Calibration and Comeback Risk in NBA In-Game Win-Probability Models

Dylan Mealey* Sean Mealey† Steven Miller‡
Serigne N'Dione§

Department of Mathematics, Williams College

August 3, 2026

Abstract

Real-time win-probability numbers appear everywhere in NBA broadcasts. But the accuracy of these models are rarely scrutinized in academic settings or NBA media discourse, despite their evaluations often failing in the extremes. We pulled 6,958 regular-season game moments from the 2019-20 through 2024-25 seasons, where ESPN's model gave one team at least an 80 percent win probability. We then subjected these figures to interrogation, surfacing any patterns between the leads that endured and those that collapsed. Teams in the sample won 84.4 percent of the time, against an average stated ESPN probability of approximately 81 percent. Eighty percent of observations are concentrated in the 80-83 percent band, but this aggregate conceals substantial variation beneath its surface. We identified subsets of games with higher and lower true comeback rates. Conditioning on the threshold, a logistic regression trained on alternating seasons produced an out-of-sample AUC of 0.666 where AUC, or area under the curve, measures a model's predictive power by assigning a number between .5 and 1. An AUC of 0.666 means that given a random comeback game and a non random comeback game, the model assigns the higher comeback probability to the actual comeback game about 66.6 percent of the time. We found that the strongest predictors of a comeback are pre-game measures of team quality; in-game snapshot variables at the threshold moment contribute little, consistent with the Brownian nature of NBA basketball.

Acknowledgements

The first, second, and fourth named authors gratefully acknowledge support from the DRFC at Williams College. We also thank Rick Cleary for his contributions.

*dmm10@williams.edu
†sdm6@williams.edu
‡sjm1@williams.edu
§sn11@williams.edu

1 Introduction

Win-probability estimates have become a standard feature of NBA broadcasts, providing a numerical summary of how likely each team is to win as a game unfolds. Public discussion of these estimates, however, is often shaped by memorable exceptions. Dramatic comebacks receive disproportionate attention, while the far more common cases in which high-probability teams ultimately win rarely become part of the conversation. Individual examples therefore say little about whether a model is well calibrated overall. That question requires evaluating its predictions across a large sample of comparable game states.

Game 4 of the 2026 NBA Finals provides a motivating example in which, at the same game state with a 29-point deficit, ESPN assigned a 99 percent win probability to the leading team, while our model assigned a 46 percent chance of comeback under the same conditions. The Finals comeback serves only as motivation. Whether ESPN's estimate or ours better reflected the underlying situation cannot be established from a single game. The broader question is whether published win probabilities remain well calibrated across thousands of similar game states, and whether systematic patterns emerge when they do not.

This paper examines the gap between ESPN's predictions and actual outcomes. It asks whether such models are well calibrated on average, when their judgment begins to fail, and in which direction those errors tend to move. Our dataset includes 6,958 NBA regular-season game moments from the 2019–20 through 2024–25 seasons, focusing on the game state when ESPN assigns one team an in-game win probability of at least 80 percent. Those teams won 84.4 percent of the time. We also note that 38.4 percent of games crossed the 80 percent threshold with an assigned win probability between 80–81 percent. The true win probability in this band was 82.1 percent. Additionally, 41.8 percent of games crossed the win probability threshold between 81–83 percent, boasting a true win probability of 85.5 percent, and 14.8 percent of games crossed between 83–85 percent with a true win probability of 86.5 percent. On the surface, this appears to be proper calibration. But aggregate accuracy can obscure where errors are concentrated. Across 21 probability thresholds between 80 and 100 percent, we find statistically significant miscalibration at eight of them. In the 81–91 percent range, the model is too conservative. The winning teams in that range win more often than the displayed probability suggests. Above 95 percent, the pattern reverses. The model overstates its confidence. At the 98 percent threshold, the divergence is the largest, with realized outcomes falling nearly 8 percent short of predictions.

This paper makes three contributions. First, it evaluates the calibration of ESPN's in-game win-probability model across high-probability game states, identifying thresholds and game contexts where miscalibration is statistically significant. Wilson (1927) confidence

intervals are computed for each probability bucket to ensure consistent inference across the distribution. Second, it identifies the game-state and organizational factors that best predict comeback outcomes conditional on the 80 percent threshold being reached. The variable set is organized into four groups. In-game snapshot variables are computed from play-by-play logs at the threshold moment. They capture shooting efficiency and deviation from season averages, turnover rate, pace, defensive rebounding rate, foul accumulation and bonus status, and remaining timeouts. Pre-game team strength is measured by each team's win-loss record up to that point in the season, along with the prior season's net rating differential, crunch-time net rating differential, and starter quality. Star-player variables indicate whether each team's highest-impact player is on the floor at the snapshot moment and the degree to which team performance is dependent on that player. Together, these variables capture systematic variation in comeback probability that is not explained by the baseline win-probability model. Third, we selected 9 of these features and trained a linear regression model on the 2021, 2023, and 2025 seasons, testing it on the 2020, 2022, and 2024 seasons. We find that our model consistently outperforms conventional probability metrics while capitalizing on mispricings in predictive markets: generating backtested returns of 64, 45, and 29 percent across the 2020, 2022, and 2024 seasons. Thus, we have constructed a lens that captures a predictive edge where market forecasts fail.

