Analysis of Betting Strategies in the NFL Playoffs Through Bayesian Hierarchical Simulation

Submitted by

Brandon Owens

Department of Mathematics and Statistics

Oakland University

November 6, 2025

A competency project in fulfillment of the requirements for MS in Applied Statistics

Committee Members:

Dr. Dorin Drignei (Advisor/Chair)

Dr. Hon Yui (Henry) So

Dr. Gary McDonald

Dr. Subbaiah Perla

TABLE OF CONTENTS

ABSTRACT

The main goal of this competency report is to evaluate a Bayesian hierarchical modeling approach using game data from the 2024–2025 NFL regular season. We develop and analyze a Bayesian hierarchical count regression model to estimate team strengths and predict scoring outcomes in the context of the NFL Playoffs for the aforementioned season. Offensive and defensive abilities are modeled separately, allowing for a flexible representation of team performance. To capture the uncertainty within game outcomes, Monte Carlo simulations are employed, generating predictive distributions for playoff matchups and point differentials. Although profitability proved to be extremely variable due to the small sample size of the playoffs, the analysis demonstrates how hierarchical modeling and simulation can formalize betting evaluations in highly liquid markets. By integrating statistical modeling with simulation-based inference and decision-making frameworks, this competency report contributes to the growing literature on sports analytics, as well as demonstrating the practical evaluation of betting strategies in professional retail sports markets.

1. INTRODUCTION

The legalization of online sports betting in the United States following the repeal of the Professional and Amateur Sports Protection Act (PASPA) in 2018 has fundamentally reshaped the landscape of both sports consumption and sports analytics. Once limited to Nevada and a small number of jurisdictions, sports wagering is now widely available across the United States, and its rapid growth has created new opportunities for fans, bettors, and analysts alike. The expansion of betting markets has paralleled the rise of advanced quantitative methods in professional sports, where statistical modeling has become an increasingly central tool for evaluating team performance, informing front-office decisions, and engaging with fans.

The NFL provides an attractive and challenging setting for statistical modeling, frequently appearing in the *Journal of Quantitative Analysis in Sports*. As an official journal of the American Statistical Association, it formalizes and validates sports data analysis as an emerging and important interdisciplinary area at the intersection of statistics and sports. As one of the most liquid sporting markets, the NFL playoffs combine widespread interest with high economic stakes. These win-or-go-home games, culminating in the Super Bowl, attract intense attention and serve as a natural test case for evaluating the predictive validity of statistical models. The discrete structure of football scoring—touchdowns, field goals, and safeties—further lends itself to probabilistic, count-based approaches, enabling researchers to capture both tabular scoring events and broader team-level performance dynamics.

This report evaluates a Bayesian hierarchical model for predicting NFL game outcomes using data from the 2024–2025 regular season and applying it to subsequent playoff contests. In these models, team performance is decomposed into offensive and defensive strengths, alongside baseline scoring tendencies, with the hierarchical structure enabling the pooling of information across games to improve estimation accuracy. The outputs of these models include scoring intensities, win probabilities, and spread/total outcomes, which are then explored through Monte Carlo simulations to capture the uncertainty inherent in competitive sports. Finally, the predictive results are applied to established sports betting strategies, enabling an assessment of profitability, risk, and consistency across one of the most liquid betting markets in professional sports.

The significance of this competency project has been designed to demonstrate the application of Bayesian frameworks within professional sporting outcomes, linking modelling outputs with real-world betting strategies, and bridging the gap between statistical evaluation and applied-decision making.

2. LITERATURE REVIEW

Although this report is centered on American football, a rich body of prior work has applied Bayesian hierarchical modeling and related count-based approaches across a range of sports. Much of this research focuses on modeling scoring outcomes as stochastic processes, providing a foundation for measuring team strengths and forecasting results.

An original contribution in this area is Baio and Blangiardo (2010), who introduced a Bayesian hierarchical model to analyze outcomes in the Italian Serie A football championship (1991–1992). In their framework, the number of goals scored by each team is modeled as a Poisson random variable, with parameters linked to latent team-specific offensive and defensive abilities. Their work demonstrated the effectiveness of Bayesian methods for combining prior beliefs with observed data, while also addressing overdispersion and structural dependencies inherent in sports outcomes. This framework has since been widely adapted to other domains, establishing a cornerstone for Bayesian applications in sports analytics.

More recent work by Attard et al. (2023) extends these approaches to professional basketball, using Bayesian hierarchical modeling to evaluate scoring intensities in the NBA. Their study explored Poisson and Negative Binomial formulations, as well as logistic regression in a Bayesian context, to capture team strengths and assess predictive accuracy from a win-totals perspective. Importantly, they decomposed scoring processes by play type—two-point shots, free throws, and three-point attempts—illustrating how hierarchical models can adapt to sport-specific structures. While their analysis did not focus on the NFL, the methodological framework and descriptive evaluation of team strengths serve as a key inspiration for this report.

In the context of American football, Mack (2024) provided a comprehensive treatment of Bayesian modeling in his book *Bayesian Sports Models in R*. His work applied hierarchical approaches to NFL scoring outcomes. Notably, Mack modeled conditional events as independent processes, enabling a detailed representation of football scoring but raising questions about interdependencies between conditional scoring opportunities. His approach has strongly influenced the methodological choices in this project, particularly in balancing interpretability with predictive performance.

Beyond statistical modeling, this study also draws from literature on betting strategies. One widely studied approach is the Kelly Criterion (Kelly, 1956), which prescribes optimal bet sizing by maximizing the long-term expected logarithmic growth of wealth. The Kelly framework has been frequently applied in sports wagering research due to its theoretically grounded balance of risk and reward. Another relevant approach is value betting, which emphasizes identifying and wagering on bets where the bettor's estimated probability of success exceeds that implied by bookmaker odds. These strategies provide a natural testing ground for evaluating the practical efficacy of predictive models in betting markets.

Taken together, prior research establishes a strong foundation for applying Bayesian hierarchical count models to sports forecasting and highlights the importance of linking statistical methods to practical betting applications. This project highlights the growth in the domain of sports analytics by applying altered Bayesian frameworks to the 2024–2025 NFL season.

3. METHODOLOGY

3.1 Data and Dataset

The dataset utilized in this competency project was obtained through nflreadR, an R-based API providing comprehensive NFL data. Using this API, team-level play-by-play data for the 2024–2025 NFL regular season was retrieved. From the complete play-by-play dataset, a subset was extracted, focusing specifically on scoring plays and attempt types, in order to support count-based modeling and team performance analyses. Additionally, Week 18 game results were ignored as it is common practice for teams that have clinched playoff positions to rest their starters so that injury is avoided for games that have more significance.

