Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

Score Optimization and Decision-Making in Competitive Reality Shows Based on Bayesian State Space

Jingjie Zhou

How to cite: ZHOU J. Score Optimization and Decision-Making in Competitive Reality Shows Based on Bayesian State Space. Textile & Leather Review. 2026; 9:2523-2548. https://doi.org/10.31881/TLR.2026.2523

How to link: https://doi.org/10.31881/TLR.2026.2523

Published: 25 April 2026

Score Optimization and Decision-Making in Competitive Reality Shows Based on Bayesian State Space

Jingjie Zhou
School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen, Shenzhen 518000, China
chingchiehchou@163.com

Article
https://doi.org/10.31881/TLR.2026.2523
Published 25 April 2026

ABSTRACT

Addressing the challenges of opaque audience voting data and variable elimination rules in competitive reality shows, this study constructs a quantitative framework integrating latent variable reconstruction, rule auditing, and dynamic optimization. First, a Bayesian state space model is employed to perform latent state reconstruction on audience support rates across thirty-four seasons. Using Laplace approximation techniques to quantify estimation uncertainty, the analysis reveals significantly heightened sensitivity of voting fluctuations to late-stage decision-making. Subsequently, counterfactual simulations under different scoring rules were conducted using the reconstructed data, deeply analyzing the systemic differences between percentage-based and ranking-based systems in addressing expert bias versus public popularity bias. Furthermore, a panel regression model was applied to deconstruct the influence mechanisms of factors like professional dancer background, contestant occupation, and age on advancement probabilities, identifying significant divergences in the driving logic between expert judging and public voting. Finally, addressing the vulnerability of existing mechanisms to noisy data, we propose an uncertainty-aware weighted moving average mechanism. By dynamically adjusting voting weights, this approach simultaneously enhances the robustness and fairness of elimination decisions. Experiments demonstrate that this framework effectively restores latent competitive dynamics, providing a robust algorithmic foundation for optimizing dual-track judging systems.

KEYWORDS
bayesian inference, counterfactual simulation, dynamic weight allocation

INTRODUCTION

In modern competitive reality shows integrating professional judging with public interaction, balancing artistic expertise and audience engagement forms the core design logic. However, persistent opacity in audience voting shares and frequent revisions to elimination rules have fueled public skepticism regarding fairness [1,2].

When technically proficient contestants face elimination due to lack of popularity, systemic strategic imbalances emerge as an urgent issue. Previous studies predominantly focused on static performance evaluation, lacking in-depth quantification of uncertainty under missing data conditions and exploration of rule-evolution mechanisms. This research innovatively integrates Bayesian inference and state space modeling to reconstruct opaque audition data, establishing an integrated framework encompassing state space inversion, dynamic attribution analysis, and multi-objective optimization [3]. This achieves a logical closed loop from latent variable reconstruction to predictive optimization. The overall research approach follows a rigorous data-driven methodology: First, a state space model incorporating random walk priors is established to reconstruct audience voting trajectories. Subsequently, reconstructed data simulate season evolution under different rules to identify specific rule impacts on controversial contestants' fates. Weighted regression models then analyze the influence weights of sociodemographic characteristics [4,5]. Finally, an adaptive weighted scoring mechanism is designed based on estimated uncertainty metrics, aiming to establish a scientific evaluation paradigm for complex dual-track decision systems [6].

A BAYESIAN STATE SPACE RECONSTRUCTION STUDY OF LATENT PUBLIC PREFERENCES

A priori of State Space During the Season

In the design of the state-space prior, we integrate both anchoring and smoothing mechanisms:
Firstly, an anchoring prior is introduced by using the judges' proportion as a weakly informative starting point, setting the prior mean of the latent variable θ to $(r + \varepsilon)$.

$$v_{i,s,t} \sim N(m_{i,s,t}, \tau^2), m_{i,s,t} = \log(j_{i,s,t} + \varepsilon) \tag{1}$$

It should be emphasized that setting the review ratio as the prior mean serves merely as a starting point for the algorithm's convergence. By adjusting the larger variance τ^2, this "weak information" constraint allows the posterior estimate to deviate significantly under the influence of observed data, thereby enabling the quantitative capture of the decoupling between expert reviews and public popularity—which is the core premise for identifying "strategic imbalance" in this study. This ensures that in the absence of observational data, the estimated fan voting proportion aligns approximately with the judges' structure, while still allowing reasonable deviations in the actual data [7-9].

Secondly, a smoothing prior is imposed by assuming that the θ values for the same contestant across adjacent participating weeks follow a random walk prior:

$$v_{i,z,t} - v_{i,t-1} \sim N(0, \sigma_w^2) \text{ (only when } \ i \in A_{z,t} \cap A_{z,t-1}) \tag{2}$$

The variance parameter σ^2 controls the smoothness of the trajectory, thereby suppressing drastic week-to-week fluctuations, enhancing temporal coherence, and automatically avoiding invalid smoothing when a contestant has gaps in participation [10,11].

Observation Layer Model (Elimination Mechanism Modeling)

1 Percent Combination Mechanism (Seasons 3–27): Linear Weighting by Proportions.

$$q_{i,s,t} = (1-\lambda) j_{i,s,t} + \lambda p_{i,s,t} \tag{3}$$

The elimination probability is modeled using softmax, making the tendency for elimination negatively correlated with q.

$$\Pr(y_{s,t} = i \mid p_{s,t}) = \frac{\exp(\beta_{\%} bad_{i,s,t})}{\sum_{k \in A_{s,t}} \exp(\beta_{\%} bad_{k,s,t})} \tag{4}$$

2 Rank Combination Mechanism (Seasons 1–2): Judge Ranking + Fan Ranking. The rank mechanism requires fan rankings, but what is modeled is P. To maintain differentiability and avoid the instability of hard ranking, a soft rank is introduced to approximate the fan ranking:

$$R^F_{i,s,t} \approx \frac{1}{2} + \sum_{k \in A_{s,t}} \sigma\left(\frac{p_{k,s,t} - p_{i,s,t}}{\tau_r}\right) \tag{5}$$

A larger value indicates a worse comprehensive ranking and a higher likelihood of elimination. Similarly, the elimination probability is modeled using softmax:

$$\Pr(y_{s,t} = i \mid p_{s,t}) = \frac{\exp(\beta_r bad_{i,s,t})}{\sum_{k \in A_{s,t}} \exp(\beta_r bad_{k,s,t})} \tag{6}$$

3 For the “bottom two judges choose one” mechanism introduced from Season 28 onward, a two-stage modeling approach is adopted. First, the two candidates for elimination are screened based on the comprehensive ranking. Then, the judges’ decision to eliminate the contestant with the lower score between the two is described using a sigmoid function. Finally, the results from the stages are integrated using softmax [12,13].

$$\Pr(y_{s,t} = i \mid p_{s,t}) \propto w_{i,s,t} \sum_{j \neq i} w_{j,s,t} S_{i \leftarrow j} \tag{7}$$

4 For weeks with multiple eliminations, the likelihood is approximated as the product of the individual probabilities of the eliminated contestants. This preserves information while avoiding the complexity of handling dependencies without replacement.

$$\log \Pr(E_{s,t} \mid p_{s,t}) \approx \sum_{i \in E_{s,t}} \log \Pr(y_{s,t} = i \mid p_{s,t}) \tag{8}$$

The Posterior Inference Stage

The problem is formulated as minimizing the negative log-posterior objective function (i.e., the sum of the prior term and the likelihood term), and the solution is carried out in three steps:

In the posterior inference stage, the prior and likelihood are combined to obtain the posterior distribution of the latent variables (expressed in negative log form for easier optimization) [14].

