Deep Learning Forecasting of NFL Game Outcomes Relative to Betting Odds

Quinlin Gregg

Betting

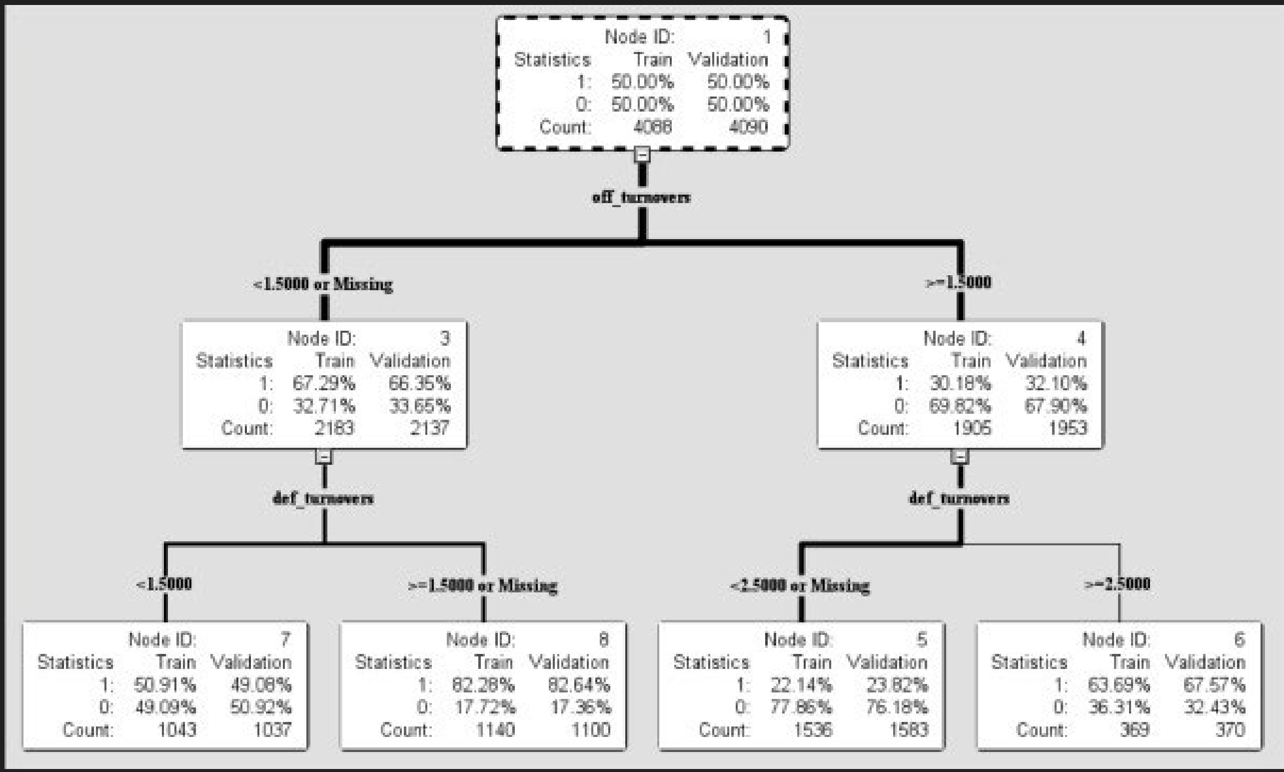
- Two types of bets were considered:
  - Over/Under: A “50-50” bet on whether the combined score of both teams will be over or under a target
    - Away Score + Home Score $\overset{?}{>}$ Target
  - Spread: A “50-50” on whether the home team will win be a least a given number of points
    - Home Score – Away Score $\overset{?}{>}$ Target
- With repeated bets, one can expect to lose 3-4.5%
- Can the outcome of an NFL game be forecasted, and can those forecasts be used to make profitable bets?

Golec and Tamarkin (1991)

- Used linear regression and F-tests to test for biases in betting odds
- $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$
  - X_1 is a vector of ones, X_2 is the betting spreads, X_3 is a dummy variable representing whether a team is at home, and X_4 is a dummy variable representing whether a team is favored
  - If the odds are efficient and have no bias, then $\beta_1 = \beta_3 + \beta_4 = 0$ and $\beta_2 = 0$
- All else being equal, bets on underdogs or home teams win more often than those on favorites or visiting teams

Gifford and Bayrak (2023)

- Focused on building binary logistic regression models and decision trees to predict where a team will win based on stats like turnovers, rush yards, and whether the game went to overtime
- They found that a logistic regression could be made that predicts the outcome of a game with a 16.9% misclassification rate (21.6% for decision trees)



Model Formulation

- Part 1: Predict proportion of drives ending in touchdowns, field goals, and punts
  - Analyze previous 16 games of each team to determine offensive and defensive outcomes
  - Create 8 time series: touchdown/field goal, home/away, offense/defense
- Part 2: Build team point total distribution using predicted proportions
  - Average respective offense and defense forecasts together
  - Assume drive outcomes are independent
  - Model game outcome with trinomial distribution and Monte Carlo simulation

