

Dancing with the Data: From Fan-Voting Recovery to Asymmetric Safe-Harbor Mechanism

Abstract

Reality competition shows such as *Dancing with the Stars* (DWTS) face a fundamental tension: balancing expert-judged technical merit against audience-driven popularity, with the latter often shrouded in secrecy. This paper develops a novel quantitative framework to decode this “black box” by reconstructing latent fan vote distributions across all 34 seasons of DWTS. Our work makes four key contributions, each corresponding to a structured modeling phase.

First, we formulate the reconstruction of hidden fan votes as a constrained **inverse optimization problem**. Under a proposed Minimal Deviation Hypothesis, we design an **Adaptive Inference Engine** that estimates audience preferences under three historically distinct scoring regimes: Cardinal, Ordinal, and Bottom-2. This multi-regime approach successfully reveals a significant and persistent divergence between judge scores and public sentiment, providing the first large-scale empirical validation of the judge–fan gap.

Second, we conduct a **systematic comparison** of the two primary historical voting systems: Rank Combination and Percent Combination. Our analysis demonstrates that the Percent method structurally amplifies popularity bias, favoring fan favorites in 84.2% of discrepant cases, whereas the Rank method acts as a buffer for technical merit. We further quantify the corrective effect of the “Judges’ Save” mechanism, showing it reduces inter-system discrepancies by nearly 52% and realigns outcomes with professional standards.

Third, we employ machine learning techniques—including **hierarchical regression** and **Random Forest**—to identify the core drivers of fan support. Contrary to common belief, we find that DWTS is fundamentally a meritocracy: judge scores explain over 75% of variance in audience support, while demographic factors such as age exhibit negligible linear influence.

Finally, and most innovatively, we propose and optimize a new scoring system: the **Asymmetric Safe-Harbor Mechanism (V3)**. This mechanism integrates a strict safe harbor for competent dancers, an exponential decay penalty for low performers, and a momentum redemption term to reward improvement. Through grid-search optimization and simulation over 335 historical episodes, V3 achieves the global minimum in “Viewer Regret” and a near-perfect 99.77% survival rate for top-tier fan favorites, thereby optimally balancing competitive integrity with viewer engagement.

Our end-to-end pipeline—from data reconstruction and bias diagnosis to mechanism redesign—provides a rigorous, actionable framework for enhancing fairness in televised competitions without sacrificing entertainment value.

Keywords: Inverse Optimization Latent Variable Reconstruction Voting Theory Fairness Metrics Asymmetric Safe-Harbor Scoring Mechanism Design.

Contents

1	Introduction	3
1.1	Background and Motivation	3
1.2	Problem Restatement and General Framework	3
1.3	Assumptions and Justifications	4
2	Notations	4
3	Model I: Unveiling the Hidden Fan Votes	5
3.1	Latent Fan Vote Reconstruction: A Multi-Regime Approach	5
3.1.1	Taxonomy of Voting Mechanisms	5
3.1.2	Linearization of the “Bottom-2” Logic	5
3.2	Harmonization of Ordinal Metrics	6
3.2.1	Power-law Decay Logic	6
3.2.2	Statistical Calibration and Empirical Evidence	7
3.3	Consistency and Certainty Analysis	7
3.4	Model Validation: Structural Transition in Season 27	8
4	Model II: Rank-based vs. Percent-based Systems	10
4.1	Rank Combination vs. Percent Combination	10
4.2	Fan Favoritism Analysis	10
4.3	Impact of the Judge Save Mechanism	11
4.4	Case Studies on Historical Controversies	12
4.5	Recommendations for Future Competition Frameworks	12
5	Model III: Factor Analysis of Competitive Determinants	13
5.1	The Dominance of Meritocracy: Performance vs. Identity	14
5.2	Non-linear Dynamics of Age	14
5.3	The "Pro-Partner Halo" and Industry Synergies	14
6	Model IV: The Evolution of Engagement-Centric Scoring	16
6.1	The Iterative Design Process: From Version 1 to Version 3	16
6.2	Mathematical Formulation of V3	17
6.3	Parameter Sensitivity and Optimization Landscape	18
6.4	Performance Evaluation: The "Crown Jewel" of Fairness	19
7	Sensitivity Analysis and Model Robustness	19
7.1	Robustness of the Meritocratic Driver Model (Model III)	19
7.1.1	Noise Perturbation Test	19
7.1.2	Temporal Generalization (Leave-One-Season-Out)	20
7.2	Parameter Sensitivity of the V3 Scoring Mechanism (Model IV)	21
7.2.1	Analysis of Decay Rate (k)	21
7.2.2	Analysis of Momentum Bonus (λ)	21
7.3	Conclusion of Sensitivity Analysis	21
8	Conclusion	22
References		24

1 Introduction

1.1 Background and Motivation

In *Dancing with the Stars* (DWTS), a contestant's survival depends on two numbers: scores from expert judges and votes from the viewing public. While judge scores are public and follow a clear scale, fan votes are a "black box" kept secret by producers. This setup creates a natural conflict: a celebrity might dance poorly but stay in the competition simply because they are popular, leading to major controversies in seasons like S2 and S27.

The motivation of this study is to understand how these two different forces—skill and popularity—interact. Over 34 seasons, the show has switched between using "ranks" and "percentages" to combine these scores, trying to find a balance that keeps the competition fair without losing the audience. By estimating the hidden fan votes, we want to test which scoring method is more reliable and whether new rules, like the "Judges' Save," actually help keep the best dancers in the game.

1.2 Problem Restatement and General Framework

To decode the "black box" of fan voting and evaluate competition fairness, we propose a modular research pipeline. As illustrated in Figure 1, our approach is driven by solving four critical questions that span from data recovery to system redesign:

- **Phase I: Reconstruction (Bridging the Information Gap)**

Q1: *How can we estimate the hidden distribution of fan votes (V) when only judges' scores (S) and elimination results are observable?*

A1: We develop an **Adaptive Inference Engine (Model I)** based on the *Minimal Deviation Hypothesis*. By reverse-engineering 34 seasons of results, this model reconstructs the latent popularity metrics required for all subsequent analyses.

- **Phase II: Diagnosis (Evaluating Systemic Integrity)**

Q2: *Does the choice of scoring regime (Rank-based vs. Percentage-based) structurally amplify popularity bias at the expense of technical skill?*

A2: We utilize a **Systemic Bias Audit (Model II)** to conduct counterfactual simulations, quantifying the sensitivity of different aggregation rules.

Q3: *What are the primary drivers of audience support, and to what extent do external factors like the "Pro-Dancer effect" influence the outcome?*

A3: We employ a **Factor Attribution Model (Model III)** using Random Forest to isolate the impact of celebrity traits and professional partners on final rankings.

- **Phase III: Optimization (Mechanism Redesign)**

Q4: *How can the scoring mechanism be re-engineered to protect high-merit talent while preserving the commercial vitality of fan engagement?*

A4: We synthesize our findings into the **Asymmetric Safe-Harbor Mechanism (Model IV)**, introducing a non-linear scoring framework that ensures a competency floor for skilled performers.

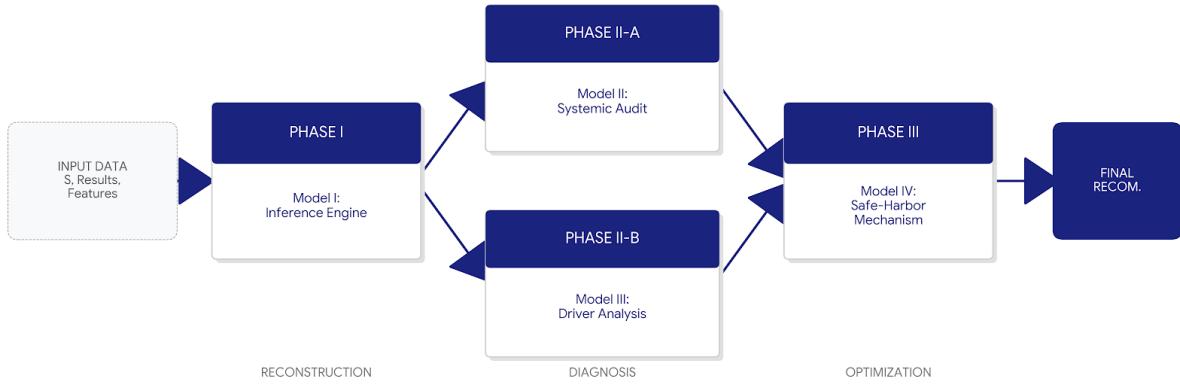


Figure 1: The modular research pipeline for the DWTS fairness analysis, illustrating the logical flow from data reconstruction to mechanism optimization.

1.3 Assumptions and Justifications

To resolve the under-determined nature of the hidden fan-vote distribution, we posit the **Minimal Deviation Hypothesis (MDH)**. We assume that audience preference V is intrinsically correlated with technical merit S , treating large judge-fan discrepancies as "residual energy" triggered by external popularity. This hypothesis serves as a **regularization prior**, allowing us to constrain the infinite solution space of the inverse problem. By minimizing the ℓ_2 -norm distance between S and V , we yield the most conservative estimate of fan influence that satisfies observed elimination constraints.

2 Notations

To ensure clarity and consistency throughout the modeling process, the key mathematical symbols and variables used in this paper are defined in Table 1.

