NBA Shot Distribution & Winning: A Bayesian Model

Ethan Babel

Johns Hopkins Whiting School of Engineering

Why Shot Distribution Matters

- The modern NBA increasingly values floor spacing, rim shots, and 3-pointers over long mid-range jumpers.
- Conventional wisdom (and analytics narratives) say:

  “More 3s + rim attempts → more wins.”
- Does the team’s shot-distance distribution truly predict wins?
- Has that relationship changed over time (2001 vs 2025)?

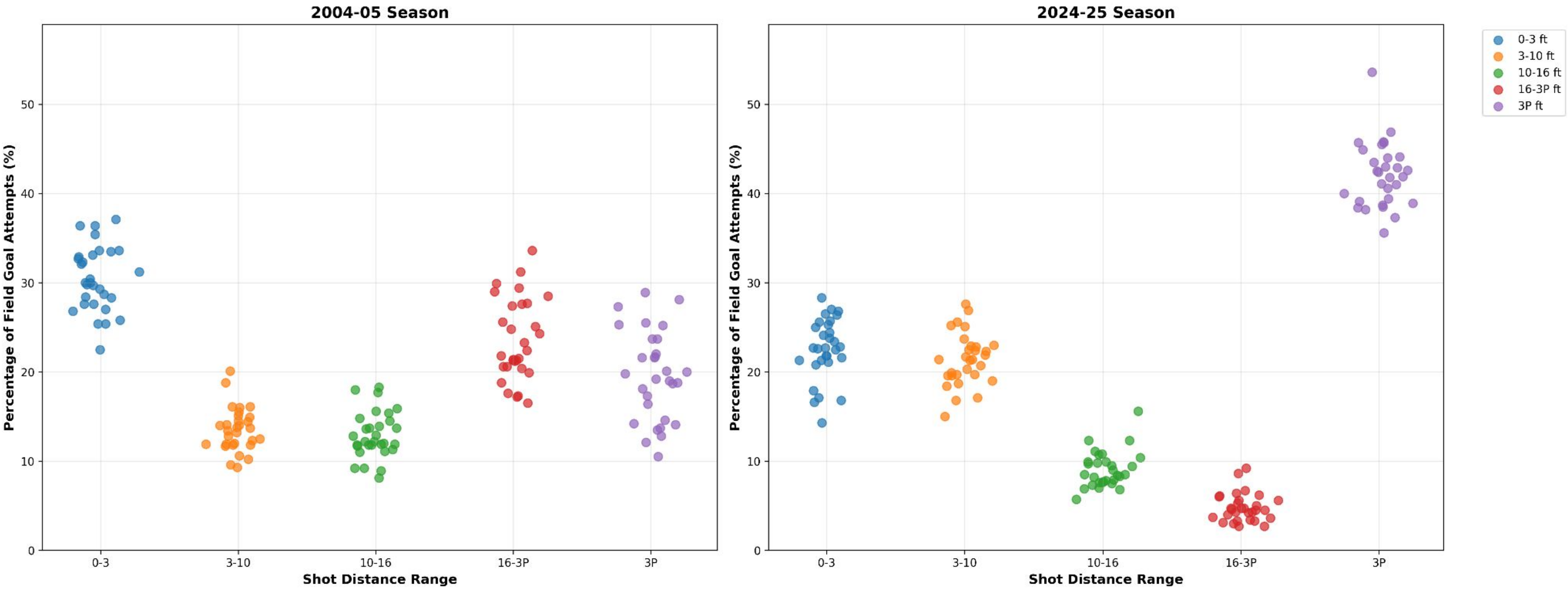
Why a Bayesian Approach

- Frequentist regressions give point estimates but little about uncertainty or predictive distributions.
- Bayesian modeling allows full posterior over shot-zone effects, credible intervals, uncertainty-aware predictions.
- Enables formal comparison across seasons (via posterior distributions) — not just point-estimates.
- New contribution: non-conjugate Bayesian logistic regression + shot-distribution data.

Data & Preprocessing

- Shot distribution features include the percentage of field-goal attempts (FGA) and percentage of FGA allowed defensively from 5 distance bands:
  - 0–3 ft
  - 3–10 ft
  - 10–16 ft
  - 16 ft–3P
  - 3P+
- Team outcome target → Wins (out of 82 games)
- Seasons analyzed: 2000–01 and 2024–25