To prepare the data for analysis, game-level aggregation was performed, organizing the dataset by individual games and separating metrics for home and away teams. Team names were standardized to ensure consistency across all observations, and extensive data cleaning procedures were applied to remove corrupted, missing, or inconsistent entries.

In addition to play-by-play metrics, historical sports betting odds were incorporated to contextualize team performance relative to market expectations. These odds were centered around home and away point spreads and were sourced from SportsOddsHistory.com, which compiles historical data referencing BetMGM via Kansas Crossing Casino and Hotel. All data processing, cleaning, and exploratory analysis were conducted using R, while Bayesian modeling and inference were performed using the Hamiltonian Monte Carlo algorithm (specifically the No-U-Turn Sampler) via RStan, leveraging its flexibility to estimate hierarchical count models and account for team and game-level variability.

3.2 Scoring in the National Football League

An important distinction regarding the scoring abilities of sports should be noted against referenced literature – American football has numerous vehicles of scoring as opposed to other sports such as soccer, where goals are the only method of accruing points, and basketball, where teams only have three opportunities to put the ball in the basket (free throws, two point shots, and three point shots).

In the NFL, these scoring methods are captured in the following manner:

- Touchdowns
    - Passing touchdowns
    - Rushing touchdowns
    - Return touchdowns
    - Defensive touchdowns
- Field Goals
- Safeties
- Extra-Points (*)
- Two-Point Conversions (*)

It is also important to note that touchdowns can be broken into the four established groups above and that modelling these abilities of teams separately is important – a team can have a strong receiver core with excellent pass blocking, resulting in plenty of passing touchdowns, but struggle while running the ball. Additionally, the scoring methods with an asterisk(*) are conditional on touchdowns being scored. Thus, after a touchdown is achieved by a team, said team has a choice to either attempt an extra point or two point conversion.

3.3 Poisson-Centric Modeling

In this competency report, multiple models separately shall be considered for the scoring types provided:

$$PassingTouchdowns_{gj} \mid \theta_{gj_{PassTD}} \sim Poisson\ (\theta_{gj_{PassTD}}) \quad (1A)$$

$$RushingTouchdowns_{gj} \mid \theta_{gj_{RushTD}} \sim Poisson\ (\theta_{gj_{RushTD}}) \quad (1B)$$

$$ReturnTouchdowns_{gj} \mid \theta_{gj_{RetTD}} \sim Poisson\ (\theta_{gj_{RetTD}}) \quad (1C)$$

$$DefensiveTouchdowns_{gj} \mid \theta_{gj_{DefTD}} \sim Poisson\ (\theta_{gj_{DefTD}}) \quad (1D)$$

$$FieldGoals_{gj} \mid \theta_{gj_{FG}} \sim Poisson\ (\theta_{gj_{FG}}) \quad (1E)$$

$$Safeties_{gj} \mid \theta_{gj_{Safety}} \sim Poisson\ (\theta_{gj_{Safety}}) \quad (1F)$$

where g represents the index of the game, j represents whether the team was away or home (1 as the home team and 0 as the away team). Thus, $PassingTouchdowns_{gj}$, $RushingTouchdowns_{gj}$, $ReturnTouchdowns_{gj}$, $DefensiveTouchdowns_{gj}$, $FieldGoals_{gj}$, and $Safeties_{gj}$ represent the observed count for touchdown types, field goals, and safeties by team j in the gth game of the 2024-2025 season, respectively. Each of these observed counts of scores are modelled via the Poisson distribution:

$$P(M = m) = (e^{-\theta gj_M} * \theta gj_M^{\ m}) / (m!) \quad (2)$$

where $m \in \{0,1,2,\dots\}$ is the observed number of events, and θgj_M is the parameter representing both the mean and variance (i.e., the “rate”) for the distribution for team j in the gth game of the season. This function models the probability of observing exactly m events of scoring type M (passing touchdowns, rushing touchdowns, return touchdowns, defensive touchdowns, field goals, or safeties). Substituting each scoring type M into the distribution specifies the probability of the corresponding observed score count m, occurring with rate θ. The parameter θ is of primary interest, as it captures the latent attacking and defending strengths of teams. We will dive into this further.

It is worth noting that football scoring types, especially different categories of touchdowns often exhibit overdispersion. Overdispersion is a statistical phenomenon where the observed variance in a data set is greater than the variance of the proposed statistical model and is commonly occurring in count data, leading to inflated standard errors. While the Negative Binomial serves as an alternative to account for this overdispersion, the focus of the Poisson assumption stems from Mack's evaluation of Poisson versus Negative Binomial, as well as Attard's Poisson-based basketball scoring models. Our choice of a Poisson framework reflects the precedent, though it may sacrifice some fit in exchange for interpretation.

Thus, in further detail, $\theta_{gj_{PassTD}}$, $\theta_{gj_{RushTD}}$, $\theta_{gj_{RetTD}}$, $\theta_{gj_{DefTD}}$, $\theta_{gj_{FG}}$, and $\theta_{gj_{Safety}}$ represent the rate of scoring intensity of a team in the gth game by considering latent attack and defensive strength, home advantage, and an intercept common to the scoring rate. In other words, these parameters are modelled using a log-linear random effect model on a game by game basis. We model team scoring intensities through a log-linear specification indexed by scoring type M:

$$log(\theta_{g1_M}) = \alpha_{g1,M} + \beta_{g0,M} + int_M + home_M \qquad (3)$$

$$log(\theta_{g0_M}) = \alpha_{g0,M} + \beta_{g1,M} + int_M \quad (4)$$

Here, $\alpha_{g1,M}$ is home team's latent attacking strength for scoring type M in the gth game of the season, $\beta_{g0,M}$ is away team's latent defensive strength against scoring type M in the gth game, int_M is a scoring-type-specific intercept that captures the baseline scoring propensity across the league, and $home_M$ captures the home advantage effect.

For a given game g, g1 indexes the home team and g0 the away team, Thus, the expected intensity of home-team scoring depends on its own attack parameter $\alpha_{g1,M}$, the away team's defense parameter $\beta_{g0,M}$, and the scoring-type baseline/intercept adjustments.

A key identification issue arises because $\alpha_{g1,M}$ and $\beta_{g0,M}$ appear additively in the log-intensity. Without constraints, the model cannot distinguish whether high scoring comes from stronger home-team attack ($\alpha_{g1,M}$) or weaker away-team defense ($\beta_{g0,M}$). To address this confounding, we impose identifiability constraints of centering such that each $\Sigma\ \alpha_{T,M} = 0$ and $\Sigma\ \beta_{T,M} = 0$ (where T references team T's seasonal latent parameter, and M references the scoring type), as well as hierarchical priors that shrink attack and defense parameters around a common league-average baseline. With these constraints, the interpretation is preserved with each of the 32 teams having its own latent attack and defense parameters for each scoring type M, and the model cleanly attributes scoring intensity to the interaction of offensive and defensive strengths, home advantage, and league-wide tendencies.