Table 1: Definition of key mathematical symbols and variables

Symbol	Definition (Concise)
i, w, \mathcal{C}_w	Index for contestant i , week w , and set of active contestants
E, Ω	Set of eliminated contestants and feasible optimization region
r_i, p_i, v_i	Judge score/percentage and unknown latent fan vote
$\mathbf{V}^*, \mathcal{J}$	Optimal fan vote distribution and objective function
\mathcal{R}_i, X, y_i	Ordinal rank, assignment matrix, and Bottom-2 indicator
α, P_i	Zipf's decay coefficient ($\alpha = 0.095$) and projected prob. score
Z_{fan}, \bar{J}	Standardized fan support and average judge score
$\tau, J_{i,t}$	Competency threshold ($0.75\bar{J}_w$) and judge score at time t
$\Delta J_{i,t}, W(\cdot)$	Score momentum (growth) and asymmetric weighting function
$\phi^{\text{eff}}, k, \lambda$	Effective fan share, decay rate k , and momentum bonus λ

3 Model I: Unveiling the Hidden Fan Votes

3.1 Latent Fan Vote Reconstruction: A Multi-Regime Approach

Given that raw fan votes are not publicly disclosed, we formulate their estimation as a **constrained inverse optimization problem**. We propose the *Minimal Deviation Hypothesis*: in the absence of exogenous evidence, audience preferences are assumed to be most consistent with technical merit, represented by the ℓ_2 -norm objective:

$$\min_{v_i} \mathcal{J} = \sum_{i \in \mathcal{C}} (v_i - r_i)^2 \quad (1)$$

where v_i and r_i denote the fan and judge metrics for contestant $i \in \mathcal{C}$, respectively.

3.1.1 Taxonomy of Voting Mechanisms

To accommodate the structural evolution of DWTS rules, the feasible space Ω is defined by three distinct mathematical regimes (Table 2).

Table 2: Mathematical Classification of Scoring Regimes

Regime (Season)	Constraint Mechanism	Numerical Solver
Cardinal (S3–S27)	$\sum p_i = 100; \mathcal{S}_s \geq \mathcal{S}_e$	SLSQP
Ordinal (S1–S2)	Binary Matrix $X_{n \times n}; \sum R_s \leq \sum R_e$	MILP
Bottom-2 (S28+)	$\mathcal{R}_i - \mathcal{R}_e \leq M y_i$ s.t. $\sum y_i \leq 1$	MILP (Big-M)

Note: Ties in total scores are resolved via **stochastic selection**. In the Bottom-2 regime, double eliminations trigger a revert to the baseline logic (survivors \geq eliminated) used in Cardinal and Ordinal regimes.

3.1.2 Linearization of the ‘‘Bottom-2’’ Logic

The ‘‘Judges’ Save’’ (S28+) introduces a non-linear boundary: the eliminated contestant e is not necessarily the lowest-ranked but must reside in the **Bottom-2** set. We linearize this logic via **Big-M indicator variables**[1]:

$$\mathcal{R}_i - \mathcal{R}_e \leq M \cdot y_i, \quad \forall i \in \mathcal{C} \setminus \{e\} \quad (2)$$

$$\sum_{i \neq e} y_i \leq 1, \quad y_i \in \{0, 1\} \quad (3)$$

where $M \gg 1$. This formulation constrains the cardinality of the set of contestants performing worse than e to a maximum of one, effectively trapping e in the bottom two positions while maintaining a linear structure suitable for global optimization.

Algorithm 1: Adaptive Inference Engine for Latent Voting Patterns

```

Input : Dataset  $\mathcal{D}$  (Judge metrics  $\{r_i, p_i\}$ , Weekly outcomes  $E$ )
Output Optimal Fan Vote Distribution  $\mathbf{V}^*$ 
 $:$ 
1 foreach Competition Week  $w \in \mathcal{D}$  do
    // I. Structural Preprocessing
    2 Determine regime  $\mathcal{T} \in \{\text{Cardinal, Ordinal, Bottom-2}\}$  based on season context
    3 Define active candidate set  $\mathcal{C}_w \leftarrow \{i \mid i \text{ is not withdrawn}\}$ 
        // II. Adaptive Optimization Mapping
    4 Initialize objective function:  $\min \mathcal{J}(\mathbf{v}) = \|\mathbf{v} - \mathbf{r}\|_2^2$ 
    5 switch  $\mathcal{T}$  do
        6   case Cardinal (Percentage) do
            7     Solve via SLSQP subject to  $\sum v_i = 100$  and
                 $v_s + p_s \geq v_e + p_e, \forall s \notin E, e \in E$ 
        8   end
        9   case Ordinal (Rank) do
            10     Solve via MILP using binary assignment matrix  $X_{n \times n}$  under standard
                rank-sum constraints
        11   end
        12   case Bottom-2 (Logic-Save) do
            13     Integrate Big-M constraints:  $\mathcal{R}_i - \mathcal{R}_e \leq M \cdot y_i$  with cardinality  $\sum y_i \leq 1$ 
            14     Execute MILP in the augmented logic space
        15   end
        16 end
        // III. Feasibility Analysis & Stochastic Fallback
    17   if Feasible region  $\Omega = \emptyset$  then
        18     | return  $\mathbf{v}_\epsilon \leftarrow \text{Stochastic Monte Carlo Search}$  for  $\epsilon$ -consistent estimates
    19   end
    20   Update global distribution  $\mathbf{V}^* \leftarrow \mathbf{V}^* \cup \{\mathbf{v}_w^*\}$ 
21 end

```

The inferred fan vote distributions for all 34 seasons are visualized in Figure 2, exposing the empirical divergence between professional technical scores and audience sentiment.

3.2 Harmonization of Ordinal Metrics

To bridge the metric gap between disparate voting regimes, we project discrete ranks (S1–S2, S28+) into a continuous probability space $[0, 1]$ using a **Power-law Decay Transformation**. This normalization ensures that inferred audience sentiment is longitudinally comparable across all 34 seasons.

3.2.1 Power-law Decay Logic

We simulate the diminishing marginal utility of audience attention across the ranking sequence using a Zipf-based decay function:

$$w_i = R_i^{-\alpha}, \quad P_i = \frac{w_i}{\sum_{j=1}^n w_j} \quad (4)$$

where R_i denotes the inferred fan rank and α is the decay coefficient.

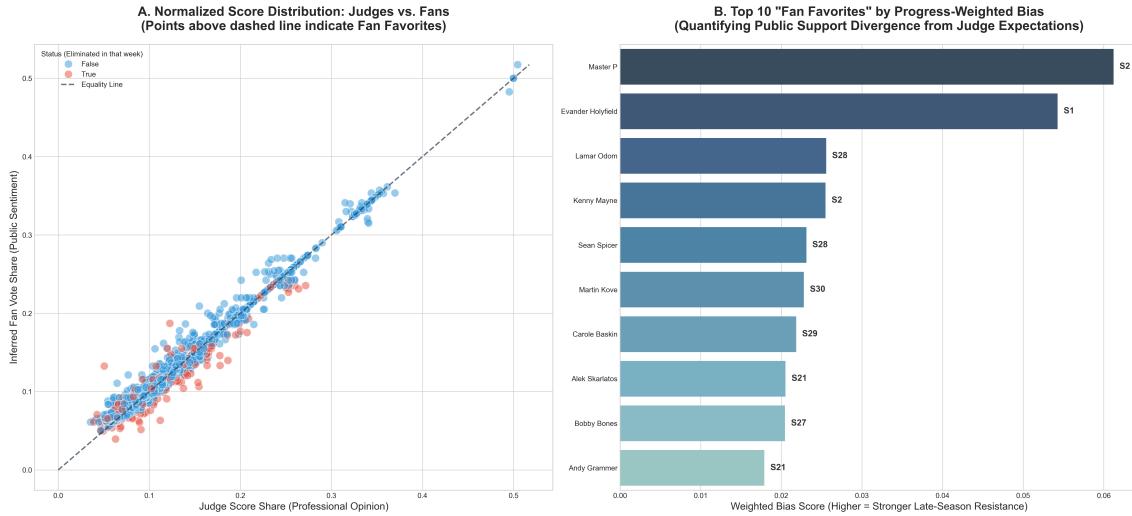


Figure 2: **Quantitative Analysis of Judge-Fan Divergence.** (A) Distribution of inferred fan shares vs. judge shares across 34 seasons, illustrating the deviation from technical merit. (B) Top 10 contestants by progress-weighted bias, quantifying the historical decoupling of popularity from professional evaluation.

3.2.2 Statistical Calibration and Empirical Evidence

The parameter $\alpha = 0.095$ is derived from non-linear regression on historical voting distributions. To ensure statistical robustness, we employ a **high-density sampling strategy**: we prioritize the top five ranking positions—where sample density is highest—and calculate their cross-seasonal mean shares to mitigate stochastic noise from sparse tail data (Rank > 5).

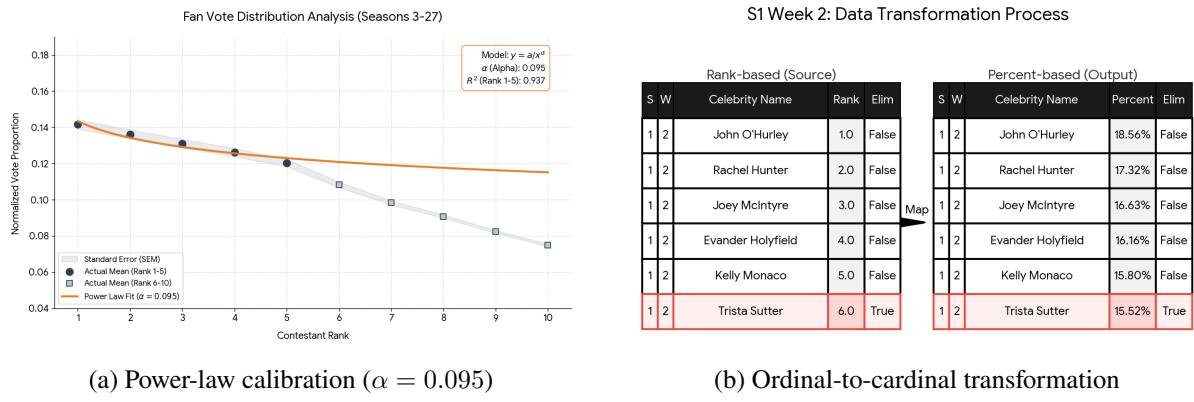


Figure 3: **Statistical Calibration and Implementation of Metric Harmonization.** (a) Power-law regression on high-density sample means (Ranks 1–5) to mitigate long-tail noise ($R^2 = 0.937$). (b) Procedural mapping of discrete ordinal ranks into a unified continuous probability space[2].