2 Related Work

Cooper et al. (1992) established a baseline that has largely held. Across 200 professional basketball games, alongside data from baseball, hockey, and football, they identified the point at which game outcomes become effectively settled. Teams leading after three quarters won approximately 80 percent of the time. The authors purport that this finding is consistent across football after the third quarter and hockey after the second period, suggesting that it reflects something general about how leads behave in timed team sports. Applying this to the 6,958 games we analyzed, we found that teams leading after the 3rd Quarter won 82.8 percent of the time. However, we note that score margin affects the win rate as seen below.

Using the 2020–2021 season as a reference point, Ganz & Allsop (2024) find that home teams win by an average of 2.13 points when fans are present, compared with 0.44 points when no fans are present—equivalent to approximately 2.2 additional home wins over the course of a regular season. Our analysis is consistent with the existence of a home-court advantage of this magnitude: across our sample, home teams won 55.3% of the time.

Stern (1994) built the formal foundation for in-game win probability in basketball. It modeled the score differential as a Brownian motion process with drift, deriving the win

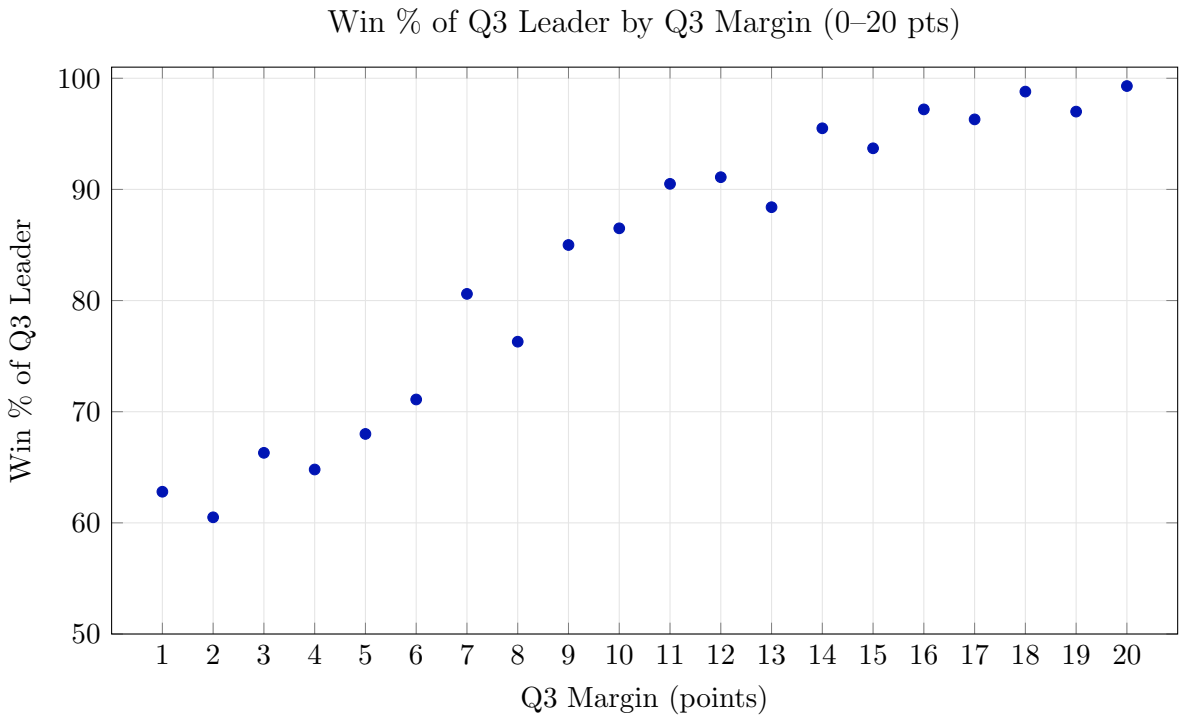

Figure 1: Win % of Q3 Leader by Q3 Margin (0–20 pts).

probability as a function of the current lead and the remaining time. Calibrated on NBA games from the 1991–92 season, it reproduced observed score distributions with considerable accuracy. The paper is limited by conditioning win probability solely on score margin and game clock, thereby assigning the same probability structure to every team and opponent.