Part 1: Drive-Level Outcomes

- Models Tested:
  - Least-Squares Regression
  - Holt Exponential Smoothing
  - Recurrent Neural Network (RNN)
  - Long Short-Term Memory (LSTM)
  - Multi-Layer Perceptron (MPL)
  - Transformer Model
- Training Data:
  - Data from the 2020-2023 NFL seasons (regular season games)
  - 1,706 time series built from offensive and defensive drive outcomes

Part 2: Drive-Level Outcomes

- Monte Carlo Simulation for Game Prediction:
  - Use predicted drive-level probabilities (touchdowns, field goals, punts).
  - Trinomial distribution for game outcomes: $X \sim trinomial(p_{touchdown}, p_{field\ goal}, p_{punt})$ where $p_{punt} = 1 - p_{touchdown} - p_{field\ goal}$
- Expected Score: $7 * p_{touchdown} + 3 * p_{field\ goal} + 0 * p_{punt}$
- Forecast real 2024 game outcomes using data from the 2023 and 2024 NFL seasons

2024 Forecast Results

Model	TD Perc MAE	FG Perc MAE	Score MAE	Home Line Accuracy	Over/Under Accuracy
LSR	0.108	0.099	7.378	0.495	0.552
RNN	0.115	0.106	8.214	0.494	0.490
LSTM	0.108	0.098	7.367	0.490	0.500
MLP	0.116	0.105	8.480	0.454	0.516
Transform	0.107	0.101	7.581	0.495	0.516
Holt	0.132	0.120	9.512	0.469	0.485

Conclusions and Further Research

- 2024 NFL Forecasting
  - The models were typically off by about 7 points per team in score predictions
  - This variance is somewhat expected given the inherent unpredictability of sports outcomes
- Betting Accuracy
  - The models did not outperform random guesses in predicting over/under or game spread outcomes, generally staying close to a 50/50 chance
  - Some models, like LSR, MPL, and Transform, show promise if their accuracy can be sustained
- Future Enhancements
  - There are opportunities for improvement by adjusting the window length to better account for player injuries
  - Separating offensive and defensive data and ncorporating advanced metrics like EPA and DVOA could further refine the models.

References

[1] M. Gifford and Tuncay Bayrak, “A predictive analytics model for forecasting outcomes in the National Football League games using decision tree and logistic regression,” Decision Analytics Journal, vol. 8, pp. 100296–100296, Aug. 2023, doi: https://doi.org/10.1016/j.dajour.2023.100296.
[1] J. Golec, “The degree of inefficiency in the football betting market Statistical tests,” Journal of Financial Economics, vol. 30, no. 2, pp. 311–323, Dec. 1991, doi: https://doi.org/10.1016/0304-405x(91)90034-h.

2020-2023 Test Fit Results (MAE)

Model	TD Offense	TD Defense	FG Offense	FG Defense
LSR	0.105	0.104	0.091	0.095
RNN	0.113	0.113	0.103	0.108
LSTM	0.110	0.104	0.091	0.095
MLP	0.115	0.116	0.101	0.108
Transform	0.121	0.105	0.094	0.095
Holt	0.135	0.140	0.118	0.120

