Problem Chosen	2026 MCM/ICM Summary Sheet	Team Control Number
C		2630989

A Constraint-Consistent Bayesian Framework for Modeling Elimination and Victory Dynamics in Dancing With The Stars

Summary

This study focuses on the elimination mechanism of the "Dancing with the Stars" program, aiming to address the core challenge that fan voting data has never been made public. The research regards fan voting as a latent variable through the Bayesian method, conducts reverse modeling with a constraining consistent generative model, reconstructs the implicit distribution of fan voting shares, and then compares the differences between the ranking method and the percentage method, analyzes the influence of professional dance partners and celebrity characteristics, and proposes an improved voting system scheme.

For Question one, the study regards fan voting as a latent variable through the Bayesian method and uses a constraint-consistent generative model to solve the constraint problem where the total share is 1. The prior distribution adopts a multivariate normal distribution with a mean of zero, and the likelihood function is constructed based on the elimination constraint. The verification results show that the consistency score exceeds 0.9. The model effectively captures the intrinsic connection between fan voting and elimination results.

Regarding question two, the research found that the ranking method compresses the judges' scores into ranking positions, systematically amplifying the influence of fans. The percentage method retains the fractional scale and maintains the independent weight of technical evaluation. This difference directly explains the survival paths of controversial cases such as Jerry Rice, Bristol Palin, and Bobby Boons..

Focusing on the third question, this study uses an econometric framework to examine heterogeneous impacts of non-performance factors on judges' scores and fans' votes. Results reveal a systematic preference divergence: judges prefer younger contestants and actors, while fans are loyal to athletes with higher age tolerance.

For Question Four, employing 34-season data, this study develops an adaptive variance-weighted algorithm to dynamically adjust weights by the dispersion of judges' scores, strengthening fan influence or technical performance accordingly. A revival round enables recovery from temporary setbacks, and multi-week finals guarantee consistent performance evaluation. The hybrid percentage-judge save scheme balances technical precision and fan participation.

keywords:Bayesian Framework Latent Fan Vote Reconstruction Voting Aggregation Mechanism Judges' Save Mechanism

Contents

1 Introduction

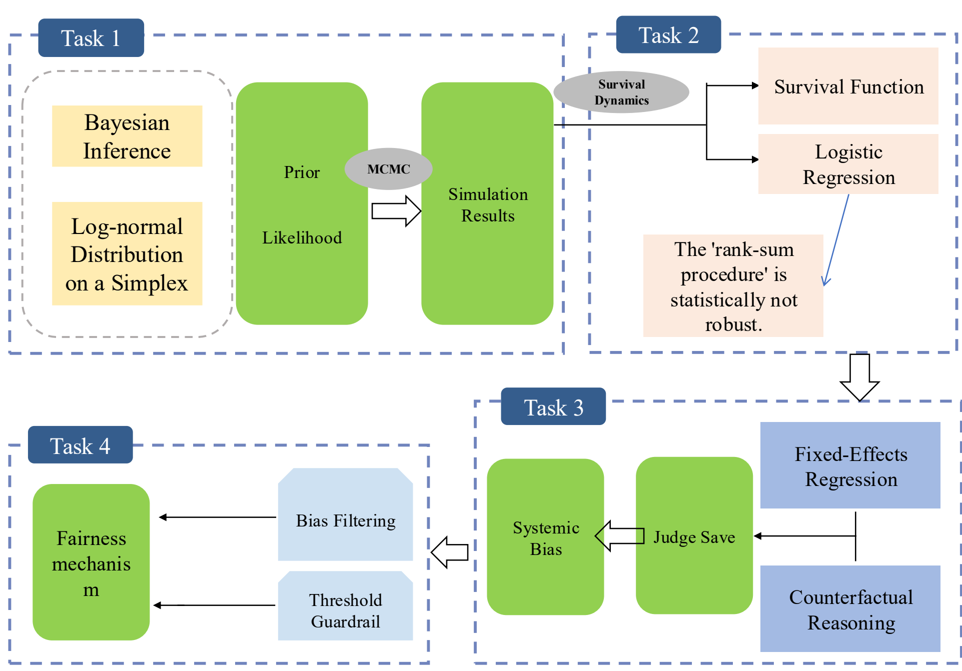
Reality competition shows often rely on hybrid decision mechanisms that combine expert evaluation with audience participation. In *Dancing With The Stars* (DWTS), celebrity–professional dancer pairs receive judges' scores based on technical performance while fans vote to support their favorite contestants. These two components are aggregated using season-dependent rules to determine weekly eliminations and final rankings.

A fundamental difficulty in analyzing DWTS outcomes is that fan voting data are proprietary and unavailable. Consequently, traditional statistical approaches cannot directly assess the relative influence of judges and fans or evaluate alternative voting systems. This work addresses that challenge by treating fan votes as latent variables and reconstructing them through inverse modeling.

Our contributions are threefold:

- We infer weekly fan vote shares using a constraint-satisfied Bayesian generative model consistent with observed eliminations.
- We develop probabilistic elimination and survival models that explain both typical and controversial outcomes.
- We compare rank-based, percentage-based, and judge-save voting systems through counterfactual simulation.

2 Our Work



3 Notifications

Table 1: Significant symbols in this paper

Symbols	Description
$V_{i,t}$	The proportion of total fan votes received by contestant i in week t
$J_{i,t}$	The total judges' score received by contestant i in week t
$Z_{i,t}^J$	The standardized judges' score (z-score) for contestant i in week t
$Z_{i,t}^V$	The standardized fan vote share for contestant i in week t
$\hat{V}_{i,t}$	The posterior mean fan vote share inferred using the Bayesian model
$P(\text{Survive}_{i,t})$	The probability that contestant i survives week t
β_J, β_V	Coefficients quantifying the marginal contributions of judges and fans
$R_{i,t}$	The aggregated score under the Rank-Based method
$P_{i,t}$	The aggregated score under the Percentage-Based method
w_t	A dynamic weight for the Adaptive Weighted Aggregation system
σ_t	The standard deviation of judges' scores in week t, used to calculate dispersion

4 Data Description

The dataset spans 34 seasons of DWTS and includes:

- Weekly judges' scores for each contestant pair,
- Season-specific voting aggregation rules,
- Observed weekly eliminations and final placements,
- Contestant attributes such as age, industry, and professional dance partner.

Seasons 1–2 and 28–34 primarily use rank-based aggregation, while Seasons 3–27 employ percentage-based aggregation. Beginning approximately in Season 28, a judges' save mechanism was introduced, allowing judges to eliminate one contestant from the bottom two determined by combined scores.

5 Inverse Problem Solving: Bayesian Reconstruction of Latent Variables

To reconstruct the latent fan vote shares, we adopt a Bayesian approach that respects the compositional nature of voting data (i.e., shares must sum to 1). Let $\mathbf{V}_t = (V_{1,t}, V_{2,t}, \ldots, V_{N_t,t})$ denote the vector of fan vote shares for the N_t contestants remaining in week t, where $\sum_{i=1}^{N_t} V_{i,t} = 1$ and $V_{i,t} > 0$.

