# Beating the Casinos by Predicting Home Win Probability in the NFL

Matt Coates 314102189, Daniel Guerrero 48592068, Maggie Puckett 48661722, Calvin Skalla 48704960, Julio Suriano 48682730

12/1/2021

**Introduction**

The purpose of this report is to determine the probability an NFL team will win at home based on statistics from the games played from 2010 to 2019. In the NFL data we pulled, there were over 300 statistics that contained information from the play by play of all 2,670 games, and we used the data from various areas of the game to extract time of possession, penalty yards, turnovers, and third down conversions rates. We can analyze the relationship between a home team's time of possession, turnovers, turnovers forced, penalty yards, and third down conversion rate to predict their probability of winning at home. By using these extractions, we can create a model to predict whether teams will win at home and what impacts a win at home.
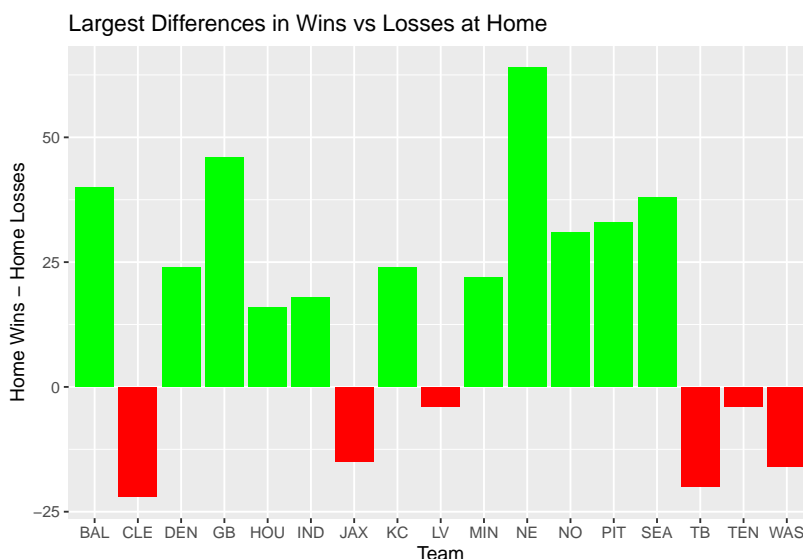


For the graph above, we aggregated all of the home wins and subtracted all of the home losses away for each team. We see that some teams won several more games at home than lost but some teams actually lost a lot more games at home than won. Seems like New England, Green Bay, and Baltimore have an advantage at home while Cleveland, Tampa Bay, and Washington have a weak home field advantage.

---

Our goal was to be able to build a model that with given inputs, will shoot out a probability that a team will win at home. The idea is, to find money lines from sports books that pay out higher than the lower bound of our 95% prediction interval. This way, if our model is accurate, in the long run, we will be making money on our bets because the casino is setting the money lines lower than our projected probability of a win. Casinos usually set their initial odds at what they think will happen in the game, and then while bettors start placing bets on either side, the casino has to tinker with the odds to produce the closest amount of money on each side. They do not care about having the closest line to the actual score but rather having equivalent amounts of money on both sides of the bet to maximize their profit. So, when we see games that start to produce moneylines with higher payout probabilities than our prediction models, we will want to

target those games because we are confident that we will make money. There is always the chance that the lower probability team does end up winning, but in the long run we will be profitable due to choosing lines where the probability is lower than our 95% prediction interval lower bound.

For example, say we have the Dallas Cowboys playing at home against the New York Giants. The casino places the moneyline for the Cowboys at -165, which means we will be paid out our wager plus our wager divided by 1.65. Meaning that if our prediction interval lower bound from our model says the Cowboys will win by a probability of over 62.3%. We should place that bet because in the long run, we will be profiting money from these wagers. So now, we will get into how we created the model for predicting the probability a home team will win the game.

---

**Logistic Regression Equation**

We decided to use a Logistic Regression model to predict the probability of a home team winning the game. To predict the probability of a home team winning, we need independent variables that we believe have an impact on whether the home team wins. For those independent variables, we decided to have home team's time of possession, accrued penalty yards, turnovers, turnovers forced, 3rd down conversion rate, and which home team itself out of the 32 different options in the NFL. After running those impactful statistics through the model, we received this model equation to predict probabilities.

$$\log\left[\frac{p}{1-p}\right] = -5.97 + 0.22 * TOP - 0.01 * PenYards - 0.72 * TO + 0.37 * TOForced$$

$$+2.57 * ThirdConRate - 0.65ATL + ... - 1.04 * WAS$$

So, interpreting the numbers above, the numbers attached to the variable name correspond to the change in the log odds of the probability of winning when there is a one unit increase in that variable. For a first-hand example below, the Arizona Cardinals are our baseline group when not designating a specific team. So, the probability of the Arizona Cardinals winning at home, assigning each other statistic to the NFL average, is 64%. To see the change in probabilities when changing teams, refer to the table below. We see that when changing the home team to these 11 different NFL teams (which are drastic from the graph above), we can see the marginal effect of each team since we are assigning the other independent variables to the league average consistently.