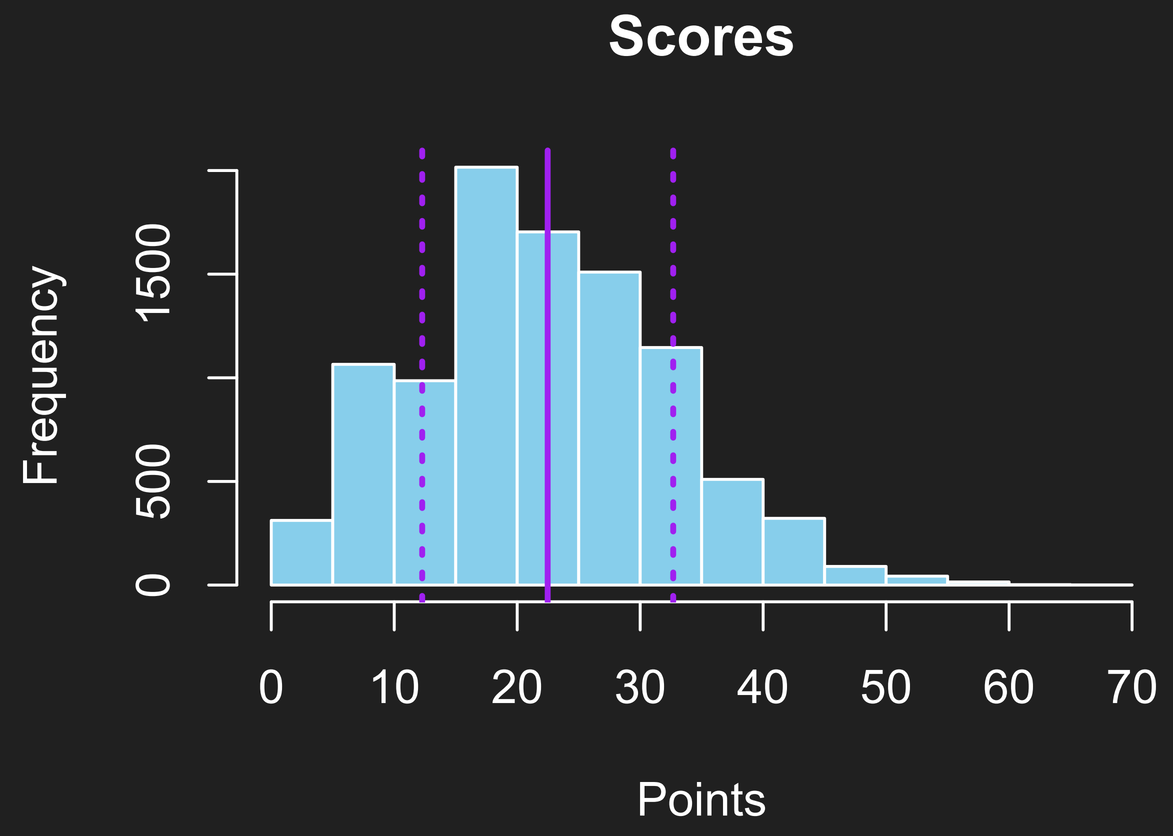
EFIN 301 Presentation

Background

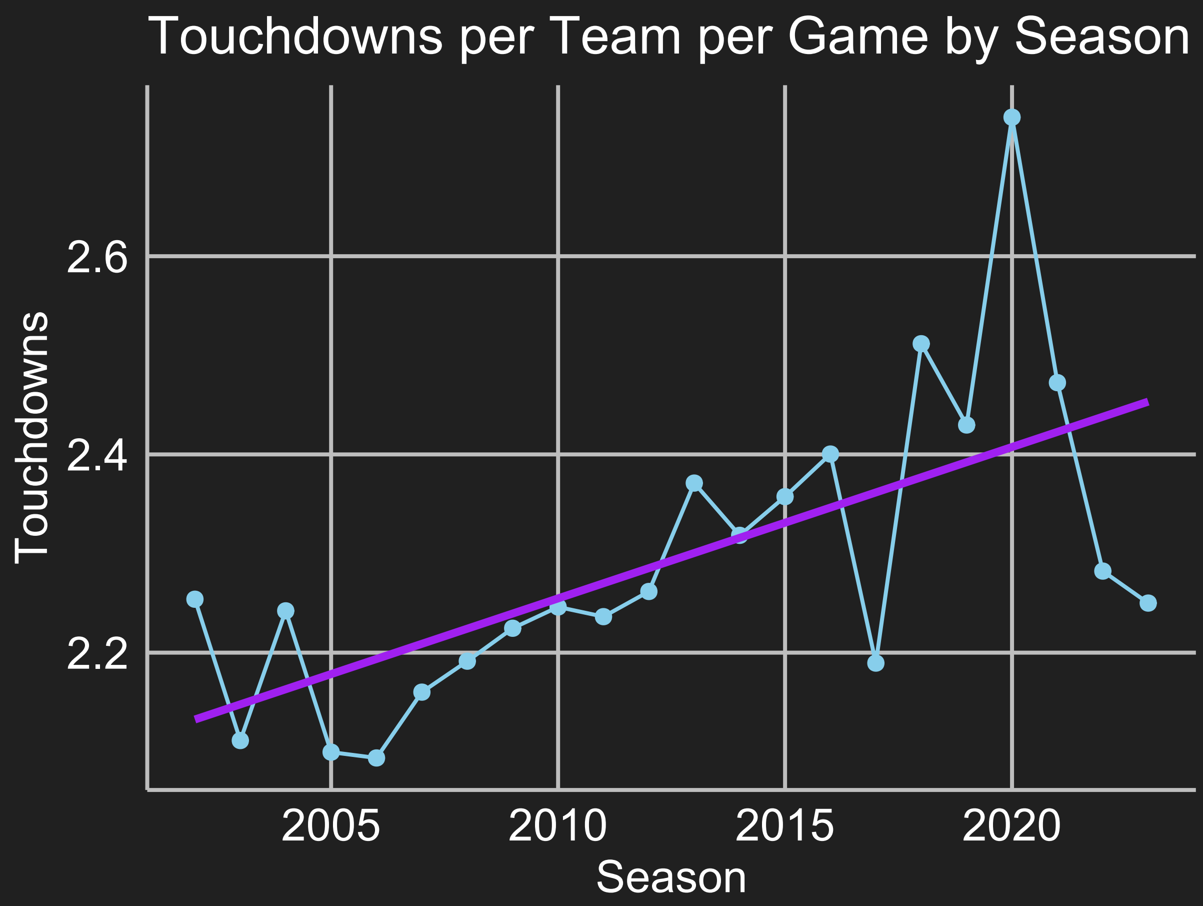
- Sports betting is currently legal in 30 states, Puerto Rico, and the District of Columbia
- Some sites claim to use arbitrage between different oddsmakers to guarantee returns
- Simulation:
  - Using drive-level data, simulate the expected outcomes of a game
  - Using previous drive-level data, predict future drive-level data