While the baseline scoring intensities are modeled using a Bayesian Poisson regression framework, the structure of American football requires special attention to conditional scoring types. In particular, following a touchdown, a team may choose either to attempt a two-point conversion or to kick an extra point. These conditional scoring events naturally lend themselves to Binomial modeling because they constitute discrete successes with a finite number of trials. Formally, we model the number of attempts of scoring type M by team j in game g as:

$$N_{gj_M} \sim Binomial(TotalTouchdowns_{gj}, \pi_{jM}) \tag{5}$$

where N_{gj_M} is the observed count of scoring type M attempted by team j in game g, $TotalTouchdowns_{gj}$ is the total number of touchdowns scored by team j in that game, and π_{jM} is the latent probability that team j selects conditional scoring type M (either a two point conversion or an extra point) after having scored a touchdown. This formulation allows us to estimate team-specific tendencies for choosing different scoring types while accounting for the total number of touchdown opportunities.

Conditional on attempting the scoring type, the number of successes is modeled as

$$x_{gj_M} \sim Binomial(N_{gj_M}, \gamma_{j_M}) \tag{6}$$

where x_{gj_M} is the number of successful outcomes of conditional scoring type M for team j in team g, N_{gj_M} follows from Equation 5 and is the number of attempts of that scoring type, and γ_{j_M} is the latent probability of success for team j on scoring type M. This formulation models the likelihood of success conditional on a team attempting either a field goal or extra point following a touchdown.

This structure mirrors established practice in sports statistics and generalized linear modeling, where Poisson distributions capture counts of scoring events (Maher, 1982), and Binomial distributions are used for repeated discrete attempts with success/failure outcomes (Maher, 1982). In particular, the Binomial specification ensures that the decision process and success probability are estimated at the team level, yielding interpretable parameters that can vary across organizations.

By embedding these Binomial processes within the broader Bayesian framework, we can capture uncertainty in team strategies and uncertainty in execution, while allowing for hierarchical priors that shrink team-level parameters toward league-wide tendencies.

Within this model framework, the following priors and hyperpriors are specified:

$$home_M \sim Normal(0,\ 0.05), \quad int_M \sim Normal(0,\ 0.05) \tag{7-8}$$

where the $home_M$ captures the home advantage and int_M represents the scoring-type intercept. The relatively tight prior variance penalizes uncommon scoring types (such as return touchdowns, defensive touchdowns, and safeties), reflecting the lower frequency of these events and is inspired by previous literature (Attard, 2023).

For team-level attack and defensive parameters, we follow a hierarchical Bayesian approach similar to that in Baio & Blangiardio (2010) and extended in Attard (2023). For scoring type M:

$$\alpha_{tM} \sim Normal(\mu_{\alpha M},\ \sigma_{\alpha M}),\ \beta_{tM} \sim Normal(\mu_{\beta M},\ \sigma_{\beta M}) \tag{9-10}$$

with hyperpriors:

$$\mu_{\alpha M},\ \mu_{\beta M} \sim Normal(0,\ 0.05) \tag{11}$$

$$\sigma_{\alpha M},\ \sigma_{\beta M} \sim Gamma(0.1,\ 0.1) \tag{12}$$

Here, α_{tM} still represents team t's attacking strength for scoring type M, while β_{tM} represents its defensive strength against scoring type M. The Normal priors on the means ensure that team-level parameters are centered around a league-wide baseline, while the Gamma priors allow for adaptive pooling, shrinking estimates toward the overall mean.

Finally, with respect to the conditional scoring options of extra points (EXP) and two-point conversions (2P), we assign the following priors:

$$\delta_{jEXP},\ \delta_{j2P} \sim Dirichlet(9, 1) \tag{13}$$

which govern the relative decision probability of attempting an extra point versus a two-point conversion for team j. A Dirichlet prior was chosen because probabilities for each conditional scoring options are constrained to the simplex of $\delta_{jEXP} + \delta_{j2P} = 1$ ($\delta_{jEXP}>0$ and $\delta_{j2P}>0$), and which directly encodes exclusivity of the decision. Additionally, the weight of the dirichlet reflects the empirical imbalance of the decision – with the vast majority of conditional attempts coming in the form of extra points.

After a decision is made on the post-touchdown conditional scoring type to attempt, team level success probabilities conditional on attempting an extra-point or two-point conversion are identified as

$$\psi_{jEXP} \sim Beta(9, 1), \qquad \psi_{j2P} \sim Beta(1, 1) \tag{14-15}$$

where Beta distribution is a natural choice here because it is conjugate to the Binomial likelihood, allowing for straightforward updating of beliefs with observed data. The hyperparameters reflect empirical scoring tendencies in the league: extra-point attempts are nearly always successful, hence the Beta(9,1) prior concentrates probability near 1 for the success that an extra point attempt will be successful, whereas two-point conversions are less reliable and more variable, motivating the uninformative Beta(1,1) prior (which is the same as Uniform(0,1). In this way, these priors encode both strategic realities of the game and maintain desirable statistical properties for inference.

As similarly provided in Attard (2023), the following Directed Acyclic Graphs (DAGs) may be helpful to visualize the setup of this model:

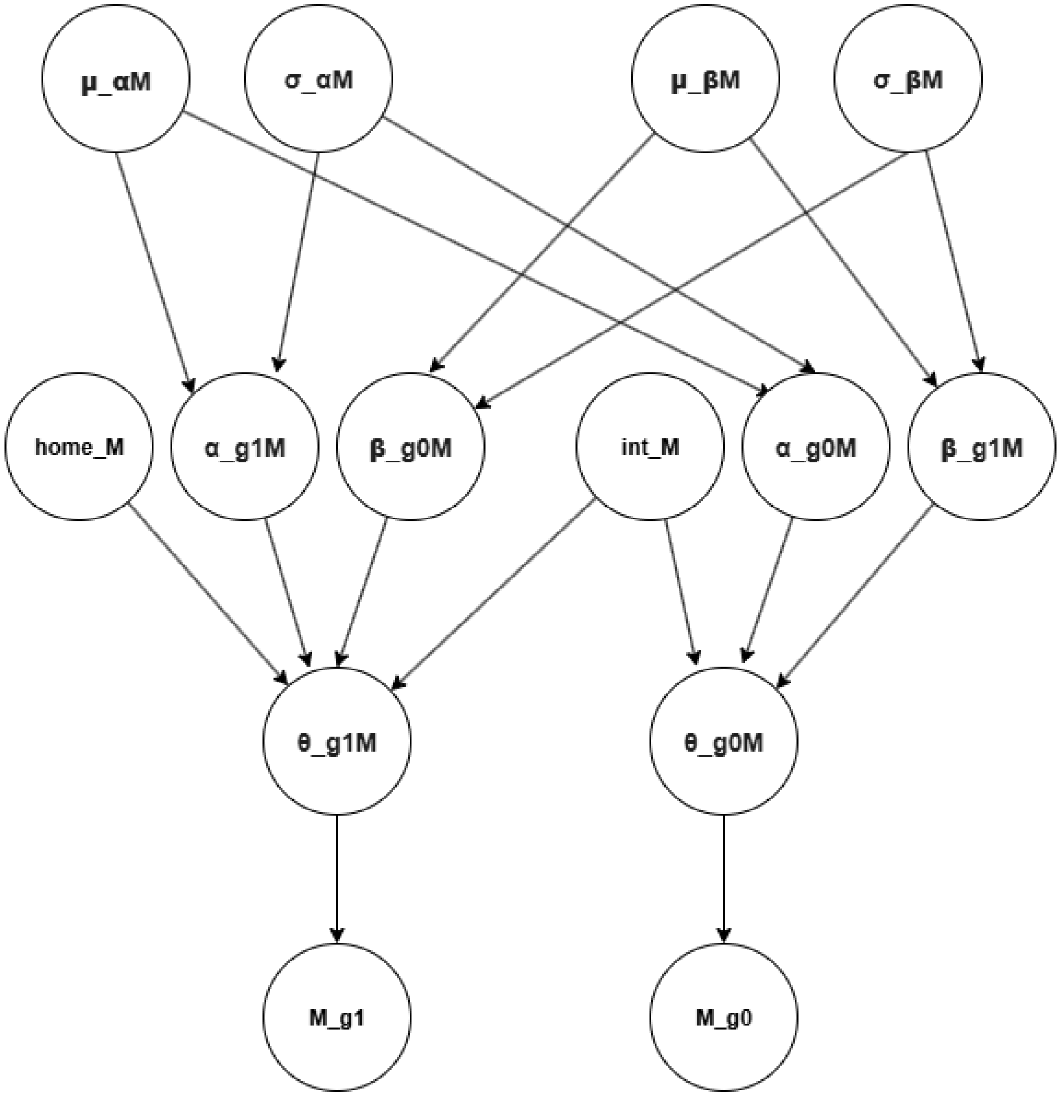


Figure 1: DAG of the general case for scoring models using the Poisson Regression.

Here, g1 and g0 still represent the denotation of the home and away team for a specified game, respectively and M is the scoring type of interest. As provided above, Equations (11) and (12) provide our hyperpriors $\mu_{\alpha M}$, $\mu_{\beta M}$, $\sigma_{\alpha M}$, and $\sigma_{\beta M}$ that serve to partially pool teams and inform the prior distributions in (9) and (10): α_{g1M} / β_{g0M} (representing the offensive latent strength for the home team in game g and the defensive latent strength of the away team in game g) and α_{g0M}

/ β_{g1M} (representing the offensive latent strength for the away team in game g and the defensive latent strength of the home team in game g). Additionally, priors for home advantage and intercept terms are fit for all scoring types as denoted by home_M and int_M (and can be referenced in Equations (7) and (8)). These parameters are used within Equations 3 and 4 general linear poisson link, where the computed $\theta_{gj_{PassTD}}$ models the likelihood of the Poisson distribution for the observed actual count of scoring event M (Equation 1).

Additionally, the DAG for conditional scoring types is provided as:

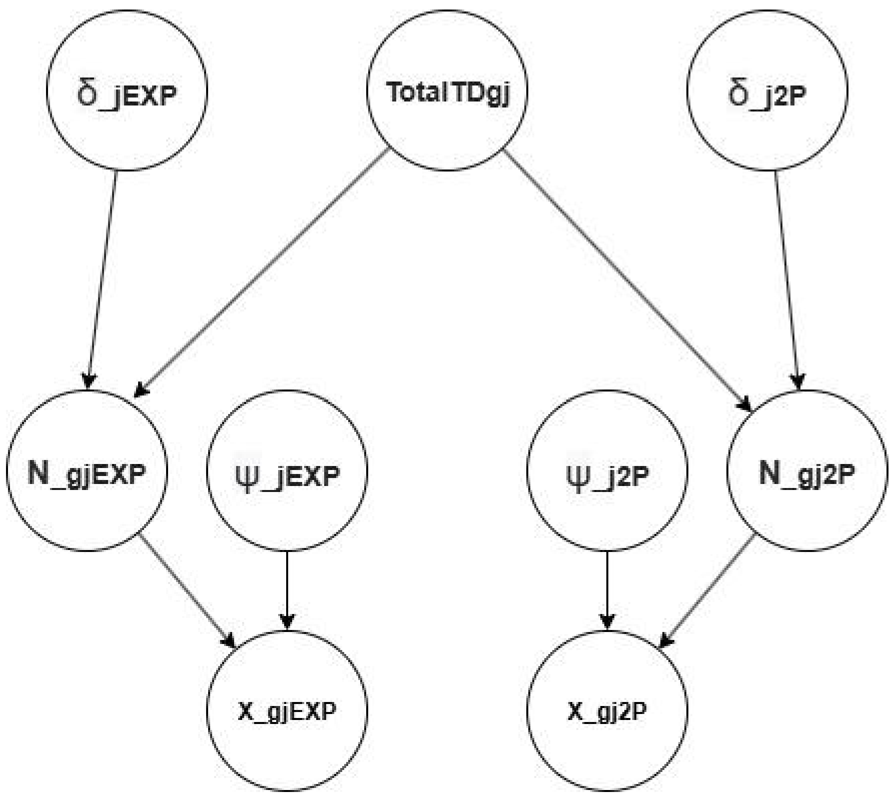


Figure 2: DAG of conditional cases after a touchdown is scored.

Similar to Figure 1, this DAG models the δ_{jEXP} hyper prior (found in Equation (13)) of choice between an EXP (extra point) and 2P (two point conversion) from team j to assume the likelihood of N_{gj_M} in equation five to these respective scoring types. From the N_{gj_M} number of extra point and two-point tries attempted, ψ_{jEXP} and ψ_{j2P} (from Equations 14 and 15) serve as prior distributions of a team's ability to successfully score an extra point or two point success relative to their N_{gj_M} attempts tried. These follow the binomial likelihood in Equation 6, that contains observations of x_{gj_M}, the observed successes from conditional scoring attempts.

