NBA Shot Distribution & Winning: A Bayesian Model

Ethan Babel
Johns Hopkins University
ebabel1@jh.edu

Abstract

Understanding how shot selection contributes to winning has become a central theme in modern basketball analytics, especially as the NBA has shifted toward a more prominent three point focus offensively over the past two decades. In this project, I study how the distribution of field-goal attempts a team both takes on offense and concedes on defense across standard distance bands affects the team's probability of winning games. I develop a Bayesian logistic regression model that uses each team's offensive and defensive shot-distance distributions as predictors. Posterior inference is done via Metropolis–Hastings sampling, yielding full distributions over the effects of each shot band and enabling comparisons across eras.

I apply the model to two NBA seasons separated by twenty years (2004-05 and 2024-25) to evaluate how the value of shot distribution has evolved over time. The model achieves substantially higher predictive skill in the modern season (Brier Skill Score: 0.26) compared to the earlier era (0.04), suggesting that shot selection has become a far more impactful determinant of success in today's spacing and efficiency driven league. Key findings indicate that in 2004-05, teams that take a greater share of their shots from three-point range and the rim while limiting the same areas defensively tend to win significantly more games. By 2024-25, these effects lessened or even reversed, suggesting that teams have adapted to prioritize three-point shooting more universally, and indicating that some teams today may benefit from diversifying their shot profiles back toward mid-range attempts.

1 Introduction

The past two decades have seen a dramatic transformation in how NBA teams structure their offenses. Long mid-range jumpers have steadily declined, while three-point attempts and shots at the rim have risen sharply. These changes are well documented in the basketball analytics literature and are often attributed to the growing influence of efficiency-based decision making, where the expected value of each shot type guides offensive schemes. Although this historical trend is clear, this project delves into determining to what extent a team's shot distribution can actually predict its success, as well as comparing the importance of shot distribution across eras.

I address these questions by developing a Bayesian model that links team-level win probabilities to offensive and defensive shot-distance distributions. Distance-based shot charts (such as the five standard NBA zones 0–3 ft, 3–10 ft, 10–16 ft, 16 ft – 3 point line, and three-pointers) provide a compact representation of a team's spatial tendencies. Teams differ not only in where they choose to shoot but also in the types of shots they allow opponents to take. A comprehensive analysis therefore requires modeling both dimensions.

While there are undoubtedly far more factors influencing wins (e.g., shooting efficiency/ability, turnovers, rebounding, pace, etc.), this study isolates the role of shot distribution to understand its standalone predictive power. I use a binomial logistic regression model in which the log-odds of winning are determined by a linear combination of the proportion of shots taken/allowed by a team each offensive and defensive shot band. Because the posterior distribution of the model parameters is not available in closed form, I employ a Metropolis–Hastings sampler to obtain posterior draws, enabling full uncertainty quantification and posterior predictive evaluation. This Bayesian framework offers advantages over classical approaches, mainly that it provides interpretable posterior distributions for each shot band's effect, naturally incorporates parameter uncertainty into predictive estimates, and allows comparison of structural effects across seasons.

The analysis focuses on two seasons separated by twenty years: 2004–05 and 2024–25. This time span captures both the pre–three-point-revolution era and the fully modern, spacing-oriented NBA. Comparing results across these seasons allows us to quantify whether shot distribution has become more (or less) predictive of success over time.

Overall, this study provides a Bayesian perspective on the evolution of shot selection in professional basketball, revealing how the value of different shot types has shifted and how both offensive and defensive spatial tendencies jointly shape team performance.

2 Related Works

A substantial body of research examines how the spatial distribution of shots, rather than merely shooting efficiency, relates to team success in modern basketball. This literature can be divided into three major themes: (1) long term evolution of shot patterns, and (2) the rise of the three-point shot, the decline of the mid-range, and the effects of these trends on winning.

2.1 Evolution of NBA Shot Distribution

Wang and Zheng [Wang and Zheng, 2022] document league-wide trends in shooting patterns over the past decade, showing sharp declines in long mid-range attempts and concurrent increases in three-point volume. They also note that offensive and defensive field-goal percentages remain the strongest predictors of winning, linking shot location to macro-level team outcomes.

