

EXTENDED KEY-NUMBER BETTING TOOLKIT: TEASERS, ALT LINES, ML PIPELINES, AND BAYESIAN LAYERS

RICHARD OLDHAM

ABSTRACT. We extend the discrete-margin football modeling framework by surveying betting strategies that depend on key-number edges. Building on calibrated margin distributions, we contrast teaser pricing, alt-spread parlays, correlated markets, and live middling tactics, then layer in machine-learning and Bayesian approaches for forecasting and uncertainty quantification. The result is a deployable workflow for identifying positive expected value opportunities across NFL and NCAA markets.

1. INTRODUCTION

In the first part of this work, we developed a discrete-margin modeling framework for football scores, highlighting the outsized impact of key numbers (particularly 3 and 7 points in NFL games) on the distribution of final score margins. American football scoring is modular (3-point field goals and 7-point touchdowns), causing certain margins of victory to occur far more often than others. Indeed, about 15% of NFL games end with a 3-point margin and about 9% with a 7-point margin, making these by far the most common outcomes. These “key numbers” occur roughly 10% (for three) and 5% (for seven) more frequently than the next most common one-score margins. In contrast, other margins (e.g. 2, 5, 8 points) are significantly less frequent. In college football (NCAA), scoring is higher and more variable; the top two key margins (3 and 7) together occur only 16.5% of the time, compared to 24.3% in the NFL. The five most common margins in college account for only 28.5% of games, versus 42% in the NFL, reflecting greater parity and scoring variability in the collegiate game. Our discrete margin model explicitly incorporates these empirical frequencies, ensuring that probabilities for spreads and alternative lines are “bumped” at the key numbers to reflect reality.

One immediate application of a well-calibrated margin distribution is in evaluating NFL teaser bets. *Wong teasers*, named after gambling author Stanford Wong, refer to teaser bets that capture the value of those key numbers. In a standard 6-point teaser, a favorite of -7.5 is teased down through 7 and 3 to -1.5 , or an underdog of $+1.5$ is teased up through 3 and 7 to $+7.5$. Wong (2001) observed that such 6-point, 2-team teasers were historically profitable at the typical odds offered (often -110 or even odds). The reason is that crossing 3 and 7 captures a disproportionate increase in win probability, given how often games land on those numbers. However, sportsbooks eventually adjusted: whereas two-team, 6-point teasers were once offered at even money or -110 juice, many books now charge around -130 for the same teaser. This higher price undermines the bettor’s edge. At -130 for a 2-team teaser, the break-even combined win probability is about 56.5% (75.2% per leg) compared to 52.4% (72.4% per leg) at -110 , essentially erasing the historical edge which hovered around 75–76% win rate. In today’s market, beating teasers requires both a sharper model of game outcomes and a more strategic selection of lines and books. In the following sections, we present an expanded toolkit built on our discrete-margin engine: we explore alternative ways to exploit the discrete scoring increments (beyond the classic Wong teaser), describe machine-learning pipelines to better estimate win probabilities for any line, and incorporate Bayesian hierarchical models to improve predictions and quantify uncertainty. We then outline an end-to-end workflow for pricing and betting these edges, and discuss which strategies perform best in various contexts.

2. ALTERNATIVE STRATEGIES LEVERAGING KEY-NUMBER EDGES

The traditional Wong teaser is just one way to capitalize on the irregular distribution of NFL score margins. Here we detail several alternative betting strategies that similarly exploit “price-for-probability” inefficiencies created by discrete scoring values. All these approaches assume we have a reliable model of the game’s margin-of-victory probability mass function (pmf) – such as the one developed in Part I – to evaluate the true chance of different bets winning. Each strategy effectively repackages the same underlying edge (the extra probability of covering when crossing key numbers) in a different betting format, often to take advantage of better pricing or looser odds in related markets.

2.1. Dynamic Teaser Pricing and Shopping. Sportsbooks have introduced variable pricing for teasers, adjusting payouts based on the number of legs, points teased, and even the specific game or spread. The classic two-team, 6-point teaser at -110 is no longer universal; many books offer worse odds (e.g. -130) for the same teaser, especially when the teased lines cross 3 and 7. Our model allows us to compute a *fair value* for any teaser leg: by integrating the discrete margin pmf from the perspective of the teased line, we obtain the probability p_i that leg i wins. For example, if a favorite is -8.5 and we tease to -2.5 , $p_i = P(\text{favorite wins by } 3+)$. Given two independent legs, the fair parlay odds for a teaser are $\frac{1}{p_1 p_2} - 1$ (in decimal format). We compare this fair price to the book’s payout. If the book’s implied probability (after juice) is higher than $p_1 p_2$, the teaser is -EV; if lower, we have an edge. In practice, books differ: some have fixed teaser pricing by table, while others dynamically alter payouts per leg. By shopping around and cherry-picking legs at books that underprice certain lines, one can still find profitable teasers. For instance, one book might offer -120 for a 2-team teaser while another is -130 ; or a certain underdog might be mispriced in a teaser combo at one book. The strategy here is to use our model to compute p_i for each potential leg and identify where the product $\prod p_i$ is significantly higher than the implied break-even of the offered odds. It may also involve mixing legs from different sportsbooks (if possible) to maximize payout. The key improvement over naive Wong betting is being price-sensitive and dynamic: only play teasers when you’re getting sufficient value for each leg and the overall parlay. Contemporary advice suggests holding out for -120 or better on two-team, 6-point NFL teasers, or alternatively constructing three- or four-team teasers which sometimes have more favorable payouts (e.g. $+160$ for a three-teaser). In short, a model-driven shopper can still capture edges by selectively playing teasers at books and prices where the old magic (crossing 3 and 7) hasn’t been fully priced out.

