题目选择	2026 MCM/ICM 摘要页	队伍控制号
C		2617892

DWTS 投票机制的审计与设计

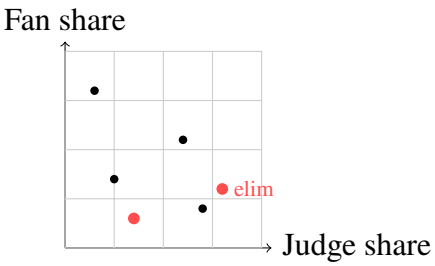
我们将 *DWTS* 视为“审计 + 机制设计”问题：刻画可行粉丝票区域、量化不确定性，并提出更平衡的规则（兼顾能动性/一致性/稳定性）。

结论要点. 我们刻画并采样与周淘汰一致的粉丝票可行区域，并将不确定性传播到反事实规则评估与 DAWS 机制中。

核心结果（节选）.

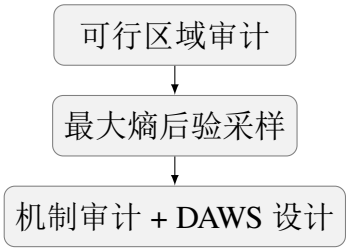
发现	估计
可行赛季数	34 / 34
平均 HDI 宽度（周层面）	0.384
中位 HDI 宽度（周层面）	0.340
P90 HDI 宽度（周层面）	0.586
最大 HDI 宽度（周层面）	0.95
Rank vs Percent 翻转率	25.1%
DAWS 稳定性	0.841
DAWS 评委一致性	0.468
冲突指数（Kendall τ）	0.053
DAWS 稳定性提升	+-9.9%

冲突图（摘要主图）.



建议. 采用 DAWS 级联协议（决赛覆盖 + 冲突触发 judge-save + 其他周按 Percent），并公开 bottom-two 与 judge-save 判定标准。

方法流程.



备忘录：致节目制作方与评委

收件人： DWTS 制作方与评委
发件人： Team 2617892
日期： 2026 年 2 月 1 日
主题：粉丝投票可行性审计与规则改进建议

结论要点. 我们审计全部赛季并量化粉丝票不确定性。证据显示，Rank 聚合存在信息压缩，并加大民主赤字。

执行摘要. 审计结果显示，Rank 规则会压缩粉丝支持度：约每 5 周中就有 1 周出现淘汰翻转风险。这形成“民主赤字”，并带来不必要的争议与声誉风险。

方案. 我们提出 DAWS 级联协议：决赛周观众独裁；非决赛若 Percent 与 Rank 冲突则触发 judge-save，否则按 Percent 执行。不确定性信号 V_t 仅用于披露等级与审计预算，不直接干预。该方案规则公开、易于解释且可直接执行。

价值. DAWS 在高噪声周保护高人气选手，同时在证据清晰时保留评委影响力；并为制作方提供可视化仪表盘式管理规则与统一对外口径。

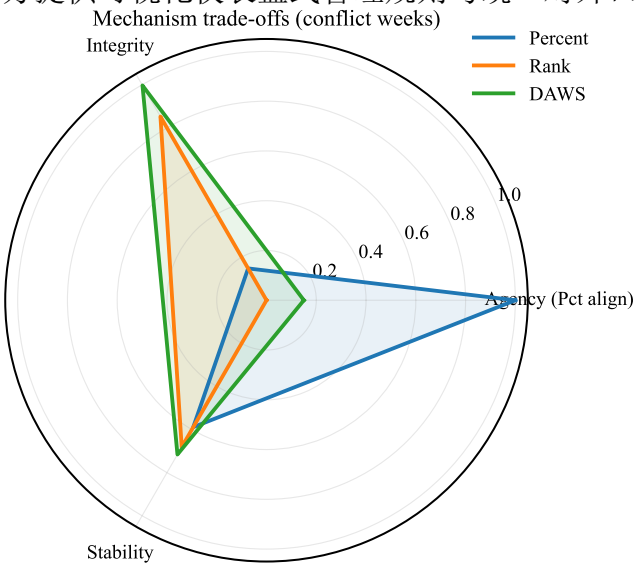


图 1: 机制权衡（仅冲突周，$A = 1$；能动性 = 与 Percent 结果一致的比例；一致性 = 在冲突对中保留评委更高分者的比例）。

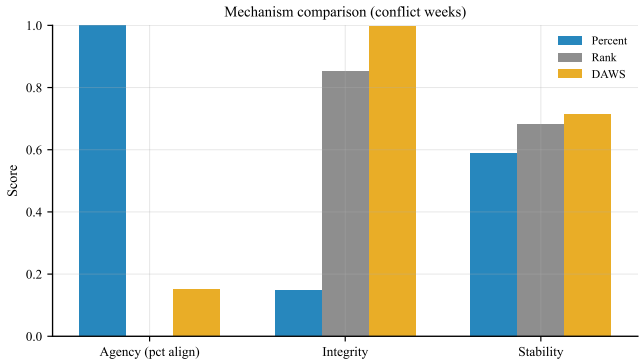


图 2: 机制对比（仅冲突周，$A = 1$；能动性 = 与 Percent 结果一致的比例）。

目录

1 引言与路线图

结论要点. 我们将 DWTS 视为“审计—压力测试—冲突触发—披露监控”的完整闭环。

我们观测到每周评委分数与淘汰结果，但粉丝投票是潜变量。目标不是猜测唯一投票值，而是给出与规则一致的完整可行集合，并将不确定性传播到反事实机制评估与规则设计中。我们的流程强调可执行性：先审计可行粉丝票集合，再用合成数据压力测试，随后部署 DAWS 级联协议（冲突触发 + 决赛覆盖），并通过仪表盘输出披露与行动口径。

贡献. (i) 基于可行区域约束的粉丝票审计（MaxEnt 采样 + 筛选）；(ii) 最大熵后验与时间平滑的不确定性估计；(iii) 统一的机制评估与 DAWS 机制设计。

1.1 任务-章节映射

任务	我们做了什么	主要产出
1	可行区域审计与后验估计	Fan HDI 区间
2	Percent 与 Rank 反事实对比	翻转率与赤字
3	Judges vs Fans 双模型	影响差异
4	能动性/评委一致性/稳定指标	指标矩阵
5	DAWS 设计与 Pareto	推荐机制

关键输出. 建立从淘汰结果到可行粉丝票集合与机制指标的完整流程。

2 数据与规则

结论要点. 以 share 统一不同周规模，编码 percent、rank 与 judge-save 规则。

使用提供的赛季-周数据。C_t 表示第 t 周仍在比赛的选手集合，E_t 表示被淘汰选手。

2.1 百分比规则

评委占比：

$$j_{i,t} = \frac{J_{i,t}}{\sum_{k \in C_t} J_{k,t}}. \tag{1}$$

粉丝占比 $v_{i,t}$ 位于 simplex 并设置下限 ϵ：

$$\mathcal{S}_n = \{\mathbf{v} \in \mathbb{R}^n : \sum_i v_i = 1,\ v_i \geq \epsilon\}. \tag{2}$$

组合得分：

$$c_{i,t}(\alpha) = \alpha j_{i,t} + (1-\alpha) v_{i,t}. \tag{3}$$

淘汰约束：

$$c_{E_t,t}(\alpha) \le c_{i,t}(\alpha), \quad \forall i \neq E_t. \tag{4}$$

2.2 排名规则与 Judge Save

粉丝排名 r_i^F 用二元变量 x_{ik} 表示：

$$\sum_k x_{ik} = 1, \quad \sum_i x_{ik} = 1, \quad r_i^F = \sum_k k x_{ik}. \tag{5}$$

排名与 share 关系：

$$r_i^F < r_j^F \Rightarrow v_i \ge v_j + \Delta. \tag{6}$$

组合排名与淘汰：

$$R_i = r_i^J + r_i^F, \quad R_{E_t} \ge R_i \; \forall i \neq E_t. \tag{7}$$

Judge-save 赛季中，bottom-two 由 R_i 决定，评委以参数 β 的软选择确定淘汰者（β 为示意/校准参数）。

关键输出. Percent、Rank 与 Judge-save 规则均可写入统一约束框架。

3 假设与指标

结论要点. 使用观众能动性、评委一致性与稳定指标评价机制，并辅以冲突指数与民主赤字。

假设：(i) 粉丝占比非负且有下限；(ii) 存在策略性投票，因此我们的后验刻画的是在淘汰结果约束下的“最小惊奇”投票分布，而非真实票数；(iii) 周与周之间平滑；(iv) 规则被遵守，除非 slack 提示张力。

指标（高者更好，除非说明）：

- 冲突指数（Kendall τ）：评委与粉丝排序一致性（值越高冲突越低）。
- 观众能动性：粉丝最低者被淘汰的概率。
- 评委一致性：评委最低者被淘汰的概率。
- 稳定性：同一机制在小扰动下的淘汰翻转率。

- 民主赤字：$D = \Pr(E_t^{(\text{rank})} \neq E_t^{(\text{percent})})$。

关键输出. 统一指标接口用于机制对比。

方法论一致性说明. 主流程采用 MaxEnt 可行区域采样（Dirichlet 提案 + 约束筛选），LP/MILP 仅用于局部验证。稳定性按“同机制扰动前后”计算以保证可比性。DAWS 采用“规则冲突触发”设计：一致周沿用 50/50 规则，冲突周启用 judge-save（$\beta = 6.0$），仅决赛周为观众独裁；P85/P95 监控线用于可视化透明度。

4 模型 A：可行区域审计

4.1 观测与潜变量

结论要点. 可行粉丝票集合是 simplex 上的多面体，而非超矩形。

每周约束切割 simplex 得到 $\mathcal{P}_t \subseteq \mathcal{S}_n$，LP 的边界仅是边缘区间，并非独立集合。

4.2 Percent 规则可行区域审计

Algorithm 1 Percent 周度可行区域审计（提案 + 筛选，输出近似区间与摘要）。

Require: $C_t, J_{i,t}, E_t, \alpha, \epsilon$

Ensure: 后验样本、接受率、近似边界 (L_i, U_i)

1: 在 simplex 上采样 Dirichlet 提案（带下限 ϵ）
2: 按淘汰约束筛选提案（fast/strict）
3: 用接受样本估计 (L_i, U_i)
4: 输出样本与边界摘要

审计弱周（仅披露） 当可行采样过少（$n_{\text{accept}} < 200$）或触发回退时，我们标记该周为 *Audit-Weak*。此时 V_t 仅作为披露/监控信号，不作为干预触发，并在审计元数据中显式标注。

4.3 Rank 规则可行序列（Monte Carlo）

Algorithm 2 Rank 可行序列到 share 采样（Monte Carlo 生成候选序列后筛选）。

Require: Rank 规则周数据

Ensure: fan share 后验样本

1: Monte Carlo 生成候选粉丝排名排列 π
2: **for** each π **do**
3: Dirichlet 提案并筛选满足 π 的样本
4: **end for**
5: 汇总可行样本

