \\ Y_1, & G = 1. \end{cases}$$

这些变量是可测的，因为组事件和所有分量映射均可测；这里不使用 $\mathcal{Y}$ 上的向量空间结构。

~对 P 的观测数据投影恰好是 P。这些定义还逐路径地满足无预期效应和处理一致性。由于 U 与（G，观测数据）相互独立，且两个组的概率均为正，

$$\mu^0_{1,1,\widetilde{P}} = \mu^\star, \qquad \mu^0_{g,t,\widetilde{P}} = \mu^0_{g,t,P} \quad (t \le 0), \qquad \mu^0_{0,1,\widetilde{P}} = \mu^0_{0,1,P}.$$

最后，$\mu^\star$ 的表示给出

$$\mu^0_{1,1,\widetilde{P}} - \mu^0_{0,1,\widetilde{P}} = \sum_{s \in S} \lambda^\star_s (\mu^0_{1,s,\widetilde{P}} - \mu^0_{0,s,\widetilde{P}}),$$

作为有符号测度相等。因此，该扩展属于 $\mathfrak{Q}_S(P)$，并实现所选的目标分布。

对于重复横截面，设 $\mathcal{T}$ 为观测日期的有限集，$p_g = P(G = g)$、$p_{t|g} = P(T = t \mid G = g)$，以及每个进入单元的$K_{g,t} = \mathcal{L}_P(Y \mid G = g,\ T = t)$。正概率单元条件使每个$K_{g,t}$ 由 P唯一确定。对于概率为零、且不受所维持限制约束的单元，在 $\mathcal{Y}$上任取一个概率律。考虑乘积空间

$$\{0,1\} \times \mathcal{T} \times [0,1]^{2|\mathcal{T}|+1}$$

赋予其乘积σ代数。以概率$(p_0,\ p_1)$抽取 G。给定 $G = g$后，以概率$(p_{t|g})_{t \in \mathcal{T}}$抽取 T；令所有均匀坐标与$(G,\ T)$相互独立。引理S1.3 给出可测函数$q_{g,t}$，使得$Z_{g,t} = q_{g,t}(U_{g,t}) \sim K_{g,t}$ 对每个 g，t均成立。令$Y = Z_{G,T}$。对于每个 g，t和 $A \in \mathcal{B}$，

$$\widetilde{P}(G = g, T = t, Y \in A) = p_g p_{t|g} K_{g,t}(A) = P(G = g, T = t, Y \in A).$$

集合 $\{g\} \times \{t\} \times A$ 构成 $\sigma(G,\ T,\ Y)$的生成π系，因此这些等式和概率测度的唯一性表明，所构造的观测数据分布恰好为P。令 $Y_t(0) = Y_t(1) = Z_{G,t}$，其中 $t \le 0$。在日期 1，令 $Y_1(0) = Z_{0,1}$，其中 $\{G = 0\}$，并且

$Y_1(1) = Z_1$，在 $\{G = 1\}$上取值。在 $\{G = 1\}$上，令 $Y_1(0) = q_\star(U_\star)$，其中最后一个均匀坐标与$(G, T, (U_{g,t})_{g,t})$相互独立，且 $q_\star(U_\star) \sim \mu^\star$。在 $\{G = 0\}$上任意指定 $Y_1(1)$，例如 Z_0，1。每个潜在结果坐标都是可测的。一致性和无预期效应逐路径地成立；正的组概率和进入单元概率能够识别未处理组的历史条件分布K_g、$_s$、对照组的处理后分布K_0，1，以及缺失的处理组未处理分布 $\mu^\star$。因此，用于面板情形的同一有符号测度等式验证了假设2。由于每个个体仅贡献一个观测日期，这一构造不会对重复横截面中可观测的任何跨日期依赖关系施加限制。

□

独立性仅用于展示一种可接受的补全方式。该定理确定的是边际分布；它既不假设也不推断缺失的未处理结果与处理结果相互独立。对于处理单元，只要保留所要求的边际分布和观测分布，$Y_1(0)$ 与 $Y_1(1)$ 之间的任何联结函数都是可接受的。

实值结果的可行权重。对于实值结果，记作

$$H_\lambda(y) = F^0_{0,1}(y) + \sum_{s \in S} \lambda_s \{F^0_{1,s}(y) - F^0_{0,s}(y)\}.$$

命题 S1.3（相容权重集合）。对于有限的 S，

$$\Lambda_S(P) = \Delta_S \cap \bigcap_{a<b} \{\lambda : H_\lambda(b) - H_\lambda(a) \geq 0\}$$

是闭的、凸的且紧的。

证明。 H_λ 的尾部性质和右连续性由基本累积分布函数推出；而具有累积分布函数尾部性质的右连续函数，当且仅当其为非递减函数时，才是累积分布函数。对于固定的$a < b$，所显示的增量关于 λ 是仿射的，因此每个约束都是一个闭半空间。将这些闭半空间与紧单纯形相交即可得到结论。 □

命题 S1.4（紧精确累积分布函数集合）。**</style}该映射 从 到**（$\ell^\infty(\mathbb{R})$ $\|\cdot\|_\infty$）是连续的，因此

$$\mathcal{F}_S(P) = \{H_\lambda : \lambda \in \Lambda_S(P)\}$$

是紧且凸的。

证明。对于 λ、$\lambda' \in \Delta_S$，

$$\|H_\lambda - H_{\lambda'}\|_\infty \leq \sum_{s \in S} |\lambda_s - \lambda'_s| \, \|F^0_{1,s} - F^0_{0,s}\|_\infty.$$

因此，该映射是利普希茨的。由于关于 λ 的线性性以及 $\Lambda_S(P)$ 的凸性，可得凸性。 □

S2 相容性何时重要：证明与补充

S2.1 完整带符号凸包

设 v_s、 $\mathcal{H}_S$ 和 $\mathcal{H}_S^+$ 如第3节所述。

命题 S2.1 （完整概率律的无增益判据）。 $\mathcal{H}_S^+ = \mathcal{H}_S$ 当且仅当每个 v_s 都是概率律。在这种情况下，相容性不会改变凸包上任何泛函的值域。如果某个 v_s 无效，则包含关系 $\mathcal{H}_S^+ \subsetneq \mathcal{H}_S$ 是严格的，但某个给定泛函的值域仍可能保持不变。

证明。 概率律构成凸集，因此有效顶点给出有效的凸包。反之，等式意味着每个顶点都属于 P(Y)。集合相等意味着所有泛函像集相等；严格包含则不意味着这一点。例如，在包含三个单元的情况下，若 $v_1 = (-0.1,\ 0。6,\ 0。5)$ 且 $v_2 = (0。2,\ 0。3,\ 0。5)$，在 v_1 上权重至多为 $2/3$ 的混合分布是相容的，但对于每个候选分布，第三个单元的概率都等于 0。5。 □

忽略相容性所损失的宽度，可分解为命题1中刻画的两个端点增益：

命题 S2.2 (忽略相容性的精确宽度代价)。 假设相容集合 Λ_S 非空，且 ℓ 是一个在历史顶点上取有限值的线性泛函。令 $a_s = \ell(v_s)$，定义 $L^+ = \min_{\lambda \in \Lambda_S} \sum_s \lambda_s a_s$ 和 $U^+ = \max_{\lambda \in \Lambda_S} \sum_s \lambda_s a_s$。那么，无约束有符号凸包值域的宽度缩减为

$$\mathsf{G}_\ell = \left(\max_s a_s - U^+\right) + \left(L^+ - \min_s a_s\right) \geq 0.$$

两个加项各自的最优化面刻画见命题 *1*。特别地，$\mathsf{G}_\ell > 0$ 当且仅当至少一个无约束最优化面与 Λ_S的交集为空时成立。仅仅因为某个历史顶点不相容，或因为历史偏差密度线性无关，并不能保证增益严格为正。

证明。 无约束值域的端点为 $\min_s a_s$ 和 $\max_s a_s$，而相容值域的端点为 L^+， U^+。两者宽度相减即得所示结果，且每一项均非负，因为 $\Lambda_S \subseteq \Delta_S$。命题 1刻画了相应项何时为零。作为不存在普遍严格增益这一主张的反例，设政策评分函数在所有历史顶点下的期望值均相同；那么，即使某些顶点无效，两个值域也都是单点集。

□

S2.2 三单元校准

考虑三个有序工资类别，政策后对照分布为（.35，.35，.30），观测到的处理组分布为（.20，.40，.40），以及两个迁移后的历史候选分布

$$v_1 = (.45, .30, .25), \qquad v_2 = (.10, -.35, 1.25).$$

每个候选分布都是在当前对照组概率律上加上一个真实的历史处理组—对照组有符号测度差得到的，但 v_2 本身并非概率律。以权重 t 赋予v_2 时，该候选分布为

$$p(t) = (.45 - .35t, .30 - .65t, .25 + t), \qquad 0 \le t \le 1.$$

概率非负要求 $t \le 6/13$。示例性工资下限以下，反事实份额减去观测份额的最小相容值为

$$\min_{0 \le t \le 6/13} \{.45 - .35t - .20\} = 23/260 \simeq .08846.$$

若在不检查概率一致性的情况下对带符号凸包进行优化，则会得到 $-.10$，这是一个数学上不允许的下界。在一个限制较少、允许针对不同阈值采用不同历史权重且 施加累积分布函数单调性的模型下，同一个负下界也能达到，下一步将对此加以验证。因此，政策排序既取决于概率一致性，也取决于是否要求所有阈值共用同一个历史混合分布。两个下界之差为 $49/260 \simeq 0.18846$，即18.85个百分点。在既定模型下，带符号凸包的端点是不允许的。这个构造示例表明，排除无效候选分布后，下界的符号会发生变化。

阈值特定权重。 对于该示例的第一个阈值，使用第二个历史候选分布： $F(y_1) = .10$。对于第二个阈值，在第二个候选分布上使用权重 $t = .65$ ： $F(y_2) = .75 - .65 = .10$。由此得到的两个坐标是非递减的，并确定了概率律（。10， 0，。90）。该概率律给出的反事实减观测第一类别份额为 $-.10$。它不满足所维持的共同权重迁移限制，因为其第一个阈值权重和第二个阈值权重不同。因此，该示例区分了跨阈值限制与边际累积分布函数单调性。

S2.3 见证程序与累积分布函数-距离检验

命题 S2.3 （有限支持裕度和线性规划）。 设 $\mathcal{Y} = \{y_1, \dots, y_K\}$且 $p_s = (M_s\{y_k\})_{k=1}^K$。令 $p_\lambda = \sum_s \lambda_s p_s$。则

$$d_S^{\mathrm{TV}} := \min_{\lambda \in \Delta_S} \sum_{k=1}^{K} [-p_{\lambda,k}]_+ = \max_{f \in [0,1]^K} \min_{s \in S} (-f^{\mathsf{T}} p_s). \tag{S.1}$$

两个极值都能达到。左侧是线性规划的值

$$\min_{\lambda, r} \sum_{k=1}^{K} r_k \quad \textit{subject to} \quad \lambda \in \Delta_S, \quad r_k \geq 0, \quad r_k \geq -\sum_s \lambda_s p_{s,k} \quad (k = 1, \dots, K),$$

右侧是以下表达式的值

$$\max_{f, t} \; t \quad \textit{subject to} \quad 0 \leq f_k \leq 1 \quad (k = 1, \dots, K), \quad t \leq -\sum_{k=1}^{K} f_k p_{s,k} \quad (s \in S).$$

特别地，当且仅当 $d_S^{\mathrm{TV}} > 0$ 时，历史凸包是不相容的。

证明：对于每个 λ，在立方体上逐坐标最大化可得

$$\sum_k [-p_{\lambda,k}]_+ = \max_{f \in [0,1]^K} (-f^{\mathsf{T}} p_\lambda).$$

单纯形和立方体都是紧凸集，而收益函数是连续双线性的。因此，在 λ 上取最低工资与在 f 上取最大值可以交换。由于单纯形上的线性泛函会在某个顶点处取得其最低值，

$$\min_{\lambda \in \Delta_S} (-f^{\mathsf{T}} p_\lambda) = \min_{s \in S} (-f^{\mathsf{T}} p_s),$$

证明（S.1）。固定 λ，最小可行 r_k 为 $[-p_{\lambda,k}]_+$，由此得到第一个项目。固定 f，最大可行 t为 $\min_s(-f^{\mathsf{T}} p_s)$，由此得到第二个项目。紧致性保证（S.1）中的极值可以取到；尽管在所示项目中，t 在形式上没有下界，但其等价的紧致表述保证了最优解存在。最后，带符号单位质量向量恰好在其所有负部均为零时才是概率向量；连续性和紧致性确保空的相容集合具有 $d_S^{\mathrm{TV}} > 0$。

□

这里，d_S^{TV}是与相容性的总体总变差距离。它不应替代主论文中的经验CDF上确界范数筛查：经验分布与无原子分布在全变差意义下彼此分离，而抽样分析使用的是不同的范数。

S2.4 抽样校正的有限支持集证书

假设结果变量取值于一个有序且预先指定的 K点支持集。经验顶点向量 $\widehat{p}_s$ 和总体向量 p_s 的总质量均为一。令 $F_s(k)=\sum_{i=1}^{k} p_{s,i}$以及 $\widehat{F}_s(k)=\sum_{i=1}^{k}\widehat{p}_{s,i}$，其中 $k<K$。令 $\eta=\max_{s,k<K}|\widehat{F}_s(k)-F_s(k)|$。对于 $f\in[0,1]^K$，记作$V(f)=\sum_{k<K}|f_k-f_{k+1}|$。

命题 S2.4 （一致得分校正与可观测规划）。对于每个实现以及每个 $\varepsilon\geq\eta$，

$$0\leq\widehat{L}_S(\varepsilon):=\max_{f\in[0,1]^K}\left\{\min_s(-f^{\mathsf{T}}\widehat{p}_s)-\varepsilon V(f)\right\}\leq d_S^{\mathrm{TV}}. \tag{S.2}$$

该统计量是以下线性规划的最优值

$$\max_{f,a,t}\quad t-\varepsilon\sum_{k=1}^{K-1}a_k$$

满足 $0\leq f_k\leq 1$、$a_k\geq f_k-f_{k+1}$、$a_k\geq f_{k+1}-f_k$、$a_k\geq 0$ 以及 $t\leq -f\intercal\widehat{p}_s$对每个 s均成立。若一个$\widehat{\ }$可测且可能依赖于数据的 ε 满足 $\Pr\{\eta\leq\varepsilon\}\geq 1-\alpha$，则在相容性条件下，对 $\{L_S(\varepsilon)>0\}\widehat{\ }$进行拒绝的检验尺寸至多为 α。若 $L_S(\varepsilon)>0$，则经验候选凸包与 *CDF* 多面体之间的累积分布函数上确界范数距离严格大于 ε。

$\widehat{\ }$证明。令 $e_{s,\,k}=F_s(k)-F_s(k)$，并注意到 $e_{s,\,0}=e_{s,\,K}=0$，后者由总质量相等可得。离散分部求和给出如下精确恒等式

$$f^{\mathsf{T}}(\widehat{p}_s-p_s)=\sum_{k=1}^{K}f_k(e_{s,k}-e_{s,k-1})=\sum_{k=1}^{K-1}(f_k-f_{k+1})e_{s,k}.$$

由此可得，对所有顶点和所有得分，$|f\intercal(\widehat{p}_s-p_s)|\leq\eta V(f)$同时成立。因此，只要 $\varepsilon\geq\eta$，

$$\min_s(-f^{\mathsf{T}}\widehat{p}_s)-\varepsilon V(f)\leq\min_s(-f^{\mathsf{T}}p_s)\leq d_S^{\mathrm{TV}},$$

其中，最后一个不等式用到了命题S2.3。取最大值得到(S.2)中的上界；$f=0$ 则给出其下界。对于固定的 f，最大的可行 t 为 $\min_s(-f\intercal\widehat{p}_s)$，最小的可行 a_k为 $|f_k-f_{k+1}|$。若 $\varepsilon=0$，则仍可取到一个最小的 a。这证明了该项目，并且由得分立方体上的连续性可知，取到得以保证。

在相容性原假设下，$d_S^{\mathrm{TV}}=0$。在 $\{\eta\leq\varepsilon\}$ 上，不等式（S.2）因此排除了拒绝的可能，无论 f 是如何选取的。反之，若概率向量 q 与某个经验混合分布 ε 的累积分布函数上确界范数距离至多为 $\widehat{p}_\lambda$，则应用同一个分部求和恒等式

于 $\widehat{p}_\lambda - q$可得 $-f\intercal\widehat{p}_\lambda \le \varepsilon V(f)$，因为 $f\intercal q \ge 0$。不等式 $\min_s(-f\intercal\widehat{p}_s) \le -f\intercal\widehat{p}_\lambda$ 蕴含 $L_S(\varepsilon) = 0$。根据有限支持距离规划中的取到性，严格为正的证书因此意味着距离严格大于 ε。 □

事实上，相对于有限支持距离筛选，该证书有一个精确的逆命题，接下来将予以证明。若支持大小固定，$d_S^{\mathrm{TV}} > 0$，$\eta = o_P(1)$，且 $\varepsilon = o_P(1)$，则选取命题 $f^\star$中的一个总体最大化元S2.3。于是 $V(f^\star) \le K - 1$并且

$$\widehat{L}_S(\varepsilon) \ge d_S^{\mathrm{TV}} - (\eta + \varepsilon)V(f^\star) > 0$$

其成立的概率趋近于一。这项一致性主张仅涉及固定的粗化；在粗化后消失的违反情形不在其适用范围内。

定理 S2.1 (与 CDF 距离筛选的精确对偶性**)。** 假设 $K \ge 2$, S 有限且非空，并且每个经验向量 $\widehat{p}_s \in \mathbb{R}^K$ 的总质量均为一；其坐标可以取负值。记 $\widehat{F}_\lambda(k) = \sum_{i=1}^k \widehat{p}_{\lambda,\, i}$，其中 $\widehat{p}_\lambda = \sum_s \lambda_s \widehat{p}_s$，并令 $\mathbb{D}_K$ 为这些 K 分箱上的 *CDF* 多面体。那么

$$\widehat{d}_{S,K} := \min_{\lambda\in\Delta_S,\, Q\in\mathsf{P}_K} \max_{k<K} |\widehat{F}_\lambda(k) - F_Q(k)| = \max_{\substack{f\in[0,1]^K \\ V(f)\le 1}} \min_{s\in S}(-f^{\mathsf{T}}\widehat{p}_s). \tag{S.3}$$

对于每个 $\varepsilon \ge 0$，命题 *S2.4* 中的调整后得分规划满足

$$\widehat{L}_S(\varepsilon) > 0 \quad\Longleftrightarrow\quad \widehat{d}_{S,K} > \varepsilon.$$

严格不等式至关重要：当管状区域在距离 ε处与 *CDF* 多面体相切时，两种规则都不会拒绝。

证明。对于任意零质量向量 $a \in \mathbb{R}^K$，令 $A_k = \sum_{i=1}^k a_i$，则 $A_0 = A_K = 0$。令

$$\mathcal{V}_0 = \{f \in \mathbb{R}^K : f_K = 0,\ V(f) \le 1\}.$$

这是一个非空紧凸集：由 $f_K = 0$ 和 $V(f) \le 1$ 可得 $|f_k| \le 1$。分部求和给出 $f\intercal a = \sum_{k<K}(f_k - f_{k+1})A_k$。$\mathcal{V}_0$ 中的得分相邻差分取遍 ℓ_1^{K-1}的闭单位球，因为在固定 $f_K = 0$ 后，它们唯一确定一个得分。因此

$$\max_{f\in\mathcal{V}_0} f^{\mathsf{T}} a = \max_{k<K} |A_k|. \tag{S.4}$$

单纯形 Δ_S、概率单纯形P_K，以及 $\mathcal{V}_0$ 都是紧凸集。由（S.4）以及

将有限维双线性极小极大应用于（λ，Q）和 f，

$$\widehat{d}_{S,K} = \max_{f \in \mathcal{V}_0} \min_{\lambda, Q} f^{\mathsf{T}}(Q - \widehat{p}_\lambda) = \max_{f \in \mathcal{V}_0} \left\{ \min_k f_k - \max_{s \in S} f^{\mathsf{T}} \widehat{p}_s \right\}.$$

每个 $\widehat{p}_s$ 的总质量均为一。将 f 替换为 $g = f - (\text{最小值}_k\, f_k)\mathbf{1}$ 不会改变大括号内的表达式以及 $V(f)$。新的得分最低为零，其最大值至多等于其变差，因此至多为一。由此可知，前一个最大值不大于（S.3）右侧的值。反之，任何 $g \in [0,1]^K$ 满足 $V(g) \leq 1$ 的对象都可以替换为$f = g - g_K \mathbf{1} \in \mathcal{V}_0$，此时大括号内的表达式变为最小值$_k\, g_k$ –最大值$_s\, g^{\mathsf{T}} \widehat{p}_s \leq$最小值$_s(-g^{\mathsf{T}} \widehat{p}_s)$。事实上，后一个问题的任一最优解都可以减去其最小值，从而使目标值增加该最小值，同时保持可行性。因此存在一个最小值为$_k\, g_k = 0$的最优解，两个最大值相等。这就证明了（S.3）。

如果 $\widehat{d}_{S,\,K} > \varepsilon$，则（S.3）中的最大化得分 g满足 $V(g) \leq 1$且最小值$_s(-g^{\mathsf{T}} \widehat{p}_s) = \widehat{d}_{S,\,K}$。其$\widehat{\quad}$$\widehat{\quad}$在调整后得分规划中的值至少为 $d_{S,\,K} - \varepsilon V(g) \geq d_{S,\,K} - \varepsilon > 0$。反之，命题 S2.4表明，正的调整后得分会排除在$\widehat{\quad}$累积分布函数距离 ε以内的经验凸包中的概率向量。定义 $d_{S,\,K}$ 的最低值在紧集上可以取得，因此它必定严格大于 ε。在等号成立时，后一蕴含关系排除了正的调整后得分，而零得分给出 $L_S(\varepsilon) = 0$。

□

公式（S.3）还给出了原始距离计算的直接对偶线性规划：最大化t，满足 $t \leq -f_{\mathsf{T}}\, \widehat{p}_s$对每个 s、$0 \leq f_k \leq 1$均成立，并且辅助变量 $a_k \geq |f_k - f_{k+1}|$满足 $\sum_{k<K} a_k \leq 1$。最优得分可识别出哪些分箱（可能是多个分箱的组合）导致候选分布凸包与任何累积分布函数均不相容。该恒等式仅适用于固定网格。它保留了论文原有的累积分布函数距离门槛，不会改变其有限样本校准或设计渐近校准。

S2.5 在标量偏差稳定性下与稳健双重差分的比较

Ban 等人 (2026年) 将标量选择偏差稳定性与几种不同的松弛相结合：无预期效应下的平行趋势稳健性、较强平行趋势条件下的有界预期效应、联合放宽的模型，以及最小数据一致性放宽。他们的参数是处理组平均处理效应的标量。其尖锐性构造针对的是这一标量目标，而非整个缺失的边际分布。他们修订版的附录 F.1.4 还将“分布鲁棒”一词用于一种敏感性包络，该包络通过在决策损失下改变标量基线选择偏差的概率分布而得到。这一构造恢复了他们的标量 ATT 区间；它不是未处理结果分布的精确识别集。二者的区别在于被分布的对象不同，而不只是术语的差异。

本文保留无预期效应假设，并对有符号测度施加历史限制。两个研究项目的识别目标不同，其识别条件不应混为一谈。

下面的初等等价关系给出了“概率律层面的限制强于标量偏差凸包限制”这一论断的确切边界。令 $\mathcal{B}_b(\mathcal{Y})$ 表示定义在 $(\mathcal{Y},\ \mathcal{B})$ 上的有界可测实函数，并令 S 为非空有限集。命题中的所有测度都是有限有符号测度。

命题 S2.5 （完整标量族与测度凸包）。给定有符号测度 ν_*、ν_s、$s \in S$，以下各项等价：

1. 存在一个 $\lambda \in \Delta_S$，使得 $\nu_* = \sum_{s\in S} \lambda_s \nu_s$ 为有符号测度。

2. 对任意 $f \in \mathcal{B}_b$（$\mathcal{Y}$）</style>，

$$\min_{s\in S} \int f\, d\nu_s \le \int f\, d\nu_* \le \max_{s\in S} \int f\, d\nu_s.$$

对于固定的有限个示性函数，类似的陈述通常并不成立。

证明。第一项通过积分推出第二项。反之，对于任意有限个可测集 A_1，.....，A_m，令 $v_* = (\nu_*(A_j))_{j=1}^m$，并令 $v_s = (\nu_s(A_j))_{j=1}^m$。如果 v_* 位于 v_s 的紧凸包之外，则有限维严格分离定理将给出 $a \in \mathbb{R}^m$，使得 $a^\top v_* > \max_{s\in S} a^\top v_s$（如有必要，取逆 a）。有界可测评分函数 $f = \sum_{j=1}^m a_j \mathbf{1}_{A_j}$ 将与第二项矛盾。因此，该集合

$$L_{A_1,\dots,A_m} = \left\{ \lambda \in \Delta_S : \nu_*(A_j) = \sum_{s\in S} \lambda_s \nu_s(A_j) \text{ for } j = 1,\dots,m \right\}$$

对每个有限列表都非空。这些集合在紧单纯形 Δ_S 中是闭集，并具有有限交性质。因此，它们对所有可测 A 取交后仍非空。交集中的任意 λ 都能在每个可测集合上实现相等，因而有符号测度相等。对于固定的有限个示性函数，交集未必包含能匹配该集合之外事件的权重；逐事件权重在所报告的示性函数之间也可能不同。有限列表的界限并不能推出完整的有符号测度相等。

□

例如，取三个结果类别，以及历史有符号偏差质量向量 $\nu_1 = (0.2, -0.2, 0)$ 和 $\nu_2 = (0, 0.2, -0.2)$。候选分布 $\nu_* = (0.15, 0, -0.15)$ 的每个坐标都介于对应历史最小值和最大值之间。然而，其第一个坐标要求对 ν_1 使用 0.75 的权重，而第三个坐标要求对 ν_1 使用 0.25 的权重。因此，分别

对全部三个类别示性函数都有效的界限，仍允许出现一个无法由任何共同历史混合分布生成的偏差向量。利用这些类别上的均匀控制分布，可以将这三个偏差向量都构造为概率向量之差。缺少的限制是共同权重相容性，而非各个示性函数界限失效。

该命题并不意味着 Ban 等人(2026年的标量框架能够计算分布锐界集合或概率有效的历史混合：其主要目标和假设各不相同。但该命题表明，在比较“强于标量界限”时，必须说明所比较的标量评分函数族。对于给定的有界政策评分函数，不受限制的历史凸包优化已经可以得到相应的标量极值。只有当某个优化解被排除在概率有效凸包之外时，概率一致性才能收紧这些极值。此处的无前瞻性分布模型并不能推出作者提出的前瞻性和稳健于模型错设的构造；这些构造需要单独的识别限制和推断。

S2.6 精确共同见证对偶与最低工资总变差松弛

设 S 为有限集， μ_s 为定义在同一可测空间上的有限符号单位质量测度。记 $\mu_\lambda = \sum_s \lambda_s \mu_s$ 为 $\lambda \in \Delta_S$ 和 $b(\mu) = \mu^-(\mathcal{Y})$。以下等式不要求有限结果支持集。

命题 S2.6 （精确符号测度见证和最小松弛）。设$\rho = \sum_s |\mu_s|$。则

$$\rho_* = \min_{\lambda \in \Delta_S} b(\mu_\lambda) = \max_{f \in L^\infty(\rho),\, 0 \le f \le 1} \min_{s \in S} \left\{ - \int f \, d\mu_s \right\}.$$

最大值可以取到。此外， ρ_* 是满足存在 $\lambda \in \Delta_S$和 $Q \in \mathsf{P}(\mathcal{Y})$ 且 $d_{\mathrm{TV}}(Q, \mu_\lambda) \le \tau$的最小 $\tau \ge 0$ 。对于划分为 J个单元的情形，后一个可行性问题是一个线性规划。

证明。 哈恩分解给出，对于每个带符号的单位质量μ， $b(\mu) = \max_{0 \le f \le 1} - \int f \, d\mu$ 。单纯形 Δ_S 是紧且凸的；序区间 $\{f \in L^\infty(\rho) : 0 \le f \le 1\}$是弱*紧且凸的。每个带符号的 μ_s 都有一个 $L^1(\rho)$密度，因此，双线性收益关于 λ 连续，并关于 f弱*连续。极小极大定理交换$\min_\lambda$ 与$\max_f$。在线性泛函上对单纯形取最低工资会选出一个顶点，从而得到所显示的恒等式。紧致性保证取到。最后，概率相容性定理将$\min_{Q \in \mathsf{P}} d_{\mathrm{TV}}(Q, \mu_\lambda)$与 $b(\mu_\lambda)$对应起来。对 λ取最低工资即可证明松弛结论。在 J单元上，引入$z = \sum_s \lambda_s z_s$、 $q \ge 0$ 以及$\sum_j q_j = 1$，并引入非负松弛变量 $u_j \ge q_j - z_j$、 $u_j \ge z_j - q_j$。约束$\sum_j u_j \le 2\tau$强制满足总变差邻域。所有约束都是线性的。

□

ρ_* 的零值意味着至少存在一个相容混合分布；个别历史顶点仍可能不相容，政策端点也仍可能收紧。正值意味着，在所考察的总体分布下，所维持的历史凸包模型被证伪。此时，最小总变差松弛只是透明的*与模型一致性的距离*，并不意味着实质性的反事实确实已经移动了该距离。