NBA Shot Distribution Comparison: 2004-05 vs 2024-25



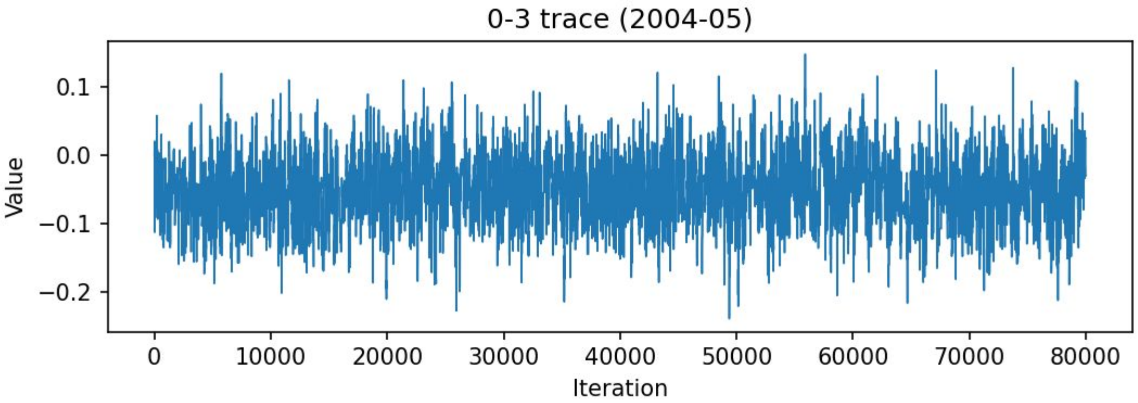
Bayesian Model Specification

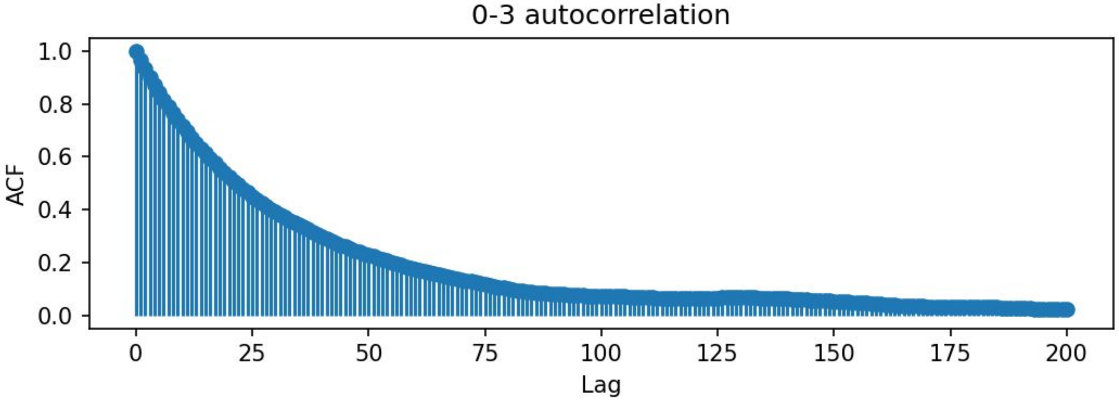
$$W_{is} \mid p_{is} \sim \text{Binomial}(n_{is}, p_{is})$$

$$p_i = \sigma(\alpha + \Sigma_j \beta_j x_{ij})$$

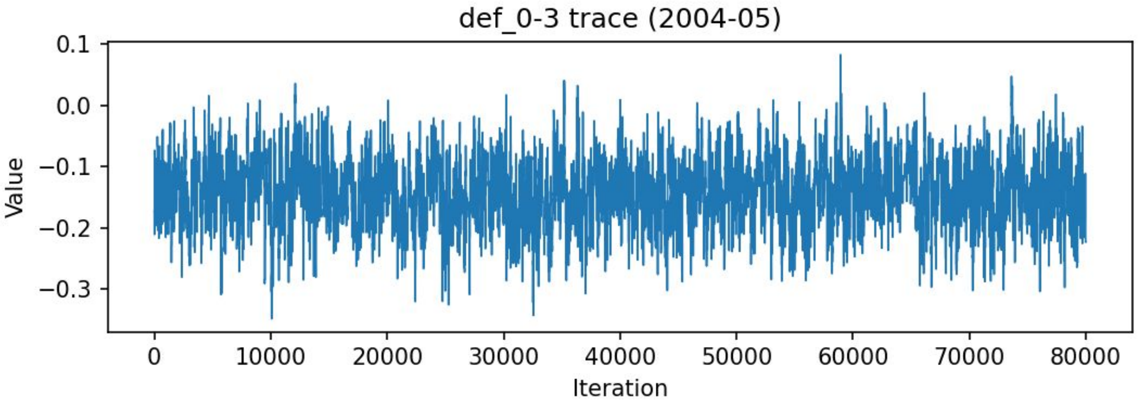
- Weak normal priors
- To avoid collinearity → drop 3P band as baseline (for both offense and defense)
- Logistic link + Normal priors → non-conjugate posterior → no closed-form solution
- Random-walk Metropolis sampler to draw posterior samples for alpha, beta.
- Post-processing yields posterior means, SDs, credible intervals, posterior predictive distributions, etc.