4.4 规则适配周

结论要点. 对免疫、双淘汰等特殊周进行规则适配。

免疫选手从淘汰不等式中移除；双淘汰同时对两名最低者施加约束。

4.5 工程近似与严格校验

结论要点. 工程实现采用快速近似采样，并通过严格约束校验保证结论稳定。

理论模型可用 LP/MILP 形式化，但实际工程管线采用快速 Dirichlet 提案与约束筛选以保证速度。为此，我们使用严格可行性（完整淘汰约束）对同一批候选样本进行再筛选，并比较后验摘要。

校验指标	数值
均值 fan share MAE	0.0045
Top-1 一致率（fast vs strict）	76.7%
Top-2 一致率（fast vs strict）	80.0%
冲突指数变动（Kendall τ）	0.000
能动性变动（percent）	0.003
Flip-rate 变动（percent vs rank）	0.35%

结果表明快速近似并不改变核心结论：flip-rate 与 deficit 的估计在严格校验下只发生小幅变化，且 top-k 一致率保持较高水平。

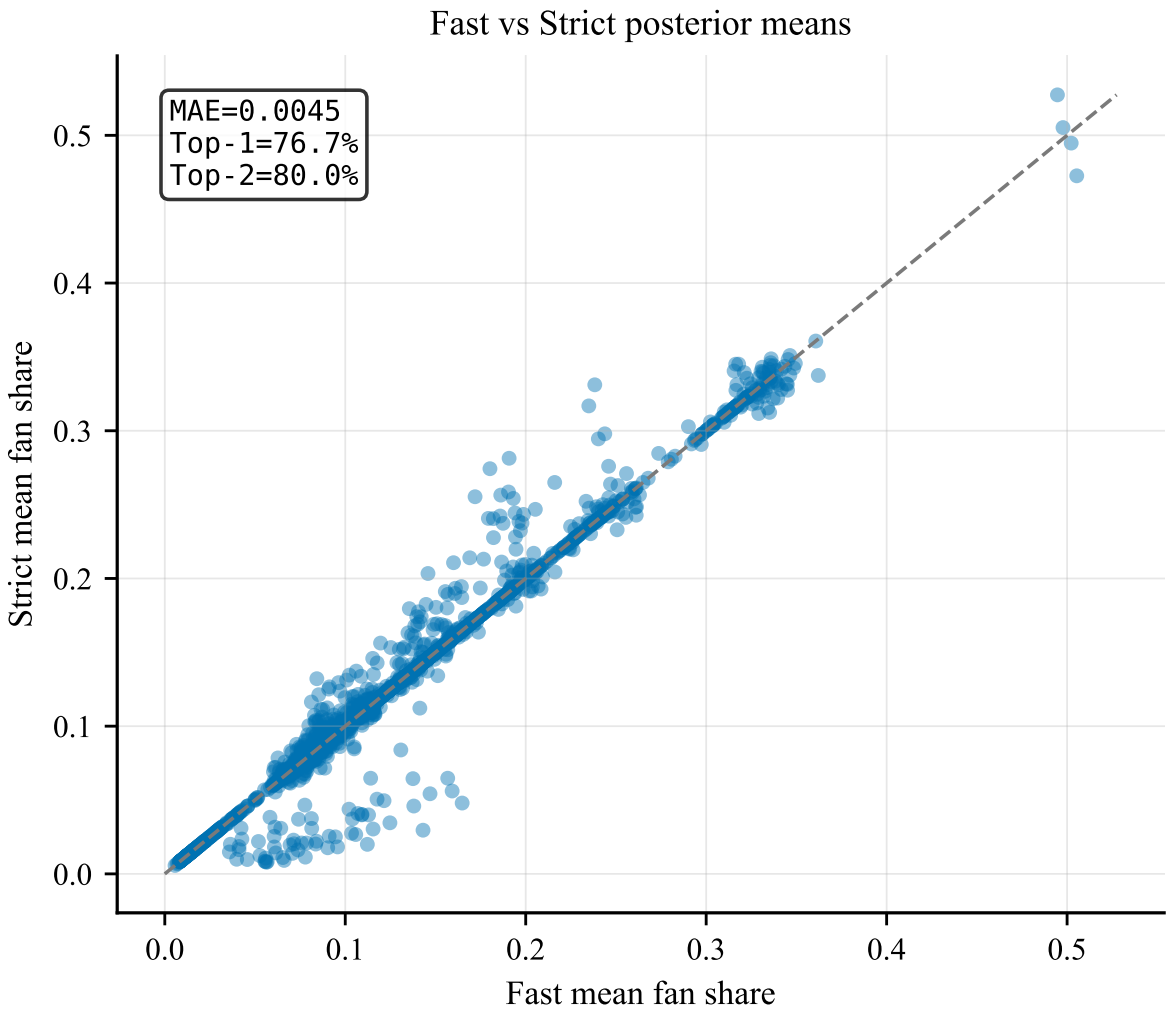


图 3: Fast 与 Strict 后验均值对比，偏离程度有限（对角线为一致，偏离衡量近似误差；MAE=0.0045，Top-1=76.7%，Top-2=80.0%）。