2.2. Alternative-Spread Parlays vs. Teasers. A teaser is essentially a bundle of alternate spreads sold at a fixed cost. For example, teasing a favorite from -8.5 to -2.5 is equivalent to parlaying that team’s alternate spread of -2.5 with another team’s alternate line, typically with a fixed payout. If a book’s teaser menu is tightly priced (high juice), one can often do better by directly parlaying two alternate lines. In our running example, instead of a 6-point teaser on a -8.5 favorite and a $+2.5$ underdog, we could consider betting the favorite’s alt spread -2.5 and the underdog’s alt spread $+8.5$ in a regular parlay. Some sportsbooks offer reasonably priced alternate spreads (often with near-zero hold on each alt line), which means the parlay might yield a higher payout than the fixed teaser odds. Using our discrete-margin curve, we can price each alternate line individually: for any team at spread s , our model gives $P(\text{cover alt } s')$ for all s' . We then compute the parlay probability and implied fair odds. We compare this to the standard teaser payout. We often find cases where the books “shade” the teaser odds (because they know the standard Wong legs are popular) but leave the alt spreads at market price. For instance, a two-leg 6-point teaser might pay -130 , but taking the equivalent alt spreads separately might pay equivalent to -120 or better. This happens especially when one leg is in a lower-total game where 6 points are very valuable, but the book’s generic teaser pricing doesn’t account for that. A methodology for evaluating this, described by Unabated and others, is to calculate the expected value of a teaser by comparing to the parlay of corresponding alt lines. If the parlay offers a higher return, you should prefer it over the teaser. In practice, we automate this comparison: for each candidate teaser leg, find the closest alt-spread price that

yields the same adjusted line, and then check parlay EV. Often, especially in high-total games or at books with aggressive teaser juice, the alt-line parlay dominates. The advantage of this strategy is flexibility: you can tease by non-integer amounts (e.g. effectively 5 or 7 point moves via alt lines) and you are not constrained to the book’s teaser structure. It also allows mixing sides and totals more freely if desired. In summary, always check if a teaser’s benefit can be replicated or exceeded by a direct alt-line parlay, and take whichever has the higher expected value.

2.3. Totals-Conscious Teaser Selection. The success of the original Wong teaser was strongly tied to low-total games. When the expected total points in a game are small (say in the low 40s or 30s), each point is relatively more valuable and the distribution of final scores is tighter around the mean. Fewer points scored implies it’s more likely the game outcome falls within a touchdown or field-goal margin. Thus, teasing through 3 and 7 in a low-total game yields a larger boost in win probability than in a shootout. Our model incorporates total points as a key feature (for instance, via the variance or by using separate historical distributions by total range). A practical heuristic is to rank prospective teaser legs by the game’s total: lower totals first. All else equal, a 6-point teaser leg in a game with a 37 point total is far more valuable than one in a game with a 51 point total. We also consider how many key numbers the teaser crosses: an ideal Wong teaser crosses both 3 and 7. Some variants cross one key number plus some secondary ones (e.g. teasing +2 to +8 crosses 3,7 and lands on 8). Our model quantifies the half-point value at each number, so we can compute exactly how much win probability is gained by moving the spread from +2.5 to +8.5, for instance. For selection, we can create a priority score for each leg: perhaps something like $(P_{\text{teaser win}} - P_{\text{no-teaser win}})$ or the expected value of including that leg in a teaser at baseline odds. This will naturally prioritize games with low totals and spreads in the Wong zone. Market data supports this: historically, Wong teasers perform best when game totals are below roughly 49, with optimal performance in the 30s and low 40s. Additionally, our model might flag if a particular “near-miss” spread is still worth teasing: for example, a favorite of -9.5 down to -3.5 doesn’t cross 3, but does cross 7 and 6; if the total is extremely low, this might still be +EV at the right price. In essence, this strategy is about *selective* teasing – only using it when the conditions are ideal (low totals, multiple key numbers crossed) – rather than blanket teaser betting. It aligns with the original Wong logic but refines it with a quantitative model and acknowledges that with higher scoring environments (the NFL average total in recent years has hovered in the mid-40s), one must be even more selective or adjust the approach.

2.4. “Reverse Teasers” (Pleasers) for Off-Market Tails. Just as moving the line in your favor (teasing) can yield value when the market underestimates the probability of covering key numbers, moving the line *against* you for a larger payout can be lucrative if the market underprices extreme outcomes. A *pleaser* is a bet where you lay extra points (or take fewer points) in exchange for a big payoff (essentially the opposite of a teaser). For instance, a two-team 6-point pleaser might take a -7 favorite to -13 and pay out at +600 or higher if both cover. Pleasers are generally considered “sucker bets” because they lose much more often than they win. However, our discrete margin model can identify situations where a team is very unlikely to be in the middle of the distribution. Imagine a game with a heavy underdog: our model might show that either the favorite covers easily (wins by 14+), or the underdog could win outright, but it’s relatively unlikely for the favorite to win by only a small margin. Perhaps the underdog’s defense is weak enough that if they lose, they lose big, and if they don’t lose big, it means they were competitive enough to possibly win. In such cases, the probability of the favorite winning by more than the standard spread might be higher than implied by the pleaser odds. As an example, if a +12 underdog has a significant chance to either win or get blown out, a pleaser that takes the favorite from -12 to -18 might capture value. Our model will tell us $P(\text{favorite wins by } 18+)$ directly from the margin pmf. If the book is paying, say, +300 for that 6-point pleaser (which implies only a 25% break-even probability for each leg in a two-team pleaser), and our model says each leg actually has a 28

2.5. Exotic Key-Number Angles: Derivatives and Correlations. Because the discrete-margin model gives us a full distribution of outcomes, we can seek value in various derivative markets that hinge on similar principles:

- **First/Second Half and Quarter Lines:** Key numbers are mostly discussed for full-game spreads, but similar logic applies on a smaller scale. For instance, a halftime score is roughly half a game, so the key numbers for a half might be 3 and 4 (since 7 points usually requires two scores in a half). Our model can be adapted (or a separate model trained) for half-time margins, capturing mini key values. Books may not price half-point moves in these markets as sharply, so buying a half point around 3 in a first-half line could be underpriced. Additionally, teasing within a single game (e.g., a first-half and second-half line together) could exploit if a team often starts slow and finishes strong or vice versa.
- **Live Betting and Middles:** In-game odds update continuously, but often with simple models that assume scoring probabilities remain roughly constant. If a game's total drops significantly (say due to a slow first quarter or bad weather), the distribution of final margins tightens – which enhances the probability mass around 3 and 7. A bettor watching live could exploit this by “middling” the game: for example, if you took a favorite -7 pre-game and the game starts low-scoring, the live line might move to -3. You could take the other side at +3 live. Now you've created a middle where both bets win if the favorite wins by exactly 4–6 points, but more importantly, you've locked in that one of the most common margins (3 or 7) will at worst push one leg. Using our model's conditional updates (it can re-calc the margin distribution given a new total or time remaining), we can identify these middle opportunities when key numbers become especially likely given the game state. Another example: if a favorite was -10 and goes up by a field goal early (so live line maybe -13) but the total has plunged, the probability of exactly 10 or 13 as final margin might jump; one could take +13 live on the dog and hope for a 10 or 13 point win.
- **Same-Game Parlays (SGPs) as Synthetic Teasers:** Many books now heavily promote SGPs – parlays of correlated outcomes in the same game. They often restrict obvious positive correlations or adjust payouts to account for them (e.g. if you parlay a favorite and the under, which are correlated, you get a worse payout than if they were independent). However, these engines can be inconsistent. Our model can jointly simulate side and total, giving us the true probability of, say, (Team A +7.5 and Under 45.5) both hitting – which is essentially capturing a scenario like a classic teaser (dog +7.5) plus an under. If the sportsbook's SGP algorithm misprices this correlation (and many casual bettors do, often overestimating or underestimating how outcomes relate), we can gain an edge. For example, a heavy underdog is more likely to cover +7.5 if the game is low-scoring (since fewer points makes it harder to blow them out). The book might naively boost the parlay payout as if side and total are independent, or not fully account for the correlation. By pricing the joint distribution with our model, we know the true fair odds. If the SGP payout is higher (which occasionally it is, due to promotional boosts or simply model error by the book), we can pounce. In essence, this recreates a teaser-like bet (getting extra points on the dog and a correlated lower total) but possibly at a payout the book would never offer if it realized how correlated they are. Caution: some books heavily reduce payouts on SGPs exactly because they fear skilled bettors exploiting correlation. But even in those cases, they might overcorrect on some combos and undercorrect on others, so a meticulous comparison using our simulation model can highlight +EV SGP components.

The common theme in all these derivative strategies is leveraging our understanding of the distribution's shape – whenever the book's pricing (or a bettor's intuition) diverges from the true probabilities given by our model, we find an opportunity. Key numbers remain central in these considerations: e.g., knowing that a 3-point lead at halftime is far more common than a 2-

or 4-point lead, or knowing that a 7-point final margin is far likelier in a 40-point total game than a 54-point game. By quantifying these, we monetize them in any market available.

2.6. Adjustments for College Football. Our discussion so far has been NFL-centric, because the NFL’s scoring distribution shows clear, stable key numbers over decades. College football (NCAA FBS) presents a more chaotic landscape: higher scoring games, larger disparities between teams, and more diverse styles of play (from triple-option service academies to Air Raid offenses). As a result, the margin distribution in college is flatter – the classic key numbers of 3 and 7 are less dominant. In fact, as noted, only 16.5% of FBS games end on 3 or 7 combined, versus 24.3% in the NFL. The most common margins in college include some larger numbers (10, 14, 17, 24) reflecting how blowouts and unbalanced scores happen frequently. Our discrete-margin model must therefore use college-specific frequency adjustments. We cannot simply port NFL key number weights onto NCAA games. For example, the probability of a game ending exactly 3 is roughly 8–9% in college (compared to 14–15% NFL). Seven is around 7.5% (vs 10% NFL). Meanwhile, margins like 17 or 24 (three scores) occur more often in college due to mismatches. The implication for teasers and alt-lines: the traditional Wong teaser (capturing 3 and 7) is far less effective in college; the odds boost you get from +1.5 to +7.5 is smaller. Some sportsbooks don’t even allow college teasers or price them prohibitively, which is telling. A bettor focusing on college should instead look at *alternative lines around the key ranges that **do** matter* – for instance, if a team is favored by 28, a teaser of -28 to -22 (crossing 24 and 21) might be analogous to an NFL teaser crossing 7 and 3, but one must verify the frequencies. Our model can learn the college distribution: perhaps 24 and 21 are moderately common margins (21 due to three touchdowns, 24 due to three TDs + FG). If so, teasing through those could be useful. However, because college games have higher variance, a more profitable angle is often to sell points (take an alt line) when you have a strong read. For example, instead of teasing a -14 favorite down to -8 (crossing 14 and 10), it might be better to take the favorite -20 at a big plus payout if you believe the talent gap is huge. The model will indicate that the probability of a 20+ point win might be higher than the market-implied probability. In essence, in college, *variance is your ally* if you have good predictions. Also, correlation plays are big in college – e.g., if you expect a blowout, the over might be correlated (as one team might score 50 by itself), whereas in NFL a blowout often correlates with under (the losing team gives up). Our toolkit’s simulation component (Section 3) can be tailored to college by adjusting team strength distributions and total points. Conference-specific quirks can be modeled with hierarchical priors (Section 4). The bottom line: college football betting should deemphasize classic Wong teasers and instead leverage the model for alt spreads and totals that align with each game’s likely script. Many sharp college bettors avoid teasers entirely and focus on straight bets and alt-lines, unless they find a rogue line that crosses a college key number cheaply. Our model supports this by outputting a half-point value table for college (similar to NFL) – one might find that, say, buying from -6.5 to -6 (to avoid the TD) is hardly worth 10 cents in NCAA, whereas buying off -3 might still be somewhat valuable, but buying off -7 is less so than in NFL. These nuances are quantitatively backed by our data-driven frequencies.