Zajac et al. [Zajac et al., 2023] expand this view by studying forty consecutive NBA seasons. They find rising efficiency and volume from three, as well as stability in two-point accuracy, concluding that effective field-goal percentage (eFG%), a metric that adjusts traditional field goal percentage by fairly weighting 3-point attempts overwhelmingly determines win probability.

$$eFG\% = \frac{FGM + 0.5 \times 3PM}{FGA}$$

These results reinforce the idea that spatial shot tendencies shape league-wide offensive performance.

2.2 Three-Point Revolution and Decline of Mid-Range Shooting

Kilcoyne [Kilcoyne, 2020] analyzes the decline of mid-range attempts in the context of basketball analytics. He finds that mid-range volume is negatively associated with team wins once efficiency is accounted for, supporting the common interpretation that these shots are dominated by higher-value alternatives (rim attempts or three-pointers).

In contrast, Winn [Winn, 2023] shows that three-point attempt rate is not always a statistically significant predictor of team success once other factors are controlled for. This highlights that volume alone is insufficient without context.

2.3 Positioning of This Work

The literature demonstrates that NBA shot distribution has changed dramatically over time, and that winning correlates more strongly with efficiency than raw volume.

This project adds to the existing literature by modeling team-level win probabilities directly as a function of *both* offensive and defensive shot-distance distributions using a Bayesian logistic framework. The Bayesian approach improves on the works above by providing full posterior uncertainty quantification for each shot band's effect, enabling richer interpretation and comparison across eras.

3 Model

3.1 Data and notation

We analyze two NBA regular seasons, $s \in \{2004\text{–}05,\ 2024\text{–}25\}$. For each season s we observe $n_s = 30$ teams indexed by $i = 1, \dots, n_s$. Let

$$y_{is} \in \{0, 1, \dots, N\}$$

denote the number of regular-season wins for team i in season s, where $N = 82$ is the number of regular-season games.

For each team we summarize its offensive and defensive shot distribution using the proportion of field-goal attempts (FGA) taken or allowed from four distance bands:

$$\text{Offense:} \quad x^{\text{off}}_{is,0\text{–}3},\ x^{\text{off}}_{is,3\text{–}10},\ x^{\text{off}}_{is,10\text{–}16},\ x^{\text{off}}_{is,16\text{–}3P},$$

$$\text{Defense:} \quad x^{\text{def}}_{is,0\text{–}3},\ x^{\text{def}}_{is,3\text{–}10},\ x^{\text{def}}_{is,10\text{–}16},\ x^{\text{def}}_{is,16\text{–}3P}.$$

These correspond respectively to 0–3 ft, 3–10 ft, 10–16 ft, and 16 ft to the three-point line. The offensive and defensive three-point shares,

$$x^{\text{off}}_{is,3P}, \quad x^{\text{def}}_{is,3P},$$

are omitted from the regression design matrix to avoid perfect collinearity, but are later reconstructed from the results of the included features for diagnostic and interpretive calculations.

For season s, we collect the eight included shot-share features into a vector

$$\mathbf{x}_{is} = \left(x^{\text{off}}_{is,0\text{–}3}, x^{\text{off}}_{is,3\text{–}10}, x^{\text{off}}_{is,10\text{–}16}, x^{\text{off}}_{is,16\text{–}3P}, x^{\text{def}}_{is,0\text{–}3}, x^{\text{def}}_{is,3\text{–}10}, x^{\text{def}}_{is,10\text{–}16}, x^{\text{def}}_{is,16\text{–}3P}\right)^{\top} \in \mathbb{R}^8,$$

so that the (i, s) design row is $\mathbf{x}_{is}$ and the response is y_{is}.