Later work found systematic departures from those assumptions. Gabel and Redner (2012), working with play-by-play data from 6,087 NBA games, identified two patterns. They documented a weak restoring force, whereby large leads tend to contract over time while trailing teams score at a slightly higher rate. The restoring force is a natural human response to an unbalanced game; a team with a large lead eases off, while a lagging team plays with greater urgency. They also found anti-persistence from possession changes; a made basket transfers possession, so scoring runs are naturally self-limiting. Taken together, comebacks are more likely than a pure random-walk process would predict. Because ESPN's complete model specification is proprietary, this paper does not attempt to reconstruct or evaluate its internal architecture. Instead, it evaluates the calibration of the model's published outputs and examines whether additional observable information systematically explains residual variation.

Modern win probability models extend well beyond score and clock. Commercial implementations, including ESPN's Basketball Power Index (BPI) and the NBA's official win

probability model, incorporate team strength, home-court advantage, roster availability, rest, and other pre-game factors alongside the evolving game state. ESPN has described BPI publicly as accounting for game-by-game efficiency, strength of schedule, pace, days of rest, game location, and preseason expectations.

Štrumbelj and Vračar (2012), using a possession-level Markov model of NBA basketball, estimate transition probabilities from play-by-play data across two seasons. This generates win-probability estimates through repeated game simulation. They estimate a transition matrix in which each state captures which team holds the ball, how that possession was obtained, and how many points were scored on the prior exchange. Each game was then simulated repeatedly to produce outcome probabilities, with 10,000 matches run per game.

Brill, Yurko, and Wyner (2025) raise an important methodological issue for such an evaluation. They show that statistical win-probability estimators fitted to play-by-play data are subject to substantial uncertainty because observations within the same game share a common outcome. As a result, effective sample sizes are often considerably smaller than their nominal counts suggest. Their simulation study of American football, with a dataset comprising 4,101 games, yielded an effective sample size of 2,291 independent outcomes. The concern extends beyond the particulars of their analysis. Win probability models are often treated as reliable because of their statistical sophistication, yet the uncertainty surrounding their estimates is rarely examined directly. The possibility of systematic error, therefore, remains an open empirical question.

Our findings suggest that this question is still relevant in the basketball setting. While ESPN's model performs well in the aggregate, statistically significant miscalibration persists at eight of the twenty-one probability thresholds examined.

The aggregate behavior of scoring runs and lead changes is reasonably well understood. The conditional prediction problem is less so. Clauset et al. (2015) asked a specific version of this problem: given a lead of a certain size with a certain amount of time left, how likely is it to survive. They used a random-walk model and assessed roughly 40,000 games across four sports, including 11,744 NBA games from 2002 through 2010. They derived a closed-form survival probability that depends only on lead size, time remaining, and the sport's diffusion coefficient. The random-walk model's worst-case error across all lead sizes was under 7 percent. They also showed that three properties of lead dynamics—how long one team holds the lead, when the last lead change happens, and when the largest lead occurs—all concentrate probability near the very beginning and end of games. The NBA persistence parameter they estimated was $p = 0.360$; the team that just scored goes on to score next only about a third of the time. This levies a natural ceiling on how long any single run can extend.

What Clauset et al. (2015), nor Gabel & Redner (2012) address is what separates individual outcomes. They characterize average game behavior and aggregate comeback frequency, but they don't isolate the conditions under which a lead becomes a comeback. When a team reaches an 80 percent win probability, we identify the game-state and organizational factors associated with comeback outcomes and assess whether those risks are reflected in ESPN's published probabilities.

Oliver (2004) identified the four factors that compose a team's offensive performance: effective field goal percentage, turnover rate per possession, offensive rebounding percentage, and free throw rate. Kubatko et al. (2007) standardized how each of these measures is measured by anchoring them to possessions, rather than games or minutes. They derived offensive and defensive efficiency ratings and showed their superiority over raw counting statistics in evaluating team performance. The present study uses this strategy to construct its snapshot variables. At the moment each game crosses the 80 percent win probability threshold, we compute these components for both teams using play-by-play data up to that point, yielding game-state measures that season-level aggregates cannot capture.

Deshpande & Jensen (2016) further motivated our use of this design. They show that traditional plus-minus statistics overweight low-leverage situations, where outcomes are already largely determined. Conditioning our measures on the 80 percent threshold ensures that all observations occur in high-leverage game states.

This paper addresses the conditional question: given that a team has already crossed 80 percent win probability, what determines whether the lead holds, and does ESPN's model reflect that probability accurately? It examines this question across multiple NBA seasons, incorporating variables constructed upon coaching execution, star-player dependency, and team quality into a setting previously treated as largely stochastic. It further evaluates the calibration of ESPN's published outputs directly.

3 Calibration

When a team reaches an 80 percent win probability while being outmatched in talent, it has likely arrived there through an anomalous stretch of play, one that is difficult to sustain. In those moments, the more talented trailing team still retains the capacity to correct it before the game is decided.