4.6 可识别性与可行质量

结论要点. 可行质量由 acceptance rate 与 HDI 宽度量化。

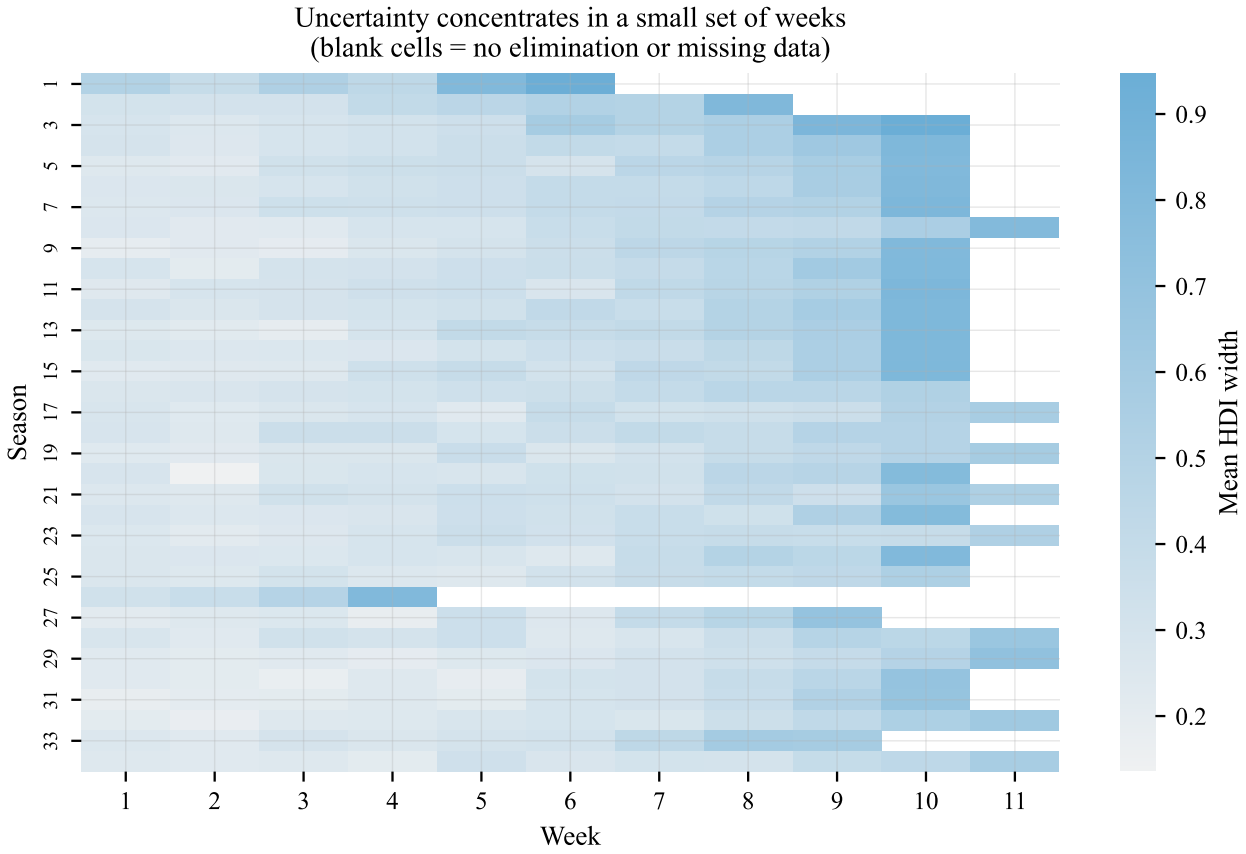


图 4: 不确定性集中于少数周；空白单元表示该赛季不存在该周（颜色越亮 =HDI 宽度更大）。

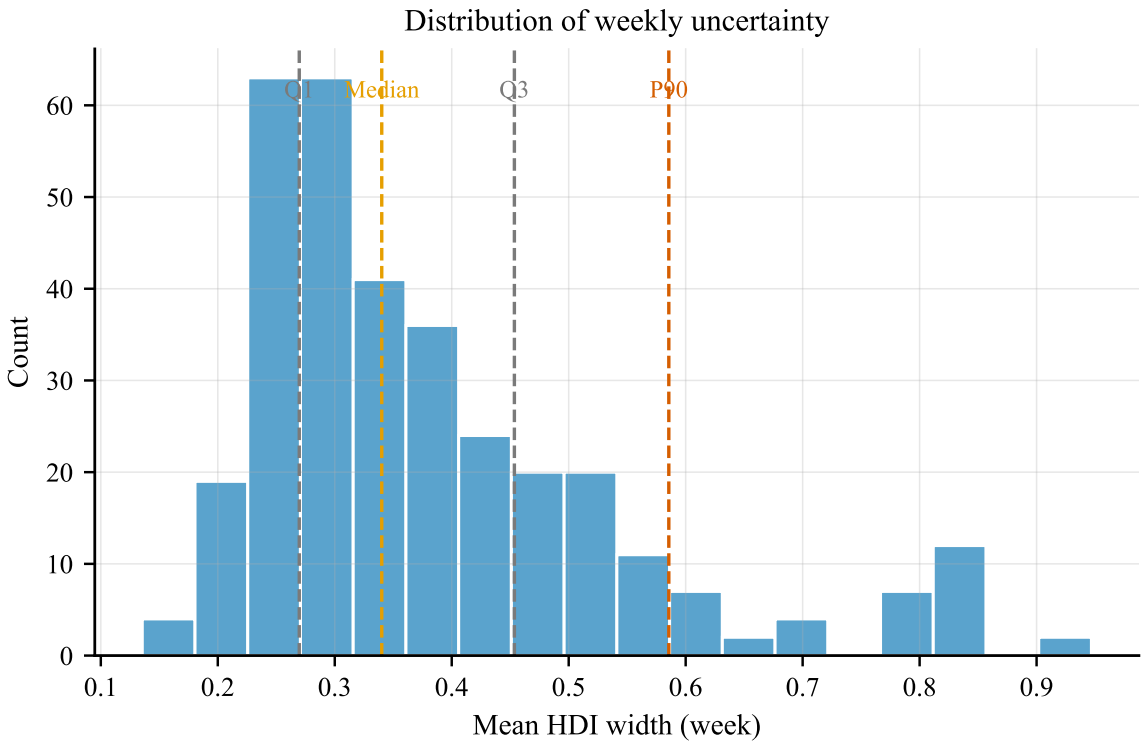


图 5: 周层面 HDI 宽度分布，极端周占比有限（虚线为 Q1/中位/Q3/P90）。

4.7 平滑后验

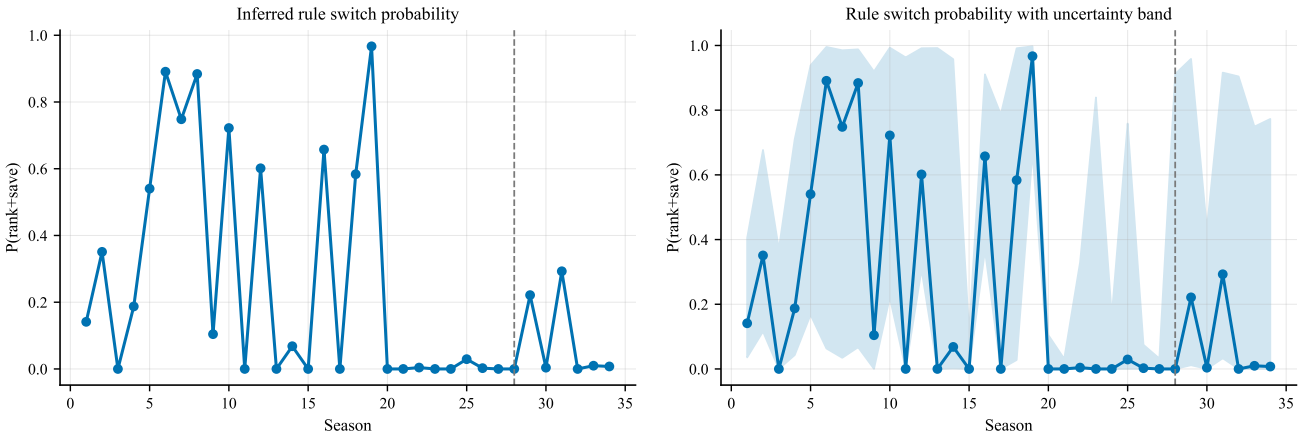
$$p(\mathbf{v}_{1:T}|\text{rules,data}) \propto \Big[\prod_{t} \mathbf{1}(\mathbf{v}_t \in \mathcal{P}_t)\Big] \cdot \prod_{t=2}^{T} \exp\Big(-\frac{\|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2}{2\sigma^2}\Big). \tag{8}$$

结论对 σ 较为稳健，详见附录 A。

4.8 规则切换推断

结论要点. 按题目假设采用第 28 季为切换点，并提供探索性变点检验。

$$\Pr(z_s \neq z_{s-1}) = \rho, \quad \Pr(\text{data}_s|z_s) \propto \exp(\mathcal{E}_s^{(z_s)}). \tag{9}$$