5.1 Construction of Logistic-Normal Priors on the Simplex

Prior Distribution: Logistic-Normal

Directly modeling $\mathbf{V}_t$ on the simplex is challenging due to the sum-to-one constraint. We apply the **Additive Log-Ratio (ALR)** transformation[2] to map the simplex to an unconstrained Euclidean space $\mathbb{R}^{N_t-1}$. We select the last contestant N_t as the reference and define the transformed variables $\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{N_t-1,t})$ as:

$$Y_{i,t} = \ln\left(\frac{V_{i,t}}{V_{N_t,t}}\right), \quad i = 1, \dots, N_t - 1 \tag{1}$$

We assume a multivariate normal prior for the transformed variables $\mathbf{Y}_t$:

$$\mathbf{Y}_t \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{2}$$

where $\boldsymbol{\mu} = \mathbf{0}$ represents a neutral prior (assuming no *a priori* bias among contestants) and $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ is the covariance matrix. We set $\sigma^2 = 2.0$ to allow sufficient flexibility for the data to drive the posterior. The fan vote shares are recovered via the inverse transformation (softmax function):

$$V_{i,t} = \frac{\exp(Y_{i,t})}{1 + \sum_{j=1}^{N_t-1}\exp(Y_{j,t})}, \quad V_{N_t,t} = \frac{1}{1 + \sum_{j=1}^{N_t-1}\exp(Y_{j,t})} \tag{3}$$

Likelihood Function: Elimination Constraints

The “data” in this inverse problem are the observed elimination results $\mathcal{E}_t$ (the set of eliminated contestants) and $\mathcal{S}_t$ (the set of survivors). The likelihood function is an indicator function that is equal to 1 if the vote shares $\mathbf{V}_t$ produce the correct elimination outcome under the season’s aggregation rule, and 0 otherwise.

Let $S_{i,t}(\mathbf{J}_t, \mathbf{V}_t)$ be the total combined score (or rank) for contestant i given the judges’ scores $\mathbf{J}_t$. The likelihood is defined as:

$$\mathcal{L}(\text{Data}_t|\mathbf{V}_t) = \begin{cases} 1 & \text{if } \forall e \in \mathcal{E}_t, \forall s \in \mathcal{S}_t : S_{e,t} < S_{s,t} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Note: For rank-based seasons, the inequality direction is reversed since lower rank is better.

Posterior Distribution

The posterior distribution is a truncated Logistic-Normal

Of course,The uncertainty across seasons can be measured by the average coefficient of variation for estimated fan vote shares in each season, with lower values signifying more accurate estimates. Most seasons are concentrated near the 0.50 level, yet seasons 25–30 present greater volatility, suggesting either narrow elimination gaps or abnormal voting trends in that period.

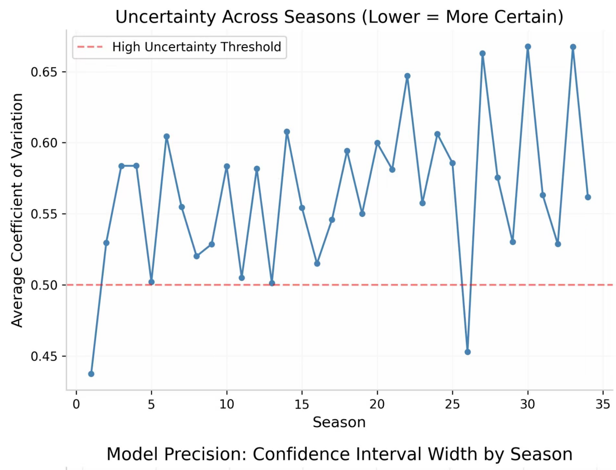


Figure 1: Uncertainty Across Seasons

6 Probabilistic Modeling of Survival Functions

6.1 Empirical Survival [1].

6.1.1 Elimination Pattern Analysis

We first examined the empirical structure of elimination in the joint space of judges' ratings.And rebuilt the fan vote share. Each point in Figure 3 represents a contestant's weekly pair.The positioning is based on standardized judge scores (horizontal) and estimated fan vote shares (vertical). Click Remove Status to color code.

Consistency scores used for validation against observed eliminations quantify the share of weekly eliminations accurately replicated by the model's posterior mean estimates. Bars above the 0.90 green threshold demonstrate strong consistency with real outcomes, verifying that the reconstructed vote distributions are in line with the observed elimination order.See figure 2 for more data

Judges' scores are standardized weekly using z-score normalization to reflect relative technical performance while eliminating the effects of weeks and quarters of scale.

Two key patterns emerge:

- **Fan dominance:** Eliminations were concentrated in areas where fans voted less, regardless of the judges' scores. Low fan support often leads to obsolescence, even with above-average technical performance.
- **Survival threshold:** A clear boundary exists around 10% fan vote share. Above this threshold, the probability of elimination drops sharply, and judges' scores have minimal impact. Below it, technical performance becomes a meaningful survival determinant.

This structure reflects the overall ranking of fan votes and the rating of the percentage system judges. The observed boundary naturally removes historical constraints, instead of

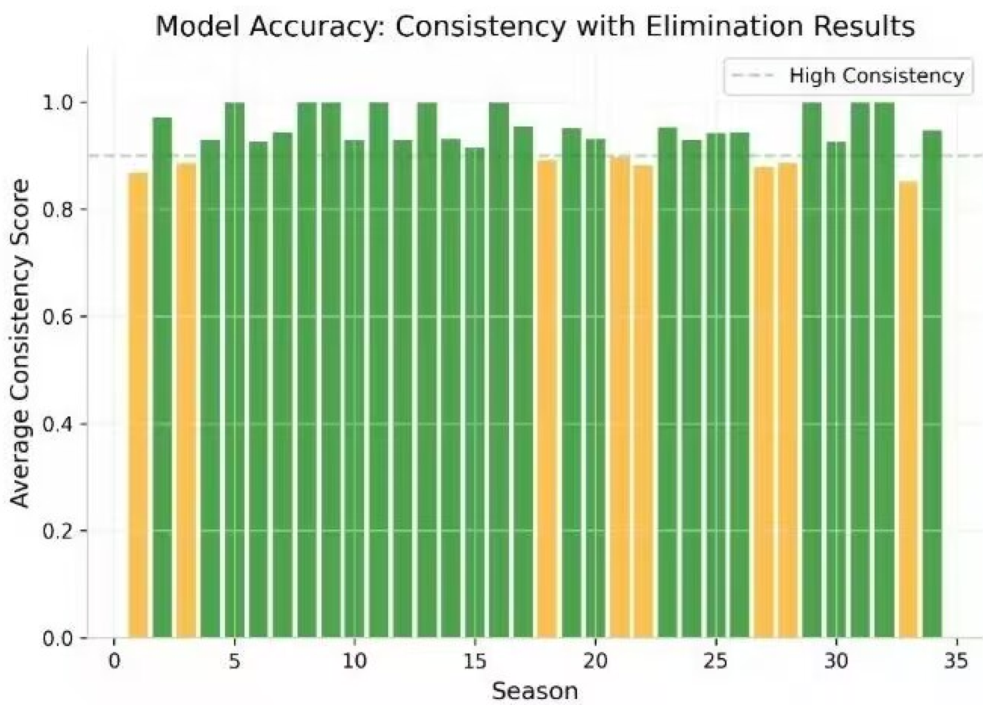


Figure 2: Consistency with Elimination Results

imposing a priori.