Win Rate Above Expected (WAE) is a measure accounting for the historic likelihood of a team to exceed what pre-game betting markets expected of them. Because those markets already account for roster strength, injuries, and home-court advantage, consistent overperformance suggests an edge the market does not fully capture. While no single metric

can isolate coaching from the many factors that shape performance, persistent outperformance across seasons may reflect advantages in preparation, tactical adjustments, player development, and game management. As a result, WAE serves as one component within a broader coaching evaluation framework, where it is combined with complementary indicators to estimate a coach's overall effectiveness percentile relative to peers.

Among teams in the bottom quartile of WAE, leading teams still win 87 to 92 percent of such games, exceeding the stated probability by 6–8 points. In contrast, teams in the top WAE quartile close out only 73 to 82 percent, falling short by as much as 7 points.

Transition offense efficiency measures how effectively a team converts defensive rebounds into scoring opportunities before the defense can set itself, within a brief 14-second window after possession changes. Teams that consistently exploit this window force the defense into decisions under time pressure. The Indiana Pacers and Golden State Warriors, led by the perimeter gravity of Tyrese Haliburton and Stephen Curry, are particularly effective at exploiting retreating defenses within that window. This tends to generate higher-quality possessions while disrupting the leading team's ability to run out the clock.

Leading teams in the bottom quartile of the transition offense efficiency measure win 85.6 percent and 89.2 percent of games when ESPN displays respective probabilities of 80 percent and 81 percent. The stated expectation is exceeded by 6 to 8 points on average. In contrast, leading teams in the top quartile win only 75.0 percent and 80.8 percent of games at those same thresholds, falling short by as much as 5 points.

Pre-game strength variables separate outcomes across otherwise similar game states. Each measure is drawn strictly from the prior season, preserving predictive integrity by constraining inputs to information available before tip-off.

Isolation efficiency measures the points scored per possession on isolation plays. It measures what might be called offensive problem-solving: when set actions fail and the half-court offense stagnates, it becomes a star player's capacity to draw contact fouls or finish through contested space. Thus, it is also an indicator of resilience under pressure and maps directly to comeback outcomes.

When the trailing team's star player exhibits superior prior-season isolation efficiency, the leading side wins roughly 80.3 percent of the time, four points below the sample average. When that same responsibility falls to a bottom-quartile isolation player, the closeout rate rises to 87.5 percent, a 7.2 percentage-point spread.

For example, the Dallas Mavericks repeatedly converted late deficits into wins because of Luka Dončić's ability to score under isolation. The Atlanta Hawks (2021), advancing to the Eastern Conference Finals as a 5-seed behind Trae Young's ability to draw fouls and collapse defenses when the half-court offense had nowhere else to go further exemplifies this

pattern. Both cases show how a proven offensive problem-solver makes a lead structurally fragile in ways a stated win probability cannot capture.

Net rating differential, points scored minus points allowed per 100 possessions averaged across a season, serves as the most comprehensive single measure of team quality.

When the leading team carries an eight-point or greater prior-season advantage, they close out 90.5 percent of such games, closely aligning with the implied probability. When that same advantage belongs to the trailing team, the closeout rate for the leading team drops to 77.2 percent, a divergence of roughly 14 percentage points.

As the trailing team's prior-season quality increases, the leading team's actual closeout rate declines, regardless of the score at the moment of observation. A team ranking higher but trailing late in a game has most likely arrived there through variance rather than a true shift in competitive balance. Over sufficient possessions, player quality asserts itself. Teams like the Denver Nuggets (2023) or the Boston Celtics (2024), who led the league in net rating and went on to win the championship, were rarely placed in a deficit their underlying quality could not overcome over sufficient possessions.

The subgroup patterns above point to a consistent set of forces shaping outcomes: talent distribution, game-state structure, coaching execution, and prior-season team quality. Each influences comeback probability in ways the model does not account for. We next outline how we used these factors to construct our dataset.

4 Variable Construction and Regression Model

Scraping from NBA.com and ESPN.com, we collected over one hundred variables for each of the 6,958 games. Using a logistic regression, we trained our model on the 2019–2020, 2021–2022, and 2023–2024 NBA seasons and used our results to predict the 2020–2021, 2022–2023, and 2024–2025 NBA seasons. The key features our model used were as follows: net rating differential between teams, clutch net rating differential, box plus minus differential between starting players, and the losing team's three-point percentage, all calculated from the end of the previous season. We also calculated the difference in the number of prior-season All-NBA players on each team's roster. The last features our model included were the score margin at the eighty percent threshold, the time elapsed, and margin multiplied by time elapsed. From this regression, we were able to predict the likelihood of a comeback for each game. We split these games into bands, grouping the teams we predicted to come back 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, and 30+ percent of the time. Our results are summarized in Table 1.