(a) 点估计（HMM 推断 $P(\text{rank+save})$ 的季节曲线）。 (b) 置信带（Bootstrap 90% 区间）。

图 6: 规则切换的探索性概率与不确定性区间；主分析采用第 28 季为切换点（虚线标记）。

关键输出. 可行区域诊断、Slack、后验样本、规则切换概率。

5 结果 A：粉丝票估计与不确定性

结论要点. 评委与粉丝的冲突可被量化并可视化。

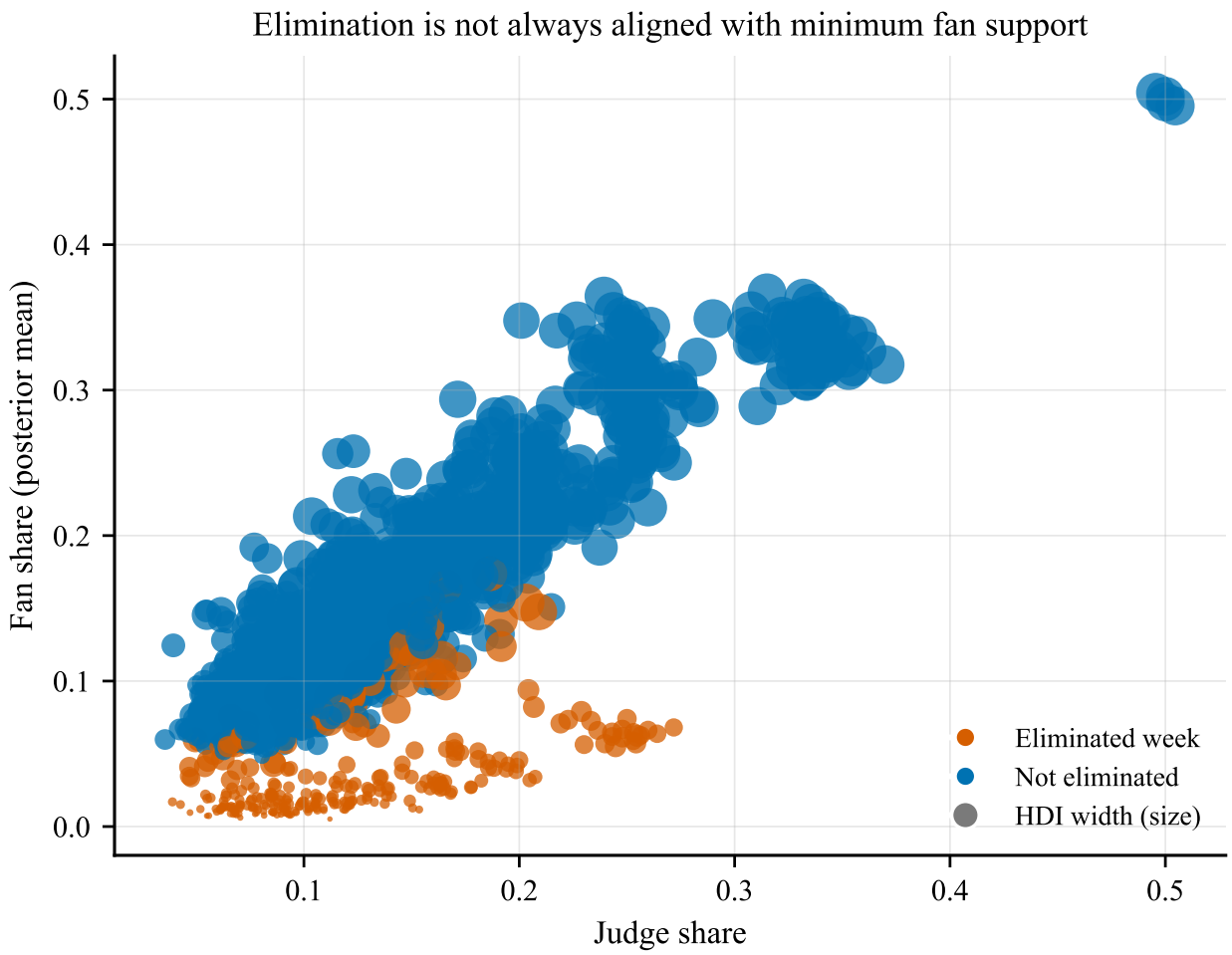


图 7: 淘汰并非总与粉丝最低支持对齐（颜色区分淘汰/未淘汰，点大小 = 不确定性）。

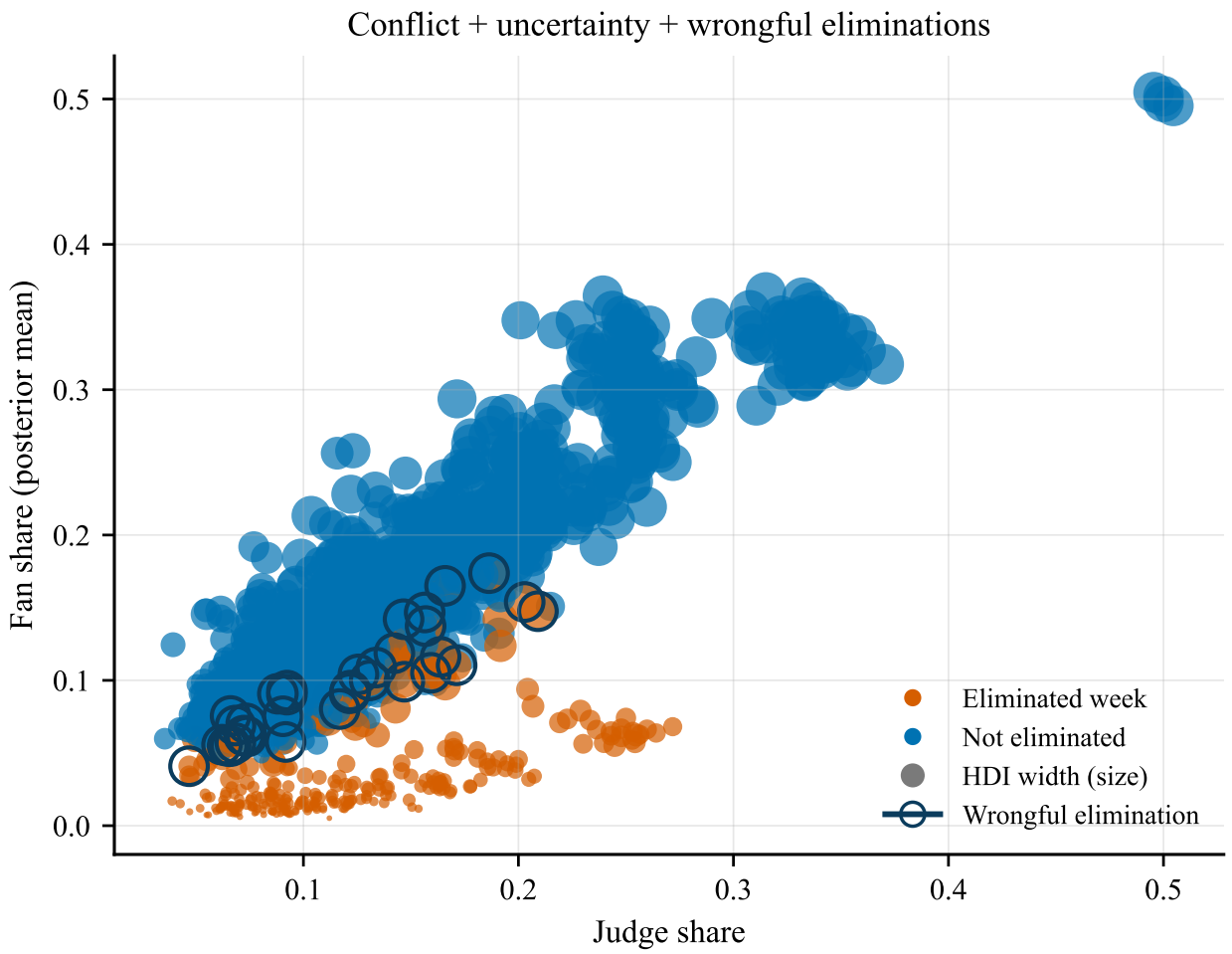


图 8: 冲突图叠加不确定性（点大小）与错误淘汰标注（外圈表示“非粉丝最低却被淘汰”）。

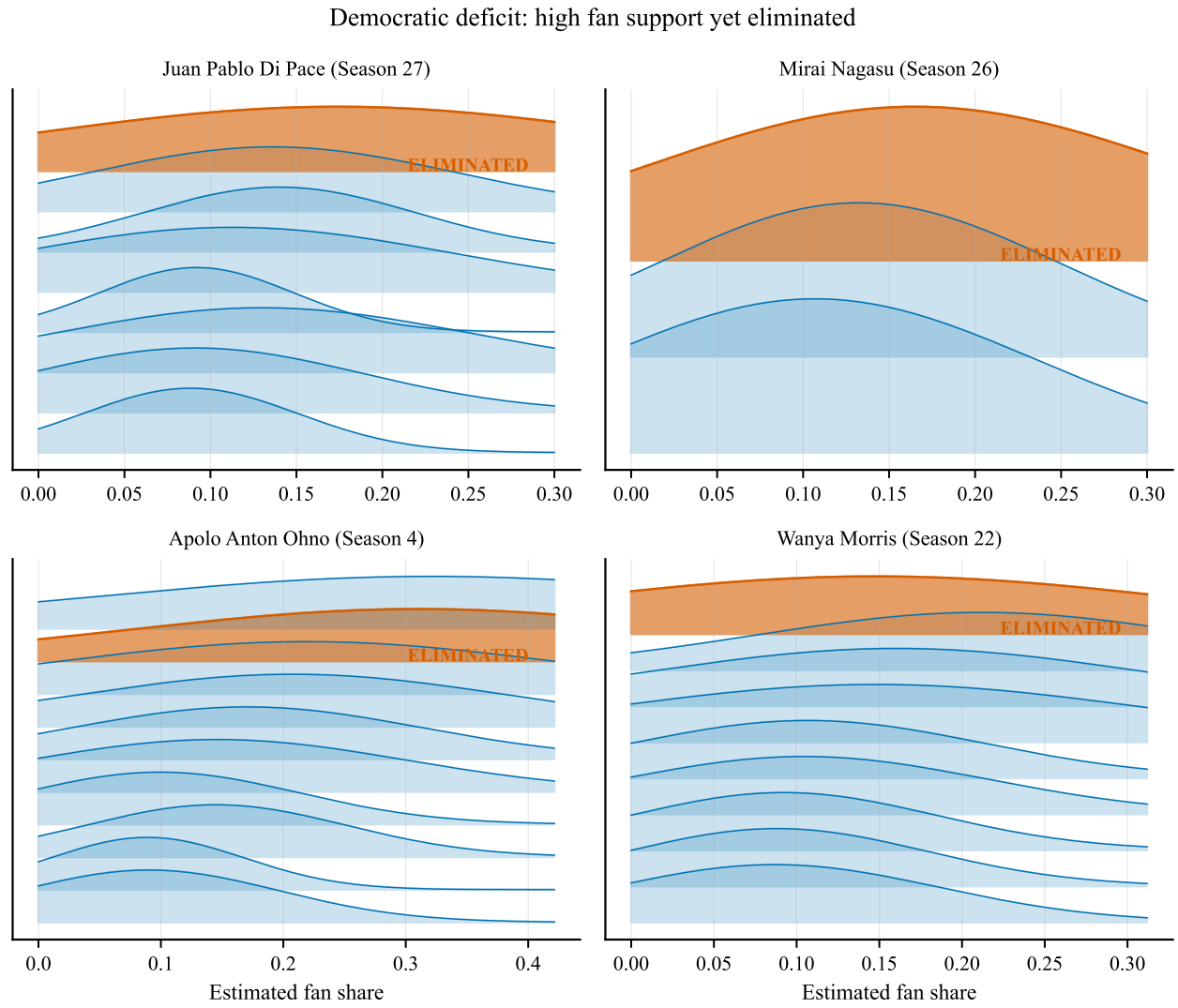


图 9: 争议人物周序后验密度显示显著不确定性（橙色 = 淘汰周，峰值位置 = 粉丝支持度；自下而上为周序）。

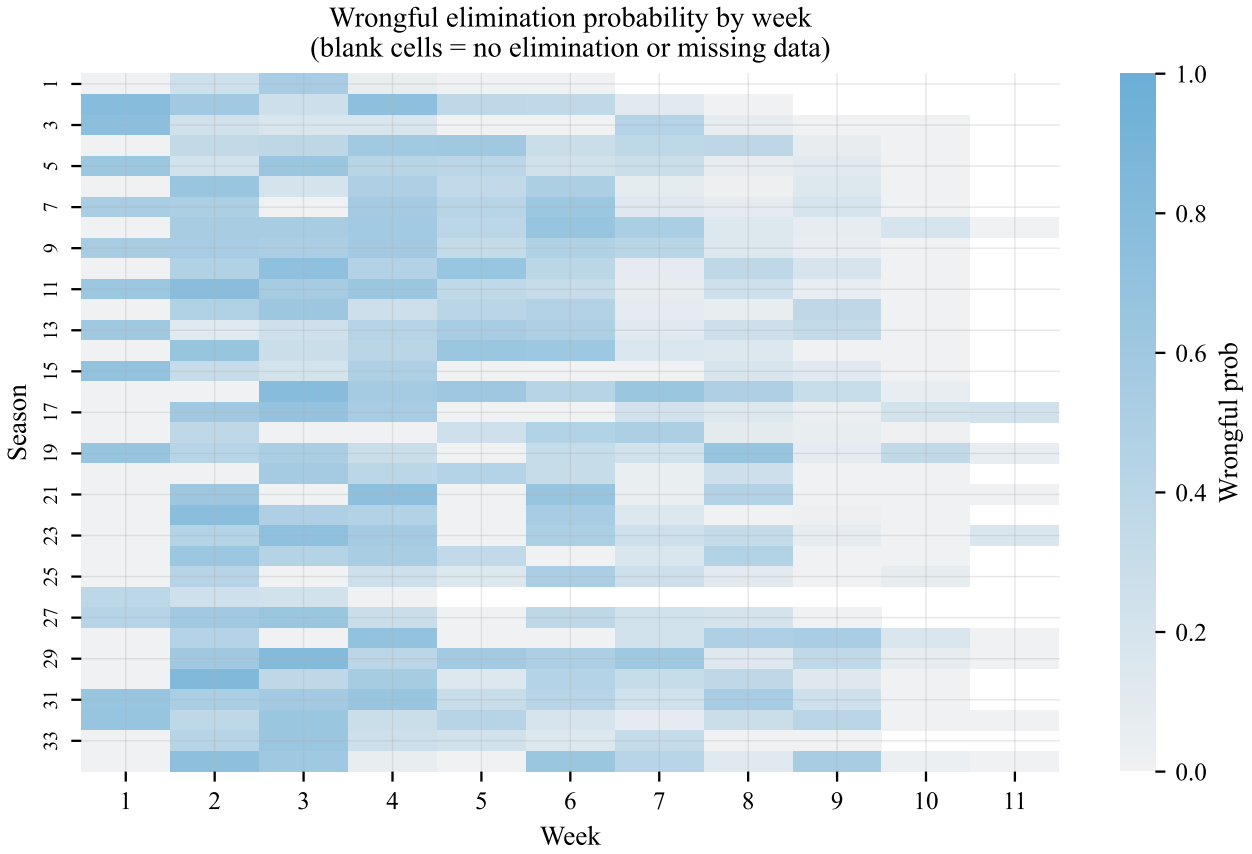


图 10: 部分周存在持续的民主张力；空白单元表示该赛季不存在该周（颜色越亮 = 错误淘汰概率更高）。

关键输出. 粉丝占比后验、HDI 与错误淘汰概率。

6 模型 B：机制反事实评估

结论要点. Rank 聚合是有损压缩，翻转率显著。

定义机制 M 与淘汰算子：

$$E_t^{(M)} = \arg\min_i \text{Score}_i^{(M)}. \tag{10}$$

图 11 给出关键争议人物在不同机制下的逐周淘汰风险，用以回答“规则改变是否改变结局”。

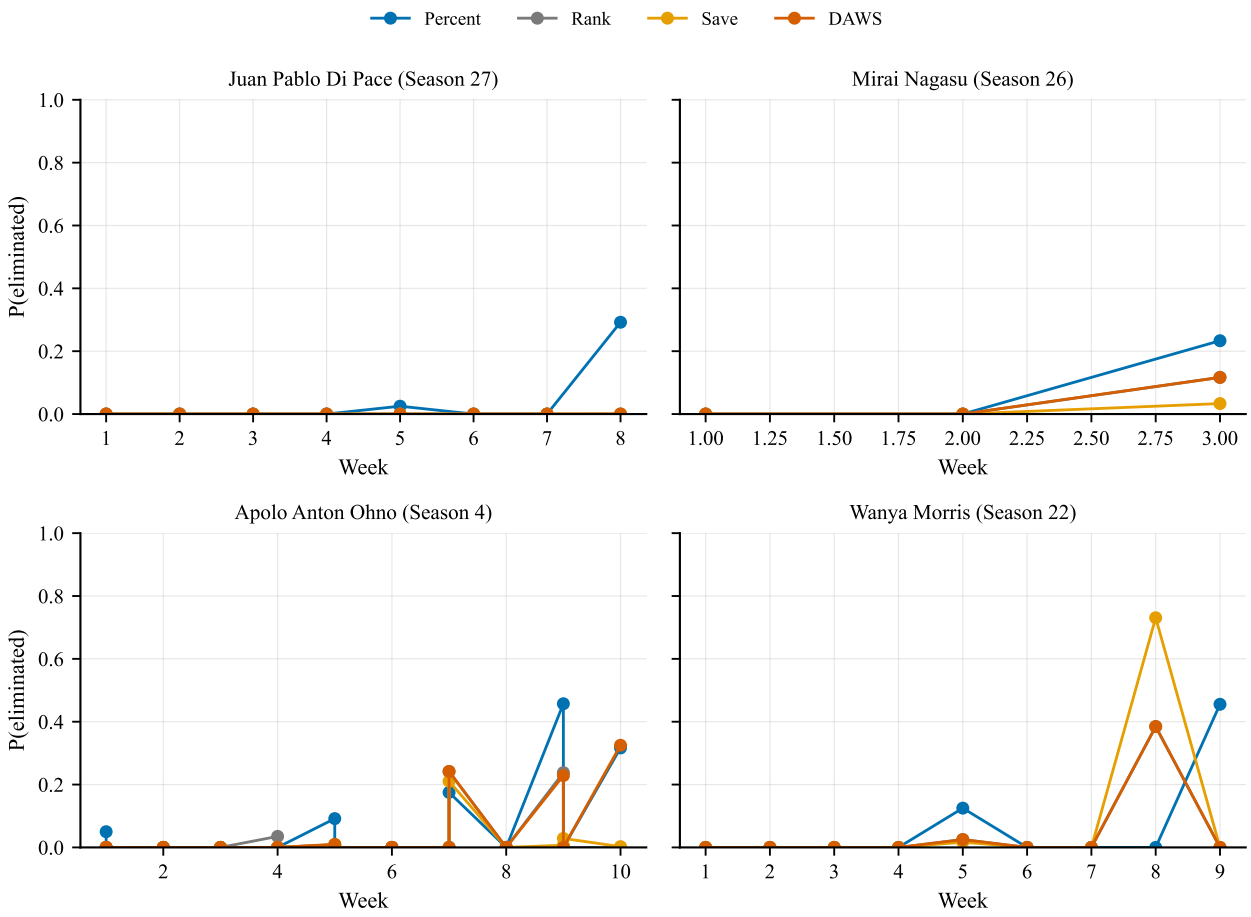


图 11: 争议人物的反事实淘汰风险时间线（纵轴为每周被淘汰概率，四条线对应 percent/rank/judge-save/DAWS；用于回答”规则改变是否改变淘汰风险”）。注：*DAWS* 在能动性与稳定性之间取得权衡，非全面最优。