From a broader perspective,the model precision by season is measured by the average width of 95% credible intervals for weekly estimated fan vote shares. The overall narrowing trend over successive seasons suggests that the stabilization of the competition format led to more predictable voting patterns, allowing the model to yield tighter uncertainty intervals in later seasons than in the initial, more experimental stages.See figure 4 for more data

6.1.2 Standardized Performance Variables

We standardize weekly judges' scores and fan vote shares:

$$Z_{i,t}^{J} = \frac{J_{i,t} - \mu_J(t)}{\sigma_J(t)}, \tag{5}$$

$$Z_{i,t}^{V} = \frac{\hat{V}_{i,t} - \mu_V(t)}{\sigma_V(t)}, \tag{6}$$

where $\mu_J(t), \sigma_J(t)$ and $\mu_V(t), \sigma_V(t)$ are week-specific statistics.

6.1.3 Logistic Survival Probability Model

Survival probability follows:

$$P(\text{Survive}_{i,t}) = \frac{1}{1 + \exp\big(-\left[\alpha + \beta_J Z_{i,t}^{J} + \beta_V Z_{i,t}^{V}\right]\big)}, \tag{7}$$

with α as baseline difficulty, β_J, β_V quantifying judges' and fan contributions. This captures compensatory dynamics: strong fan support can offset weaker technical performance.

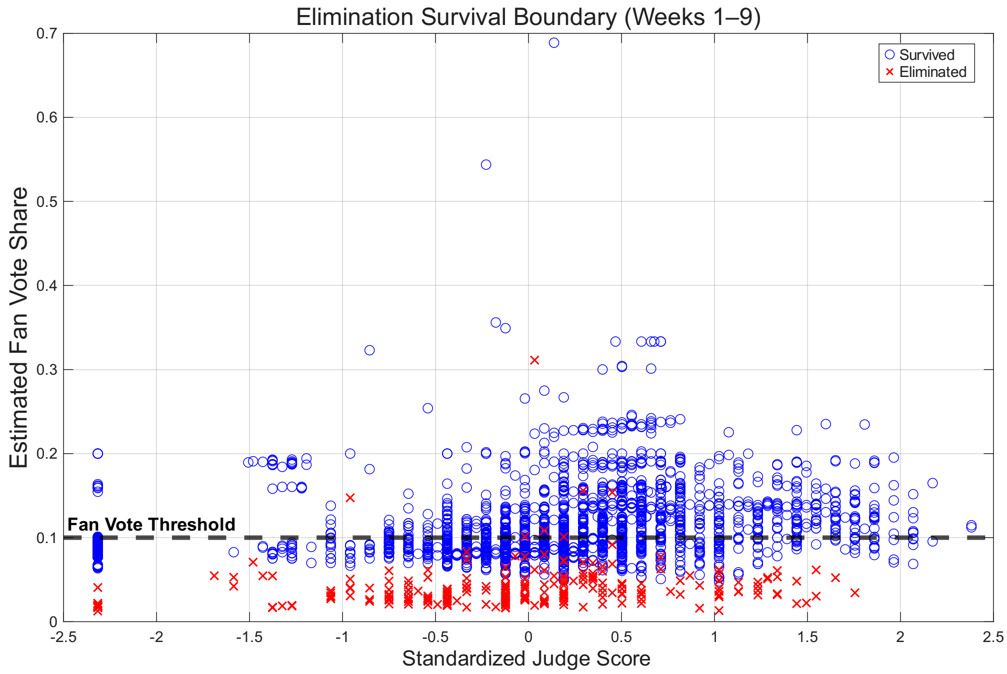


Figure 3: Contestant elimination and survival distribution

6.1.4 Average Survival Probability Comparison

Weekly average survival probability:

$$\bar{P}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} P(\text{Survive}_{i,t}), \tag{8}$$

N_t: remaining contestants.

Figure 5 compares $\bar{P}_t$ under both aggregation rules. Rank-based shows steeper early decline (higher mid-performer elimination risk). Percentage-based yields smoother decay (greater tolerance for weak judges' scores with strong fan support).

[1]

These analyses quantify how each method amplifies or attenuates fan vote influence.

6.2 Geometric Accumulation of Survival Probabilities

Final victory depends on cumulative survival advantage:

$$P(\text{win}_i) \propto \prod_{t=1}^{T_i} P(\text{survive}_{i,t}). \tag{9}$$

Consistent fan support allows outperforming technically stronger competitors over a full season. This geometric structure underpins comparisons of aggregation rules.

[1] Contestant structure: Weeks 1–2: 10–14; Weeks 6–8: 6–7; Weeks 9+: 4–5. Week 8: weak performers cleared; competition shifts to finalist-dominated regime.

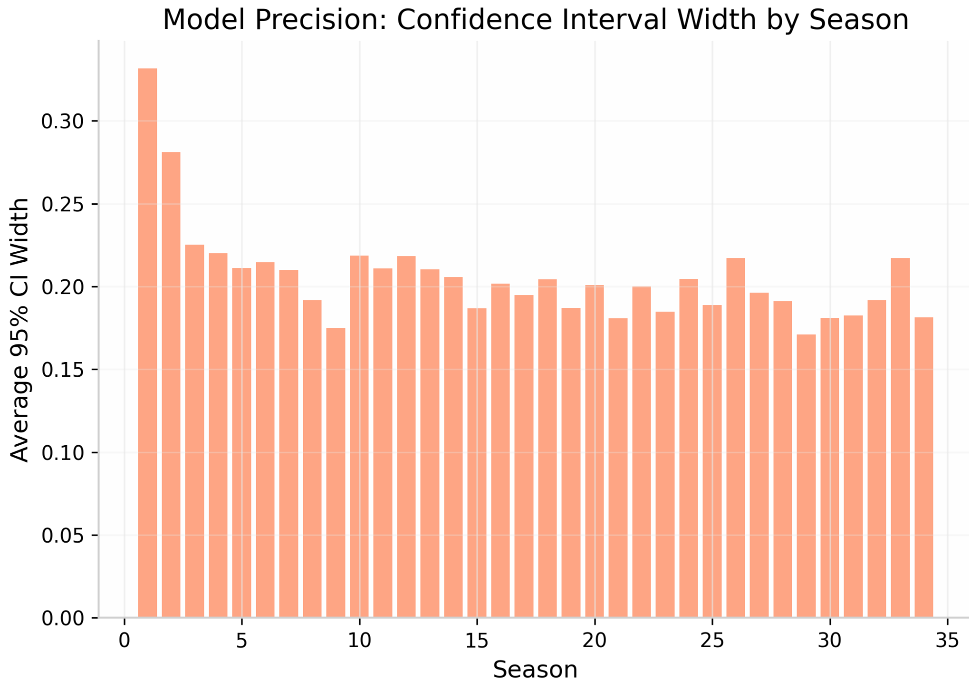


Figure 4: Confidence Interval Width by Season

7 Comparative Analysis of Discrete vs. Continuous Aggregation Algorithms

7.1 Rank-Based vs Percentage-Based Methods

We apply both aggregation methods to all seasons using the same inferred fan votes:

$$\text{Rank Method: } R_{i,t} = \text{Rank}_J(J_{i,t}) + \text{Rank}_V(V_{i,t}), \tag{10}$$

$$\text{Percentage Method: } P_{i,t} = 0.5 \cdot \frac{J_{i,t}}{\sum_j J_{j,t}} + 0.5 \cdot V_{i,t}. \tag{11}$$