3.4 Monte Carlo Simulation

After the data was cleaned and the model was designed, posterior samples were obtained using Hamiltonian Monte Carlo (HMC) with the No-U-Turn Sampler (NUTS). HMC is a Markov Chain Monte Carlo (MCMC) method that uses concepts from physics – specifically Hamiltonian dynamics – to explore complex posterior distributions. Unlike traditional MCMC, HMC treats the parameters as positions of a particle moving in a potential energy landscape defined by the negative log-posterior (Alter & Wainwright, 1959). The particle's trajectory is simulated, and HMC proposes new parameter values that are far from the current state, but still have a high probability of acceptance. Meanwhile, NUTS is an extension of HMC that determines the appropriate trajectory of length for each sampling iteration, making HMC much more efficient. NUTS avoids trajectories that "turn back" on themselves, ensuring efficient exploration of the posterior (Hoffman & Gelman, 2011). Together, HMC and NUTS provide a powerful and efficient framework for obtaining high-quality posterior samples. This sampling methodology was conducted over 10,000 iterations per chain, with 1,000 warm-up iterations, across 4 parallel chains. To reduce storage requirements and mitigate autocorrelation in the draws, the chains were thinned at a fixed interval.

Convergence was evaluated using multiple diagnostics. The Gelman–Rubin statistic $((\hat{R}) \sim \sqrt{1 + m/ESS}$, where m is the number of chains and ESS is the effective sample size), was equal to 1.0 for all parameters, indicating good chain mixing and convergence to the target distribution. The computed effective sample sizes (ESS) for both bulk and tail estimates were sufficiently large given the number of iterations, ensuring stable estimation of posterior means and credible intervals. Finally, trace plots were examined to confirm proper mixing and to verify that the sampler did not exhibit divergences or become stuck in localized regions of the posterior space. These trace plots visualize the progression of parameter values for each chain over specified iterations and are used to assess the convergence of a Markov chain. Convergence is indicated by the thick, fuzzy horizontal bands, indicating that the parameter is fluctuating around its stationary distribution.

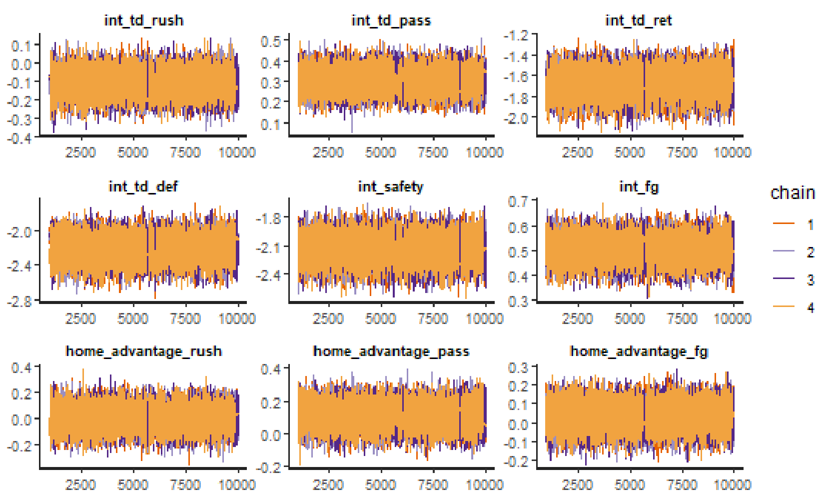


Figure 3: Trace Plots of Posterior Sampling

We estimated latent parameters representing each team's offensive and defensive strengths across all scoring types. After each playoff round—the Wild Card, Divisional, and Conference

Championships—we updated the posterior predictive distributions to incorporate the most recent data:

$$f(\tilde{Y} \mid Y, \theta) = \int f(\tilde{Y} \mid \theta)\, f(\theta \mid Y, prior)\, d\theta \tag{16}$$

Through this process, we obtained posterior samples of team-specific offensive and defensive parameters, which were then used to update the corresponding parameters from the previous playoff round for each scoring type. These updated estimates enabled us to quantify both the relative offensive and defensive capabilities of all 32 NFL teams and the probabilities of specific post-touchdown scoring decisions, along with their expected success rates.

Building on these estimates, we developed a Monte Carlo simulation framework to model playoff matchups. For a given home team and defensive opponent, the algorithm simulates each game 10,000 times. Each simulation incorporates team-specific latent strengths and probabilistic scoring decisions sampled from the posterior distributions. By aggregating results across all iterations, we generated estimates of win probabilities, expected points, and scoring distributions for each matchup—offering a probabilistic foundation for forecasting playoff outcomes.

3.5 Evaluation

Building on the probabilistic outcomes generated from our Monte Carlo simulations, we evaluated the performance of two distinct bankroll management strategies within the 2024 playoff scenario. Each strategy was applied consistently across all simulated games to avoid confounding effects from strategy interactions. Only bets that focused on teams to cover the spread of each playoff game was analyzed. The strategies were as follows:

- *Kelly Criterion* – Strategy that adjusts wager size proportionally based on the calculated edge and probability of success.

$$f = (bp - q) / b \tag{17}$$

where f is the fraction of the bankroll dedicated towards a transaction, b represents the decimal odds, p represents the probability of the bet being successful, and q represents 1 - p. Thus, the simulated probabilities from each game serve to determine the portion of bankroll being deployed.

- *Flat Unit*– An approach that allocates a fixed wager amount on every bet, independent of confidence or probability of success.

where every bet receives the same percentage of the overall bankroll.

Both strategies assumed an initial bankroll of $100, reflecting a standard wager scale in American odds. These strategies were evaluated using two decision-making frameworks:

- *Value Based Decision* – Bets are only placed when the simulated probability of success exceeded the sportsbook's implied odds after adjusting for hold percentage:

$$P_{underdog} = P- = O_{-} / (O_{-} + 100) * 100 \tag{18}$$

$$P_{favorite} = P+ = 100 / (O_{+} + 100) * 100 \tag{19}$$

$$Hold\,\% = [1 - (P_{favorite} + P_{underdog})] / 2 \tag{20}$$

where $P_{underdog}$ is the implied probability of the underdog bet being successful, $P_{favorite}$ is the implied probability of the favorite side of the bet being successful, and the hold percentage shows the percentage of profitability taken by a market maker. Thus, bets were only placed when "value" is found – the simulated odds for a bet need to be greater than the sportsbook's implied probability aggregated with their hold percentage to constitute a bet placement.