$$-\log p(v_z \mid D_z) = \underbrace{\sum_{t \in A_v} \frac{(v_{i,z,t} - m_{i,z,t})^2}{2\tau^2}}_{\text{Anchoritem}} + \underbrace{\sum_{i} \sum_{t \to t-1} \frac{(v_{i,z,t} - v_{i,z,t-1})^2}{2\sigma_{rw}^2}}_{\text{SmorthingTerm}} \tag{9}$$

We formulate the problem as minimizing the negative log-posterior objective function (i.e., the sum of the prior term and the likelihood term), and complete the solution in three steps:

First, the maximum a posteriori (MAP) estimate is solved using the quasi-Newton method (L-BFGS), yielding the point estimate of the latent variable θ and the corresponding point estimate of the fan proportion [15,16].

$$\hat{v}_s = \arg\min_{v_s} [-\log p(v_s \mid D_s)] \tag{10}$$

Next, the Hessian matrix is computed at the MAP estimate, and a Gaussian approximation of the posterior distribution is constructed using the Laplace approximation.

$$H_s = \nabla^2_{v_s}[-\log p(v_s \mid D_s)] \mid_{\hat{v}_s} \tag{11}$$

$$\sum_s \approx (H_s + \eta I)^{-1} \tag{12}$$

$$v_s \mid D_s \approx N(v_s, \sum_s) \tag{13}$$

Finally, Monte Carlo sampling is performed based on this Gaussian distribution. The samples are transformed into fan voting proportions (p) , and the 95% confidence intervals for the estimated values of each contestant in each week are derived from the quantiles of these samples:

$$p^{lo}_{i,s,t} = Q_{0.025}(\left\{p^{(b)}_{i,s,t}\right\}), p^{hi}_{i,s,t} = Q_{0.975}(\left\{p^{(b)}_{i,s,t}\right\}) \tag{14}$$

$Q_{\alpha}(\cdot)$ represents the α quantile of the sample p. The interval width serves directly as a quantitative measure of the uncertainty in the voting estimate for that contestant in that week [17,18].

$$w_{i,s,t} = p^{hi}_{i,s,t} - p^{lo}_{i,s,t} \tag{15}$$

Consistency and Uncertainty Index Testing

Consistency Check

Due to the lack of direct labels for the original vote counts, this section conducts a goodness-of-fit check, which verifies the extent to which the reconstructed latent variables can replicate the actual elimination results used to construct the observed data. As shown in Figure 1 and Figure 2, in most weeks, the weak agreement probability ranges between 0.55 and 0.75, indicating that the model can reasonably explain the elimination outcomes. However, there are certain periods with notably low consistency, which are often associated with differences in competition rules, weeks with multiple eliminations, or other specific structural factors, reflecting the model's limitations when dealing with more complex scenarios [19,20].

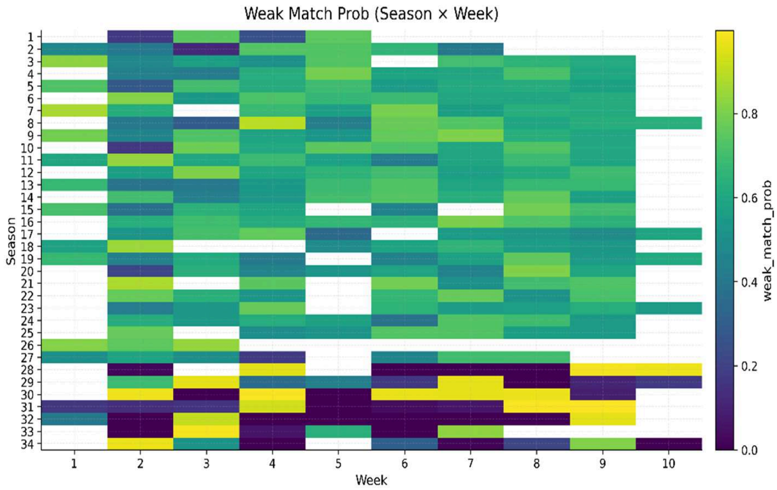

Figure 1. Heat map of weak match probability of season × week

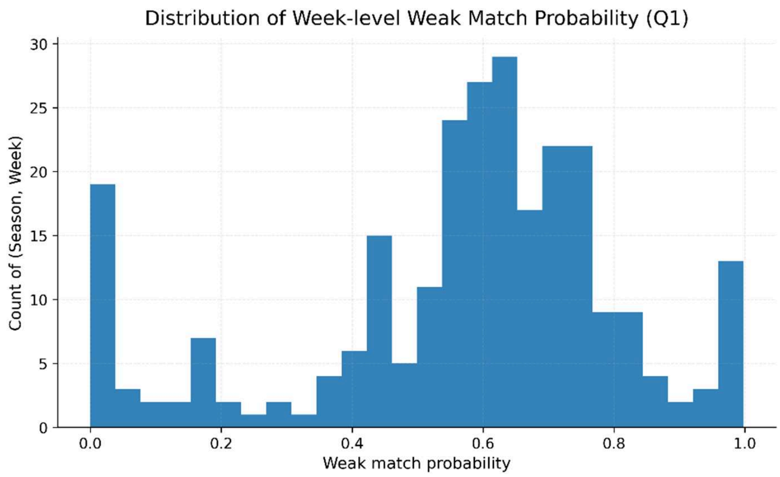

Figure 2. Distribution of weak match probability at weekly level

Changes in Uncertainty

As shown in Figure 3, the uncertainty in model estimates, as measured by the width of confidence intervals, is relatively low in the early stages but increases significantly in the mid-to-late phases, particularly around Weeks 10–11. At the same time, the entropy of voting shares gradually decreases over the weeks, suggesting that fan support shifts from being dispersed to becoming more concentrated, and the competitive landscape becomes clearer-consistent with the narrative logic of the show.

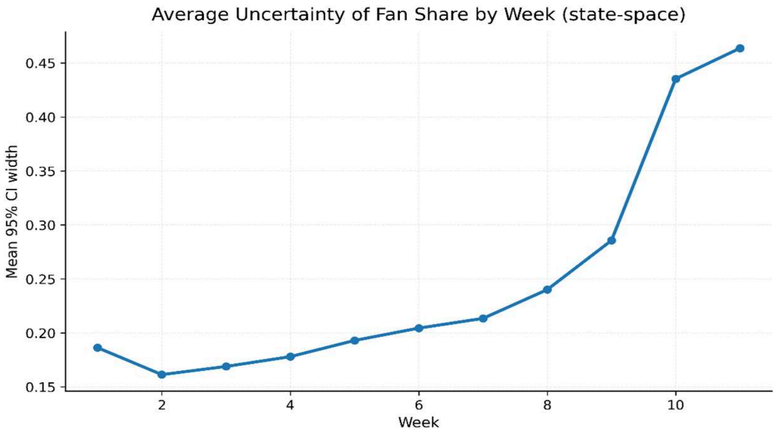

Figure 3. Mean of the uncertainty of fan share estimates for each week

Impact of Different Rules

Different competition rules clearly affect model performance: seasons using the Percent mechanism generally show robust performance, characterized by medium-to-high consistency paired with low-to-medium uncertainty. In contrast, the Rank mechanism is associated with higher uncertainty and only limited improvement in consistency. Meanwhile, the Rank + Save mechanism, which introduces a judge-rescue element, leads to even lower consistency and more dispersed uncertainty, indicating that this rule is relatively less interpretable within the model framework.

Validation of Behavioral Rationality

As shown in Figure 4, validation of behavioral rationality shows a positive correlation between fan voting shares and judges' scores, but the relationship is not a simple replication- significant vertical dispersion exists between the two. Elimination outcomes depend more on whether a contestant's fan share is relatively low; even with decent performance from the judges, insufficient fan support can still lead to elimination. This empirically supports the dual-mechanism design of the show, where both judge evaluations and audience voting jointly determine outcomes.