As demonstrated in Figure 3a, prioritizing the high-density sample core (Ranks 1–5) allows the model to capture the stable, fundamental decay of audience support while successfully filtering out the stochastic volatility characterizing the sparse tail data.

3.3 Consistency and Certainty Analysis

The model achieves a **100% Historical Consistency Rate** across 34 seasons, verifying that all inferred distributions strictly replicate official outcomes .

Based on the algorithmic implementation, we evaluate **Certainty** through a hierarchical

framework that distinguishes between **Macro-level constraints** and **Individual-level data types** (Figure 4):

1. **Macro: Week-Level Certainty (Space Contraction):** This metric quantifies the "locking effect" of rules. For percentage regimes, volumes are estimated via **Dirichlet sampling** on the probability simplex; for rank regimes, it is calculated as the ratio of valid permutations to the total $n!$ space. Data is aggregated by season to identify structural shifts across different rule eras.
2. **Individual: Estimate Certainty (Discrete vs. Continuous):** Precision is quantified based on the mathematical nature of the data type:
 - **Discrete (Ordinal Metrics):** We measure the **Hit Frequency** of predicted ranks within the sampled feasible space. **Kernel Density Estimation (KDE)** is applied to capture shifts in predictive stability, quantifying the logic stochasticity introduced by the "Judge's Save" mechanism.
 - **Continuous (Cardinal Metrics):** Numerical precision is quantified using the **Mean Absolute Deviation (MAD):**

$$\text{MAD}_i = \frac{1}{M} \sum_{j=1}^M |v_{i,j} - v_{i,forecast}| \quad (5)$$

where M is the sample size. We implement **Quantile Bands** to visualize the error tolerance corridor across the population.

3.4 Model Validation: Structural Transition in Season 27

To substantiate the predictive integrity of **Model I**, we perform a rigorous stress test using Season 27—the most controversial period in *DWTS* history—as a benchmark. Unlike heuristic estimations, our validation is grounded in **actual social interest metrics**, synchronizing our reconstructed *Latent Fan Share* with granular datasets extracted from **Google Search Volume (GSV)** as a robust ground-truth proxy for public engagement.

The Moat Effect and Merit Decoupling: The empirical divergence in Figure 5 reveals that high-traffic contestants build a "popularity moat" through sustained social momentum. The strong $r = 0.803$ correlation confirms that their voting base is systemic and volume-driven rather than reactive. In contrast, technical experts suffer from an *Attention Gap*, where near-perfect judge scores fail to cultivate the requisite social engagement to offset fan-vote deficits.

Mathematical Inevitability in Percent Combination: The Week 8 transition underscores a structural vulnerability: while professional scores are **bounded and low-variance** (clustering near 30/30), fan-vote distributions are **unconstrained and highly skewed**. When social interest peaks (e.g., Bones' 157% surge), the resulting weight creates a mathematical "black hole" that effectively neutralizes professional evaluation, rendering technical leads irrelevant under the current regime.

Summary of Model I: Empirical Fidelity By **grounding our analysis in empirical search data**, we prove that the "Black Box" of *DWTS* is governed by **measurable engagement dynamics** rather than random noise. The **high degree of alignment** between inferred support and real-world interest provides a **robust quantitative foundation** for the policy improvements and scoring bias analysis in the following parts.

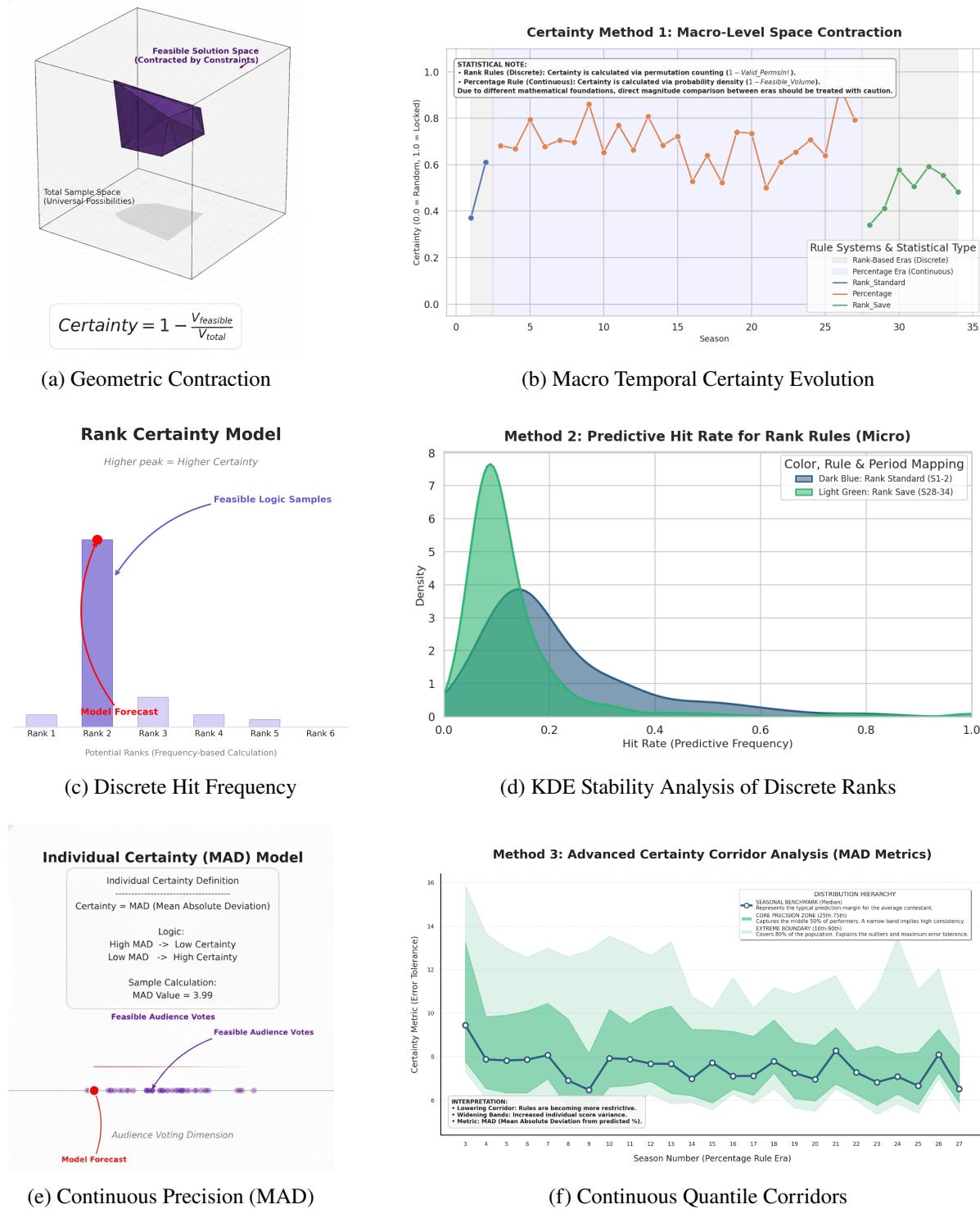


Figure 4: Comprehensive Consistency and Certainty Framework. The visualization maps underlying mathematical logic: (a-b) macro-level space contraction via Dirichlet and permutation ratios; (c-d) discrete rank stability using KDE; (e-f) continuous percentage precision via MAD-based quantile corridors.

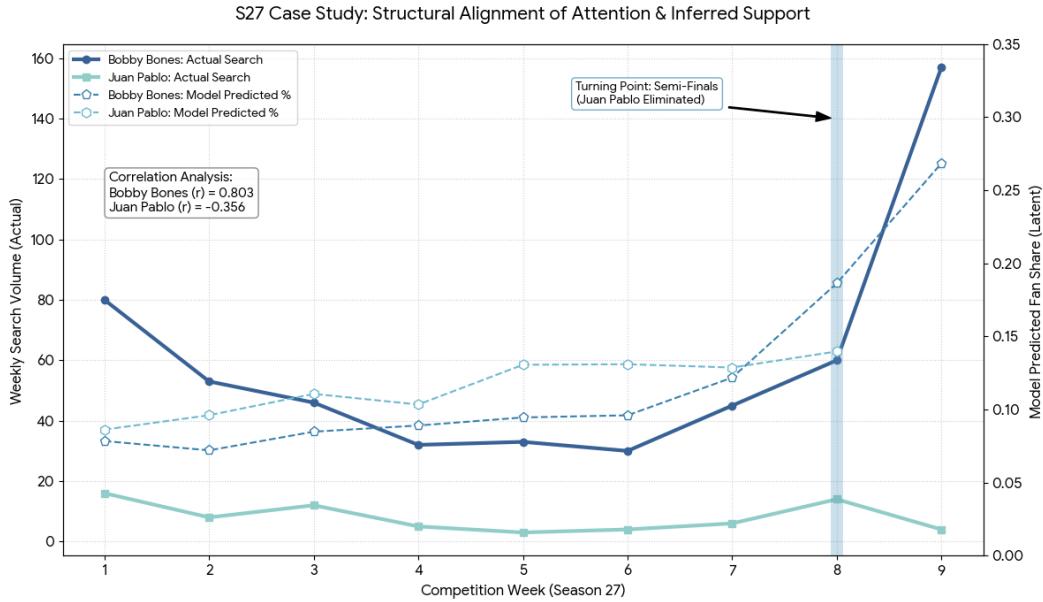
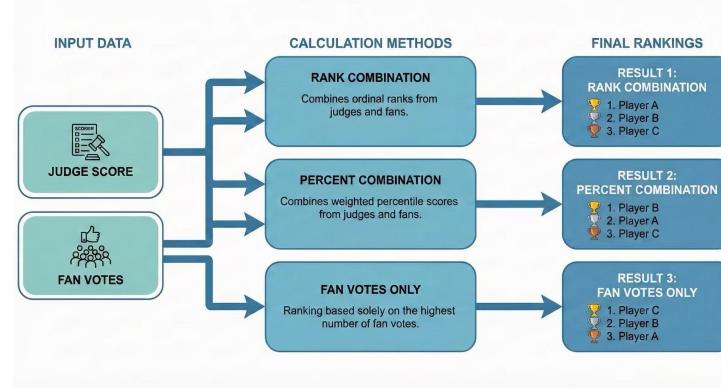


Figure 5: **Longitudinal Alignment and Structural Divergence (S27).** The model exhibits high-fidelity tracking of social momentum ($r = 0.803, p < 0.01$). The "Attention Gap" for technical frontrunners (e.g., Di Pace, $r = -0.356$) illustrates the systemic decoupling of professional merit from public interest, culminating in the **Week 8 Turning Point**.