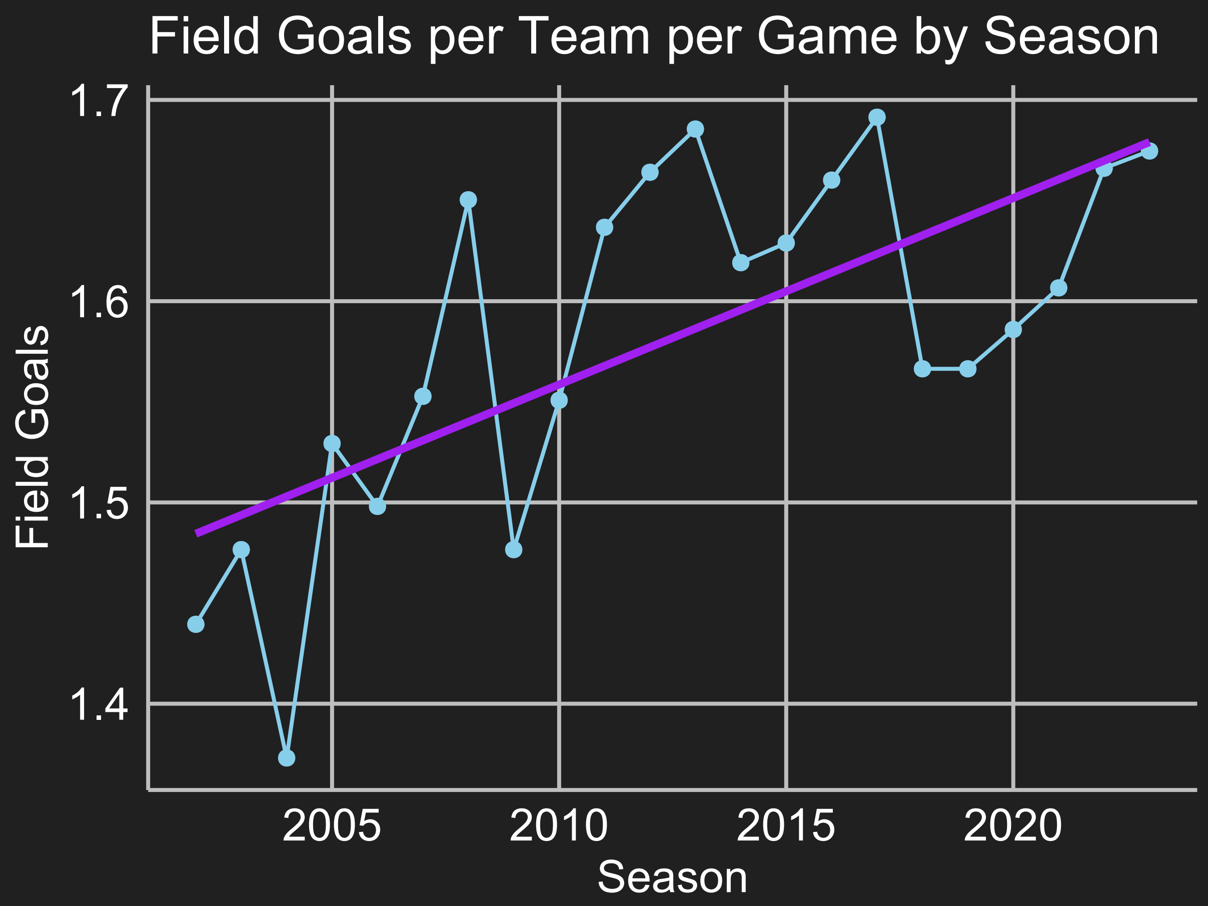
Data

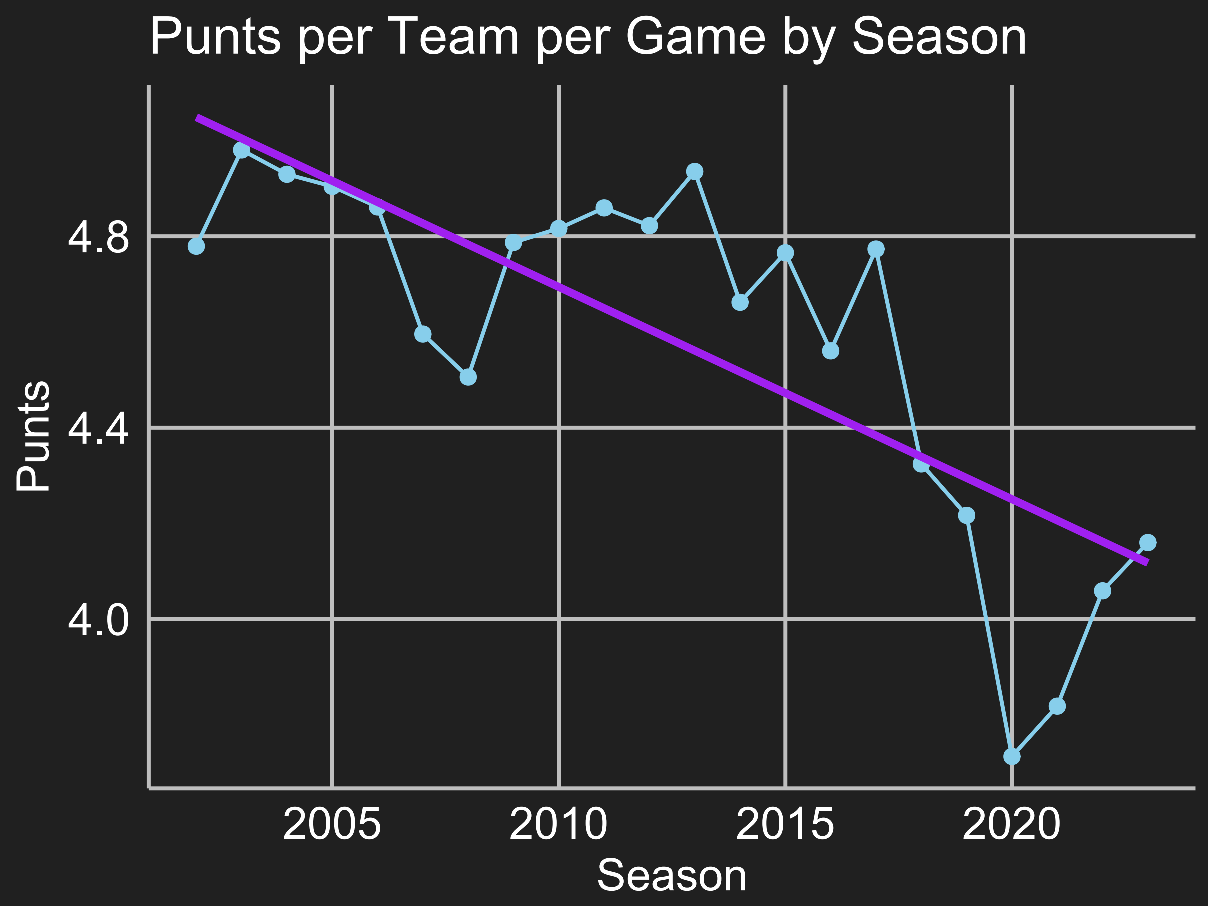
- Drive-level data (offense and defense) on all 32 NFL teams since 2002 (n=704)
  - Touchdowns
  - Field Goals
  - Punts
  - Turnovers
  - Total Drives
- Betting odds and outcomes for all NFL games since 2007 (n=4,861)

