Analyzing Sports Bets with Quadrinomial Trees

Quinlin Gregg

Background

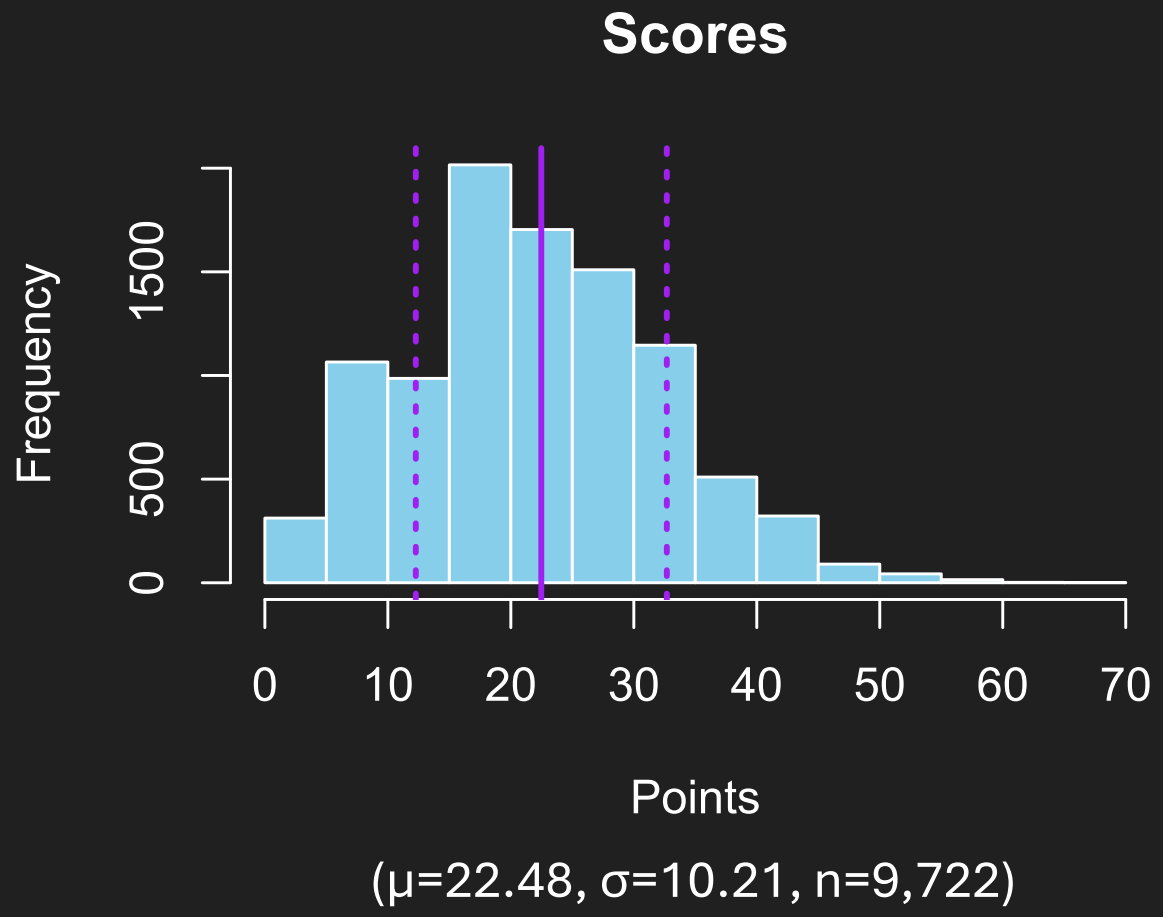
- Sports betting is currently legal in 30 states, Puerto Rico, and the District of Columbia
- Some sites claim to use arbitrage between different oddsmakers to guarantee returns
- Simulation:
  - Using drive-level data, simulate the expected outcomes of a game
  - Using previous drive-level data, predict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