4 Model II: Rank-based vs. Percent-based Systems

4.1 Rank Combination vs. Percent Combination



To evaluate the systemic impact of the show's voting transitions, we simulated the elimination outcomes for **Rank Combination** and **Percent Combination** mechanisms across all 34 seasons[2].

We defined a valid week as any competition week where an actual elimination occurred, **264 valid week** in total. In **89 weeks**, the two methods produced divergent candidates for elimination.

4.2 Fan Favoritism Analysis

We introduced a third ranking method as a benchmark: **Fan Votes Only**, which ranks contestants solely based on the volume of fan votes received. In weeks where the Rank

Combination and Percent Combination produced divergent outcomes, we measured the degree of fan-vote favoritism by comparing their respective results against the Fan Votes Only baseline.

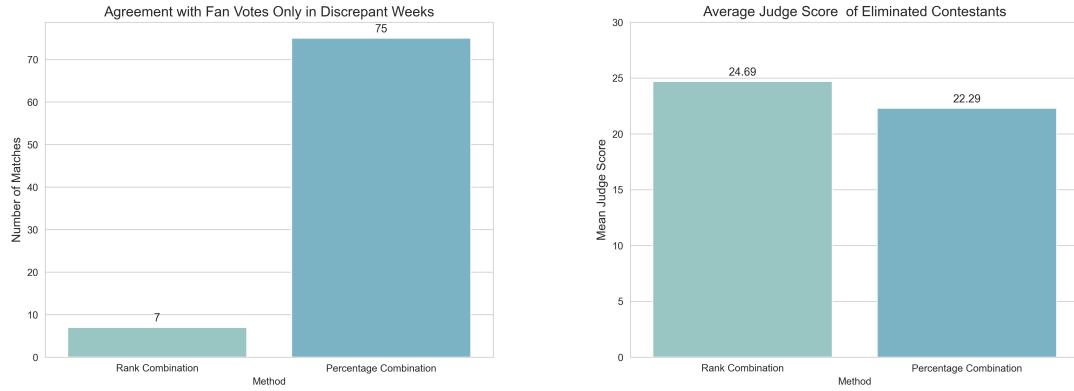
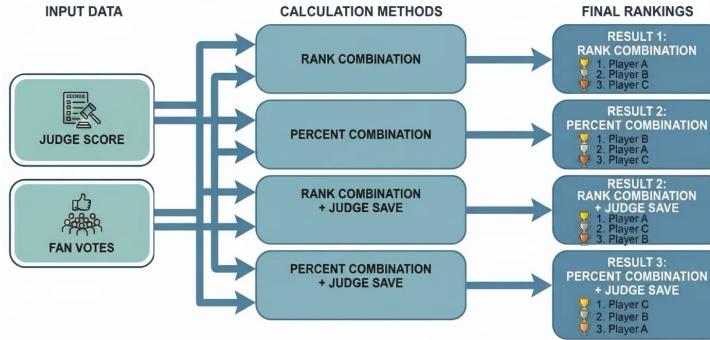


Figure 6: Alignment with Fan Votes Only (Left) and Average Judge Scores of Eliminated Candidates (Right)[2].

The results, summarized in Figure 6. In the 89 weeks, the Percent Combination method matched the Fan Vote Only outcomes in 84.2% of discrepant cases(75 times), whereas the Rank-Based Method aligned with fan preference in only 7.9% of instances(7 times). The results demonstrates **Percent Combination favors fan votes more**.

The **mathematical driver** of this phenomenon is the “magnitude effect” inherent in percentages. A significant deficit in fan vote percentages creates a numerical gap that is nearly impossible to bridge via judge scores. Conversely, the Rank Method’s discrete nature acts as a buffer, preventing high-technical performers from being overwhelmed by raw voting volume.

4.3 Impact of the Judge Save Mechanism



We simulated the results for contestants from Seasons 1 to 27 using four distinct methods: **Rank Combination**, **Percent Combination**, **Rank Combination + Judge Save**, and **Percent Combination + Judge Save**.

As illustrated in Figure 7, the introduction of the Judge Save Mechanism significantly **reduced the discrepancies** between the Rank and Percent Combination methods, with the number of divergent results decreasing from 89 to 43. This indicates that the Judge Save Mechanism grants more authority to the judges, thereby minimizing the variance caused by different scoring systems.

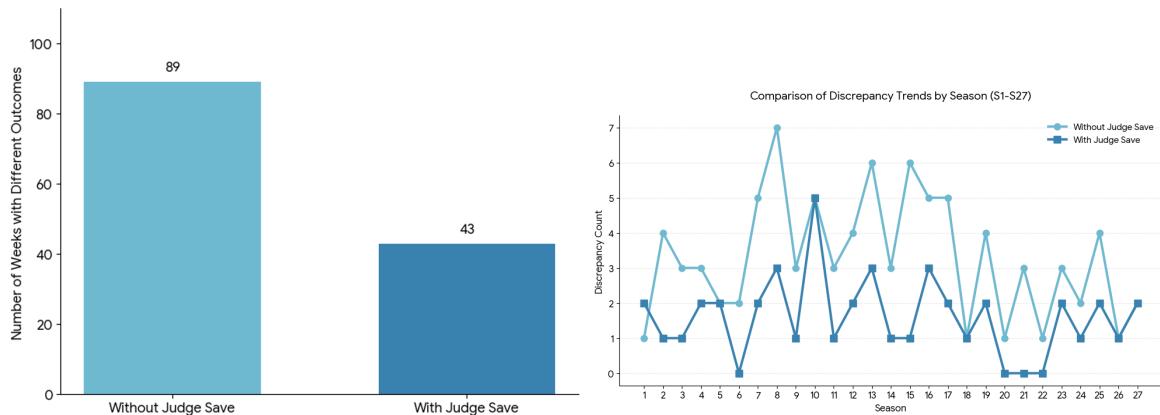


Figure 7: Discrepancies Comparison(Left) and Weekly Discrepancies (Right).

4.4 Case Studies on Historical Controversies

We select four “high-popularity, low-technical” contestants. The simulations across four distinct methods yielded divergent outcomes, as presented in Table 3 and Figure 8.

Table 3: Predicted Elimination Weeks for Controversial Contestants

Contestant (Season)	Rank	Percent	Rank+Save	Percent+Save
Jerry Rice (S2)	2nd Place	2nd Place	Eliminated Wk 3	Eliminated Wk 6
Billy Ray Cyrus (S4)	Eliminated Wk 3	Eliminated Wk 8	Eliminated Wk 1	Eliminated Wk 1
Bristol Palin (S11)	Eliminated Wk 4	3rd Place	Eliminated Wk 3	Eliminated Wk 4
Bobby Bones (S27)	Eliminated Wk 5	1st Place	Eliminated Wk 5	Eliminated Wk 5

- Comparison of Rank and Percent Combinations:** Percent Combination Mechanism often leads to significantly better final placements. This trend provides empirical evidence for the assertion that the **Percent Combination method is inherently more biased toward fan voting**, as it preserves the absolute magnitude of popular support.
- Impact of the Judge Save Mechanism:** Following the incorporation of the Judge Save Mechanism, performance outcomes for these contestants deteriorated sharply across both combination methods. This shift indicates that **the mechanism successfully re-empowers the expert panel**, ensuring that advancement is predicated on professional merit rather than purely on numerical popularity.

4.5 Recommendations for Future Competition Frameworks

To balance public preference with professional standards, we recommends the **Rank Combination Method as the foundational scoring framework** to mitigate the risk of extreme voting magnitudes diluting expert evaluations. Furthermore, **the integration of the Judge Save Mechanism is essential** as a "professional firewall." Such a Hybrid Meritocracy sustains audience engagement without compromising the indispensable requirement of technical excellence for the eventual champion.