Table 1 reports the reliability of the predicted probabilities, and Figure **??** plots them

against observed comeback rates. Predicted and observed rates agree within roughly two percentage points across every band except the sparse 30%+ tail, where the model slightly overstates risk. The model is well calibrated and discriminates leads by fragility rather than declaring any specific lead doomed—its predictions never exceed about 50 percent, appropriate given that comebacks from 80 percent occur only about 15 percent of the time.

Table 1: Reliability of predicted comeback probabilities, out-of-sample (test seasons 2020–21, 2022–23, 2024–25; $n = 2,953$).

Predicted band	**Count**	**Mean predicted**	**Observed rate**	**Difference**
0–5%	129	4.2%	2.3%	-1.9
5–10%	766	7.7%	9.4%	+1.7
10–15%	830	12.4%	10.4%	-2.0
15–20%	548	17.3%	19.0%	+1.7
20–25%	322	22.2%	23.6%	+1.4
25–30%	168	27.0%	27.4%	+0.4
30%+	190	36.2%	31.6%	-4.6

5 Results

The key finding of this paper is that ESPN's in-game win probability model performs well in aggregate but exhibits systematic miscalibration at the extremes of the probability distribution. In the 81–91 percent range, the model understates the leading team's true win probability. Above 95 percent, the direction reverses, and the model overstates confidence.

This miscalibration is concentrated in specific, predictable contexts. The subgroup analysis suggests why. A more granular representation of talent distribution, score-margin structure, coaching quality, and prior-season team strength explains residual variation in comeback. A team holding an 80 percent win probability while trailing in roster quality is structurally different from one holding the same probability with the talent advantage, and those differences remain evident in the observed outcomes.

The regression results reinforce this. An AUC of 0.666 with the strongest predictors being pre-game measures of team quality indicates that a meaningful signal exists before tip-off. In-game snapshot variables at the 80 percent threshold contributed little to model performance. What distinguishes a lead that holds from one that collapses is largely determined by the composition of the teams, rather than the game state at the moment the threshold is crossed.

This is consistent with the broader literature. Gabel and Redner (2012) demonstrated that basketball scoring approximates a memory-less process, with events at any given moment carrying limited forward-looking information. Clauset et al. (2015) showed that lead

survival probabilities conform closely to random-walk predictions in the aggregate. The present paper contributes the conditional layer: given that the 80 percent threshold has been reached, residual variation in outcomes is systematically related to organizational factors the baseline model does not incorporate.

We note that several variables carry measurement noise, and while the directional findings are consistent, they are not definitive. Second, ESPN's model features are not publicly available; miscalibration may reflect deliberate modeling choices, input constraints, or true modeling error—we cannot separate them.

The practical implication is real, though bounded. Two teams assigned the same win probability do not necessarily occupy equivalent game states. Differences in roster quality, coaching, and organizational structure continue to influence comeback risk even after the score and clock are held constant. Published win probabilities summarize much of that uncertainty, but they do not exhaust it.

Taken together, these findings suggest that ESPN's published probabilities are broadly well calibrated across high-probability game states. The departures identified here should not be interpreted as evidence that the model lacks predictive value. Rather, they indicate that meaningful residual variation remains even after sophisticated real-time estimation. That residual variation is systematic, arising under identifiable organizational and game-state conditions.

Since basketball scoring is a process of discrete possessions, win probability evolves through a sequence of bounded revisions rather than continuous change. Teams rarely entered the 80 percent range through a single dramatic revision. The median increase in ESPN's published win probability when a team first crossed the threshold was only 2.8 percentage points, suggesting that high-confidence states generally emerged through successive possessions rather than isolated events.

