An Examination of Win Probability in NCAA Division I Football Overtime and Its Applications for Evaluating Coaching Decisions

Matt Ulishney
Advised by Dr. Drew Pasteur

Abstract

The increasing use of analytics in sports has changed the way organizations, coaches, and players approach game planning, strategies, and evaluation of performance. One common metric to describe a team's chances of winning in progress contests is win probability, or WP. Win probability is comprised of many game state variables, including down, distance, yards to goal, and score differential, along with team strength ratings such as point spread and offensive or defensive ratings. In this study, we will examine what variables are useful in determining WP for in progress NCAA Division I football games in the overtime periods. We use logistic regression and decision trees to create WP models using various game state and team strength variables. Our results are promising, as they perform well in describing a team's WP given various overtime game state situations. We can use our results to evaluate coaching decisions made in the overtime periods, in addition to determining which plays had the greatest impact on a team's chances of winning the game.

Overtime Rules

- Overtime occurs when the score is tied after the 4th quarter
- Coin toss is used to determine possession
- Possessions start with a 1st and 10 at the opponent's 25 yard line
- Each team gets one possession per period
- The team with the most points at the end of a quarter wins
- The 2nd overtime period has reversed order of possession
- Teams must attempt a two-point conversion after a touchdown in the second period
- Alternating two-point conversions determine a winner if tied after two periods

Data Source

- Acquired from R package *cfbfastR*
- Advanced play-by-play data for games starting in the 2014 season
- Final dataset contained 4130 plays from 299 games
- Data cleaning was required for over 100 games
- Reordering plays, fixing data errors, and manual data entry required

Variables of Interest

Variable	Description
down	Play number in the current 4 down series
distance	Distance from the line of scrimmage to the yard to gain
yards_to_goal	Distance from the line of scrimmage to the end zone
pre_diff	Pre-snap score differential between the offense and defense
home_team	Indicates if the current offense is the home team
def_first	Indicates which team started overtime on defense
fav_team	Indicates if the current offense is the favorite
pts_favored	Spread amount by which the current offense is favored
momentum	Indicates if the offense was the last to score in regulation
off_4th	Total points scored in the 4th quarter by the team on offense
def_4th	Total points scored in the 4th quarter by the team on defense
diff_4th	off_4th minus def_4th
off_4pct	Proportion of all 4th quarter points scored by the offense
second_pos	Indicates if the current possession is second in the period
off	Offense red zone rating
def	Defense red zone rating

Logistic Regression

- Goal: Create a model that produces an estimated WP for the current offense given a possible overtime situation
- Logistic regression produces estimated WP value given a set of predictors
- Fit simple logistic regression models, then added combinations of predictors
- AIC, RMSE, drop in deviance, and model calibration used to determine most descriptive models
- Repeated k-fold cross validation 1000 times to determine final model

Final Model

Variable	Name
x_1	down
x_2	yards_to_goal
x_3	pre_diff
x_4	pts_favored
x_5	second_pos
x_6	home_team

Equation 1 shows the equation for the final model with the variables listed above

$$ln\left(\frac{p}{1-p}\right) = 1.04 - 0.31x_1 - 0.03x_2 + 0.41x_3 + 0.03x_4 + 2.01x_5 + 0.18x_6 \quad (1)$$

Model Calibration

- Used to assess descriptive ability of WP model
- If we produce a WP of 80%, we expect that team to win 80% of the time
- Binned WP values at 2.5% and 5% intervals plotted against actual winning percentage for each bin
- Perfect model results in the line y=x
- Model calibration plot in Figure 1 shows our model accurately describes WP for the current offense