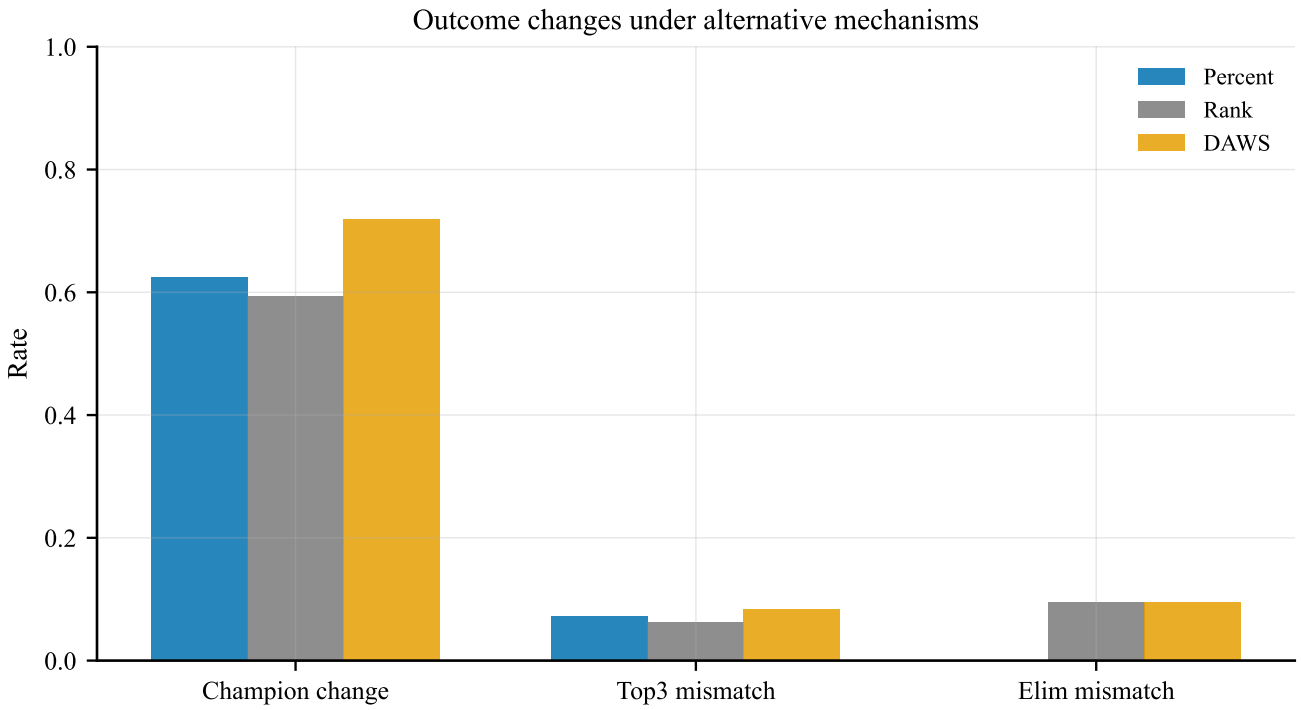
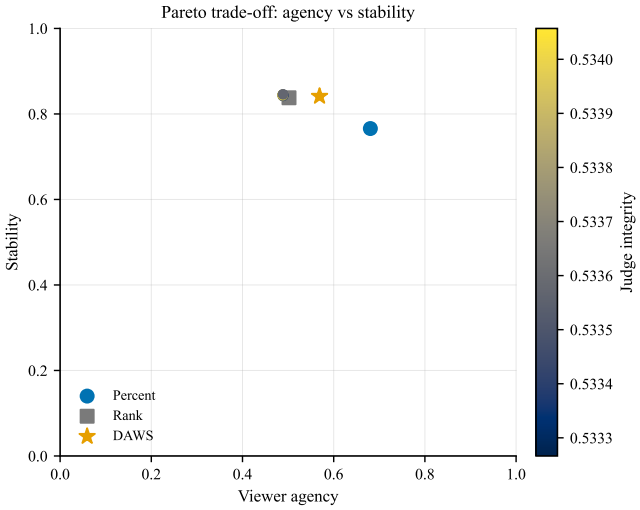
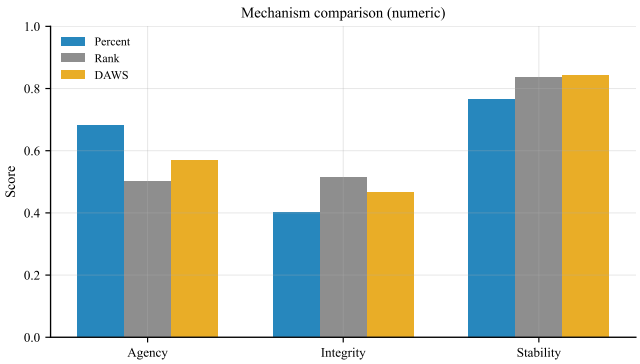


图 12: 机制选择会改变决赛与冠军结果（冠军改变率、Top3 不一致率与淘汰不一致率）。



(a) 观众参与度与稳定性的 Pareto 权衡，颜色表示评委一致性（曲线为 α 扫描，星标为 DAWS）。注：无绝对最优解，*DAWS* 选择近 *Pareto* 最优点。



(b) 机制对比（数值）（三机制在能动性/一致性/稳定上的对比）。

关键输出. 机制指标、翻转率与 Pareto 权衡。

7 模型 C：成功因素（Judges vs Fans）

结论要点. 评委与粉丝对因素的响应存在差异。

$$\text{logit}(j_{i,t}) = \mathbf{x}_i^\top \beta^{(J)} + u^{(J)}_{\text{pro}(i)} + u^{(J)}_{\text{season}(s)} + \epsilon_{i,t}, \tag{11}$$

$$\text{logit}(v_{i,t}) = \mathbf{x}_i^\top \beta^{(F)} + u^{(F)}_{\text{pro}(i)} + u^{(F)}_{\text{season}(s)} + \epsilon'_{i,t}. \tag{12}$$