Literature and Market Context. Before moving on, we note a few references and tools that informed these strategies. The concept of key numbers and basic strategy teasers was first popularized by Stanford Wong in *Sharp Sports Betting*. Numerous articles (e.g. Covers.com guides) have documented the prevalence of 3- and 7-point margins and how rule changes (like the NFL moving the extra point in 2015) have slightly shifted those frequencies. Our model’s key-number weighting is grounded in such empirical studies. Unabated’s NFL Teaser Calculator and educational content illustrate how to compute teaser EV and emphasize the importance of shopping for better teaser odds. In the academic realm, the general idea of exploiting pricing inefficiencies due to discretization connects to research on *middles* and *arbitrage* in sports betting; while we focus on positive EV betting rather than risk-free arbitrage, the principle of value from distribution tails is similar. By extending these ideas with a modern modeling toolkit, we can adapt as sportsbooks become more efficient.

3. MACHINE-LEARNING PIPELINES FOR WIN PROBABILITY ESTIMATION

Accurate pricing of spreads, alt-lines, teasers, and correlated parlays all starts with a robust model for game outcomes. In Part I, we described building a discrete margin-of-victory distribution using historical data and key-number adjustments. We now outline two complementary machine-learning approaches to generate those probabilities, plus a simulation-based method for capturing game dynamics. These pipelines are designed for both NFL and NCAA (with appropriate modifications as noted). The goal is to produce calibrated win probabilities for any bet (spread, total, teaser leg, etc.), which we can then plug into the betting formulas above.

3.1. Ratings Model with Discrete Margin Output. One approach is a two-stage ML model. First, we predict the expected margin of victory (and possibly total points) using a host of features; second, we transform that prediction into a full discrete probability distribution over possible margins.

For the first stage, a classic approach is to use team *ratings* as features. This could be a simple Elo rating for each team, or offense and defense ratings akin to Bradley–Terry or Massey models. We might have $R_{\text{team}} = \text{offense}_{\text{team}} - \text{defense}_{\text{opponent}} + H$ (if home) as a baseline predictor. More sophisticated is to train an ML regression model on historical game outcomes. Features can include:

- Team strength metrics: win rates, point differentials, drive success rates, etc., possibly decayed for recency.
- Injuries and roster changes: if a star QB is out, that’s a significant feature. Weather, travel distance, rest days, and other situational factors can be encoded.
- Market information: the closing spread and total (implying the market consensus for expected margin and points) can be included as features, as well as how they moved from open (to capture sharp action).
- Matchup variables: e.g. Team A’s passing offense vs Team B’s pass defense ranking, coaching tendencies on 4th down (which affect variance), etc.

A gradient boosted tree model (XGBoost, LightGBM) is suitable for handling such mixed inputs. It can output a prediction for the mean margin $\hat{\mu}$ and possibly an uncertainty $\hat{\sigma}$. Alternatively, one could train a model to predict win probability directly for various spreads (classification), but it’s easier to start with margin as a continuous outcome.

Once we have a predicted mean margin and perhaps a predicted total (or separate model for total points), we need to derive a full distribution of the margin. Historically, many have assumed a normal distribution of scoring margin around the spread. A normal with mean μ and variance σ^2 can give approximate win probabilities for a given spread. Hal Stern (1991) showed that this approach is a decent first approximation for NFL outcomes. However, the normal fails to capture the spikes at certain integers. We improve this by *discrete calibration*: we start with a continuous distribution (like Normal or perhaps a skewed distribution if the favorite is very strong), then adjust the probabilities at key margins.

One technique is to use a flexible discrete distribution such as a Skellam distribution. A Skellam arises as the difference of two independent Poisson scores (often used in soccer). But NFL scoring is not exactly Poisson – it’s a mixture of 3s and 7s. Instead, we might:

- Fit a baseline distribution: e.g., take $\hat{\mu}$ and assume margin $\sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$. This gives a smooth density $f(m)$ over real m .
- Convert that to a discrete pmf: $P(M = k) = \int_{k-0.5}^{k+0.5} f(m) dm$ for each integer k . Now we have an initial estimate for each margin.
- Apply multiplicative adjustments to key numbers: We know from data that $P(M = 3)$ is higher than the normal would predict. So we can compute the ratio $r_3 = \frac{\text{empirical freq of 3 given similar spreads/totals}}{\text{model predicted freq}}$, and similarly r_7 for 7, etc. We then set $P^*(M = 3) = r_3 \cdot P(M = 3)$, and renormalize the pmf (taking a little probability mass away from other outcomes to compensate).

We can learn these adjustment factors r_k from historical data by stratifying games by total and spread. For instance, in low-total NFL games, the mass on 3 and 7 might be even more

pronounced than average; in high-total games, slightly less so (because more points means more chance to move off those margins). We therefore can have a small lookup table for half-point values: e.g., in a 37-point total game, $P(M = 3)$ might be say 12% instead of 9%, etc. Our model can incorporate that by making r_3 a function of the total.

An alternative approach is to train a categorical classifier directly for $P(M = k)$ over a range of k . For example, classify the game outcome into bins (like -30 to $+30$ margin). This is a multi-class prediction problem which something like a neural network or gradient boosting could handle. However, class imbalance and the need for smoothing make it tricky. A compromise is to predict a few key probabilities (like the probability to cover -2.5 , cover -3.5 , etc.) using binary classifiers, and derive the pmf from those. But the two-stage (regression then adjustment) approach is more transparent and ensures consistency.