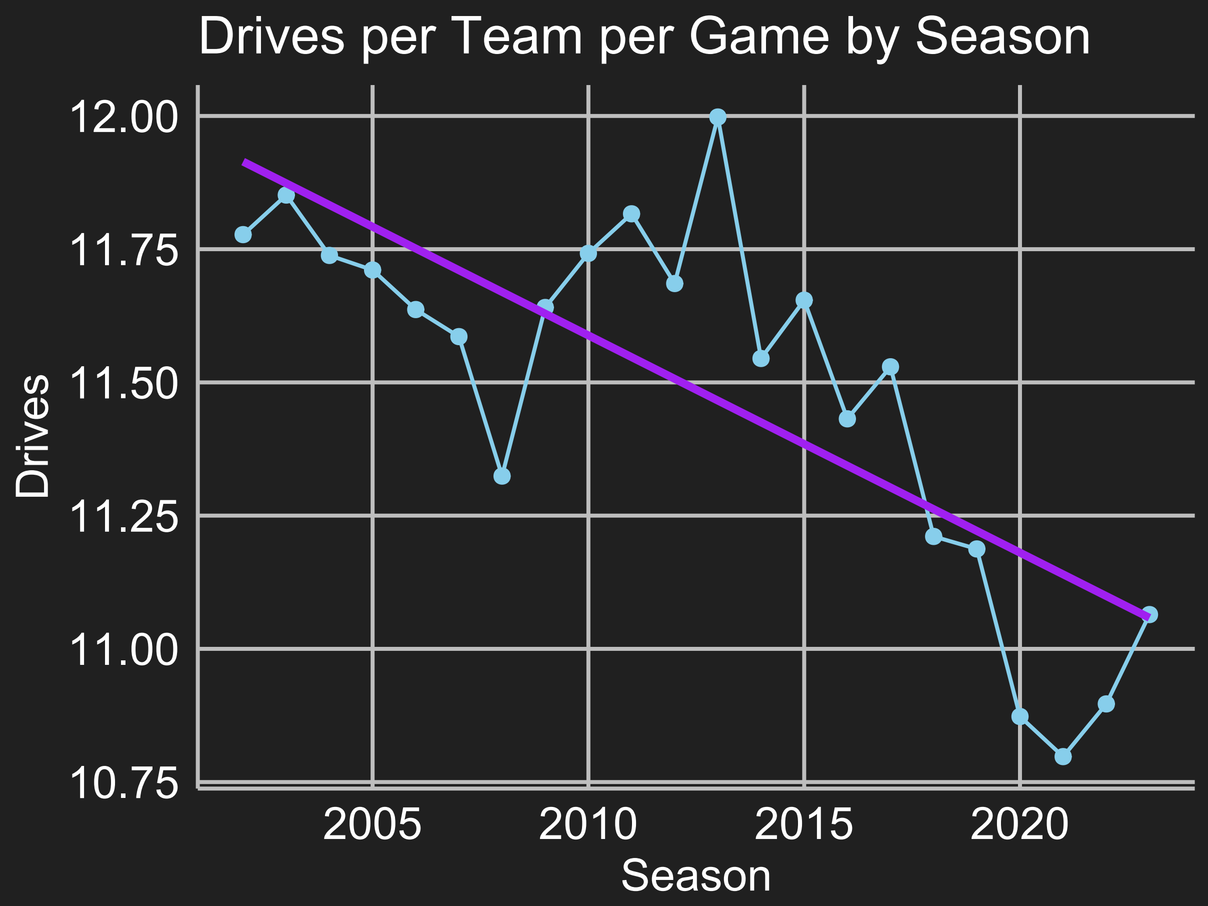


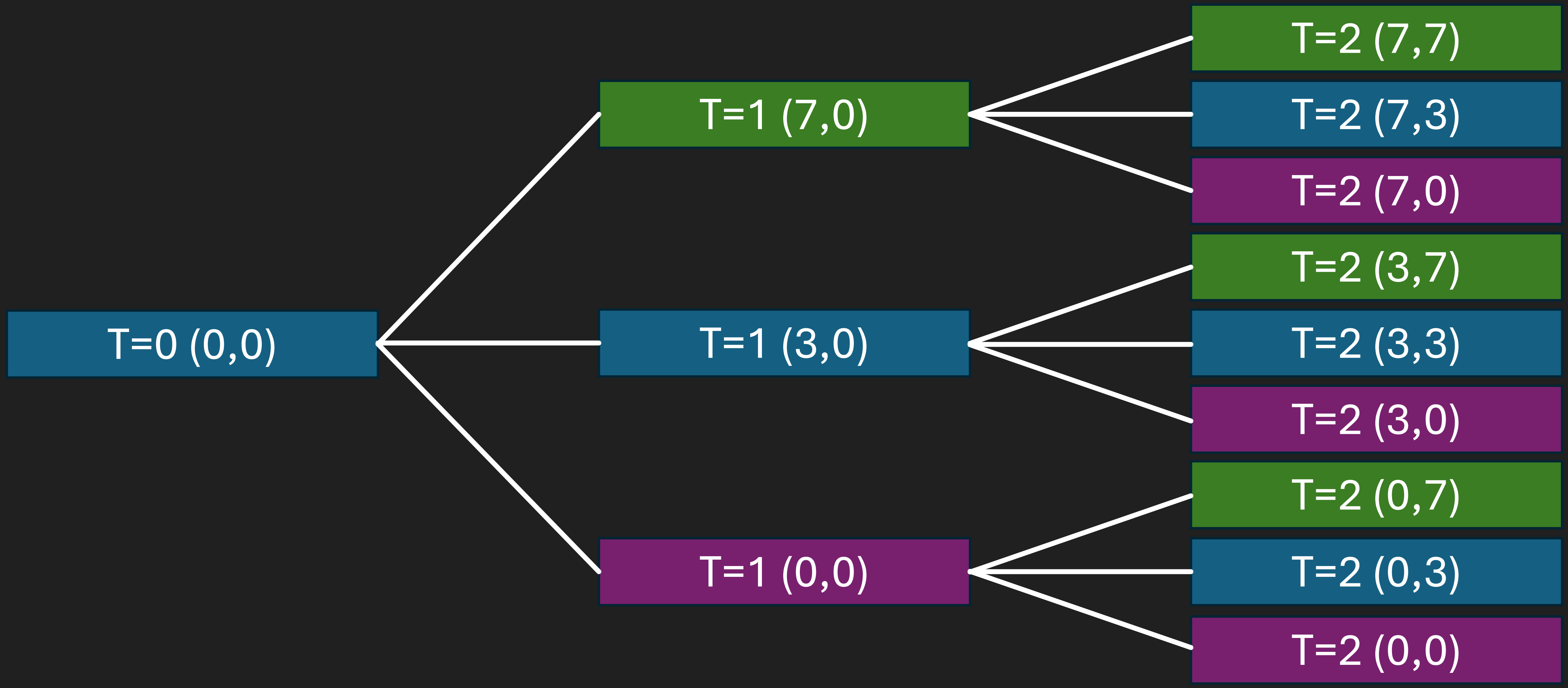
(μ=22.48, σ=10.21, n=9,722)


Touchdowns per Team per Game by Season
Touchdowns
2.6
2.4
2.2
2005
2010
2015
2020
Season