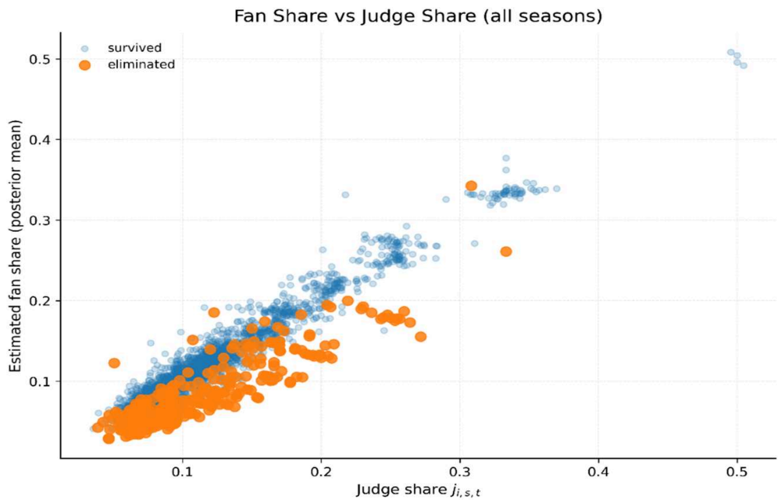

Figure 4. Scatter plot of the relationship between estimated fan share and judge share

COMPETITIVE EVOLUTION SIMULATION AND BIAS AUDIT UNDER DIFFERENT SCORING RULES

Model Establishment

We first designed a controlled experimental framework. Its core principle is to fix the fan vote share as a unified “audience-side” input:

$$p_{i,s,t} \in (0,1), \sum_{i \in A_{s,t}} p_{i,s,t} = 1 \tag{16}$$

Within the same set of active contestants each week, the Percent rule and the Rank rule are applied separately to predict eliminations, ensuring that any observed differences stem solely from the rules themselves.

Percent Rule: This rule computes a combined score by taking a weighted sum of the fan share and the judge share. The contestant with the lowest total score is predicted for elimination. It directly reflects numerical advantage and records the two contestants with the lowest scores as the bottom-two candidate set:

$$\hat{B}_{s,t}^{(\%)} = \arg 2 \min_{i \in A_{s,t}} q_{i,s,t}^{(\%)} \tag{17}$$

where returns the set of indices for the two contestants minimizing q.

Rank Rule: This rule first converts both fan share and judge share into rankings, then transforms these rankings into points which are summed. The contestant with the lowest total points is predicted for elimination. It is insensitive to absolute numerical differences and reflects only relative ordering, also generating a bottom-two candidate set:

$$\hat{B}^{(r)}_{s,t} = \arg 2 \min_{i \in A_{s,t}} S^{(r)}_{i,s,t} \tag{18}$$

Judges' Save Mechanism: In the "bottom-two candidate set" generated by each rule, it is assumed that the judges eliminate the contestant with the lower judge score. This constructs the Percent + Save and Rank + Save variants to assess how the judges' final say moderates the influence of fan votes. It should be noted that this model assumes the judges' decision-making logic is based solely on their scores. Although in actual production, judges may retain "controversial" contestants for narrative or ratings considerations, this framework currently treats such non-technical factors as model constraints due to the lack of quantifiable public data to support them. Future research could explore the use of social media sentiment analysis to further quantify a contestant's "buzz value.".

$$\hat{y}^{(save)}_{s,t} = \arg \min_{i \in \hat{B}_{s,t}} J_{i,s,t} \tag{19}$$

Thus, the elimination predictions for the two variants are obtained: percent + save: $\hat{y}^{(\%+save)}_{s,t}$ predicted by $\hat{B}^{(\%)}_{s,t}$, rank+save: $\hat{y}^{(r+save)}_{s,t}$ predicted by $\hat{B}^{(r)}_{s,t}$ and J.

Fifth, we established a comprehensive evaluation index system to quantify and compare the effects of the rules. This includes:

(1) Elimination Reproduction Accuracy measuring the consistency between the rule's predicted results and historical actual outcomes.

(2) Rule Divergence Rate statistically tracking the frequency with which the two rules generate different predictions.

(3) Bias Proxy Indicator determining rule preference by calculating the gap between the eliminated contestant's "judge rank minus fan rank" (a negative Gap implies elimination is more driven by lack of fan support, i.e., more "fan-favored").

(4) Confidence Interval Analysis examining how the uncertainty in the estimated fan vote shares affects the interpretation of results.

Finally, we constructed a case study framework. For controversial contestants mentioned in the problem statement, we plotted trend curves of their fan vote shares and judges' score shares over the entire season, attaching confidence intervals to the estimated values to display uncertainty. By comparing these curves and identifying the specific weeks where rule divergence occurred, we can intuitively analyze how the specific "divergence pattern between judges' and fans' evaluations" for that contestant could lead to different elimination fates under different rules, thereby explaining the concrete impacts of rule differences at the micro level.

Model Solution and Result Analysis

Contrast in Rule Explanatory Power is Significant.

As shown in Figure 5, the Percent rule demonstrates high adaptability and stability, reproducing the actual elimination results with near-perfect accuracy (close to 1.0) in its native format seasons. In contrast, the Rank rule shows highly volatile reproduction rates (0.37–1.0) under the same conditions. The introduction of the judges' save mechanism further complicates prediction, causing a universal drop in reproduction rates across all rules to a wide range of 0.22– 0.90, indicating that increased format complexity substantially raises prediction uncertainty.

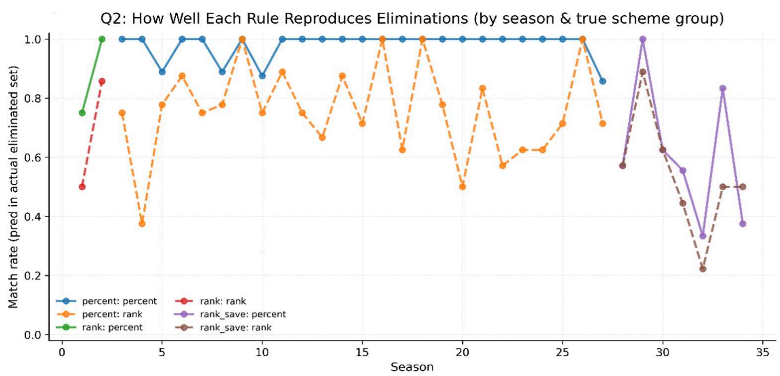

Figure 5. Degree of elimination result reproduction by each rule

Disagreements Exhibit Specific Spatiotemporal Patterns.

Divergences between the two rules predominantly cluster in the early to mid-season weeks (Weeks 2–8), a period characterized by dense competition where minor differences in data processing between rules are amplified. While the judges' save mechanism reduces the overall frequency of disagreements, it does not eliminate them entirely, signifying that rule differences can still indirectly influence the final outcome by altering the composition of the “bottom-two” candidate pool.

Rule Bias is Quantitatively Confirmed.

As shown in Figure 6, analysis using the “judge rank minus fan rank” difference (Gap) for predicted eliminations reveals distinct biases: the Percent rule (median Gap) clearly favors fan sentiment; the Rank + Save rule (median Gap $Gap \approx +0.5$) reinforces judges’ authority; while the pure Rank and Percent + Save rules exhibit a near-neutral bias with Gaps close to 0.