3.2 Standardization of predictors

Before fitting, we standardize each feature dimension within a season. Let $\mu_{s,j}$ and $\sigma_{s,j}$ denote the empirical mean and standard deviation of the j-th feature across teams in season s:

$$\mu_{s,j} = \frac{1}{n_s} \sum_{i=1}^{n_s} x_{is,j}, \qquad \sigma_{s,j} = \sqrt{\frac{1}{n_s} \sum_{i=1}^{n_s} \big(x_{is,j} - \mu_{s,j}\big)^2}.$$

We then define standardized covariates

$$z_{is,j} = \frac{x_{is,j} - \mu_{s,j}}{\sigma_{s,j}}, \qquad j = 1, \dots, 8.$$

In matrix notation we write

$$\mathbf{z}_{is} = \big(z_{is,1}, \dots, z_{is,8}\big)^{\top},$$

and stack these into a season-specific design matrix $Z_s \in \mathbb{R}^{n_s \times 8}$.

3.3 Likelihood

We model team wins via a Binomial likelihood with a logistic link. For each team i in season s,

$$y_{is} \mid p_{is} \sim \text{Binomial}(N, p_{is}), \tag{1}$$

where $p_{is} \in (0, 1)$ is the underlying win probability for team i.

The logit of p_{is} is modeled as a linear function of the standardized offensive and defensive shot shares:

$$\text{logit}(p_{is}) \;=\; \alpha_s + \mathbf{z}_{is}^{\top} \boldsymbol{\beta}_s \;=\; \alpha_s + \sum_{j=1}^{8} \beta_{s,j} z_{is,j}, \tag{2}$$

where α_s is a season-specific intercept and $\boldsymbol{\beta}_s = (\beta_{s,1}, \dots, \beta_{s,8})^\top$ are the season-specific slope coefficients for the eight included bands (four offensive and four defensive).

Rearranging (2), we arrive at an expression for the win probability:

$$p_{is} = \sigma\big(\alpha_s + \mathbf{z}_{is}^\top \boldsymbol{\beta}_s\big) = \frac{\exp\big(\alpha_s + \mathbf{z}_{is}^\top \boldsymbol{\beta}_s\big)}{1 + \exp\big(\alpha_s + \mathbf{z}_{is}^\top \boldsymbol{\beta}_s\big)}, \tag{3}$$

where $\sigma(\cdot)$ denotes the logistic sigmoid function.

Letting $\boldsymbol{\theta}_s = (\alpha_s, \boldsymbol{\beta}_s)^\top \in \mathbb{R}^9$, the season-$s$ log-likelihood for all teams is

$$\ell_s(\boldsymbol{\theta}_s) = \sum_{i=1}^{n_s} \Big[y_{is} \log p_{is}(\boldsymbol{\theta}_s) + (N - y_{is}) \log(1 - p_{is}(\boldsymbol{\theta}_s)) \Big], \tag{4}$$

where $p_{is}(\boldsymbol{\theta}_s)$ is given by (3).

3.4 Priors

We place independent Normal priors on the intercept and slope coefficients for each season. For the intercept,

$$\alpha_s \sim \mathcal{N}(0, \sigma_\alpha^2), \qquad \sigma_\alpha = 0.5, \tag{5}$$

and for each slope coefficient,

$$\beta_{s,j} \sim \mathcal{N}(0, \sigma_\beta^2), \qquad \sigma_\beta = 0.1, \qquad j = 1, \dots, 8. \tag{6}$$

These priors are weakly informative on the standardized scale, encoding the prior belief that:

1. The league-average team has roughly 50% win probability.
2. A one-standard-deviation change in any shot-share feature is unlikely to change the log-odds of winning by more than roughly ± 0.2 to ± 0.3.

While the parameters undoubtedly have some correlation (particularly between parameters on the same side of the ball), we assume independence for simplicity reasons. Thus the log-prior density for $\boldsymbol{\theta}_s$ is

$$\log \pi(\boldsymbol{\theta}_s) = \log \mathcal{N}(\alpha_s \mid 0, \sigma_\alpha^2) + \sum_{j=1}^{8} \log \mathcal{N}(\beta_{s,j} \mid 0, \sigma_\beta^2). \tag{7}$$

3.5 Posterior and MCMC sampling

Combining (4) and (7), the unnormalized log-posterior for season s is

$$\log \pi(\boldsymbol{\theta}_s \mid \{y_{is}, \mathbf{z}_{is}\}) = \ell_s(\boldsymbol{\theta}_s) + \log \pi(\boldsymbol{\theta}_s). \tag{8}$$

Because of the logistic link and normal priors, the model is non-conjugate and there is no closed-form expression for the posterior distribution of $\boldsymbol{\theta}_s$.