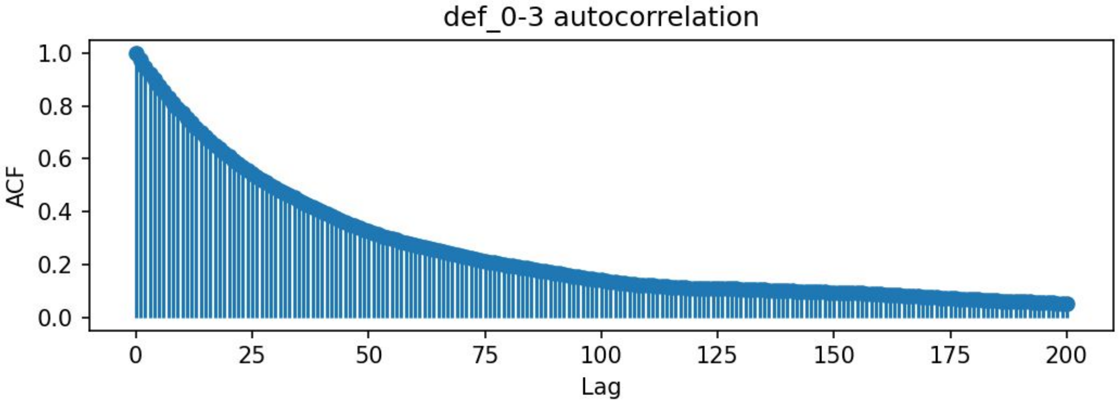
Diagnostics and Model Fit





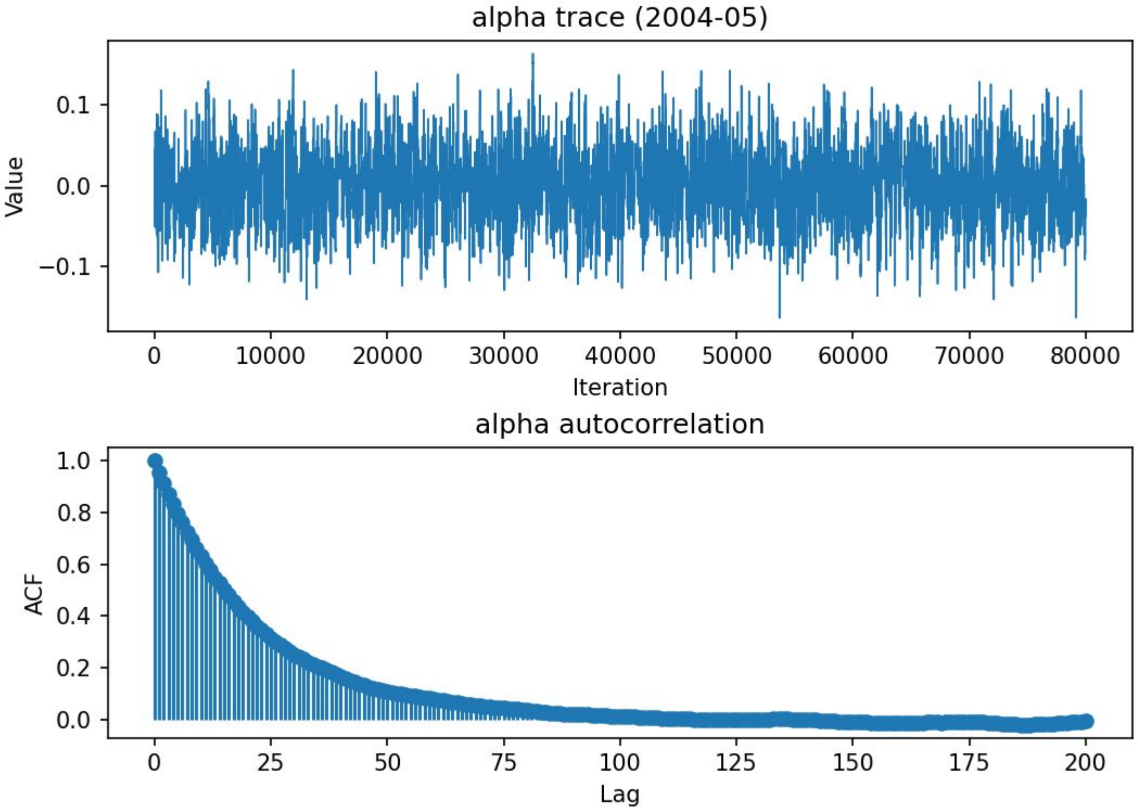
Diagnostics and Model Fit



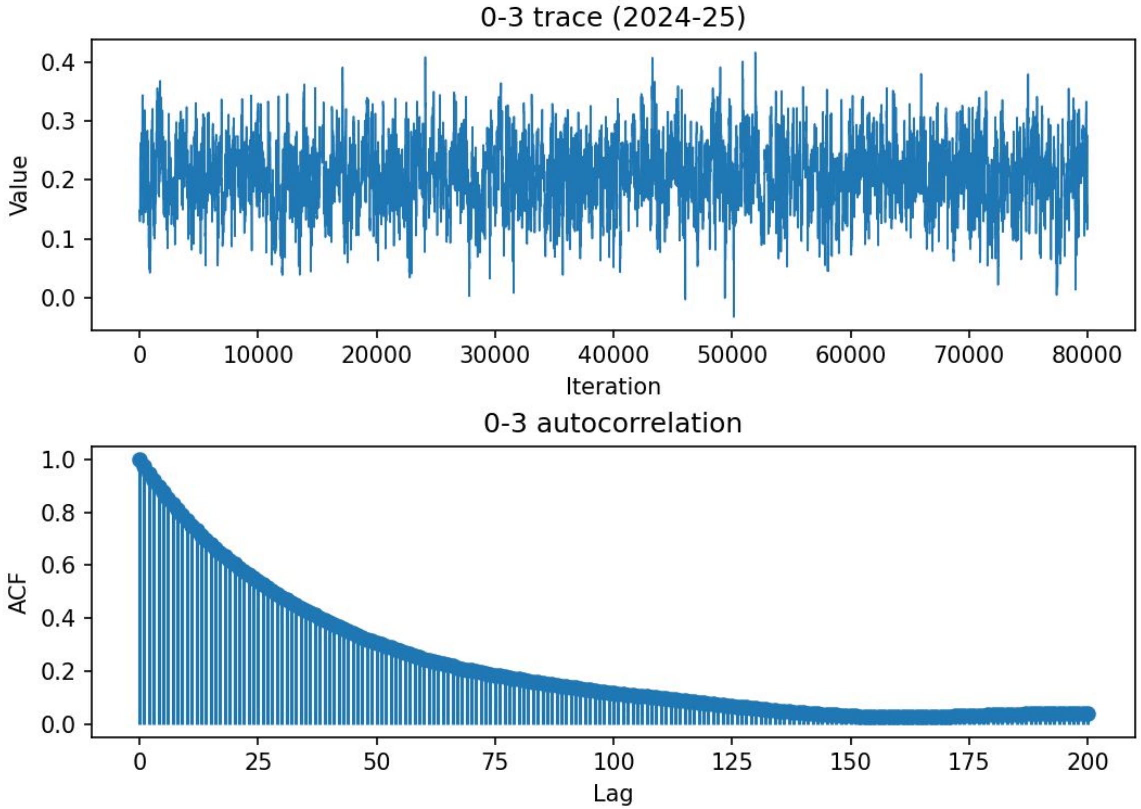


JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Diagnostics and Model Fit