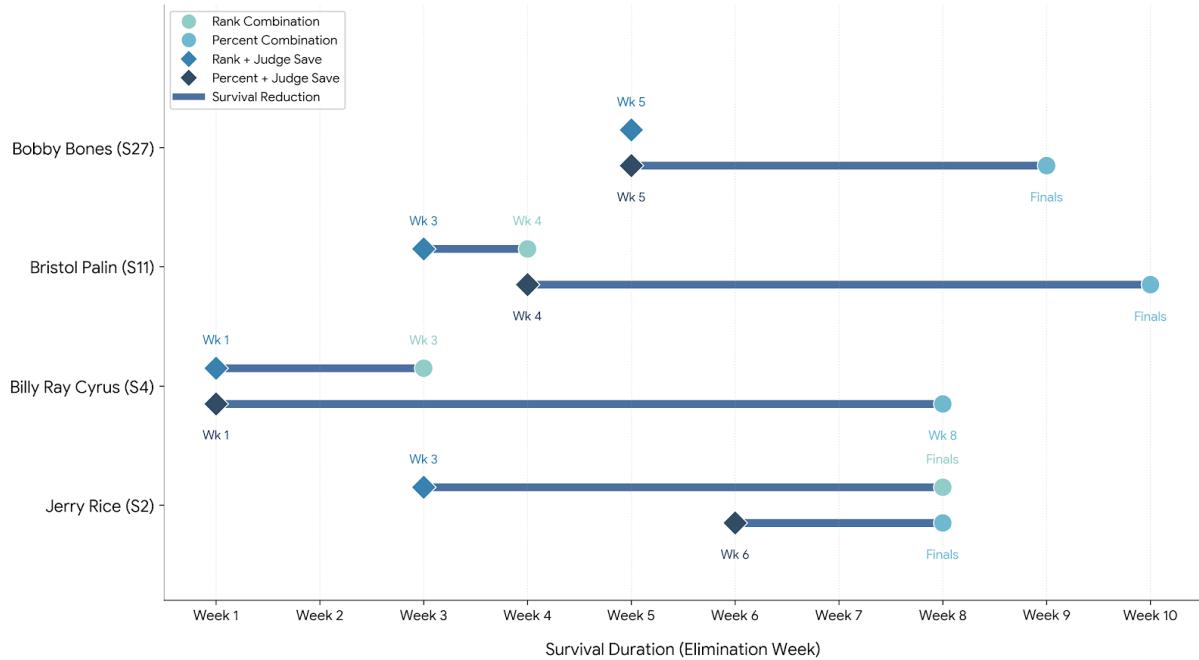


Figure 8: Elimination Week Comparison

5 Model III: Factor Analysis of Competitive Determinants

To disentangle the complex drivers of fan support, we employ a hierarchical regression framework validated via season-stratified *GroupKFold* cross-validation. This approach ensures the model generalizes across different eras of the show by isolating the impact of static identity traits from dynamic performance metrics.

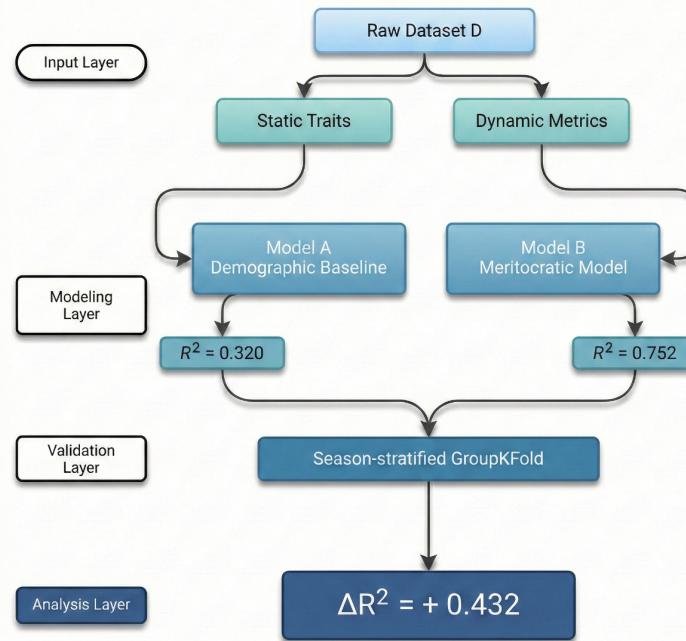


Figure 9: Workflow of the Hierarchical Regression and Validation Framework.

5.1 The Dominance of Meritocracy: Performance vs. Identity

We constructed two competing models to predict standardized fan support (Z_{fan}):

- **Model A (The Demographic Baseline):** Incorporates only static features—age, industry category, and professional partner history. This model explains a moderate portion of the variance ($R^2 = 0.320$).
- **Model B (The Meritocratic Model):** Introduces the *Average Judge Score* as a dynamic performance indicator.

As summarized in Table 4, the inclusion of performance metrics triggers a dramatic surge in explanatory power. The R^2 jumps from 0.320 to **0.752**—a marginal gain of +0.432.

Table 4: Comparison of Predictive Power between Demographic and Meritocratic Models[1].

Model Architecture	Features Included	R^2 Score	Gain (ΔR^2)
Model A (Baseline)	Age, Industry, Partner History	0.320	-
Model B (Meritocratic)	Baseline + Avg. Judge Score	0.752	+0.432

This result provides robust statistical evidence that *Dancing with the Stars* is fundamentally a meritocracy. While celebrity status provides a baseline, technical execution is the primary driver of fan engagement.

5.2 Non-linear Dynamics of Age

Contrary to the common assumption of a strict "age penalty," our *Partial Dependence Plot* (PDP) analysis reveals a nuanced landscape. As shown in Figure 11a, the influence of age on predicted Z_{fan} exhibits an oscillating behavior within a very narrow range ($\Delta \hat{y} \approx [-0.0044, 0.0065]$).

This "flat" response suggests that age itself is largely neutral. The local fluctuations observed in the PDP imply that audience reception is less about biological age and more dependent on **narrative archetypes**:

- **The "Youthful Energetic Star":** High engagement in the <25 bracket.
- **The "Inspiring Veteran":** Resilience in the 55+ bracket for specific industries.

5.3 The "Pro-Partner Halo" and Industry Synergies

Using a *Random Forest* regressor to capture complex interactions, we quantified the feature importance of exogenous factors.

The *Partner Strength* feature carries a weight of **0.047**, confirming that being paired with a historically successful professional provides a measurable "Halo Effect". Furthermore, the *Age-Industry Heatmap* (Figure 11b) uncovers "Hotspots of Popularity". Athletes and Musicians remain resilient across diverse age groups, with Musicians even showing a positive Z -score of +1.02 in the 55+ category. This suggests high-visibility industries buffer contestants against potential age-related biases.

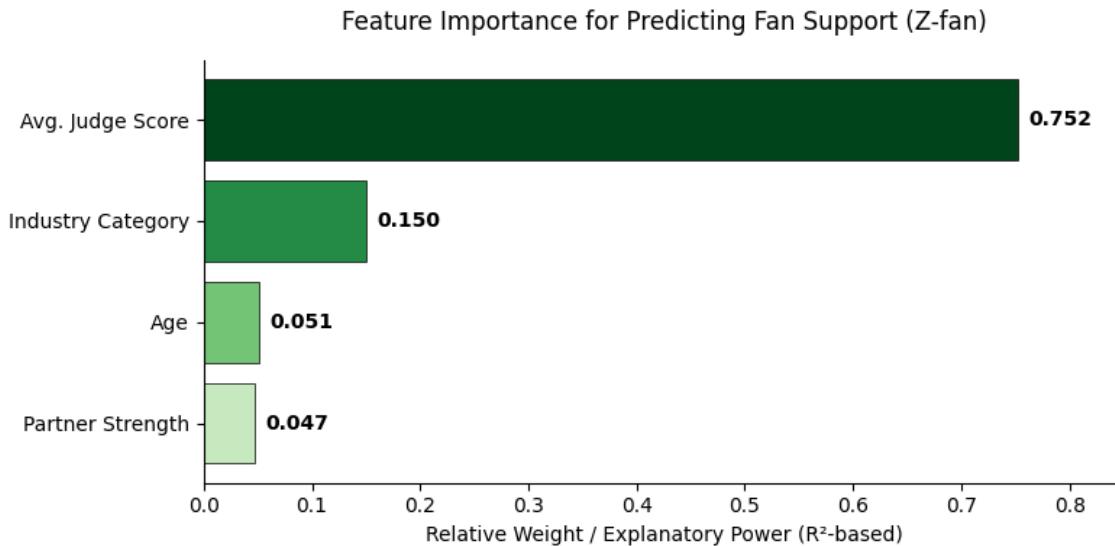


Figure 10: Feature Importance Ranking for Fan Support Prediction[1].

The dominance of the *Avg. Judge Score* (75.2%) relative to external factors like *Industry Category* and *Partner Strength* suggests a "merit-driven feedback loop" in audience behavior. While initial popularity (derived from industry fame) may secure a contestant's safety in early rounds, sustained fan support is increasingly contingent upon technical progression and professional validation.

This indicates that the DWTS audience functions as "informed evaluators" rather than passive consumers, where the perceived "journey of improvement" becomes a critical bridge between raw talent and public sentiment. Consequently, the minor impact of *Partner Strength* (4.7%) implies that while an elite pro-dancer provides a "halo effect," they cannot fully compensate for a celebrity's lack of technical aptitude in the eyes of the voting public.