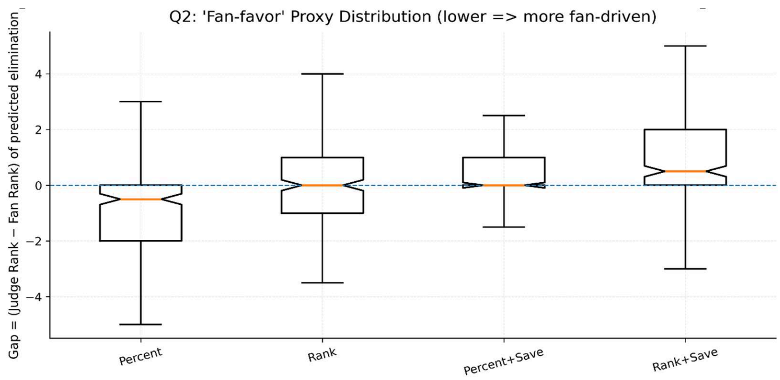

Figure 6. Distribution of fan-favor proxy indicators

Case-Specific Pathways are Clearly Illustrated.

As shown in Figure 7, Jerry Rice (S2) benefited under the Percent rule due to his rising fan share in the latter stages (reaching ~0.26). As shown in Figure 8, Billy Ray Cyrus (S4), with closely matched fan and judge shares, resided in a rule-sensitive zone. As shown in Figure 9, Bristol Palin (S11) enjoyed greater protection under the Percent rule thanks to her sustained popularity lead (fan share 0.19–0.33). As shown in Figure 10, Bobby Bones’ (S27) trajectory, defined by a persistent but modest fan advantage, was highly dependent on how different rules processed these subtle disparities.

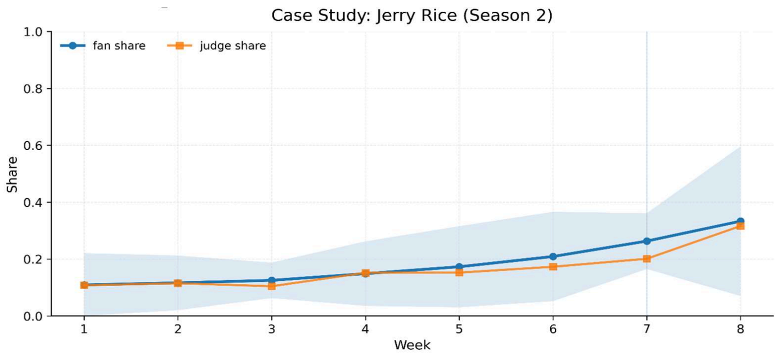

Figure 7. Case study results - Jerry Rice

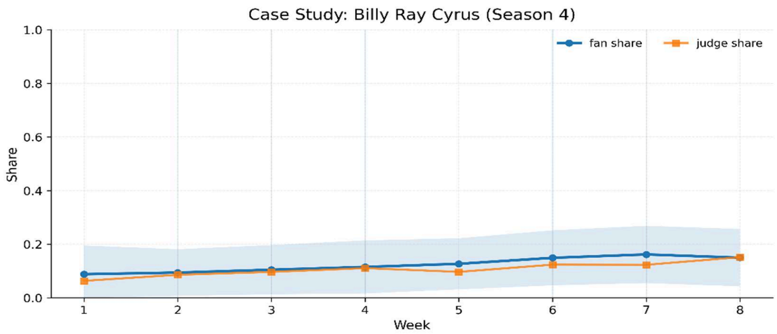

Figure 8. Case study results - Billy Ray Cyrus

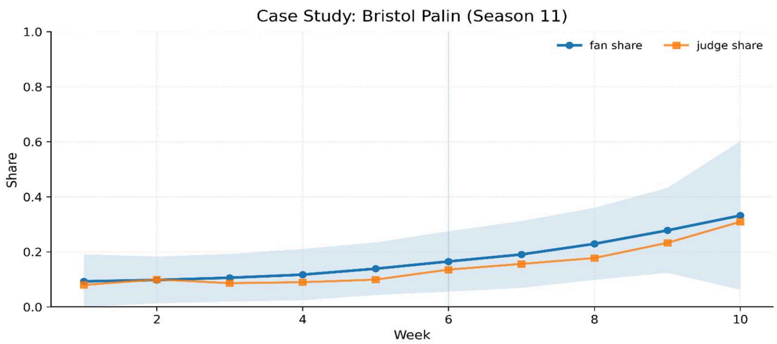

Figure 9. Case study results - Bristol Palin

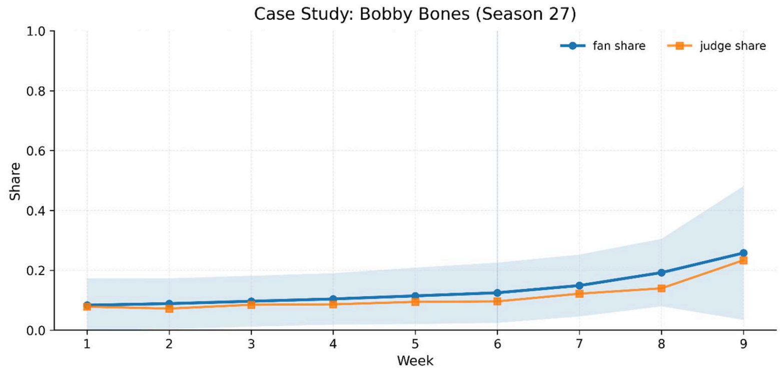

Figure 10. Case study results - Bobby Bones

Systematic Patterns and Global Impact.

As shown in Figure 11, controversial contestants are consistently concentrated in a characteristic zone of "mid-to-low judge scores (18–24) coupled with high fan votes (1.3–2.0)". Historical data reveals even more extreme cases in early seasons (e.g., Season 1) with controversy indices as high as 4.0, confirming that this evaluative misalignment is a structural phenomenon. Consequently, the choice of rules directly dictates the fate of such contestants, thereby fundamentally shaping the narrative arc and competitive outcome of an entire season.

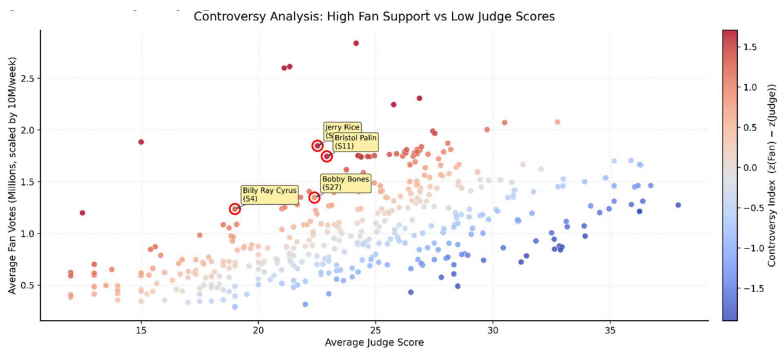

Figure 11. Scatter plot of high fan support vs low judge scores

MECHANISM ANALYSIS OF HOW COMPETITIVE PERFORMANCE AND SOCIAL BACKGROUND CHARACTERISTICS INFLUENCE PROMOTION OUTCOMES

Model Establishment

In the data integration and variable construction phase

The study utilized the "contestant–week" characteristic table as the foundational dataset, integrating judges' scores, competition status, and the fan vote estimates, merged by the triple:

$$(s,t,i) \Rightarrow (\overline{J}_{i,s,t}, p_{i,s,t}, w^{CI}_{i,s,t}, \text{covariates }) \tag{20}$$

To ensure model applicability and robustness, a logit transformation was applied to the fan support measure to mitigate boundary effects:

$$F_{i,s,t} = \log \frac{p_{i,s,t}}{1-p_{i,s,t}} \tag{21}$$

The width of the confidence interval from the inferred results was used to construct weights for weighted least squares (WLS) in order to address heteroskedasticity. Concurrently, continuous variables were standardized:

$$z(x) = \frac{x - E[x]}{\sqrt{\text{Var}(x) + \varepsilon}} \tag{22}$$

Additionally, low-frequency categories-such as industry, region, and professional dancer status-were consolidated to avoid overfitting and enhance model stability. Categories with fewer than 15 observations in industry were grouped as “Other,” those with fewer than 20 in region as “Other,” and professional dancers with fewer than 25 observations (in weekly models) or 20 (in season-level models) as OtherPro.