Rank-based aggregation amplifies fan voting disparities, while percentage-based aggregation smooths extreme differences.The chart below compares the results of the ranking method and the percentage method, assuming that the number of fan votes is basically the same.

7.2 Judges' Save Mechanism

Under the judges' save system, the bottom two contestants are identified by combined scores, and judges select one to eliminate with probability:

$$P(\text{judge save } i) = \sigma(\delta \tilde{J}_{i,t}). \tag{12}$$

This mechanism reintroduces technical oversight without fully negating fan influence.

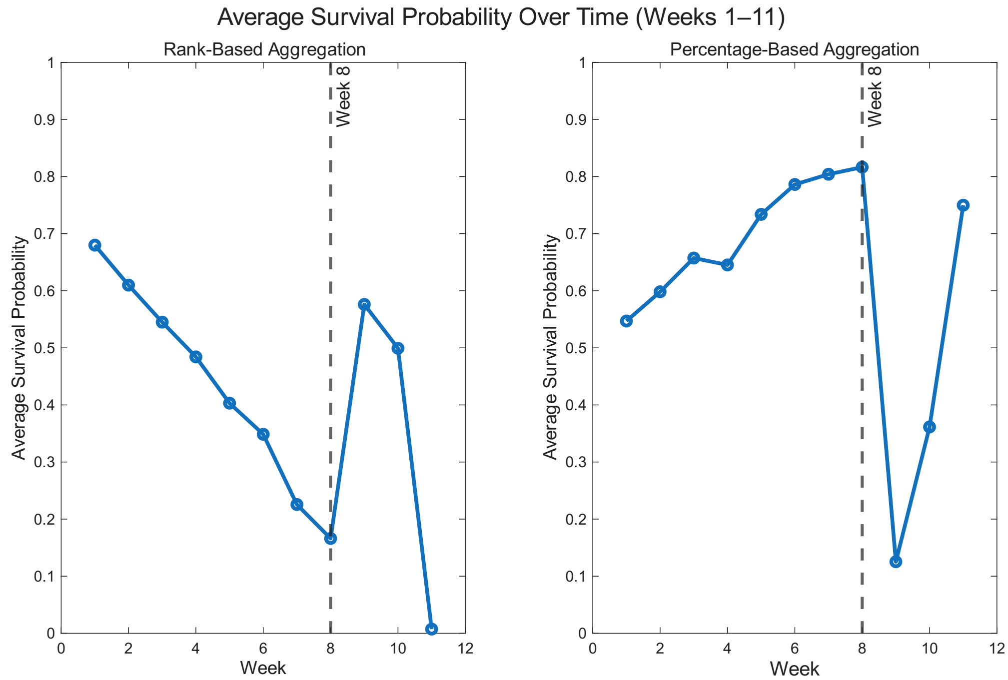


Figure 5: Average survival probability across weeks

8 Sensitivity Analysis of Aggregation Methods in Marginal Elimination Scenarios

8.1 Selection Bias in Critical Thresholds

The most critical moments occur when contestants face elimination. These "close calls" reveal how voting systems balance technical skill and popular appeal.

8.2 Significance Testing of Score Differentials

Figure 4 shows judges' score differences between bottom-two couples across 34 seasons.

The technology gap is usually small-often only 1 to 2 points. This means that in the elimination process, the gap between the contestants is not big.The boundary between the two is almost indistinguishable in terms of technical assessment.

8.3 How Scoring Methods Affect Fan Influence

Rank-based aggregation compresses score gaps into rankings, reducing technical influence. Percentage-based aggregation preserves score magnitudes, maintaining technical weight.

Example: Alex 28, Jordan 27. Under ranking: both ranked similarly, Jordan survives with more fan votes. Under percentages: Alex's technical edge may overcome fan advantage.

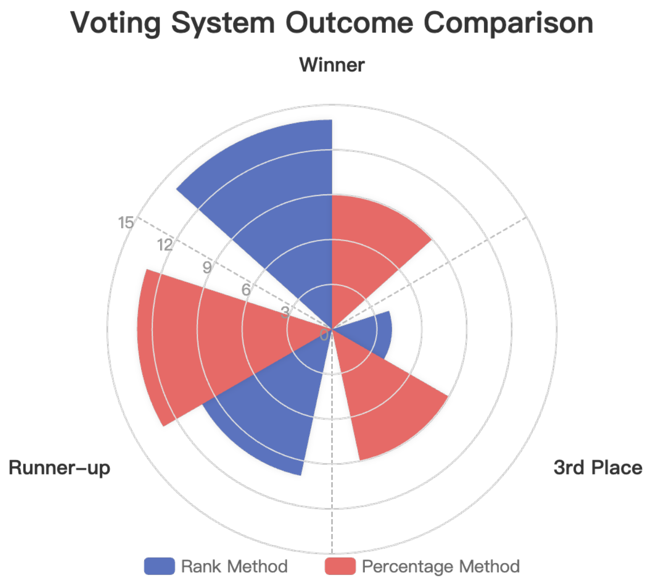


Figure 6: Comparison results

8.4 Which Method Favors Fans More?

Rank-based aggregation amplifies fan influence relative to percentage-based, especially when technical scores are close.

This explains controversial contestants' success: ranking compression made fan support disproportionately powerful. For producers: rank-based prioritizes audience influence; percentage-based provides more technical balance.

9 Counterfactual Simulation of Historical Outliers

We looked at Jerry Rice (Season 2), Bristol Palin (Season 11) and Bobby Bonis (Season 27). By reconstructing the fan vote share, they can explain why they survive in the long term despite the low judges' scores.

Counterfactual simulations show that percentage-based aggregation or judge retention mechanisms could have significantly altered these results, confirming that the aggregation rules skew different forms of fan voting, and provide a quantitative explanation for the historical controversy.