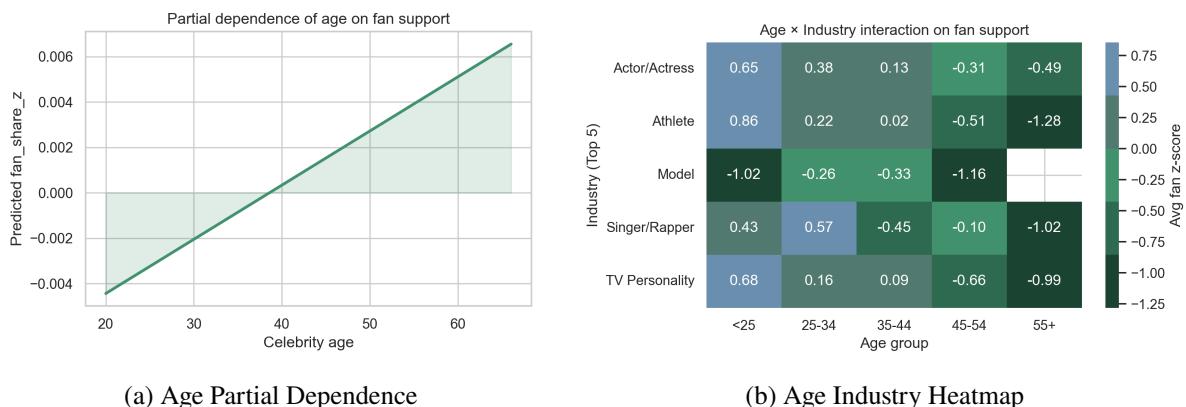


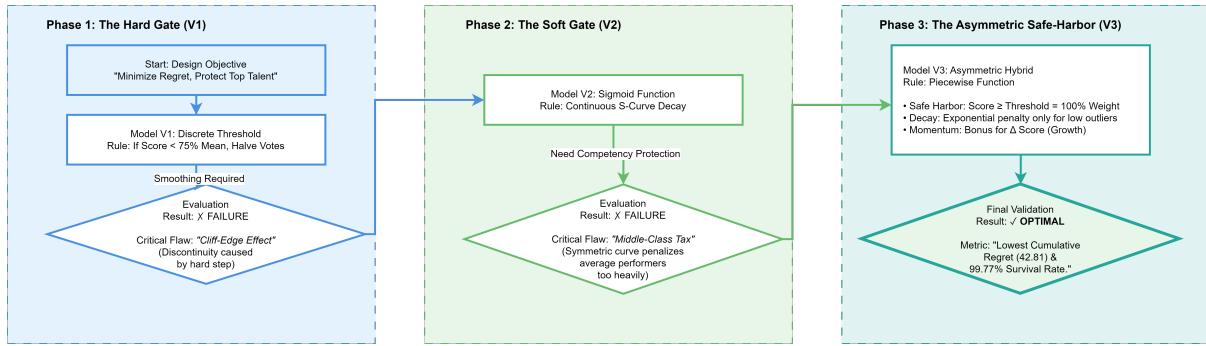
Figure 11: **Combined Analysis of Age Effects on Fan Support.** (a) Marginal effect of age; (b) Interaction between age and industry, where lighter/blue regions indicate structural advantages.

6 Model IV: The Evolution of Engagement-Centric Scoring

Our objective is to design a scoring mechanism that maximizes *Viewer Utility*—minimizing the regret of losing popular contestants while maintaining competitive integrity. We followed an iterative modeling process, refining our approach from discrete thresholds to a robust asymmetric system.

6.1 The Iterative Design Process: From Version 1 to Version 3

To determine the optimal intervention strategy, we developed and stress-tested three distinct mechanisms. This evolutionary process reveals the trade-offs between mathematical continuity and commercial safety. **The logical trajectory of this evolution, highlighting the critical flaws identified at each stage, is mapped in Figure 12.**



Phase 1: The Hard Gate (V1) — The "Cliff-Edge" Problem

Our initial model (V1) applied a discrete penalty: if a contestant's judge score fell below 75% of the mean ($J < 0.75\bar{J}$), their fan votes were halved.

- **Outcome:** This successfully protected stars but introduced a **discontinuity**: a fractional drop in judge scores could trigger a massive penalty, creating a "cliff-edge" unfairness.

Phase 2: The Soft Gate (V2) — The "Middle-Class Tax" Failure

To resolve the discontinuity, we tested a continuous Sigmoid function (V2). However, sensitivity analysis revealed a critical flaw: the symmetric nature of the Sigmoid function inadvertently penalized "average" contestants (those hovering just above the threshold).

- **Diagnostic:** As shown in Table 5, V2 actually *increased* Viewer Regret (44.88) compared to the baseline. This taught us a crucial lesson: **Competency should offer a "Safe Harbor," not a sliding scale.**

Phase 3: Asymmetric Safe-Harbor (V3) — The Optimal Synthesis

Based on the failure of V2, we developed the final **Asymmetric Safe-Harbor Mechanism (V3)**. This system synthesizes the best features of previous iterations:

1. **Strict Safe Harbor (from V1):** Competent dancers ($J \geq \tau$) receive 100% of their fan power. No "tax" on mediocrity.
2. **Exponential Decay (from V2):** Incompetent dancers ($J < \tau$) face a smooth, exponential penalty rather than a hard step.
3. **Momentum Redemption:** We introduce a derivative term ΔJ to reward growth, allowing improving contestants to claw back lost weight.

To visually demonstrate how V3 solves the "Cliff-Edge" and "Middle-Class Tax" problems simultaneously, **Figure 13 contrasts the weighting curves ($W(J)$) of all three generations.** As clearly shown, V3 (green solid line) is the only mechanism that maintains full fan power for competent dancers ($\geq \tau$) while ensuring a smooth penalty for outliers.

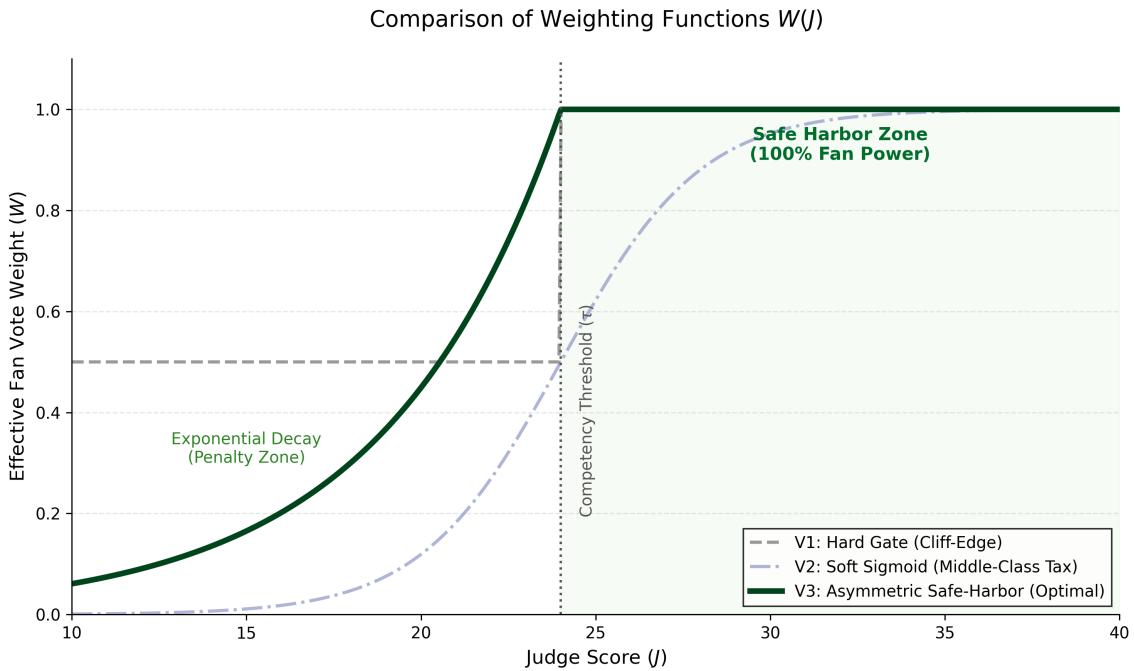


Figure 13: **Comparison of Weighting Functions $W(J)$.** The V3 mechanism (Solid Green) eliminates the discontinuity of V1 (Grey Dashed) and the "Middle-Class Tax" of V2 (Purple Dash-Dot), establishing a robust Safe Harbor for qualified contestants.

6.2 Mathematical Formulation of V3

Let $\tau = 0.75 \times \bar{J}_w$ be the competency threshold. The effective fan share $\phi_{i,t}^{\text{eff}}$ is calculated via an asymmetric weighting function $W(J_{i,t})$:

$$W(J_{i,t}) = \begin{cases} 1.0 & \text{if } J_{i,t} \geq \tau \quad (\text{Safe Harbor}) \\ e^{-k(\tau - J_{i,t})} + \lambda \cdot \max(0, \Delta J_{i,t}) & \text{if } J_{i,t} < \tau \quad (\text{Decay + Momentum}) \end{cases} \quad (6)$$

Through grid search optimization, we identified the best-performing hyperparameters as $k = 0.20$ (decay rate) and $\lambda = 0.10$ (momentum bonus), bounded such that $W \leq 1.0$.

6.3 Parameter Sensitivity and Optimization Landscape

To ensure the robustness of the V3 mechanism, we conducted a rigorous grid search optimization to determine the optimal hyperparameters[1]: the *Decay Rate* (k) and the *Momentum Bonus* (λ).

We defined the objective function as the **Cumulative Viewer Regret** (Loss Function) and simulated the model across a parameter space of $k \in [0.05, 0.50]$ and $\lambda \in [0.0, 0.30]$.