Finally, we calibrate the distribution. This means checking that, for example, when our model says a team has a 75% chance to cover -2.5 , it really happens 75

The output of this pipeline is a fully specified pmf for each game: $P(M = k)$ for say $k = -40, -39, \dots, +40$ (practically, games rarely exceed 40-point blowouts). With this pmf, we can price any bet:

- Point spread cover probability at spread s : $\sum_{k \geq s} P(M = k)$ (if favorite's margin M).
- Moneyline: $P(M > 0)$ for favorite win (or include tie for overtime rules).
- Teaser leg: e.g. favorite -8.5 teased to -2.5 is $P(M \geq 3)$ under our distribution.
- Alt spreads: $P(M \geq x)$ for any x can be read off.
- Totals: we can also produce a distribution for total points; our model might separately predict expected total λ and apply a similar discretization (though total points have "key numbers" too, they are softer as noted).
- Props related to margin: e.g. the probability of an "exact margin" bet (some books offer "win by exactly 3" at long odds) can be directly taken from $P(M = 3)$.

Notably, this model can highlight how the value of half-point changes depending on context. We can create a table of *half-point values*: for each possible point spread and total, how much does the win probability change if the spread moves by 0.5. This is extremely useful for bettors deciding whether buying a half-point is worth it, or comparing two lines (one book has -7 (-110), another has -7.5 ($+100$), etc.). Our table would confirm, for instance, that buying off -3 or -7 in the NFL is worth significantly more than buying off -4 or -8 ; whereas in NCAA, buying off -3 is worth somewhat less (since 3 is not as frequent) and buying off -7 even less so, often not worth the cost.

In summary, the ratings+pmf pipeline combines the power of machine learning for prediction with domain-based post-processing to ensure the discrete output matches known football realities. This forms the backbone for quickly evaluating any standard wager.

3.2. Drive-Based Monte Carlo Simulation. An alternative and illuminating approach is to simulate the game play-by-play (or drive-by-drive) to derive the distribution of outcomes. Monte Carlo simulation can incorporate the actual mechanics of football scoring, which naturally reproduces key number frequencies without having to enforce them post hoc.

In a drive-based simulator, we model each possession of the game as a sequence of events: starting field position, series of plays, result of drive (touchdown, field goal, turnover, punt, etc.), time elapsed, and then repeat for the other team. By assigning probabilities to each event based on team strengths, we can simulate thousands of whole games to see the distribution of final scores.

For example, we might parameterize: - Team A's probability of scoring a touchdown on a given drive vs Team B's defense. - Likelihood of field goal attempt (and success) if a drive stalls in field goal range. - Probability of turnover (which ends the drive with zero points and gives the opponent possession). - Time consumed by each drive (a quick three-and-out vs a long drive). - Special situations: end-of-half sequences, overtime if tied.

We utilize team-specific metrics: drive success rate (percentage of drives resulting in scores), turnover rate, average starting field position (which could incorporate special teams). These can be derived from historical data or an advanced metrics model like expected points. We

can also calibrate against the betting total and spread: for instance, ensure the average points in simulation matches the total, and the fraction of simulations Team A wins by more than X matches the implied probability from the spread (this can be achieved by adjusting team strength inputs).

One big advantage of simulation is the natural emergence of key numbers. In an NFL simulation, a team often scores 7 or 3 points on a successful drive, so combinations of those scores will produce final margins of 3, 7, 10, 14, etc., more frequently. If we simulate enough games, the histogram of margins will show spikes at those numbers without us manually imposing them. We might observe, say, 9–10

This simulator also allows us to model conditional scenarios, which is useful for live betting or derivative markets. For instance, we can simulate just the second half of a game given a halftime score to update the distribution of final margins. Or simulate a game state where one team leads by 14 in the 3rd quarter to see how often the final lands on 7 versus a larger blowout.

To implement such a Monte Carlo, one could use a state-based approach (a Markov chain for football drives, as in football analytics literature). Define states like “1st and 10 at own 25” and transition probabilities (e.g., probability of gaining a first down, scoring, turning over, etc.). Work by [Lock & Nettleton \(2014\)](#) and others in *Journal of Quantitative Analysis in Sports* have used Markov models to simulate football drives realistically. By fitting these probabilities to each team (Team A offense vs Team B defense), we capture team styles – a slow, run-heavy team will have longer drives with fewer total possessions (tending to lower totals and tighter margins), whereas an up-tempo passing team will create more scoring opportunities (possibly more variance in outcomes).

The output of the simulation is again a distribution of margins, which we can treat just like the pmf from the previous section. We should confirm that the simulation’s average outcomes line up with expected spread/total inputs (if not, adjust the simulation parameters or scale team strengths). One can also glean interesting insights: e.g., if the simulation consistently shows a higher probability of blowouts than the normal distribution model, it might indicate that tail outcomes are fatter-tailed (perhaps due to turnover cascades or teams giving up when down big). This could justify pleaser bets or alternate spread plays as discussed.

While simulation is computationally heavier than a direct model, it has the benefit of being very flexible. Want to know the chance a team wins by exactly 5? Just count it in the sim outcomes. Want to incorporate a specific matchup factor (Team A has a bad kicker, lowering FG success rates)? Just tweak that in the sim. It’s also easier to extend to Same-Game Parlay scenarios: we can simulate not just final scores but also player stats, drive outcomes, etc., to evaluate correlation between, say, a team covering -6.5 and the total going under 48.

We include this method in our toolkit both to validate the ML model’s outputs (they should roughly agree if both are calibrated to the same inputs) and to power those more complex bets that a simple model struggles with (like the joint distribution of side and total, or the time dynamics for live betting).

3.3. Correlation Engine for Same-Game Parlays. The third component of our modeling toolkit is a specialized extension of the simulation: a joint outcome generator for a game, which produces not just the final score but other correlated events. This is motivated by the rise of Same-Game Parlays (SGPs), where bettors can combine picks like “Team A +7.5, and Under 45.5, and Quarterback X under 2.5 touchdowns” in one bet. The challenge is that these events are not independent, and sportsbooks use crude correlation adjustments that sometimes misprice the combination.

