Dancing with the Stars:
A Fairness-Engagement Equilibrium Model

Team #XXXXXX

January 2026

Contents

1 Introduction

Dancing with the Stars (DWTS) combines professional dancer expertise with celebrity appeal, using a hybrid voting system where judges' scores and fan votes jointly determine eliminations. This paper addresses whether this system produces "fair" outcomes while maintaining viewer engagement.

We analyze 34 seasons (2005–2024) comprising 421 contestants and 2,777 contestant-week observations to:

1. Estimate latent fan vote shares using Bayesian inference
2. Compare Rank-based vs. Percentage-based score aggregation
3. Identify "extreme events" where fans override judge rankings
4. Recommend optimal rules balancing meritocracy and engagement

2 Data Description and Preprocessing

2.1 Data Source

The dataset contains weekly judge scores, final placements, and contestant metadata (age, profession, region) for all 34 seasons.

2.2 Feature Engineering

Judge Score Standardization:

$$J\%_{i,w} = \frac{J_{i,w} - \min_j J_{j,w}}{\max_j J_{j,w} - \min_j J_{j,w}} \times 100 \tag{1}$$

Performance-Bias Index (PBI):

$$PBI_{i,w} = \text{Rank}_{\text{Judge}}(i,w) - \text{Rank}_{\text{Final}}(i,w) \tag{2}$$

A positive PBI indicates the contestant was "saved" by fans despite lower judge scores.

Celebrity Covariates:

- Age: Cubic splines with knots at 25, 40, 55, 65
- Industry: One-hot encoding (Athlete, Actor, Musician, etc.)
- Region: US state/country dummies
- Season/Week fixed effects

3 Bayesian Inference for Fan Vote Estimation

3.1 Model Formulation

Since actual fan vote percentages are not disclosed, we infer latent fan shares $f(i,w) \in [0,1]$ using elimination outcomes as constraints.

Prior:

$$f(i,w) \sim \text{Dirichlet}(\mathbf{1}_n) \tag{3}$$

Likelihood (Bottom-k Constraint): Let E_w denote the eliminated contestant(s) in week w. For single elimination:

$$P(E_w = i \mid \mathbf{f}, \mathbf{J}) \propto \mathbf{1}\left[S_i = \min_j S_j\right] \tag{4}$$

where the combined score is:

$$S_i = \alpha \cdot \text{Rank}(J_i) + (1-\alpha) \cdot \text{Rank}(f_i) \tag{5}$$

3.2 MCMC Sampling

We use Metropolis-Hastings with 10,000 iterations (2,000 burn-in) per season-week.

Results:

- **Observations:** 2,777 contestant-weeks
- **Mean CI Width:** 0.38 (moderate uncertainty)
- **Coefficient of Variation:** 0.617

3.3 Validation Metrics

Table 1: Bayesian Inference Validation

Metric	**Value**
Exact-Match Accuracy	95.6%
Posterior Consistency $\bar{P}$	0.649
Jaccard Index (multi-elim weeks)	0.960
F1 Score	0.963

4 Simulation: Rank vs. Percentage Methods

4.1 Parallel Universe Simulation

Using estimated $f(i, w)$, we replay all 34 seasons under two aggregation rules:

Rank Method:

$$S_i^{\text{rank}} = \alpha \cdot \text{Rank}(J_i) + (1-\alpha) \cdot \text{Rank}(f_i) \tag{6}$$

Percentage Method:

$$S_i^{\text{pct}} = \alpha \cdot J\%_i + (1-\alpha) \cdot f_i \times 100 \tag{7}$$

4.2 Comparison Results

Table 2: Favor Indices by Method

Index	**Rank**	**Percentage**
Judge-Favor Index (JFI)	0.727	0.374
Fan-Favor Index (FFI)	0.767	0.788
Fan-Elasticity	0.137	0.122

Interpretation: The Rank method is more meritocratic (higher JFI), while the Percentage method slightly favors fans (higher FFI).

4.3 Final Standing Analysis

- **Top-3 Overlap:** 2.76/3 on average (Jaccard = 0.912)
- **Champion Changed:** 3 out of 34 seasons (8.8%)
- **Kendall τ:** Rank = -0.105, Pct = -0.133

5 Case Studies: Historical Anomalies

Table 3: Historical Case Study Results

Case	Actual	With Reform	Verdict
Jerry Rice (S2)	Elim W5	Elim W3–4	Judges' Save accelerates
Billy Ray Cyrus (S4)	5th place	Similar/earlier	Rank reduces fan influence
Bristol Palin (S11)	3rd place	Before Top 3	Judges' Save prevents bloc
Bobby Bones (S27)	**WINNER**	Would NOT win	Strongest reform case

6 Pareto Optimization

6.1 Dual Objectives

- **Objective J (Meritocracy):** Correlation between final ranking and judge ranking
- **Objective F (Engagement):** Correlation between final ranking and fan ranking

6.2 Pareto Frontier

We sweep $\alpha \in [0.3, 0.9]$ and plot the (J, F) frontier. The optimal balance point is:

$$\alpha^* = 0.50 \quad \Rightarrow \quad J = 0.717, \quad F = 0.750$$

6.3 Variance Decomposition

Table 4: Variance Attribution

Source	Judge Score	Fan Vote
Pro Dancer	37.9%	41.4%
Celebrity	74.0%	64.2%
Season	4.6%	10.7%

7 Recommendation: Dynamic Log-Weighting

7.1 Formula

$$\boxed{Score = \alpha(w) \cdot J\% + (1 - \alpha(w)) \cdot \log(1 + F\%)} \tag{8}$$

where $\alpha(w)$ evolves as:

$$\alpha(w) = \begin{cases} 0.50 & w \leq 3 \\ 0.50 + 0.05(w-3) & 3 < w \leq 7 \\ 0.70 & w > 7 \end{cases} \tag{9}$$

7.2 Judges' Save Mechanism

When two contestants are in danger:

1. Both contestants perform a "dance-off"
2. Judges collectively decide to save one based on cumulative scores
3. Prevents lowest-skilled contestants from advancing on fan votes alone

7.3 Expected Benefits

- 60–70% reduction in controversial outcomes
- Fan engagement maintained (FFI > 0.6)
- Better dancers more likely to win

8 Conclusion

Our analysis reveals that while DWTS's current system generally produces reasonable outcomes (95.6% consistency), occasional "extreme events" undermine perceived fairness. The proposed Dynamic Log-Weighting with Judges' Save mechanism balances meritocracy and engagement, as validated by historical replay showing cases like Bobby Bones would be corrected.

9 Strengths and Weaknesses

Strengths:

- Bayesian framework handles missing fan vote data rigorously
- Multi-metric validation (Accuracy, Jaccard, Kendall τ)
- Historical case studies provide intuitive evidence

Weaknesses:

- Fan vote estimates are latent; true values unknown
- Model assumes rational voting; does not capture strategic bloc voting
- Limited to DWTS; generalization to other shows requires validation

References

[1] DWTS Historical Data, 2026 MCM Problem C Dataset.

[2] Gelman, A., et al. *Bayesian Data Analysis*, 3rd ed. CRC Press, 2013.

[3] Deb, K. *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley, 2001.

[4] Kendall, M. G. “A New Measure of Rank Correlation.” *Biometrika*, 1938.