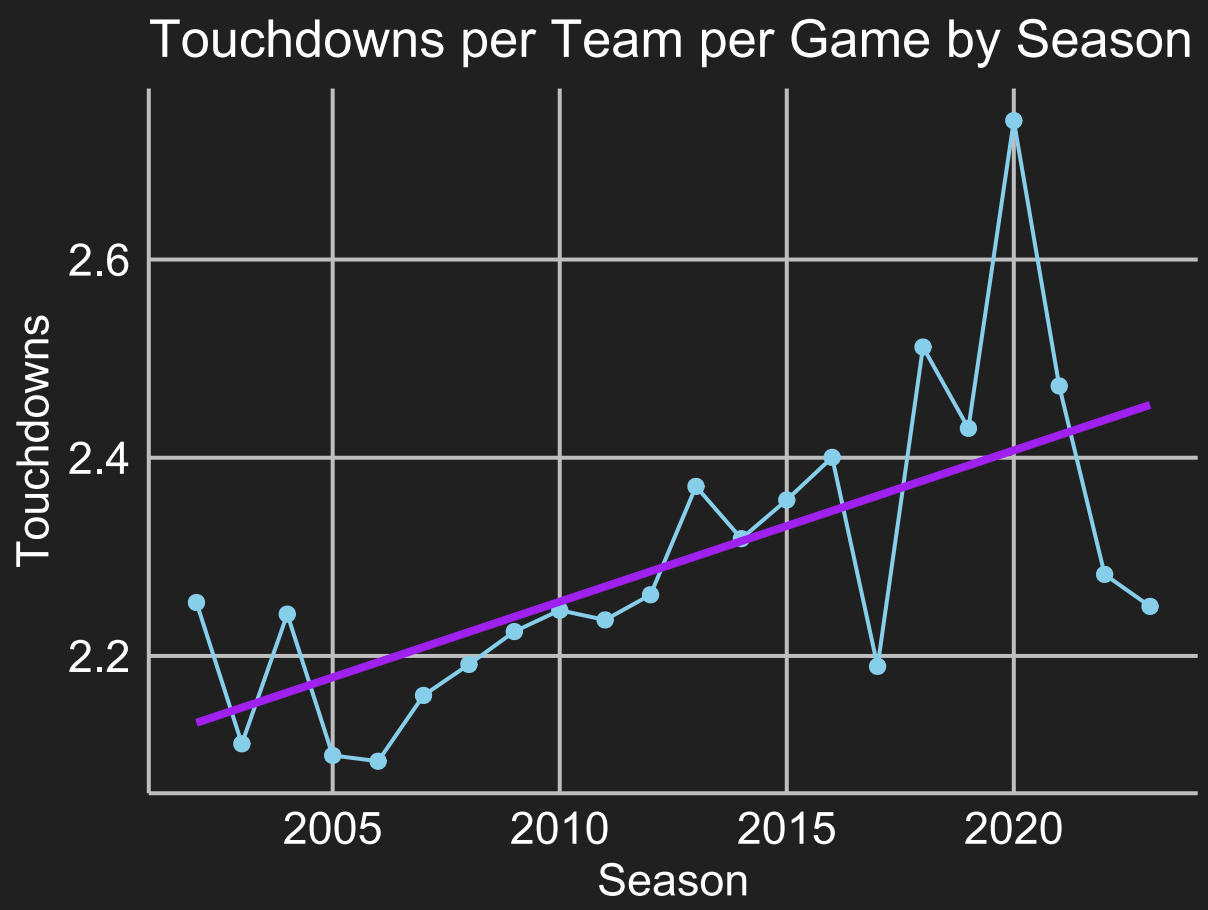
Betting

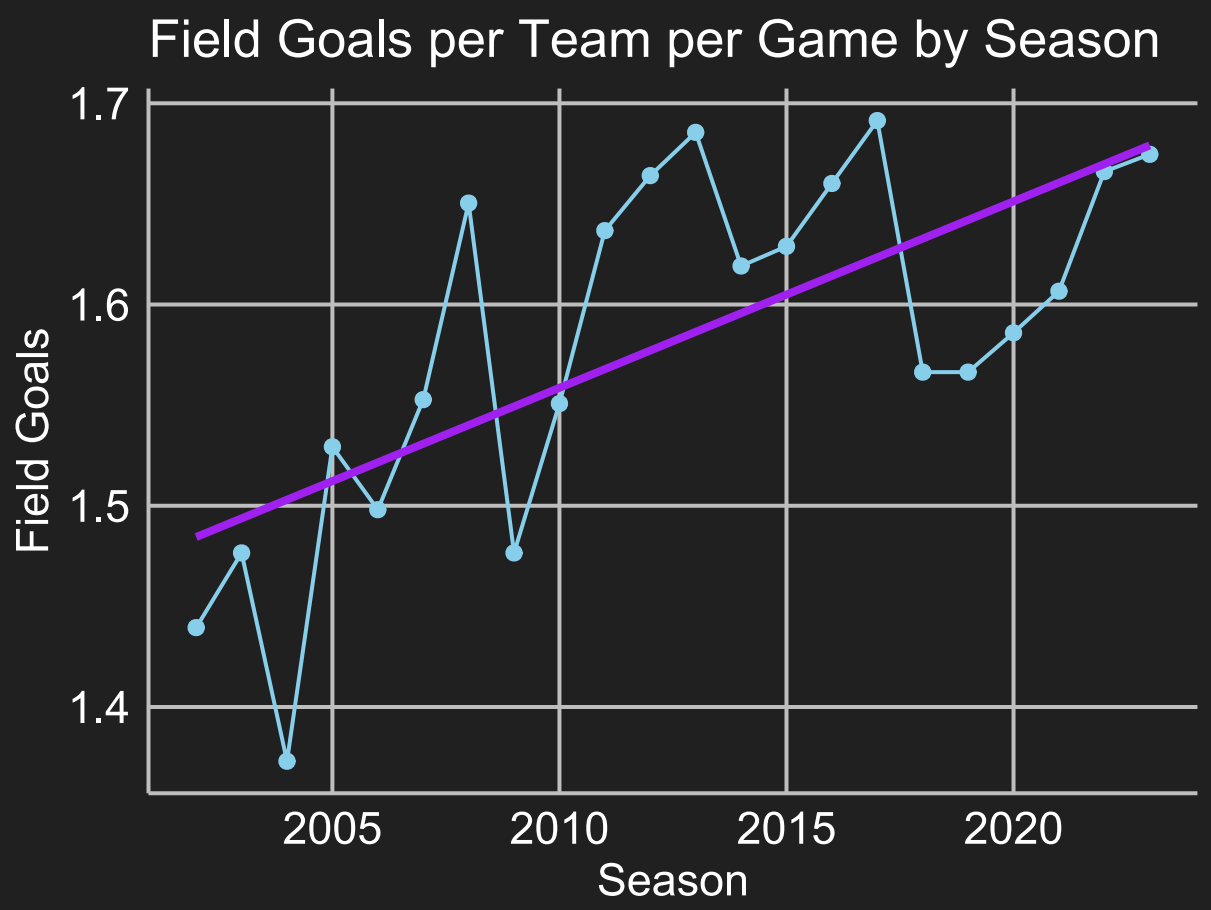
- Three main types of bets:
  - Money Line: A bet on which team will win—returns are based on the likelihood of a given team winning
  - Over/Under: A “50-50” bet on whether the combined score of both teams will be over or under a target
  - Spread: A “50-50” on whether the favorited team will win be a least a given number of points
- With repeated bets, one can expect to lose 3-4.5%