S2.7 堆叠与逐事件相容性

现在，我们明确说明：若要将历史限制应用于发生在不同日历季度的事件平均值，需要满足哪些条件。所有对象都是定义在同一个可测结果空间（$\mathcal{Y}$，$\mathcal{B}$）上的概率分布或有限符号测度。令 $e = 1, \ldots\ldots, E$ 索引政策事件。在事件 e处，将 C_e 记为结果日期对照总体的概率分布，将$D_{eh} = T_{eh} - C_{eh}$ 记为偏移量 $h \in H$处历史处理组减对照组符号测度，其中 T_{eh} 和 C_{eh} 是概率分布，且 H 是一个固定的有限偏移量集合。特别地，每个 D_{eh} 的质量都为零。令 Δ_H 为历史权重构成的单纯形。取非负事件权重 w_e，使其总和为一，并且权重固定且独立于结果变量。在工资应用中，$E = 5$ 和 $w_e = 1/5$。定义

$$Q_\lambda = \sum_{e=1}^{E} w_e \Big(C_e + \sum_{h \in H} \lambda_h D_{eh} \Big), \qquad \lambda \in \Delta_H.$$

只有当各单元均非负时，聚合表达式才是候选概率分布。即使聚合候选分布是概率分布，事件特定候选分布对于每个 λ也未必都是概率分布。

命题 S2.7（异质事件权重未必能够汇总）。 假设某个事件的未处理分布满足

$$Q_e^0 = C_e + \sum_{h \in H} \lambda_{eh} D_{eh} + R_e, \qquad \lambda_e \in \Delta_H,$$

其中 R_e 是一个零质量的有限符号测度。定义堆叠的未处理分布 $Q^0 = \sum_e w_e Q_e^0$，并对于一个零质量有符号测度 ν，$\|\nu\|_{\mathrm{TV}} =$ 上确界$_{A \in \mathcal{B}} |\nu(A)|$。那么对于每个$\lambda \in \Delta_H$，

$$\|Q^0 - Q_\lambda\|_{\mathrm{TV}} \le \sum_{e=1}^{E} w_e \left\{ \|R_e\|_{\mathrm{TV}} + \sum_{h \in H} |\lambda_{eh} - \lambda_h| \, \|D_{eh}\|_{\mathrm{TV}} \right\}. \tag{S.5}$$

即使对所有事件均有 $R_e = 0$，且每个事件都满足精确的历史混合分布限制，也可能不存在满足 $Q^0 = Q_\lambda$ 的共同 λ。

证明。将两个表达式相减并利用线性性质： $Q^0 - Q_\lambda = \sum_e w_e\{R_e + \sum_h (\lambda_{eh} - \lambda_h) D_{eh}\}$。有符号测度全变差半范数的三角不等式给出 (S.5)。为说明该不等式可能严格不成立，取两个权重相等的事件、两个历史偏移量，以及二元结果空间 $\{0, 1\}$。在每个事件中，比较分布在结果日期和两个偏移量处均为 δ_0 ， $C_e = C_{e1} = C_{e2} = \delta_0$。令 $(T_{11}, T_{12}) = (\delta_0, \delta_1)$，且 $(T_{21}, T_{22}) = (\delta_1, \delta_0)$。两个事件的未处理事件分布均为 $Q_e^0 = \delta_0$，其事件特定权重分别为 $\lambda_1 = (1, 0)$ 和 $\lambda_2 = (0, 1)$，且 $R_1 = R_2 = 0$。它们的堆叠分布为 δ_0。但两个聚合历史顶点均为$\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$，因此任何共同权重 Q_λ 都是同一个混合分布，无法等于 Q^0。此构造中的所有分量分布都是有效的概率分布。因此，概率可行性并不能消除事件权重聚合误差。

□

该命题给出的是一个界限，而非估计得到的约束：事件级权重向量λ_e 和残差 R_e 均不可观测。若从经济学论证出发，认为所有事件共享同一历史响应机制，则会施加约束 $\lambda_e = \lambda$；较弱的论证则可以为它们之间的差异提供实质性的上界。仅凭政策前拟合良好的堆叠 历史，既不能确立上述任一条件。

命题 S2.8 （事件层面相干性的精确有限网格检验）。 固定 E、 H、 $J < \infty$、满足 $\sum_e w_e = 1$ 的事件权重 $w_e \geq 0$，以及单位质量有符号向量 $v_{eh} \in \mathbb{R}^J$。对于共同历史向量 $\lambda \in \Delta_H$，令

$$p_e(\lambda) = \sum_{h=1}^{H} \lambda_h v_{eh}, \qquad \bar{p}(\lambda) = \sum_{e=1}^{E} w_e p_e(\lambda).$$

Let

$$\Lambda_{\text{stack}} = \{\lambda \in \Delta_H : \bar{p}_j(\lambda) \geq 0 \text{ for all } j\},$$
$$\Lambda_{\text{event}} = \{\lambda \in \Lambda_{\text{stack}} : p_{e,j}(\lambda) \geq 0 \text{ for all } e, j\}.$$

假设 $\Lambda_{\text{stack}} \neq \varnothing$。定义

$$m_{ej} = \min_{\lambda \in \Lambda_{\text{stack}}} \sum_{h=1}^{H} \lambda_h v_{eh,j}.$$

当且仅当对于每个事件—单元对（e、 j）， $m_{ej} \geq 0$ 成立时， $\Lambda_{\text{event}} = \Lambda_{\text{stack}}$ 成立。每个 m_{ej} 都是一个线性规划所达到的最优值，该线性规划包含 H 个单纯形变量和 J 个堆叠非负约束。若某个 $m_{ej} < 0$，其最小化权重便是严格包含关系 $\Lambda_{\text{event}} \subsetneq \Lambda_{\text{stack}}$的一个明确见证。若每个 v_{eh} 都是概率向量，则无需求解这些线性规划，等式即成立。该充分条件的逆命题不成立。

证明。 每个 $p_e(\lambda)$ 和 $\bar{p}(\lambda)$ 的总质量均为一。堆叠可行集是非空紧致多面体，因此定义 m_{ej} 的每个线性目标函数都能达到其最低值。每个 $\lambda \in \Lambda_{\text{stack}}$

都满足逐事件不等式，当且仅当该集合上的每个最低值都非负。如果某个最低值为负，其优化解属于 Λ_{stack} ，但违反了某个定义逐事件不等式。如果每个单独的历史顶点都是概率向量，则由凸性可知，对于所有 $\lambda \in \Delta_H$， $p_e(\lambda) \geq 0$ 。为说明逆命题不成立，取两个聚合权重相等的事件、一个历史，以及向量 $v_{11} = (-1/10,\ 11/10)$ 和 $v_{21} = (2/10,\ 8/10)$：它们的堆叠平均值为 $(1/20,\ 19/20)$，尽管第一个事件向量并非概率有效。反过来，即使存在无效顶点，两个权重集合也可能相等：对于两个历史和相同的事件， $v_{e1} = (-1/10,\ 11/10)$ 和 $v_{e2} = (2/10,\ 8/10)$，堆叠非负性与逐事件非负性等价，尽管第一个历史顶点无效。

□

对于固定的穷尽划分，该检验是一种总体刻画。若在未进行联合抽样误差控制的情况下将其应用于估计的单元质量，则它只是代入式诊断工具。它并未断言粗划分上的相容性在细化后仍然成立。

命题 *S2.9*（事件特定权重）。假设每个事件 e 都有其自身的权重 $\lambda_e \in \Delta_S$ ，且 $\mu_{e、\lambda_e} \in P(\mathcal{Y})$，并记 $\Lambda_e = \{\lambda \in \Delta_S : \mu_{e、\lambda} \in P(\mathcal{Y})\}$。（*i*）平均未处理分布的尖锐集合是闵可夫斯基平均 $\mathcal{Q}^{\mathrm{sp}} = \sum_e \omega_e \{\mu_{e,\lambda} : \lambda \in \Lambda_e\}$；当且仅当每个 Λ_e 均非空时，该集合非空。对于有界得分函数 u ，其值域为 $\sum_e \omega_e [min_{\Lambda_e} \int u\, d\mu_{e、} max_{\Lambda_e} \int u\, d\mu_{e\lambda}]$。（*ii*）逐事件集合 $\{\bar{\mu}_\lambda : \lambda \in \Lambda^{\mathrm{ev}}\}$ 包含于 $\mathcal{Q}^{\mathrm{sp}}$。（*iii*） $\mathcal{Q}^{\mathrm{sp}}$ 和堆叠集合 $\{\bar{\mu}_\lambda : \lambda \in \Lambda^{\mathrm{st}}\}$ 均不一定包含对方。

证明。(i) 各事件的权重彼此不受限制，因此可达平均概率律的集合是各事件层级集合的加权和，且线性得分可按事件逐一优化。(ii) 取 $\lambda_e = \lambda$ 对所有 e均成立。(iii) 在定理 4(ii) 的双单元示例中， $\Lambda_1 = \varnothing$，因此 $\mathcal{Q}^{\mathrm{sp}}$ 为空，而堆叠集合非空。反过来，对于 $\{0,1\}$ ，有两个事件、两个日期、相等的权重，且顶点为 $v_{1,\ 1} = v_{2,\ 2} = \delta_0$ 和 $v_{1,\ 2} = v_{2,\ 1} = \delta_1$，每个顶点都是概率律，因此 $\mathcal{Q}^{\mathrm{sp}}$ 是 $\{0,1\}$上所有概率律的集合，而两个平均顶点都等于 $\frac{1}{2}(\delta_0 + \delta_1)$，堆叠集合则仅包含该概率律。

□

在最低工资应用中，在118单元网格上，每个 Λ_e 都为空：新泽西、加利福尼亚州、特拉华州、密歇根州和明尼苏达州的事件层级混合的最低负质量分别为0.05、0.52、0.23、0.03和0.15个百分点。堆叠集合之所以非空，只是因为不同事件的负值区间在平均值中相互抵消。

命题 S2.10 (近似共同权重)。 对于 $\delta \geq 0$ 令 Λ^δ 为数组的集合 $(\bar{\lambda}, \lambda_1, \ldots, \lambda_E) \in \Delta_S^{E+1}$，满足 $\|\lambda_e - \bar{\lambda}\|_1 \leq \delta$ 且 $\mu_{e,\lambda_e} \in \mathsf{P}(\mathcal{Y})$ 对每个 e均成立。*(i)* 平均未处理分布的尖锐集合为 $\mathcal{Q}^\delta = \{\sum_e \omega_e \mu_{e,\lambda_e} : (\bar{\lambda}, \lambda_1, \ldots, \lambda_E) \in \Lambda^\delta\}$；在有限网格上，其线性对比的界限可通过线性规划求得。*(ii)* $\mathcal{Q}^\delta$ 关于 δ非递减，在 $\delta = 0$处等于逐事件集合，并且当 $\delta \geq 2$时等于事件特定集合 $\mathcal{Q}^{\mathrm{sp}}$ 。*(iii)* $\delta^* =$ 最低$\{\delta : \Lambda^\delta \neq \varnothing\}$ 可达到，在有限网格上是一个线性规划的值，并且当且仅当每个 Λ_e 均非空时为有限值。

证明。(i) 对于一个固定的可行数组，定理 1 逐事件应用后可得一个与观测分布相容的数据生成过程，其事件-e未处理分布为μ_{e,λ_e}；这些事件涉及不同的总体，因此可以将这些过程组合起来。反过来，模型中的每个¯分布都具有这种形式。在网格上，引入 $t_{e,s} \geq |\lambda_{e,s} - \lambda_s|$，使其满足两个线性不等式 $t_{e,s} \geq \pm(\lambda_{e,s} - \bar{\lambda}_s)$，并施加 $\sum_s t_{e,s} \leq \delta$；相容性条件为 $Z_e^\top \lambda_e \geq 0$；目标函数¯是线性的。(ii) 约束集随 δ递减嵌套。在 $\delta = 0$ 处，每个 λ_e 都等于 λ。对于任意¯ ¯，Δ_S, $\|\lambda_e - \lambda\|_1 \leq \|\lambda_e\|_1 + \|\lambda\|_1 = 2$中的两个点，因此对 $\delta \geq 2$而言，该约束不起作用。(iii)_ _δ^* =是在 $\lambda \in \Delta_S$和 $\lambda_e \in \Lambda_e$上对 $_e \|\lambda_e - \lambda\|_1$取最小值、最大值，这是连续函数在紧集上的最小值；当且仅当某个 Λ_e 为空时，该集合为空；引入变量 $t_{e,s}$ 及一个上图集变量后，它就是一个线性规划。 □

在最低工资应用中，每个 Λ_e 都是空集，因此 $\delta^* = \infty$：任何程度的权重异质性都无法使这五个事件与历史迁移相协调。

S2.8 协变量

设 X 是一个预先确定的协变量，其有限支撑为 $\mathcal{X}$，不受处理影响，并记 $p_{g,\,t}(x) = \Pr(X = x \mid G = g,$ 日期 $t)$ 和 $\mu_{g,\,t}(\cdot \mid x)$ 为条件分布。条件传输指出，对每个 x，都存在 $\lambda(x) \in \Delta_S$，使得

$$\mu^0_{1,1}(\cdot \mid x) - \mu^0_{0,1}(\cdot \mid x) = \sum_s \lambda_s(x)\{\mu^0_{1,s}(\cdot \mid x) - \mu^0_{0,s}(\cdot \mid x)\}.$$

命题 *S2.11*（条件传输）。*(i)* 在权重可随 x 变化的条件传输下，政策日期处理组未处理分布的尖锐集合为闵可夫斯基和 $\sum_x p_1$*,1(x)*$\mathcal{F}_S(P \mid x)$，其中 $\mathcal{F}_S(P \mid x)$ 是根据条件分布计算得到的定理 *1* 的尖锐集合；当且仅当每个条件集均非空时，该集合才非空。若采用共同权重 $\lambda(x) = \lambda$，则尖锐集合为 $\{\sum_x p_{1,1}(x)\mu_\lambda(\cdot \mid x)$*:* $\lambda \in \bigcap_x \Lambda_S(P \mid x)\}$。*(ii)* 假设各组的协变量分布不随时间变化，$p_{g,t} = p_g$，且条件传输在共同权重 λ 下成立。那么

$$\nu_1 - \sum_s \lambda_s \nu_s = \sum_x \{p_1(x) - p_0(x)\}\Big\{\mu^0_{0,1}(\cdot \mid x) - \sum_s \lambda_s \mu^0_{0,s}(\cdot \mid x)\Big\},$$

因此，当两个组具有相同的协变量分布，或每个条件比较分布都是其自身过去状态的相同混合分布时，带权重 λ 的无条件迁移成立；否则可能不成立。

证明。(i) 处理组的未处理分布为 $\sum_x p_{1,\,1}(x)\mu^0_{1,\,1}(\cdot \mid x)$，其中 $p_{1,\,1}$ 可以识别，因为 X 不受处理影响。模型逐个单元约束条件分布，因此，将协变量单元替代事件后，命题 S2.9(i) 的论证仍然适用。(ii)

写作 $\nu_t = \sum_x \{p_1(x)\mu^0_{1,t}(\cdot \mid x) - p_0(x)\mu^0_{0,t}(\cdot \mid x)\}$，并加上再减去 $p_1(x)\mu^0_{0,t}(\cdot \mid x)$：$\nu_t = \sum_x p_1(x)\nu_t(x) + \sum_x \{p_1(x) - p_0(x)\}\mu^0_{0,t}(\cdot \mid x)$，其中 $\nu_t(x)$ 是条件偏差。条件传输给出 $\sum_x p_1(x)\nu_1(x) = \sum_s \lambda_s \sum_x p_1(x)\nu_s(x)$；减去 $\sum_s \lambda_s \nu_s$ 后即得到所述项。

□

S2.9 细化

命题 3 的证明。设 S 为有限集，设 Π_n 为嵌套的有限可测划分，其 σ-代数 $\mathcal{A}_n = \sigma(\Pi_n)$ 生成结果σ-代数 $\mathcal{B}$，并设 μ_s $(s \in S)$为单位质量的有限有符号测度。对于 $\lambda \in \Delta_S$，令 $\mu_\lambda = \sum_s \lambda_s \mu_s$，并令 $b(\mu_\lambda) = \sup_{A \in \mathcal{B}} [-\mu_\lambda(A)]_+$。定义划分亏损 $b_n(\mu_\lambda) = \sum_{A \in \Pi_n} [-\mu_\lambda(A)]_+$。每个粗单元都是细单元的并，因此 $0 \le b_n(\mu_\lambda) \le b_{n+1}(\mu_\lambda) \le b(\mu_\lambda)$。相应的可行权重集合构成递减序列，并包含完整结果变量的可行集。

令 $\rho = \sum_{s \in S} |\mu_s|$ 和 $g_s = d\mu_s / d\rho$。由于每个 μ_s 都是有限的，ρ 也是有限的。相对于 $\rho/\rho(\mathcal{Y})$的条件期望给出

$$b_n(\mu_\lambda) = \int \Big[- \sum_s \lambda_s \mathbb{E}_\rho(g_s \mid \mathcal{A}_n) \Big]_+ d\rho.$$

递增的 σ-代数生成 $\mathcal{B}$，因此由条件期望收敛可得 $\|\mathbb{E}_\rho(g_s \mid \mathcal{A}_n) - g_s\|_{L^1(\rho)} \to 0$ 对每个 s 都成立。负部映射是 1-Lipschitz，因此

$$\sup_{\lambda \in \Delta_S} |b_n(\mu_\lambda) - b(\mu_\lambda)| \le \max_{s \in S} \|\mathbb{E}_\rho(g_s \mid \mathcal{A}_n) - g_s\|_{L^1(\rho)} \longrightarrow 0.$$

对 Δ_S 取最小值，即可得到优化后的划分亏损收敛。最后，若一个固定的有符号测度在每个划分的每个单元上均非负，则它在由这些单元生成的代数上也非负。该代数在度量 $\rho(A \triangle B)$ 下对可测集稠密；通过逼近任意可测事件可知，其有符号测度值非负。因此，所有划分可行权重集合的交集等于完整的结果变量可行集。

□

S2.10 粗化

命题 *4* 的证明。(i) 设 $\mu_\lambda = P_{0,1} + \sum_s \lambda_s P_{1,s} - \sum_s \lambda_s P_{0,s}$，其中 $P_{g,t} = \mu^0_{g,t}$。对于每个事件 A，$\mu_\lambda(A) \ge -\sum_s \lambda_s P_{0,s}(A) \ge -1$，因此 $b(\mu_\lambda) \le 1$。为证明可以取到，令 $x \ne y$ 位于 Π 的一个单元中，取一个历史日期，并设 $P_{0,1} = P_{1,s} = \delta_x$, $P_{0,s} = \delta_y$。候选分布$2\delta_x - \delta_y$ 的单元质量与 δ_x 相同，因此其网格修复为零，而其负质量为一。

(ii) 固定一个密度为 g 的混合分布以及一个长度为 ℓ 的单元 A，并令 $\bar{g}$ 为 g 在 A 上的平均值。A 上的粗化损失为

$$\int_A [-g]_+ - \Big[-\int_A g\Big]_+ = \tfrac{1}{2}\Big(\int_A |g| - \Big|\int_A g\Big|\Big) \le \tfrac{1}{2}\int_A |g-\bar{g}|,$$

因为 $\int_A |g| \le \int_A |g-\bar{g}| + |\int_A g|$。函数 $h = g - \bar{g}$ 在 A 上的均值为零，且是L-Lipschitz 的，因为 L-Lipschitz 密度的混合仍是 L-Lipschitz 的。我们断言 $\int_A |h| \le L\ell^2/4$。令 h^* 为 h在长度为 ℓ的区间上的非递增重排。重排保持值的分布不变，因此 h^* 的均值为零，且 $\int |h^*| = \int_A |h|$；在一维情形下，重排不会增大 Lipschitz 常数。令 c 为 h^* 变号的点。于是，对于 $x < c$，有 $0 \le h^*(x) \le L(c-x)$；对于 $x > c$，有 $0 \ge h^*(x) \ge -L(x-c)$，因此$\int h^*_+ \le Lc^2/2$且 $\int h^*_- \le L(\ell-c)^2/2$。均值为零使两个积分相等，因此各自至多为 $L\min\{c,\ell-c\}^2/2 \le L\ell^2/8$，且 $\int |h^*| \le L\ell^2/4$。因此，单元损失至多为$L\ell^2/8$。对各单元求和并利用 $\sum_A \ell_A^2 \le h(b-a)$，可得 $b(\mu_\lambda) - b_\Pi(\mu_\lambda) \le Lh(b-a)/8$，对任意 λ均成立；对共同单纯形取最小值后，不等式仍然成立：$d_S - d_{S,\Pi} \le b(\mu_{\lambda_\Pi}) - b_\Pi(\mu_{\lambda_\Pi})$，其中 λ_Π 为网格修复的最小化解。对于常数，在 $[0,\ell]$ 上，$g(x) = L(x-\ell/2)$的单元损失为 $\int_0^{\ell/2} L(\ell/2-x)\,dx = L\ell^2/8$。

□

其结果之一是，细化网格能够一致地衡量总体缺口，而不必将其强行降为零：如果 $d_S > 0$，那么每个网格宽度为 $h < 8d_S/\{L(b-a)\}$ 的划分都有严格为正的网格缺口。第 (i) 部分说明了为什么任何此类结论都需要一定的光滑性。

S2.11 平滑结果

概率约束并非离散结果所特有。

命题 S2.12 (平滑分布可能具有持续的相容性增益)。 存在真实观测到的政策前概率律，其定义在 $[0,1]$，上；另有一个具有有界密度的政策后对照分布，以及两个迁移后的历史候选分布，它们的累积有符号测度函数均为平滑函数，并满足以下条件：*(i)* 相容历史混合集合非空；*(ii)* 对于一个有界连续的政策评分函数，相容上界与无约束上界相差 $1/2$；*(iii)* 对于网格宽度趋于零的均匀结果划分，相应的有限网格上界差收敛至 $1/2$。因此，无论是有界密度还是利普希茨累积分布函数，都不能保证细网格上的相容性增益趋于零。

证明。 设政策实施后对照组密度在 $[0,1]$ 上为 1。一种历史处理组与对照组差异为零，可由两个均匀分布得到。对于另一种差异，分别取真正的历史处理组和对照组密度 $1+\sin(2\pi x)$ 和 $1-\sin(2\pi x)$。两者均非负

且积分为 1。因此，迁移后的候选分布密度为 $g_0(x) = 1$ 和$g_1(x) = 1 + 2\sin(2\pi x)$。对第二个候选分布赋予权重 λ，得到 $g_\lambda(x) = 1 + 2\lambda\sin(2\pi x)$。由于 $\min_{x\in[0,1]}\sin(2\pi x) = -1$，当且仅当 $0 \le \lambda \le 1/2$ 时，该候选分布才是概率密度。对于 $\ell(g) = \int_0^1 \sin(2\pi x)g(x)\,dx$，直接积分可得 $\ell(g_\lambda) = \lambda$。因此，有符号凸包上界为 1，相容上界为 $1/2$。

对于等长分割 Π_K，令 λ_K 为满足每个单元质量均非负的最大权重。正弦函数 $2\pi x$ 的单元平均值可取得趋近于 -1 的值，当 $K \to \infty$ 时；因此 $\lambda_K \to 1/2$。将分割得分定义为 $\sin(2\pi x)$ 在单元内的平均值。相应的线性泛函在双候选凸包上均匀收敛于 ℓ，因为这些得分均匀收敛于 $\sin(2\pi x)$。其不受约束的上端点趋于 1，而其相容上端点趋于 $1/2$。因此，有限网格增益趋于$1/2$。所有底层观测密度均有界且光滑，所有相关累积分布函数均为利普希茨连续。

□

命题 S2.13 (精确密度振幅边界)。令 f 为相对于公共 σ-有限测度 m 在 $\mathcal{Y}$ 上的概率密度，并令 ν 为满足 $\int \nu\,dm = 0$ 的可积有符号密度。对于 $a \ge 0$，置 $g_a = f + a\nu$ 并定义

$$a_\star = \operatorname*{ess\,inf}_{\{y:\nu(y)<0\}} \frac{f(y)}{-\nu(y)},$$

其中，空集上的本质下确界定义为 $+\infty$。当且仅当 $0 \le a \le a_\star$ 时， g_a 才是概率密度。若 $\nu \not\equiv 0$，则负值集的测度为正，且 $a_\star < +\infty$。更一般地，对于历史有符号密度顶点 $g_s = f + \nu_s$， $s \in S$，全分布相容权重集恰好为

$$\Lambda_S = \left\{ \lambda \in \Delta_S : f(y) + \sum_{s\in S} \lambda_s \nu_s(y) \ge 0 \quad m\text{-almost everywhere} \right\}.$$

整个历史凸包相容，当且仅当对于每个 $s \in S$， $\nu_s \ge -f$ 几乎处处成立。

对于积分为零的可积有符号密度 ν 和基准概率密度 f，每个 $f + a\nu$ 的总质量均为 1。对于 $a \ge 0$，在 $\{\nu \ge 0\}$ 上的非负性自动成立；对于 $\{\nu < 0\}$，则要求 $a \le f/(-\nu)$ 几乎处处成立。取本质下确界即可得到精确的振幅阈值。若 $\nu \not\equiv 0$ 且积分为零，则其负值集的测度为正。比值 $f/(-\nu)$ 在该集合上几乎处处有限，因此其本质下确界有限。对于多个顶点，整个凸包非负等价于每个顶点均非负，这可由考察单纯形极端权重得出。

阈值	当负向历史偏移发生在基线密度为零的位置时，可能为零	l

它并不由最大绝对偏移决定：正向偏移不会违反相容性。振幅条件刻画的是概率律有效性，而不是某个特定政策端点是否发生移动。

S3 分位数与政策排序

以下始终假设 S 是有限集，且 $\Lambda_S(P) \neq \varnothing$。

命题 S3.1（多面体相干性下分位数端点的取到）。设相容权重集 $\Lambda_S（P）$是一个非空多面体，并令 V 为其极点构成的有限集。对于每个 $\tau \in$ （0，1），

$$\min_{\lambda \in \Lambda_S(P)} Q_{H_\lambda}(\tau) = \min_{v \in V} Q_{H_v}(\tau), \qquad \max_{\lambda \in \Lambda_S(P)} Q_{H_\lambda}(\tau) = \max_{v \in V} Q_{H_v}(\tau).$$

In 特别是，当结果变量的支持集有限时，两个端点都能取到 .