Field Goals per Team per Game by Season
Field Goals
1.7
1.6
1.5
1.4
2005
2010
2015
2020
Season

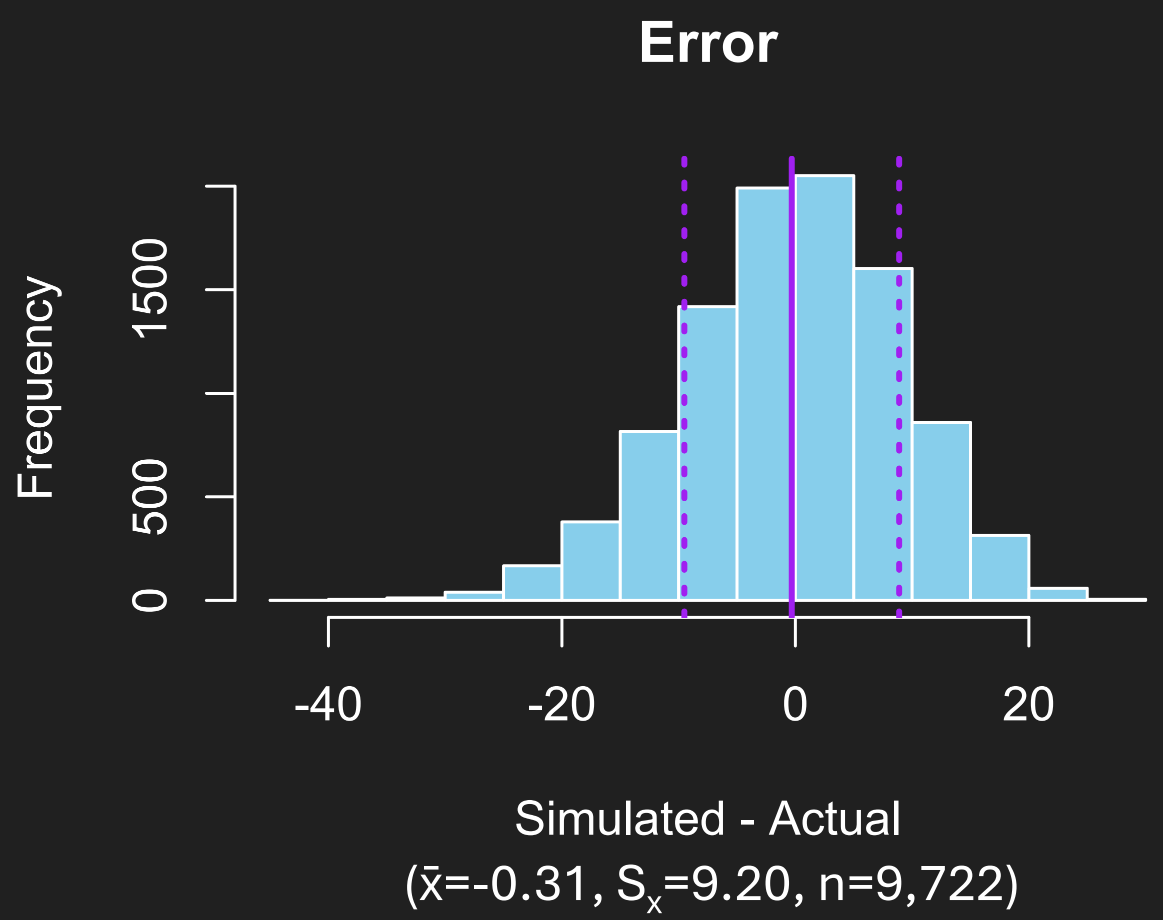

Punts per Team per Game by Season
Punts
4.8
4.4
4.0
2005
2010
2015
2020
Season

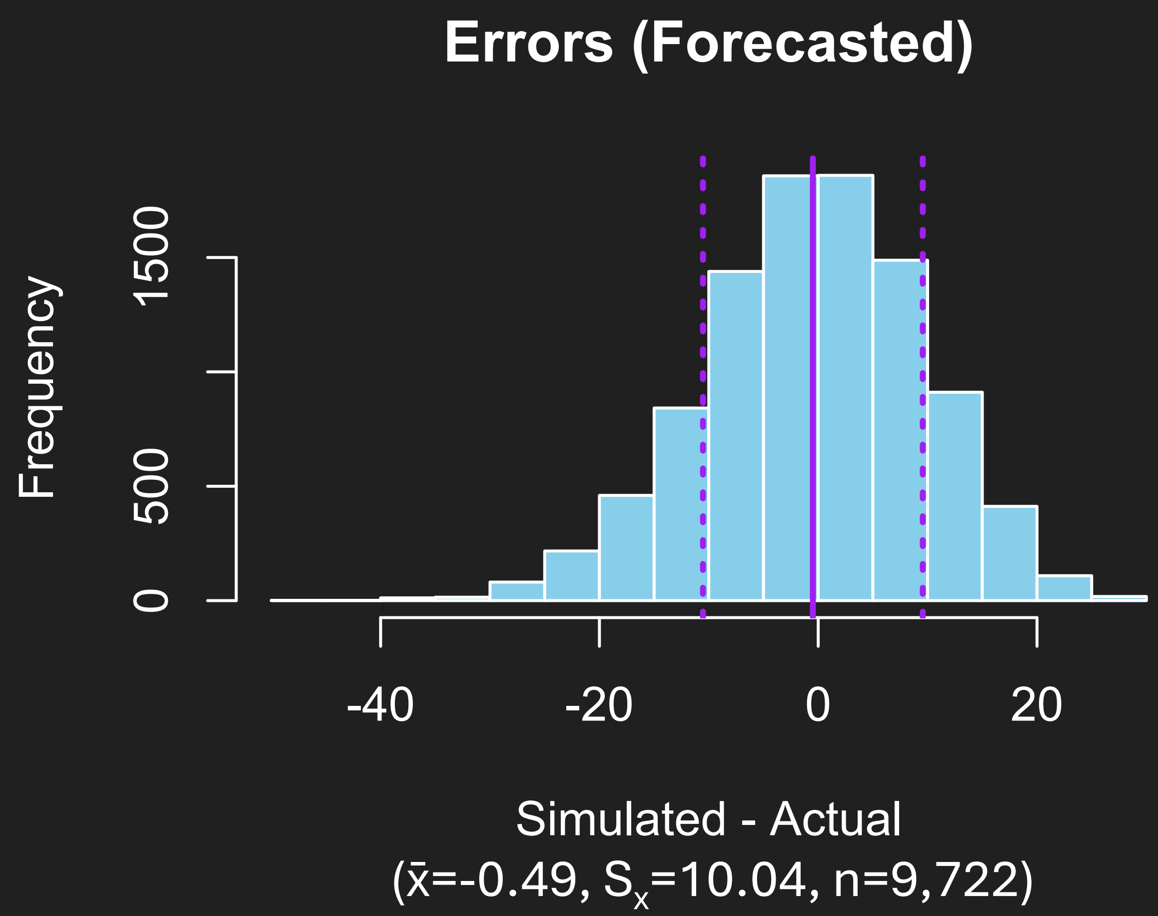

Drives per Team per Game by Season
Drives
12.00
11.75
11.50
11.25
11.00
10.75
2005
2010
2015
2020
Season


T=0 (0,0)
T=1 (7,0)
T=1 (3,0)
T=1 (0,0)
T=2 (7,7)
T=2 (7,3)
T=2 (7,0)
T=2 (3,7)
T=2 (3,3)
T=2 (3,0)
T=2 (0,7)
T=2 (0,3)
T=2 (0,0)

$$\sqrt{td\_per\_drive_t} = \hat{\beta}_0 + \hat{\beta}_1\sqrt{td\_per\_drive_{t-1}} + \hat{\beta}_2 playoff_{t-1} + \hat{\beta}_3 season$$

- *td_per_drive*$_t$: Number of touchdowns scored per drive in season *t*
- *playoff*$_t$: A dummy variable representing whether a team made the playoffs in season *t*
- *season*: The season to account for time trends


Error
Frequency
0
500
1500
-40
-20
0
20
Simulated - Actual
(x̄=-0.31, Sx=9.20, n=9,722)


Errors (Forecasted)
Frequency
0
500
1500
-40
-20
0
20
Simulated - Actual
(x̄=-0.49, Sx=10.04, n=9,722)

Over/Under Bet Returns