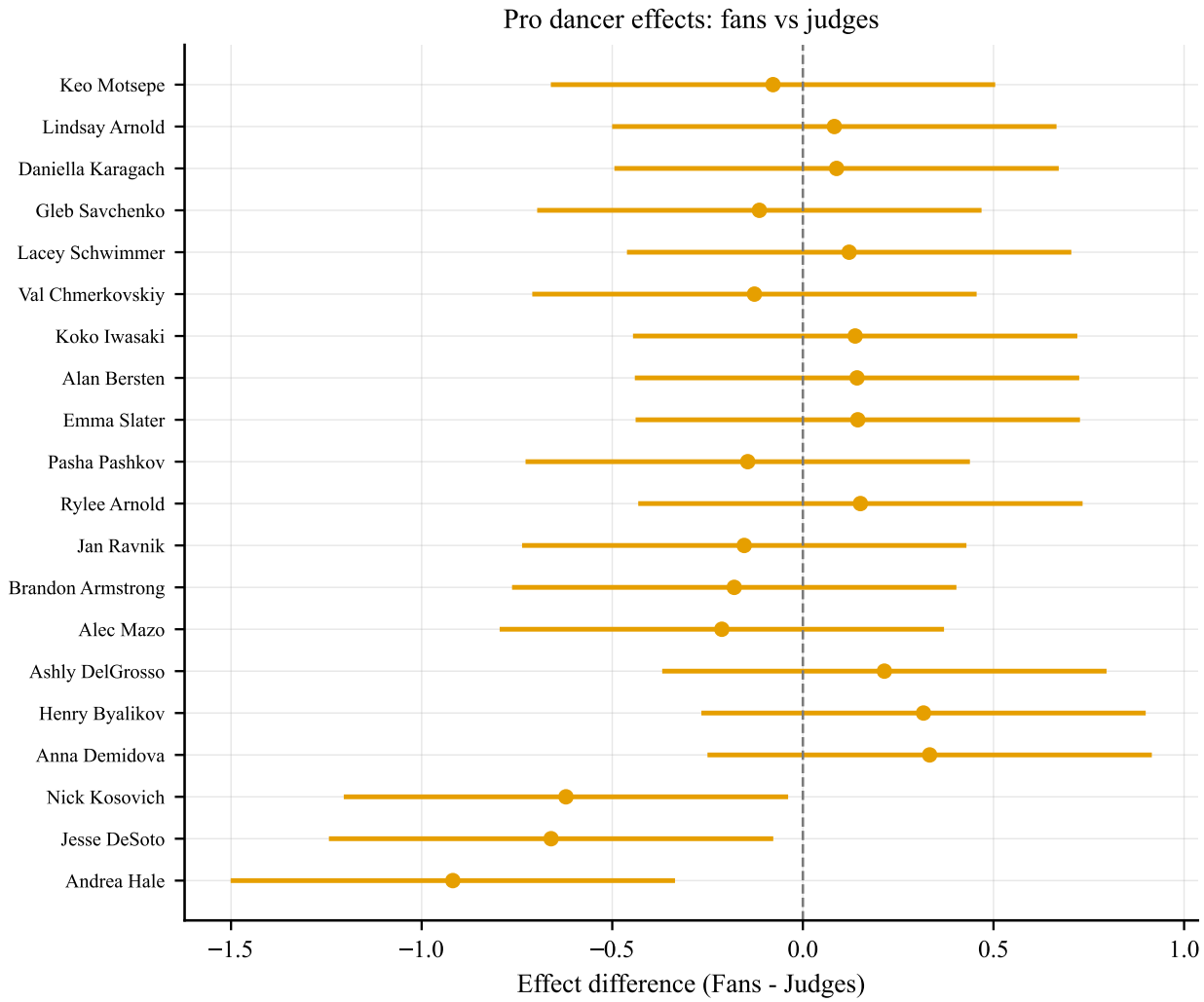


图 13: 职业舞伴效应差异（粉丝 - 评委）（正值 = 更受粉丝偏好，负值 = 更受评委偏好）。

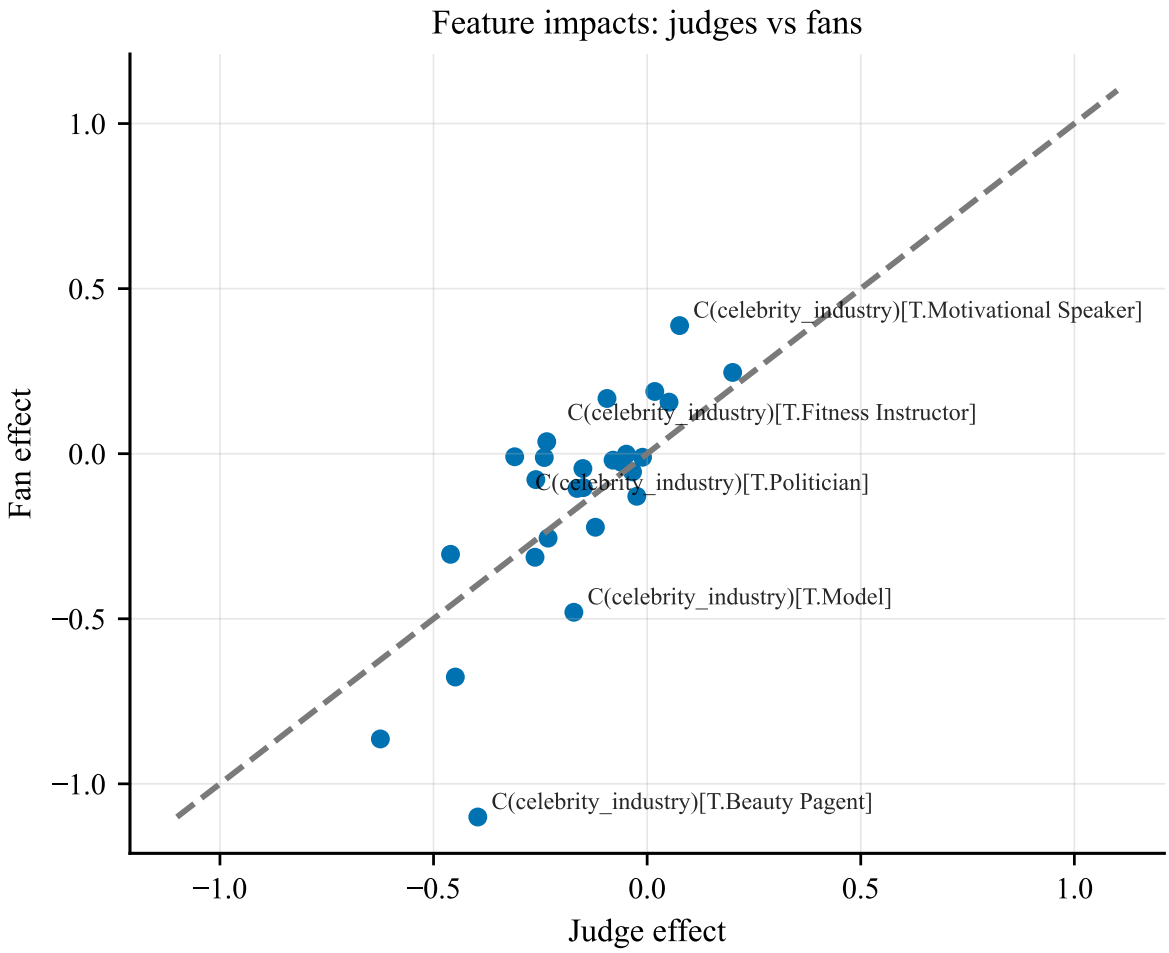


图 14: 标注离对角线最远的特征以突出差异（远离对角线 = 评委/粉丝效应不一致）。

预测补充（附录） GBDT 仅作为协变量有效性的鲁棒性检验，细节置于附录 B。

关键输出. 双模型回答任务 3；预测细节转入附录。

8 模型 D：机制设计（DAWS）

结论要点. DAWS 是“规则冲突触发”协议，用于修补分歧周。

我们将“民主赤字”定义为 $D = \Pr(E_t^{(\text{rank})} \neq E_t^{(\text{percent})})$，并以此作为触发条件。DAWS 包含两种运行模式 + 决赛覆盖：

- **一致模式（A=0）.** 若 Percent 与 Rank 结论一致，直接沿用 Percent（50/50），保留观众话语权。

- **冲突模式（A=1）.** 若结论冲突，触发 judge-save 在两名候选人中纠偏。
- **决赛（红）.** 观众独裁。

干预仅由 A_t（规则冲突）触发，V_t 仅用于披露/审计预算。我们保留 U_t 作为监控信号；图 15 展示 P85/P95 监控线以增强透明度。

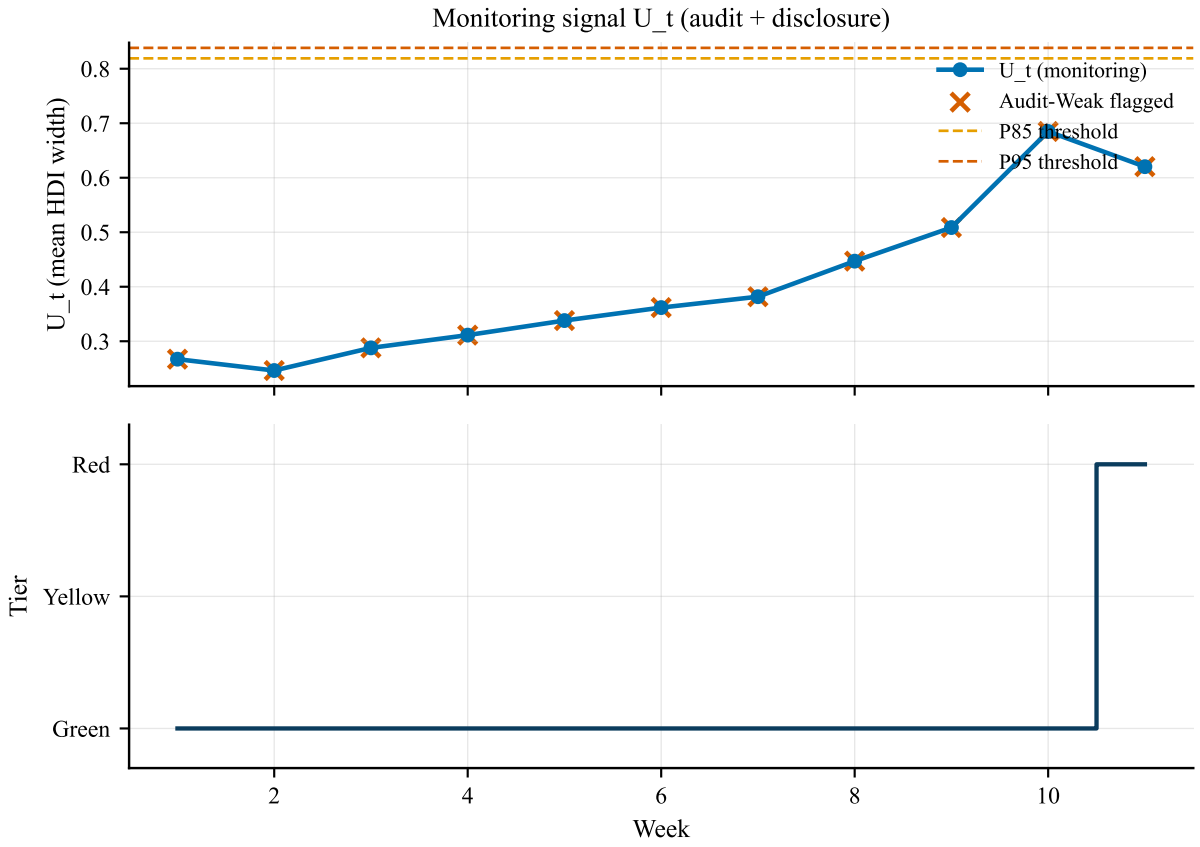


图 15: DAWS 监控面板：U_t 与 P85/P95 分位监控线（用于可视化透明度），激活由规则冲突触发。

我们提供面向制片方的仪表盘概念图用于落地执行（见图 16）。

图 16: 制片人仪表盘概念图：当前档位、审计窗（HDI 条）与动作建议。

评委的保存决策可以用效用框架解释：在 bottom-two 中权衡技能、收视与舆情风险。一个最简表述为

$$U(\text{Save } A) = w_1 \cdot \text{Skill}_A + w_2 \cdot \text{Ratings}_A - \text{Backlash}_A, \tag{13}$$