证明。每个可行的 λ 都是有限个顶点 $v \in V$的凸组合，因此H_λ是相应顶点累积分布函数的同一凸组合。令 $a =$为$_{v \in V} Q_{H_v}(\tau)$的最小值， $b =$为$_{v \in V} Q_{H_v}(\tau)$的最大值。对于任意 $y < a$，每个顶点累积分布函数都满足 $H_v(y) < \tau$，因此 $H_\lambda(y)$也满足该条件；于是 $Q_{H_\lambda}(\tau) \geq a$。在 $y = b$ 处，每个顶点累积分布函数都满足 $H_v(b) \geq \tau$，所以每个混合分布都满足 $H_\lambda(b) \geq \tau$ 和 $Q_{H_\lambda}(\tau) \leq b$。取到 a 和 b 的顶点分位数属于可行集，因此两个等式均成立。在有限结果支持集上，相容性由有限个线性单元约束定义，因此 $\Lambda_S(P)$是一个多面体。

□

该命题讨论的是单个分位数的极值，而不是由单个混合分布同时取到累积分布函数包络。它也不意味着两个分位数极值之间的每个值都可以取到。

S3.1 包络正则性

定义

$$\underline{F}(y) = \min_{\lambda \in \Lambda_S} H_\lambda(y), \qquad \overline{F}(y) = \max_{\lambda \in \Lambda_S} H_\lambda(y).$$

引理 S3.1（历史单纯形上的一致右连续性）。若 $y_n \downarrow y$，则

$$\sup_{\lambda \in \Delta_S} |H_\lambda(y_n) - H_\lambda(y)| \to 0.$$

类似的均匀分布尾部极限也成立，如 $y \to \pm\infty$。

证明。根据三角不等式,

$$|H_\lambda(y_n) - H_\lambda(y)| \le |F^0_{0,1}(y_n) - F^0_{0,1}(y)| + \sum_{s\in S} \lambda_s \{|F^0_{1,s}(y_n) - F^0_{1,s}(y)| + |F^0_{0,s}(y_n) - F^0_{0,s}(y)|\}.$$

右侧由有限个原始累积分布函数增量之和界定，而该和趋于零。尾部结论同样成立。

□

命题 S3.2（累积分布函数包络）。和 F 都是 F 累积分布函数。

证明。 非递减函数的逐点最小值和最大值仍是非递减的。右连续性和尾部极限由前述一致引理推出。 □

S3.2 广义逆

对于累积分布函数 F，定义 $Q_F(\tau) = \inf\{y: F(y) \ge \tau\}$。

引理 S3.2（广义逆等价性）。 对于 $\tau \in (0, 1)$,

$$Q_F(\tau) \le y \quad \Longleftrightarrow \quad F(y) \ge \tau.$$

证明。 若 $F(y) \ge \tau$，则定义集合包含 y，因此 $Q_F(\tau) \le y$。反之，假设 $F(y) < \tau$。由右连续性，存在 $\delta > 0$ 使得 $F(y+\delta) < \tau$。由单调性可知，对于每个 $x \le y + \delta$，都有 $F(x) < \tau$，因此定义集合包含于 $(y+\delta, \infty)$，且 $Q_F(\tau) \ge y + \delta > y$。取逆否命题即可证明正向蕴含。 □

命题 S3.3（尖锐分位数极值）。 对于每个 $\tau \in (0, 1)$,

$$\inf_{\lambda\in\Lambda_S} Q_{H_\lambda}(\tau) = Q_{\overline{F}}(\tau), \qquad \sup_{\lambda\in\Lambda_S} Q_{H_\lambda}(\tau) = Q_{\underline{F}}(\tau).$$

下确界可以取到。在没有进一步限制的情况下，不声称上确界可以取到。

证明。 令 $q_\lambda = Q_{H_\lambda}(\tau)$。由引理 S3.2,

$$\{y : \exists\lambda,\ q_\lambda \le y\} = \{y : \max_\lambda H_\lambda(y) \ge \tau\} = \{y : \overline{F}(y) \ge \tau\}.$$

左侧集合的下确界为 $\inf_\lambda q_\lambda$，右侧集合的下确界为 $Q_F(\tau)$。为证明可以取到，令 $\lambda_n \to \lambda$。S 的仿射表示和有限性蕴含 $\|H_{\lambda_n} - H_\lambda\|_\infty \to 0$。若 $x < Q_{H_\lambda}(\tau)$，则 $H_\lambda(x) < \tau$，因此 $H_{\lambda_n}(x)$ 最终 $< \tau$。于是

$$\liminf_{n\to\infty} Q_{H_{\lambda_n}}(\tau) \ge Q_{H_\lambda}(\tau),$$

且广义逆映射在 Λ_S 上下半连续。由于 Λ_S 是紧集，它可以取到最低工资。这证明了第一个等式以及所述的最低工资结论。

同样，

$$\{y : q_\lambda \le y \text{ for every } \lambda\} = \{y : \min_\lambda H_\lambda(y) \ge \tau\} = \{y : \underline{F}(y) \ge \tau\}.$$

左侧集合的下确界是 $\sup_\lambda q_\lambda$，而右侧集合的下确界是 $Q_F(\tau)$。这证明了第二个等式。下半连续性无法保证最大值点的存在，而广义逆映射在没有额外限制的情况下并非上半连续；因此，此端点表述为上确界。

□

S3.3 严格交叉下的连续性

命题 *S3.4*（分位数映射的连续性）。固定 $\tau \in (0,1)$，并令 $I = [a,b]$ 为紧集。假设 Λ_S 非空，$(\lambda, y) \mapsto H_\lambda(y)$ 在 $\Lambda_S \times I$，$H_\lambda(a) < \tau < H_\lambda(b)$ 上对每个 $\lambda \in \Lambda_S$ 都连续，且方程 $H_\lambda(y) = \tau$ 具有唯一解 $q(\lambda) \in (a,b)$。则 $Q_{H_\lambda}(\tau) = q(\lambda)$，映射 $\lambda \mapsto q(\lambda)$ 是连续的，其像是一个端点均能取到的紧区间。

证明。固定 λ。H_λ的单调性和连续性，加上唯一性，意味着对于 $y < q(\lambda)$，$H_\lambda(y) < \tau$，且 $H_\lambda(q(\lambda)) = \tau$；因此广义逆等于 $q(\lambda)$。令 $\lambda_n \to \lambda$。由于 I的紧致性，$q(\lambda_n)$的每个子序列都有一个进一步的子序列收敛到某个 $q^* \in I$。联合连续性给出 $H_\lambda(q^*) = \tau$，而唯一性迫使 $q^* = q(\lambda)$。因此$q(\lambda_n) \to q(\lambda)$。最后，$\Lambda_S$ 是紧的且为凸集。其连续像是紧的且连通的，因而是一个紧区间，并且其端点均能取到。

□

S3.4 福利规划与同时排序

令共同的有限支撑为 $y_1 < \cdots < y_K$。将当前对照组概率向量记为$p_{0,\,1}$并将历史处理组与对照组的概率差异置于 $K \times |S|$ 矩阵 D_S的各列中。对于向量 $u = (u(y_1) \ldots, u(y_K))'$，定义

$$\Lambda = \{\lambda \in \mathbb{R}^{|S|} : \lambda \ge 0,\ \mathbf{1}'\lambda = 1,\ p_{0,1} + D_S\lambda \ge 0\}.$$

如果 Λ 非空，则精确的反事实福利上限和下限满足

$$\max_{\lambda\in\Lambda}\{u'p_{0,1} + u'D_S\lambda\}, \qquad \min_{\lambda\in\Lambda}\{u'p_{0,1} + u'D_S\lambda\}.$$

由于 Λ是紧多面体，因此可以达到这些端点。若 g 是处理组后期概率向量，则福利差异的尖锐区间端点分别为 $u'g$ 减去上述最大值

和最小值。特别是，严格为正的下端点表明，对于既定模型中的每一种完整数据分布，差异都为正。

对于该支持集上的一阶随机占优，只需比较前$K-1$ 个支撑点处的累积分布函数。处理后概率律支配每个相容反事实概率律这一断言，等价于

$$G(y_k) \le \min_{\lambda \in \Lambda} \left\{ F_{0,1}(y_k) + \sum_{s \in S} \lambda_s [F_{1,s}(y_k) - F_{0,s}(y_k)] \right\}, \quad k = 1, \dots, K-1.$$

每个最低值都对应一个线性规划，但其最优解可能取决于 k。这些逐点最小值用于检验普遍排序；它们并不意味着存在一个权重向量能够实现整个下包络。

为考虑抽样误差，令 $\mathfrak{C}_{T,N}$ 覆盖整个处理后累积分布函数，并令 $\mathfrak{C}_{F,N}$包含整个尖锐反事实集合，且该联合事件 E_N的概率至少为 $1-\alpha$。投影后的福利区域

$$\mathfrak{C}_{W,N}(u) = \left\{ \int u \, dG' - \int u \, dF' : G' \in \mathfrak{C}_{T,N}, \ F' \in \mathfrak{C}_{F,N} \right\}$$

在 E_N 上包含 $\mathcal{W}_{u,S}(P)$。因此，只要该包含事件可测，它就以至少 $1-\alpha$ 的概率覆盖完整的识别区间；否则，其内概率至少为$1-\alpha$ 。同一联合事件同时验证每个预先指定的有界福利指标；无需为每个指标单独设定临界值。只有当联合累积分布函数区域支持其普遍不等式时，支配关系的结论才有依据。

S4 放宽与失效

本节证明第5节背后的有限网格项目，给出一个反例，说明失效半径中的两个障碍必须共享同一个历史权重，并陈述第9节所用的排除限制结果。在含有 J 个单元的网格上，记 Z 为顶点行 $H \times J$矩阵， $u \in \{0,1\}^J$为结果事件的指示变量，并记$d_{\mathrm{TV}}(q, \ z) = \frac{1}{2}\|q - z\|_1$。

–

$$\mathcal{Q}_\tau(Z) = \left\{ q \in \Delta_J : \ \exists \lambda \in \Delta_H, \ \tfrac{1}{2}\|q - Z^\top \lambda\|_1 \le \tau \right\}, \qquad \Theta_\tau = \{ g - u^\top q : \ q \in \mathcal{Q}_\tau(Z) \}.$$

命题 S4.1 （最低一致性与符号崩溃规划）。 定义 $\rho_* := \inf\{\tau \ge 0 : \ \mathcal{Q}_\tau(Z) \ne \varnothing\}$ 且 $\tau_{\mathrm{sign}} := \inf\{\tau \ge 0 : \ \exists q \in \mathcal{Q}_\tau(Z) \ \ g - u^\top q \ge 0\}$。那么：

1. 最小一致性松弛为

$$\rho_* = \min_{\lambda \in \Delta_H} \sum_{j=1}^{J} \left(-[Z^\top \lambda]_j\right)_+ .$$

2. 每当 $\tau \geq \rho_*$、 Θ_τ 是通过在 $q \in \mathcal{Q}_\tau$ 上最小化和最大化 $g - u^\top q$ 而得到的具有精确端点的非空紧区间（Z）时，它都嵌套于 τ 中。引入带有 $-t_j \leq q_j - [Z^\top \lambda]_j \leq t_j$ 和 $\sum_j t_j \leq 2\tau$ 的 $t_j \geq 0$ 后，该项目成为线性项目。

3. 最小 松弛（允许非负效应）由以下线性规划给出

$$\tau_{\text{sign}} = \min_{q \in \Delta_J,\, \lambda \in \Delta_H} \frac{1}{2}\|q - Z^\top \lambda\|_1 \quad \textit{subject to } u^\top q \leq g.$$

如果 Θ_0 非空且具有严格负的上端点，那么对于 $\tau < \tau_{\text{sign}}$， $\tau_{\text{sign}} > 0$ 及 Θ_τ 的每个成员都严格为负。该边界可以达到。

证明。 对于每个总质量为一的固定带符号向量 z，其负质量 $b(z) = \sum_j (-z_j)_+$ 等于 $\min_{q \in \Delta_J} \frac{1}{2}\|q - z\|_1$：将其负质量移至零，必然会使负部和正部各自至少改变 $b(z)$，而一个支持集位于正坐标上的概率向量可以达到这一数值。在紧单纯形 Δ_H 上进行最小化即可证明第一个结论。满足所显示线性约束且半径有限的三元组集合 $(q,\ \lambda,\ t)$，在将无关松弛限制为 $0 \leq t_j \leq 2\tau$ 后，是一个紧投影；由于原可行集是凸集，它在仿射映射 $q \mapsto g - u^\top q$ 下的像是一个紧区间。不等式等价于所述的 ℓ_1 界限，并给出精确极值。单调性由嵌套性推出。最后，事件 $g - u^\top q \geq 0$ 是线性约束 $u^\top q \leq g$。连续函数在 $\Delta_J \times \Delta_H$ 上的约束最小值存在；其目标函数的上图集由相同的 t_j 不等式给出。如果严格负的上端点在 $\tau = 0$ 处取得，则紧致性排除了 $\tau_{\text{sign}} = 0$。在 $\tau < \tau_{\text{sign}}$ 处，不存在具有非负效应的可行对，从而证明最后一个断言。

□

命题 S4.2（事件面距离与精确符号崩解分解）。设 $z \in \mathbb{R}^J$ 满足 $\sum_j z_j = 1$，并令 A 为 J 单元的非空真子集。记

$$b(z) = \sum_{j=1}^{J} (-z_j)_+, \qquad z^+(A) = \sum_{j \in A} (z_j)_+, \qquad K_g = \{q \in \Delta_J : q(A) \leq g\}, \quad g \in [0,1].$$

Then

$$\min_{q \in K_g} \frac{1}{2}\|q - z\|_1 = \max\left\{b(z),\ z^+(A) - g\right\}. \tag{S.6}$$

因此，对于单位质量的历史顶点 z_h ，

$$\tau_{\mathrm{sign}}(Z,g)=\min_{\lambda\in\Delta_H}\max\left\{b(z_\lambda),\ z_\lambda^+(A)-g\right\},\qquad z_\lambda=\sum_{h=1}^{H}\lambda_h z_h. \tag{S.7}$$

两种最小化均能达到其最小值。引入正、负坐标部分的上包络后，第二种最小化是一个有限线性规划。 A是真子集这一条件至关重要：当 A 是整个结果空间时，对每个概率律都有 q（A） $=1$ ，因此对于 $g<1$而言，该符号事件不可能发生。该公式针对一个预先指定的事件 A；对多个嵌套事件进行联合认证则属于另一种优化问题。

证明。 记 $p=z^+$ 为逐坐标正部，并令 $P_A=p(A)$、 $P_{A^c}=p(A^c)$；于是 $P_A+P_{A^c}=1+b(z)$。对于概率向量 q，等质量恒等式给出

$$\frac{1}{2}\|q-z\|_1=\sum_{j=1}^{J}(p_j-q_j)_+.$$

事实上，右侧是从正坐标中移除的质量，而为了得到单位质量的概率向量，同样多的质量会被添加到负坐标或质量减少的坐标中。每个可行的 q 都必须至少移除 $b(z)$，因为 p 的总质量为 $1+b(z)$。它还必须从 A中的正坐标至少移除 P_A-g ，因为 $q(A)\le g$。这就证明了 (S.6) 中的下界。

为证明取到，令 $r=\max\{b(z)\ P_A-g\}$ ，并从 p中移除 r 个质量单元，每个坐标最多移除 p_j。若 $P_A-g\le b(z)$，我们可以移除总计 $b(z)$，使 A 中的剩余质量至多为 g: 从 A 中移除 $\max\{0,P_A-g\}$ ，并将剩余需要移除的质量分配到 A 和 A^c ，且不超过它们各自的正质量。这是可行的，因为正质量总量为 $1+b(z)$，且 $g\ge 0$。剩余质量为一，因此它本身就是 K_g 中的一个元素，其距离为 $b(z)$。若 $P_A-g>b(z)$，则恰好从 A中的正坐标移除 P_A-g ；所得剩余正向量的总质量为 $1+b(z)-(P_A-g)<1$ ，且在 A中的质量恰好为 g。将缺少的 $P_A-g-b(z)$ 个单元添加到 A^c中的任意一个单元中；该集合非空。所得概率向量的事件质量为 g ，总质量为一。在两种情形下，从正坐标移除的质量均等于 r，由此可得等号。最后，对每个 z_λ 应用 (S.6)，并在紧单纯形上取最小值。分段线性保证最小值可取；引入 $(z_{\lambda j})_+$ 和 $(-z_{\lambda j})_+$ 的辅助变量，即得到所述线性规划。

□

解释。 $b(z_\lambda)$衡量生成任意概率律的代价，而 $z_\lambda^+(A)-g$衡量为使效应非负，必须从策略事件中移除多少正质量。二者的最大值而非总和，才是精确的修复成本

因为一次质量转移即可同时解决这两项障碍。对 λ 取最低值，意味着最优历史混合分布可以随提出的事件而变化，但绝不允许针对各个结果单元分别采用不同的混合分布。

命题 S4.3 (分离筛查不一定有共同见证**)。** 令 $A = \{1\}$、$g = 3/10$，并设三单元带符号的历史顶点为

$$z_1 = (4/5, 1/5, 0), \qquad z_2 = (1/10, 11/10, -1/5).$$

Both 顶点的总质量为一。对于 $t \in [0,1]$，令 $z_t = tz_1 + (1-t)z_2$。于是

$$\min_{t\in[0,1]} b(z_t) = 0, \qquad \min_{t\in[0,1]} \left\{ z_t^+(A) - g \right\} = -1/5,$$

而联合失效半径为

$$\min_{t\in[0,1]} \max\{b(z_t), z_t^+(A) - g\} = 1/9. \tag{S.8}$$

更一般地，若 $f(\lambda) = b(z_\lambda)$ 且 $h(\lambda) = z_\lambda^+(A) - g$，则

$$\min_{\lambda\in\Delta_H} \max\{f(\lambda), h(\lambda)\} = \max_{\eta\in[0,1]} \min_{\lambda\in\Delta_H} \{\eta f(\lambda) + (1-\eta)h(\lambda)\}. \tag{S.9}$$

这是一个等价的 标量极小极大表述，而非额外的识别假设 .

证明。 该混合分布为 $z_t = (1/10 + 7t/10,\ \ 11/10 - 9t/10,\ \ -1/5 + t/5)$。前两个坐标在整个 $[0,1]$，上均为正，第三个则非正。因此

$$b(z_t) = (1-t)/5, \qquad z_t^+(A) - g = 7t/10 - 1/5.$$

第一个函数随 t 增大而减小，第二个则随之增大。二者的最大值在它们相等时达到最低点，即在 $t = 4/9$ 处，其共同值为 $1/9$。然而，每个障碍都有各自有利的权重：$b(z_1) = 0$，而 $z_2^+(A) - g = -1/5$。因此，分别进行优化只能得到没有信息量的零下界，而非联合最低工资 $1/9$。

对于一般恒等式，注意到 $\max\{f(\lambda)\ h(\lambda)\}$ 等于 $\max_{\eta\in[0,1]}\{\eta f(\lambda) + (1-\eta)h(\lambda)\}$。$f$ 和 h 关于 λ均为凸且连续，因为它们是仿射坐标正部之和，而目标函数关于 η是仿射的。单纯形是紧且凸的。因此，有限维凸凹极小极大定理允许交换对 λ 的最低工资与对 η的最大值。这就证明了 (S.9)。该恒等式并不能证明可以将两个分别进行的最低工资运算与最大值交换：这一操作恰恰是该例所否定的。

□

反例所证明的内容。它检验的是一种可能的捷径，而不是实际的经验性违例。用户可以分别求解一致性筛选和事件份额筛选，并发现两者的结果都令人满意，但这些解可能采用彼此不相容的历史权重。原始的分布识别限制要求进行联合优化。经过适当优化的标量极小极大规划是一种有效的计算替代方案。

S4.1 最小修复和符号崩溃半径的稳定性

整个优化问题的界限很有用，因为随着估计单元频率的变化，使目标函数达到最低值的历史权重可能会发生不连续变化。以下稳定性结果不要求唯一性、可微性或活跃约束之间存在间隔。

令 Z，$Z' \in \mathbb{R}^{H\times J}$ 具有总质量为一的行 z_h，z'_h，并令 g，$g' \in [0,1]$。令 $u \in \{0,1\}^J$ 选择一个至少包含一个单元的事件，且其补集也至少包含一个单元。记作

$$d(z,z') = \tfrac{1}{2}\|z - z'\|_1, \qquad \varepsilon = \max_{1\le h\le H} d(z_h, z'_h), \qquad \delta = |g - g'|.$$

带符号的行和下面的概率向量的总质量均为一。令 $\rho_*(Z)$ 以及 $\tau_{\mathrm{sign}}(Z,\ g)$ 分别表示命题 S4.1中定义的最优值。

命题 S4.4 （优化敏感性值的统一扰动界）。根据上述定义，

$$|\rho_*(Z) - \rho_*(Z')| \le \varepsilon, \qquad |\tau_{\mathrm{sign}}(Z,g) - \tau_{\mathrm{sign}}(Z',g')| \le \varepsilon + \delta.$$

特别地，在任何同时抽样事件中，只要估计行和事件份额满足这些确定性界限，总体最小修复值和符号崩溃值就属于围绕代入估计解的相应闭区间。如果总体模型在松弛 τ 和 $\widehat{\tau}_{\mathrm{sign}} > \tau + \varepsilon + \delta$ 下可行，则该松弛所允许的每个总体效应都严格为负。

证明。设 $z_\lambda = \sum_h \lambda_h z_h$ 和 $z'_\lambda = \sum_h \lambda_h z'_h$。对于每个 $\lambda \in \Delta_H$，由三角不等式可得 $d(z_\lambda, z'_\lambda) \le \sum_h \lambda_h d(z_h, z'_h) \le \varepsilon$。对于任意非空子集 $K \subseteq \Delta_J$，距离函数 $D_K(z) = \min_{q\in K} d(q,z)$ 关于 d 是 1-Lipschitz 的，这是由三角不等式得到的；反向结论可通过交换 z 和 z' 得出。对 λ 取最低值，在使用相同 K 时仍保持界限 ε。取 $K = \Delta_J$ 即可证明第一个断言。

对于第二种情况，令 $K_g = \{q \in \Delta_J :\ u^\top q \le g\}$。对于每个 $q \in K_g$，都有一个 $q' \in K_{g'}$ with $d(q,q') \le |g - g'|$。如果 $g' \ge g$，则使用 $q' = q$。否则，当 $u^\top q > g'$ 时，将质量 $u^\top q - g'$ 从事件单元转移到任意一个固定的补集单元；这是可行的，因为 $u^\top q \le g$，并且得到 $d(q,q') = u^\top q - g' \le g - g'$。交换 g 和 g' 的角色进行相同构造

可证明，在 d 下， K_g 与 $K_{g'}$ 之间的豪斯多夫距离至多为 δ。因此，对于任意 z， $|D_{K_g}(z) - D_{K_{g'}}(z)| \le \delta$。将此界限与 $d(z_\lambda, z'_\lambda) \le \varepsilon$ 结合，并在共同权重单纯形上取最小值，即可证明关于 τ_{sign} 的所述界限。最后一个结论由符号崩溃半径的定义和严格不等式得出；总体可行性另行假定，因为若识别集为空，关于符号的断言便是空泛的。

□

S4.2 排除限制

假设分析者知道一类事件 $\mathcal{U} \subseteq \mathcal{B}$，对于处理组，政策对这些事件没有效应，$\mu^1_{1,\,1}(B) = \mu^0_{1,\,1}(B)$ 对于 $B \in \mathcal{U}$。由于 $\mu^1_{1,\,1} = \mu_{1,\,1}$ 已被观测到，这限制了缺失分布。

命题 S4.5 （排除限制下的尖锐集合）。 在假设 *1–2* 和排除限制下，未处理结果分布的尖锐集合为

$$\{\mu_\lambda : \ \lambda \in \Lambda_S(P),\ \mu_\lambda(B) = \mu_{1,1}(B) \ \textit{for all}\ B \in \mathcal{U}\}.$$

在由单元或单元并集生成的 $\mathcal{U}$ 有限网格上，可容许权重的集合是由 Λ_S、Π 的不等式以及附加线性等式定义的多面体，而线性对比的界限属于线性规划。如果该集合为空，则假设 *2* 与排除限制在 P 处被联合驳斥。

证明。 必要性是直接的。对于充分性，定理 1（第 S1.5节）证明中的构造将缺失的处理组未处理结果赋予凸包中的任一相容概率律，该构造独立于观测数据，并保持观测到的处理组结果不变；排除限制仅涉及这两个边际分布，因此只要 μ_λ 满足所显示的等式，它就对所构造的概率律成立。

□

在第 9 节中，我们施加单个等式 $\mu_\lambda(\{w \ge W\}) = \mu_{1,\,1}(\{w \ge W\})$，该等式表示，政策不会改变工资至少为 W 时的就业人口占比。对高于 W 的每个单元分别施加等式约束更为严格；如果分箱较细，这主要检验的是抽样噪声。

S5 推断：证明与补充

S5.1 为什么代入相容性会失效

令 $\widehat{F}_{g,\,t}$ 为组内经验累积分布函数，且 $\widehat{H}_\lambda = \sum_s \lambda_s(\widehat{F}_{0,\,1} + \widehat{F}_{1,\,s} - \widehat{F}_{0,\,s})$。

命题 S5.1 (一般经验不相容性)。 假设对每个 $s \in S$, $F_{0,s}$ 在某一点发生跳跃，而 $F_{0,1}$ 及任何 $F_{1,u}$, $u \in S$均不在该点跳跃。那么，对于任何 $\lambda \in \Delta_S$， H_λ 都不是累积分布函数。若结果变量服从连续分布，且每个组别—日期样本均非空，则该条件以概率为一成立。

证明。 某个 $\lambda_s > 0$。在 $F_{0,s}$ 的指定跳跃处，项 $-\lambda_s F_{0,s}$ 会产生向下跳跃，其他带负号的项只会使跳跃幅度更大，而没有带正号的项会在此处跳跃。因此，H_λ 在该点处递减。对于连续结果，不同观测值之间出现并列情形的概率为零。

□

S5.2 分位数效应的联合投影

令 $\mathfrak{C}_{T,N}$ 为处理组政策后累积分布函数 F_T 的置信域，并令 $\mathfrak{C}_N$ 为相容性管。定义 $\mathfrak{C}_{\Theta,N}(\tau) = \{Q_G(\tau) - Q_F(\tau) : G \in \mathfrak{C}_{T,N},\ F \in \mathfrak{C}_N\}$ 以及 $\Theta_S(\tau) = \{Q_{F_T}(\tau) - Q_{H_\lambda}(\tau) :\ \lambda \in \Lambda_S(P)\}$。记 $\mathbb{P}_*$ 为内概率。

推论 S5.1 (分位数处理效应)。 若 $\mathbb{P}_*\{F_T \in \mathfrak{C}_{T,N},\ \mathcal{F}_S(P) \subseteq \mathfrak{C}_N\} \geq 1-\alpha$，则 $\mathbb{P}_*\{\Theta_S(\tau) \subseteq \mathfrak{C}_{\Theta,N}(\tau)\} \geq 1-\alpha$。若这两个事件可测，且分别在 α_T 和 α_C 水平上校准，则该界限在 $1-\alpha_T-\alpha_C$ 下成立。

证明。 在联合事件上，每个 $\theta \in \Theta_S(\tau)$ 都等于 $Q_{F_T}(\tau) - Q_{H_\lambda}(\tau)$，且 $F_T \in \mathfrak{C}_{T,N}$ 和 $H_\lambda \in \mathfrak{C}_N$，因此 $\theta \in \mathfrak{C}_{\Theta,N}(\tau)$。内概率的单调性给出第一个结论，并集界给出第二个结论。

□

从投影得到的反事实分位数集合中减去代入式处理组分位数，并不能构成针对 $\Theta_S(\tau)$ 的置信程序；处理组分位数也需要估计。由于投影是针对整个累积分布函数进行的，因此在存在原子和平坦区间时仍然有效。

S5.3 基准检验

命题8在给定同时事件的条件下是确定性的。第 7 节采用了该事件的两种校准方法。

有限样本。 假设基准样本、处理组样本和对照样本相互独立，且每个样本中的单元均为独立同分布，并具有完整历史。$M+1 = 2|S|+2$ 基本累积分布函数为 $F_{0,1}$, $F_{g,s}$ ($g = 0,1$, $s \in S$)，以及 F_R。按照S6.5命题所述，对每个累积分布函数应用置信水平为 $\alpha/(M+1)$、带有 Massart 常数的双侧 DKW 不等式，并对 $(F_R - F_R) - (H_s - H_s)$ 应用三角不等式，可得满足 $c = r_R + \max_s(r_{0,1} + r_{1,s} + r_{0,s})$ 的事件。单元内不同日期之间的依赖关系不受限制。

自助法。 设 $\mathbb{G}_{N,s} = \sqrt{N}\{(\widehat{F}_R - \widehat{H}_s) - (F_R - H_s)\}$， $s \in S$。当样本比例有正下界时，将第 S6 节的联合函数中心极限定理分别应用

于独立样本，可得 $(\mathbb{G}_{N,\,s})_{s\in S}$ 在 $\ell^\infty(\mathbb{R})^{|S|}$ 中弱收敛至一个紧高斯过程。根据适用于 Donsker 类的自助经验过程定理，在每个样本内对完整历史进行重抽样的非参数自助法对这一极限是一致的，并且 $\max_s \|\cdot\|_\infty$ 是连续的。如果 $\max_s \|\mathbb{G}_{N,\,s}\|_\infty$ 的极限分布在其 $1-\alpha$ 分位数处连续，则自助法临界值对该事件给出渐近覆盖率 $1-\alpha$，因此检验的渐近检验大小为 α，并为下置信界 $D-c$ 提供覆盖率。当通过估计的倾向得分对比较单位重新加权时，必须在每次抽样中重新估计权重；此外，有效性还要求加权累积分布函数具有渐近线性展开；我们仅将此版本用于稳健性检验。

统计量 D 可通过以下线性规划精确计算

$$\min_{\lambda\in\Delta_S,\,t\ge 0}\; t \quad \text{subject to} \quad -t \le \widehat{F}_R(y_j) - \sum_s \lambda_s \widehat{H}_s(y_j) \le t, \quad j=1,\dots,n,$$

其中，$y_1<\cdots<y_n$ 为合并样本点。在相邻的合并点之间，所有涉及的函数均为常数；在最小点以下，它们均为零，因此，$\mathbb{R}$ 上的上确界等于合并样本点上的最大值。

S5.4 有限网格：稳定性、夹逼与证书

对于一个包含 K 个单元的固定划分，令 S 为非空的有限历史日期集合，并令 $q_s\in\mathbb{R}^K$ 表示从日期 $s\in S$ 迁移而来的单位质量有符号单元向量。定义单纯形 $\Delta_S=\{\lambda:\ \lambda_s\ge 0,\ \sum_{s\in S}\lambda_s=1\}$，$q(\lambda)=\sum_s\lambda_s q_s$，以及相容集合 $\Lambda=\{\lambda\in\Delta_S:\ q_k(\lambda)\ge 0\ \forall k\}$。对于相应的估计值 $\widehat{q}_s$，通过将 q_s 替换为 $\widehat{q}_s$ 来定义 $\widehat{\Lambda}$，并令 $\varepsilon=$ 为 $_s$ 与 $_k|\widehat{q}_s$、$_k-q_s$ 与 $_k|$ 的最大值。这里，$\widehat{a}_s$ 估计顶点 s 的预先指定线性评分系数 a_s；d_H,1 表示相对于单纯形 ℓ_1 距离的豪斯多夫距离。对于有限评分向量 $a=(a_s)_{s\in S}$，记 $\mathrm{osc}(a)=\max_s a_s-\min_s a_s$。对于确定性的 $e\ge 0$,$\widehat{\ }$，定义 $\Lambda_{\mathrm{in}}(e)=\{\lambda\in\Delta_S:\ \sum_s\lambda_s\widehat{q}_{s,k}\ge e\ \forall k\}$，以及 $\Lambda_{\mathrm{out}}(e)=\{\lambda\in\Delta_S:\ \sum_s\lambda_s\widehat{q}_{s,k}\ge -e\ \forall k\}$。同时顶点误差事件为 $\mathcal{E}_e=\{\max_{s,k}|\widehat{q}_{s,k}-q_{s,k}|\le e\}$。对于下文的概率陈述，$e$ 是预先指定的非负误差容许量，满足 $\Pr(\mathcal{E}_e)\ge 1-\alpha$,$\widehat{\ }$，其中 $\alpha\in(0,1)$ 是指定的误差概率。“内”集合和“外”集合分别指 $\Lambda_{\mathrm{in}}(e)\widehat{\ }$ 和 $\Lambda_{\mathrm{out}}(e)$。