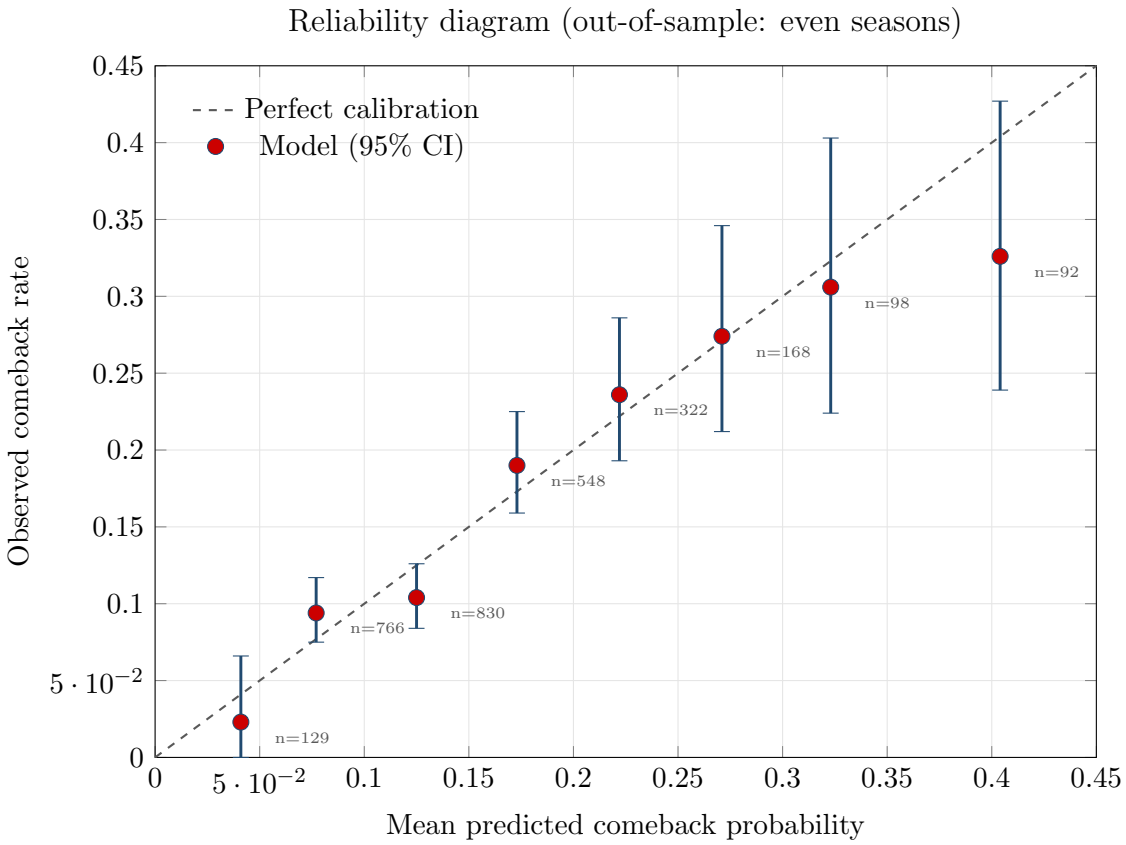

Figure 2: Reliability diagram for out-of-sample comeback predictions (even seasons, $n = 2,953$). Points show observed comeback rate versus mean predicted probability within each band; error bars are Wilson 95% confidence intervals; the dashed line denotes perfect calibration. Observed rates did not differ significantly from predicted in any band; a Hosmer–Lemeshow test indicated no significant miscalibration ($\chi^2_6 = 10.93, p = 0.090$).

The few large revisions in our sample occurred only when games remained effectively even until a decisive late possession. Jamal Murray's go-ahead jumper with four seconds remaining against Los Angeles, for example, increased Denver's win probability from 51.2 to 87.8 percent by resolving nearly all remaining uncertainty at once. Once teams approached the 80 percent threshold, however, individual possessions produced substantially smaller revisions.

We note that our model was trained across 2,829 games in the odd seasons and tested on 2,953 games. Of the training games, a comeback occurred 16.1 percent of the time while a comeback occurred 15.1 percent of the time on the test games. We report a Log Loss score of 0.4028 compared to a baseline score of 0.4251. The standardized logistic regression coefficients, where a positive value indicates a higher comeback probability, were as follows: net rating differential, 0.376; clutch-time net rating differential, 0.305; starter box plus/minus differential, 0.235; margin $\times$ time elapsed, -0.154; score margin, 0.137; All-NBA player differential, -0.135; time elapsed, 0.088; and trailing-team three-point percentage, -0.081.

Trailing-team three-point percentage was positively associated with comeback probability at the margin ($\beta = +14.8$, $p < .001$), but this association was fully attributable to overall team strength: once season net-rating differential was included, the coefficient reversed sign and became non-significant ($\beta = -6.2$, $p = .14$), indicating no independent effect.

Calibration also varied systematically across the course of the game. The largest departures were observed during the early portion of the fourth quarter. Among game states occurring with six to twelve minutes remaining, teams won approximately 3.8 percentage points more often than ESPN's published probabilities suggested ($p < 0.001$), indicating that 80 percent win probabilities during this interval were systematically conservative.

A similar pattern appeared in the final two minutes. Teams assigned an 80 percent win probability won approximately 3.9 percentage points more frequently than expected ($p = 0.01$), suggesting that late-game estimates also understated the security of the leading team. By contrast, observations from the first quarter showed the opposite tendency. Teams reached the 80 percent threshold approximately 2.6 percentage points less often than the published probabilities implied, although this difference was not statistically significant, reflecting the comparatively small number of early-game observations.

Taken together, these results suggest that calibration varies over time rather than uniformly across the game. The largest systematic departures occurred during the transition from competitive to high-confidence game states in the fourth quarter, while the earliest stages of the game exhibited only tentative evidence of overconfidence. A more detailed examination of temporal calibration across the full probability distribution represents a natural

direction for future research.