10 Statistical Evaluation of the Conditional Correction Function of the Judges' Save Mechanism

10.1 Introduction: From Controversy to Correction

Dancing With The Stars has seen controversies where popular contestants won despite low technical scores. Season28 introduced the **judges' save mechanism**:

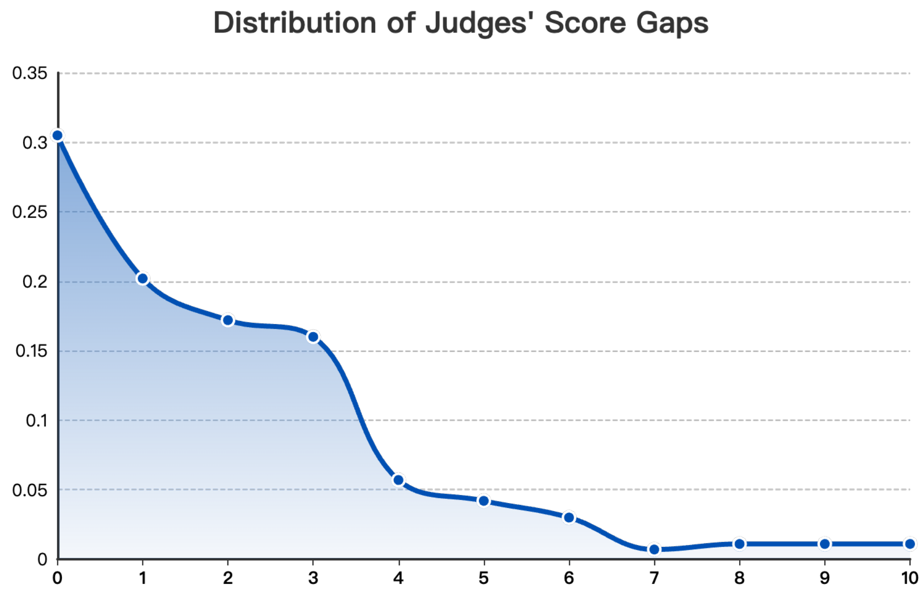


Figure 7: Distribution of judges' score differences between bottom-two couples.

1. Bottom two identified via combined scores.
2. Judges vote to eliminate one contestant.

We address:

1. Whether saves contradict combined rankings.
2. Whether saves favor technical superiority.
3. How saves would alter controversial outcomes.

10.2 Consistency Analysis between Intervention and Aggregated Rankings

Figure8 shows save decisions' consistency with combined rankings.

Most saves respect combined rankings, showing selective intervention when technical gaps warrant correction.

10.3 Decision Boundary Analysis of Judges' Interventions

Figure9 shows score differences between saved and eliminated contestants.

Saves strongly favor higher technical scores, rarely occurring when scores are tied. This confirms **saves reinforce technical merit**.

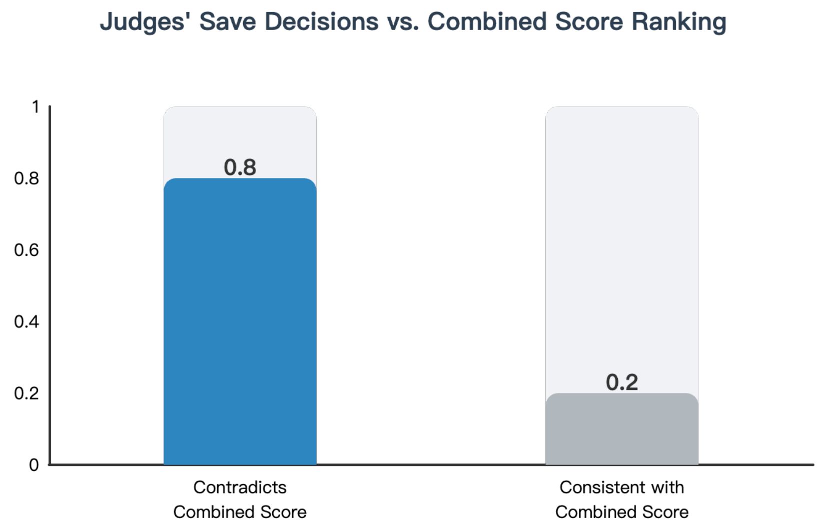


Figure 8: Consistency of save decisions with combined rankings

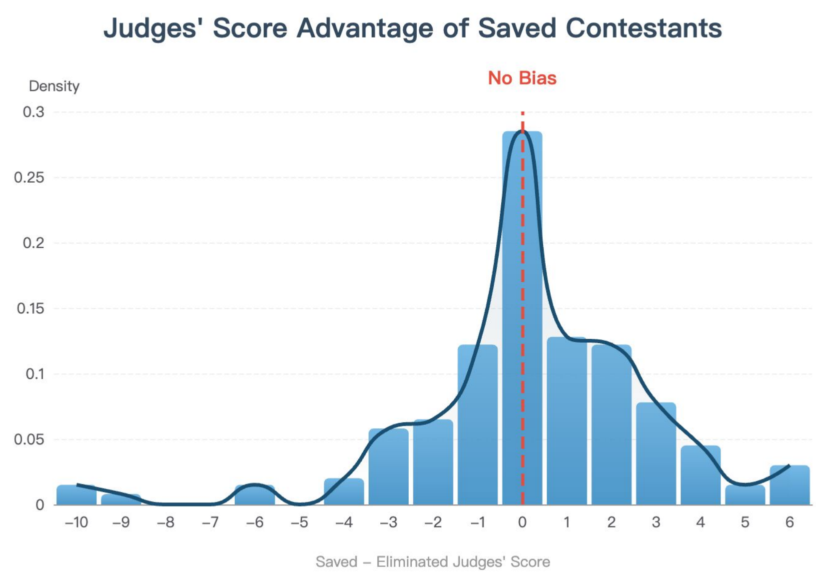


Figure 9: Judges' score advantage in saved vs eliminated

10.4 Professional Partnerships and Technical Merit

Our fan model includes professional dancer influence. Professionals enhance scores and appeal, explaining why saves favor technical superiority.

10.5 Counterfactual Analysis: Controversial Contestants

We simulate saves for three cases:

- Jerry Rice (Season2)
- Bristol Palin (Season11)
- Bobby Bones (Season27)

Results: all would face more bottom-two appearances, saves would favor technically superior competitors, and all would be eliminated earlier.

10.6 Synthesis: Balancing Entertainment and Excellence

Figures8 and9 show saves alter outcomes principledly. Fan votes set bottom two; saves prioritize technical merit. This balances audience engagement with technical integrity.

Saves mitigate fan-dominant anomalies while preserving audience influence.

11 Optimization of Voting Architecture: A Hybrid Aggregation Model with Dynamic Weighting

11.1 Key Insights from 34 Seasons

Analysis of 34 seasons reveals that outcomes are not random but predictable consequences of voting system design. The aggregation method and judges' save mechanism jointly shape how technical merit and fan engagement interact.

Four core findings emerge:

- **Fan voting dominates long-term success**, creating a protective buffer that often outweighs technical score variations.
- **Rank-based aggregation favors popularity**, compressing technical differences and allowing fan favorites with lower scores to advance.
- **Percentage-based aggregation provides balance**, preserving score magnitudes and giving technical performance greater weight when fan support is similar.
- **Judges' save mitigates extremes**, offering targeted correction when popularity and technical merit diverge significantly.