- *Straight Decision*: Bets were placed by directly comparing the median simulated outcomes to the evaluative lines to infer expected results.

meaning that if the simulated probability for a game is greater than fifty percent for one side of the spread, that side is wagered on.

Thus, starting with an initial standardized bankroll of $100 (as is American standard – all betting lines are described in spending/earning structures of $100), each simulated game involved selecting two opposing teams, sampling their posterior parameters 10,000 times, and generating full Monte Carlo simulations of scoring events. These simulated scoring events were then translated into actual point outcomes—touchdowns (6 points), field goals (3 points), two-point conversions and safeties (2 points each), and extra points (1 point)—to reconstruct realistic final scores.

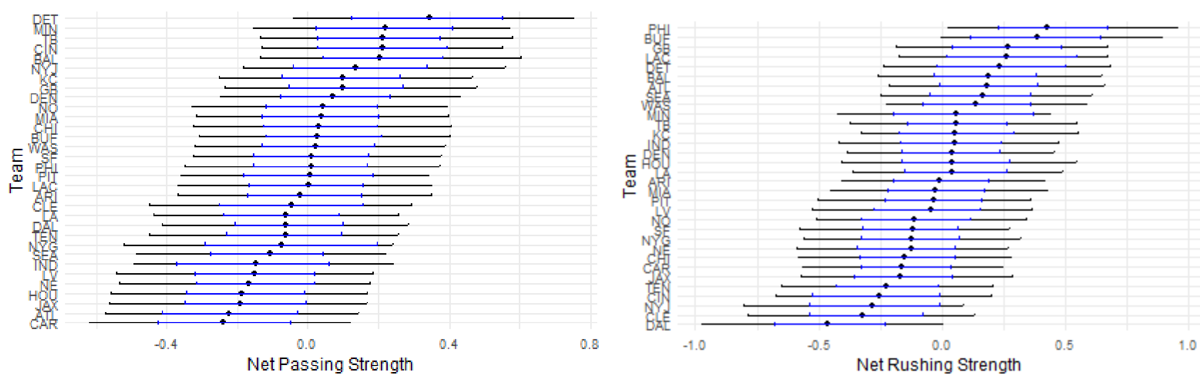
From these distributions, we derived team-level win probabilities and point differentials, which were subsequently compared to sportsbook evaluative lines. Applying the defined decision-making frameworks, each betting strategy (Kelly Criterion and Flat Unit) produced a sequence of wagers across the simulated playoffs.

Ultimately, this approach allows for a quantitative evaluation of how effectively the model's posterior predictive distributions capture real betting outcomes—specifically, the accuracy and profitability of spread predictions when integrated into realistic wagering scenarios.

4. RESULTS AND ANALYSIS

4.1 Visualizing Team Strengths

Before digging into validation and betting strategy with the model, it is useful to visualize team abilities across different scoring contexts. Here, we focus on the main scoring methods: passing touchdowns, rushing touchdowns, field goals, and extra point attempts.



Figures 4, 5: Team Net Passing/Rushing Strengths

Figures 4 and 5 display team net strengths in passing and rushing, where net strength is defined as a team's attacking ability minus its defensive ability. Teams with strong overall rushing and passing abilities will have latent ability distributions that lie to the right of zero, while teams with weaker rushing and passing abilities will fall to the left. As discussed previously, the latent attacking and defensive parameters are centered so that their overall sum equals zero. This implies that teams with strong attacking ability will have positive values, whereas teams with strong defensive ability against scoring type M will have negative values. Therefore, when we subtract a negative defensive parameter (indicating strong defense against scoring type M) from a positive attacking parameter (indicating strong attack for scoring type M), the resulting value will be positive. In other words, stronger teams will tend to produce more positive values, while weaker teams will tend toward more negative ones. Each visualization shows 95%

credible intervals for these net strengths. Specifically, posterior draws of defensive abilities are subtracted from attacking ability to obtain net strength samples, from which the 2.5th percentile, posterior mean, and 97.5th percentile were calculated to construct the intervals.

As a sanity check, these intervals can be compared to observed MOV, or ‘Margin of Victory’ – a commonly used benchmark that measures the average amount of points a team outscores their opponents per game. As shown in Figure 5, our estimated net strength parameters align closely with observed team MOV values.

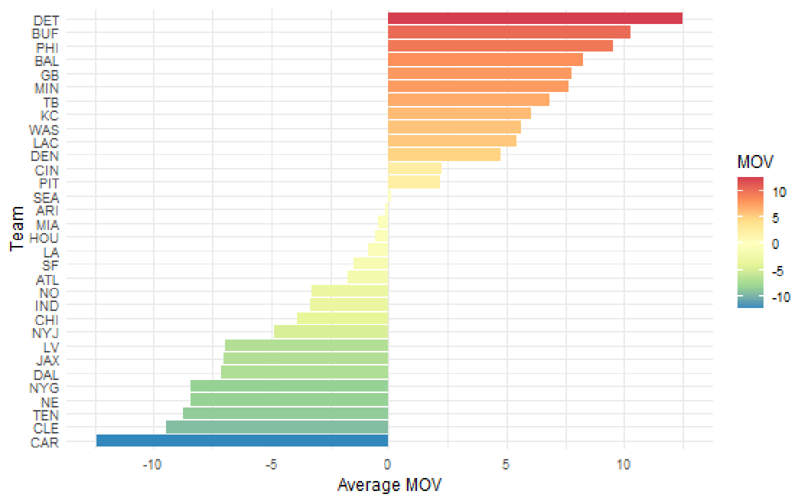
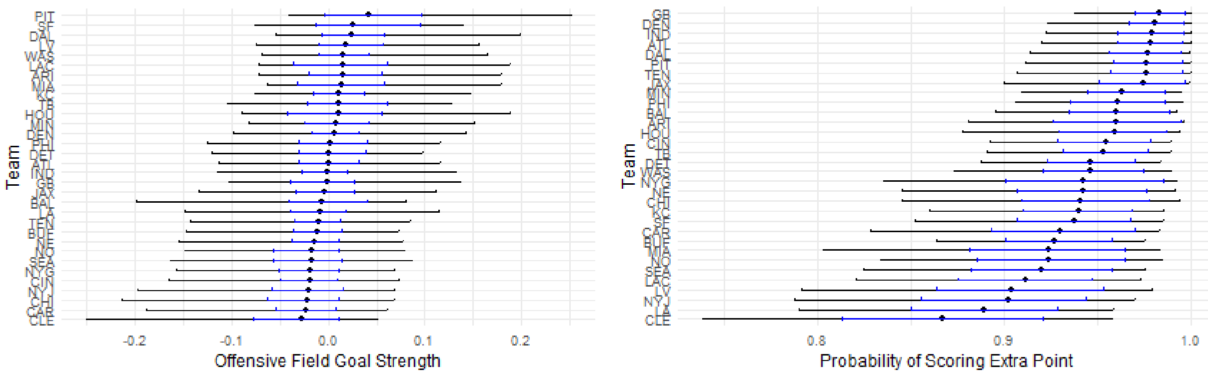