Our SGP correlation engine builds on the drive simulator. In each simulated game, we can track additional quantities: - Did the game go over or under a certain total? - Player statistics, if we incorporate a player performance model (e.g., yards gained per drive can feed into a QB’s yardage total). - Intermediate outcomes like halftime score (for a halftime bet), or whether a team covered various alternate spreads.

By recording these in each simulation run, we can estimate probabilities like $\Pr(\text{cover} + 7.5, \text{total} < 45.5)$. Suppose our simulation (using the appropriate team strengths and perhaps weather adjustments) finds this probability to be 0.30 (30

From an implementation perspective, building a full SGP model is complex. It may involve simulating play-by-play (to get player stats) or at least drive-by-drive with some player stat attribution. There is work in the sports analytics space on joint distributions of scores and yardages (often via Gaussian copulas or bivariate Poisson models). For our purposes, we focus on team-level outcomes: side and total correlation. We find that our engine can produce realistic joint distributions of (spread outcome, total outcome). For instance, it might show that the correlation between a favorite covering and the game going over is +0.3 (meaning favorites cover more in high-scoring games), or an underdog cover correlates with under, etc., depending on teams. We can then input these into analytical formulas or just use simulation frequencies directly to price parlays.

In summary, the ML pipelines and simulations work in tandem: the ML model (with discrete adjustments) is fast and good for most pricing tasks (straight bets, simple parlays, teaser legs), while the simulation and correlation engine handle more intricate situations (dynamic/live betting and multi-outcome parlays). Both are calibrated to align with observed key number frequencies and can be updated as the sport evolves (for instance, if a rules change made 2-point conversions more common, leading to more 1- or 2-point margins, our model and sim would adjust accordingly in their key number weighting).

4. BAYESIAN HIERARCHICAL MODELING FOR TEAM STRENGTHS

So far, our modeling approach has been largely frequentist or algorithmic (machine learning and simulation). We now introduce a Bayesian hierarchical model as a powerful extension for our predictions. The Bayesian approach is particularly useful for:

- (1) Incorporating prior information and partial pooling across teams and seasons.
- (2) Quantifying uncertainty in our predictions (e.g., we get a full posterior distribution of win probabilities, not just a point estimate).
- (3) Updating our model in a principled way as new data (games) come in, which is valuable for sequential betting through a season.

A typical Bayesian hierarchical model for football scores might look like the following. We model the points scored by each team in a game as random variables depending on team-specific parameters:

$$\text{Points}_A \sim f(\text{attack}_A, \text{defense}_B, \dots),$$

$$\text{Points}_B \sim f(\text{attack}_B, \text{defense}_A, \dots),$$

where attack_A is team A's offensive strength, defense_B is team B's defensive strength. We include home-field advantage parameter h if applicable. The function f could be something like Poisson (leading to a bivariate Poisson or Skellam model for the score difference), or a normal with mean difference $\text{attack}_A - \text{defense}_B + h$, or even a custom likelihood that directly targets the score difference.

The hierarchical part comes from treating each team's parameters as random effects. For example, we might say:

$$\text{attack}_i \sim \mathcal{N}(\mu_{\text{attack}, \text{conf}(i)}, \sigma_{\text{attack}}^2),$$

$$\text{defense}_i \sim \mathcal{N}(\mu_{\text{defense}, \text{conf}(i)}, \sigma_{\text{defense}}^2),$$

meaning that team i 's offense and defense ratings are drawn from a distribution specific to its conference (in college) or just a league-wide distribution (in NFL). This *partial pooling* shrinks extreme estimates toward the group mean, which is crucial when we have limited data on a team (e.g., early season or smaller conferences). In the NFL, we might not need a conference layer (since there are only 32 teams and plenty of games), but in college, grouping by conference improves estimates (SEC teams collectively might have a different baseline than Sun Belt teams, for instance).

We also allow these strengths to vary over time (dynamic modeling). A common approach is a random walk:

$$\text{attack}_{i,t} = \text{attack}_{i,t-1} + \epsilon_{i,t},$$

$$\text{defense}_{i,t} = \text{defense}_{i,t-1} + \delta_{i,t},$$

with $\epsilon_{i,t}, \delta_{i,t}$ as zero-mean shocks each week. This means we believe team strengths evolve gradually (accounting for injuries, form, etc.). Bayesian dynamic models like this have been successfully applied in soccer and football predictions. Recent work by [Macri-Demartino et al. \(2025\)](#) introduces an adaptive weighting in such models, allowing certain periods (e.g., recent games) to influence the estimate more when a team undergoes drastic changes, while borrowing strength from a longer history when the team is stable. Adopting such an approach would let our model quickly adjust if, say, a star quarterback gets injured – the offensive strength might drop significantly with a larger uncertainty, and over a few games stabilize to a new level.

The Bayesian model yields a posterior distribution for each team’s offense/defense rating and for each game’s outcome. We can simulate from the posterior to get a predictive distribution of score margins for each matchup. Essentially, this provides another way to get the game outcome pmf, but now with full accounting of parameter uncertainty. For example, if two middling teams play, we might be very uncertain about who is better – the posterior predictive might be very spread out, perhaps giving a high probability of a close game. If an elite team plays a weak team, the posterior predictive will be more skewed toward a blowout, but with an uncertainty that might fatten the tails a bit compared to a fixed-effects model.

One valuable output of the Bayesian approach is the ability to simulate season outcomes and measure risk. For betting, this means we can estimate, say, the distribution of our betting portfolio’s returns given the uncertainties. For instance, when we identify a +EV teaser, that EV is usually computed using point estimates of win probability. A Bayesian would instead integrate over the uncertainty in those win probabilities, giving us a distribution of possible outcomes. This could inform how we size our bets (Kelly criterion under parameter uncertainty, etc.).