We therefore approximate the posterior via Markov chain Monte Carlo (MCMC), using the random-walk Metropolis algorithm. For each season s we run a single-chain sampleras follows:

- Number of iterations: $T = 105{,}000$ total proposals.
- Burn-in: the first $25{,}000$ iterations are discarded.
- Proposal distribution: for the current state $\boldsymbol{\theta}$,

  $$\boldsymbol{\theta}' = \boldsymbol{\theta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}\big(\mathbf{0}, \Sigma_{\text{prop}}\big),$$

  with $\Sigma_{\text{prop}} = \mathrm{diag}(s_0^2, s_1^2, \dots, s_8^2)$, where $s_0 = 0.05$ for the intercept and $s_j = 0.05$ for all slopes.
- Acceptance probability: with $\log \pi(\cdot)$ given by (8), we accept the proposal with probability

  $$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\big\{1,\ \exp\big[\log \pi(\boldsymbol{\theta}') - \log \pi(\boldsymbol{\theta})\big]\big\} = \min\left\{1,\ \frac{\boldsymbol{\theta}'}{\boldsymbol{\theta}}\right\}.$$

We initialize the chain at $\boldsymbol{\theta}^{(0)} = \mathbf{0}$ and fix the random seed to facilitate reproducibility. For each season we record the acceptance rate and the trace of the log-posterior values $\log \pi(\boldsymbol{\theta}^{(t)})$.

3.6 MCMC diagnostics and effective sample size

To assess convergence and mixing of the random-walk Metropolis sampler, we examine trace plots, autocorrelation functions (ACFs), and effective sample size (ESS) estimates for all parameters. Figure 1 shows representative trace and ACF plots for the parameter corresponding to the 0–3 ft distance in the 2004–05 and 2024–25 seasons. The chain appears well-mixed, with no visible nonstationarity and rapidly decaying autocorrelation.

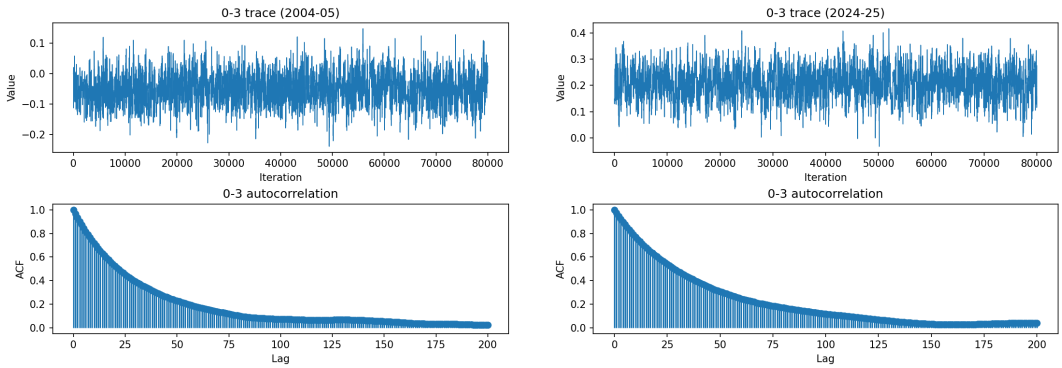

Figure 1: Trace and autocorrelation for the 0–3 ft parameter (β_1) for both seasons.

Effective sample size (ESS) values for the 2004–05 and 2024–25 seasons are shown in Table 1. In both seasons, ESS values are sufficiently large (near a thousand or more), indicating that the posterior summaries are reliable.

Table 1: ESS for 2004–05 and 2024–25 seasons.