Figure 6: Team Observed MOV

Additionally, with respect to modelling important scoring capacities of teams in a non-touchdown capacity, we look at field goals and extra point success rates:



Figures 7, 8: Team Kicking Ability

Figures 6 and 7 highlight team kicking ability (created similarly to Figures 3 and 4). Notably, the left skew shape found within the Cleveland Browns' distributions highlight the weakness of their kicking units. This makes sense as witnessed in the 2024-2025 NFL regular season, the Browns performed far worse than any other team, converting only 66.7% of field goals (last in the league) on 27 total makes (last in the league) and 85.7% of extra points (again– last in the league). This outcome matches expectations from their season performance.

As described, each visualization illustrates a 95% credible interval for a team's estimated parameter. The narrower blue band represents one standard deviation around the posterior mean, obtained from the 16th and 84th percentiles of the posterior draws. Overlapping intervals across teams highlight the uncertainty captured through Monte Carlo simulation. For instance, while the Browns' mean field-goal strength overlaps with other teams, their posterior distribution is skewed to the right, emphasizing greater variability and inconsistency despite a similar central estimate.

4.2 Playoff Game Values from Monte Carlo Simulations

Across the 13 games (which include all games from the 2024-25 playoffs– 6 wildcard games, 4 divisional games, 2 conference games, and the super bowl), the Monte Carlo simulations provided a probabilistic assessment of team outcomes across a wide range of potential scoring scenarios. By repeatedly simulating each matchup, we were able to quantify expected point differentials and the likelihood of teams covering the spread.

For example, running 10,000 simulations of the Super Bowl produced the following scoring distributions for the Kansas City Chiefs and the Philadelphia Eagles:

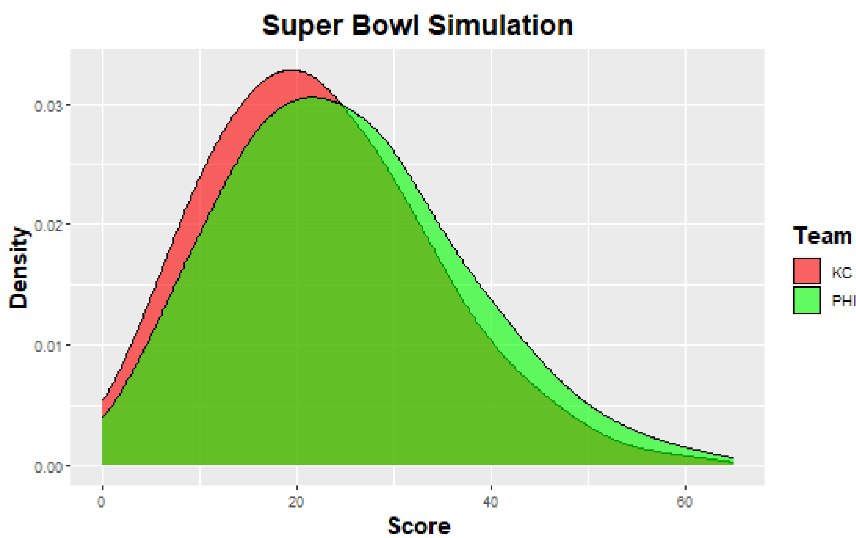


Figure 9: SuperBowl Scoring Simulation

As shown, the distributions are relatively close, with the Eagles – the eventual Super Bowl champions who were not favored according to the books – holding a slight scoring edge. Beyond final score outcomes, we can also examine latent parameters specified in the methodology section. One such parameter is Team Attack Rushing Strength, where simulations indicated that the Eagles possessed the strongest rushing offense in the NFL:

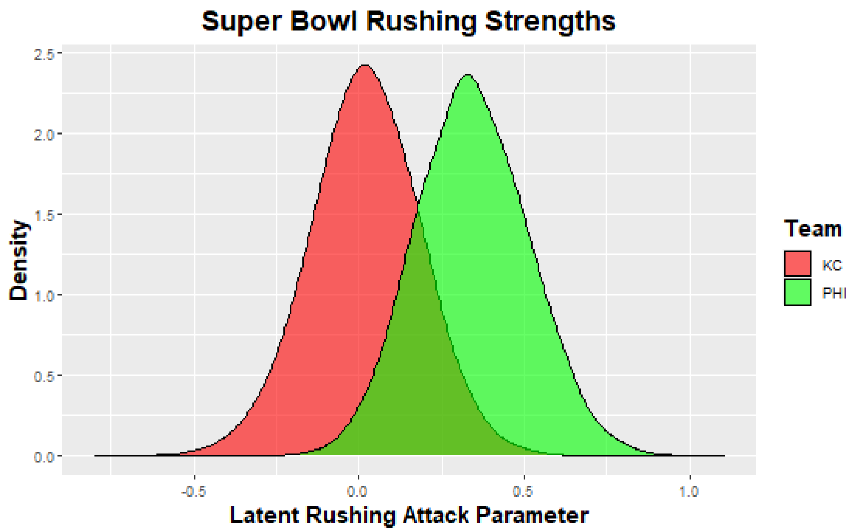


Figure 10: SuperBowl Latent Offensive Rushing Strengths

Additionally, it is important to emphasize that the analysis is based on a limited playoff sample of predictions. With only 13 games being simulated, the stability of any single betting strategy derived from these results is bound to have excessive variance. Therefore, the findings should not be interpreted as offering a prescriptive or reliable sports betting framework. Instead, the purpose of this analysis is to illustrate how statistical tools – specifically a Bayesian hierarchical Poisson model paired with Monte Carlo simulations – can be applied within a sports betting context.