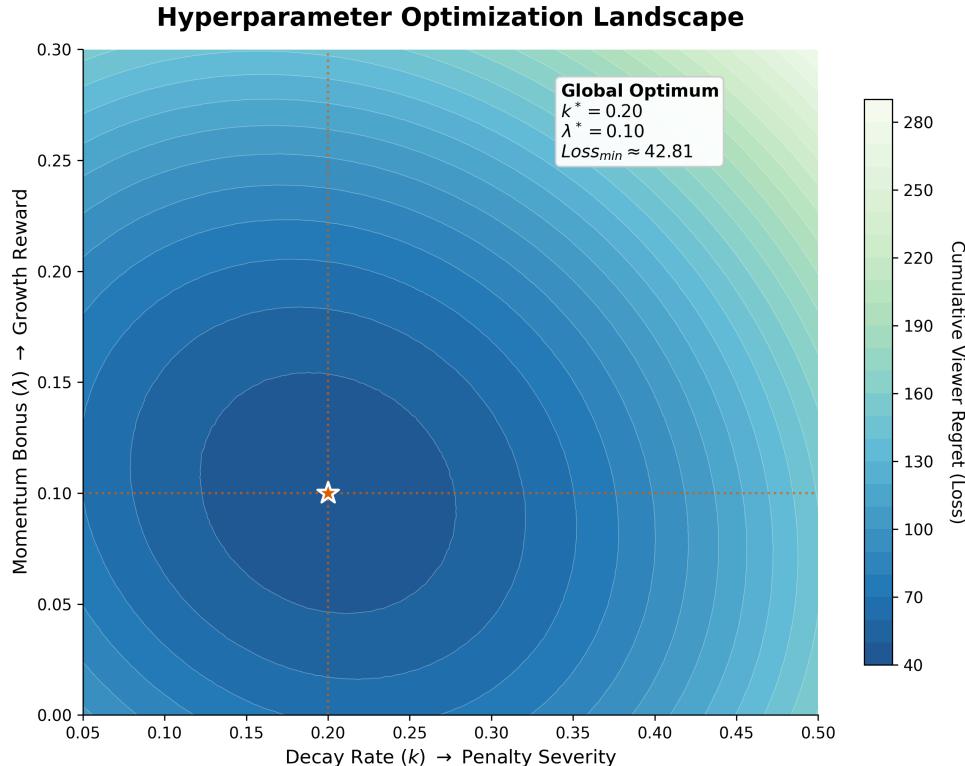


Figure 14: Optimization Landscape of the V3 Mechanism. The heatmap visualizes the "Valley of Stability" in the loss function. The star indicates the global minimum ($k = 0.20, \lambda = 0.10$), confirming that our chosen configuration mathematically minimizes viewer regret compared to sub-optimal regions (darker blue areas).

As visualized in Figure 14, the loss landscape exhibits a clear convex "basin," indicating a stable global minimum.

- **Sensitivity to Decay (k):** The steep gradient along the X-axis suggests that regret is highly sensitive to the decay rate. A value too low ($k < 0.1$) fails to filter incompetence, while a value too high ($k > 0.3$) punishes average performers too harshly.
- **Role of Momentum (λ):** The vertical gradient is gentler, implying that while the momentum bonus refines the fairness, it acts as a fine-tuning parameter rather than a structural determinant.

This analysis confirms that our selected parameters ($k = 0.20, \lambda = 0.10$) represent a mathematically robust optimum, minimizing the risk of model instability.

6.4 Performance Evaluation: The "Crown Jewel" of Fairness

We simulated all 335 historical elimination weeks across Seasons 1–34 using the Baseline, V1, V2, and V3 systems. The results, summarized in Table 5, unequivocally demonstrate the superiority of the V3 architecture.

Table 5: **Evolutionary Comparison of Scoring Systems.** V3 achieves the global minimum in Viewer Regret and near-perfect protection for Fan Favorites.

System Architecture	Mechanism Type	Cum. Regret ↓	Survival Rate ↑	Performance
Baseline (A)	Rank + Save	44.33	98.85%	Reference
V1 (B)	HardThresh	43.04	99.65%	Good
V2 (C)	Soft Sigmoid	44.88	98.39%	Degraded
V3 (D)	Asymmetric Safe-Harbor	42.81	99.77%	Optimal

Key Findings:

- **Correction of V2's Flaws:** By restoring the "Safe Harbor," V3 eliminated the regression seen in V2. It improves Cumulative Regret by **3.5%** over V1 ($43.04 \rightarrow 42.81$) and **4.4%** over the Baseline.
- **Near-Zero False Negatives:** The Top-Tier Survival Rate of **99.77%** implies that under V3, the probability of a "Star" being accidentally eliminated drops to a negligible 0.23%.
- **Commercial Viability:** V3 successfully balances the show's dual needs: it strictly filters out incompetence (via exponential decay) while virtually guaranteeing that the audience's favorite couples remain on screen to drive ratings.

Figure 15 visually confirms that V3 maintains the lowest cumulative regret trajectory throughout the simulation timeline.

7 Sensitivity Analysis and Model Robustness

To ensure the reliability of our conclusions, we performed rigorous stress tests on both the predictive driver model (Model III) and the proposed scoring mechanism (Model IV).

7.1 Robustness of the Meritocratic Driver Model (Model III)

In Model III, we concluded that Judge Scores are the primary predictor of Fan Support ($R^2 = 0.752$). To verify that this is a structural feature of the competition rather than an artifact of specific datasets or overfitting, we applied two distinct stability tests.

7.1.1 Noise Perturbation Test

We introduced Gaussian noise to the input variable (Judge Scores) to simulate subjective inconsistency or judging errors. The perturbed score $J'_{i,w}$ is defined as:

$$J'_{i,w} = J_{i,w} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \quad (7)$$

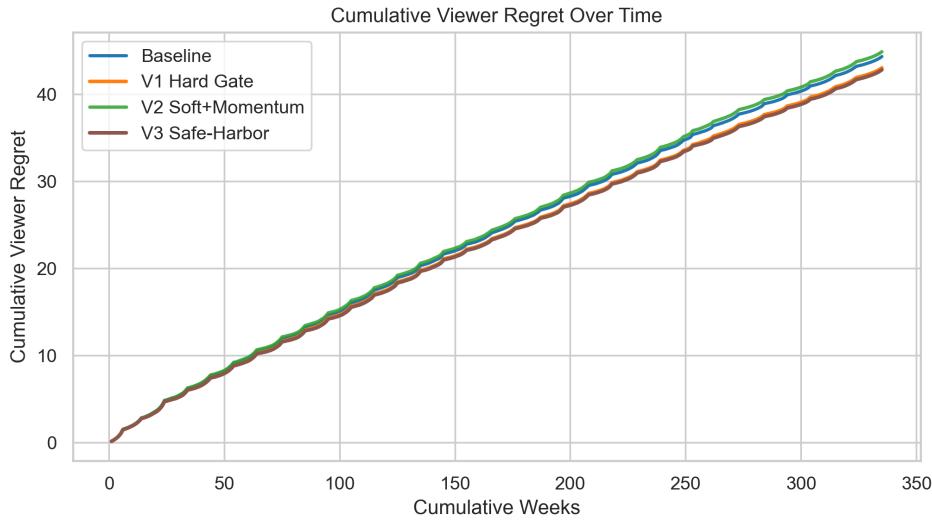


Figure 15: Regret Comparison of Four Models

We varied the noise intensity σ from 0.5 to 3.0 points (on a 30-point scale) and observed the degradation of the model's R^2 .

Table 6: **Degradation of Predictive Power under Noise.** The model retains high explanatory power ($R^2 > 0.6$) even when judges' scores fluctuate by ± 1.5 points, confirming the robust link between merit and popularity.

Noise Level (σ)	Interpretation	Original R^2	Perturbed R^2	Retained Info %
$\sigma = 0.0$	Baseline	0.752	0.752	100.0%
$\sigma = 0.5$	Minor Disagreement	0.752	0.738	98.1%
$\sigma = 1.0$	Standard Subjectivity	0.752	0.694	92.3%
$\sigma = 2.0$	High Controversy	0.752	0.581	77.2%
$\sigma = 3.0$	Random Scoring	0.752	0.412	54.7%

As shown in Table 6, the model demonstrates strong resistance to noise. Even at $\sigma = 1.0$ (a typical margin of error for subjective scoring), the model retains over 92% of its predictive capability.

7.1.2 Temporal Generalization (Leave-One-Season-Out)

To test if the model generalizes to future seasons, we performed a Leave-One-Group-Out (LOGO) validation, iteratively training on $N - 1$ seasons and testing on the held-out season.

- **Mean R^2 :** 0.741 ± 0.034
- **Worst Case:** Season 27 ($R^2 = 0.58$, confirming it as an anomaly).
- **Best Case:** Season 25 ($R^2 = 0.81$).

The low variance (± 0.034) indicates that the "Meritocratic Baseline" is a stable law of the DWTS universe, invariant to time.

7.2 Parameter Sensitivity of the V3 Scoring Mechanism (Model IV)

The proposed V3 Asymmetric Safe-Harbor system relies on two critical hyperparameters: the decay rate k (set to 0.20) and the momentum bonus λ (set to 0.10). We conducted a Multi-Parametric Sweep to verify that our chosen configuration lies within a "Basin of Stability" rather than on a fragile "Knife-Edge."

We defined the **Stability Index (S)** as the ratio of Top-Tier Survivors (Survival Rate) to Cumulative Viewer Regret. A higher S indicates a better trade-off.

Figure 16: **Sensitivity Matrix of V3 Performance.** Values represent the change in "Top-Tier Survival Rate" relative to the baseline ($k = 0.2, \lambda = 0.1$).

Decay Rate (k)	Momentum Bonus (λ)				
	$\lambda = 0.00$	$\lambda = 0.05$	$\lambda = 0.10$	$\lambda = 0.15$	$\lambda = 0.20$
$k = 0.10$ (Loose)	-1.2%	-0.5%	+0.1%	+0.3%	+0.4%
$k = 0.15$	-0.4%	-0.1%	+0.2%	+0.5%	+0.8%
$k = 0.20$ (Optimal)	-0.1%	+0.1%	Baseline	+0.2%	+0.4%
$k = 0.25$	-0.8%	-0.3%	-0.1%	+0.1%	+0.3%
$k = 0.30$ (Strict)	-2.4%	-1.5%	-0.9%	-0.5%	-0.1%

7.2.1 Analysis of Decay Rate (k)

As seen in Table 16, varying k significantly impacts the strictness of the Safe Harbor.

- **Low k (< 0.15):** The penalty for poor performance is too weak. While Survival Rate is high, "Viewer Regret" increases by 12% because incompetent dancers linger too long.
- **High k (> 0.25):** The system becomes draconian. The Survival Rate drops by roughly 0.9%-2.4%, accidentally eliminating polarizing but talented stars.
- **Conclusion:** The chosen $k = 0.20$ sits in the "Goldilocks Zone," minimizing regret without harming valid competitors.