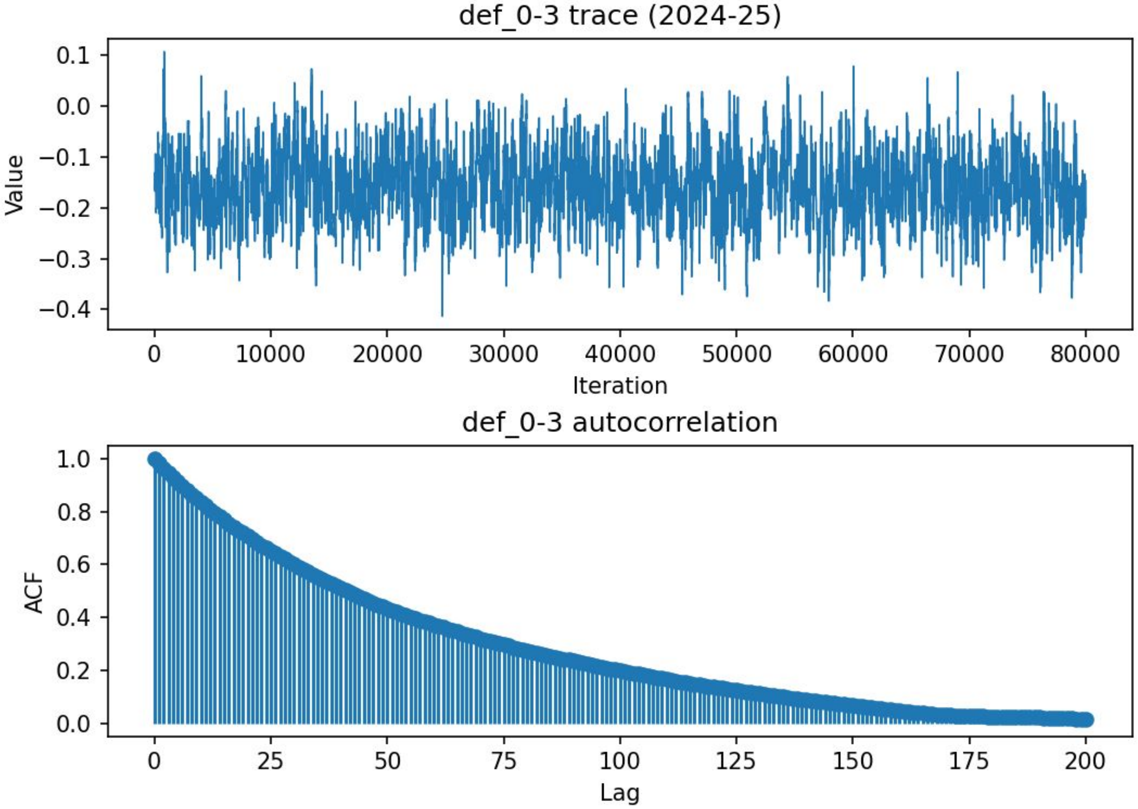


Diagnostics and Model Fit

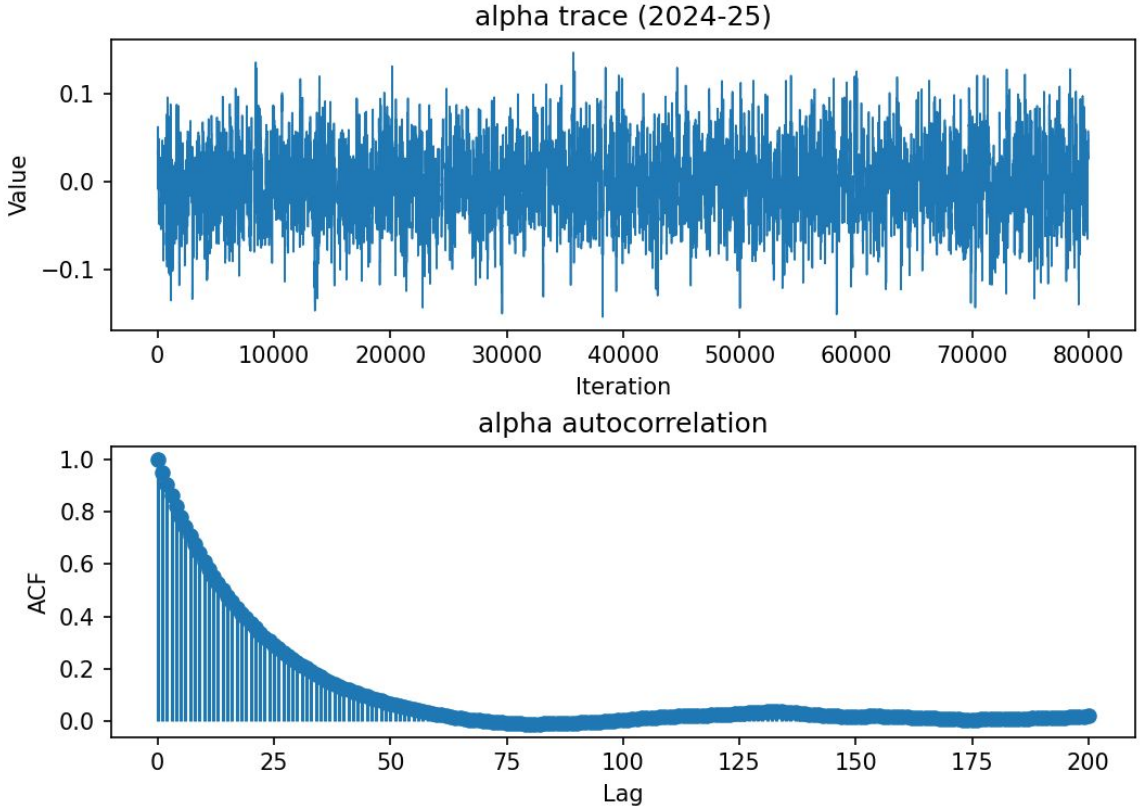


JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

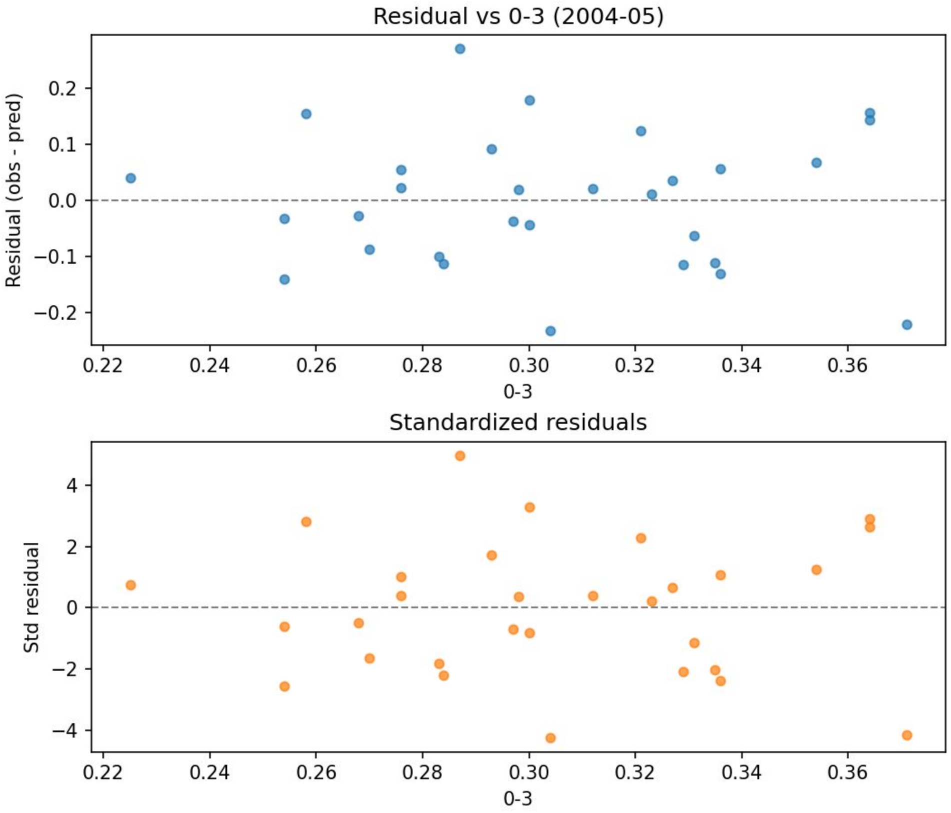
Diagnostics and Model Fit



Diagnostics and Model Fit