Difference in Odds	Percent of Games (in 1 Season)	Returns	Returns (Dynamic)
>0%	100% (272)	-3.9%	-4.4%
>5%	77.5% (211)	-3.5%	-4.1%
>10%	44.0% (120)	-5.3%	-5.0%
>20%	8.4% (23)	-4.0%	-4.4%
>25%	2.3% (6)	-1.1%	-2.3%
>30%	0.5% (1)	-26.5%	-26.6%

Money Line Bet Returns

Difference in Odds	Percent of Games (in 1 Season)	Returns	Returns (Dynamic)
>0%	100% (272)	-3.0%	-4.3%
>5%	76.3% (208)	-3.9%	-4.6%
>10%	53.2% (145)	-7.0%	-6.1%
>20%	18.3% (50)	-8.5%	-6.4%
>25%	8.2% (22)	1.9%	3.0%
>30%	3.4% (9)	1.8%	3.8%

Spread Bet Returns

Difference in Odds	Percent of Games (in 1 Season)	Returns	Returns (Dynamic)
>0%	100% (272)	-11.5%	-0.6%
>5%	97.1% (264)	-11.7%	-0.6%
>10%	91.5% (249)	-11.2%	-0.4%
>20%	77.3% (210)	-5.6%	2.8%
>25%	68.2% (186)	0.5%	7.1%
>30%	58.2% (158)	8.6%	13.2%

Conclusions

- The model is better at determining the difference between the teams' scores than determining the combined score
- Scaling bets based on expected returns may or may not increase returns/decrease losses
- There is a Goldilocks zone of differences to exploit
- Football games are inherently difficult to accurately predict