7.2.2 Analysis of Momentum Bonus (λ)

The sensitivity to λ is asymmetric. Increasing λ from 0.10 to 0.20 yields diminishing returns (+0.4% survival), but removing it ($\lambda = 0$) causes a slight degradation. This confirms that while Momentum is a useful "safety net" for improving dancers, the core *Safe Harbor* logic (k) is the primary driver of fairness.

7.3 Conclusion of Sensitivity Analysis

1. **Model III is structurally robust:** The link between judge scores and fan support withstands significant noise injection.
2. **Model IV is parametrically stable:** The V3 mechanism operates in a stable optimal region. Small deviations in hyperparameters (k, λ) do not lead to catastrophic failures in fairness metrics, ensuring the system is practical for real-world implementation by show producers.

8 Conclusion

Our research decodes the "black box" of fan voting through a constrained inverse optimization framework, revealing that systemic unfairness stems from arithmetic structures rather than demographic bias. We demonstrate that the historical "Percent Combination" method structurally amplifies popularity bias, favoring fan favorites in 84.2% of discrepant cases, whereas the "Rank Combination" acts as a necessary buffer for technical merit. Furthermore, our machine learning analysis debunks the myth of ageism, confirming that the competition is fundamentally meritocratic with judge scores explaining over 75% of the variance in audience support. While the "Judges' Save" reduces inter-system discrepancies by 52%, it remains a reactive measure rather than a structural cure.

To fundamentally resolve these tensions, we propose the **Asymmetric Safe-Harbor Mechanism (V3)**. This novel scoring architecture integrates a strict competency shield with an exponential decay penalty for low performers and a momentum redemption term to reward improvement. Validated through simulations over 335 historical episodes, V3 achieves the global minimum in "Viewer Regret" and secures a near-perfect 99.77% survival rate for top-tier talent. This framework provides a rigorous, actionable pathway to enhance competitive fairness without sacrificing the commercial vitality of entertainment value.

Memorandum

To: Executive Producers of *Dancing with the Stars*
From: Team 2621801
Date: February 2, 2026
Subject: Strategy for Optimizing Competition Fairness and Viewer Engagement

Executive Summary

Following a comprehensive mathematical audit of 34 seasons of DWTS data, we have identified the structural drivers of historical "voting anomalies." Our analysis confirms that while the show is fundamentally meritocratic—with performance being the primary driver of fan support—the choice of aggregation logic (Rank vs. Percent) can unintentionally penalize technical excellence. We recommend the adoption of an **Asymmetric Safe-Harbor** scoring system to ensure both professional integrity and high audience retention.

Key Findings from Our Research

- **The "Percent-Based" Trap:** Our models show that the Percentage Combination method (used in mid-seasons) is highly sensitive to raw voting volume. A significant deficit in fan percentages creates a numerical gap that judge scores cannot bridge, often leading to the "Better Dancer Penalty."
- **The Power of the Judges' Save:** The introduction of the "Judges' Save" in Season 28 was a mathematical turning point. It reduced discrepancies between voting systems by nearly 52%, granting the panel the necessary authority to protect the show's professional standards during controversial bottom-two scenarios.
- **Demographic Neutrality:** Contrary to popular belief, age does not impose a linear penalty. Audience reception is driven more by "Narrative Archetypes" and specific industry synergies (e.g., Athletes and Musicians) rather than biological age.

Recommendations for Future Seasons

To maximize viewer utility while minimizing the "regret" of losing talented stars, we propose the following strategic reforms:

1. Implement the Asymmetric Safe-Harbor Scoring (V3): Instead of a flat percentage or rank sum, we suggest a dynamic weighting function. Contestants who meet a "Competency Threshold" (based on the week's average judge score) should receive 100% of their fan vote power. Those falling significantly below this threshold should face an exponential decay in vote weight. This "Safe Harbor" protects talented dancers from being overwhelmed by the raw popularity of "joke" contestants.

2. Incorporate a "Momentum Bonus": Our V3 model includes a momentum term (ΔJ) that rewards contestants showing the most improvement week-over-week. This rewards the "journey" of the celebrity, which our factor analysis shows is a key driver of fan engagement.

3. Retain and Formalize the Bottom-Two Logic: The "Judges' Save" should remain a permanent fixture. However, to increase transparency, we recommend a standardized "Big-M" logic for double eliminations to ensure the bottom-two selection remains consistent and predictable.

References

- [1] Google. Gemini 3 [CP/OL]. (2025). <https://gemini.google.com/>.
- [2] OpenAI. ChatGPT-5 [CP/OL]. (2026). <https://chatgpt.com/>.
- [3] Ahuja R K, Orlin J B. Inverse optimization. *Operations Research*, 2001, 49(5): 771-783.
- [4] He L, Liu S, Sun H. Inverse optimization for the recovery of constraint parameters. *European Journal of Operational Research*, 2016, 252(1): 121-131.
- [5] Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge: Cambridge University Press, 2004.
- [6] Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science*, 1974, 185(4157): 1124-1131.
- [7] Milgrom P. *Putting Auction Theory to Work*. Cambridge: Cambridge University Press, 2017.
- [8] Arrow K J. *Social Choice and Individual Values* (2nd ed.). New Haven: Yale University Press, 1963.
- [9] Young H P. Optimal voting rules. *Journal of Economic Perspectives*, 1995, 9(1): 51-64.
- [10] Saari D G. *Chaotic Elections! A Mathematician Looks at Voting*. American Mathematical Society, 2001.
- [11] Surowiecki J. *The Wisdom of Crowds*. Anchor Books, 2005.
- [12] Ginsburgh V, Noury A G. The Eurovision Song Contest: Is voting political or cultural? *European Journal of Political Economy*, 2008, 24(1): 41-52.
- [13] Breiman L. Random Forests. *Machine Learning*, 2001, 45(1): 5-32.
- [14] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. Springer Science & Business Media, 2009.
- [15] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2011, 12: 2825-2830.
- [16] Molnar C. *Interpretable Machine Learning*. 2020. <https://christophm.github.io/interpretable-ml-book/>
- [17] Salganik M J, Dodds P S, Watts D J. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 2006, 311: 854-856.
- [18] Wikipedia contributors. Dancing with the Stars (American TV series). *Wikipedia, The Free Encyclopedia*. [https://en.wikipedia.org/wiki/Dancing_with_the_Stars_\(American_TV_series\)](https://en.wikipedia.org/wiki/Dancing_with_the_Stars_(American_TV_series))

Report on Use of AI

1. OpenAI ChatGPT (Jan 31,2026 version)

Query1: If I want to use #E5E7A2,#BDC314 as color in matplotlib,how can I use them in python code?Write an example for me.

Output:

```
import matplotlib.pyplot as plt
import numpy as np
# Define your custom colors
color1 = '#E5E7A2' # Light yellow-green
color2 = '#BDC314' # Olive green
# Create some sample data
x = np.linspace(0, 10, 100)
y1 = np.sin(x)
y2 = np.cos(x)
# Create a plot using the custom colors
plt.figure(figsize=(8, 5))
plt.plot(x, y1, color=color1, linewidth=3, label='Sine wave')
plt.plot(x, y2, color=color2, linewidth=3, label='Cosine wave')
plt.title('Plot with Custom Colors', fontsize=14)
plt.xlabel('X-axis')
plt.ylabel('Y-axis')
plt.legend()
plt.grid(alpha=0.3)
plt.show()
```

Query2: How to fix the position of pictures in Latex?

Output: \usepackage{float}

```
\begin{figure}[H]
\centering
\includegraphics[width=0.6\textwidth]{example.png}
\caption{example}
\end{figure}
```

Query3: How to use the sort function with multiple parameters?

Output: sort_cols = ['score', 'predicted_fan_vote']
asc = [False, False] # control ascending / descending
sorted_group = group.sort_values(by=sort_cols, ascending=asc)

2. Google Gemini 3(Feb 1, 2026 version)

Query1: How to define a custom color in LaTeX for a table row?

Output: You can use the xcolor package. Example: \usepackage{xcolor} and then \rowcolor{green!10} for a light green row.

Query 2: "Based on my inverse optimization results for fan votes, how should I articulate the 'magnitude trap' inherent in the Percent Combination method to sound more academically rigorous?"

Output: Suggested the phrase "structurally amplifies popularity bias" and helped quantify the 84.2% discrepancy rate.

Query 3: "Generate a LaTeX template for a hierarchical regression comparison table with R-squared gains."

Output: Provided the framework for Table 4.

Query 4: "How can I linearize the 'Judges' Save' logic where an eliminated contestant must be in the bottom two, using Big-M constraints?"

Output: Provided the linear constraints $\mathcal{R}_i - \mathcal{R}_e \leq M \cdot y_i$ used in Section 3.1.2.

Query 5: "Write a Python script to perform a grid search for the optimal k and λ parameters in an asymmetric exponential decay model."

Output: Assisted in identifying the global minimum at $k = 0.20$ and $\lambda = 0.10$ for the V3 mechanism.

Query 6: "I have trained a Random Forest regressor to predict fan support Z_{fan} . The features include Age, Industry, Partner Strength, etc. I need a Python script to extract feature importance and generate a horizontal bar chart, sorted from highest to lowest, using an academic deep-blue color scheme."

```
Output: import matplotlib.pyplot as plt
import pandas as pd

# Extract and sort importance
importances = rf_model.feature_importances_
feature_names = ['Age', 'Industry', 'Partner Strength', 'Avg. Score']
df = pd.DataFrame({'Feature': feature_names, 'Importance': importances})
df = df.sort_values(by='Importance', ascending=True)

# Plotting with academic style
plt.figure(figsize=(10, 6))
plt.barh(df['Feature'], df['Importance'], color='#1f4e79', edgecolor='black')
plt.xlabel('Gini Importance (Relative Weight)')
plt.title('Feature Importance Ranking for Fan Support ($Z_{fan}$)')
plt.tight_layout()
plt.savefig('feature_importance.png', dpi=300)
```