命题 S5.2（严格有限网格可行性下的稳定性）。假设 $q_k(\lambda^\circ)\ge\kappa>0$ 对$\widehat{\ }$每个单元以及某个 $\lambda^\circ\in\Delta_S$都成立，且 $\varepsilon<\kappa$。那么 Λ 和 $\widehat{\Lambda}$ 均非空，且它们在单纯形 ℓ_1 范数下的豪斯多夫距离满足

$$d_{H,1}(\Lambda,\widehat{\Lambda})\le 2\varepsilon/\kappa.$$

对于顶点系数为 a_s 、由 $\widehat{a}_s$ 估计的线性得分，令 $\zeta = \text{最大值}_s\,|a_s - \widehat{a}_s|$。总体和估计可行集上的最低与最高得分之差至多为 $\zeta + (\varepsilon/\kappa)$(最大值$_s\,\widehat{a}_s$ − 最低值$_s\,\widehat{a}_s$)。

命题 S5.3（有限网格可行性夹逼）。 在 $\mathcal{E}_e$ 上，内集包含于 Λ，而 Λ 包含于外集。如果总体 Λ 非空，那么当外集为空时拒绝其非空性假设的概率至多为 α。反之，不拒绝并不能证明存在相容的总体模型。如果联合事件 $\mathcal{E}_e \cap \{\max_s |\widehat{a}_s - a_s| \le \zeta\}$ 的概率至少为 $1-\alpha$ ，对于预先指定的 $\zeta \ge 0$，线性得分的整个识别区间包含于

$$\left[\min_{\lambda \in \widehat{\Lambda}_{\text{out}}(e)} \sum_s \lambda_s \widehat{a}_s - \zeta, \quad \max_{\lambda \in \widehat{\Lambda}_{\text{out}}(e)} \sum_s \lambda_s \widehat{a}_s + \zeta\right]$$

其概率至少为 $1-\alpha$，前提是 Λ 非空。

命题 S5.4（数据依赖的区间划分的累积分布函数置信带）。 令 $F_{0,1}$ 为未处理组政策后对照组结果的总体累积分布函数，并令 $F_{1,s}$ 和 $F_{0,s}$ 表示两个组在历史日期 $s \in S$ 的总体累积分布函数。用帽标记定义它们的经验对应量。假设在事件 $\mathcal{D}$ 上，同时界

$$\|\widehat{F}_{0,1} - F_{0,1}\|_\infty \le \delta_{0,1}, \qquad \|\widehat{F}_{g,s} - F_{g,s}\|_\infty \le \delta_{g,s} \quad (g \in \{0,1\},\ s \in S)$$

固定，其中每个 $\delta_{g,t} \ge 0$ 均独立于实际误差指定。这里 $\|H\|_\infty = \sup_{y \in \mathbb{R}} |H(y)|$。考虑实数轴上任意一个有限划分，将其划分为端点可由同一组观测值决定的半开区间。分别用 $q_{s,k}$ 和 $\widehat{q}_{s,k}$ 表示输运后的符号测度 $F_{0,1} + F_{1,s} - F_{0,s}$在其第 k 个区间上的总体单元质量和经验单元质量。则对于每个这样的划分，以下结论均同时成立：

$$\max_k |\widehat{q}_{s,k} - q_{s,k}| \le 2(\delta_{0,1} + \delta_{1,s} + \delta_{0,s}) =: e_s, \qquad s \in S.$$

因此，只要 $\Pr(\mathcal{D}) \ge 1-\alpha$，有限网格无收益证书就适用于实际实现的、可能依赖数据的区间划分，且使用 $e = \max_{s \in S} e_s$。经证实的结论针对的是在实际实现的随机分区上评估的固定总体概率律。因此，尽管底层概率律不变，统计目标仍会随选定截点而变化。该结果并未将这些概率质量与另一预先指定的分区进行比较。具体而言，记 $\mathcal{N}_e$ 为实际分区上的可观测无收益证书，并令 $\mathcal{I}$ 表示至少有一个总体历史候选项在同一分区上具有负单元这一事件。如果 $\Pr(\mathcal{D}) \ge 1-\alpha$，那么

$$\Pr(\mathcal{N}_e \cap \mathcal{I}) \le \alpha.$$

该界限控制了针对实际分区上受限总体分布作出的错误无收益声明。它并不能确立未划分总体分布的相容性，验证历史传输，或在未提供同时评分误差保证的情况下，为经数据考察后选定的政策评分函数提供有效推断。

证明。 对于任意固定区间或由数据选定的区间（a， b]，其在累积分布函数 F 下的总体质量等于 F（b） $-F$（a）。该区间上的两个累积分布函数误差的绝对值，均不超过其均匀累积分布函数误差的两倍，且与 a 和 b 的位置无关。对于首尾两个无界区间（无穷端点处的误差为零），以及原子，这一界限同样成立，因为半开区间约定使用右连续的累积分布函数。对每个迁移后的有符号单元质量中的三项应用三角不等式。由于在单一事件 $\mathcal{D}$上，累积分布函数界限对每个实数阈值都成立，因此对于从数据中选出的每个随机端点也成立。由此，该界限同时适用于所有区间划分，无需再对单元或候选截断值应用并集界。在 $\mathcal{D}$ 上，无增益事件$\mathcal{N}_e$ 根据单元误差界排除了 $\mathcal{I}$。因此 $\mathcal{N}_e \cap \mathcal{I} \subseteq \mathcal{D}^c$，这就证明了该概率结论。利用所得的坐标误差界，也可得到有限划分的结论，见推论 S5.2 。

□

有限样本独立单元基准。如果在给定已实现的样本量 $n_{g、\,t} \geq 1$ 的条件下，每个基本经验累积分布函数都基于来自相应组—日期概率律的独立同分布观测值（且这些概率律不因以样本量为条件而改变），那么根据德沃雷茨基–基弗–沃尔福维茨不等式和并集界，有 $\Pr(\mathcal{D}) \geq 1-\alpha$ 对于

$$\delta_{g,t} = \sqrt{\frac{\log(2m/\alpha)}{2n_{g,t}}}, \qquad m = 1 + 2|S|,$$

以上使用的 m 个不同基本累积分布函数（相关索引为 (0、1) 和 (g、 s)）。完整历史单元内允许存在跨日期依赖关系；但用于每个基本经验累积分布函数的观测值仍须满足所述独立单元条件。该基准不能为州级聚类数据或复杂调查复制权重提供符合研究设计的保证。

推论 S5.2 （同时适用的有限网格无相容性增益证书）。 沿用命题*S5.3*的有限网格记号。假设每个总体候选项 q_s 的单元质量总和为一，且 $\Pr(\mathcal{E}_e) \geq 1-\alpha$。定义可观测事件

$$\mathcal{N}_e = \left\{ \min_{s \in S,\, 1 \leq k \leq K} \widehat{q}_{s,k} \geq e \right\}.$$

在 $\mathcal{E}_e \cap \mathcal{N}_e$ 上，每个 q_s 都是一个概率向量，因此 $\Lambda = \Delta_S$。特别地，对于此预先指定的划分上的任意评分向量，总体相容界限与不受约束界限均相等。对于任何包含无效历史顶点的固定总体数组，

$$\Pr(\mathcal{N}_e) \le \alpha.$$

在同时误差事件上，对于定义在同一划分上的所有政策评分，凸包相等；这一确定性的集合相等结论无需对各评分进行多重性校正。单凭这一点，并不能为数据选定的政策评分值提供同时置信界。这并不构成完整结果相干性或历史混合识别假设的证书。对于某个特定政策端点而言，这是充分而非必要条件：无效顶点可能无法达到该端点。估计的正单元格小于 e时，无法得出上述结论。对于采用数据依赖截断值的区间，在有同时 *CDF* 置信带可用时，命题 *S5.4*会对实际的随机划分进行控制。此时，其误差保证针对的是在该实际划分上错误地认证不存在增益，而非针对偏离另行预先指定的总体划分。一般抽样设计需要各自有效的置信带。

证明。 在 $\mathcal{E}_e$, $q_{s,k} \ge \widehat{q}_{s,k} - e$ 上，对每个顶点和单元均成立。因此， $\mathcal{N}_e$ 蕴含 $q_{s,k} \ge 0$ 对所有 s, k均成立。再结合总质量为 1，每个 q_s 都是概率向量。概率单纯形的凸性推出 $\Lambda = \Delta_S$，因此每个评分的相干极值与不受约束的极值相同。如果某个总体顶点包含负单元，则 $\mathcal{E}_e \cap \mathcal{N}_e$为空。因此 $\Pr(\mathcal{N}_e) \le \Pr(\mathcal{E}_e^c) \le \alpha$。</final> □

推论 S5.3 （策略端点增益的有限样本认证）。 沿用命题 *S5.3* 的记号和同时误差事件。假设总体相容集Λ非空，并且

$$\Pr\left(\mathcal{E}_e \cap \left\{\max_{s \in S} |\widehat{a}_s - a_s| \le \zeta\right\}\right) \ge 1 - \alpha.$$

在此处所示的事件上，外集非空。将总体下界和上界相容性增益定义为

$$\Delta_L = \min_{\lambda \in \Lambda} a^\top \lambda - \min_{s \in S} a_s, \qquad \Delta_U = \max_{s \in S} a_s - \max_{\lambda \in \Lambda} a^\top \lambda.$$

两个增益均为非负数。当估计的外集非空时，构造下界认证

$$B_L = \min_{\lambda \in \widehat{\Lambda}_{\mathrm{out}}(e)} \widehat{a}^\top \lambda - \min_{s \in S} \widehat{a}_s - 2\zeta,$$
$$B_U = \max_{s \in S} \widehat{a}_s - \max_{\lambda \in \widehat{\Lambda}_{\mathrm{out}}(e)} \widehat{a}^\top \lambda - 2\zeta.$$

当外集为空时，令 $B_L = B_U = -\infty$ 。当估计的内集非空时，还定义

$$U_L = \min_{\lambda \in \widehat{\Lambda}_{\text{in}}(e)} \widehat{a}^\top \lambda - \min_{s \in S} \widehat{a}_s + 2\zeta,$$
$$U_U = \max_{s \in S} \widehat{a}_s - \max_{\lambda \in \widehat{\Lambda}_{\text{in}}(e)} \widehat{a}^\top \lambda + 2\zeta.$$

当内集为空时，设置 $U_L = U_U = +\infty$ 。然后

$$\Pr\left\{ \max(0, B_L) \le \Delta_L \le U_L, \quad \max(0, B_U) \le \Delta_U \le U_U \right\} \ge 1 - \alpha.$$

因此， $B_L > 0$ 或 $B_U > 0$ 可以证明增益严格为正；在相应的零增益原假设下，任何错误声明的概率均被 α 界定。较小的有限上界证书可以限制增益的大小，但不能证明端点完全相等；非正的下界证书无法确定增益的符号。在概率边界处，内集可能仍为空，从而使上界证书无法提供有用信息。两个增益由同一事件联合覆盖；不存在针对每个端点的单独误差预算。这是关于增益幅度的置信度声明，而不是对历史混合识别假设本身的覆盖率声明。对于多个预先指定的政策评分，如果所假定的评分误差事件同时界定所有相关顶点系数，则该声明可以扩展。否则，单独的置信度声明需要进行多重性调整。证书本身不要求独立先导试验：其保证源于所述联合误差事件，而该事件必须针对实际估计量和抽样设计加以论证。

^证明。 在同时事件上，命题 S5.3 给出 $\Lambda \subseteq \Lambda_{\text{out}}(e)$。因此，

$$\min_{\lambda \in \Lambda} a^\top \lambda \ge \min_{\lambda \in \widehat{\Lambda}_{\text{out}}(e)} \widehat{a}^\top \lambda - \zeta, \qquad \min_{s \in S} a_s \le \min_{s \in S} \widehat{a}_s + \zeta.$$

相减得 $\Delta_L \ge B_L$。类似地，

$$\max_{s \in S} a_s \ge \max_{s \in S} \widehat{a}_s - \zeta, \qquad \max_{\lambda \in \Lambda} a^\top \lambda \le \max_{\lambda \in \widehat{\Lambda}_{\text{out}}(e)} \widehat{a}^\top \lambda + \zeta,$$

^由此可得 $\Delta_U \ge B_U$。若内集非空，命题 S5.3 给出 $\Lambda_{\text{in}}(e) \subseteq \Lambda$。因此

$$\min_{\lambda \in \Lambda} a^\top \lambda \le \min_{\lambda \in \widehat{\Lambda}_{\text{in}}(e)} \widehat{a}^\top \lambda + \zeta,$$
$$\max_{\lambda \in \Lambda} a^\top \lambda \ge \max_{\lambda \in \widehat{\Lambda}_{\text{in}}(e)} \widehat{a}^\top \lambda - \zeta.$$

将这些不等式与 $\min_s a_s \geq \min_s \widehat{a}_s - \zeta$ 和 $\max_s a_s \leq \max_s \widehat{a}_s + \zeta$ 结合，可得 $\Delta_L \leq U_L$ 和 $\Delta_U \leq U_U$。当内集为空时，上界是空泛的；当外集为空时，下界是空泛的。最后，$\Delta_L, \Delta_U \geq 0$ 是确定性的。所有不等式在所假设的事件上同时成立，从而得到所声称的覆盖率和错误声明控制。

□

命题 S5.2 的证明。设 $q(\lambda) = \sum_s \lambda_s q_s$ 和 $\widehat{q}(\lambda) = \sum_s \lambda_s \widehat{q}_s$。顶点误差 ε 意味着，对于每个 $\lambda \in \Delta_S$，都有 $\|\widehat{q}(\lambda) - q(\lambda)\|_\infty \leq \varepsilon$。首先，对于任意 $\lambda \in \Lambda$，令 $t = \varepsilon/(\kappa + \varepsilon)$ 和 $\lambda' = (1-t)\lambda + t\lambda^\circ$。于是 $q_k(\lambda) \geq -\varepsilon$ 和 $q_k(\lambda^\circ) \geq \kappa$，从而 $q_k(\lambda') \geq -(1-t)\varepsilon + t\kappa = 0$。因此 $\lambda' \in \Lambda$ 和 $\|\lambda' - \lambda\|_1 \leq 2t \leq 2\varepsilon/\kappa$。反过来，对于任意 $\lambda \in \Lambda$，令 $u = \varepsilon/\kappa < 1$ 和 $\lambda' = (1-u)\lambda + u\lambda^\circ$。估计质量满足 $\widehat{q}_k(\lambda) \geq -\varepsilon$ 和 $\widehat{q}_k(\lambda^\circ) \geq \kappa - \varepsilon$，因此 $\widehat{q}_k(\lambda') \geq -(1-u)\varepsilon + u(\kappa - \varepsilon) = 0$。这也证明了 $\Lambda \neq \varnothing$ 并给出 $\|\lambda' - \lambda\|_1 \leq 2\varepsilon/\kappa$。这两个界限共同确立了所述豪斯多夫不等式。

对于向量 $a = (a_s)_s$ 和概率权重 λ，λ'，$|a^\top(\lambda - \lambda')| \leq \frac{1}{2}\operatorname{osc}(a)\|\lambda - \lambda'\|_1$。因此，对于任意总体可行权重，其邻近的估计可行权重会使估计目标函数的变化至多为 $\varepsilon \operatorname{osc}(\widehat{a})/\kappa$。反向的可行权重近似也给出相同的界限，只是方向相反；并且 $|a^\top\lambda - \widehat{a}^\top\lambda| \leq \zeta$ 对每个单纯形权重都成立。比较下确界和上确界即可证明所述极值不等式。$\varepsilon = 0$ 的情形显然成立。

命题 S5.3 的证明。在同时事件上，每个 $\lambda \in \Delta_S$ 都满足 $|q_k(\lambda) - \widehat{q}_k(\lambda)| \leq e$，对所有 k 均成立。若对每个 k 都有 $\widehat{q}_k(\lambda) \geq e$，则 $q_k(\lambda) \geq 0$，从而证明内包含。若对每个 k 都有 $q_k(\lambda) \geq 0$，则 $\widehat{q}_k(\lambda) \geq -e$，从而证明外包含。因此，在总体相干性原假设下，外集为空只能发生在同时事件之外，其拒绝概率至多为指定的事件失效概率。在与福利系数事件的交集上，每个可行的 λ 都满足 $|a(\lambda) - \widehat{a}(\lambda)| \leq \zeta$。在包含该外集的集合上对后一个不等式取最小值和最大值，即得到所报告的覆盖整个识别值域的区间。这是关于同时事件的无条件概率陈述；以数据选出的 PASS 决策为条件进行推断并无依据。

未经缓冲的端点增益筛查不是检验。 考虑两个结果单元和两个单位质量历史候选。总体向量为 $p_1 = (0,1)$ 和 $p_2 = (1,0)$。对于等于第一个单元概率的得分，不受限制总体的下端点和相容总体的下端点都等于零：使其最小化的历史候选分布 p_1 本身就是一个概率向量。

为估计这些向量，让每个试点单位贡献联合得分 (Z_1, Z_2)，其中 $Z_2 =$ 确定性地取值为 $(1, 0)$，而 Z_1 以相同概率取值为 $(-1, 2)$ 和 $(1, 0)$，且各单元之间相互独立，

。这些有界分数对 p_1， p_2 是无偏的，并且每次实现的得分之和都为1。对于奇数样本量 n 的试点样本，将其均值记为 $\widetilde{p}_1 = (X_n,\ 1 - X_n)$。这里， X_n 是 n 个独立拉ademacher变量的均值，因此$\Pr(X_n < 0) = 1/2$ 且$\Pr(X_n = 0) = 0$。

当 $X_n < 0$ 时，不受约束的插件下端点为 X_n，且仅在第一个历史顶点处取得。第一个顶点不相容，因为其估计的首单元质量为负。尽管如此，插件相容混合分布并非空集：沿着连接 $\widetilde{p}_1$ 和 p_2 的线段，首单元概率可以等于零。其相容下端点为零。因此，未经缓冲的程序错误地报告了严格增益。当 $X_n > 0$ 时，两个顶点都相容，插件下端点没有相容性增益。由此可知

$$\text{对于每个奇数 } n\text{.，即使总体增益为零，}\Pr\{\text{仍会判定插件存在严格增益}\} = \frac{1}{2}$$

试点估计是一致的，但在这个边界处，严格增益决策并非始终可靠。要进行受控推断，需要使用同时坐标界和外部可行域，而非未经缓冲的端点面计算。这个例子涉及抽样诊断工具，而非通过经济学方法估计的处理效应。

S6 经验过程基础

S6.1 组经验分布函数的比率展开

设 $W_i = (G_i,\ \{Y_{it}\}_{t\in\mathcal{T}})$ 为独立同分布，并假设 $\pi_g = P(G_i = g) > 0$。定义

$$\widehat{F}_{g,t}(y) = \frac{\mathbb{P}_N 1\{G = g, Y_t \le y\}}{\mathbb{P}_N 1\{G = g\}}.$$

在事件 $\mathbb{P}_N 1\{G = g\} = 0$ 上，任意定义该比率；该事件的概率为 $(1 - \pi_g)^N$，因此不会影响所述随机展开。

引理 S6.1 （一致影响展开）。对于固定的有限日期集合，

$$\sqrt{N}(\widehat{F}_{g,t} - F_{g,t}) = \mathbb{G}_N \psi_{g,t} + o_p(1)$$

在 y 中均匀分布，其中

$$\psi_{g,t,y}(W) = \frac{1\{G = g\}}{\pi_g}\{1\{Y_t \le y\} - F_{g,t}(y)\}.$$

证明。 设 $A_y = 1\{G = g, Y_t \le y\}$ 和 $B = 1\{G = g\}$。由于 $PA_y = \pi_g F_{g,\,t}(y)$，

$$\widehat{F}_{g,t}(y) - F_{g,t}(y) = \frac{(\mathbb{P}_N - P)(A_y - F_{g,t}(y)B)}{\mathbb{P}_N B}.$$

类 $\{A_y - F_{g,t}(y)B : \ y \in \mathbb{R}\}$ 是有界的 VC 型。因此，其经验过程在上确界范数下为 $O_p(1)$。在 $E_N = \{\mathbb{P}_N B \ge \pi_g/2\}$ 上，其概率趋于 1，此时替换随机分母产生的余项恰好为

$$R_N(y) = \mathbb{G}_N\{A_y - F_{g,t}(y)B\}\left\{\frac{1}{\mathbb{P}_N B} - \frac{1}{\pi_g}\right\},$$

因此

$$\sup_y |R_N(y)| \le \sup_y |\mathbb{G}_N\{A_y - F_{g,t}(y)B\}| \frac{2|\mathbb{P}_N B - \pi_g|}{\pi_g^2} = O_p(1)O_p(N^{-1/2}) = o_p(1).$$

由于日期集合是有限的，同一界限对每个进入的 $(g,\ t)$ 坐标均成立。

□

S6.2 联合 Donsker 极限

命题 S6.1 (联合顶点函数中心极限定理)。 对于固定的有限 S，向量过程

$$\left\{\sqrt{N}(\widehat{H}_s - H_s) : s \in S\right\}$$

在乘积上确界范数空间的一个可分版本中弱收敛于由顶点影响函数生成的中心高斯过程

$$\psi_{s,y} = \psi_{0,1,y} + \psi_{1,s,y} - \psi_{0,s,y}.$$

特别是，其协方差核为

$$\mathrm{Cov}\{\mathbb{Z}_s(y), \mathbb{Z}_u(z)\} = P[\psi_{s,y}\psi_{u,z}], \qquad s, u \in S, \quad y, z \in \mathbb{R}.$$

证明。 相关的函数类是有界半直线指示函数类的有限并及其有限线性组合。因此，经过标准的有理数阈值可分修改后，该类逐点可测，具有平方可积包络，并且是 P-Donsker 类。应用前述联合扩张以及形成顶点组合的连续线性映射。经验过程协方差公式给出所示核，其中包括所有单位内跨日期协方差。

□

本命题中的观测值是完整单位历史 W_i；因此，必须使用一个乘子 ξ_i 乘以单位 i的所有日期坐标。对于独立的重复横截面，不存在共同的单位历史，因此影响过程改由各自具有抽样比例的单独组别—日期样本构成。确定性的单纯形约化仍然有效，但命题 S6.1以及单位层面的乘子实现必须替换为相应的设计特定联合第一阶段极限。

S6.3 乘子近似

假设 S6.1 (乘子矩)。 乘子独立同分布，且独立于样本，均值为零、方差为一，并且 $E|\xi|^{2+\delta} < \infty$ 对于某个 $\delta > 0$成立。

在 $\{\widehat{\pi}_g > 0\}$ 上定义

$$\widehat{\psi}_{g,t,y}(W_i) = \frac{1\{G_i = g\}}{\widehat{\pi}_g}\{1\{Y_{it} \le y\} - \widehat{F}_{g,t}(y)\}, \qquad \widehat{\psi}_{s,y} = \widehat{\psi}_{0,1,y} + \widehat{\psi}_{1,s,y} - \widehat{\psi}_{0,s,y}.$$

那么，$\mathbb{P}_N\psi_{g,\ t,\ y} = 0$ 恒成立。在进入组为空这一概率趋于零的事件上，将乘子临界值设为 $+\infty$。

命题 S6.2 (固定数据生成过程下的乘子有效性)。 在基线独立同分布单位实验、固定有限 S、组重叠以及假设 *S6.1*条件下，定义

$$\widehat{Z}^*_{N,s}(y) = N^{-1/2}\sum_{i=1}^{N}\xi_i\widehat{\psi}_{s,y}(W_i), \qquad s \in S.$$

设 $\mathbb{Z}$ 为有限乘积上确界范数空间的可分版本 $\mathbb{B}$ 上经验顶点过程的紧高斯极限，且令 d_{BL} 为 $\mathbb{B}$上概率分布的有界利普希茨度量。那么

$$d_{BL}\Big\{\mathcal{L}_\xi(\widehat{Z}^*_N \mid W_1, \ldots, W_N), \mathcal{L}(\mathbb{Z})\Big\} \to_p 0.$$

因此，对于 $T^*_N =$ 最大值$_{s\in S}\,\|Z^*_{N,\ s}\|_\infty$ 以及 $T =$ 最大值$_{s\in S}\,\|\mathbb{Z}_s\|_\infty$，$T^*_N$ 的条件累积分布函数在每个连续点处依概率收敛于 T 的累积分布函数。

证明。 令 Ψ 为所有进入的 (g,t) 对所对应的类 $\{\psi_{g,t,y} :\ y \in \mathbb{R}\}$ 的有限并。组重叠保证存在一个固定的有界包络。半直线指标函数属于 VC 类，乘以有限值组别指示变量以及有限线性组合均保持 VC 型；因此 Ψ 是 P-Donsker 且 P-预高斯。由乘子中心极限定理和 $E|\xi|^{2+\delta} < \infty$，可得其在外概率意义下条件弱收敛至与以 为索引的经验过程相同的紧布朗桥。

$$\left\{N^{-1/2}\sum_{i=1}^{N}\xi_i[\psi_{g,t,y}(W_i) - \mathbb{P}_N\psi_{g,t,y}] : (g,t), y\right\}$$

收敛至与以 Ψ 为索引的经验过程相同的紧布朗桥。

剩下需要论证的是，用估计的影响函数替换真实影响函数。我们直接处理这一点，避免诉诸针对样本依赖类的极大不等式。令 $I_{ig} = 1\{G_i = g\}$。在 $\{\widehat{\pi}_g > 0\}$ 上，直接代数运算可得

$$\widehat{\psi}_{g,t,y}(W_i) - \psi_{g,t,y}(W_i) = (\widehat{\pi}_g^{-1} - \pi_g^{-1}) I_{ig}\{1\{Y_{it} \le y\} - F_{g,t}(y)\} - \widehat{\pi}_g^{-1} I_{ig}\{\widehat{F}_{g,t}(y) - F_{g,t}(y)\}.$$

第一个乘子余项关于 y均匀有界于

$$|\widehat{\pi}_g^{-1} - \pi_g^{-1}| \sup_y \left| N^{-1/2} \sum_{i=1}^{N} \xi_i I_{ig}\{1\{Y_{it} \le y\} - F_{g,t}(y)\} \right|.$$

系数为 $O_p(N^{-1/2})$。由针对固定有界半直线 VC 类的乘子极大不等式可知，该上确界在条件外概率下为 $O_p^*(1)$。第二个乘子余项有界于

$$\widehat{\pi}_g^{-1} \|\widehat{F}_{g,t} - F_{g,t}\|_\infty \left| N^{-1/2} \sum_{i=1}^{N} \xi_i I_{ig} \right|.$$

这三个因子依次为 $O_p(1)$、$O_p(N^{-1/2})$ 和 $O_p^*(1)$；中间的收敛速率来自同一半直线类的Donsker性质（等价地，可由标准经验过程极大界限推出）。由于进入的 (g,t) 块数固定，两个结论联合成立。因此

$$\max_{g,t} \sup_y \left| N^{-1/2} \sum_{i=1}^{N} \xi_i [\widehat{\psi}_{g,t,y}(W_i) - \psi_{g,t,y}(W_i)] \right| = o_p^*(1).$$

这里的 $o_p^*(1)$ 表示条件外概率收敛：对任意 $\eta > 0$，条件乘子概率超过 η 的外概率大于 η 的概率趋于零。

最后，$\mathbb{P}_N \widehat{\psi}_{g,t,y} = 0$ 恒等成立，而 $\sup_y |\mathbb{P}_N \psi_{g,t,y}| = O_p(N^{-1/2})$ 和 $N^{-1/2} \sum_i \xi_i = O_p^*(1)$。因此，对不可行乘子过程进行经验中心化会使其变化 $o_p^*(1)$。所以，可行过程与中心化后的不可行过程在乘积上确界范数下渐近等价。构造有限个顶点组合是一个连续线性映射，因此由条件Slutsky定理可得所断言的联合极限。

□

映射 $z \mapsto \max_{s \in S} \|z_s\|_\infty$ 在 $\mathbb{B}$上是利普希茨的，因此最后的断言也可直接由条件连续映射定理推出。如果极限上确界统计量具有

在其（$1-\alpha$）分位数处的连续分布，则条件自助法分位数使同时顶点事件具有渐近概率 $1-\alpha$。随后，主要的确定性定理给出全体集合覆盖，其下极限至少为 $1-\alpha$，但不一定相等：校准还控制了不相容顶点，而这些顶点并不索引尖锐集合。在缺乏连续性的情况下，稳妥的表述仍然是基于适当上分位数的下极限覆盖结果。命题S6.2既不是关于 P 均匀成立的结果，也不能据此将调查复制权重视为乘子抽样。

S6.4 相容性管的闭性

令 $\ell^\infty(\mathbb{R})$ 赋予上确界范数，并令 $\mathbb{D}$ 表示其由累积分布函数构成的子集。对于有限的 S、有界函数 $h=(h_s:\ s\in S)$ 以及 $\varepsilon\geq 0$，令 $h_\lambda=\sum_s \lambda_s h_s$ 并且

$$\mathcal{C}(h,\varepsilon)=\left\{F\in\mathbb{D}:\ \inf_{\lambda\in\Delta_S}\|F-h_\lambda\|_\infty\leq\varepsilon\right\}.$$