References

[1] Brill, R. S., Yurko, T., & Wyner, A. J. (2025). Analytics, have some humility: A statistical view of win probability. *Journal of Quantitative Analysis in Sports*.

[2] Clauset, A., Kogan, M., & Redner, S. (2015). Safe leads and lead changes in competitive team sports. *Physical Review E*, 91(6), 062815.

[3] Cooper, H., DeNeve, K. M., & Mosteller, F. (1992). Predicting professional sports game outcomes from intermediate game scores. *Chance*, 5(3-4), 18-22.

[4] Deshpande, S. K., & Jensen, S. T. (2016). Estimating an NBA player's impact on his team's chances of winning. *Journal of Quantitative Analysis in Sports*, 12(2), 51-72.

[5] Gabel, A., & Redner, S. (2012). Random walk picture of basketball scoring. *Journal of Quantitative Analysis in Sports*, 8(1).

[6] Ganz, S. C., & Allsop, K. (2024). A mere fan effect on home-court advantage. *Journal of Sports Economics*, 25(1), 30-53.

[7] Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports*, 3(3).

[8] Oliver, D. (2004). *Basketball on Paper: Rules and Tools for Performance Analysis*. Potomac Books.

[9] Stern, H. S. (1994). A Brownian motion model for the progress of sports scores. *Journal of the American Statistical Association*, 89(427), 1128-1134.

[10] Štrumbelj, E., & Vračar, P. (2012). Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. *International Journal of Forecasting*, 28(2), 532-542.

[11] Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158), 209-212.

Table 2: Strongest individual predictors of a comeback (top 12 by univariate AUC).

Feature	Description	AUC
starter_bpm_diff	Difference in starting-lineup Box Plus/Minus (favorite minus opponent)	0.677
net_rtg_diff	Season net-rating difference (favorite minus opponent)	0.663
clutch_net_rtg_diff	Clutch-time net-rating difference	0.639
wp_net_rtg	Favorite season net rating	0.616
opp_net_rtg	Opponent season net rating	0.613
opp_iso_x_net_rtg	Opponent iso share × net rating	0.611
opp_clutch_net_rtg	Opponent clutch-time net rating	0.603
wp_net_rtg_last_month	Favorite net rating, last month	0.602
allnba_total_diff	Net count of All-NBA players (favorite minus opponent)	0.599
wp_clutch_net_rtg	Favorite clutch-time net rating	0.598
opp_rts_pct	Opponent recent (rolling) true-shooting %	0.596
pts_needed_x_rts	Points-needed × recent TS% interaction	0.594

Pre-game context (51 features)

Feature	Description	AUC
Team strength (season)		
starter_bpm_diff	Difference in starting-lineup Box Plus/Minus (favorite minus opponent)	0.677
net_rtg_diff	Season net-rating difference (favorite minus opponent)	0.663
clutch_net_rtg_diff	Clutch-time net-rating difference	0.639
wp_net_rtg	Favorite season net rating	0.616
opp_net_rtg	Opponent season net rating	0.613
opp_clutch_net_rtg	Opponent clutch-time net rating	0.603
allnba_total_diff	Net count of All-NBA players (favorite minus opponent)	0.599
wp_clutch_net_rtg	Favorite clutch-time net rating	0.598
opp_starter_bpm	Opponent starters average BPM	0.591
wp_starter_bpm	Favorite starters average BPM	0.587
Recent form (trailing month)		
wp_net_rtg_last_month	Favorite net rating, last month	0.602
opp_rts_pct	Opponent recent (rolling) true-shooting %	0.596