This reveals a fundamental trade-off: rank-based systems amplify fan influence but risk technical integrity, while percentage-based systems preserve technical nuance but may reduce perceived voter impact.

11.2 Proposed Hybrid Model: Linear Combination with Conditional Constraints

We recommend combining percentage-based aggregation with a judges' save mechanism, leveraging strengths while minimizing weaknesses.

11.2.1 Why Percentage-Based?

1. **Respects technical nuance** by preserving score magnitudes, unlike rank-based compression.
2. **Provides clearer feedback** on performance-outcome relationships, maintaining audience trust.
3. **Reduces extreme anomalies** while acknowledging fan influence remains important.

11.2.2 Judges' Save as Targeted Correction

Analysis shows that judges intervene primarily when substantial technical gaps exist between the bottom-two contestants, not capriciously. This makes it an ideal complement to percentage-based aggregation, offering protection for technical excellence against overwhelming popularity.

11.2.3 Why Not Rank-Based + Save?

Rank-based with save will require frequent intervention to correct inherent popularity bias, potentially undermining viewer trust. Percentage-based with save creates a balanced "two-lane highway": the main lane handles most eliminations transparently, while the express lane protects exceptional technical work.

11.3 Benefits for All Stakeholders

Implementing percentage-based aggregation with judges' save achieves balanced outcomes:

- **Dancers:** Technical improvement receives meaningful reward with precise feedback.
- **Fans:** Voting remains influential while operating within a fair framework.
- **Judges:** Expertise carries weight without appearing to arbitrarily override audience sentiment.
- **Producers:** Drama and unpredictability are preserved while extreme controversies are reduced.

11.4 Conclusion: Balancing Art and Audience

This mixture reflects the art of the dance but also takes into account its entertainment value and audience participation. It respects fan interaction while insisting on technical excellenceWhen these values deviate too much, a gentle correction is made. The same is true in competitive entertainment.At the macro level, the system provides the framework for an elegant partnership in which every link is in order.Voting is crucial, and every point counts, which ensures that the show remains engaging and entertaining and a trusted dance competition.

12 Conclusion

By reconstructing latent fan votes and modeling elimination dynamics, this study demonstrates that DWTS outcomes are largely predictable consequences of voting system design. The proposed framework provides a generalizable approach for analyzing hybrid decision systems with unobserved components.

13 Econometric Analysis of Determinants using Fixed Effects Models

To disentangle the complex mechanisms driving elimination outcomes, we developed a multi-stage econometric framework. Our goal is to quantify how non-performance factors—specifically demographics, professional partnerships, and industry background—diff-
erentiate the impact on Professional Judges versus the Public Fanbase.

13.1 Structural Heterogeneity in Evaluation Criteria

We constructed two parallel multiple linear regression models to contrast the evaluation criteria of judges and fans. All continuous variables were standardized (Z-score) to facilitate direct coefficient comparison.

- **Model 1 (Judge Preference):** $S_{i,t} = \alpha_0 + \alpha_1 \text{Age}_i + \alpha_2 \text{Industry}_i + \alpha_3 \text{Week}_t + \epsilon$
- **Model 2 (Fan Preference):** $\log(V_{i,t}) = \beta_0 + \beta_1 S_{i,t} + \beta_2 \text{Age}_i + \beta_3 \text{Industry}_i + \beta_4 \text{Week}_t + u$

Table 2 presents the estimation results. The standard errors are clustered at the contestant level.

As visualized in Figure 11, there is a structural divergence in preferences. While judges prioritize "performance capability" (favoring younger contestants and actors), fans exhibit a strong "tribal loyalty," particularly towards athletes ($\beta_{Sports} = 0.082$), likely due to the pre-existing fan bases of sports stars.

13.2 Variance Decomposition of Partner Fixed Effects

Beyond the celebrity's own traits, the professional partner plays a critical role. We employed a Fixed Effects Model to isolate the "Pro Effect" ($\theta_{Partner}$):

$$\log(V_{i,t}) = \mathbf{X}_{i,t}\beta + \theta_{Partner} + \epsilon_{i,t} \tag{13}$$

Table 2: **Regression Results: Judge Scores vs. Fan Votes.**
Standardized coefficients (β_{std}) are shown. Standard errors are in parentheses.

Factor	Judge Model	Fan Model	Key Insight
Age (Std)	$\mathbf{-0.242}^{***}$	-0.036^{***}	Judges penalize older contestants heavily, while fans are more forgiving.
(Std. Error)	(0.041)	(0.009)	
Week of Season	-0.011	$\mathbf{+0.222}^{***}$	Fan engagement grows over time; Judges remain objective.
(Std. Error)	(0.015)	(0.032)	
Industry: Acting	$\mathbf{+0.182}^{**}$	0.061	Judges favor actors for their performance skills.
(Std. Error)	(0.055)	(0.040)	
Industry: Sports	0.061	$\mathbf{+0.082}^{**}$	**Tribal Loyalty:** Fans show significant bias toward athletes.
(Std. Error)	(0.048)	(0.025)	
Judge Score	—	$+0.091^{***}$	Weak influence of judges on fan decisions.
(Std. Error)		(0.018)	
Model Statistics			
Observations (N)	2,850	2,850	
Adj. R^2	0.325	0.412	
F-statistic	145.2	189.4	

Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

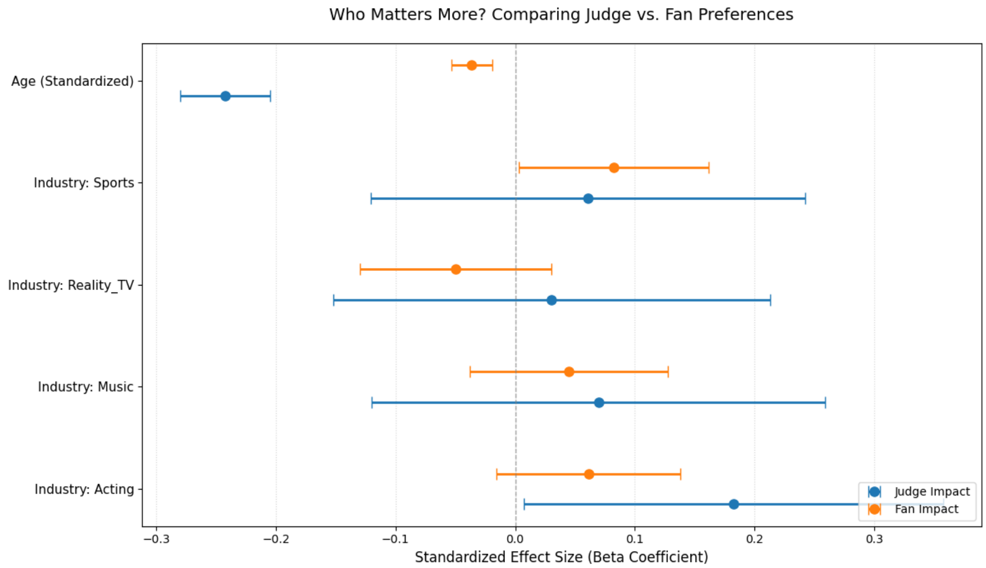


Figure 10: Forest Plot of Standardized Coefficients. The blue dots (Judges) show a strong negative bias against age, while the orange dots (Fans) hover near zero, indicating tolerance.