命题 S6.3（实现的相容管的闭性）。对于每个实现的（h、ε），关于 Δ_S 的下确界可以取得，$\mathcal{C}$（h、ε）在上确界范数下是闭的，并且

$$\mathcal{C}(h,\varepsilon)=\bigcup_{\lambda\in\Delta_S}\{F\in\mathbb{D}:\|F-h_\lambda\|_\infty\leq\varepsilon\}.$$

这一确定性陈述本身并不能断言，在代入估计顶点和随机半径后具有 *Effros* 可测性。

证明。映射 $\lambda\mapsto h_\lambda$ 是连续的，因为

$$\|h_\lambda-h_{\lambda'}\|_\infty\leq\sum_{s\in S}|\lambda_s-\lambda'_s|\|h_s\|_\infty.$$

对于固定的 F，紧单纯形上的连续性保证存在一个最小值点，这也证明了并集恒等式。现在令 $F_n\in\mathcal{C}(h,\ \varepsilon)$ 一致收敛于 F，并为每个 n 选取一个使其达到最小值的 λ_n。紧致性给出一个子序列 $\lambda_{n_k}\to\lambda$。由于 $\|F_{n_k}-h_{\lambda_{n_k}}\|_\infty\leq\varepsilon$，$\lambda\mapsto h_\lambda$ 的连续性以及 F_{n_k} 的一致收敛性意味着

$$\|F-h_\lambda\|_\infty\leq\|F-F_{n_k}\|_\infty+\|F_{n_k}-h_{\lambda_{n_k}}\|_\infty+\|h_{\lambda_{n_k}}-h_\lambda\|_\infty\leq\varepsilon+o(1).$$

令 $k\to\infty$，得到 $\|F-h_\lambda\|_\infty\leq\varepsilon$。均匀极限保留单调性和右连续性。它们也保留累积分布函数尾部：用一个 F_n一致地逼近 F，利用该累积分布函数的两个尾部，然后令逼近误差趋于零。因此，$\mathbb{D}$ 在上确界范数下是闭的，所以 $F\in\mathbb{D}$，进而 $F\in\mathcal{C}(h,\ \varepsilon)$。

□

命题 S6.4 (可测的全体集合包含**)**。 设 K 是 $\mathbb{D}$ 的固定非空紧$\widehat{}$子集。假设随机顶点 H_s 具有有界右连续样本路径，$\widehat{}$且对每个有理数 q， $H_s(q)$ 可测；再设 ε 是一个可测的、取值于 $[0,+\infty]$ 的半径。$\widehat{}$ $\widehat{}$那么 $\{K \subseteq \mathcal{C}(H, \varepsilon)\}$ 是可测的。只要固定的 H_s 有界且右连续，顶点事件 $\{$最大值$_s \|H_s - H_s\|_\infty \le \varepsilon\}$ 就是可测的。如果 K 为空，那么包含事件就是整个样本空间。

证明。 对于任意有界右连续的 f 和 g，它们的差也是右连续的，因此$\|f-g\|_\infty = \sup_{q\in\mathbb{Q}} |f(q) - g(q)|$。选取紧度量$\widehat{}$空间 K 的一个可数稠密子集 D，以及 Δ_S的一个可数稠密子集 L。对于固定的 $F \in K$，映射 $\lambda \mapsto \|F - H_\lambda\|_\infty$是连续的，因此

$$d(F, \widehat{\mathcal{H}}) = \inf_{\lambda \in L} \sup_{q \in \mathbb{Q}} \left| F(q) - \sum_{s \in S} \lambda_s \widehat{H}_s(q) \right|$$

是可测的。到实现的凸包的距离作为 F的函数是 1-Lipschitz。因此

$$\{K \subseteq \mathcal{C}(\widehat{H}, \varepsilon)\} = \left\{ \sup_{F \in D} d(F, \widehat{\mathcal{H}}) \le \varepsilon \right\},$$

是一个可测事件。同样的有理阈值恒等式以及 S 的有限性，确立了顶点事件的可测性。无需对环境非可分空间中的集值映射 $\mathcal{C}$ 作任何可测性断言。

□

精确识别集是紧的，并在 K 定理 6中提供。

S6.5 有限样本 DKW 边界情形

DKW 不等式用于控制经验累积分布函数（累积分布函数）。该结果适用于具有完整观测历史的独立同分布观测单位；不涵盖调查重复权重抽样或依赖状态抽样。令 $J = |S|$ 并令

$$\mathcal{I} = \{(0,1)\} \cup \{(1,s), (0,s) : s \in S\}, \qquad M = |\mathcal{I}| = 1 + 2J.$$

对于 $N_g = \sum_{i=1}^{N} 1\{G_i = g\} > 0$，定义

$$r_{g,t} = \left\{ \frac{\log(2M/\alpha)}{2N_g} \right\}^{1/2}, \qquad (g,t) \in \mathcal{I}.$$

如果进入组计数为零，则任意定义其经验累积分布函数，并为每个相应的原始量设置 $r_{g,t} = +\infty$。

命题 S6.5（可操作有限样本 DKW 半径）。 在基线独立同分布单位抽样实验下，

$$P\Big\{\|\widehat{F}_{g,t}-F_{g,t}\|_\infty \le r_{g,t}\ \textit{for every}\ (g,t)\in\mathcal{I}\Big\}\ge 1-\alpha.$$

因此，满足

$$\varepsilon_N^{DKW}=\max_{s\in S}\{r_{0,1}+r_{1,s}+r_{0,s}\},$$

$$P\left\{\max_{s\in S}\|\widehat{H}_s-H_s\|_\infty\le\varepsilon_N^{DKW}\right\}\ge 1-\alpha.$$

证明。 以完整的组别标签向量为条件。若两组的计数均为正，则构成每个固定的 (g,t) 经验累积分布函数的观测值，均为来自 $F_{g,t}$ 的独立同分布抽样。双侧 DKW 不等式给出

$$P\Big\{\|\widehat{F}_{g,t}-F_{g,t}\|_\infty > r_{g,t}\mid G_1,\dots,G_N\Big\}\le 2e^{-2N_g r_{g,t}^2}=\frac{\alpha}{M}.$$

对组计数为正的基本累积分布函数应用并集界，可证明条件同时覆盖率至少为 $1-\alpha$，因为这类坐标至多有 M 个。属于空组的坐标不构成限制：其半径为无穷大，因此其个体覆盖事件必然成立。其他坐标仍受其 DKW 界限约束。此论证不依赖同一观测单位内各日期之间的独立性。对组别标签取条件界限的平均值，即可证明第一个式子无条件成立。最后，对每个 $s\in S$，由三角不等式可得

$$\|\widehat{H}_s-H_s\|_\infty\le\|\widehat{F}_{0,1}-F_{0,1}\|_\infty+\|\widehat{F}_{1,s}-F_{1,s}\|_\infty+\|\widehat{F}_{0,s}-F_{0,s}\|_\infty.$$

对 s 取最大值，即可证明第二个结论。 □

S6.6 到累积分布函数类的距离

设 $h:\ \mathbb{R}\to\mathbb{R}$ 为右连续函数，且 $h(-\infty)=0$， $h(+\infty)=1$。设 $\mathbb{D}$ 为累积分布函数类。

定理 S6.1（到累积分布函数类的精确上确界范数距离）。 定义

$$L=\sup_y[-h(y)]_+,\quad U=\sup_y[h(y)-1]_+,\quad V=\sup_{x<y}[h(x)-h(y)]_+.$$

Then

$$\inf_{F\in\mathbb{D}}\|F-h\|_\infty=\max\{L,U,V/2\}.$$

证明。令 δ 表示右端项。半径为 r 内的任何 CDF近似都要求$r \geq L$ 和 $r \geq U$。此外，对于 $x < y$,

$$h(x) - h(y) \leq |h(x) - F(x)| + F(x) - F(y) + |F(y) - h(y)| \leq 2r,$$

因此 $r \geq V/2$。这给出了下界。

若 $\delta = +\infty$，下界意味着每个累积分布函数与h的上确界范数距离均为无穷大，因此所述等式成立。若 $\delta = 0$，则 h 本身已经是一个累积分布函数。此后假设$0 < \delta < +\infty$ 并定义

$$G(y) = \max\left\{0, \sup_{x \leq y}[h(x) - \delta]\right\}.$$

L，$U \leq \delta$ 所蕴含的值域界限保证了 $0 \leq G \leq 1$。该函数是非递减的。对于每个 $x \leq y$，根据 V的定义，有 $h(x) - \delta \leq h(y) + \delta$，因此

$$h(y) - \delta \leq G(y) \leq h(y) + \delta.$$

为验证右连续性，取 $y_n \downarrow y$。由单调性可知，极限 $G(y_n) \downarrow \ell \geq G(y)$存在。对于每个 $\epsilon > 0$，h的右连续性给出 $\eta > 0$，使得当 $y < x \leq y + \eta$时，$h(x) \leq h(y) + \epsilon$。因此，对于所有足够大的 n,

$$\sup_{x \leq y_n}\{h(x) - \delta\} \leq \max\left\{\sup_{x \leq y}\{h(x) - \delta\}, h(y) - \delta + \epsilon\right\},$$

其正部至多为 $G(y) + \epsilon$。因此 $\ell = G(y)$。为证明下尾部，固定 $\epsilon > 0$并选取 M_-，使得对每个 $x \leq M_-$都有 $|h(x)| < \epsilon$。对于 $y \leq M_-$，$0 \leq G(y) \leq [\epsilon - \delta]_+ \leq \epsilon$，因此当 $y \to -\infty$时，$G(y) \to 0$。由于 $h(y) \to 1$，选取 M，使得对所有 $y \geq M$都有 $h(y) \geq 1 - \delta$。定义

$$F(y) = G(y) \quad (y < M), \qquad F(y) = 1 \quad (y \geq M).$$

函数 F 非递减、右连续，且极限分别为零和一。对于 $y < M$，其误差可通过 G 的不等式界定为 δ；对于 $y \geq M$，两个界限 $h(y) \geq 1 - \delta$ 和 $U \leq \delta$ 蕴含 $|1 - h(y)| \leq \delta$。因此，F 是一个累积分布函数，且 $\|F - h\|_\infty \leq \delta$。 □

前述公式还给出了精确的有限表示，前提是保留尾部区间，而不是悄然将其舍弃。

命题 S6.6 (精确合并跳跃约化)。 假设 h 右连续，且在由不同的有限断点 $x_1 < \cdots < x_K$ 所确定的区间上为常数，其中 $h(y) = 0$对于 $y < x_1$，而 $h(y) = 1$ 对于 $y \geq x_K$。记 $h_j = h(x_j)$ 为 $j = 1$、.....、K，并令 $h_0 = 0$ 为

左尾值。那么

$$L = \max_{1\le j\le K}[-h_j]_+, \quad U = \max_{1\le j\le K}[h_j - 1]_+, \quad V = \max_{0\le i<j\le K}[h_i - h_j]_+.$$

索引 $i = 0$ 至关重要：它保留了零左尾与第一个或后续有限断点处负值之间的比较。该公式也适用于 $K = 1$，此时唯一的一对是 $(0,\ 1)$ 以及 $h_1 = 1$。此外，

$$\inf_{F\in\mathbb{D}} \|F - h\|_\infty = \min_{0\le z_1\le\cdots\le z_K=1} \max_{1\le j\le K} |z_j - h_j|,$$

且最低工资可以达到。

证明。 对于值域项，两端尾部只贡献零和一，因此每个违反情形都在某个有限断点处达到。对于向下增量项，将索引 0 赋给左尾区间，将索引 $1,\ldots,K$ 赋给从有限断点开始的各区间。如果 $x < y$ 位于不同区间，则其值分别为 h_i 和 h_j，其中某个 $0 \le i < j \le K$；如果它们位于同一区间，则差值为零。反过来，对于每个 $i < j$，都可以在相应区间中选取具有所述值的点。这证明了全部三个有限最大值公式，包括左尾比较。

对于任意可行向量 z，取一个右连续阶梯函数：在 x_1 以下等于零，在 $[x_j, x_{j+1})$ 上等于 z_j，并在 $[x_K, \infty)$ 上等于一。该函数是一个累积分布函数，其一致误差就是所示的最大值。反过来，任取一个累积分布函数，记录其在 x_1、……、x_{K-1} 处的取值，并令$z_K = 1$；所得向量可行。由于 $h = 1$ 在 $[x_K, \infty)$ 上，其最大误差不超过该累积分布函数的一致误差。取下确界即可证明二者相等。可行集是紧集，且目标函数连续，因此最低值可以取到。

□

对于经验累积分布函数，断点是合并样本值中的不同取值，而仿射组合会保留共同的尾部取值零和一。新增的索引 0 记录的是左尾值，而非人为添加的数据点。因此，该命题给出了精确的化简，无需进行结果离散化，也无需指定单侧取值。

S7 样本拆分校准：有限网格

该程序使用预先指定的结果网格，以及支持集已知且联合顶点得分有界的独立先导样本和估计样本。其保证仅适用于该抽样实验。

固定整数 $J \ge 1$ 和 $K \ge 2$。在一个 K 单元划分上，设 $p_j \in \mathbb{R}^K$ 为历史有符号概率向量，其中 $\sum_{k=1}^{K} p_{j,\ k} = 1$。记 $\Delta_J = \{\lambda \in \mathbb{R}^J :\ \lambda_j \ge 0,\quad \sum_j \lambda_j = 1\}$ 表示

历史权重的单纯形。其相容权重集为

$$\Lambda = \{\lambda \in \Delta_J : p_{\lambda,k} := \sum_{j=1}^{J} \lambda_j p_{j,k} \geq 0 \text{ for all } k\}.$$

对于下文的覆盖定理，我们将假设 $\Lambda \neq \varnothing$；后续的设定检验也会考虑 $\Lambda = \varnothing$。每个观测值都是一个向量 $Z = (Z_{j,\,k})_{j\leq J,\,k\leq K}$ 满足$EZ_{j,\,k} = p_{j,\,k}$且 $Z \in \mathcal{Z}$ 几乎必然成立，其中 $\mathcal{Z}$ 是 $\mathbb{R}^{JK}$的一个已知非空紧子集。我们使用来自该概率律、样本量分别为 n_1，$n_2 \geq 1$的独立先导试验和估计样本，这些样本均为独立同分布。分别用 $\widetilde{p}_j$ 和 $\widehat{p}_j$ 表示相应的逐坐标样本均值。定义已知的坐标值域

$$D_{j,k} = \sup_{z\in\mathcal{Z}} z_{j,k} - \inf_{z\in\mathcal{Z}} z_{j,k}, \qquad D = \max_{j,k} D_{j,k},$$

并且对于 $0 < \beta < 1$，令

$$\eta = D\sqrt{\frac{\log(2JK/\beta)}{2n_1}}, \qquad \widetilde{\Lambda}^{+} = \{\lambda \in \Delta_J : \sum_{j} \lambda_j \widetilde{p}_{j,k} \geq -\eta \ \forall k\}.$$

~ ~ ~如果该多面体为空，则令 $\Lambda = \Delta_J$；否则令 $\Lambda = \Lambda^{+}$。对于 $\lambda \in \Delta_J$，定义联合支撑范围

$$R_{\lambda,k} = \sup_{z\in\mathcal{Z}} \sum_{j} \lambda_j z_{j,k} - \inf_{z\in\mathcal{Z}} \sum_{j} \lambda_j z_{j,k}.$$

使用精确的联合 范围可以保留边际范围之和可能掩盖的抵消效应 .

~**定理 S7.1 (**有限样本外部权重校准**)**。 假设 $\Lambda \neq \varnothing$。令 $V = \mathrm{ext}(\Lambda)$为其有限、非空的顶点集。给定先导样本，定义

$$r_{n_2} = \max_{v\in V,\,1\leq k\leq K} R_{v,k}\sqrt{\frac{\log(2K|V|/\alpha)}{2n_2}}, \qquad 0 < \alpha < 1.$$

定义概率向量的随机区域

$$\mathcal{C}_{n_1,n_2} = \left\{ q \in \Delta_K : \ \exists \lambda \in \widetilde{\Lambda} : \ \left\| q - \sum_{j} \lambda_j \widehat{p}_j \right\|_{\infty} \leq r_{n_2} \right\}.$$

那么，对于满足上述假设的每个概率律，

$$P(\{p_\lambda : \lambda \in \Lambda\} \subseteq \mathcal{C}_{n_1,n_2}) \geq 1 - \alpha - \beta.$$

保证为 在该有界得分类上均匀分布，其中指定的支持集 $\mathcal{Z}$ 是一个

有效的共同界限。不要求相容集合具备内点性、唯一性或可微性。

证明。对于每个先导坐标，霍夫丁不等式给出 $P(|\widetilde{p}_{j,\,k}-p_{j,\,k}|>\eta)\le 2\exp(-2n_1\eta^2/D_{j,\,k}^2)\le \beta/(JK)$，值域为零的坐标则视为精确值。由并集界可知，事件 $E_1=\{\max_{j,\,k}|\widetilde{p}_{j,\,k}-p_{j,\,k}|\le\eta\}$ 发生的概率至少为 $1-\beta$。对于每个 $\lambda\in\Lambda$ 和 k，凸加权的先导误差至多为 η，因此在 E_1 上有 $\sim\Lambda\subseteq\Lambda^+\subseteq\Lambda$。特别地，只有在 E_1之外才会出现回退情形。

给定先导样本后，V 是固定的，且估计样本相互独立。对于每个$v\in V$和 k，$\sum_j v_j Z_{j,\,k}$ 的第二样本均值的期望为 $p_{v,\,k}$，且其已知值域为$R_{v,\,k}$。根据霍夫丁不等式，并对 $K|V|$ 顶点-坐标对应用并集界，以至少 $1-\alpha$的条件概率可得

$$\max_{v\in V,k}\left|\sum_j v_j(\widehat{p}_{j,k}-p_{j,k})\right|\le r_{n_2}.$$

~任意 $\lambda\in\Lambda$ 都是 V中各点的凸组合。每个坐标的线性性质和三角不等式可将同一误差界推广到每个这样的 λ。在与$\sim E_1$的交集中，每个 $\lambda\in\Lambda$ 都属于 Λ，且真实概率向量 p_λ 位于$\mathcal{C}_{n_1,\,n_2}$。根据独立性和两项条件误差保证，覆盖率至少为$(1-\beta)(1-\alpha)\ge 1-\alpha-\beta$。□

命题 S7.1（有限样本拒绝与不可行性裕度）。保留得分支持集、先导数据独立性和集中性假设，以及定理 J、K、D、β、$\widetilde{\eta}$和 Λ^+ 之前给出的 *S7.1* 定义，但允许总体相容权重集合 Λ 为空。定义总体网格不可行性裕度

$$\gamma=\min_{\lambda\in\Delta_J}\max_{1\le k\le K}[-p_{\lambda,k}]_+,$$

其中 $[x]_+=\max\{x,0\}$。因此，$\gamma=0$ 当且仅当 $\Lambda\neq\varnothing$。使用该定理中的先导数据和阈值 $\widetilde{\eta}$，并在应用用于构造置信区域的全单纯形回退方法之前，当 $\Lambda^+=\varnothing$时拒绝网格相容性。在 $\Lambda\neq\varnothing$ 下，该规则的拒绝概率至多为 β。若 $\gamma>2\eta$，则其拒绝概率至少为 $1-\beta$。特别是，当 $D>0$且 $\gamma>0$时，只要满足以下条件，后一结论就成立

$$n_1>\frac{2D^2}{\gamma^2}\log\frac{2JK}{\beta}.$$

这些概率在满足所述联合得分支持集和独立先导假设的概率律中均匀成立；它们并不检验历史迁移假设是否能够识别真实的未处理反事实。

证明。 根据定理证明中的先导集中论证S7.1（该论证不使用相干性假设），以至少 $1-\beta$ 的概率，每个坐标都满足 $|\widetilde{p}_{j,\,k}-p_{j,\,k}| \leq_{\sim} \eta$。在此事件上，每个总体相干权重都属于 Λ^{+}，从而证明检验尺寸结论。对于任意 $\lambda \in \Delta_J$，由权重的凸性可得 $|\sum_j \lambda_j(\widetilde{p}_{j,\,k}-p_{j,\,k})| \leq \eta$，对每个 k 均成立。若先导外集包含 λ，则对所有 k 均有 $\sum_j \lambda_j \widetilde{p}_{j,\,k} \geq -\eta$，因此对所有 k 均有 $p_{\lambda,\,k} \geq -2\eta$。由此可得 $\gamma \leq 2\eta$。因此，当 $\gamma > 2\eta$ 时，在集中事件上先导外集必为空，从而证明检验功效界。对 $\gamma > 2D\sqrt{\log(2JK/\beta)/(2n_1)}$ 求解 n_1，即可得到最后一个显示式。单纯形的紧致性以及有限维最大值的连续性确保，当且仅当存在总体相干权重时，$\gamma = 0$。

□

S8 一致推断

对于实数轴上的概率分布， d_{Pr} 表示 Prokhorov度量。对于任意此类概率分布R，定义其广义分位数如下：

$$q_R(p) = \inf\{x \in \mathbb{R} : R((-\infty, x]) \geq p\}, \qquad p \in (0,1).$$

对于具有联合可测累积分布函数的随机条件律 Q^{*}，用 $q^{*}(p)$ 表示其广义分位数。在下述紧性条件下，用于校准的分位数均为有限值。

设 $\mathcal{P}_0$ 是完整历史抽样实验中指定的一个非空观测数据概率律类。对于每个 $P \in \mathcal{P}_0$，所有历史顶点累积分布函数及相应的经验过程均在该概率律下定义。本节说明了为确保在 $\mathcal{P}_0$ 上的一致有效性（而非逐点有效性）所需的额外逼近条件。整个过程中，历史指标集 S 固定且有限，并且该类中的每个成员均使用相同的范数和顶点统计量。

假设 S8.1 (U-EP)。对于每个 $P \in \mathcal{P}_0$ 和样本量 N，令

$$T_{N,P} = \max_{s \in S} \|\sqrt{N}(\widehat{H}_s - H_s)\|_{\infty}, \qquad T_P = \max_{s \in S} \|Z_{P,s}\|_{\infty},$$

其中 $Z_P = (Z_{P,\,s})_{s \in S}$ 是一个取值于上确界范数空间有限乘积中的紧的、*Borel* 可测且可分的高斯顶点过程。记

$$T^{*}_{N,P} = \max_{s \in S} \|Z^{*}_{N,s}\|_{\infty}$$

并令 $Q^{*}_{N,\,P}$ 表示给定数据时 $T^{*}_{N,\,P}$ 的一个正则条件律，并选择该条件律，使其累积分布函数关于观测数据和标量自变量联合可测。存在确定性

序列 $\rho_N \downarrow 0$ 和 $\eta_N \downarrow 0$，使得

$$\sup_{P \in \mathcal{P}_0} d_{Pr}\{\mathcal{L}_P(T_{N,P}), \mathcal{L}(T_P)\} \le \rho_N$$

and

$$\sup_{P \in \mathcal{P}_0} P\big[d_{Pr}\{Q^*_{N,P}, \mathcal{L}(T_P)\} > \rho_N\big] \le \eta_N.$$

假设 S8.1 是一个高层次的一致逼近条件。其第一条款要求可测经验统计量在 $P \in \mathcal{P}_0$ 上均受到控制；第二条款要求可测随机条件乘子律以概率趋近于同一高斯参考分布。

条件累积分布函数的联合可测性使下文使用的条件分位数可测。右连续经验累积分布函数轨道和有理阈值约化使$T_{N,\,P}$ 可测。对于真正不可测的过程，则需要改用内概率覆盖表述。

引理 S8.1 (缓冲分位数传递)。 设 X、Y 为实值随机变量，Q 为 $\mathbb{R}$上的固定概率律，并假设

$$d_{Pr}\{\mathcal{L}(X), \mathcal{L}(Y)\} \le \rho, \qquad d_{Pr}\{Q, \mathcal{L}(Y)\} \le \rho.$$

若 $2\rho < \alpha$，则

$$P\{X \le q_Q(1 - \alpha + 2\rho) + 2\rho\} \ge 1 - \alpha,$$

其中 q_Q 是 Q下的广义分位数。若 Q^* 是一个随机条件律，且第二个距离界限在事件 A上成立，则相应的无条件陈述为

$$P\{X \le q^*(1 - \alpha + 2\rho) + 2\rho\} \ge 1 - \alpha - P(A^c).$$

广义逆在原子处有定义。条件 $2\rho < \alpha$ 使分位数索引保持在(0，1)内。

证明。 将 F_R 记为概率律 R的累积分布函数。Prokhorov度量的定义性扩张不等式给出，对于每个 $x \in \mathbb{R}$,

$$F_Q(x) \le F_Y(x + \rho) + \rho.$$

因此，对每个 $p \in (\rho,\ 1)$,

$$q_Q(p) \ge q_Y(p - \rho) - \rho.$$

由于两个分位数都是广义逆，因此这一推论在原子处仍然成立。令$p = 1 - \alpha + 2\rho$。另一个 Prokhorov界给出 $F_X(x) \geq F_Y(x - \rho) - \rho$。在$x = q_Y(1 - \alpha + \rho) + \rho$处求值，得到

$$P\{X \leq q_Y(1 - \alpha + \rho) + \rho\} \geq 1 - \alpha.$$

分位数比较表明，引理中的阈值弱大于 $q_Y(1 - \alpha + \rho) + \rho$，从而证明了固定-$Q$ 主张。对于随机 Q^*，令 $B = \{X \leq q_Y(1 - \alpha + \rho) + \rho\}$。在 A 上，条件分布距离逐点给出 $q^*(1 - \alpha + 2\rho) + 2\rho \geq q_Y(1 - \alpha + \rho) + \rho$。因此， $A \cap B$ 包含在引理所示的事件中，其概率至少为$1 - \alpha - P(A^c)$。

□

定理 S8.1 (一致缓冲覆盖)。 假设 *S8.1* 成立，且 $0 < \alpha < 1$。对于每个充分大的 N 满足 $2\rho_N < \alpha$，定义

$$\varepsilon_{N,P}^{\text{buf}} = N^{-1/2}\{q_{Q^*_{N,P}}(1 - \alpha + 2\rho_N) + 2\rho_N\},$$

其中，条件分位数是在每个样本实现下根据观测到的乘子分布定义的。近似事件不可观测，仅用于覆盖证明。于是

$$\inf_{P \in \mathcal{P}_0} P\left\{\mathcal{F}_S(P) \subseteq \mathfrak{C}_{F,N}(\varepsilon_{N,P}^{\text{buf}})\right\} \geq 1 - \alpha - \eta_N.$$

因此，

$$\liminf_{N \to \infty} \inf_{P \in \mathcal{P}_0} P\left\{\mathcal{F}_S(P) \subseteq \mathfrak{C}_{F,N}(\varepsilon_{N,P}^{\text{buf}})\right\} \geq 1 - \alpha.$$

证明。 固定 $P \in \mathcal{P}_0$ 并令

$$A_{N,P} = \left\{d_{Pr}\{Q^*_{N,P}, \mathcal{L}(T_P)\} \leq \rho_N\right\}.$$

U-EP 的第二个子句给出 $\sup_{P \in \mathcal{P}_0} P(A^c_{N,\,P}) \leq \eta_N$。应用引理 S8.1，其中 $X = T_{N,\,P}$, $Y = T_P$, $Q^* = Q^*_{N,\,P}$，以及 $A = A_{N,\,P}$。于是，U-EP 的第一个子句给出

$$P\left\{T_{N,P} \leq q_{Q^*_{N,P}}(1 - \alpha + 2\rho_N) + 2\rho_N\right\} \geq 1 - \alpha - \eta_N.$$

花括号内的事件恰好是半径为 $\varepsilon^{\text{buf}}_{N\,,\,P}$ 的同时顶点事件。主论文中的确定性覆盖定理将该事件映射为整体集合包含。由于界限与 P 无关，对 $\mathcal{P}_0$ 取下确界即可证明有限-N 结论。由 $\eta_N \to 0$ 可得极限情形的结论。

□

该定理不使用 T_P的概率律的反集中性质或连续性。缓冲区是根据可观测条件分布定义的，但其数值取决于一个经过认证的

ρ_N的上界。均匀有效性的主张不能仅基于逐点乘子中心极限定理或调查复制权重。本文的实证应用并未建立这样的界限。

S8.1 重复横截面

确定性覆盖定理不要求纵向观测。记

$$H_s = F_{0,1} + B_s,$$

其中，$F_{0,\,1}$ 是处理后比较组累积分布函数，B_s 是历史的处理组-对照组累积分布函数差异。

命题 S8.1（区块半径转移）。假设随机半径在事件 E 上满足

$$\|\widehat{F}_{0,1} - F_{0,1}\|_\infty \le r_1, \qquad \|\widehat{B}_s - B_s\|_\infty \le r_s \quad \forall s \in S.$$