Feature	Description	AUC
wp_rts_pct	Favorite recent (rolling) true-shooting %	0.593
net_rtg_monthly_diff	Net-rating difference over the trailing month	0.566
opp_net_rtg_last_month	Opponent net rating, last month	0.566
opp_close_game_wpct_last_month	Opponent close-game win % over last month	0.508
Opponent style / personnel		
opp_fg3_pct	Opponent season three-point %	0.573
opp_dreb_percentile	Opponent defensive-rebound percentile	0.554
wp_tov_percentile	Favorite turnover-rate percentile	0.546
wp_dreb_percentile	Favorite defensive-rebound percentile	0.543
opp_athleticism_pct	Opponent athleticism percentile	0.524
opp_stl_percentile	Opponent steal-rate percentile	0.520
opp_contested_shots_pg	Opponent contested shots per game	0.518
opp_stl_top75	Indicator: opponent steal rate in top quartile	0.516
opp_deflections_pg	Opponent deflections per game	0.516
opp_cut_pct	Opponent cut possession share	0.515
opp_iso_pct	Opponent share of possessions in isolation	0.513
opp_loose_balls_pg	Opponent loose balls recovered per game	0.508
opp_pnr_pct	Opponent pick-and-roll possession share	0.507
opp_spotup_pct	Opponent spot-up possession share	0.507
opp_charges_pg	Opponent charges drawn per game	0.507
wp_contested_shots_pg	Favorite contested shots per game	0.504
opp_3pa_rate	Opponent three-point attempt rate	0.502
Set-play / situational efficiency		
dreb_full_ortg	Defensive-rebound full offensive rating	0.564
slob_cqm	Sideline-out-of-bounds clean-quality-make rate	0.556
dreb_qh_ortg	Defensive-rebound (quarter-half) offensive rating	0.556
slob_organic_ppp	SLOB organic points per possession	0.543
dreb_eo_ortg	Defensive-rebound (end-of) offensive rating	0.541
ato_qs_ppp	After-timeout points per possession, quick-strike	0.538
ato_overall_ppp	After-timeout points per possession	0.536
slob_lategame_ppp	SLOB late-game points per possession	0.532
ato_designed_rate	Rate of designed after-timeout plays	0.517
dreb_qh_rate	Defensive-rebound quarter-half rate	0.505
Rest, comeback context & interactions		
opp_iso_x_net_rtg	Opponent iso share × net rating	0.611
pts_needed_x_rts	Points-needed × recent TS% interaction	0.594
opp_iso_x_monthly_net_rtg	Opponent iso share × monthly net rating	0.564
b2b_x_bpm_diff	Back-to-back × BPM-diff interaction	0.527
opp_spotup_x_fg3	Opponent spot-up share × 3P%	0.515

Feature	Description	AUC
pts_per_min_needed	Points per minute the opponent must score to catch up	0.513
wp_back_to_back	Indicator: favorite on a back-to-back	0.506

In-game state at the snapshot (42 features)

Feature	Description	AUC
Game state		
wp_pct	Live win-probability of the favorite at the snapshot	0.541
lead_changes	Cumulative lead changes so far	0.520
time_elapsed_pct	Fraction of game elapsed at the snapshot	0.512
score_margin	Favorite lead (points) at the snapshot	0.507
Foul & bonus state (point-in-time)		
snap_foul_diff	Foul differential so far	0.529
snap_def_foul_rate	Defensive foul rate so far	0.523
snap_opp_fouls	Opponent fouls so far	0.513
wp_fouls_this_qtr	Favorite team fouls in current quarter	0.507
opp_fouls_this_qtr	Opponent team fouls in current quarter	0.506
wp_starters_foul_trouble	Count of favorite starters in foul trouble	0.506
snap_total_fouls	Total fouls so far	0.506
opp_in_bonus	Indicator: opponent in the bonus	0.505
snap_wp_fouls	Favorite fouls so far	0.503
wp_in_bonus	Indicator: favorite in the bonus	0.502
Timeout state		
timeout_diff	Timeout differential (favorite minus opponent)	0.515
opp_timeouts_used	Opponent timeouts used so far	0.514
opp_timeouts_left	Opponent timeouts remaining	0.514
wp_timeouts_used	Favorite timeouts used so far	0.502
wp_timeouts_left	Favorite timeouts remaining	0.502
Box-score accumulation so far (snap)		
snap_wp_blocks	Favorite blocks so far	0.517
snap_wp_fta	Favorite free-throw attempts so far	0.516
snap_wp_oreb	Favorite offensive rebounds so far	0.513
snap_opp_fga	Opponent field-goal attempts so far	0.513
snap_wp_dreb	Favorite defensive rebounds so far	0.512

Feature	Description	AUC
snap_wp_fga	Favorite field-goal attempts so far (vs. league pace)	0.508
snap_opp_oreb	Opponent offensive rebounds so far	0.508
snap_opp_dreb	Opponent defensive rebounds so far	0.507
snap_wp_tov	Favorite turnovers so far	0.502
snap_wp_steals	Favorite steals so far	0.501
snap_opp_fta	Opponent free-throw attempts so far	0.501
snap_opp_tov	Opponent turnovers so far	0.501
Efficiency rates so far (snap)		
snap_opp_ft_rate	Opponent free-throw rate so far	0.516
snap_wp_ft_rate	Favorite free-throw rate so far	0.514
snap_wp_oreb_rate	Favorite offensive-rebound rate so far	0.511
snap_opp_tov_rate	Opponent turnover rate so far	0.510
snap_wp_ft_pct	Favorite free-throw % so far	0.507
snap_wp_tov_rate	Favorite turnover rate so far	0.507
snap_opp_ts_pct	Opponent true-shooting % so far	0.506
snap_opp_ft_pct	Opponent free-throw % so far	0.505
snap_wp_ts_pct	Favorite true-shooting % so far	0.505
snap_def_reb_rate	Defensive rebound rate so far	0.502