The consolidation of the low-frequency range here is intended to ensure the statistical robustness of the panel regression model when estimating parameters and to avoid overfitting. This does not contradict the subsequent reference to “extreme cases” from earlier seasons (such as Season 1): regression analysis focuses on identifying general patterns across seasons, while case studies aim to demonstrate the model’s diagnostic performance regarding its limits when dealing with structural biases; the two approaches complement each other by examining macro trends and micro mechanisms, respectively.

The weekly panel model (consists of two sub-models)

Model A uses the average score per judge as the dependent variable to examine the effects of week, age, industry, region, professional dancer status, and other factors on judges’ scores, while controlling for season fixed effects and professional dancer fixed effects. “Season– contestant” clustered robust standard errors are applied to account for panel correlation.

The weekly judge-scoring model is specified as:

Model B uses the log-odds ratio of fan support as the dependent variable and introduces the current week's judge score as a performance control among the independent variables:

Weighted least squares (WLS) is further employed to correct for inversion uncertainty, thereby isolating the net effect of professional dancers on fan support.

Season-Level Outcome Model

By aggregating each contestant's weekly data within a season, the model constructs average judge score and average fan-support intensity as core features.

$$\overline{J}_{i,s} = \frac{1}{n_{i,s}} \sum_t \overline{J}_{i,s,t}, \overline{F}_{i,s} = \frac{1}{n_{i,s}} \sum_t F_{i,s,t} \tag{23}$$

Two outcome variables are defined: rank percentile and survival ratio, reflecting final ranking and the ability to remain in the competition, respectively.

$$place_{pctt_{i,s}} = \frac{placement_{i,s} - 1}{N_s - 1 + \varepsilon} \tag{24}$$

On this basis, regression models are established to systematically evaluate the independent and combined effects of technical performance, fan support, and professional dancer status on contestants' seasonal outcomes, thereby revealing the mechanism of each factor from a results-oriented perspective.

$$place_{pctt_{i,s}} = \alpha_P + \beta_{P,J} mean_J avg_z z + \beta_{P,F} mean_f anlogitz + \beta_{P,a} age_z \tag{25}$$

$$\begin{aligned} survival_{ratio_{i,s}} =& \alpha_S + \beta_{S,J} mean_J avg_z + \beta_{S,F} mean_f anlogit_z + \beta_{S,a} age_z \\ &+ \theta_S^u 1\{industry2\} + \phi_S 1\{region2\} + \psi_S^{ton} 1\{season\} + \eta_S 1\{proj2\} + \varepsilon_{i,s}^S \end{aligned} \tag{26}$$

Clustered robust covariance method

To ensure the reliability of statistical inference, the study employs a clustered robust covariance method (i.e., the "sandwich" estimator) in both the weekly and season-level models. For the weekly model, clustering is performed by "season–contestant" to control for within-individual correlation across weeks, while the season-level model is clustered by "season" to absorb common shocks within the same season.

$$\widehat{\mathrm{Var}}(\hat{\beta}) = (X^{\mathrm{ton}}X)^{-1}(\sum_{g} X_g \hat{u}_g \hat{u}_g^{\mathrm{ton}} X_g)(XX)^{-1} \tag{27}$$

This approach allows for arbitrary forms of heteroskedasticity and autocorrelation within clusters, resulting in more conservative standard errors for parameter estimates and enhancing the statistical robustness of the conclusions.

Model Solution and Result Analysis and Verification

The determinants of fan support primarily include the time accumulation effect, the performance level in the given week, and structural identity differences. After controlling for weekly performance, certain occupational and regional categories still show significant effects, indicating the presence of systematic influence stemming from pre-existing popularity or cultural background differences. Fan popularity model is shown in figure 12.

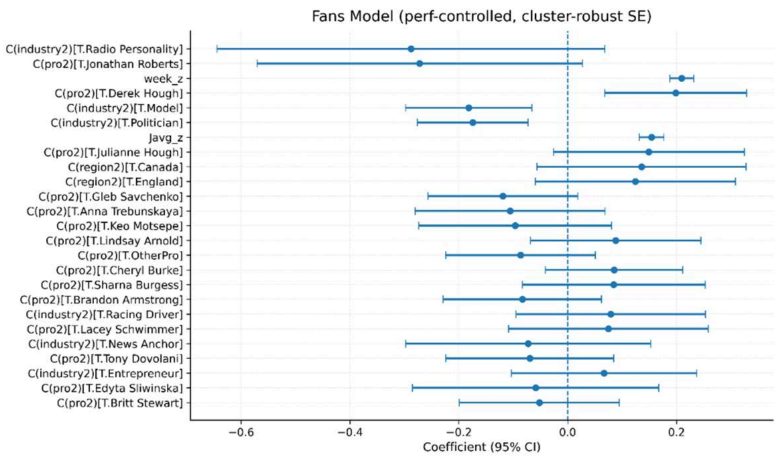

Figure 12. Fan popularity model

The influence of professional dancers on judges' scores and fan support exhibits clear mechanistic differences. Regarding judges' scores, the effects vary significantly among different professional dancers, with some showing positive and significant contributions. In contrast, their impact on fan support is more dispersed and inconsistent in direction. This suggests that judges focus more on technical execution, while fan voting is more easily driven by factors such as narrative expression, interactive appeal, and emotional connection. Pro fixed effects for judges' scores and Impact of professional dancers on fan popularity are shown in figure 13 and 14.

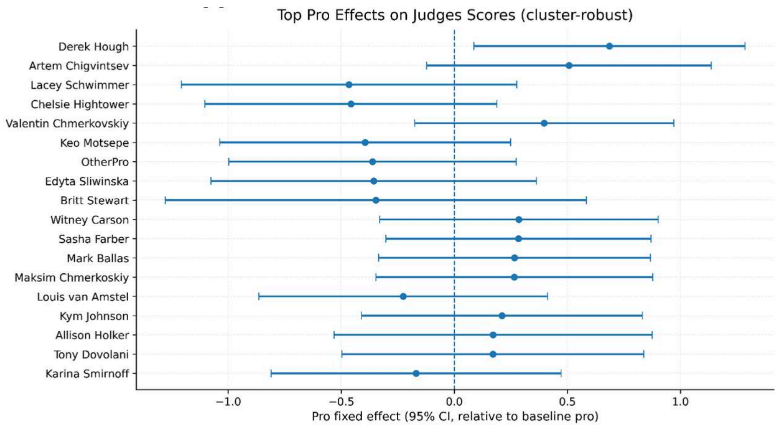

Figure 13. Pro fixed effects for judges' scores

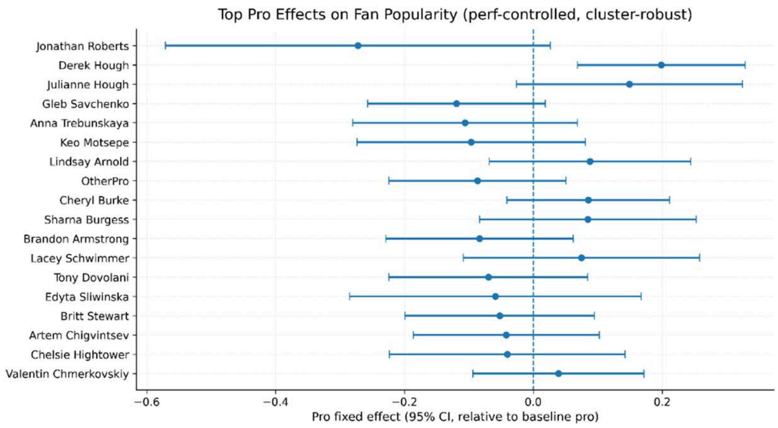

Figure 14. Impact of professional dancers on fan popularity

In terms of the advancement/elimination and final ranking mechanisms, both fan support and judges' scores have a significantly positive impact on contestants' survival ratio. Fan support also significantly affects the final ranking, although its mechanism differs somewhat from the elimination stage. Comprehensive analysis indicates that fan support plays a direct protective role in "avoiding elimination," while the final ranking is shaped by a combination of multiple factors including fan support, technical performance, and competitive intensity as the season progresses. Survival ratio model with cluster-robust SE and Placement percent model with cluster-robust SE are shown in figure 15 and 16.