那么在 E 上，

$$\max_{s\in S} \|\widehat{H}_s - H_s\|_\infty \le r_1 + \max_{s\in S} r_s.$$

因此，任何针对区块半径的联合构造或邦费罗尼构造都可以纳入主要的相干管定理中。

邦费罗尼版本不要求跨年度独立性。

S8.2 整群抽样

政策数据通常以簇为单位抽样：州、地方区域或调查初级抽样单位；每个簇内的单元、日期和结果单元之间可能存在任意依赖关系。主论文中的覆盖定理一旦有了顶点单元质量的同时置信带，便是确定性的，因此在整群抽样下，改变的只有该置信带的校准方式。本小节介绍固定网格上的两种校准方法，一种在有限样本中精确，另一种渐近且均匀，然后给出相干性检验功效的界限。下文中，网格 $\Pi = \{A_1, \dots, A_K\}$ 和历史日期集合 S（其中 $J = |S|$）均为固定。顶点单元质量构成 $J \times K$ 数组 Z，其元素为 $z_{s,\,k} = \mu_{0,\,1}(A_k) + \mu_{1,\,s}(A_k) - \mu_{0,\,s}(A_k)$，$Z$ 的每一行之和均为一，且 $Z_\lambda = \sum_s \lambda_s Z_{s\cdot}$。对于数组 V，记 $\|V\|_\infty = \max_{s,\,k} |v_{s,\,k}|$。我们用 C 表示簇的数量；组别指示变量仍为 G。

研究设计。处理组群组 $c \in \mathcal{T}$和比较组群组 $c \in \mathcal{U}$（$C_1 = |\mathcal{T}|$ 和$C_0 = |\mathcal{U}|$）彼此独立。每个处理簇报告其自身的估计单元

份额 $\widehat{\mu}^{c}_{1,\,s}(A_k)$、 $s \in S$，且每个比较簇报告 $\widehat{\mu}^{c}_{0,\,1}(A_k) - \widehat{\mu}^{c}_{0,\,s}(A_k)$，这些值均根据簇内任意样本计算得出。该估计量是等权平均值

$$\widehat{z}_{s,k} = \frac{1}{C_0} \sum_{c \in \mathcal{U}} \{\widehat{\mu}^{c}_{0,1}(A_k) - \widehat{\mu}^{c}_{0,s}(A_k)\} + \frac{1}{C_1} \sum_{c \in \mathcal{T}} \widehat{\mu}^{c}_{1,s}(A_k),$$

而目标 Z 是其期望：簇平均未处理分布，其中的簇来自各自的总体。对于一个簇在不同日期、组或单元上的报告，其联合分布不作任何假设。若对单元赋予相同权重（例如汇总所有簇的记录），则 Z 会变成簇总量之比。此时，每个单元都是 C 独立有界向量（簇按组、日期和单元划分的计数，以及各组的规模）均值的光滑函数。若簇规模有界且其均值远离零，一阶展开便给出下文假设S8.2 的第 (i) 条，其中 ψ_c 是簇的线性化贡献，且 $R_C = O_P(C^{-1})$，因此适用高斯校准；非参数簇自助法是相应的重抽样方法。有限样本半径不适用。

在许多应用中，两组都在相同的簇中观测：母亲和无子女女性居住在相同的州。在这种共享簇设计中，每个 C 独立簇都会报告完整的数组 V_c，其中各项为 $v_{c,\,s,\,k} = \widehat{\mu}^{c}_{0,\,1}(A_k) + \widehat{\mu}^{c}_{1,\,s}(A_k) - \widehat{\mu}^{c}_{0,\,s}(A_k)$，这些数据根据簇内居民计算得出，且 $\widehat{Z} = C^{-1} \sum_c V_c$。于是，同一簇内两组的估计误差可能以任意方式相关，因此必须按簇将两组一起重抽样。

命题 S8.2（有限样本聚类半径）。对于 $0 < \alpha < 1$，$\mathbb{P}(\|Z - Z\|_\infty \le t_C) \ge 1 - \alpha$，其中

$$t_C = \begin{cases} \left\{ \dfrac{\log(2JK/\alpha)}{2} \left(\dfrac{1}{C_1} + \dfrac{4}{C_0} \right) \right\}^{1/2} & \textit{separate clusters}, \\ \left\{ \dfrac{9 \log(2JK/\alpha)}{2C} \right\}^{1/2} & \textit{shared clusters}. \end{cases}$$

因此，定理 *6* 在网格上的置信域，半径为 t_C，以至少 $1 - \alpha$的概率包含由单元向量构成的相容集合；当时拒绝相容性的检验 $d(Z) > t_C$，其中 $d(V) =$ 最小值$_{\lambda \in \Delta_S}$ 最大值$_k [-(V\lambda)_k]_+$，其检验尺寸至多为 α。

证明。固定 (s, k)。偏差 $\widehat{z}_{s,k} - z_{s,k}$ 是由相互独立的零均值项组成的和，每个簇对应一项。处理簇的项是 $C_1^{-1} \widehat{\mu}^{c}_{1,s}(A_k)$ 减去其均值，而 $\widehat{\mu}^{c}_{1,s}(A_k) \in [0, 1]$，因此该项的值域是长度为 $1/C_1$ 的区间；对照簇的项的值域是长度为 $2/C_0$ 的区间，因为两个份额之差落在 $[-1, 1]$ 内。根据霍夫丁不等式，$\mathbb{P}(|\widehat{z}_{s,k} - z_{s,k}| > t) \le 2\exp\{-2t^2/(C_1 \cdot C_1^{-2} + C_0 \cdot 4C_0^{-2})\}$，在 $t = t_C$ 处等于 $\alpha/(JK)$。在共享簇设计中，每个 $v_{c,s,k}$ 都落在 $[-1, 2]$ 内，因此每个簇的项的值域是长度为 $3/C$ 的区间，界限为 $2\exp(-2Ct^2/9)$。并集界

在 JK 个坐标上证明了第一个结论。在该事件中，每个相容的 Z_λ 都满足$\widehat{\ }$ $\widehat{\ }$ $\|Z_\lambda - Z_\lambda\|_\infty \le \|Z - Z\|_\infty \le t_C$ 根据引理 1，这给出了该区域。如果 λ_0 相容，$\widehat{\ }$ $\widehat{\ }$ $(Z_{\lambda_0})_k \ge 0$，因此 $[-(Z_{\lambda_0})_k]_+ \le |(Z_{\lambda_0} - Z_{\lambda_0})_k| \le t_C$ 对每个 k 以及 $d(Z) \le t_C$。 □

半径仅取决于簇的数量，而不取决于每个簇中的单元数量，并且由较小的组主导。对于这种形式的研究设计，在最低工资应用中有五个处理组州、四个历史日期和118个单元格，$t_C \ge 0.99$ at$\alpha = 0.05$，单元格本身的尺度（此处使用分簇公式，因为处理州和对照州彼此不同）；日历时间上重叠的事件窗口还会在此基础上增加事件之间的依赖关系。因此，整群抽样无法使该应用具备推断性，这就是我们仅对其进行描述性报告的原因。

高斯校准。更尖锐的临界值来自簇中心极限定理。令 $\widehat{Z} - Z = C^{-1}\sum_{c=1}^{C}\psi_c + R_C$，其中 $C = C_0 + C_1$，ψ_c是相互独立、均值为零的$J \times K$ 数组；在上述研究设计中，ψ_c 是 C/C_1 或 C/C_0 乘以簇 c和 $R_C = 0$的中心化报告，且这些报告有界，因此当 C_1/C 和 C_0/C远离零时，以下条件 (i) 成立。

假设 S8.2 (簇)。 对于数据生成过程类 $\mathcal{P}_C$ 和常数 $B < \infty$、$\sigma_{\min} > 0$，一致地对所有 $P \in \mathcal{P}_C$ 成立：*(i)* J 和 K 固定，且 ψ_c 相互独立，满足 $\|\psi_c\|_\infty \le B$；*(ii)* 平均协方差 $\Omega_C = C^{-1}\sum_c \mathrm{Var}\{\mathrm{vec}(\psi_c)\}$ 的每个对角元素都至少为 $\sigma^2_{\min}$，几乎必然为常数的坐标除外；这些坐标会从置信带中移除，并被精确施加；*(iii)* $\sqrt{C}\|R_C\|_\infty \le r_C$，其中 $r_C \to 0$ 为确定性值；*(iv)* $\widehat{\ }$ $\widehat{\ }$ 估计量 $\overline{\Omega}$ 满足 $\mathbb{P}(\max_{i,j}|\Omega_{ij} - \Omega_{C,ij}| > \Delta_C) \le \eta_C$，其中 Δ_C、$\eta_C \to 0$ 为确定性值。

共享簇设计满足条件 (i)，其中 $\psi_c = V_c - \mathbb{E}V_c$，界限为 3，且$R_C = 0$。在任一研究设计中，条件 (iv) 均成立，其中 $\widehat{\Omega} = C^{-1}\sum_c \widehat{\psi}_c\widehat{\psi}_c^\top$，其中 $\widehat{\psi}_c$是该簇的报告减去其所在组（或共享设计中的所有簇）的平均报告；当每组内的簇服从相同分布时，对有界乘积应用霍夫丁不等式可得 Δ_C 的阶为 $(\log C/C)^{1/2}$。当簇的均值不同时，$\widehat{\Omega}$ 估计 Ω_C 再加上一个半正定项；根据安德森定理（Anderson，1955），此时下文的对称置信带是保守的，但单侧置信带未必如此；其有效性要求使用一个在异质均值下仍一致的 Ω_C 估计量，例如移除经模型拟合的簇特定均值的估计量。

定理 S8.2 (一致簇校准)。 在假设 *S8.2* 下，设 $\widehat{q}_{\mathrm{sym}}$ 和 $\widehat{q}_{\mathrm{one}}$ 是给定数据后，$\|W\|_\infty$ 和 $\max_{s,k}[-W_{s,k}]_+$ 关于 $\mathrm{vec}(W) \sim$ 的分位数。

$\sim N(0, \Omega)$。然后

$$\liminf_{C\to\infty} \inf_{P\in\mathcal{P}_C} \mathbb{P}\big(\sqrt{C}\|\widehat{Z} - Z\|_\infty \le \widehat{q}_{\mathrm{sym}}\big) \ge 1-\alpha,$$
$$\limsup_{C\to\infty} \sup_{P\in\mathcal{P}_C:\,\Lambda_{S,\Pi}(P)\neq\varnothing} \mathbb{P}\big(\sqrt{C}\, d(\widehat{Z}) > \widehat{q}_{\mathrm{one}}\big) \le \alpha.$$

证明。两个事件都具有 $\{\sqrt{C}(\widehat{Z}-Z)^{\top}\in\mathcal{R}\}$ 的形式，其中 $\mathcal{R}$ 是 $\mathbb{R}^{JK}$ 中的一个超矩形（忽略余项），因为 $\{v:\ \|v\|_\infty \le q\}$ 和 $\{v:\ \min_{s,k} v_{s,k} \ge -q\}$ 都是矩形；对于检验，我们使用下文命题 S8.3 第 (ii) 步的界限，即 $d(Z) \le \max_{s,k}|-(Z-Z)_s|$，在原假设下。令 $W\sim N(0,\Omega_C)$。(a) 根据 Chernozhukov、Chetverikov 和 Kato (2017) 关于独立（但不一定同分布）的有界向量在超矩形上的中心极限定理，$\sup_{\mathcal{R}} |\mathbb{P}\{C^{-1/2}\sum_c \psi_c \in \mathcal{R}\} - \mathbb{P}(W\in\mathcal{R})| \le \kappa_{1,C}$，其中 $\kappa_{1,C}$ 的阶为 $\{B^2\log^7(JKC)/C\}^{1/6}$，常数仅取决于 $\sigma_{\min}$。(b) 余项使统计量的偏移至多为 r_C；根据 Chernozhukov、Chetverikov 和 Kato (2015) 关于高斯最大值的反集中不等式（该不等式只需要 (ii) 中的方差下界），这会使概率最多改变 $\kappa_{2,C} = r_C\{(2\log(2JK))^{1/2}+2\}/\sigma_{\min}$。(c) 在条款 (iv) 中的事件上，Chernozhukov、Chetverikov 和 Kato (2015) 关于高斯最大值的比较不等式，将 $N(0,\widehat{\Omega})$ 与 $N(0,\Omega_C)$ 下两个最大值的分布函数之差界定为 $\kappa_{3,C}$，其阶为 $\Delta_C^{1/3}\{1\vee\log(JK/\Delta_C)\}^{2/3}$，常数仅取决于 $\sigma_{\min}$ 和 B。步骤 (a)–(c) 将统计量的概率律与其条件高斯近似之间的柯尔莫哥洛夫距离界定为 $\kappa_C = \kappa_{1,C}+\kappa_{2,C}+\kappa_{3,C}$，除去概率为 η_C 的事件后，该界对 $\mathcal{P}_C$ 均匀成立，因为每个常数都仅通过 $(B,\sigma_{\min},J,K)$ 依赖于 P。随后，标准的分位数转移给出置信带的覆盖率至少为 $1-\alpha-2\kappa_C-\eta_C$，以及检验尺寸至多为 $\alpha+2\kappa_C+\eta_C$；两项误差都趋于零。

□

将每个坐标除以 $\widehat{\sigma}_{s,k}\vee f$ 其中下限 $f>0$ 预先固定的学生化带，在 $f\le\sigma_{\min}$ 且 $\widehat{\sigma}_{s,k}$ 一致相合时，经过坐标重缩放后也可用同样的论证覆盖；若 $f>\sigma_{\min}$，则仍然有效，但对受影响的单元而言较为保守。若某个单元的方差没有正下界，则假设 S8.2(ii) 不成立。只有当研究设计使某个单元成为常量时，才能将其精确地施加约束，例如，顶端编码以上且任何报告都无法达到的单元；样本方差为零并不能证明这一点，这种单元应改为单独置于 Hoeffding 带中，并分配一部分 α。若命题 8的基准与对照组共享簇，则必须将基准报告与对照组报告联合重抽样，而非独立重抽样。

检验功效。 检验统计量为亏损 d。以下性质适用于任何估计量，且不需要分布假设。

命题 *S8.3*（亏损性质与功效界）。*(i)* $|d(V)-d(V')| \le \|V-V'\|_\infty$ 对所有 $J\times K$ 数组 V, V' 均成立。*(ii)* 若 $\Lambda_{S,\Pi}(P) \neq \varnothing$，则 $d(Z) \le$ 最大值为 $s,k[-(Z-Z)_s, {}_k]_+ \le \|Z-Z\|_\infty$。*(iii)* 对任意阈值 t，当 $d(Z) > t$ 时拒绝的检验满足

$$\mathbb{P}\big(\|\widehat{Z}-Z\|_\infty < d(Z)-t\big) \le \mathbb{P}\big(d(\widehat{Z}) > t\big) \le \mathbb{P}\big(\|\widehat{Z}-Z\|_\infty > t-d(Z)\big).$$

(iv) 在局部备择 Z_C 下，若 $\sqrt{C}\, d\,(Z_C)^{-} \to \gamma$，阈值满足 $\sqrt{C}t_C \to q,^{-}$且$\sqrt{C}\,(\widehat{Z^{-}} - Z_C) \Rightarrow W$满足 W为高斯且非退化，则在连续点处，极限拒绝概率介于 $\mathbb{P}(\|W\|_\infty < \gamma - q)$ 与 $\mathbb{P}(\|W\|_\infty > q-\gamma)$ 之间。特别地，阈值取为 $\|W\|_\infty$的 $1-\alpha$分位数的检验，对于满足 $\gamma = 0$ 的备择假设，其渐近功效至多为 α；当 $\gamma \to \infty$时，功效趋于 *1*。

证明。(i) 对每个 λ，$\max_k[-(V_\lambda)_k]_+ \le \max_k[-(V'_\lambda)_k]_+ + \|V_\lambda - V'_\lambda\|_\infty$，因为 $x \mapsto [-x]_+$是 1-利普希茨的；而 $\|V_\lambda - V'_\lambda\|_\infty \le \|V-V'\|_\infty$，因为 λ 是一个概率向量。对两边的 λ 取最小值，并交换 V 与 V'，即可得到 (i)。(ii) 设 λ_0 相容。对每个 $\sim k$，$(Z_{\lambda_0})_k \ge (Z_{\lambda_0} - Z_{\lambda_0})_k$，因此 $[-(Z_{\lambda_0})_k]_+ \le [-(Z_{\lambda_0} - Z_{\lambda_0})_k]_+ \le \sum_s \lambda_0,\ {}_s[-(Z-Z)_s,\ {}_k]_+ \le \max_s[-(Z-Z)_s,k]_+$，这是由 $x \mapsto [-x]_+$的凸性得出的；而 $d(Z)$不超过对 k取最大值的结果。(iii) 根据 (i)，$d(Z) \ge d(Z) - \|Z-Z\|_\infty$ 且 $d(Z) \le d(Z) + \|Z-\overline{Z}\|_\infty$；由此得到两个事件包含关系。(iv) 将 (iii) 乘以 $\sqrt{C}$，用 Z_C 代替 Z，并在连续点处应用连续映射定理取极限。当 $\gamma = 0$ 时，上界为$\mathbb{P}(\|W\|_\infty > q) = \alpha$；当 $\gamma \to \infty$ 时，下界趋于 1。 □

第 (ii) 部分表明，单侧临界值 $\widehat{q}_{\mathrm{one}}$ 有效，且绝不会大于 $\widehat{q}_{\mathrm{sym}}$；第 (iv) 部分表明，要获得功效，亏损必须达到 $C^{-1/2}$ 的量级。对于同一统计量，较小的临界值会逐路径地带来不低于原先的功效；但这并不能说明其功效相对于其他统计量如何。(iii) 和 (iv) 中的界限本身并不是功效函数。下一个结果将给出功效函数。

精确局部极限。固定 Z_0，使其满足 $d(Z_0) = 0$，并令 $\Lambda^* = \{\lambda \in \Delta_S : (Z_{0,\ \lambda})_k \ge 0$ 对所有 $k\}$ 均成立；其非空相容集合为 $Z_{0,\ \lambda} = \sum_s \lambda_s Z_{0,\ s\cdot}$。对于 $\lambda \in \Lambda^*$，令 $K^*(\lambda) = \{k : (Z_{0,\ \lambda})_k = 0\}$ 为其绑定单元，并令 $T(\lambda) = \{v \in \mathbb{R}^J : \sum_s v_s = 0,\ \ v_s \ge 0$，只要 $\lambda_s = 0\}$，它就是 Δ_S 在 λ 处的可行方向锥。对于一个 $J\times K$ 数组 H，记作 $H_\lambda = \sum_s \lambda_s H_{s\cdot}$，并定义

$$D(H) = \inf_{\lambda\in\Lambda^*} \inf_{v\in T(\lambda)} \max\Big\{0,\ \max_{k\in K^*(\lambda)} \big[-(Z_{0,v})_k - (H_\lambda)_k\big]\Big\},$$

其中，空集上的内部最大值记作 $-\infty$。

定理 S8.3 (相容性统计量的局部极限)。*(i)* 亏损 d 在 Z_0 处 *Hadamard* 方向可微，其导数为 D：对于 $t \downarrow 0$ 和 $H_t \to H$，$d(Z_0 + tH_t)/t \to D(H)$。映射 D 是正齐次的，且 $|D(H) - D(H')| \le \|H - H'\|_\infty$。*(ii)* 若

$Z_C = Z_0 + C^{-1/2}H + o(C^{-1/2})$ 和 $\sqrt{C}(\widehat{Z} - Z_C) \Rightarrow W$，则 $\sqrt{C}\, d(\widehat{Z}) \Rightarrow D(H+W)$，且阈值为 $\sqrt{C}t_C \to q$ 的检验在连续点处的极限拒绝概率为 $\mathbb{P}\{D(H+W) > q\}$。*(iii)* 若某个 $\lambda \in \Lambda^*$ 满足 $K^*(\lambda) = \varnothing$，则对于所有足够大的 C，$D \equiv 0$，Z_C 相容，且对于每个 $q > 0$ 和每个 H，极限拒绝概率均为零。*(iv)* 若对每个 C，Z_C 均相容，则 $D(H+W)$ 的概率律随机占优于 $\max_{s,k}[-W_{s,k}]_+$ 的概率律。

证明。（i）写作 $\phi(\lambda, V) = \max\{0, \max_k[-(V\lambda)_k]\}$，使得 $d(V) = \min_{\lambda\in\Delta_S}\phi(\lambda, V)$，并令 $V_t = Z_0 + tH_t$。上界。固定 $\lambda \in \Lambda^*$ 和 $v \in T(\lambda)$。由于 Δ_S 是多面体，对于足够小的 t，$\lambda + tv \in \Delta_S$。对于 $k \notin K^*(\lambda)$，有 $(Z_0\lambda)_k > 0$，因此 $\phi(\lambda + tv, V_t)$ 的第 k 项在 t 足够小时为负；对于 $k \in K^*(\lambda)$，该项等于 $-t(Z_0,v)_k - t(H_t,\lambda)_k - t^2(H_t,v)_k$。因此 $\limsup_t d(V_t)/t \le \limsup_t \phi(\lambda+tv, V_t)/t = \max\{0, \max_{k\in K^*(\lambda)}[-(Z_0,v)_k - (H\lambda)_k]\}$，对 (λ, v) 取下确界，得到 $\limsup_t d(V_t)/t \le D(H)$。下界。令 λ_t 最小化 $\phi(\cdot, V_t)$。对于任意 $\lambda_0 \in \Lambda^*$，$d(V_t) \le \phi(\lambda_0, V_t) \le t\|H_t\|_\infty$，因此 $\phi(\lambda_t, Z_0) \le \phi(\lambda_t, V_t) + t\|H_t\|_\infty \le 2t\|H_t\|_\infty$。集合 Λ^* 是由线性不等式定义的多面体，而 $\phi(\lambda, Z_0)$ 是那些尚未由 $\lambda \in \Delta_S$ 施加的不等式中的最大违反量；根据霍夫曼误差界（Hoffman，1952年），存在仅依赖于 Z_0 的 $\kappa < \infty$，以及一点 $\mu_t \in \Lambda^*$，使得 $\|\lambda_t - \mu_t\|_1 \le \kappa\phi(\lambda_t, Z_0) \le 2\kappa t\|H_t\|_\infty$。令 $v_t = (\lambda_t - \mu_t)/t$；那么 $\|v_t\|_1 \le 2\kappa\|H_t\|_\infty$ 且 $v_t \in T(\mu_t)$，因为 $\mu_t + tv_t \in \Delta_S$。由于 $\lambda_t = \mu_t + tv_t$，V_t 的第 k 个单元在 $k \in K^*(\mu_t)$ 时等于 $t\{(Z_0,v_t)_k + (H_t,\mu_t)_k\} + t^2(H_t,v_t)_k$。仅保留这些单元，并利用 $|(H_t,v_t)_k| \le \|H_t\|_\infty\|v_t\|_1$ 以及 (μ_t, v_t) 在 $D(H_t)$ 的定义中是可行的这一事实，

$$\begin{aligned}\frac{d(V_t)}{t} = \frac{\phi(\lambda_t, V_t)}{t} &\ge \max\Big\{0, \max_{k\in K^*(\mu_t)}\big[-(Z_{0,v_t})_k - (H_{t,\mu_t})_k\big]\Big\} - t\|H_t\|_\infty\|v_t\|_1 \\ &\ge D(H_t) - 2\kappa t\|H_t\|_\infty^2.\end{aligned}$$

由于对于每个概率向量 λ，$|(H_t,\lambda)_k - (H\lambda)_k| \le \|H_t - H\|_\infty$，因此 D 内的表达式变化至多为 $\|H_t - H\|_\infty$，这证明了利普希茨性质以及 $D(H_t) \to D(H)$。因此 $\liminf_t d(V_t)/t \ge D(H)$。由于 $T(\lambda)$ 是一个锥，正齐次性成立。(ii) 令 $\sqrt{C}\, d(\widehat{Z}) = \sqrt{C}\{d(Z_0 + C^{-1/2}h_C) - d(Z_0)\}$，其中 $h_C = \sqrt{C}(\widehat{Z} - Z_C) + \sqrt{C}(Z_C - Z_0) \Rightarrow W + H$；针对 Hadamard 方向可微映射的德尔塔方法（Shapiro，1991年；Fang 和 Santos，2019年）给出了该极限，而拒绝概率可由连续点处的结果推出。(iii) 对于此类 λ，内部最大值是在空集上取值，因此方括号内的值为零。对于较大的 C，Z_C 的相容性成立，因为 $Z_{0,\lambda}$ 的每个单元都严格为正，且 $\|Z_C - Z_0\|_\infty \to 0$。(iv) 根据命题 S8.3(ii)，对于每个 C，$\sqrt{C}\, d(\widehat{Z}) \le \max_{s,k}[-\sqrt{C}(\widehat{Z} - Z_C)_{s,k}]_+$；弱极限保留了分布函数的这一序关系。

□

三个推论。第一，极限分布仅通过Λ^* 和绑定单元取决于配置。当边界上的唯一权重处有一个绑定单元时，

单纯形的 $\lambda_0 =$（表 S1 研究设计中的 1, 0），锥为 $\{(-a, a)：\ a \geq 0\}$，沿锥移动只会降低该绑定单元，而 $D(H) = [-H_{1,1}]_+$。当唯一内部权重 $\lambda_0 = (1/2, 1/2)$ 处有两个绑定单元时，权重可以朝任一方向移动；在该研究设计中，移动权重会以相同速率提高一个绑定单元并降低另一个，而对其取最小值可得 $D(H) = [-\{(H_{\lambda_0})_1 + H_{\lambda_0 2}\}/2]_+$：统计量对两个绑定单元取平均，而不是取较差者。数值检验逐次抽样复现了这两个公式：第一种情形下完全一致；第二种情形下，$C = 50$ 处的平均绝对差异为 0.002，在 $C = 800$ 处降至 0.0007。第二，(iv)部分解释了模拟中的保守性：顶点临界值是所有 JK 坐标上 $\max_{s,k}[-W_{s,k}]_+$ 的分位数，而零假设极限 $D(W)$ 使用少数几个经平均的绑定单元。第三，(iii)部分描述了原假设内部：当相容集合包含一个没有绑定单元的权重时，对 Z_0 的每个局部扰动仍然相容，因此这些序列都是零假设序列，检验沿这些序列的拒绝概率趋于零，而非 α。因此，要获得非平凡的局部功效，配置必须满足每个相容权重都有一个绑定单元，并且存在一个能将它们全部推出去的方向 H。基于 $D(\cdot)$ 估计的临界值，其中 Λ^* 和 K^* 使用调节序列估计，所得检验在逐点意义上有效（Fang 和 Santos，2019），且功效更高；其一致有效性需要我们尚未证明的条件，而表 S1 中的接触集规则这一粗略版本会导致过度拒绝。

模拟。 表 S1 报告了一项蒙特卡洛研究，其中包含 $J = 2$ 个历史日期和 $K = 5$ 个单元。每个 $C \in \{20, 40, 80, 160\}$ 个簇贡献一个有界数组，其两行相关（$\rho \in \{0, 0.9\}$），且不同簇之间的尺度相差一倍或 $\sqrt{10}$。研究采用四种原假设配置：每个单元均位于内部；一个唯一相容权重，包含一个绑定单元；一个唯一权重，包含两个绑定单元；以及一组相容权重，其边缘处存在绑定单元。局部备择将数量级为 $\delta C^{-1/2}$ 的质量移出绑定单元。检验尺寸取以下16种组合中的最大值：（C、ρ、尺度），每种组合均进行2,000次重复模拟（在5%的水平下，蒙特卡洛标准误最多为0.5个百分点）；功效使用1,000次重复模拟，且 $\rho = 0.5$；高斯临界值基于1,000次抽取得出。

有三项发现值得注意。所有顶点置信带检验在每种配置下都能控制检验尺寸，最坏情况拒绝率不超过 1.3%；它们较为保守，因为亏损是权重上的最低值，而置信带覆盖了每个顶点。单侧校准的功效始终高于对称校准，在中间备择假设下高出 4 到 8 个百分点；在这些研究设计中，学生化带来的提升微乎其微。有限样本霍夫丁检验是有效的，但在这些簇数量下没有功效。接触集规则在边界附近的功效要高得多，但在单个绑定单元配置下，其最坏情形检验规模为 6.7%，超过名义水平的幅度大于模拟误差；这正是一致性文献所警示的风险。当相容集合存在松弛时，对绑定

表 S1：整群抽样下的相容性检验：最坏情形检验规模与局部功效

	$\sqrt{C}\,d$	对称	单侧 Student。		Hoeffding	接触
检验尺寸：对 C、ρ 和整群尺度取最坏情形						
内部	0	0.000	0.000	0.000	0.000	0.000
唯一权重，一个绑定单元	0	0.007	0.011	0.013	0.000	0.067
唯一权重，两个绑定单元	0	0.000	0.000	0.000	0.000	0.003
权重集合，在边缘处绑定	0	0.000	0.000	0.000	0.000	0.000
功效，$C = 160$，检验尺寸为 $\delta C^{-1/2}$ 的备择假设						
一个绑定单元，$\delta = 0.5$	0.15	0.033	0.058	0.065	0.000	0.243
一个绑定单元，$\delta = 1$	0.30	0.211	0.282	0.285	0.000	0.629
一个绑定单元，$\delta = 2$	0.60	0.861	0.910	0.918	0.000	0.989
一个绑定单元，$\delta = 4$	1.20	1.000	1.000	1.000	0.000	1.000
两个绑定单元，$\delta = 0.5$	0.08	0.000	0.000	0.000	0.000	0.009
两个绑定单元，$\delta = 1$	0.15	0.002	0.004	0.005	0.000	0.094
两个绑定单元，$\delta = 2$	0.30	0.061	0.141	0.146	0.000	0.686
两个绑定单元，$\delta = 4$	0.60	0.985	0.996	0.995	0.000	1.000

注：名义水平为 5%。对称检验和单侧检验：采用定理 S8.2 的高斯校准。学生化检验：单侧检验，坐标除以 $\widehat{\sigma}_{s,\,k} \vee 0.02$。Hoeffding：命题 S8.2。接触集：临界值取自估计权重下距离零点不超过 $(\log C)^{1/2}$ 个标准误的单元；–没有一致有效性结果。$\sqrt{C}\,d$：备择假设的缩放总体缺口。