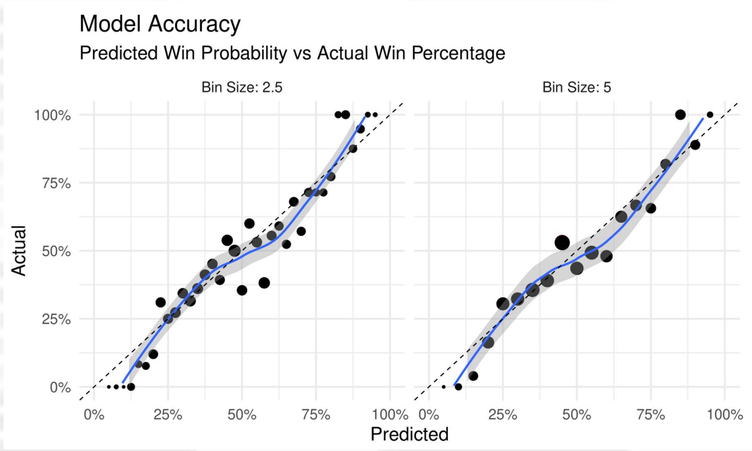

Figure 1: Model calibration plot graphing predicted WP against actual win percentage

Applications

- Results can be used to evaluate coaching decisions and provide insight for coaching decisions for games in progress
- Can calculate the expected WP for each decision
- Two point plays and fourth downs have the largest impact on WP
- Win probability added (WPA) can be used to determine which plays had the greatest impact on a team's WP

Case Study

Figure 2 shows Tennessee's win probability over the course of the overtime period against Purdue in 2021. In this game, Tennessee was faced with two separate fourth down situations. The first was a fourth and two from the 17 yard line. If Tennessee elected to kick the field goal, their expected WP would be 30%. Attempting to convert a first down results in an expected WP of at least 42%. Tennessee converted the first down, and their win probability added (WPA) from the previous play to the current play was 25%. The second situation was a fourth and one from the 1 yard line. Kicking the field goal results in an expected WP of 34%, whereas trying for the touchdown produces an expected WP of 56%. Tennessee was unable to score the touchdown, resulting in a WPA of -36%. The Tennessee head coach made the decision that resulted in the highest expected WP in each situation, but ultimately lost the game.

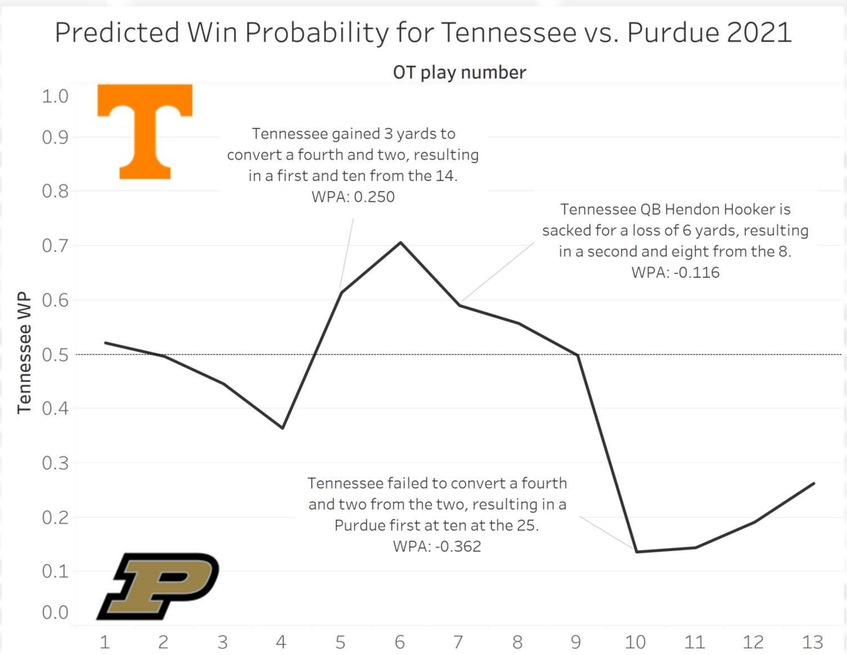

Figure 2: Tennessee's predicted WP over the course of the overtime period against Purdue in 2021

References

[1] Dan McDonald. *How College Football Overtime Works.* World Wide Web. Accessed on 4/5/2023. URL: https://www.ncaa.com/news/football/2019-07-11/how-college-football-overtime-works

WOOSTER
Mathematical & Computational Sciences