When should coaches deviate most from the play that gives the most expected points?

Emily Arbieto, Maxwell Elim, Paul (Shibo) Zhao, Jason Chen

2025-05-06

Introduction

In football, coaches face critical decisions in the 4th down with 3 options: go for it, punt, or kick a field goal. Expected points models have been introduced to quantify the rewards of each decision. However, each decision doesn't boil down to just "which would give the most points" and coaches typically will deviate from this. This asks the question- when does it make sense to deviate from the play that gives the most expected points?

With these models, we can see the expected points of each play, though these models don't take into account risk or long term strategy. A go-for-it attempt might have a higher expected value but also a significant chance of failure and a major swing in momentum. In the real world, coaches often act conservatively — perhaps not because they're ignoring analytics, but because they're accounting for variance, score differential, field position, or time remaining. a failed conversion attempt can hand the ball to the opponent in a dangerous position, or a risky pass might be intercepted. Expected value doesn't tell the whole story. Coaches may make seemingly conservative decisions not because they are misinformed, but because they are risk-aware.

For example, in the 2015 Super Bowl XLIX, Patriots vs. Seahawks, we see a pass instead of a run at the 1 yard line. The Seahawks were down 28-24, where there were about 26 seconds and 1 timeout left in the 2nd down with 1 yard to go from the 1 yard line. The logic is as follows: although they had a superstar running back, Marshawn Lynch, who is actually known for short yardage runs, Coach Pete Carroll decided to pass because the options of plays that would result if the run had failed would greatly narrow. If the decision was to run, it would have the clock running and leave very little time for you to run again if it fails. In comparison, if you have an incomplete pass, the clock would stop and you could try running in the 3rd and

4th down.

Table 1: Decision Logic on 2nd Down in Super Bowl XLIX

Play.Type	Outcome	Clock	Timeout	Options.Left
2nd Down Run	Incomplete	Running	Must use	3rd/4th Down Passes
2nd Down Pass	Incomplete	Stops	Still have a timeout	3rd/4th Down Pass or Run
2nd Down Pass	Complete	Game	Not used	Win

Given the circumstances, It was not unreasonable for the coaches to choose to pass, though much of our expected points model doesn't capture risk, such as having a star running back- Marshawn Lynch. In this example, this gives a vivid time involving managing risk instead of maximizing averages. Ultimately, this is what we are working towards, to mimic what coaches actually do.

Methodology

Traditional Expected Points (EP) models evaluate play options by assigning a deterministic expected value based on field position and down. However, actual football decisions are probabilistic. Particularly on fourth down, Outcomes hinge on the success or failure of a risky play. To better reflect this uncertainty, our project extends the EP framework to incorporate probabilistic outcomes of each decision. We compute the expected value (EV) of a fourth-down decision using the equation:

$$\text{EV} = p \cdot \text{EP}_{\text{success}} + (1 - p) \cdot \text{EP}_{\text{fail}}$$

p is the estimated probability of converting the 4th down. EPsuccess is the expected points if the conversion succeeds. EP fail is the expected points for the opponent after a turnover on downs This EV is compared to the expected points from alternative plays, such as punting or attempting a field goal. To find the threshold success probability (p) where go-for-it becomes equivalent to the alternative: p = (EPalternative - EPfail) / EPsuccess - EPfail) This threshold helps us determine whether the risk of going for it is justified, given the alternative options. Our simulation uses this logic to decide whether a team should "go for it", punt, or attempt a field goal based on estimated probabilities and expected outcomes. This modification adds a layer of realism into our simulation and makes it a closer representation of real coaching strategies, where decision-making not only depends on what the "obvious" play is, but on risk and uncertainty. When comparing the 3 options, we decided to only compare 2 decisions at a time, i.e. Punts v.s. Go-For-It (Up until 60 yard line) Field Goals v.s. Go-For-It (After 60 yard line) Our model will also be operating under the assumption that the punt will be successful at any yard line. Although this is a strong assumption, we believe this to be a reasonable one, as blocked punts are very rare. As for field goals, we consider this equation