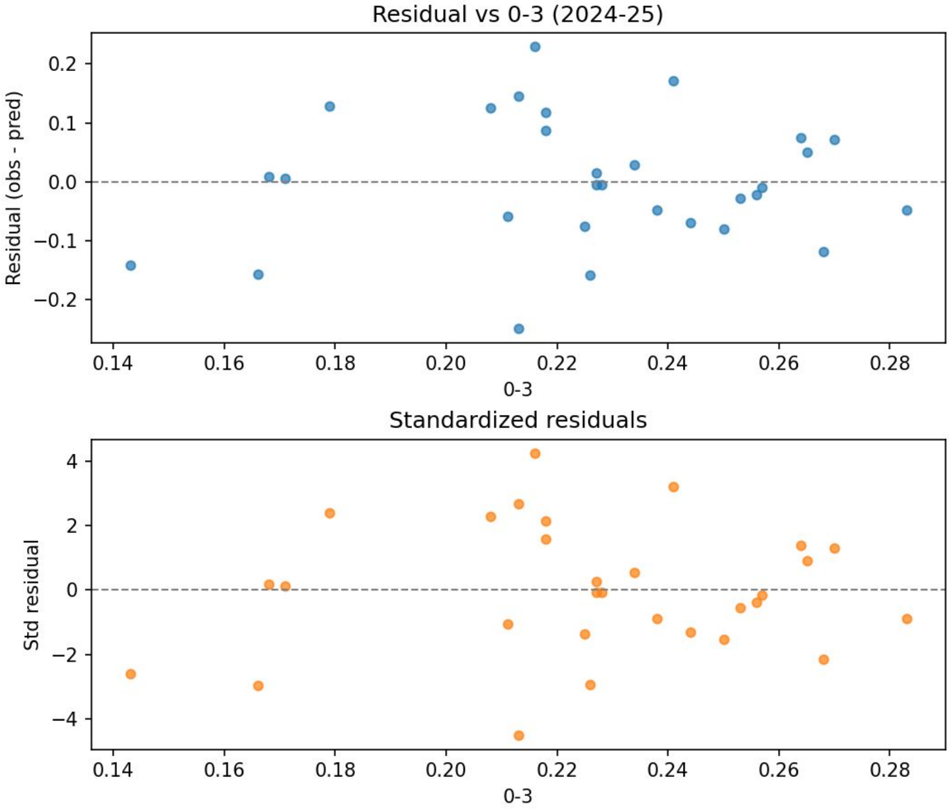


Diagnostics and Model Fit



JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

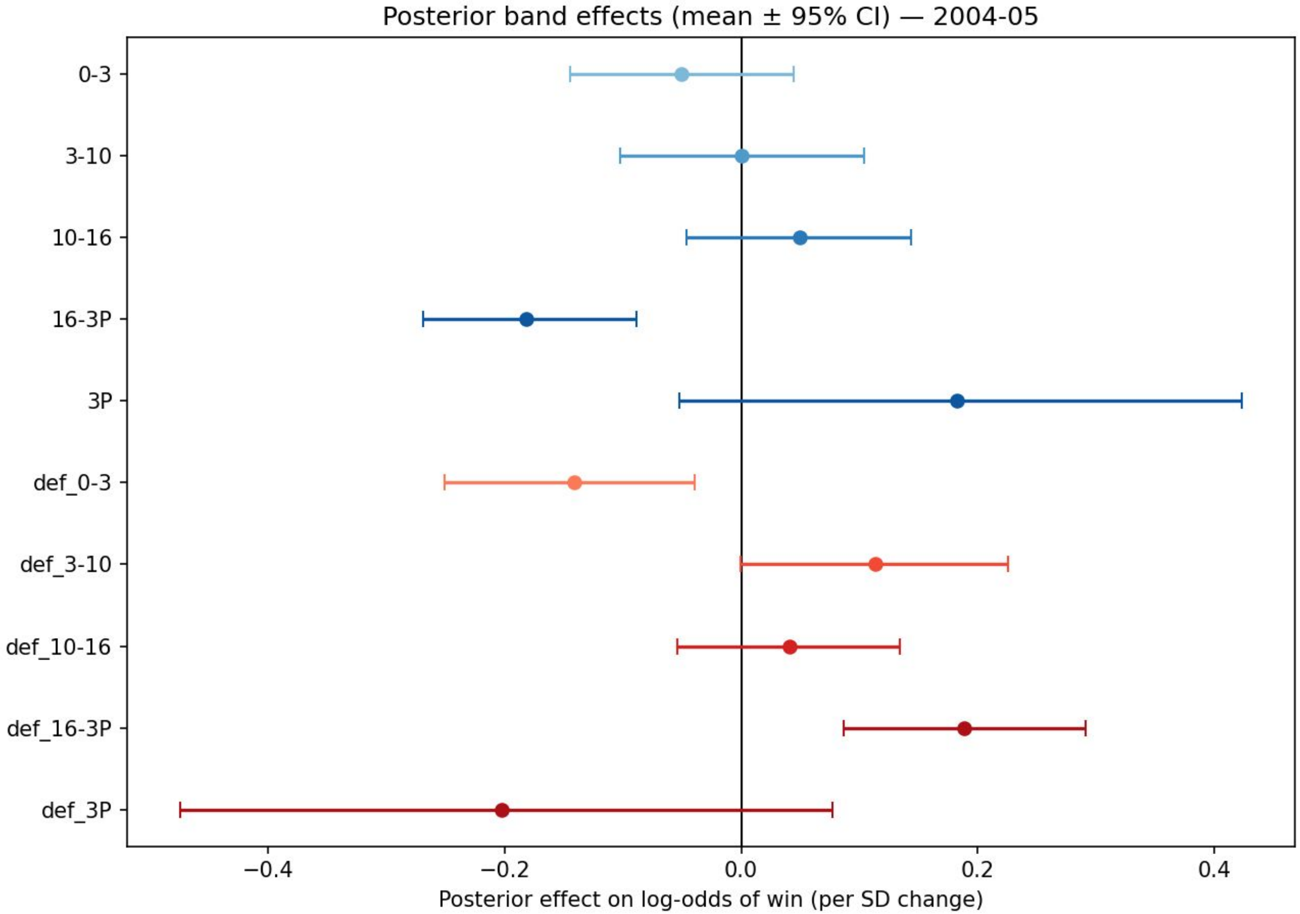
Diagnostics and Model Fit



Effective Sample Size

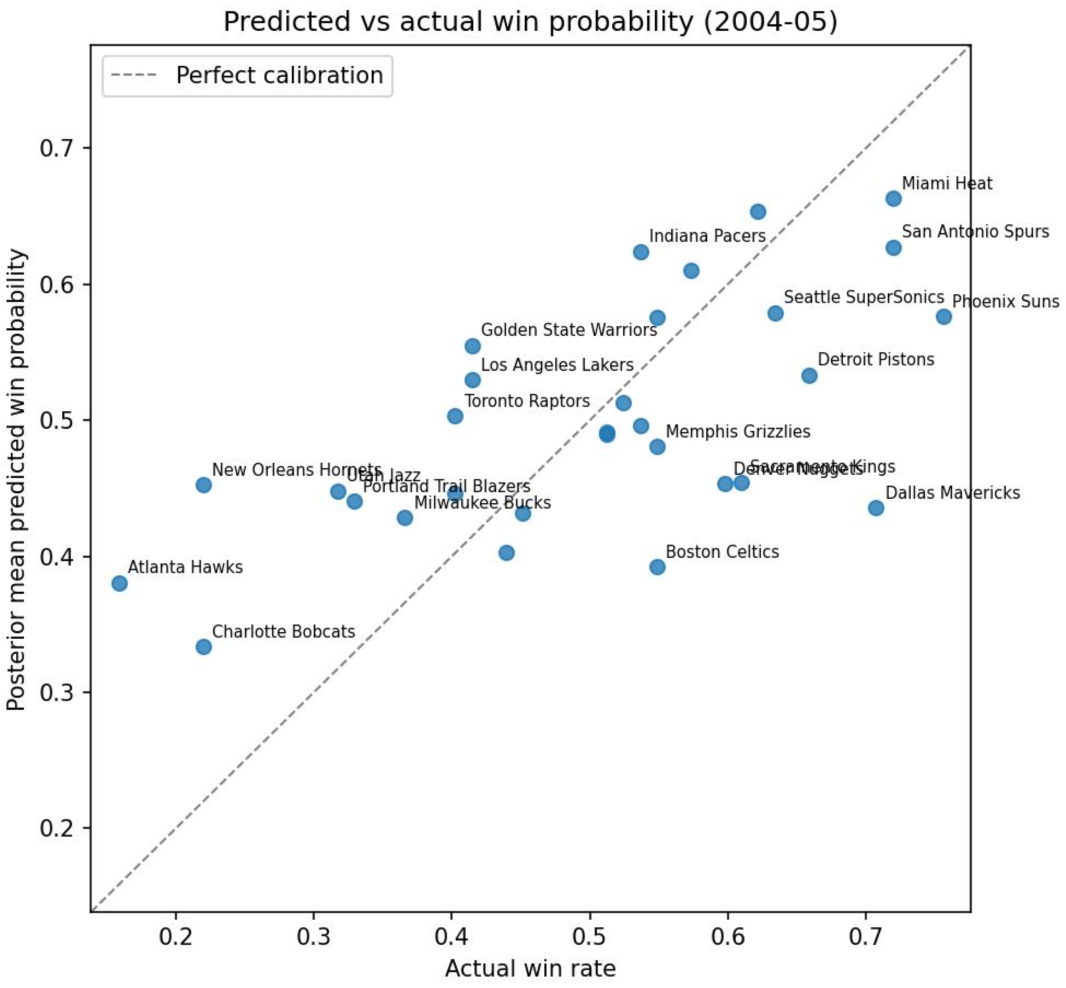
2004-05		2024-25	
Param	ESS	Param	ESS
alpha	1,798.7	alpha	2,113.3
0-3	1,094.2	0-3	918.8
3-10	1,089.0	3-10	1,166.2
10-16	980.9	10-16	1,365.5
16-3P	1,166.5	16-3P	1,119.2
def 0-3	809.8	def 0-3	704.9
def 3-10	902.2	def 3-10	797.9
def 10-16	1,160.0	def 10-16	1,462.9
def 16-3P	946.4	def 16-3P	1,264.0