4.3 Application of Betting Strategies to Results (Backtesting on the 2024-25 Playoffs)

To translate these simulation results into a practical tool, aforementioned betting strategies were applied to the projected playoff outcomes derived from 10,000 simulation realizations of every individual postseason game. The objective of this analysis was to evaluate whether the probabilistic forecasts generated by the Bayesian hierarchical model could be leveraged to produce profitability under different bankroll management and wager schemes with focus on spread betting. Taking the opening sportsbook odds from BetMGM for each game, decisions created from our strategic outline using our simulated probabilities were utilized. Table 1 presents the median profit or loss (P/L) and the 95% prediction intervals across all simulation trials for each strategy, along with a singular choice measure for teams covering the spread. This measure represents the aggregate profit from a single betting decision informed by the overall simulation-derived probabilities, rather than from per-game stochastic results. In essence, the quantiles were calculated from the distribution of profit/loss outcomes across all 13 aggregated games, repeated over 10,000 simulations, while the singular choice outcome was determined by applying each betting strategy once according to the probabilities inferred from the Monte Carlo simulations

Strategy	**Median P/L**	**95% Prediction Interval P/L**	**Summarized Singular Choice**
Kelly Criterion - Value	0.50%	(-15.45, 12.87)	-2.88%
Kelly Criterion - Straight	11.50%	(3.84, 12.87)	**7.68%**
Flat Unit - Value	2.80%	(-56.00, 47.93)	-10.87%
Flat Unit - Straight	41.70%	(14.70, 63.40)	**28.70%**

Table 1: P/L Strategy Outcomes

Interpretation and Discussion:

The Flat Unit – Straight and Kelly Criterion – Straight strategies continue to demonstrate the strongest performance, both in magnitude and consistency of returns.

The Flat Unit – Straight approach achieved the highest median return (41.7%), albeit with a wide 95% prediction interval (14.7%, 63.4%), suggesting high upside potential alongside considerable volatility. In contrast, the Kelly Criterion – Straight strategy produced a more moderate median return (11.5%) with a narrower 95% interval (3.84%, 12.87%), reflecting a more stable yet still favorable profit trajectory across simulated trials.

By comparison, the “value”-driven strategies—those leveraging discrepancies between modeled and market-implied probabilities—underperformed. The Kelly Criterion – Value approach yielded a near-neutral median return (0.5%) within a 95% interval (-15.45%, 12.87%), implying limited evidence of consistent market inefficiency. Similarly, the Flat Unit – Value strategy posted a small positive median (2.8%) but with substantial variance (-56.0%, 47.93%) and a negative summarized singular outcome, underscoring the high noise and low reliability of value-based opportunities in this playoff sample.

Additionally, the “Summarized Singular Choice” column provides an additional diagnostic of the model’s calibration. Positive summarized returns in the straight-based strategies indicate even when stochastic variability is removed, the posterior mean probabilities retain predictive integrity to inform single-decision outcomes.

Comparative Insights

Overall, these results suggest that in the specific context of playoff betting – where the number of observations is small and matchups are tightly balanced – that the Flat Unit approaches capture higher short-term gains but expose the bettor to considerable variance and downside risk. In contrast, the Kelly Criterion provides a risk-adjusted approach that moderates potential losses during unfavorable stretches.

Value-based bankroll strategies, while theoretically optimal in an efficient market, proves difficult to execute successfully in a postseason environment. The simulations reveal a limited frequency of scenarios where true market mispricings exist, implying that opportunities to exploit market differentials are rare and highly sensitive to estimation error.

It is also important to note that these findings should not be interpreted as evidence of a guaranteed profitable strategy. The limited number of playoff games, coupled with inherent uncertainty in team ability and market dynamics, restricts generalizability. Instead, these simulations provide a framework for evaluating the interaction between predictive modeling, bankroll allocation, and market structure, illustrating how probabilistic forecasts can be integrated into decision-based analysis rather than serving as a direct instruction for gain.

5. CONCLUSION

The analysis of playoff game values through Monte Carlo simulations, using Bayesian hierarchical modeling of scoring types, provided valuable insights into the behavior of different betting strategies under highly competitive conditions. Across both straight and value-based decisions, as well as proportional (Kelly Criterion) and flat unit wagers, the results demonstrated variable profitability. While flat unit bets yielded nominal gains for betting the spread, value-based strategies underperformed, producing negative returns in several scenarios. This underscores the difficulty of achieving consistent profits in the NFL playoffs, widely recognized as one of the most liquid and competitive markets in sports betting.

These results reinforce a critical point: Poisson-based count models and their Bayesian extensions, while useful as a foundation, may not be independently sufficient on their own to consistently outperform sportsbooks in major markets. Their assumptions of independence and constant scoring rates limit their capacity to capture contextual dynamics such as injuries, coaching adjustments, and game states. Moreover, sportsbooks already employ models of comparable or greater sophistication, leaving little exploitable inefficiency at the game level.

Nonetheless, the framework developed here is a valuable starting point. It establishes a proof of concept for using simulation-based inference to evaluate betting strategies and highlights next steps for more complex modeling approaches. More promising opportunities likely lie in niche markets, such as player prop bets, where sportsbooks rely on less robust modeling and struggle from data scarcity. Extending Poisson-type models to player-level outcomes, incorporating more complex statistical features, and relaxing assumptions could yield more predictive power.

REFERENCES

Alder, B. J. and Wainwright, T. E. (1959). Studies in molecular dynamics. I. General method. Journal of Chemical Physics, 31:459–466.

Attard, P., Suda, D., & Sammut, F. (2023). Bayesian hierarchical modelling of Basketball Team Performance: An NBA regular season case study. *Proceedings of the 11th International Conference on Sport Sciences Research and Technology Support*, 101–111. https://doi.org/10.5220/0012159100003587

Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, *37*(2), 253–264. https://doi.org/10.1080/02664760802684177

Comprehensive R Archive Network (CRAN). (2025, March 10). Rstan: R interface to Stan. The Comprehensive R Archive Network. https://cran.r-project.org/web/packages/rstan/index.html

Historical NFL playoff point spreads. SportsOddsHistory.com | Archived futures lines of the Super Bowl, World Series & more. (2021, January 10). https://www.sportsoddshistory.com/nfl-playoffs/?o=s&Team=&fv=&hv=&fd=&rd=&chv=&cfd=&ou=

Hoffman, M. D., & Gelman, A. (2011, November 18). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. arXiv.org. https://arxiv.org/abs/1111.4246

Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal*, *35*(4), 917–926. https://doi.org/10.1002/j.1538-7305.1956.tb03809.x

Mack, A. (2024). *Bayesian sports models in R: A Bayesian primer*.

Maher, M. J. (1982b). Modelling association football scores. *Statistica Neerlandica*, *36*(3), 109–118. https://doi.org/10.1111/j.1467-9574.1982.tb00782.x

Miller, E., & Davidow, M. (2019). *The logic of sports betting*. Ed Miller.

NFLREADR • Download NFLVERSE data. Dev status. (n.d.). https://nflreadr.nflverse.com/

RStudio desktop. Posit. https://posit.co/download/rstudio-desktop/

APPENDIX

Code: https://github.com/brandonowens24/NFL-Count-Modelling