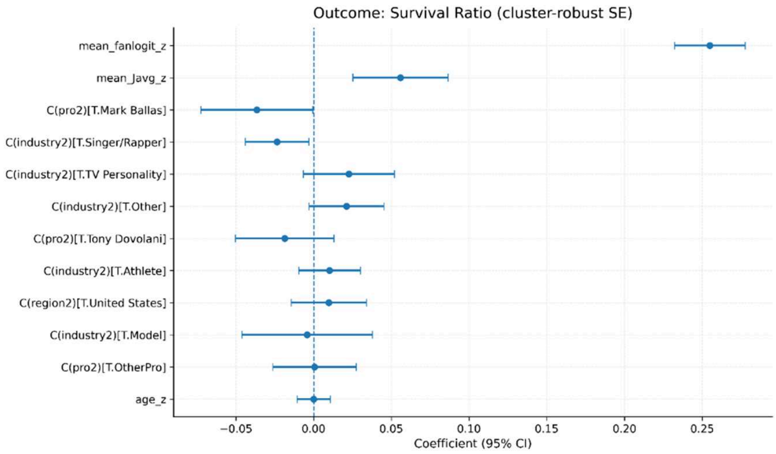

Figure 15. Survival ratio model with cluster-robust SE

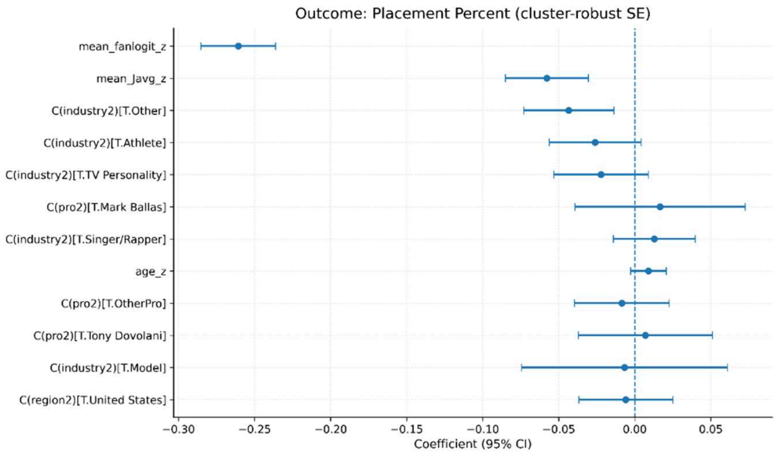

Figure 16. Placement percent model with cluster-robust SE

OPTIMAL DESIGN OF DYNAMIC WEIGHTED ELIMINATION MECHANISMS BASED ON UNCERTAINTY PERCEPTION

Model Establishment

This paper proposes a weekly elimination synthesis mechanism named “Uncertainty-Aware Weighted Moving Average” (UAW-MA), designed to enhance the robustness and fairness of the elimination process. The mechanism operates on basic units of season, week, and contestant. Inputs include the set of active contestants each week, each contestant’s share derived from judge scores, and the posterior mean of fan vote share-estimated via historical data inversion- along with its confidence interval width. The actual number of eliminations per week is used as the basis for output scale. Through the weekly uncertainty aggregation formula, the estimated uncertainty of audience voting for individual players is aggregated to the weekly level:

$$W_t = \frac{1}{\mid S_t \mid} \sum_{i \in S_i} Cl\_width(p_{i,t}) \tag{28}$$

At the heart of the model is a dynamic, uncertainty-aware weighting of fan votes.

$$W_t^{(norm)} = \frac{W_t - \min_{t \in T_s} W_t}{\max_{t \in T_s} W_t - \min_{t \in T_s} W_t} \tag{29}$$

This weight is not fixed but calculated based on the width of the confidence interval for the estimated fan share each week. Specifically, the average uncertainty at the weekly level is first aggregated and then calibrated within the season to obtain a typical uncertainty level and fluctuation scale. A logistic mapping function is then used to convert the normalized uncertainty into a week-specific fan weight bounded by preset upper and lower limits. This ensures that when the fan vote signal is highly uncertain, its weight is automatically reduced to minimize noise interference. Map the standardized uncertainty to week-specific audience voting weights through the dynamic audience weight formula:

$$\lambda_t = \lambda_{\min} + \frac{\lambda_{\max} - \lambda_{\min}}{1 + \exp(\kappa \cdot (W_t^{(norm)} - \theta_0))} \tag{30}$$

Using the dynamically determined fan weight each week, the model calculates a composite score for each contestant, which is a weighted average of the judge share and the fan share. To further smooth weekly fluctuations and introduce performance inertia, a moving average is applied to the composite scores, resulting in a smoothed score used for final ranking.

The elimination rule strictly follows a “weekly independent evaluation” principle. Each week, the model ranks only the active contestants based on their smoothed scores and eliminates the lowest-ranked contestants according to the actual number of eliminations that occurred that week. For comparison, a baseline mechanism using a fixed fan weight and no smoothing- simulating a traditional percentage-based approach-is run in parallel to measure changes introduced by the new mechanism.

To evaluate the robustness and decision clarity of the new mechanism, two types of metrics are constructed. The first is the “divergence rate” between the elimination sets of the new mechanism and the baseline each week, measuring relative volatility. The second is the “boundary margin,” defined as the score gap between the last eliminated contestant and the first safe contestant. A larger margin indicates a clearer elimination decision that is less sensitive to minor disturbances.

Finally, the model aggregates the metrics over the entire season and performs a grid scan of key parameters. This yields a trade-off frontier between “average divergence rate” and “average boundary margin,” providing decision-makers with a visual basis for selecting mechanism parameters that achieve the best balance between robustness and decision clarity.

Model Solution and Result Analysis and Verification

Temporal and Cross-Season Variation in Dynamic Fan Weight

Figure 17 presents the estimated weekly fan weight λt under the uncertainty-aware weighting framework for each season. Overall, λt remains at a relatively high level (approximately 0.50–0.65) in the early weeks of most seasons, then exhibits a systematic decline as the season progresses, eventually converging to a lower range (approximately 0.25–0.35) in later weeks. This pattern indicates that as the competition advances, the model tends to reduce the relative influence of fan votes on elimination predictions, giving greater weight to judge scores instead. Instead, we assign a higher weight to the judges’ scores. Although common sense suggests that fan engagement is higher in the later stages, the uncertainty analysis in this study (see Figure 19) indicates that noise in popularity signals and valuation volatility also surge significantly during this period. By dynamically reducing the weight of later-stage fan engagement, the UAW-MA mechanism effectively serves as a “risk-mitigation” strategy, designed to ensure that decisions regarding the finals rankings are not swayed by high-noise popularity signals, thereby striking an algorithmic balance between maximizing audience interaction and maintaining the integrity of the competition. Cross-season comparisons show that while peak levels in early weeks vary across seasons, the overall “starts higher and declines over time” trend is highly consistent across the sample. This suggests the dynamic weighting mechanism is not driven by a few outlier seasons but rather reflects systematic adjustments prompted by evolving information structures during the competition.