Results



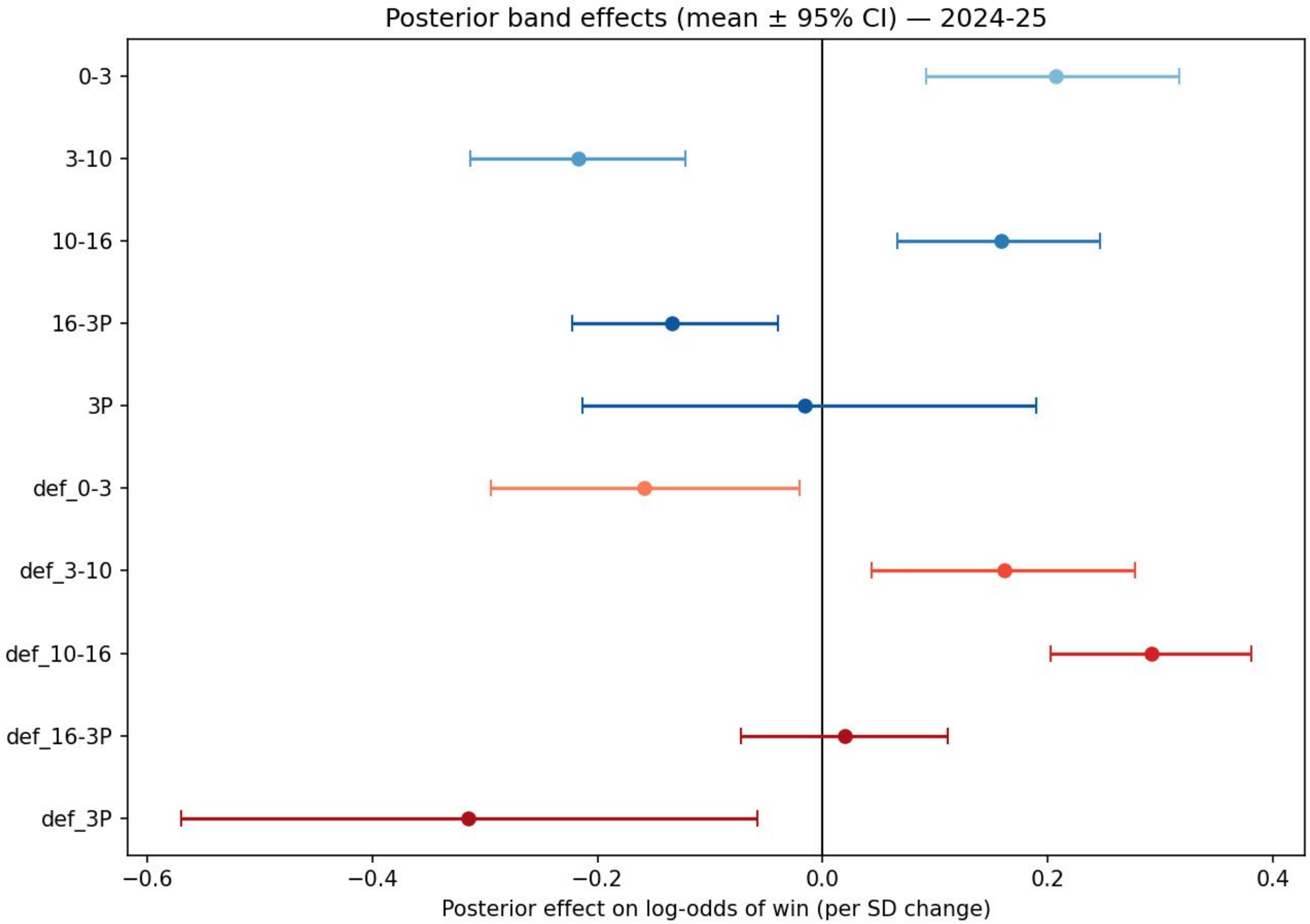
Posterior band effects (mean ± 95% CI) — 2004-05
0-3
3-10
10-16
16-3P
3P
def_0-3
def_3-10
def_10-16
def_16-3P
def_3P
−0.4
−0.2
0.0
0.2
0.4
Posterior effect on log-odds of win (per SD change)