单元的局部扰动会使亏损保持为零，因此没有任何检验具有功效，这正如定理 S8.3 的第 (iii) 部分所预测的那样。我们建议对聚类数据使用单侧高斯校准；若需要有限样本保证，且簇数量达到数百，则建议使用 Hoeffding 半径。

定理 S8.2 背后的高维中心极限定理，其误差阶为$\{\log^7(JKC)/C\}^{1/6}$，当 $K = 118$ 和 $C = 50$ 时，这一误差并不小。因此，另外两项实验将最后三个单元分别拆分为 20 个和 116 个单元，其中 $C \in \{15, 25, 50\}$，且整群尺度相差 $\sqrt{10}$ 倍。高斯顶点置信带本身在所有（K，C）条件下的覆盖率介于 0.94 和 0.98 之间；对于许多单元较多而簇较少的情况，覆盖率略高于名义水平。基于这些置信带的相容性检验在原假设下的拒绝率至多为 1.8%，包括 $K = 118$ 和 $C = 15$ 的情形。接触集规则在只有五个单元时会过度拒绝（在 $C = 25$ 时最高达 7.5% ），而单元较多时则趋于保守。若某个单元的整群报告几乎没有变化（标准差为 0.003，而下限为 0.02），则三种高斯检验仍然全部有效。这些研究设计中的报告有界且尾部较轻；若整群报告具有重尾，近似效果可能更差，此时 Hoeffding 半径是更稳妥的选择。

S9 处理前数据无法识别的内容

对于 $S \subset S'$，令 $\mathfrak{M}_S \subseteq \mathfrak{M}_{S'}$ 为相应的结构模型类，并令 $\theta(Q) = F^{0}_{1,\,1,\,Q}$。将观测样本记为 O_N。在下文关于检验和置信集的结论中，假设其在完整数据分布 Q 下的概率律为 $\mathrm{K}_N\{\mathcal{O}(Q)\}$，其中抽样核 K_N 是各个比较概率律所共有的

。对完整面板历史进行独立抽样，或在共同的固定单元抽样规则下对重复横截面进行独立抽样，均满足这一条件。该结论以抽样核为条件：仅有相同的边际观测数据分布，并不足以确定调查样本的联合设计分布（如果研究设计本身不同）。随机化程序在所比较的模型下必须使用相同的辅助随机化分布。

T定理 S9.1（精确观测等价）。 对于观测分布 P ，

$$\mathcal{F}_S(P) \subsetneq \mathcal{F}_{S'}(P),$$

则存在 $Q_a \in \mathfrak{M}_S$ 和 $Q_b \in \mathfrak{M}_{S'} \setminus \mathfrak{M}_S$ ，使得

$$\mathcal{O}(Q_a) = \mathcal{O}(Q_b) = P.$$

在共同抽样核条件下，对于任意样本量，任何基于观测数据的检验在两个概率律下的拒绝概率都相同。

证明。 选择 $F^a \in \mathcal{F}_S(P)$ 和 $F^b \in \mathcal{F}_{S'}(P) \setminus \mathcal{F}_S(P)$。 S下的尖锐性给出 $Q_a \in \mathfrak{M}_S$，使得 $\mathcal{O}(Q_a) = P$ 且 $\theta(Q_a) = F^a$。 S' 下的尖锐性给出 $Q_b \in \mathfrak{M}_{S'}$ ，使得 $\mathcal{O}(Q_b) = P$且 $\theta(Q_b) = F^b$。如果 Q_b属于 $\mathfrak{M}_S$，那么 S下的尖锐性会使 F^b落入 $\mathcal{F}_S(P)$，这就产生了矛盾。共同抽样核意味着，对于每个N，都有 $\mathcal{L}_{Q_a}(O_N) = \mathrm{K}_N(P) = \mathcal{L}_{Q_b}(O_N)$。当辅助随机化在两个模型下具有相同的条件分布时，这一等式也适用于随机化检验。

□

定理 S9.2 (不存在诚实的预言机自适应)。 固定 $0 < \alpha < 1$ 并保留共同抽样核条件。假设 C_N 是一个观测数据可测的随机非空闭子集，取累积分布函数上确界范数下由豪斯多夫距离诱导的博雷尔 σ-代数，并且对 $\mathfrak{M}_{S'}$ 上的结构目标是均匀诚实的，即

$$\liminf_{N\to\infty} \inf_{Q\in\mathfrak{M}_{S'}} Q\{\theta(Q) \in C_N\} \geq 1 - \alpha.$$

另假设，在一个观测等价的 $Q_a \in \mathfrak{M}_S$ 处，

$$d_H(C_N, \mathcal{F}_S(P)) \to_p 0.$$

若 $\mathcal{F}_S$（P） $\subsetneq \mathcal{F}_{S'}$（$P$），这些要求便互不相容。

证明。 选取 $F^b \in \mathcal{F}_{S'}(P) \setminus \mathcal{F}_S(P)$。 $\mathcal{F}_S(P)$的紧致性意味着

$$\delta = d_\infty(F^b, \mathcal{F}_S(P)) > 0.$$

构造一个观测等价的 Q_b，使其满足 $\theta(Q_b) = F^b$。令

$$E_N = \{d_H(C_N, \mathcal{F}_S(P)) < \delta/2\}.$$

如果 E_N成立且 $F^b \in C_N$，那么

$$\delta = d_\infty(F^b, \mathcal{F}_S(P)) \le \sup_{F \in C_N} d_\infty(F, \mathcal{F}_S(P)) < \delta/2,$$

矛盾。根据可测性和共同抽样核，$Q_b(E_N^c) = Q_a(E_N^c) \to 0$。因此，

$$Q_b\{\theta(Q_b) \in C_N\} \le Q_b(E_N^c) \longrightarrow 0.$$

由于 $Q_b \in \mathfrak{M}_{S'}$，诚实性下确界在每个 N 处都至多为该数量，并且其下极限不可能至少为 $1 - \alpha > 0$。 □

S10 模拟

研究设计。所有研究设计均包含六个有序单元，第一个单元为“无收入”，组别–日期样本独立地从多项分布中抽取。在设计 1 中，政策后比较分布为 c =(.10, .20, .25, .20, .15, .10)，三个历史日期的处理组分布分别为 (.55, .25, .12, .05, .02, .01)、(.45, .28, .15, .07, .03, .02)、(.40, .30, .16, .08, .04, .02)，比较组分布分别为 (.10, .20, .25, .20, .10, .15)、 c 和 (.12, .22, .25, .19, .14, .08)。第一个顶点在顶部单元中的质量为 -0.04。真实未处理分布是权重为 (0.2, 0.5, 0.3) 的相容混合分布，基准样本从该分布中抽取。目标是无收入的反事实份额，其相对于 $[0.380, 0.550]$ 的尖锐区间为 $[0.380, 0.483]$，针对带符号凸包而言。设计 2 将基准分布平移 γ(0, -1, 0, 0, 1, 0)，其中 $\gamma \in \{0.05, 0.10, 0.15\}$，使其位于表 S2 所报告的、相对于凸包的累积分布函数距离处。在设计 3 中，两个顶点位于直线 $q + t \cdot 0.1$ $(e_1 - e_6)$ 上，其中 q =(.40, .20, .15, .13, .10, .02)， e_k 为第 k 个单位向量；它们由有效的处理组分布和比较分布生成。当较近的顶点位于 $t = 0.2$ 时，相容集合为单点；当其位于 $t = 0.2 + 10d$ 时，在总体修正d 的情况下相容集合为空。

程序。 未学生化相容性检验使用 $\max_{s,\,k} |\widehat{z}^*_{s,\,k} - \widehat{z}_{s,\,k}|$的自助法分位数；DKW 检验使用有限样本单元置信带 $2\varepsilon_N^{\mathrm{DKW}}$。学生化检验以及零收入份额的置信域见第6节：该置信域在满足学生化单元约束的权重范围内对份额进行优化，并加入顶点份额的学生化置信带，所有单元和份额共用一个临界值。表中的带符号区域是在有符号凸包区间上加入未学生化置信带。基准检验

与命题 8相同，其中分布函数在单元边界处求值。

表 S2：蒙特卡洛结果（1,000 次重复模拟，299 次自助抽样）

$(N_1,\ N_0,\ N_R)$	$(200, 2000, 300)$	$(1000,\ 5000,\ 1000)$	$(5000,\ 20000,\ 5000)$
设计 *1*：迁移成立；总体界限 $[0.380, 0.550]$ 有符号，$[0.380, 0.483]$ 相容			
拒绝：学生化相干性检验	0.000	0.000	0.000
拒绝：基准检验，自助法 / DKW	0.000 / 0.000	0.000 / 0.000	0.000 / 0.000
识别区间的覆盖率	0.999	1.000	1.000
中位宽度：相容 / 带符号区域	0.375 / 0.345	0.239 / 0.250	0.166 / 0.206
设计*2*：距凸包的累积分布函数距离为 D 处的基准；拒绝率，自助法 / *DKW*			
$D = 0.031$	0.002 / 0.000	0.010 / 0.000	0.852 / 0.000
$D = 0.068$	0.098 / 0.000	0.845 / 0.000	1.000 / 0.342
$D = 0.118$	0.622 / 0.000	1.000 / 0.024	1.000 / 1.000
设计 *3*：总体修正 d；拒绝率，代入规则 / 未学生化 / 学生化			
$d = 0.000$	0.50 / 0.00 / 0.01	0.48 / 0.00 / 0.01	0.50 / 0.00 / 0.01
$d = 0.005$	0.64 / 0.00 / 0.03	0.79 / 0.00 / 0.04	0.96 / 0.00 / 0.17
$d = 0.010$	0.78 / 0.00 / 0.06	0.96 / 0.00 / 0.17	1.00 / 0.00 / 0.81
$d = 0.020$	0.95 / 0.00 / 0.19	1.00 / 0.00 / 0.69	1.00 / 0.55 / 1.00
$d = 0.040$	0.99 / 0.00 / 0.64	1.00 / 0.37 / 1.00	1.00 / 1.00 / 1.00

注：N_1 和 N_0 分别是每个处理组和对照组—日期样本的样本量，N_R 是基准样本的样本量。名义水平为 5%。当估计的相容集合为空时，代入规则会拒绝原假设。在这些研究设计中，DKW 相容性检验从未拒绝，因此省略。在设计 3 中，$d = 0$总体相容集是一个单点。

研究发现。 所有程序都能控制检验尺寸。朴素的代入规则则不能：当相容集合是一个点时，约有一半样本中的集合为空。正如原假设包含整个相容区域的检验所预期的那样，相容性检验在边界处较为保守。学生化赋予相容性检验功效，因为不可行性出现在稀有单元中，而这些单元的单元质量估计得很精确；未学生化置信带则受变异性最大的单元所主导。对于样本量与 NSW 相当的样本，相容置信区域略宽于带符号区域，因为多个学生化单元的联合临界值高于单个得分的临界值；样本量增大后，相容性会使置信区域变得更尖锐。处理单元达到一千时，基准检验对于累积分布函数距离偏离 5 至 10 个点的情况具有良好功效。有限样本 DKW 校准虽有效，但保守得多。

学生化带的有效性。令 $\sigma_{s,\,k} > 0$，并假设以至少 $1 - \alpha$ 的概率，$|\widehat{z}_{s,\,k} - z_{s,\,k}| \le c\,\sigma_{s,\,k}$ 对所有 s、k 成立。如果 $\lambda \in \Lambda_{S,\,\Pi}$，那么在该事件上，对于每个 k，都有 $\widehat{z}_{\lambda,\,k} \ge z_{\lambda,\,k} - \sum_s \lambda_s |\widehat{z}_{s,\,k} - z_{s,\,k}| \ge -c \sum_s \lambda_s \sigma_{s,\,k}$。因此，集合 $\{\lambda :\ \widehat{z}_{\lambda,\,k} + c \sum_s \lambda_s \sigma_{s,\,k} \ge 0\ \forall k\}$ 在该事件上包含 $\Lambda_{S,\,\Pi}$，而在集合为空时拒绝的检验，其检验尺寸至多为 α。使用估计的 $\sigma_{s,\,k}$ 以及针对最大学生化误差的自助法临界值时，有效性

是渐近的，并要求单元方差与零保持一定距离，或像我们的实现一样设置下限。

S10.1 异质权重下的合并

该实验处于总体层面：不进行抽样。三个事件的权重相等，每个事件有六个结果单元和两个历史日期。对于事件 e，我们从 Δ_2 上的均匀分布中抽取真实权重λ_e，并从 Δ_6上的均匀分布中抽取真实未处理分布 μ_e，然后将两个顶点设为 $\mu_e + 0.15\, d_{e,\ s}/$ 最大值 $|d_e|$，其中 $d_{e,\ s}$是经过中心化、使总质量为零的高斯方向，随后进行平移，使得 $\sum_s \lambda_{e,\ s} d_{e,\ s} = 0$。因此，每个事件的真实反事实都是其顶点的 λ_e 混合分布，并且是一个概率律，而单个顶点可以带符号：事件特定模型成立，而共同权重模型通常不成立。目标是各事件反事实概率中前两个单元的平均值。在 2,000 个设计中：

模型	包含真实值	空	不包含真实值
事件特定权重	1.00	0.00	0.00
堆叠	0.71	0.00	0.29
逐事件共同权重	0.28	0.25	0.47

非空界限的中位宽度分别为 0.077（事件特定权重）、0.059（堆叠）和 0.015（逐事件）。堆叠通过对合并历史数据施加共同权重，得到了比模型设定正确时更窄的界限；而在本研究设计中，这一限制有 29% 的情况下并不成立，并导致界限偏离真实值。

S11 应用I：数据与补充结果

S11.1 样本

这些样本来自 Imbens and Xu （2025年）随复现档案分发的数据：NSW 男性的 Dehejia–Wahba 子样本（185 名参与者、260 名对照者）、CPS-1 对照样本（15,992 名男性）和 PSID-1 对照样本（2,490 名男性）。这些样本包含与 Rajeev Dehejia 分发的文件相同的记录。收入以 1982 年美元计，CPS 收入的顶端编码为 $25,862。表 S3 给出了七单元格网格上的单元计数，这些计数足以在不使用微观数据的情况下复现所有七单元格计算。

参与者与调查样本的平均特征存在显著差异。参与者的平均年龄为25.8岁，平均受教育年限为10.4年；其中84%为黑人，19%已婚，71%没有高中学历。相应的数据为

表 S3：NSW 和对照样本按收入单元格划分的数量

	参与者			实验控制组			CPS-1			PSID-1		
收入（千美元）	74	75	78	74	75	78	74	75	78	74	75	78
0	131	111	45	195	178	92	1913	1748	2172	215	249	286
$(0, 5]$	25	54	56	28	57	70	2300	2483	1649	132	125	97
$(5, 10]$	15	15	43	21	17	56	1682	1778	1606	246	252	179
$(10, 15]$	10	3	23	5	5	29	2054	2209	1938	362	362	300
$(15, 20]$	1	1	9	4	2	9	2323	2378	2096	447	424	323
$(20, 30]$	2	1	6	4	1	3	5720	5396	6531	660	654	730
> 30	1	0	3	3	0	1	0	0	0	428	424	575

CPS-1中分别为33.2、12.0、7、71%和30%，PSID-1中分别为34.9、12.1、25、87%和31%。实验对照组与参与者相似（25.1、10.1、83%、15%和83%）。

S11.2 网格、结果与重新加权

表 S4 报告了三个网格上的代入式修复量，以及基准与带符号凸包之间的距离：正文中的七单元格网格、一个包含\$2,500 步至\$20,000 且单元数为 $(20, 25]$，$(25, 30]$ 及以上 30 的 12 单元网格，以及一个包含 \$1,000 步至 \$20,000 且单元数为 $(20, 30]$ 及以上 30 的 23 单元网格。12 单元网格和 23 单元网格并非嵌套，因此命题 3 无法对它们排序。对于“收入高于 \$5,000”这一结果变量，实验效应为 0.077；带符号凸包区间为 $[0.27, 0.32]$ (CPS-1) 和 $[0.31, 0.35]$ (PSID-1)，重新加权后为$[0.20, 0.21]$ 和 $[0.22, 0.32]$，在七单元格网格上的相容区间为 $[0.200, 0.204]$ 和 $[0.302, 0.304]$。

表 S4：NSW：按网格列出的最小修复和基准距离

	最低工资修复（百分点）			与基准的全变差距离		
单元	7	12	23	7	12	23
CPS-1	0.88	1.27	1.16	0.338	0.371	0.395
CPS-1 重新加权	0	0.93	3.86	0.287	0.293	0.356
PSID-1	3.73	4.33	4.64	0.351	0.407	0.432
PSID-1 重新加权	0	4.55	8.20	0.319	0.433	0.447

重新加权样本对每位对照组男性赋予权重 $\hat{p}(x)/\{1 - \hat{p}(x)\}$，并将权重标准化为均值为 1；其中 $\hat{p}$是参与度对年龄、年龄平方、教育、教育平方，以及黑人、西班牙裔、已婚和无学位示性函数进行逻辑回归得到的预测值，该回归在参与者和对照样本上估计。这些权重以参与者的特征分布为目标。我们有意排除项目前收入：以此为条件会机械地缩小本方法旨在利用的历史偏差。加权后的有效样本量分别为 431（CPS-1）和 145（PSID-1）。

S11.3 推断细节

DKW 临界值使用 $\alpha = 0.05$、S 节 S5.3 中列出的六个基本累积分布函数，以及实际样本量。自助法在参与者样本、实验组—对照组样本和比较样本中分别进行有放回的个体重抽样，并将每个人 1974年、1975 和 1978年 的收入数据作为整体保留；共进行 2,000 次抽样（每次抽样中重新估计权重时进行 999 次），随机种子为 20261121。一致性筛选和就业效应的外部区域使用联合带，该联合带通过邦费罗尼划分将 α 分配给顶点和单元上的最大单元误差（CPS-1 和 PSID-1 的自助法分位数均为 0.087）以及处理组就业份额的误差（0.070）。由于顶点单元误差在各单元间的和为零，顶点的就业份额误差至多为单元置信带的宽度。就业效应的实验置信区间采用两个独立比例之差的正态近似。

S11.4 无网格置信区间

令 $x_1 < \cdots < x_n$ 为合并后的不同收入值，H_s 为估计的顶点分布函数，c 为在每个样本内对完整历史进行{重抽样所得的最大$_s \, \|H_s^* - H_s\|_\infty$、$|\hat{g}^* - \hat{g}|$}的自助法 95% 分位数。在置信带覆盖的事件上，每个 $\lambda \in \Lambda_S(P)$ 都有一个取值于 $[0,1]$, 的非递减 H_λ，因此对于所有 $u < w$ 和 $-c \le H_\lambda \le 1+c$，$H_\lambda(w) - H_\lambda(u) \ge -2c$；也就是说，$\lambda \in \Lambda_\infty(c)$。命题 5 的每个区间单元约束都是这些不等式之一，因此对于每个区间划分 Π，$\Lambda_\infty(c) \subseteq \Lambda_\Pi(c)$。

精确计算。 H_λ 在相邻的 x_i 之间为常数，其左极限等于前一个点处的取值，因此只需在 $x_1, \ldots, x_n$处施加条件。引入$m_1, \ldots, m_n$ 以及约束

$$m_{i-1} \le m_i, \qquad \widehat{H}_\lambda(x_i) \le m_i, \qquad m_i - \widehat{H}_\lambda(x_i) \le 2c, \qquad -c \le \widehat{H}_\lambda(x_i) \le 1 + c.$$

若 $\lambda \in \Lambda_\infty(c)$，则运行最大值 $m_i = \max_{j \le i} H_\lambda(x_j)$满足这些条件。反之，任意可行 m都不小于运行最大值，因此 $\max_{j \le i} H_\lambda(x_j) - H_\lambda(x_i) \le m_i - H_\lambda(x_i) \le 2c$.将此多面体投影到 λ上，所得结果恰为 $\Lambda_\infty(c)$；因此，任意线性对比的区域端点可通过两个线性规划求得，其中包含 $|S| + n$ 个变量和 $5n - 1$个不等式。我们将每个右端项都扩大 10^{-9}后求解，因此计算得到的集合包含精确集合。

结果。 就业效应的区域为 $[\hat{g} - \text{最大值}\, \hat{e}_\lambda - 2c, \;\; \hat{g} - \text{最小值}\, \hat{e}_\lambda + 2c]$ 在 $\lambda \in \Lambda_\infty(c)$上，其中 $\hat{e}_\lambda = 1 - H_\lambda(0)$；它以渐近概率覆盖识别区间

至少为 0.95。对于 CPS-1（$n = 22$，339），$c = 0.092$，1974年上的可容许权重构成 $[0.108, 0.485]$，且该区域为 $[0.209, 0.615]$。范围条件排除了接近 1 的权重：在 CPS 顶端编码附近，1974年顶点超过 1 的幅度大于 c。在具有相同置信带的 7、12 和 23 单元网格上，该区域为 $[0.199, 0.666]$。对于 PSID-1，$c = 0.090$，所有权重均可容许，且在每个网格上以及不使用网格时，该区域为 $[0.191, 0.674]$ 。表 3 中的外部区域有所不同，因为其置信带是根据七个单元质量而非整个分布函数校准的。它们都排除了 0.111 的实验估计值。

S11.5 正向对照：按研究设计进行迁移

为检验当假设 2成立时，检验不会拒绝，我们将 260 名实验对照组随机分为两半，每半 130 人，一半作为对照组，另一半作为基准，并将 185 名参与者作为处理组。随机化使参与者与对照组在未处理状态下的收入历史具有相同分布，因此总体中的每一项历史偏差均为零，迁移也完全成立。在 200 次随机划分中，每次进行 199 次自助抽样，命题8 的基准检验在 1 次划分中于 5% 显著性水平拒绝（0.5%），学生化相干性检验在所有划分中均未拒绝，而七单元格网格上的插件相干集合在 94% 的划分中非空。基准组与最近的迁移后分布之间的中位累积分布函数距离为 0.10，而使用调查比较组时为 0.32 和 0.35。因此，第7节中的拒绝结果并非由这些小样本中的稀疏单元格造成。与模拟结果一样，两项检验都较为保守。

S11.6 学生化检验的敏感性

学生化检验将每个顶点单元误差除以其自助法标准误，并设置下限以避免除以零。对于 CPS-1，一个单元（收入超过 30,000 美元）具有零自助法方差：没有 CPS-1 收入超过顶端编码，因此 1975 年顶点在每次抽样中该处的质量都恰好为零。对于 PSID-1，最小的自助法标准误为 0.007。将下限从 10^{-6} 调整到 5×10^{-3}，两个样本中的 p 值均保持不变。使用不同的自助抽样种子时，PSID-1 p 值为 0.008，而非 0.013；两个样本中的 CPS-1 p 值均为 0.84。

S11.7 Wasserstein分解

在七单元格网格上，令 $x_1 < \cdots < x_7$ 表示单元代表值（0、2,500、7,500、12,500、17,500、25,000 美元，以及对于开放顶端单元的 35,000 美元）。对于概率向量 q 和单位质量有符号向量 z，定义 $W_1(q, z) = \sum_{k=1}^{6} |Q_k - Z_k|(x_{k+1} - x_k)$，其中 Q_k 和 Z_k

是累积和；对于两个概率向量，这就是相应离散分布之间的 Wasserstein 距离。第 5 节中的最低修复、符号分解半径和实际误差线性规划，在 W_1 替代全变差的情况下定义，并为绝对值设置辅助变量。对于 CPS-1，最低修复为 $44；就业效应的符号在 $1,053 处失效，而实验对照分布与历史凸包的距离为$2,630。对于 PSID-1，这三个数分别为 $ 260、$1,236 和 $3,793。在全变差下，相应的半径为 0.38 和 0.37，而实际误差为 0.34 和 0.35。因此，这两种距离对于实验所揭示的偏离是否会使符号保持不变，给出了不同结论：在全变差下，将质量从零收入转移到高收入的代价较低，而在 W_1下代价较高；NSW 的偏离主要由此类转移构成。W_1 这些数值取决于为开放顶端单元选择的代表值，该代表值只通过总和中的最后一项发挥作用。

S11.8 其他使用项目实施前差距的估计量

以单一年份为基准的分布差分中的差分 s 使用顶点 H_s，因此其就业效应是有符号区间的一个端点：CPS-1 在 1974年和 1975 年分别为 0.481 和 0.383，PSID-1 分别为 0.493 和 0.372。对于变化中的变化，我们将Athey 和 Imbens (2006，定理4.1)针对离散结果的界限应用于零收入的反事实概率，其中对照组作为控制组，基准年作为第一期。由于对照组基准年份分布在相关水平处的分位数落在该分布连续的位置，下界和上界重合。由此得出的就业效应为：CPS-1 以 1974 年为基期时为 0.465，以 1975 年为基期时为 0.405；PSID-1 分别为 0.568 和 0.508。

S11.9 学生化相干性检验

该检验使用七单元格网格。对于每个顶点和单元，$\sigma_{s,\,k}$ 是$\widehat{z}_{s\text{的自助法标准差},\,k}$（下限为 10^{-4}），根据参与者样本和对照样本中对完整历史记录进行重抽样的2,000次抽样计算得出。该统计量是满足以下条件的最小 c，其中某个 λ 满足$\widehat{z}_{\lambda,\,k} + c\sum_s \lambda_s \sigma_{s,\,k} \geq 0$ 在每个单元中均成立；对于 CPS-1 和 PSID-1，该统计量分别为 1.35 和 3.37，而自助法95%临界值分别为 3.00 和 2.93，从而得到表 p3中的数值。

S11.10 AFDC 女性

女性样本来自Cal´onico 和 Smith（2017年），包含600名参与者，以及来自AFDC目标组的585名实验对照组成员；这些数据包括1975年和1979年的收入，以及648

名PSID女性作为对照组。 $S = \{1975\}$ 由于凸包只有一个顶点，历史迁移便简化为分布差分中的差分。在七单元格网格上，该顶点的负质量为0.006；其隐含的就业效应为0.269，而实验值为0.012；与实验对照分布的全变差距离为0.29；$\widehat{}$累积分布函数距离为 $D = 0.282$，而自助法和DKW临界值分别为0.081和0.256。迁移误差的95%置信下界为0.20（自助法）和0.03（DKW）。与男性一样，其机制在项目实施之前就已显现：参与者中有73%没有1975年收入，而PSID女性中这一比例为28%；相比之下，1979年实验对照组与PSID女性之间的差距为0.20。鉴于长期福利领取是一种比当前失业更持久的状态，人们原本可能预期基于长期福利领取的选择会有利于迁移；但事实并非如此。

S11.11 针对已知误差的失效分析

对于 CPS-1，松弛模型在 $\tau \geq 0.009$ 时非空，在 $\tau = 0.383$ 时允许就业效应为零；从实验对照分布到带符号凸包的全变差，即实际迁移误差，为 0。338。在 $\tau = 0.338$ 处，就业效应区间为 $[0.046, 0.757]$。对于 PSID-1，这三个数值分别为 0。037、0。372 和 0。351；在实际误差处的区间为 $[0.020, 0.757]$。上端点，即处理组就业份额，可以达到，这是因为松弛允许每位参与者在没有该项目时都处于非就业状态。根据命题 7 ，采用联合界 $\max_s d_{\mathrm{TV}}(\widehat{z}^*_s,\ \widehat{z}_s) + |\widehat{g}^* - \widehat{g}|$，CPS-1 半径的 95% 自助法置信区间为 $[0.26, 0.51]$；PSID-1 的区间为 $[0.24, 0.51]$。

S12 应用二：劳动所得税抵免扩展，详情

S12.1 数据与预先设定

这些数据是1991—1996纳税年度3月当前人口调查记录的公开摘录，广泛用于教学中讲解EissaandLiebman (1996)的差异中的差异比较。摘录包含13,746名年龄为20至54岁、受教育年限不超过11年的女性；发布者称她们未婚，但我们无法根据原始文件核实这一描述，而且摘录不包含调查权重、家庭标识符或设计变量。所用变量包括：纳税年度、子女数量、年度收入，以及第S12.4节中的居住州。收入以1997年美元计：将每笔记录金额除以1997年与相应纳税年度消费者价格指数（CPI-U年度平均值）之比，即可还原为整数美元名义金额。州采用CPS的人口普查州代码记录（从缅因州的11到夏威夷州的95）；我们将代码映射到各州，并根据摘录所含的州失业率核实映射（例如，1992年

加利福尼亚州的失业率为9.3%）。摘录中的工作状态指标与收入为正的情况并不一致（1,351名收入为正的女性被编码为未工作），因此我们全程使用收入。该文件的SHA-256 哈希值列于复现包中。

在按组和年份计算任何结果变量之前，我们已在一份注明日期的计划中记录研究设计，复现包中也重现了该计划：各组、历史年份、结果年份、五单元网格、21 单元次级网格（从零开始，之后按 2,500 美元分箱，直到 47,500 美元，另有更高区间）、三个对比、诊断指标、对照组以及成功标准。首次估计完成后，我们又增补了三项内容，并在此明确说明：第2.2节的对照组环境检验，这是正文方案所要求的；基于顶点约化、针对每个年份计算的自助法外部区域；以及针对 1994年和1995年21 单元网格进行的学生化检验。这些内容均不改变任何界限。