	Effective Sample Size (ESS)	
Parameter	**2004–05**	**2024–25**
α	1798.7	2113.3
0–3 (off)	1094.2	918.8
3–10 (off)	1089.0	1166.2
10–16 (off)	980.9	1365.5
16–3P (off)	1166.5	1119.2
0–3 (def)	809.8	704.9
3–10 (def)	902.2	797.9
10–16 (def)	1160.0	1462.9
16–3P (def)	946.4	1264.0

3.7 Posterior predictive fit and residual diagnostics

Posterior mean win probabilities $\hat{p}_{is}$ provide a simple summary of model fit. Figure 2 compares posterior predicted win probabilities with observed "true" win rates for both seasons. The closer points lie to the 45-degree line, the better calibrated the model is. The 2024–25 season exhibits noticeably stronger predictive alignment, consistent with the higher Brier skill score.

Residual diagnostics evaluate whether the model suffers from systematic bias. For each team we compute residuals

$$r_{is} = \frac{y_{is}}{N} - \hat{p}_{is}.$$

Sample residual plots are shown in Figure 3, which displays residuals against the 0–3 ft offensive shot share for both seasons. No clear patterns of bias are visible, suggesting that the model adequately captures the relationship between shot distribution and winning.

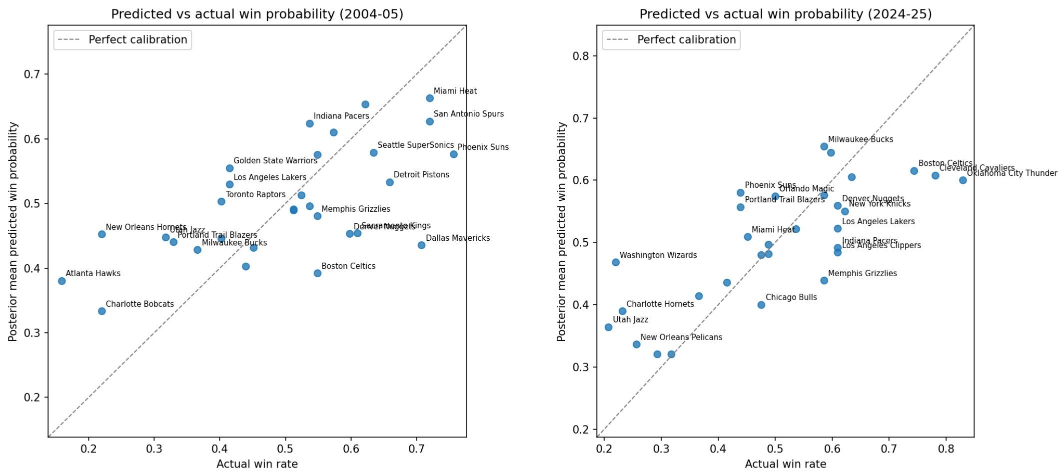

Figure 2: Posterior mean predicted win probability vs. actual win rate for both seasons.

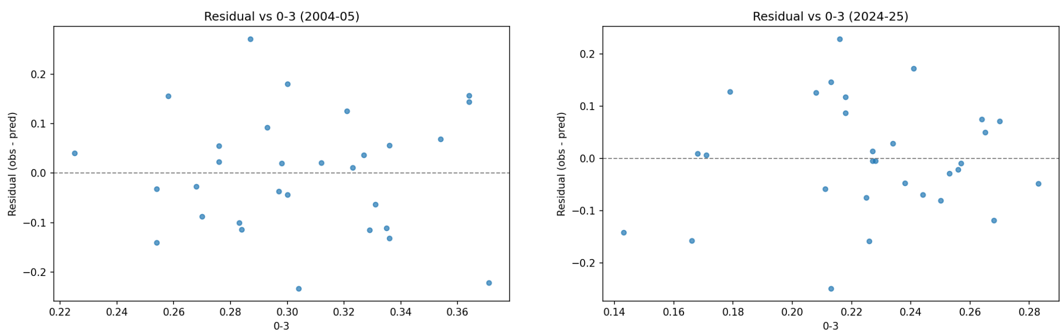

Figure 3: Residuals vs. 0–3 ft offensive shot share for both seasons.