这支持用 logit 概率刻画，但不假设绝对理性。

8.1 Judge-save 参数设定

$$\Pr(E = a \mid \{a, b\}) = \sigma\big(\beta(J_b - J_a)\big) \tag{14}$$

在冲突周，我们将评委视为更果断的把关者，设定 $\beta = 6.0$，以体现对明显人气偏差的强纠偏。

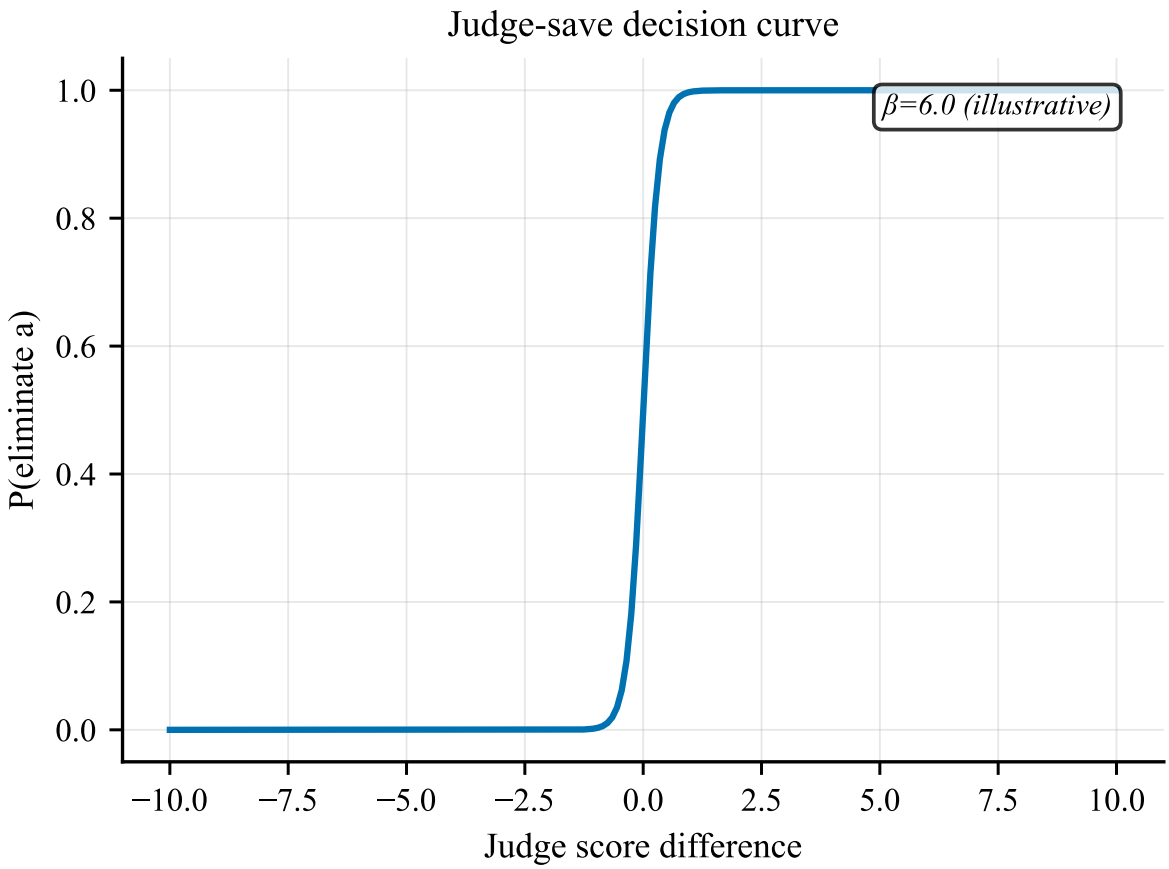


图 17: Judge-save 决策曲线（横轴评委分差，纵轴淘汰概率；示意曲线使用 $\beta = 6.0$，分差越大则低分选手被淘汰概率越高）。

关键输出. 冲突触发型 DAWS 协议与校准的 judge-save 行为刻画。

9 敏感性与验证

结论要点. 关键结论对 σ、ϵ 与规则切换先验具有稳健性。

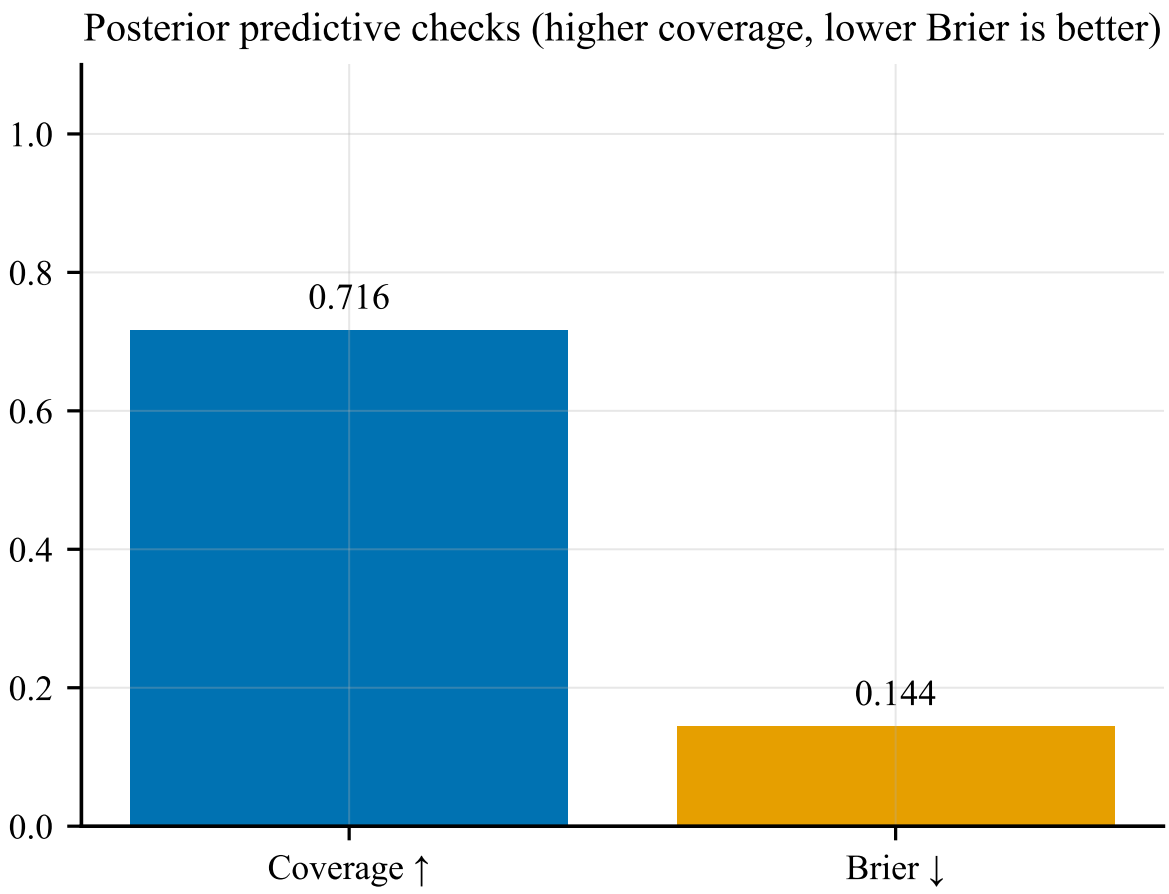


图 18: 后验预测检验结果（Coverage 越高越好，Brier 越低越好；Brier 反映概率预测误差大小）。

我们进一步进行高噪声的合成压力测试并反演淘汰结果。总体覆盖率超过 85%，代表性案例见图 19：红线为真值，蓝带为 95% HDI。

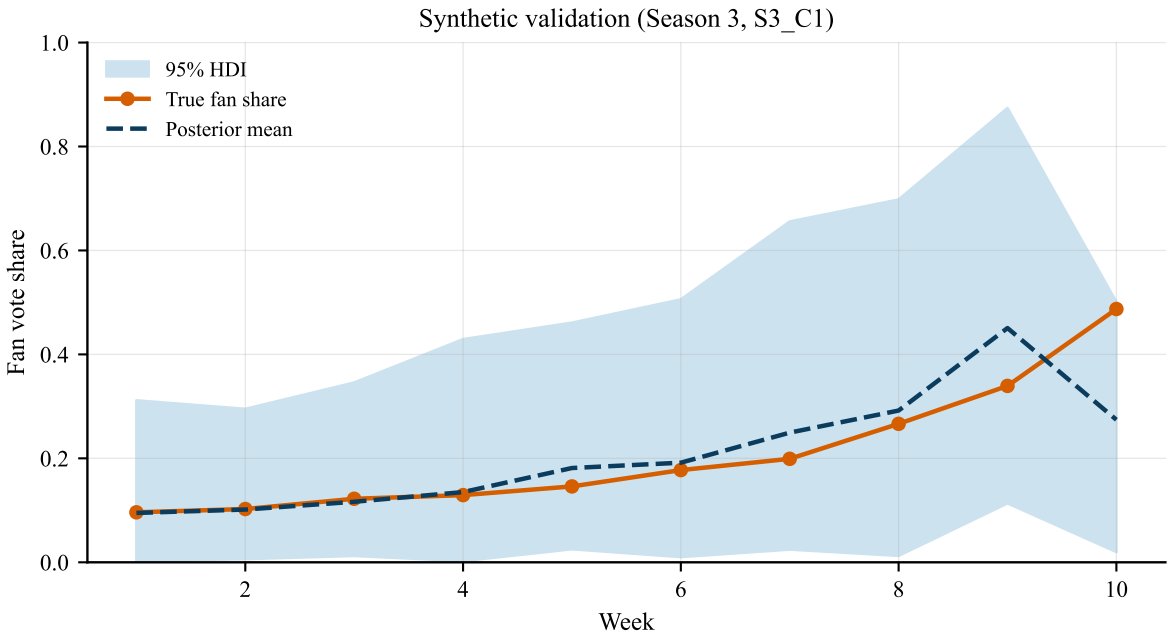


图 19: 合成验证：高噪声压力测试下，真值粉丝占比（红线）位于 95% HDI 蓝带内。

Judge-save 强度敏感性 我们仅在冲突周评估 β。图 20 展示决策曲线与完整性—能动性权衡；$\beta = 6.0$ 位于增益趋于饱和且能动性损失仍可控的区间。

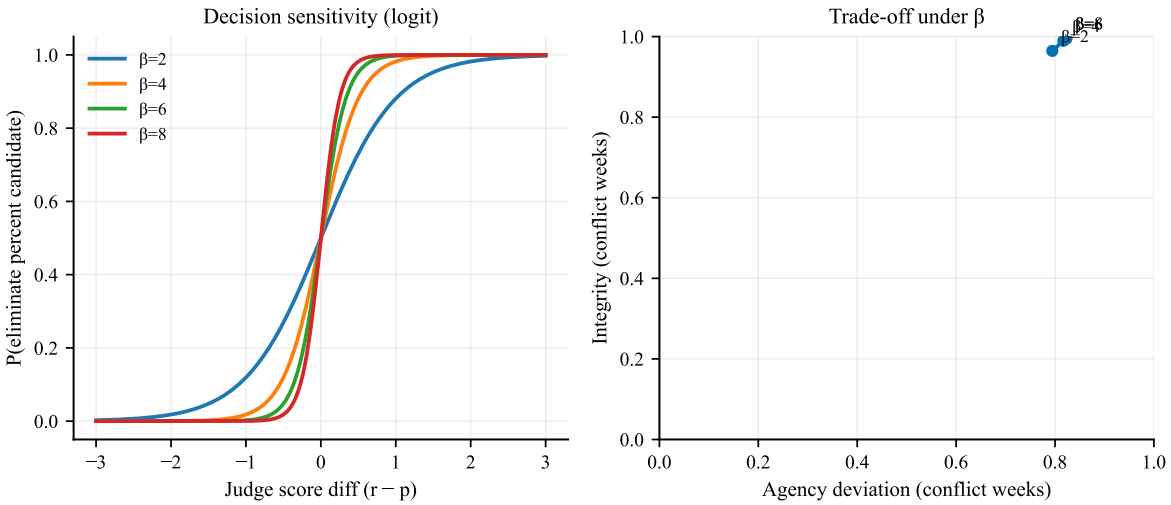


图 20: Judge-save 参数 β 的敏感性（仅冲突周）：logit 决策曲线与完整性—能动性权衡。

9.1 规模对比实验

我们在多进程设置下比较不同采样规模，记录运行时间、误差（均值 HDI 宽度）、稳定性（DAWS）与理论匹配度（Kendall τ）。结果显示误差随规模提升而趋于平缓，图中虚线标注了