S12.2 两个或两个以上子女的母亲

	1994	1995	1996
就业效应，界限	$[-0.025, 0.012]$	$[0.049, 0.086]$	$[0.134, 0.171]$
95%外部区域	$[-0.09, 0.08]$	$[-0.03, 0.16]$	$[0.06, 0.24]$
无网格95%区域	$[-0.18, 0.17]$	$[-0.11, 0.24]$	$[-0.02, 0.33]$
逐步退出份额效应	$[-0.019, -0.003]$	$[-0.004, 0.011]$	$[0.062, 0.077]$
逐步实施及稳定阶段占比效应	$[-0.014, 0.001]$	$[0.031, 0.046]$	$[0.045, 0.059]$
符号分解半径，就业	0	0.049	0.134
比较法检验， p-值	0.78	0.39	0.01

学生化相干性检验在任何年份的两个网格上均未拒绝（$p = 1.00$），而1993年的安慰剂基准显示，累积分布函数距离为0.036，自助法临界值为0.089（$p = 0.94$）。对于所有母亲，逐步实施及稳定阶段占比效应在1994年、1995年和1996年分别为 $[-0.005, 0.012]$、$[0.018, 0.035]$ 和 $[0.014, 0.031]$ 。

S12.3 敏感性分析与比较

将1993年加入作为第三个历史年份后，1996年的就业界限保持不变（$[0.115, 0.139]$ 对于所有母亲， $[0.134, 0.171]$ 对于有两个或更多子女的母亲），逐步实施及稳定阶段占比效应的范围扩大至 $[0.014, 0.043]$ 和 $[0.045, 0.093]$，一致性检验仍为 $p = 1.00$。1996年就业的单基准分布差分中的差分效应，对于所有母亲而言，分别为0.115（以1992年为基准）和0.139（以1991年为基准），即传输界限的两个端点；以1992年为基准的变化中的变化结果为 $[0.153, 0.164]$，对于有两个或更多子女的母亲则为 $[0.174, 0.182]$。

在21 单元网格上，代入式修复量对于所有母亲在1994–1996年分别为0.003、0.001和0；对于有两个或更多子女的母亲分别为0.005、0.003和0.003；学生化检验在这些情况下均不拒绝，而且在五单元格网格上，三个结果年份之间存在共同权重是可行的，但在代入估计中，在21 单元网格上则不可行。

自助法在组和年份内对女性进行重复抽样，并将记录视为独立抽取。对于独立的组–年份样本，每个顶点都是独立经验分布函数之和；按照第 S6 节逐个样本应用的论证，样本内自助法是一致的。CPS 家庭是在州内按簇抽样的；第 S12.4 节则改为对完整州进行重复抽样，结果几乎相同。

S12.4 州、聚类与福利豁免

表 S5 将表 5 扩展至1994年和有两个或两个以上子女的母亲，并比较对完整州进行重复抽样的区域与对女性进行重复抽样的区域。州级分析是在修订时加入的，当时表4的结果已经揭晓；豁免分类取自已发表的表格，且在按州计算任何结果变量之前确定。每种研究设计下的界限均根据保留州的合并记录计算，因此目标总体是这些州的人口，并按其样本量加权。州级自助法有放回地抽取州，并保留所抽中州的全部记录，因此州内跨年份、组和家庭的任何依赖关系都会得到保留。其依据是第 S8.2节的聚类理论应用于比率估计量（见命题 S8.2之前的段落）；州数为15时，该方法是近似的。

有三个特点值得注意。第一，在任何研究设计、组或年份中，以州而非女性为单位进行重抽样，都不会使外部区域的任何端点移动超过 0.035，因此在这段摘录中，州内依赖关系并非实质性问题。第二，剔除豁免州后，相容性约束变得有效。在 24 州设计中，1996 年就业效应的带符号界限为$[0.084, 0.131]$（针对所有母亲），针对有两个或更多子女的母亲则为 $[0.098, 0.148]$；相容性将上界分别压低至 0.118 和 0.125。在 15 州设计中，它将有两个或更多子女的母亲在 1995 年的界限上界从 0.075 压低至 0.049。在每种情况下，历史凸包中的一些权重都会给某个单元赋予负质量，因此被剔除。第三，在 24 州设计中，1994 年的界限也为正，即 $[0.056, 0.103]$；而在完整样本中并非如此。因此，在这些州，就业增幅似乎在扩张的第一年便已出现；其外部区域 $[-0.04, 0.19]$包含零。

命题 S8.2 的有限样本共享簇半径在此毫无用处：当 $J = 2$、$K = 5$ 以及 15 个州时，该半径为 1.34，超过任何可能的亏损。等权重的州级平均研究设计也会改变目标，并依赖于最少仅有一名女性的州–年份–群体单元（612 个单元中有 18 个包含一或两名女性）。因此，我们保留合并目标和州级自助法，并将豁免设计作为敏感性分析：批准日期

表 S5：福利豁免研究设计中的 EITC 应用

	1994	1995	1996
所有母亲			
所有州（*51*）			
就业效应，界限	$[-0.010, 0.014]$	$[0.047, 0.070]$	$[0.115, 0.139]$
95%外部区域，州重抽样	$[-0.08, 0.08]$	$[-0.01, 0.13]$	$[0.05, 0.20]$
95%外部区域，女性重抽样	$[-0.07, 0.08]$	$[-0.01, 0.13]$	$[0.05, 0.21]$
逐步退出份额效应，界限	$[-0.031, -0.008]$	$[-0.009, 0.014]$	$[0.053, 0.076]$
比较法检验，p-值	0.70	0.38	0.02
1993年安慰剂基准，p-值		0.92	
截至*1995*年未批准豁免（*24*）			
就业效应，界限	$[0.056, 0.103]$	$[0.055, 0.102]$	$[0.084, 0.118]$
95%外部区域，州重抽样	$[-0.04, 0.19]$	$[-0.04, 0.20]$	$[0.01, 0.20]$
95%外部区域，女性重抽样	$[-0.04, 0.19]$	$[-0.04, 0.20]$	$[-0.01, 0.22]$
逐步退出份额效应，界限	$[0.023, 0.029]$	$[0.009, 0.014]$	$[0.044, 0.048]$
比较法检验，p-值	0.48	0.61	0.20
1993年安慰剂基准，p-值		0.37	
截至*1996*年未批准豁免 *(15)*			
就业效应，界限	$[0.000, 0.089]$	$[-0.010, 0.079]$	$[0.031, 0.120]$
95%外部区域，州重抽样	$[-0.15, 0.23]$	$[-0.16, 0.22]$	$[-0.06, 0.21]$
95%外部区域，女性重抽样	$[-0.12, 0.21]$	$[-0.13, 0.20]$	$[-0.08, 0.23]$
逐步退出份额效应，界限	$[-0.004, 0.015]$	$[-0.038, -0.019]$	$[0.023, 0.041]$
比较法检验，p-值	0.79	0.93	0.56
1993年安慰剂基准，p-值		0.50	
有两个或两个以上子女的母亲			
所有州（*51*）			
就业效应，界限	$[-0.025, 0.012]$	$[0.049, 0.086]$	$[0.134, 0.171]$
95%外部区域，州重抽样	$[-0.10, 0.09]$	$[-0.02, 0.15]$	$[0.06, 0.25]$
95%外部区域，女性重抽样	$[-0.09, 0.08]$	$[-0.02, 0.16]$	$[0.06, 0.24]$
逐步退出份额效应，界限	$[-0.019, -0.003]$	$[-0.004, 0.011]$	$[0.062, 0.077]$
比较法检验，p-值	0.71	0.39	0.03
1993年安慰剂基准，p-值		0.90	
截至*1995*年未批准豁免 *(24)*			
就业效应，界限	$[0.058, 0.108]$	$[0.058, 0.108]$	$[0.098, 0.125]$
95%外部区域，州重抽样	$[-0.05, 0.22]$	$[-0.04, 0.21]$	$[0.00, 0.24]$
95%外部区域，女性重抽样	$[-0.04, 0.21]$	$[-0.05, 0.21]$	$[-0.01, 0.26]$
逐步退出份额效应，界限	$[0.035, 0.047]$	$[0.015, 0.027]$	$[0.048, 0.054]$
比较法检验，p值	0.46	0.58	0.20
1993年安慰剂基准，p值		0.59	
截至*1996*年未批准豁免（*15*）			
就业效应，界限	$[-0.001, 0.089]$	$[-0.015, 0.049]$	$[0.045, 0.135]$
95%外部区域，州重抽样	$[-0.16, 0.25]$	$[-0.16, 0.22]$	$[-0.09, 0.27]$
95%外部区域，女性重抽样	$[-0.13, 0.22]$	$[-0.15, 0.21]$	$[-0.09, 0.27]$
逐步退出份额效应，界限	$[-0.002, 0.006]$	$[-0.041, -0.035]$	$[0.016, 0.025]$
比较法检验，p-值	0.78	0.92	0.58
1993年安慰剂基准，p-值		0.62	

注：设计剔除了截至所示年份首次获批全州性AFDC豁免的州（Schanzenbach 和 Strain，2021年，表A3）；括号内为保留的州数。界限基于包含1991年和1992年这两个历史年份的五单元格网格；界限施加了相容性约束，除以下情况外，均与带符号界限一致：24州设计下1996年的所有母亲和有两个或更多子女的母亲，以及15州设计下1995年的有两个或更多子女的母亲（见正文）。外部区域采用顶点约化，并针对两个顶点效应构造联合自助法带；重抽样时抽取完整州（所有年份和两组）或组别与年份内的女性。比较法检验和安慰剂基准通过州重抽样。学生化相干性检验在任何设计、组别或年份下均未拒绝（$p = 1.00$）。

并非政策实施日期；此外，若干州也有自己的税收抵免，而剔除这些州会改变总体。

S13 应用 III：数据、稳健性与审计细节

S13.1 数据与时间安排

面板数据包含 $51 \times 36 \times 118 = 216$ 条2007—2015 年的州—季度—单元记录。每个州—季度内，117 个正工资单元按以 2016 年美元计的时薪四分之一美元区间统计人均劳动者数量（区间标签为下边缘），残差单元等于 1 减去总和。就业人数以季度就业与工资普查（QCEW）的就业人数为基准，并除以常住人口，因此哥伦比亚特区每个季度的残差都为负（介于-0.45 和 -0.33之间）；哥伦比亚特区不纳入任何对照组。使用Cengiz 等人 (2019年) 的存档文件中的原始 CPS 处理输入重新构建面板数据，所得结果与每个已发布单元的差异均在 5.5×10^{-8}以内；所有计算均使用已发布的面板数据。

表 S6：2014 年的五起事件

州	颁布	生效	结果季度	样本池大小
新泽西州	2013年11月5日（投票）	2014年1月1日	2014年第一季度	34
特拉华州	2014年1月30日	2014年6月1日	2014年第三季度	30
加利福尼亚州	2013年9月25日	2014年7月1日	2014年第三季度	30
明尼苏达州	2014年4月14日	2014年8月1日	2014年第四季度	30
密歇根州	2014年5月27日	2014年9月1日	2014年第四季度	30

注：结果季度是完全处于新的最低工资标准生效期内的第一个日历季度。对照组规模是指从2010年第一季度到结果季度期间最低工资没有上调的对照州数量，不包括哥伦比亚特区。明尼苏达州的法律为大型和小型雇主规定了不同的最低工资标准。

对于事件 e 、结果季度 q_e、对照组 $\mathcal{C}_e$ 和接受处理的州 j_e，偏移量为 $h \in \{-16, -12, -8, -4\}$ 处的顶点为

$$z_{eh} = \overline{C}_e(q_e) + T_{j_e}(q_e + h) - \overline{C}_e(q_e + h),$$

其中，$T_j(t)$是州 j的118个单元格向量，$C_e(t)$是 $\mathcal{C}_e$的未加权平均值。堆叠顶点为 $z_h = \frac{1}{5}\sum_e z_{eh}$，堆叠处理向量为$g = \frac{1}{5}\sum_e T_{j_e}(q_e)$。

S13.2 从7个单元到118个单元格

七单元格网格由残差以及在以下位置截断的区间组成：\$7.25、\$9.25、\$11.25、\$14.25 和 \$18.25。表S7 通过反复在中点处分割包含最多美元区间的单元来细化该网格，

并列情形取最低工资的美元区间作为分割对象，同时将五个截点保留为边界。中间各行取决于分割规则；第一行和最后一行则不受影响。

表 S7：低于 9.25 美元界限的细化路径

单元	无效顶点	相干反事实占比（百分点）	效应（百分点）
7	0	$[7.363, 7.750]$	$[-0.614, -0.228]$
15	0	$[7.363, 7.750]$	$[-0.614, -0.228]$
30	1	$[7.363, 7.750]$	$[-0.614, -0.228]$
60	3	$[7.453, 7.719]$	$[-0.584, -0.317]$
118	4	$[7.459, 7.653]$	$[-0.517, -0.324]$

S13.3 逐事件相干性、匹配窗口与松弛

在完整网格上，堆叠凸包是相容的，但不存在能使全部五个事件向量 $\sum_h \lambda_h z_{eh}$ 均为非负的共同权重；各事件负质量的最小平均值为 0.00303。命题 S2.8 的检验得出负值 m_{ej}，涉及 $5{\times}118$ 个事件—单元对；每个对都给出一个与堆叠凸包相容、却违反逐事件相容性的权重；最负的值为-0.0065，对应加利福尼亚州的 \$7.75–\$8.00 区间。使用三个偏移量（-12，-8，-4）而非四个时，政策日期凸包再次相容，符号分解半径仍为 0.00288；提前一年安慰剂检验使用冻结在 2014年定义下的对照组计算，其最小修复在使用三个偏移量时为 1.19×10^{-5}，使用四个时为 5.93×10^{-6}。这些修复都很小；有信息量的安慰剂统计量是观测到的处理组分布与带符号凸包之间的全变差距离，2013 年为 0.0492，2012 年为 0.0473。低于 \$9.25 的效应的放宽界限，在 $[-0.70, -0.19]$ 时为 $\tau = 0.001$，在 $[-0.85, -0.04]$ 时为$\tau = 0.0025$。

S13.4 审计详情

表 S8汇总了论文第 9 节总结的审计结果，图 S1则展示了安慰剂分布。

括定。 对事件取平均后，偏移量 -16、 -12、 -8 和 -4 处的比较州非就业率分别为 44.41%、44.04%、44.08% 和 43.83%，结果季度为 43.42%；处理组与比较组差距分别为 0.61、0.60、0.64 和 0.30 个百分点。历史上，比较州低于 9.25 美元的就业占比分别为 8.02%、8.59%、8.57% 和 8.93%，结果季度为 8.82%，因此从这一维度看，政策日期处于括定范围内。2014年比较州概率律与其自身历史分布凸包之间，在 118个单元格上的全变差距离为 0.120；早一年时为 0.127；在这一

表 S8：低于 9.25 美元的界限审计（百分点）

	已签署 hull	相容 hull	逐事件修复	τ^{sign}	安慰剂距离
基线	$[-0.61, -0.23]$	$[-0.52, -0.32]$	0.30	0.29	4.92
颁布前的历史日期					
丢弃 -4	$[-0.57, -0.23]$	空	0.41	–	
丢弃 -4 和 -8	$[-0.57, -0.23]$	空	0.59	–	
聚合					
人口加权事件	$[-0.56,\ \ 0.20]$	$[-0.01,\ \ 0.08]$			
对照州					
从未处理，2010–2015	$[-0.54, -0.18]$	$[-0.42, -0.27]$	0.34	0.25	4.91
不与接受处理的州接壤	$[-0.72, -0.39]$	$[-0.65, -0.54]$	0.30	0.46	5.03
两项限制	$[-0.56, -0.25]$	$[-0.46, -0.39]$	0.35	0.30	5.05
对就业比例达到或超过 W 没有影响					
$W =$\$18.25		$[-0.51, -0.37]$			
$W =$\$15.25		$[-0.45, -0.39]$			

注：“剔除 -4”是指剔除结果变量前一年的历史季度；加利福尼亚州的上调在该季度签署（2013年9月25日），新泽西州的修正案已获州议会通过。人口权重为处理组州在结果季度的总人口（加利福尼亚州占60%）。“从未处理”排除2010–2015年间有任何上调的州，因此新泽西州的比较组中不包括加利福尼亚州、特拉华州、密歇根州和明尼苏达州；剩下23个对照州。“不接壤”排除与处理组州相邻的州。最后几行施加 $\mu_\lambda(\{w \geq W\}) = \mu_{1,\ 1}(\{w \geq W\})$。

分辨率下，甚至对照州的分布也超出了其历史范围，部分原因是名义工资漂移，跨越实际四分之一美元区间。

人口增长。 在结果季度之前的四年里，密歇根州的人口增长了1.0%，新泽西州增长了3.2%，明尼苏达州增长了3.8%，特拉华州增长了5.6%，加利福尼亚州增长了5.9%，而各比较组的平均增幅为4.4%。结果变量为人均值，因此人口增长差异只会通过构成产生影响，而公共面板数据无法衡量这一点。

上尾部。 在政策日期，就业人数中收入达到或超过 \$18.25 的处理组占比为 28.67%；相干凸包的值域为 $[28.14, 28.98]$，带符号凸包的值域为 $[28.00, 29.34]$。收入达到或超过 \$15.25 时，相应数字为 33.94、 $[33.59, 34.25]$ 和 $[33.29, 34.49]$。在 2013 年安慰剂检验中，观测到的占比 27.97% 和 33.61% 均低于带符号凸包的值域 $[28.48, 29.38]$ 和 $[33.97, 34.81]$。

对照组。 从未接受处理的州构成的对照组包含 23 个在 2010–2015 年间没有增幅的辖区。剔除与任何处理组州接壤的州（俄勒冈州、内华达州、亚利桑那州、马里兰州、宾夕法尼亚州、纽约州、俄亥俄州、印第安纳州、威斯康星州、爱荷华州、南达科他州和北达科他州，以及处理组州本身）后，按基准规则剩余 23 个州；与从未接受处理的州规则结合时，则剩余 18 个州。

不确定性 i 在包夹检验中。有放回地从每个事件的样本池中重抽样对照州

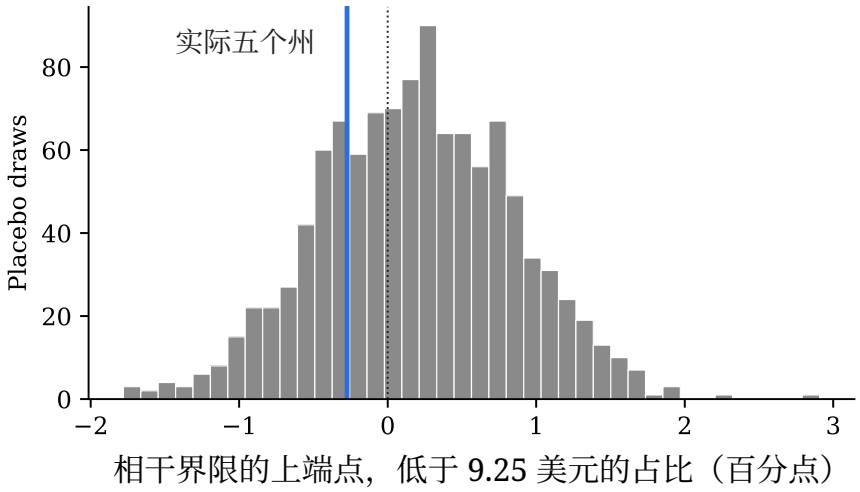


图 S1：安慰剂分配：相干界限的上端点注： 2,000 次安慰剂抽样中，有 1,090 次的相容集合非空。每次抽样都将五个实际事件日期分配给随机选出的五个从未接受处理的州，并将剩余 18 个州用作对照组。竖线：实际州及从未接受处理的对照组州集合。

在每个事件的样本池中（抽取 2,000 次），政策日期的非就业与其历史最低工资之间的差异为 -0.41 点，95%区间为 $[-0.47, -0.35]$。

人口权重。 事件等权重回答的是关于平均州级事件的问题；人口权重回答的是关于处理组州平均居民的问题。数据并未偏向任何一种权重，因此我们两者都报告。在按照结果季度处理组州的人口比例设定权重时（新泽西州 0.14、加利福尼亚州 0.60、特拉华州 0.01、密歇根州 0.16、明尼苏达州 0.09），堆叠凸包仍然相容，带符号界限为 $[-0.56, 0.20]$，相干界限为 $[-0.007, 0.078]$。

安慰剂分配。 在 2,000 次抽样中（随机种子为 7），每次都从 23 个从未接受处理的州中无放回地抽取 5 个州，并以随机顺序为其分配新泽西州、加利福尼亚州、特拉华州、密歇根州和明尼苏达州的结果季度；其余 18 个从未接受处理的州构成对照组。1,090 次抽样（54.5%）中，相容集合非空；以集合非空为条件进行分析属于选择，因此我们也报告无条件占比。安慰剂上端点（以百分点计）在第 5、10、50 和 90 百分位处的分位数分别为 -0.89、-0.62、0。19 和 1。04。使用从未接受处理的州作为对照组时，实际的上端点为 -0.27。在相容的抽样中，24.6% 的上端点不高于该值；无条件地看，全部 2,000 次抽样中有 13.4% 得到相容集合，且其上端点不高于该值。在第二项分析中，从规模最大的 11 个从未接受处理的州（按 2014年第一季度的人口计算）中抽取 5 个伪处理州，其余从未接受处理的州作为对照组；2,000 次抽样中有 77% 得到非空相容集合，其中 15% 的上端点不高于 -0.27，29% 的集合完全为负。

S13.5 一项令人信服的最低工资分析应增加什么

审计结果指出了四项要求。(i) 时间安排：根据法案颁布或可信公告的日期确定历史日期，而非根据生效日期；并检查政策日期的劳动力市场状况，在影响处理组与控制组差异的各个维度上是否处于历史范围之内（例如周期性就业，以及工资接近新最低工资的劳动者占比）。(ii) 构成：使用工人层面的 CPS 记录，在构建 118 个单元的向量之前，将处理州和对照州重新加权至固定的人口构成；并使用有效的同时置信带，逐个单元检验上尾排除。(iii) 对照组：使用从未接受处理的非相邻州，或明确对溢出效应建模，并直接衡量政策影响强度。(iv) 抽样：为州—季度单元获取基于设计的联合协方差，或以行政工资记录替代 CPS，以便将崩溃半径与覆盖率已知的抽样误差进行比较。在此之前，这些界限仅具描述性。

S13.6 负质量所在之处

对于每个州，取使其自身反事实的负质量最小的权重（命题 S2.9）。负值单元是工资区间；残差单元，即就业人口比减一，是每种情况下的正值。五个州共有40个这样的区间：新泽西州6个、加利福尼亚州13个、特拉华州13个、密歇根州2个、明尼苏达州6个。其中26个位于$8.00（2016年美元）以下，即低于或接近旧的最低工资标准、单元较为稀疏的区间；其余则分散在$10.75至$29.75之间。没有任何负计数超过总体的0.13%；最大的是加利福尼亚州的$14.75和$19.50。无论这些单元是否描述个体结果，工资区间中的就业岗位负计数都是不可能的，因此，各州反事实的不可行性并不依赖于对面板数据的个体解释。

参考文献

Abadie, A.、A. Diamond 和 J. Hainmueller（2010年）。用于比较案例研究的合成控制法：估计加利福尼亚州烟草控制计划的效应。《美国统计学会会刊》105(490)，493–505。Anderson, T. W.（1955年）。对称单峰函数在对称凸集上的积分及一些概率不等式。《美国数学学会会刊》6(2)，170–176。Andrews, D. W. K. 和 G. Soares（2010年）。使用广义矩选择对由矩不等式定义的参数进行推断。《计量经济学》78(1)，119–157。

Ashenfelter, O.（1978年）。估计培训项目对收入的效应。《经济学与统计学评论》60(1)，47–57。Athey, S. 和 G. W. Imbens（2006）。非线性双重差分模型中的识别与推断。《计量经济学》74(2)，431–497。Ban, K.、D. K´edagni、Z. Chen 和 A. Denteh（2026年）。从前趋势到界限：稳健的双重差分。工作论文，arXiv:2211.06710。Callaway, B. 和 T. Li（2019年）。面板数据双重差分模型中的分位数处理效应。《定量经济学》10(4)，1579–1618。Callaway, B. 和 P. H. C. Sant'Anna（2021年）。多个时期的双重差分。计量经济学杂志225(2)，200–230。Cal´onico, S. 和 J. Smith（2017年）。全国支持性就业示范项目中的女性。《劳动经济学杂志》35(S1)，S65–S97。Cengiz, D.、A. Dube、A. Lindner 和 B. Zipperer（2019年）。最低工资对低工资岗位的效应。《经济学季刊》134(3)，1405–1454。Chernozhukov, V.、D. Chetverikov 和 K. Kato（2015年）。高斯随机向量最大值的比较与反集中界。《概率论及相关领域》162(1–2)，47–70。Chernozhukov, V.、D. Chetverikov 和 K. Kato（2017年）。高维中的中心极限定理与自助法。《概率年刊》45(4)，2309–2352。Chernozhukov, V.、I. Fern´andez-Val 和 A. Galichon（2010年）。无交叉的分位数曲线与概率曲线。《计量经济学》78(3)，1093–1125。Chernozhukov, V.、H. Hong 和 E. Tamer（2007年）。计量经济模型中参数集的估计与置信区域。《计量经济学》75(5)，1243–1284。de Chaisemartin, C. 和 X. D'Haultfœuille（2020年）。具有异质处理效应的双向固定效应估计量。《美国经济评论》110(9)，2964–2996。Dehejia, R. H. 和 S. Wahba（1999年）。非实验研究中的因果效应：重新评估培训项目的评估。《美国统计学会会刊》94(448)，1053–1062。Dube, A.、T. W. Lester 和 M. Reich（2010年）。跨州界的最低工资效应：利用相邻县进行估计。《经济学与统计学评论》92(4)，945–964。Dvoretzky, A.、J. Kiefer 和 J. Wolfowitz（1956年）。样本分布函数与经典多项式估计量的渐近极小极大性质。《数学统计年刊》27(3)，642–669。Eissa, N. 和 J. B. Liebman（1996年）。劳动供给对劳动所得税抵免的反应。《经济学季刊》111(2)，605–637。Fang, Z. 和 A. Santos（2019年）。关于方向可微函数的推断。《经济学与统计学评论》86(1)，377–412。Fern´andez-Val, I.、J. Meier、A. van Vuuren 和 F. Vella（2024年）。分布回归双重差分。工作论文，arXiv:2409.02311。

Goodman-Bacon, A.（2021）。处理时点变差下的双重差分。计量经济学杂志225(2)，254–277。Gunsilius, F. F.（2023）。分布型合成控制。《计量经济学》91(3)，1105–1117。Hoffman, A. J.（1952年）。关于线性不等式系统的近似解。美国国家标准局研究杂志49(4)，263–265。Imbens, G. W.，以及 Y. Xu（2025年）。比较实验方法与非实验方法：LaLonde（1986年）之后四十年，我们学到了什么？《经济学展望杂志》39(4)，173–202。Kaido, H.、F. Molinari 和 J. Stoye（2019年）。部分识别参数投影的置信区间。《计量经济学》87(4)，1397–1432。Kleven, H. J.（2019年）。劳动所得税抵免与广延边际：重新评估。美国国家经济研究局工作论文26405。LaLonde, R. J.（1986 年）。利用实验数据评估培训项目的计量经济学评估。《美国经济评论》76(4)，604–620。Manski, C. F. 和 J. V. Pepper（2018年）。携枪权法律如何影响犯罪率？利用有界变差假设应对模糊性。《经济学与统计学评论》100(2)，232–244。Massart, P.（1990年）。德沃雷茨基–基弗–沃尔福维茨不等式中的紧常数。《概率年刊》18(3)，1269–1283。Masten, M. A. 和 A. Poirier（2020年）。关于崩溃前沿的推断。《定量经济学》11(1)，41–111。Meyer, B. D. 和 D. T. Rosenbaum（2001年）。福利、劳动所得税抵免与单身母亲的劳动供给。《经济学季刊》116(3)，1063–1114。Mogstad, M.、A. Santos 和 A. Torgovitsky（2018年）。使用工具变量推断政策相关处理参数。《计量经济学》86(5)，1589–1619。Rambachan, A. 和 J. Roth（2023年）。一种更可信的平行趋势方法。《经济学研究评论》90(5)，2555–2591。Romano, J. P.、A. M. Shaikh 和 M. Wolf（2014年）。检验矩不等式的实用两步法。《计量经济学》82(5)，1979–2002。Roth, J. 和 P. H. C. Sant’Anna（2023年）。平行趋势何时对函数形式敏感？《计量经济学》91(2)，737–747。Schanzenbach, D. W. 和 M. R. Strain（2021年）。劳动所得税抵免的就业效应：长期视角。《税收政策与经济》35，87–129。Schultz, T. P.（2004年）。面向贫困人口的学校补贴：评估墨西哥Progresa扶贫计划。《发展经济学杂志》74(1)，199–250。Shapiro, A.（1991年）。随机规划的渐近分析。《运筹学年刊》30(1)，169–186。

Smith, J. A. 和 P. E. Todd（2005）。匹配能否克服LaLonde对非实验估计量的批评？计量经济学杂志125(1–2)，305–353。Sun, L. 和 S. Abraham（2021年）。在处理效应存在异质性的事件研究中估计动态处理效应。计量经济学杂志225(2)，175–199。Todd, P. E. 和 K. I. Wolpin（2006）。评估墨西哥学校补贴计划的影响：利用社会实验验证关于儿童教育与生育率的动态行为模型。《美国经济评论》96(5)，1384–1417。