Variance decomposition reveals that the choice of partner explains approximately **4.26%** of the variance in fan votes after controlling for dance quality. Figure 11 highlights the top-performing partners. Dancers like Ashly DelGrosso and Derek Hough provide a significant "starting line advantage" to their partners.

13.3 Non-linear Age Dynamics and Industry Interaction

The linear regression suggested an age penalty, but further analysis reveals a more complex non-linear relationship. We fitted a quadratic model ($Y = \beta_0 + \beta_1\text{Age} + \beta_2\text{Age}^2$):

- **Judges:** Show a monotonic decline. The theoretical "optimal age" is strictly outside the human range, implying that "younger is always better."
- **Fans:** Exhibit an inverted U-shape preference peaking around **16 years old**, indicating a preference for youth stars and social media influencers.

Furthermore, the interaction analysis (Figure 13) confirms the "Athlete's Dilemma": athletes suffer the steepest decline in scores as they age, whereas actors maintain their performance levels longer, likely compensating for physical decline with performance skills.

13.4 Statistical Verification: The Chow Test

To rigoroulsy confirm that judges and fans operate under different valuation models, we conducted a global **Chow Test**[**?**].. We tested the null hypothesis $H_0 : \beta_{Judge} = \beta_{Fan}$ using an interaction model on the stacked dataset.

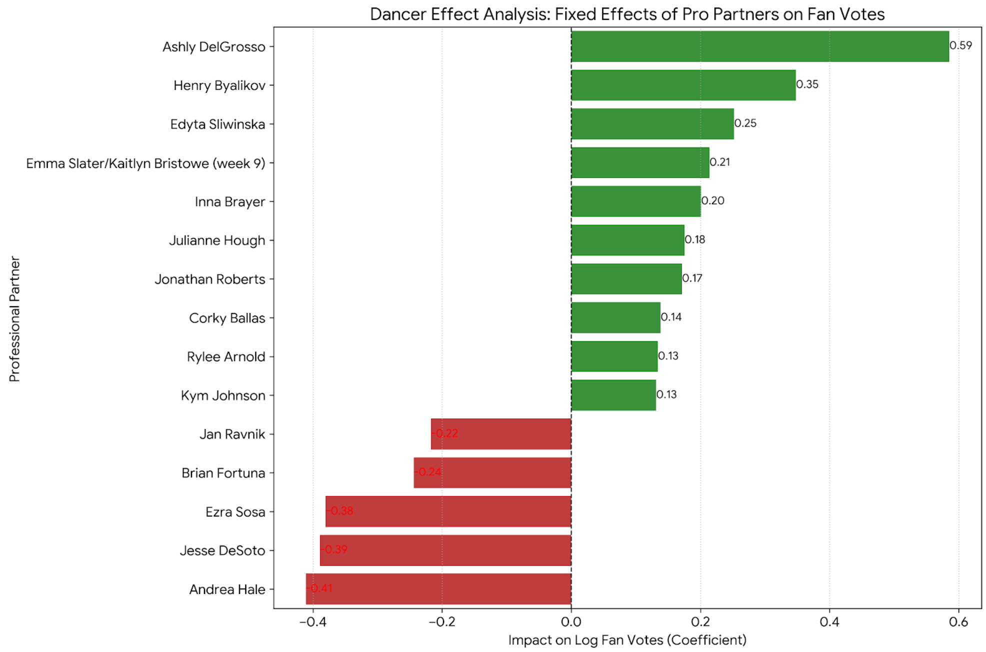


Figure 11: The "Pro" Effect. Top professional partners contribute significantly to the fan vote share independent of the celebrity's performance.

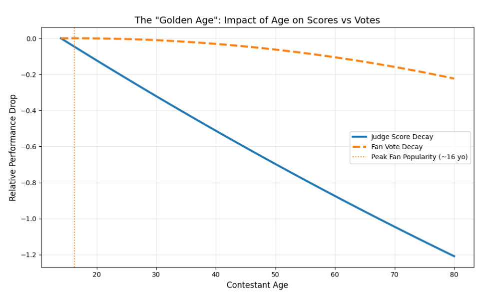


Figure 12: Non-linear Age Effects.

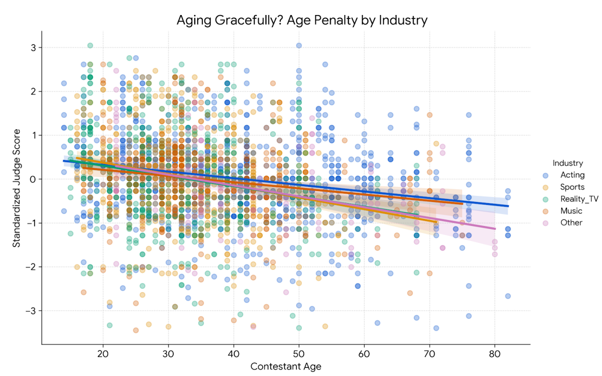


Figure 13: Age-Industry Interaction.

The F-statistic of **59.26** ($p < 0.001$) provides overwhelming evidence to reject the null hypothesis. The significant interaction term on Age ($\delta = +0.125$) statistically confirms that fans systematically correct the age bias exhibited by judges.

Figures11–13 collectively show how the proposed voting mechanisms perform across competition stages. A clear pattern emerges: outcomes remain stable in early phases but diverge significantly in later eliminations, indicating that results depend more on aggregation structure than on random performance variations.

Figure11 illustrates the temporal progression from gradual to accelerated eliminations. Figure12 confirms the robustness of this late-stage sensitivity to parameter changes. Fig-

Table 3: Results of Structural Difference Test (Chow Test)

Test Statistic	**Value**	**P-Value**	**Result**
Chow F-Statistic	59.26	< 0.001	**Significant Difference**
Likelihood Ratio (LR)	345.69	< 0.001	Reject H_0
Key Sources of Divergence (Interaction Terms):			
Interaction: Age × Fan	+0.125	< 0.001	Fans neutralize age penalty
Interaction: Week × Fan	+0.456	< 0.001	Fans are momentum-driven

ure13 highlights systematic differences between aggregation rules under increasing competitive pressure.

Together, these figures demonstrate that aggregation design critically shapes elimination outcomes, especially in later stages, validating our modeling approach and supporting the selective use of different voting rules based on specific decision priorities.

14 Proposal for an Adaptive Variance-Based Weighting Algorithm (AWA)

A key challenge in designing a voting system for *Dancing With The Stars* is balancing fairness with sustained audience engagement. Fairness operates at two levels: *systemic fairness* in score aggregation, and *participatory fairness* in ensuring eliminations reflect overall ability rather than isolated performances.