Results



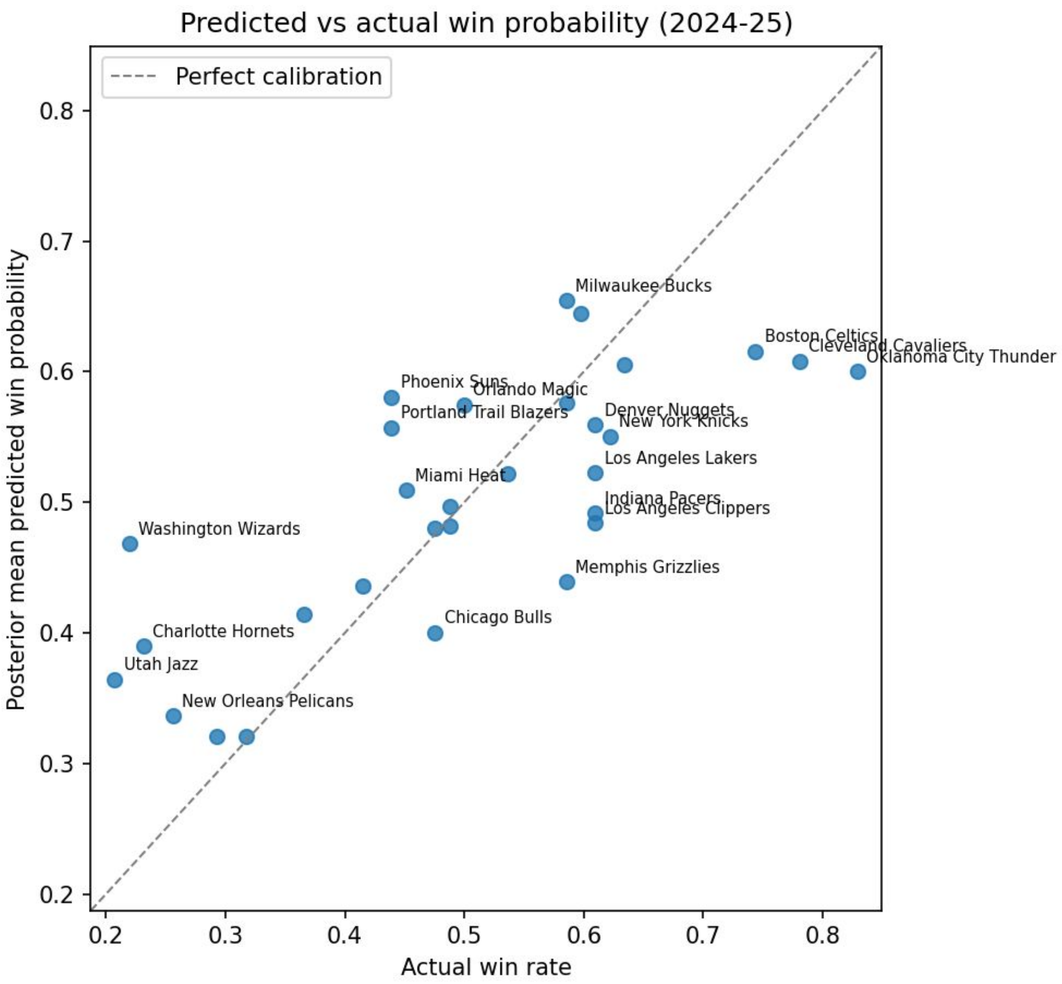
Predicted vs actual win probability (2004-05)
Perfect calibration
Posterior mean predicted win probability
Actual win rate
0.2
0.3
0.4
0.5
0.6
0.7
Miami Heat
San Antonio Spurs
Indiana Pacers
Seattle SuperSonics
Phoenix Suns
Golden State Warriors
Los Angeles Lakers
Detroit Pistons
Toronto Raptors
Memphis Grizzlies
New Orleans Hornets
Utah Jazz
Portland Trail Blazers
Milwaukee Bucks
Denver Nuggets
Sacramento Kings
Dallas Mavericks
Boston Celtics
Atlanta Hawks
Charlotte Bobcats

Results



JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Results


Predicted vs actual win probability (2024-25)
Perfect calibration
Posterior mean predicted win probability
Actual win rate
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Milwaukee Bucks
Boston Celtics
Cleveland Cavaliers
Oklahoma City Thunder
Phoenix Suns
Orlando Magic
Portland Trail Blazers
Denver Nuggets
New York Knicks
Los Angeles Lakers
Miami Heat
Indiana Pacers
Los Angeles Clippers
Washington Wizards
Memphis Grizzlies
Chicago Bulls
Charlotte Hornets
Utah Jazz
New Orleans Pelicans

Predictive Performance

- Using 32-fold (leave-one-out) cross validation on each season independently
- Interpretation → Model substantially improves over baseline in 2024–25 (≈26% skill), but only modestly in 2000–01 (≈4%).
- Suggests shot-distribution was more predictive of wins in recent era vs 25 years ago.
- Higher Baseline for 2004-05 indicates the league had more parody compared to now.

Season	Brier	Brier Baseline	Skill (BSS)
2004-05	0.0237	0.0248	0.0425
2024-25	0.0203	0.0274	0.2611

Limitations

- Only shot-distance distribution, not shot efficiency (FG%, eFG%). Distribution doesn’t capture whether shots are made or missed.
- Purely linear combination of shot distribution features → cannot capture “sweet spot” while penalizing for both too high and too low percentages.
- No control for pace, pace-adjusted possessions, or opponent strength / schedule.
- Small sample: only two seasons, one team-season per team → limited generalizability.
- Bayesian model ignores possible defensive / rebound / turnover influences.

Discussion/Further Improvements

- Add shot efficiency metrics (FG%, eFG%, true-shooting) per zone.
- Include pace, possessions per game, rebound rate, turnover rate, free-throw rate, defensive metrics.
- Expand to multiple seasons (2000–2025) and use a hierarchical Bayesian model to pool across seasons while allowing season-specific effects.
- Extend to game by game modeling to increase sample size while holding team composition and league-wide offensive/defensive tendencies (era-based) constant.
- Test posterior predictive power out-of-sample (hold out an entire game/season).
- Incorporate team-level random effects to capture unobserved heterogeneity (coaching, injuries, trades, etc.).