3.8 Cross-validated performance

To assess out-of-sample predictive accuracy, we perform leave-one-out cross-validation within each season. For each team i we refit the model on the remaining $n_s - 1$ teams and compute the posterior mean predicted win probability $\hat{p}_{is}^{(-i)}$. Predictive accuracy is quantified using the Brier score,

$$\text{Brier}_s = \frac{1}{n_s} \sum_{i=1}^{n_s} \left(\hat{p}_{is}^{(-i)} - \frac{y_{is}}{N} \right)^2,$$

and the Brier skill score, which compares the model to a baseline predictor that simply predicts the in-sample average win rate:

$$\text{Skill}_s = 1 - \frac{\text{Brier}_s}{\text{Brier}_{\text{baseline},s}}.$$

Table 2 reports the cross-validated Brier scores and skill scores for both seasons. The 2024–25 model demonstrates markedly higher predictive skill, indicating that offensive and defensive shot distribution explains substantially more variation in team success in the modern NBA than in the earlier era.

Table 2: Cross-validated Brier score summaries for both seasons.

Metric	**2004–05 Season**	**2024–25 Season**
Brier score	0.0237	0.0203
Baseline Brier score	0.0248	0.0274
Brier skill score	0.0425	0.2611

4 Results

This section presents the estimated posterior effects of offensive and defensive shot distribution on team success. Figure 4 displays the posterior mean and 95% credible interval for every shot-distance band are shown in the forest plots below. All effects are interpreted on the log-odds scale per one standard-deviation change in shot share. Note that the 3P and def_3P effects represent the implied contributions of the omitted compositional bucket.

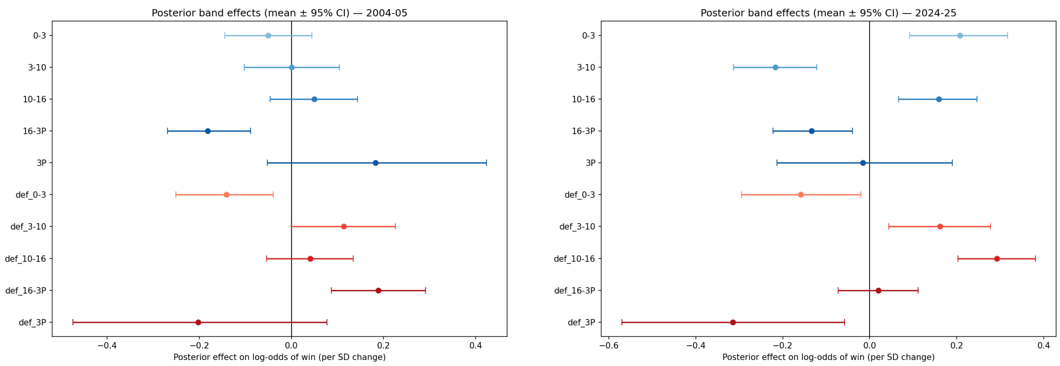

Figure 4: Posterior effects of shot distribution on log-odds of winning (per SD change).

4.1 2004–05 Season

Offensively, it is clear that three-point have a strong positive association with winning. Mid-range bands (3–10, 10–16, 16–3P) on the other hand generally show neutral or negative effects. Moreover, the graph shows that preventing opponents from shooting in the effecient 3P zone was also a major separator between winning and losing teams, with conceding three-point attempts being by far the largest defensive detriment.

One notable observation for this season is the slightly negative effect for offensive rim attempts (0–3 ft). While counterintuitive at first glance, this likely reflects that teams in this era took an excessive share of their shots at the rim. The linear nature of the model does not capture diminishing returns, so the overwhelming number of teams taking a high proportion of rim attempts due to the style of play may have led to this negative association.