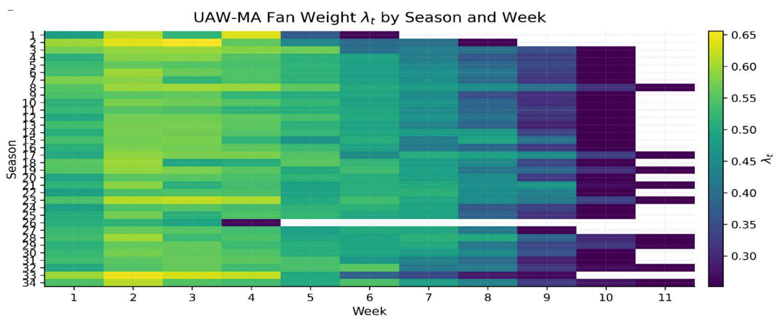

Figure 17. Heatmap of fan weight by season and week

Figure 18 supplements the above conclusion from a distributional perspective: the values of λt are primarily concentrated in the medium-to-high range (approx. 0.48–0.58), with a notable cluster near the lower bound (around 0.25). This implies that fan weight maintains strong influence during most "regular" weeks. However, in certain situations-typically corresponding to later season stages or weeks with significantly increased uncertainty-the weight is compressed toward the lower bound. This prevents the model from over-relying on fan signals in highly uncertain environments.

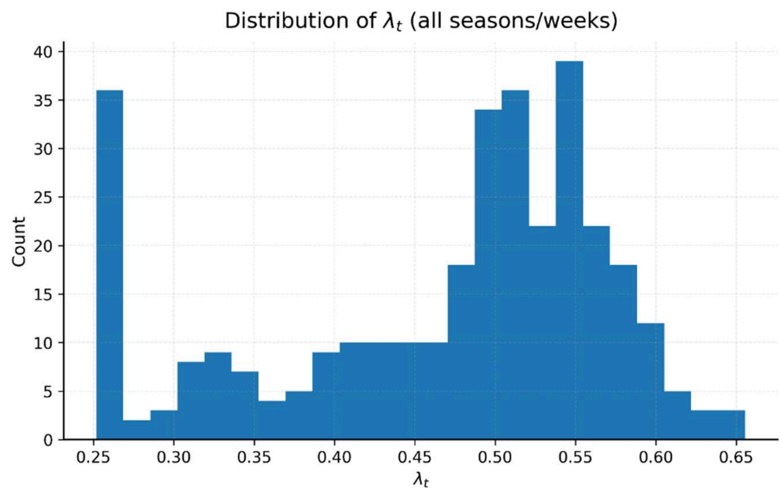

Figure 18. Histogram of fan weight distribution

Correspondence Between Weight Adjustment and Uncertainty

As shown in Figure 19, the change in weight is closely tied to the uncertainty in estimating fan share, showing a significant negative correlation. When the uncertainty of the fan signal is low, λt remains high. Once uncertainty rises, λt declines rapidly and stabilizes near a conservative lower limit after reaching a certain threshold. This nonlinear adjustment pattern- "sharp initial decline followed by plateau convergence"-

reflects the model's risk-control logic: proactively suppressing the weight of noisy signals in high-uncertainty environments to prioritize the overall robustness of elimination decisions.

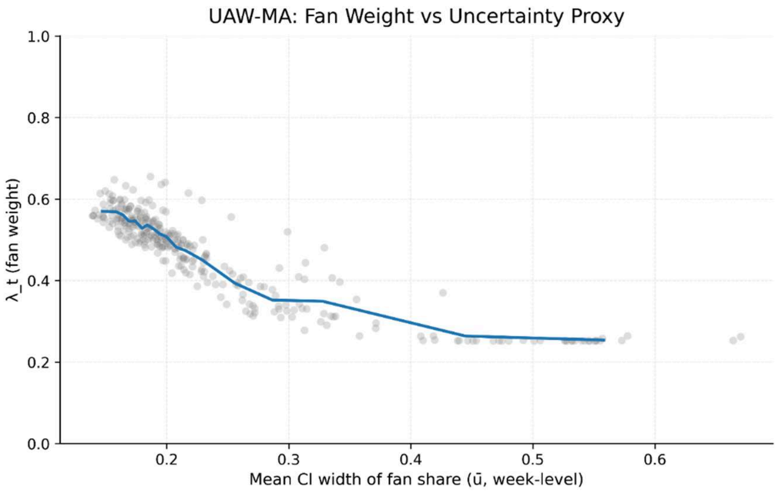

Figure 19. Relationship between fan weight and uncertainty proxy metric

Divergence Structure Relative to Baseline and Elimination Boundary Risk

As shown in Figure 20, comparison with the fixed-weight baseline mechanism shows that the new mechanism does not overturn the baseline's elimination decisions in the vast majority of weeks. High divergence rates are concentrated only in specific weeks of certain seasons. These high-divergence weeks often coincide with smaller "elimination boundary margins," meaning the elimination decision is borderline. This indicates that the scenarios where dynamic weight adjustment produces a substantive impact are precisely those "high-risk" weeks where scoring and voting signals are ambiguous, and elimination outcomes are highly sensitive to minor perturbations.

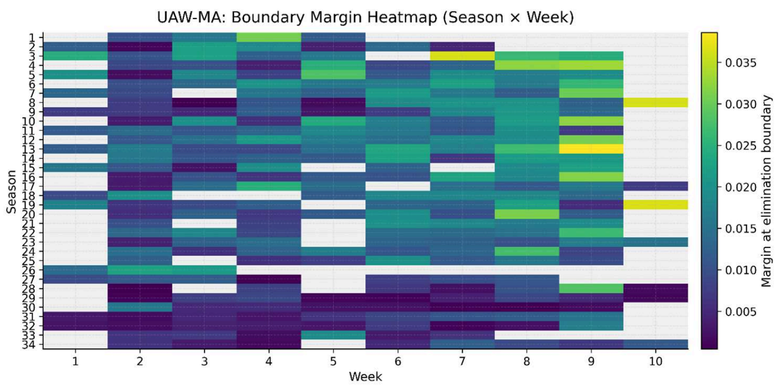

Figure 20. Heatmap of elimination boundary margin

Case Study: Dynamic Weight Trajectory and Divergence Performance in Season 10

As shown in Figure 21, a case study of Season 10 further reveals the mechanism's robustness. Although its fan weight λt declines monotonically week by week, the season's elimination outcomes are almost entirely consistent with the baseline. This shows that the adjustment of dynamic weight primarily reflects the model's "internal calibration" of information reliability. When the elimination boundary is clear or judge and fan signals are aligned, changes in weight do not necessarily alter the final output. This case underscores the model's core strength: maintaining coherence with an established baseline in most instances while preserving the capacity for robust corrections during key, contentious weeks.

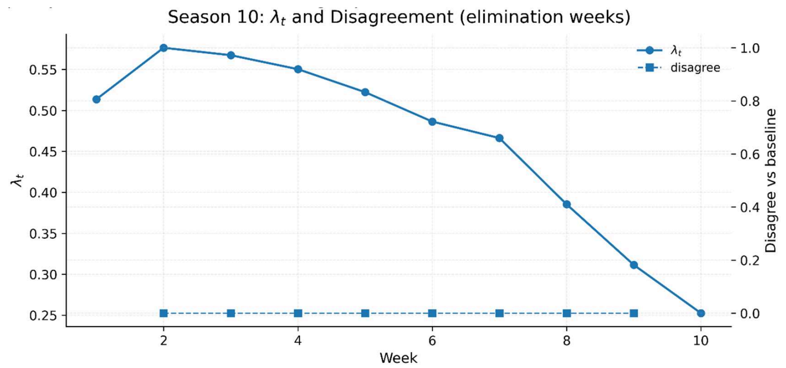

Figure 21. Trend of fan weight and baseline disagreement in season 10

CONCLUSIONS

By integrating Bayesian state space inference, counterfactual simulation, and weighted panel regression, this study systematically reveals the operational patterns of opaque evaluation factors and their impact on competitive dynamics. The research not only successfully reverse-engineered undisclosed audience support rates but also demonstrated the significant advantages of dynamic weight adjustment in suppressing evaluation noise and protecting technical excellence. However, limitations remain: due to the strict confidentiality of absolute vote counts, while the reconstructed model highly simulates elimination outcomes, it lacks direct verification of absolute vote numbers. Additionally, the model's ability to capture transient, non-stationary popularity surges requires improvement. Furthermore, current random walk priors exhibit limited sensitivity to the "explosive" surges in popularity driven by social media. Future research could explore the incorporation of non-stationary stochastic processes or jump diffusion models to capture the nonlinear fluctuations in fan support more accurately. Future research should focus on developing more refined nonlinear prior functions and extending this heterogeneous decision-making framework to real-world scenarios involving expert-public interactions, such as academic peer review and public policy consensus-building.