Implementing the Bayesian model can be done in Stan or PyMC3/PyMC. We would write a likelihood for game scores or margins. One could use a Skellam likelihood for score differential, or even directly model the categorical outcome of margin if one wanted to be very discrete – e.g., use a multinomial likelihood with logits that have a linear component for the expected score and some fixed effects for the key numbers. That latter approach is complex, but one could, for example, incorporate a parameter for “extra tendency to land on 3” which affects the likelihood of margin=3 outcomes. However, a simpler approach is to model scores (or goal counts) and let the discreteness emerge naturally.

We also include prior distributions on all parameters. Priors can encode common-sense: e.g., the prior for home-field advantage might be Normal(0, 3) (meaning 0 ± 3 points, since we know HFA is around 2.5 points historically). Priors for attack/defense might center around 0 with a certain variance that we infer (the model will learn how dispersed team strengths typically are – in NFL it might be 10 points standard deviation between best and worst, in college maybe much larger).

One must check model fit – classic references like [Stern \(1991\)](#) and [Baio & Blangiardo \(2010\)](#) validate their models by how well they predict win/loss and scores. We would do the same, comparing our Bayesian model’s predictive accuracy to the ML model and to betting market spreads. The expectation is that the Bayesian model, by pooling data, will do better early in the season or for teams with sparse data, and will give well-calibrated probabilities with honest error bars.

To illustrate the benefit: imagine in Week 1 of the NFL season, our ML model might not know that Team X is now much stronger (say they got a new superstar quarterback) except via whatever features we input. A Bayesian model could incorporate a prior that teams regress to the mean from last year plus some offseason adjustments. After a couple of games, the model will update Team X’s offensive rating upward if they score a lot, but with a bit of conservatism (partial pooling) to avoid overreaction. This kind of dynamic updating is exactly what a sharp

bettor does mentally, and our model does it automatically. It can flag edges like “the market hasn’t caught up to Team X’s new strength, so their spreads are still a point or two too low”.

Finally, the Bayesian framework is naturally suited for decision-making. Because we get a full distribution of the teaser win probability, for example, we can compute not just an EV but a probability that the teaser is positive EV (given parameter uncertainty). If we want to be risk-averse, we might only play bets that are >90% likely to be +EV by our posterior. Or we might adjust bet sizing down on ones that are +EV but with large uncertainty. This integrates with portfolio management and the concept of “edge distribution” rather than point estimate.

In terms of references, [Baio & Blangiardo \(2010\)](#) provide a foundation for hierarchical models in football (soccer) which we have adapted to American football; [Macrì-Demartino *et al.* \(2025\)](#) (an arXiv preprint) demonstrates the cutting edge of dynamic Bayesian updating for team strengths, albeit in the context of soccer leagues. These show improved prediction error and better probabilistic calibration, which in a betting context translates to more reliable edges.

5. INTEGRATING THE TOOLKIT: WORKFLOW AND USE CASES

Having outlined the menu of strategies (teasers, alt-lines, etc.) and the modeling approaches to value them, we now describe an end-to-end workflow for deploying this toolkit in a betting context. We also discuss how different situations call for different tools from the toolkit, i.e. when to use which strategy.

5.1. End-to-End Modeling and Betting Workflow. 1. Data and Feature Engineering:

We continuously gather data: game results with margins and totals, closing lines from sportsbooks, play-by-play data for simulation, and any relevant covariates (injury reports, weather forecasts, etc.). For college, we also collect conference-specific stats. This data is fed into our models. Prior to each week’s games, we update our team ratings (if using the Bayesian model, this means running the posterior update with last week’s games; if using ML, retrain or at least update recent form metrics). We also update key number frequency tables each season – e.g., if we observe an uptick in 2-point conversion attempts one year, it might slightly raise the frequency of 1- and 2-point margins, which we’d incorporate.

2. Model Prediction Stage: Using the ratings/ML model, we generate a predicted distribution for each upcoming game. Simultaneously, we can run a handful of Monte Carlo simulations for each game to sanity-check those distributions (especially for games with unusual conditions). We output for each game: probability of covering every relevant spread (say from -20 to +20 around the line), probability of every alt total in a range, joint distributions as needed (for side/total). These predictions are stored in a structured way so we can query them for any bet.

3. Line Shopping and Value Identification: We then pull the current odds from various books for spreads, totals, alt lines, teaser prices, etc. For each book and each game, we compare our predicted probability to the implied probability from the odds. For a straight bet, this is straightforward: if we think a team covers -6.5 with 55% probability and a book is offering -110 (implying 52.4

4. Bet Selection and Sizing: From the identified value bets, we decide which to play. This could involve filtering (maybe we avoid pleasers unless the edge is very large, due to their volatility) or grouping (we might not want to bet too many correlated things, like multiple different teasers involving the same team). We then size the bets according to a bankroll management scheme. A common approach is fractional Kelly betting based on our edge and variance. The nice aspect of having a full distribution via Bayesian model is we can simulate the distribution of outcomes for our portfolio to calculate things like probability of drawdown, etc., and adjust sizing to stay within risk tolerance. For example, teasers have a higher correlation between legs (especially if we reuse teams) which increases variance; our simulation can account for that.

5. Monitoring and Iteration: As games are played, we feed the results back into model updates. We also track how our bets perform to verify that realized results align with expectations. If we consistently see certain bet types doing worse than predicted, that could indicate a model misspecification (or just bad luck, which we’d assess with significance tests). We especially

monitor the frequency distribution of margins each season to catch any shifts – for instance, a few years ago the NFL moved the extra point distance, and suddenly missed PATs made 5-point and 6-point margins more common. Our system would detect an uptick in 6 and 8-point games and we’d adjust the key number weights accordingly. Keeping the model current is crucial since even minor rule changes (like the 2023 NFL kickoff rule) can have downstream effects on scoring.