Table 1: Probabilities of Winning at Home given League Averages

|  | BAL | CLE | DEN | GB | JAX | KC | NE | PIT | SEA | TB | WAS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Probability | 0.76 | 0.44 | 0.71 | 0.73 | 0.50 | 0.68 | 0.81 | 0.67 | 0.76 | 0.41 | 0.39 |
| Change from Baseline | 0.12 | -0.20 | 0.07 | 0.09 | -0.14 | 0.04 | 0.17 | 0.03 | 0.12 | -0.23 | -0.25 |

So, after looking at how changing solely the home team that is playing while keeping the other independent variables consistent, we can see large impacts of some teams in the NFL. Now, when looking through different sports books, keep an eye on the teams playing at home that have a high probability of winning but also keep an eye on the teams playing at home that have a low probability because we can either bet for or against the home team.

Now, let's switch gears to look at the marginal effects of a team's turnovers. How will the probability of a home team winning change when their number of turnovers goes from 0 to 1 to 2 all the way up to 4? Well, when holding the team, and other independent variables constant, we can see that the probability continually drops the more turnovers the home team has to eventually dropping below the 50% mark when over 2.

Table 2: Probability a Team Wins at Home while changing Turnovers

|  | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Probability | 0.89 | 0.8 | 0.66 | 0.49 | 0.32 |

Since we looked at a team's turnovers, we thought we should also look at their ability to force turnovers. We should see that as forced turnovers increase, the probability that the home team wins will increase.

Table 3: Probability a Team Wins at Home while changing Turnovers Forced

|  | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Probability | 0.45 | 0.54 | 0.63 | 0.71 | 0.78 |

Our results confirm our hypothesis because each additional turnover forced leads to a significant increase in the probability of a team winning at home. When the home team goes from forcing 1 turnover to forcing 2, the home team's probability of winning is increased by 9%. This would lead to us looking for the teams who are averaging more forced turnovers per game when deciding which teams to bet on. We should look for matchups that have a large difference in average turnovers per game.
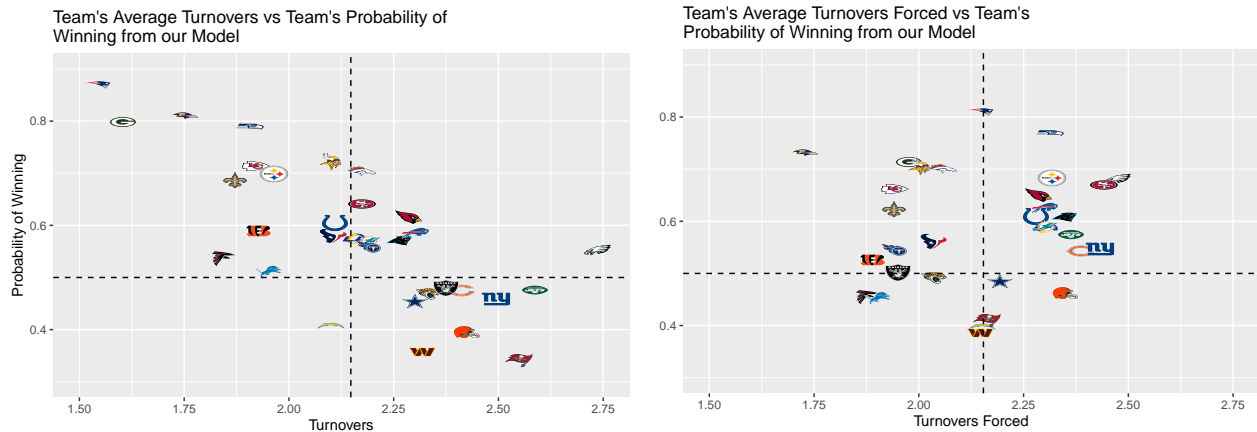


So, turning back to look at how turnovers and turnovers forced effect a home team's win percentage. We calculated each team's average amount of turnovers and turnovers forced per game from 2010 until 2019. Then using the league averages for all of the other independent variables, we created probabilities, by using our model, that the said home team would win their game given their average turnovers per game (left) and average turnovers forced per game (right) from the last 10 years. Then, we plotted all 32 NFL teams with their average turnovers and turnovers forced from the last 10 years on the x-axis with their probability of winning from our model on the y-axis.