$$3 \cdot p - (1 - p) \cdot (\text{EP}_{\text{fail}})$$

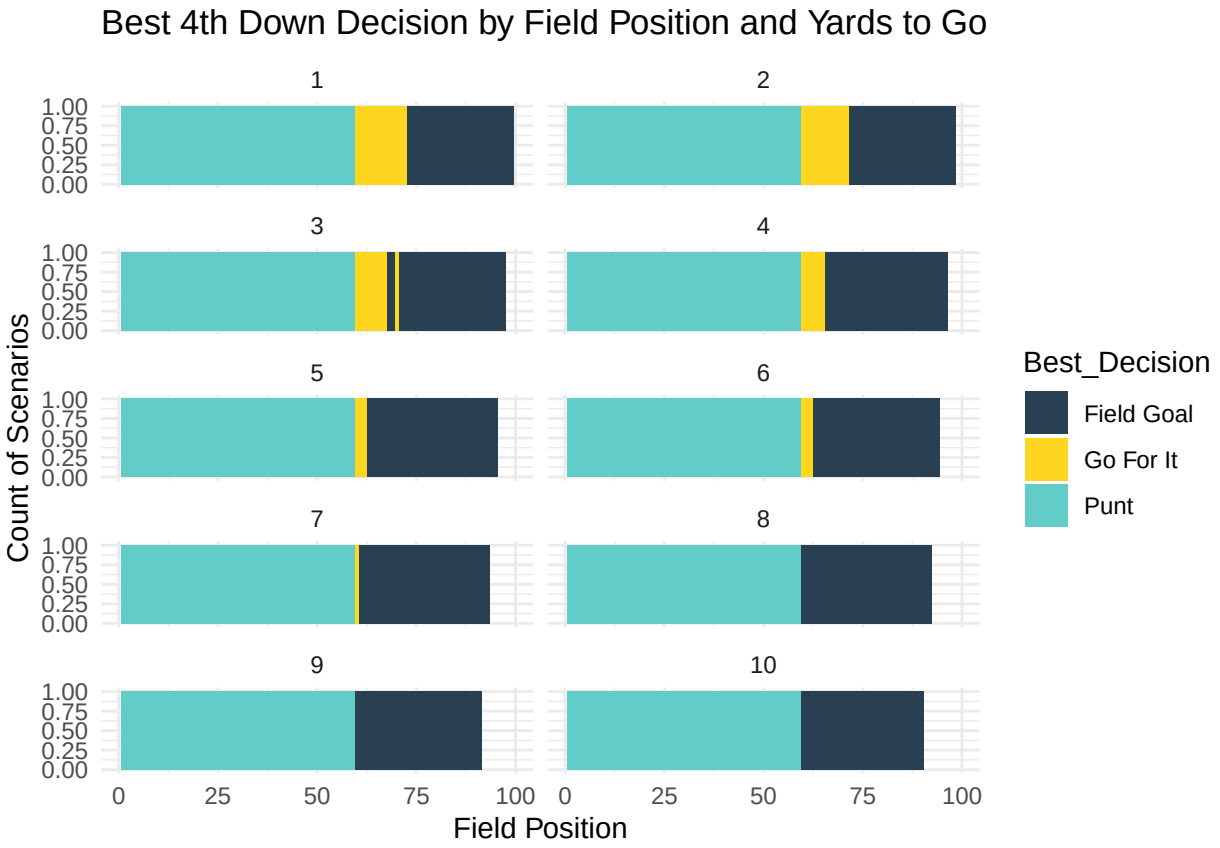
Before the 60-yard line, a field goal is unrealistic due to the distance required. Further, if a miss occurs, the opponent has a very good position to start from. After the 60-yard line, punts become less valuable, as going-for-it or a field goal will give a higher expected points value. Additionally, the EP comparison by nature makes sense if it is a pairwise comparison, using 3 complicates the logic (doesn't go into the equation cleanly). Finally, when using 3, we would have to estimate more parameters which will increase variance and risks overfitting. We can also avoid scenarios such as: "Go-for-it beats punt, but loses to field goal, which loses to punt" (cyclical preferences).