4.2 2024–25 Season

In contrast to what one might expect, the three-point effect in 2024–25 weakens significantly, with the interval mean being less than zero, indicating that aking more threes than league average no longer provides the same marginal advantage. Additionally, several mid-range bands now exhibit less negative or even slightly positive effects. The value of offensive rim attempts also explodes positively, suggesting that in the modern era where teams space the floor far more effectively, attacking the rim has become a more potent weapon. This result also somewhat validates the provided explanation for the negative rim effect in 2004–05. namely that the linear constraint of the model prevented capturing diminishing returns in an era where rim attempts were overused. These patterns suggest that as the league has converged on analytically optimized shot selection, marginal returns to additional three-point volume have diminished, and some teams may even benefit from reintroducing a more diversified shot profile.

Defensively however, the effect of conceding large quantities of three-point attempts remains negative, even exceeding the impact seen in 2004–05. This underscores that even if taking more threes may not be as advantageous as before, preventing opponents from doing so is critical. With defending the rim also as important as ever, it is clear that while the optimal offensive shot profile is still debatable, modern defensive schemes must prioritize protecting the rim and the perimeter, while funneling opponents into mid-range areas.

4.3 Interpretation Summary

Overall, these results show that in 2004–05, exploiting high-efficiency zones (rim and three) created substantial competitive differentiation.By 2024–25, league-wide strategic convergence has flattened these advantages. Modern teams may gain edges through spacing that re-opens rim opportunities, retaining credible mid-range threats, and implementing defensive schemes emphasizing rim and perimeter protection while encouraging the mid-range attempts that increasingly unfamiliar to today's players.

5 Discussion

Limitations. Although informative, this analysis is subject to several important limitations:

- The model incorporates only shot-distance distribution, not shot efficiency (FG%, eFG%, TS%), and thus cannot distinguish whether teams are good at generating or defending specific shots.
- The linear structure cannot capture nonlinear "sweet spots" or diminishing returns for particular shot types.
- Pace, possessions, opponent strength, and other contextual variables are not included.
- Only two seasons are analyzed, providing a limited sample for inference and reducing generalizability.
- Potentially important factors such as coaching, injuries, roster changes, and defensive scheme are not modeled.

Future Improvements. Several extensions could strengthen and broaden the scope of this work:

- Incorporate zone-specific shooting efficiency metrics and additional team performance indicators such as pace, turnover rate, rebound rate, and free-throw rate.
- Extend the hierarchical framework using many seasons to jointly estimate league-wide and season-specific effects.
- Move from season-level aggregates to game-by-game modeling to increase sample size and control for opponent-adjusted factors.
- Evaluate posterior predictive performance by holding out entire games or entire seasons.
- Introduce team-specific random effects to account for unobserved individual circumstances (e.g., coaching strategies or roster stability).

All code used in this analysis, including data processing pipelines, MCMC implementation, and visualization scripts, is available in the project repository:

`https://github.com/ethanbabel/Bayesian-Stats-Final-Project`.

References

Feng Wang and Guohua Zheng. What are the changes in basketball shooting pattern and accuracy in national basketball association in the past decade? *Frontiers in Psychology*, 13, 2022. ISSN 1664-1078. doi: 10.3389/fpsyg.2022.917980. URL https://www.frontiersin.org/articles/10.3389/fpsyg.2022.917980.

Tomasz Zajac, Kazimierz Mikolajec, Pawel Chmura, Marek Konefal, Michal Krzysztofik, and Piotr Makar. Long-term trends in shooting performance in the nba: An analysis of two- and three-point shooting across 40 consecutive seasons. *International Journal of Environmental Research and Public Health*, 20(3), 2023. doi: 10.3390/ijerph20031924. URL https://doi.org/10.3390/ijerph20031924.

Shawn Kilcoyne. The decline of the mid-range jump shot in basketball: A study of the impact of data analytics on shooting habits in the nba. Master's thesis, Bryant University, 2020. URL https://digitalcommons.bryant.edu/honors_mathematics/35/.

Marcamus Winn. Does the three point shot affect winning percentage?, 2023. URL https://digitalcommons.latech.edu/mathematics-senior-capstone-papers/37/.