Author Contributions

Jingjie ZHOU: Writing - original draft; Writing - review & editing; Methodology; Project administration; Software; Formal analysis; Investigation; Resources; Conceptualization; Visualization.

Conflicts of Interest

The author declares no conflict of interest.

Funding

This research received no external funding.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

Not applicable.

REFERENCES

[1] Ning Han, Yifei Li, Zhiyuan Zhang, Jiabao Gong, Bolin Zhang, Yanfeng Li. A Bayesian optimization stacked ensemble learning method: Predicting ash content in flotation cleaned coal using X-ray fluorescence oxides data. Minerals Engineering. 2026; 242:110227-110227. doi:10.1016/J.MINENG.2026.110227.

[2] Nanxiang Guo, Weiyi Cong, Yikun Su, Shoujian Zhang, Qingpeng Man. Bayesian-optimized CNN-LSTM for residential building energy consumption prediction with SHAP explainability. Case Studies in Thermal Engineering. 2026; 81:107870-107870. doi:10.1016/J.CSITE.2026.107870.

[3] Helaleh Khoshkam, Mohammad Valipour, Sayed M. Bateni, Changhyun Jun, Dongkyun Kim, Jonathan L. Deenik, Masoud Karbasi, Tongren Xu. Forecasting daily reference evapotranspiration under different hydrological conditions using a hybrid wavelet–Bayesian optimization–Gaussian process regression model. Results in Engineering. 2026; 30:110004-110004. doi:10.1016/J.RINENG.2026.110004.

[4] Jiayu Yan, Shengwu Qin, Wendi Rao, Chaobiao Zhang, Jiangfeng Lv, Zhijun Lin, Feng Wan, Jiasheng Cao. Landslide susceptibility assessment based on Bayesian optimization and geographical similarity. Natural Hazards. 2026; 122(7):317-317. doi:10.1007/S11069-026-08055-3.

[5] Lee K, Yoo Y, Mun D. Point Cloud Registration Using PCA and Hyperparameter Optimization. International Journal of Control, Automation, and Systems. 2026; (prepublish):1-17. doi:10.1007/S12555-026-00059-W.

[6] Qinchi Yue, Kun Bu, Ruzhi Wang, Changhao Wang. The study of hydrogen adsorption-induced topological surface state in-out hop in MgB2 nodal-line semimetals via physics-informed Bayesian optimization. Physical chemistry chemical physics: PCCP. 2026. doi:10.1039/D5CP03713A.

[7] Haitao Wang, Bowen Chi, Wenhua Li, Jisheng Chen, Shiwei Ge, Qian Zhao, Tianyi Guo. Research on performance evaluation methods for silver-based electrical contact materials based on multi-dimensional parameter fusion and micro-morphology verification. Measurement Science and Technology. 2026; 37(10):106203-106203. doi:10.1088/1361-6501/AE4CAC.

[8] Meng F, Qiao S, Chen D. A Data-Driven framework for precise geometric measurement of tunnel structures using 3D point clouds and Bayesian optimization. Measurement. 2026; 271:120931-120931. doi:10.1016/J.MEASUREMENT.2026.120931.

[9] Giacomo Delcaro, Johanna Menn, David Stenger, Sergio M. Savaresi, Sebastian Trimpe, Simone Formentin. Twin-in-the-Loop observer tuning via gradient-information Bayesian optimization with line search. Control Engineering Practice. 2026; 172:106891-106891. doi:10.1016/J.CONENGPRAC.2026.106891.

[10] Guo Dongwei, Zhu Yingming, Chen Yulei, Zhang Yao. An Optimization Model for Evaluating Large-Scale Innovative Competitions Based on Cross-Entropy Weighting. Chinese Journal of Management Science. 2025; 33(12):134-145. doi:10.16381/j.cnki.issn1003-207x.2023.1775.

[11] Liu Weijiang, Liu Bingqi, Li Xuan. Transformation of Promotion Evaluation for Local Officials: From the “GDP Championship” to the Competition for High-Quality Development. Journal of Central University of Finance and Economics. 2025; (10):5-25. doi:10.19681/j.cnki.jcufe.2025.10.002.

[12] Yao Jian, Long Yue. The Reform Process, Current Status, and Optimization Directions of the CBA Competition System. Journal of Shanghai University of Sport. 2025; 49(05):25-38. doi:10.16099/j.sus.2024.02.29.0008.

[13] Yi Quanyong, Wu Xiaoli, Luo Shengquan. Practical Challenges and Optimization Pathways for Disciplinary Competitions in Empowering the Cultivation of Top-tier Innovative Talent. Research on Educational Development. 2024; 44(24):9-19. doi:10.14121/j.cnki.1008-3855.2024.24.002.

[14] Mao Zhangyong, Feng Siming. Current Status and Optimization Strategies for Vocational Skills Competitions in Archival Work in the New Era: Based on a Survey of 31 Provinces in China. Zhejiang Archives. 2024; (05):17-20+25. doi:10.16033/j.cnki.33-1055/g2.2024.05.018.

[15] Xia Jing, Li Zan, Li Xin. Practical Deviations and Optimization Pathways of the Competitive Sports Competition System with Chinese Characteristics. Research in Physical Education. 2024; 38(02):39-51. doi:10.15877/j.cnki.nsic.20240201.001.

[16] Lu Kongbao, Duan Zhenhua, Zhou Changquan. Exploration and Practice of Deep Integration Between Disciplinary Competitions and Provincial Experimental Teaching Demonstration Centers. Laboratory Research and Exploration. 2023; 42(06):150-154. doi:10.19927/j.cnki.syyt.2023.06.030.

[17] Bao Mingxiao, Zhao Yilong, Gao Ying. A Study on Strategies for Expanding and Optimizing China's Sports Competition System in the Context of the COVID-19 Pandemic. Sports Science. 2022; 42(11):11-18. doi:10.16469/j.css.202211002.

[18] Yuan Weiwei, Sun Chao, Liu Li, Yi Weijing. Fabric Optimization and Pressure Comfort Evaluation of Freestyle Skiing Competition Apparel. Wool Textile Science and Technology. 2022; 50(02):8-14+7. doi:10.19333/j.mfkj.20211106507.

[19] Liu Yongqing, Chen Qi, Shen Yanfei, Sun Lixin, Xiao Shuming, Wu Yiwei, Dong Qinglai, Huang Xifa, Zhen Qingkai. A Study on Ergonomic Evaluation Indicators for Freestyle Skiing Helmets Based on the AHP Method. Journal of Beijing Sport University. 2022; 45(01):56-75. doi:10.19582/j.cnki.11-3785/g8.2022.01.006.

[20] Xia Jing, Li Zan, Tan Yangyang. The Development Process, Model Optimization, and Prospective Considerations of China's Competitive Sports System Reform. Journal of Tianjin University of Physical Education. 2022; 37(01):80-86. doi:10.13297/j.cnki.issn1005-0000.2022.01.012.