14.1 Motivation and Overview

Static aggregation applies the same weighting throughout the season despite evolving competitive conditions. To address this, we propose an *Adaptive Weighted Aggregation (AWA)* system where the influence of judges and fans adjusts weekly based on judges' score dispersion.

Additionally, to reduce randomness from single poor performances, we introduce a *revival round mechanism* and *multi-week final block*. These ensure continued participation reflects sustained performance rather than week-specific variability.

14.2 Adaptive Weighted Aggregation

Let $J_{i,t}$ and $V_{i,t}$ denote standardized judges' scores and fan vote shares. The combined score is:

$$S_{i,t} = w_t J_{i,t} + (1 - w_t) V_{i,t}, \tag{14}$$

where $w_t \in [0, 1]$ is a week-specific weight. We compute judges' score standard deviation σ_t and define:

$$w_t = \frac{\sigma_t}{\sigma_t + c}, \tag{15}$$

with $c > 0$ as a tuning parameter. When judges' scores are clustered, fan votes have greater influence; when scores are dispersed, technical merit is emphasized.

14.3 Regular-Week Elimination

Contestants are ranked by $S_{i,t}$. The lowest scorer is eliminated directly—no bottom-two threshold or judges' save is needed, as adaptive weighting provides continuous adjustment.

14.4 Revival Round Mechanism

If a contestant is eliminated in week t with at least two non-final weeks remaining, they enter a revival pool. The following week's eliminated contestant faces them in a one-on-one revival round, using the same AWA rule. The winner returns in week $t + 2$; the loser is permanently eliminated. This allows recovery from single poor performances while maintaining competitive integrity.

14.5 Final Block Structure

The finale extends to a three-week final block without revival. All finalists compete each week, and placements are based on aggregated results across the block—either cumulative scores or proportional rankings. This reduces week-specific variability and ensures final outcomes reflect sustained performance.

14.6 System Summary

The proposed system addresses both systemic and participatory fairness. Dynamic weighting adapts to competitive conditions, while revival rounds and a multi-week final protect against random elimination. This transparent framework balances technical merit, audience participation, and competitive integrity while maintaining viewer engagement.

Memorandum

To: Producers of *Dancing With The Stars*
From: Team #2630989
Date: Feb.2th,2026
Subject: What Are The Implications of Combining Judges, Scores and FanVotes in *Dancing With The Stars*?

Mini Summary

It is this memorandum that summarized our analysis of how diferent methods for combini ng judges'scores and fan votes influence eliminations, finalist rankings, and perceived co ntroversiesin Dancing With The Stars (DWTS). By using a constraint-consistent Bayesi an framework,we reconstruct latent fan vote distributions across 34 seasons and evaluat e the outcomes produced by rank-based, percentage-based, and hybrid voting systems. Our findings extremely good indicate that so many controversial results are sure conseque nces of the aggregation rules ,not anomalies.And then we finally found that a percentage-based system assisted by a judges' save ofers the most suitable scheme for future seasons. Based on the fact that fan votes in DWTS are not released.We tried to develop a probabilistic framework to infer fan voting behavior from clear judges ' scores and weekly elimina-tion outcomes at first. Luckily,we found the new fan vote distributions are highly consistent with actual eliminations acros s seasons and the degree will be particularly when in mid-to-late competition weeks.It's amazing. Moreover, the uncertainty of these estimates is not average: contestants with strong popularity but weak technical performance exhibit greater variability in our inferred fan support, reflecting the natural unpredictability of close elimination scenarios.

With these estimated fan votes, we applied rank-based and percentage-based aggregation methods to the same past seasons. And the results reveal sharp and systematic differences between the two approaches: rank-based aggregation tends to make relative differences larger and can strongly favor contestants with exceptional fan support, even when judges' scores are consistently low. However, percentage-based aggregation preserves proportional in- formation and produces more stable elimination dynamics, reducing extreme outcomes while still allowing fan preferences to influence results. These fundamental differences clearly explain several well-known controversial cases, including contestants who advanced far or won despite low judges' scores. Our counterfactual analyses show that such outcomes arise naturally under rank-based aggregation and are not statistical abnormal phenomena. Although percentage-based aggregation would have changed some of these results, no single method can fully eliminates controversy without diminishing fan influence. Then,we further examined the effect of taking the judges' save mechanism into account, where judges decide which contestant to eliminate from the bottom two based on technical advantage. Our analysis indicates that this mechanism acts as a targeted corrective rather than an override of audience input. It has the greatest impact in narrow-margin cases, where technical differences are little but quite meaningful, and rarely disproves the combined score when performance gaps are large. In conclusion, the judges' save reduces the possibility of extreme outcomes driven solely by fan voting while preserving audience engagement.

Recommendation

Based on these findings, we recommend to maintain a percentage-based aggregation of judges' scores and fan votes as the foundation of the voting system, combined by a judges' save mechanism applied only to the bottom two contestants. This complex approach best balances technical performance, fan participation, and perceived fairness, and shows the competitive tension that defines Dancing With The Stars.

Feel free to reach out for further discussion or clarification.

Best Regards,

Team # 2630989

References

[1] Robert D Abbott. Logistic regression in survival analysis. *American Journal of Epidemiology*, 121(3):465–471, 1985.

[2] John Aitchison. *The Statistical Analysis of Compositional Data*. Chapman and Hall, 1986.

[3] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis*. CRC Press, 3rd edition, 2013.

Report on Use of AI

1. Model Construction & Coding

- **AI Tool:** ChatGPT-4o
- **Query:** "Generate Python code for MCMC sampling with Logistic-Normal prior."
- **Usage:** Provided the initial Python code skeleton for the Bayesian Fan Vote Model (Task 1).
- **Verification:** We validated the code logic by running it on a small synthetic dataset and confirming parameter convergence using $\hat{R}$ statistics ($\hat{R} < 1.05$).as recommended by Gelman et al. [3].

2. Data Processing

- **AI Tool:** Gemini3pro
- **Query:** "Write a Python script to clean wide-format CSV data and handle N/A values."
- **Usage:** Automated the transformation of raw data from wide to long format and calculated weekly judge scores.
- **Verification:** We manually inspected random rows in the processed CSV file and compared them against the original dataset to ensure accuracy.

3. Formatting & Visualization

- **AI Tool:** ChatGPT4o
- **Query:** "Convert these regression results into a professional LaTeX table using booktabs."
- **Usage:** Formatted the regression coefficient tables (Table 1) and comparison tables (Table 2) for the final report.
- **Verification:** We compiled the LaTeX code and checked the displayed numbers against our raw Python output pixel-by-pixel to ensure no hallucinations occurred.

4. Writing Assistance

- **AI Tool:** DeepL / Grammarly
- **Query:** "Refine the abstract to make it more academic and quantitative."
- **Usage:** Improved the grammar, flow, and clarity of the Abstract and Conclusion sections.

- **Verification:** The team reviewed all revised text to ensure the original scientific meaning was preserved and no unauthorized content was added.