This workflow runs weekly (or daily for live betting updates). It’s essentially an analytics-driven cycle: predict, find value, bet, learn, and update.

5.2. When Each Strategy Shines. Not all edges are present all the time, and each of the strategies from Section 2 has scenarios where it’s most effective. Here we summarize when to deploy each:

- **Classic Wong Teaser (2-team, 6-pt):** Best in NFL games with low totals (say < 45) and spreads in the Wong zone (-7.5 to -8.5 favorites or $+1.5$ to $+2.5$ dogs). Only worth playing if you can get close to -120 odds. Ideal when both legs cross 3 and 7 and your model shows each leg winning 75
- **Alt-Spread Parlay:** Use this when books have tightened teaser payouts too much. If your model shows, for example, a 6-pt teaser leg has 74
- **Correlation Parlays/SGPs:** When you have a strong read on game script (e.g., if an underdog covers, it will likely be a low-scoring game), and your model quantifies that correlation as higher than the implied SGP odds, this is the play. It’s also a way to squeeze value out of highly efficient primary markets by moving into second-order markets where books are less sophisticated. For example, late in the season, you might know a particular team with a backup quarterback will only win in a slow, grindy game – taking that team $+$ points and under the total is a correlated bet that often pays better as an SGP than as separate bets. Just be wary of books that severely limit SGP payouts; use the ones with more lenient pricing for these opportunities.
- **College Football Alt-Lines:** As discussed, teasers in college are usually not worth it. Instead, if you have an edge (maybe your model sees a mismatch bigger than the market does), consider alt spreads. For instance, instead of betting the favorite -14 at -110 , maybe -21 at $+200$ if your model thinks a $21+$ win is quite likely. This shines in games with high variance – which in college is a lot of games, especially non-conference mismatches. Also, if your model finds that certain margins (like exactly 14) are less sticky (maybe the team either wins by <10 or >20 most of the time), then selling points (laying more points for higher payout) is advantageous. Our toolkit’s half-point value table for college helps identify these: e.g., it might show that going from -14 to -17 (through 15,16) costs maybe 12
- **Live Betting and Middles:** Use during games when totals and spreads swing. If a heavy favorite falls behind early, live markets might overreact – your pre-game model (updated for new score and time remaining) might still say there’s a good chance of a close finish. Middling comes into play when a key number is between the two lines you have. This is common around 3 and 7. For example, bet Team A -7 pre-game, then if Team A leads and the line moves to -14 , bet the underdog $+14$. You now have a huge middle from 7 to 13. Your model’s distribution can calculate the exact probability of landing in that middle (which might be, say, 10–15
- **Pleasers (Reverse Teasers):** Extremely situational. They work best when you identify that a point spread might be very misaligned with true probabilities at the tails. For instance, if your model thinks a favorite has, say, a 50

Ultimately, our expanded toolkit allows us to approach the betting market with a much richer set of options. Instead of only hitting the straightforward spread or total, we can find value in the margins – both figuratively and literally – by leveraging the discrete nature of scoring. Each week, depending on what our model says and how the books are pricing things, we might be primarily playing a slate of teasers and alt parlays; another week, we might see no good teasers (perhaps books all moved to -130 and our edges disappeared) so we instead focus on live betting or SGPs. The key is flexibility and rigorous modeling.

6. CONCLUSION

We have presented a comprehensive framework that extends the classic “Wong teaser” concept into a versatile betting toolkit. By constructing a discrete margin-of-victory model that respects key numbers, and enhancing it with machine learning predictions, Monte Carlo simulations, and Bayesian hierarchical updates, we gain a powerful lens to evaluate all manner of football bets. This approach turns qualitative betting wisdom (e.g. “3 and 7 matter a lot”) into quantitative calculations (precise probabilities and EV). Our toolkit not only identifies +EV opportunities in traditional markets (spreads, totals, teasers) but also adapts to modern betting formats like same-game parlays and live betting, where mispricing can often be found by those armed with superior models.

This research bridges the gap between academic sports modeling and practical betting strategy. On one hand, we incorporated rigorous techniques from the literature – such as dynamic Bayesian team ratings (Macrì-Demartino *et al.*, 2025) and hierarchical models (Baio & Blangiardo, 2010) – to improve predictive accuracy. On the other hand, we addressed the pragmatic aspects of execution: shopping lines, dealing with sportsbook pricing quirks, and selecting bets to maximize return. The result is a blueprint for advantage play in an increasingly efficient market. As sportsbooks continue to evolve (e.g., more dynamic odds, new bet types), the modular nature of our toolkit means we can always plug in new models or consider new derivatives, but the core insight remains: understanding and exploiting the discrete scoring distribution is fundamental to beating the NFL and NCAA point spread markets.

REFERENCES

- Wong, S. (2001). *Sharp Sports Betting*. Pi Yee Press.
- Stern, H. S. (1991). On the Probability of Winning a Football Game. *The American Statistician*, 45(3), 179–183.
- Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, 37(2), 253–264.
- Macrì-Demartino, R., Egidi, L., & Torelli, N. (2025). Bayesian weighted discrete-time dynamic models for association football prediction. *arXiv preprint arXiv:2508.05891*.
- Unabated Sports (2025). *Betting the NFL: Are You Placing the Right Wagers?* (Jack Andrews, contributor). Unabated.com, Sep 2025.
- Covers.com (2025). Examining NFL Key Numbers for the 2025-26 Season. Covers Betting Guides, Aug 25, 2025.
- Campus2Canton (2025). College Football Betting 101 – Understanding the Fundamentals. Posted Sep 2025, available at campus2canton.com.
- Lock, D., & Nettleton, D. (2014). Using random forests to estimate win probability before each play of an NFL game. *Journal of Quantitative Analysis in Sports*, 10(2), 197–205.