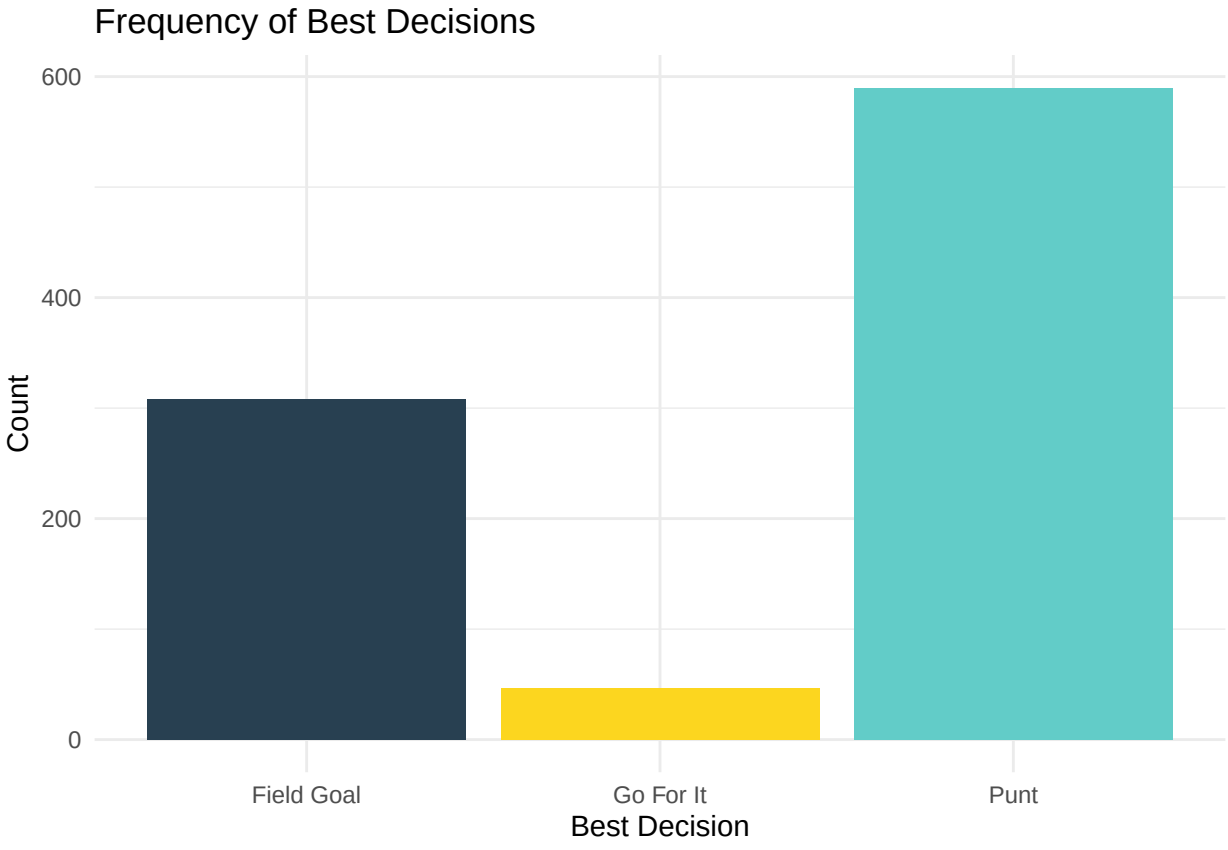
At 4th down, a coach must choose between:

- Attempting to convert (continuation of the play and possibly a score)
- Attempting a field goal (guarantees three points, but with miss risk)
- Punting (giving up possession but improving field position)

Then from the decision threshold, if the probability of converting is higher than the break-even point, the coach should go for it. And if the probability is lower than p, then the decision should lead toward punting. Additionally, we will only be looking at punts up until the 60 yard line, field goals afterwards in comparison to go-for-it. Although field goals are very dependent on who the kicker is for each team, they are less realistic/unlikely to succeed before the opponents 60 yard line. This reduces the number of comparisons that need to be made as well as keeps the multinomial and EP comparisons interpretable. We focus our simulation on fourth-down decisions due to their uniquely high-stakes nature. On downs 1 through 3, unsuccessful plays do not end the possession, allowing for recovery on subsequent downs. On fourth down, however, failure results in an immediate turnover at the line of scrimmage, dramatically changing the game's momentum and expected value. Our simulation framework models each option's success rate and corresponding EP, then evaluates when a coach should deviate from the play with the highest nominal EP in favor of one with lower risk but more favorable probabilistic expectation.