We see that there is clearly a negative relationship with turnovers and the probability of winning. So, we can conclude that teams turning the ball over will have less of a chance of winning their games at home. However, we do see that some teams with more than the league average of turnovers still have probabilities over 50% and the that Chargers have less than average turnovers but a less than 50% probability of winning at home. This may be because those teams average more or less turnovers forced than the league average. So, we want to make sure we monitor the average turnovers a team and opponent have been accruing and forcing when deciding whether the matchup is a good bet or not.

Now let us grasp the same concept above but instead of plugging in turnovers or turnovers forced, let us use the teams' average third down conversion rate from the last 10 years. This way, we will see how using the average third down conversion rate from team to team will impact the probability of winning at home.

Team's Average 3rd Down Conversion Rate vs Team's Probability of Winning from our Model
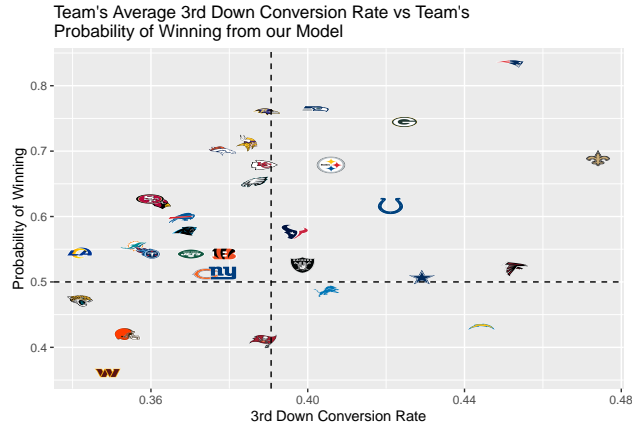
Using the graph above, it shows the general trend that as the percentage of successful 3rd down conversions increases for a given home team, that team's probability of winning the game will also increase. So, when we are looking at teams or matchups to bet on, we will want to scan over their third down conversion rates in previous games to see how well they have been converting on third downs, along with how poorly the opposing team is converting on third down.

---

### Endogeneity

After running this model and finding the marginal effect of the variables on win percentage, we wondered if there were some factors that impacted home teams winning that we did not include. Does a team's discipline and football IQ impact a win or not? Does that even matter if they are home or away? So, we wanted to find some factor that we could use that would account for a team's discipline and IQ. We decided to create a new variable by counting the number of undisciplined penalties a team had in the game. Penalties included in the undisciplined count are as follows: Neutral Zone Infraction, Defensive Offsides, Delay of Game, False Start, Illegal Motion, Taunting, Unnecessary Roughness, Illegal Substitution, Illegal Shift, Unsportsmanlike Conduct, Defensive/Offensive Too Many Men on Field, and Encroachment. We then used the number of undisciplined penalties to predict the amount of penalty yards a team had in that game. Unfortunately, our model became weaker when we used the same independent variables to to predict win probability but traded out the predicted penalty yards instead of the true penalty yards. This most likely means that the number of undisciplined penalties in a game is not a good tool for estimation of the total penalty yards a team accrued in a game. Suppose we performed an experiment to see if a team is more disciplined at home or on the road. We could track the number of undisciplined penalties they average on the road and the average amount of undisciplined penalties at home, to see if teams produce more penalty yards at home or not. We would be able to see if teams are more disciplined at home, which results in less penalty yards at home, which would result in a higher probability of winning at home.

### Conclusion

So, after extracting full game data from the 2010-2019 NFL season play by plays, we were able to build a model to predict the probability of a home team winning at their stadium dependent on the home team's time of possession, turnovers, turnovers forced, penalty yards accrued, and third down conversion rate. The idea is to carry this model into the NFL season, input the season averages for the team's considered in a bet, calculate the 95% prediction interval probability that the home team will win, then if the odds payout the casino has placed has a lower probability than our lower bound of our 95% prediction interval based off the teams' statistics, we will place the bet.

As we continually seek out matchups that have moneyline odd probabilities lower than our desired probability to win, we should continually make profit because the probability of the home team winning is higher than the risk of placing that money on the home team for the resulting payout.