拐点与最终规模选择。

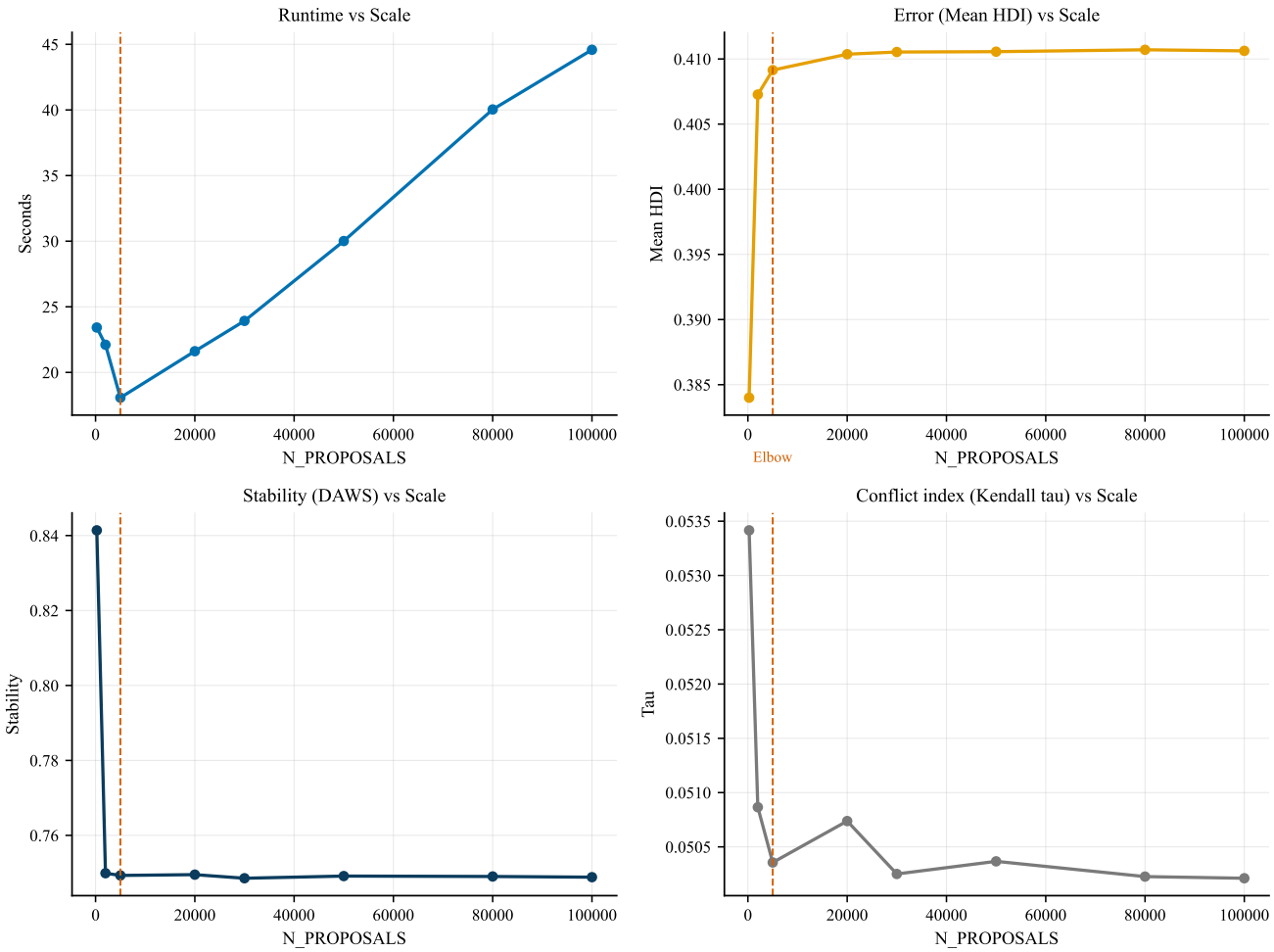


图 21: 不同采样规模下的时间、误差、稳定性与匹配度对比（虚线为折中规模）。

关键输出. 敏感性曲线与后验预测覆盖率。

10 结论与建议

结论要点. 审计先行揭示关键不确定性，DAWS 提供透明权衡。

我们完成全赛季粉丝票审计，量化 Rank 机制的民主赤字，并提出 DAWS 作为透明权衡方案以提升能动性与评委一致性，同时承认稳定性存在小幅代价。

- **可读结论：** 不确定性集中在少数周，其余周可识别性较高。
- **机制影响：** Rank 聚合提高翻转概率；DAWS 提升能动性但稳定性略有代价（见图 11 与图 15）。

- **落地建议：**公布冲突触发与披露规则及 judge-save 规则以提升透明度。

A 敏感性分析

本附录展示平滑参数 σ 的敏感性分析。主要结论在测试范围内保持稳健。

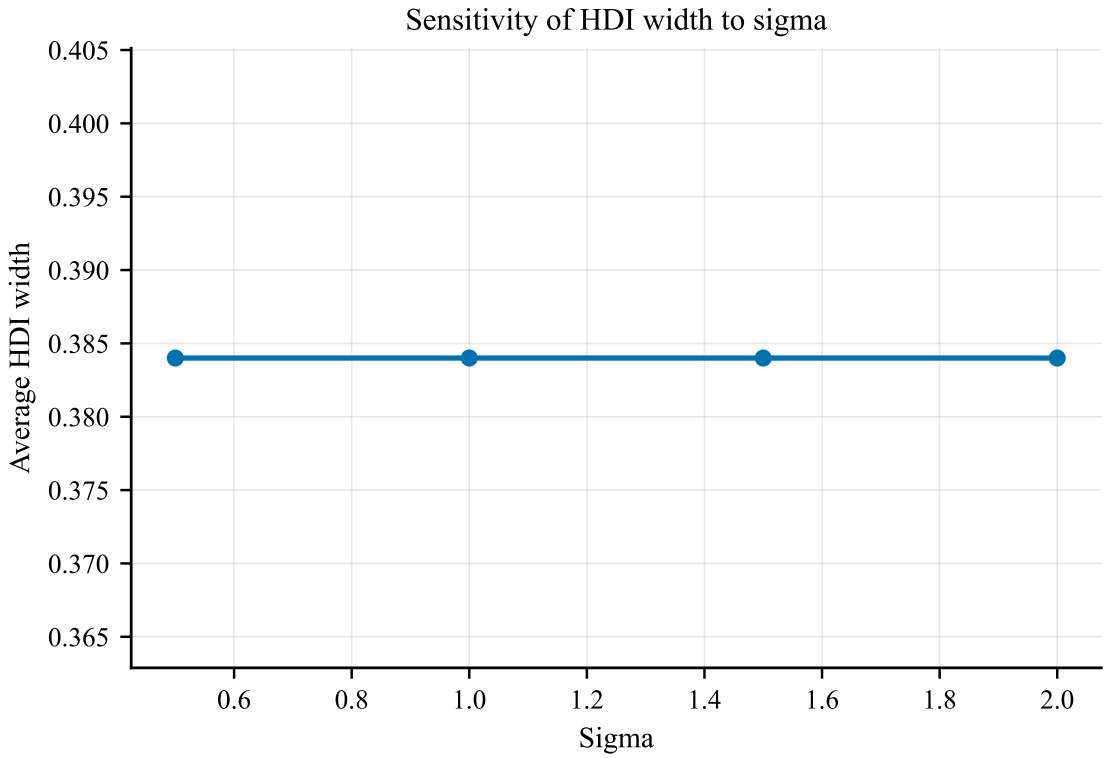


图 22: σ 敏感性：结论对平滑参数较为稳健（横轴为 σ，纵轴为平均 HDI）。

B 预测校准

本附录提供 GBDT 前向链式验证的 AUC 结果，作为协变量有效性的鲁棒性检验。

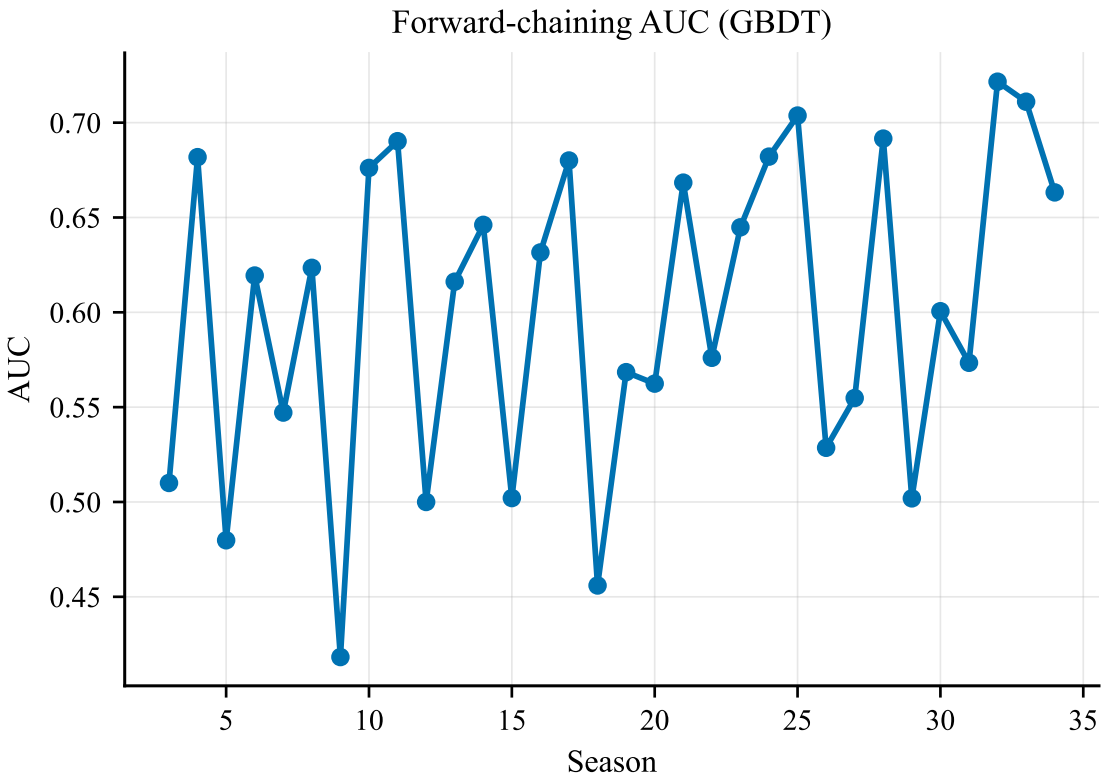


图 23: 分季 AUC 表现稳定（前向链式验证）。

参考文献

[1] COMAP. 2026 MCM/ICM Problem C: Dancing with the Stars (DWTS). Contest Problem Statement.

[2] Smith, R. (1984). Efficient Monte Carlo procedures for generating points uniformly in polytopes. *Operations Research*.

[3] Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*.

[4] Gelman, A., et al. (2013). *Bayesian Data Analysis*. CRC Press.

[5] Moulin, H. (1988). *Axioms of Cooperative Decision Making*. Cambridge Univ. Press.

AI 使用报告

我们使用 AI 协助完成论文结构草稿、LaTeX 模板与方法表述润色；所有模型选择与解释均由团队复核并最终确认。

- 可复现性：代码、图表与指标均由提供数据自动生成。
- 环境：Miniforge + mcm2026，科学计算栈已固定版本。
- 过程留痕：运行日志与汇总指标可追溯每次实验。