Results


Best 4th Down Decision by Field Position and Yards to Go
1
2
3
4
5
6
7
8
9
10
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
Count of Scenarios
Field Position
Best_Decision
Field Goal
Go For It
Punt


Frequency of Best Decisions
600
400
200
0
Count
Field Goal
Go For It
Punt
Best Decision

From the visualization, punt is usually preferred at poor field position, mostly less than 50, which reflects the high risk of giving the opponent strong field position. Go for it emerges as the best option in a narrow window, usually between FP 55 and 70, and only when YTG is low. The green zone mostly appears when the yard to go is between 1 to 4, and less likely when the YTG is higher. Then as field position advances further into opponent territory, usually past the 70 yard line, field goals become increasingly favored, especially when YTG is moderate to high. So in this sense, the probability of converting a 4th down is not high enough to justify the risk, while a field goal offers a decent chance at guaranteed points.

The result shows that risk aversion appears quite significantly particularly in the mid-to-high field position range. There are many scenarios in which the expected points for the go-for-it option are often close to or even exceed those of the alternative, but the decision still chooses a punt or field goal instead. In terms of decisions close to the opponent's end zone, the possible explanation is that a field goal offers a relatively narrower outcome distribution than go-for-it. Field goal particularly provides either 3 or 0 points, but go-for-it is a higher variance play that could either lead to a new play with potential touchdown opportunity or a turnover and eventually result in losing points. The coaches usually will prefer a low variance outcome than a high variance outcome even though it might provide more expected return, and this is consistent with risk aversion which decision-makers often prefer guaranteed smaller gains over risky larger ones.

This tendency toward risk aversion naturally leads to measurable opportunity cost, which is the value lost when a team opts for a safer but sub optimal choice. Across all fourth down scenarios, when the ball is in the midfield or opponent's territory makes the opportunity cost look more significant. By risk aversion, the team might choose "field goal" or "punt" even when the expected point of "go-for-it" is higher because of the higher risk and larger distribution of variance, but consistently choosing the conservative play accumulates lost value over the season. While a single play may not swing a game, repeated small opportunity costs compound and can influence win probabilities over time.

Conclusion

Limitations

When examining a question like "should we deviate," both field position and yards-to-go are key factors, but typically there are other measures that come to mind when it comes to decision-making. If we had more time/resources it would have been interesting to examine score differential as well as time and their influence on risk. For example, with less time on the clock and a high score differential, coaches may be more likely to go for riskier plays. Introducing score/time adds hundreds of unique game scenarios. For each, you'd need to simulate all options and weigh them, massively increasing computation and variance. Our goal is to explore when coaches should deviate from the most EP-optimal play based on risk of failure, not to fully replicate a game clock or strategy tree. Including time and score would risk burying that insight under too many conditionals.

Additionally, our plays are assumed to be independent of one another. In reality, momentum, player fatigue, play-calling patterns, and psychological factors can influence both success probabilities and decision preferences. The way that we instead do this is through the average of plays, instead of seeing previous decisions that could influence the final 4th down play.

Another factor that coaches take into account when playcalling is the personnel on the field, both their own players on offense but also what players are on defense. Certain individual matchups may alter a coach's decision. For example a strong running back and offensive line could make short fourth down conversions more likely, leading to choosing to go for it more often, regardless of field position. Unfortunately, the scope of our model was unable to look into individual players and their effects on the team's expected points. Given enough time, expanding our simulations to account for the strengths and weaknesses of different NFL teams could grant further insight into answering our question.

Finally, our study is limited in how it can capture the situation surrounding a game. Coaches may make different decisions depending on when a game is played, as the end of the season and the playoffs generally hold higher stakes for the team than the first few weeks of the season. Decision making may also be affected by the records of division and conference rivals. Both of these factors could lead to either safer or riskier decision making, especially on fourth downs. With more time and access to further datasets, we may be able to take these factors into account. Our current model is built using regular season data, but incorporating data pulled from the playoffs into our study could better represent the stakes and pressure coaches are under when making key decisions.