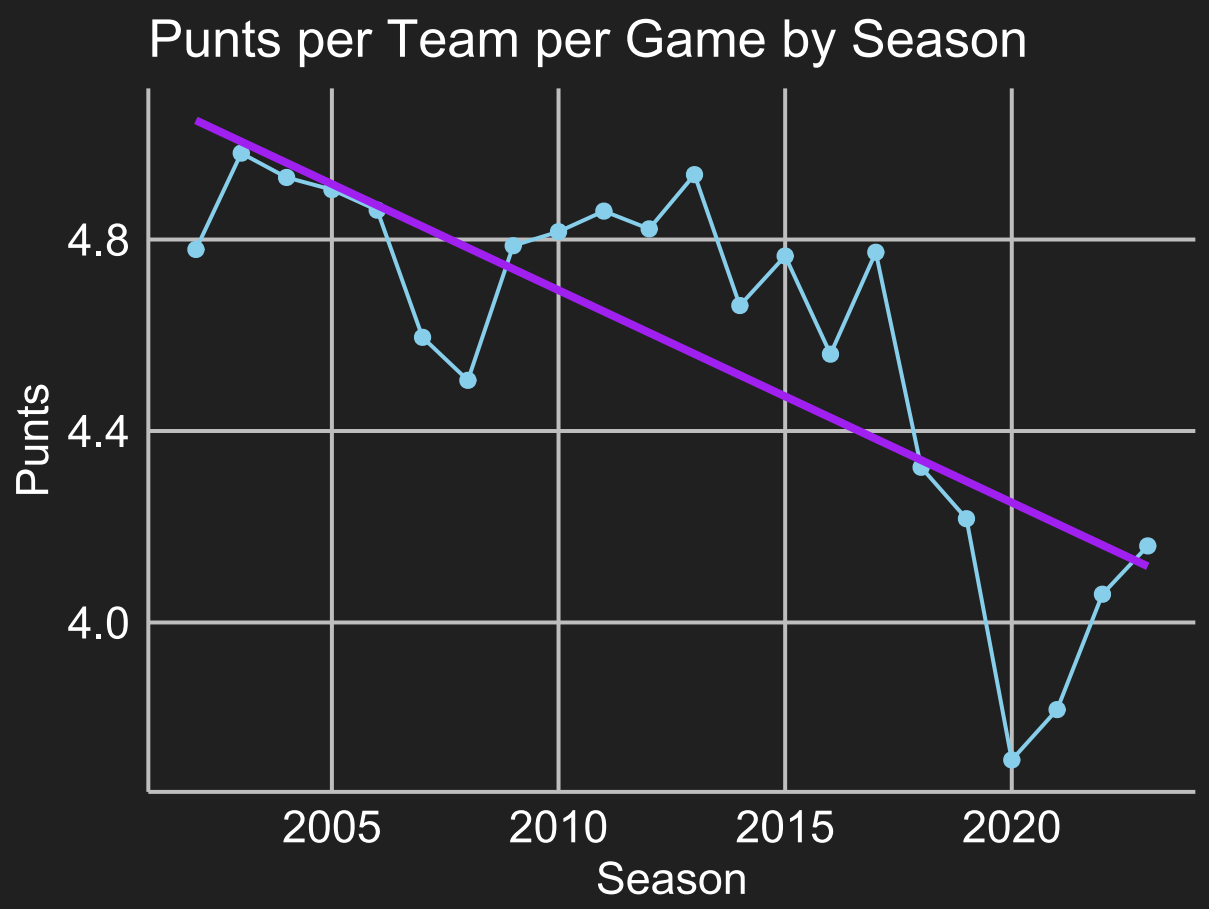
Data

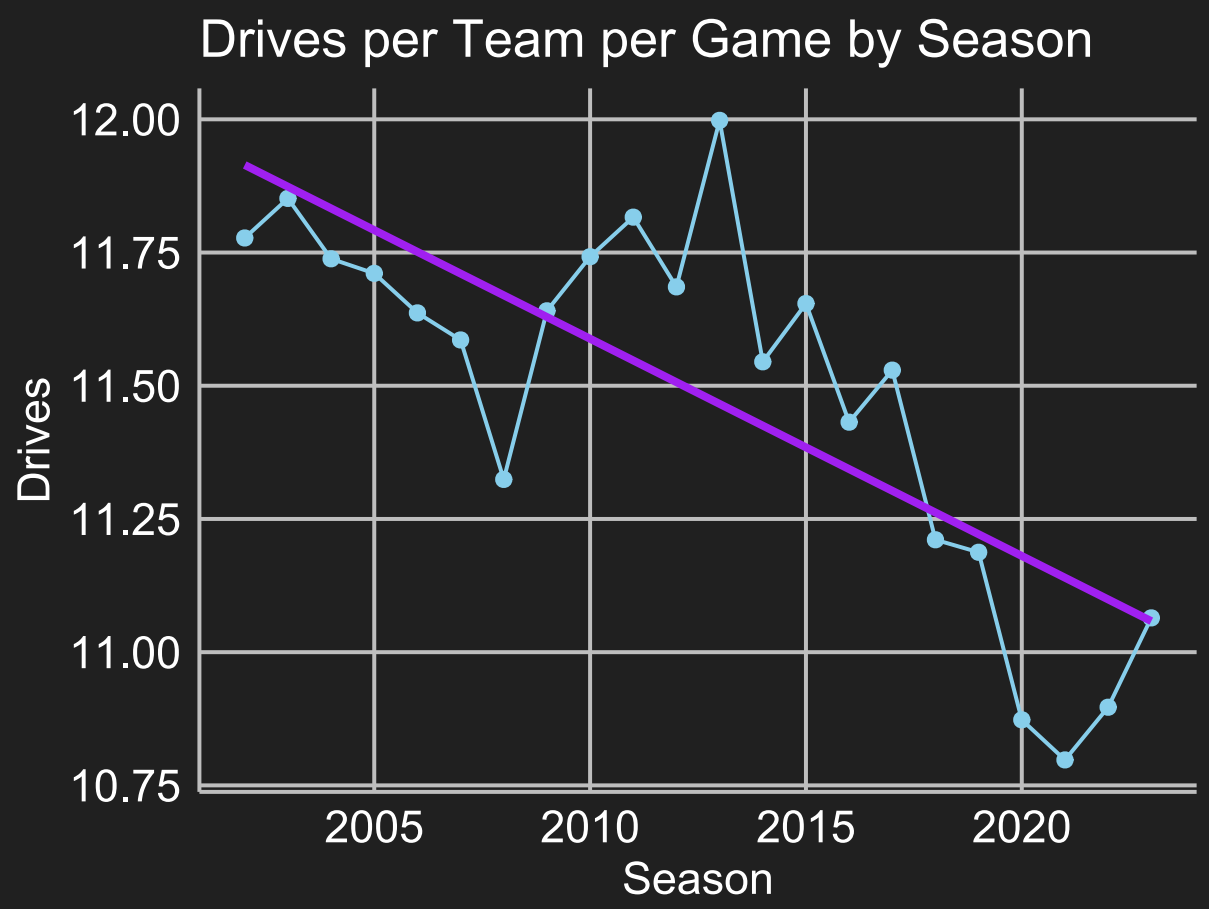
- Drive-level data (offense and defense) on all 32 NFL teams since 2002 (n=704)
  - Touchdowns
  - Field Goals
  - Punts
  - Turnovers
  - Total Drives
- Betting odds and outcomes for all NFL games since 2007 (n=4,861)

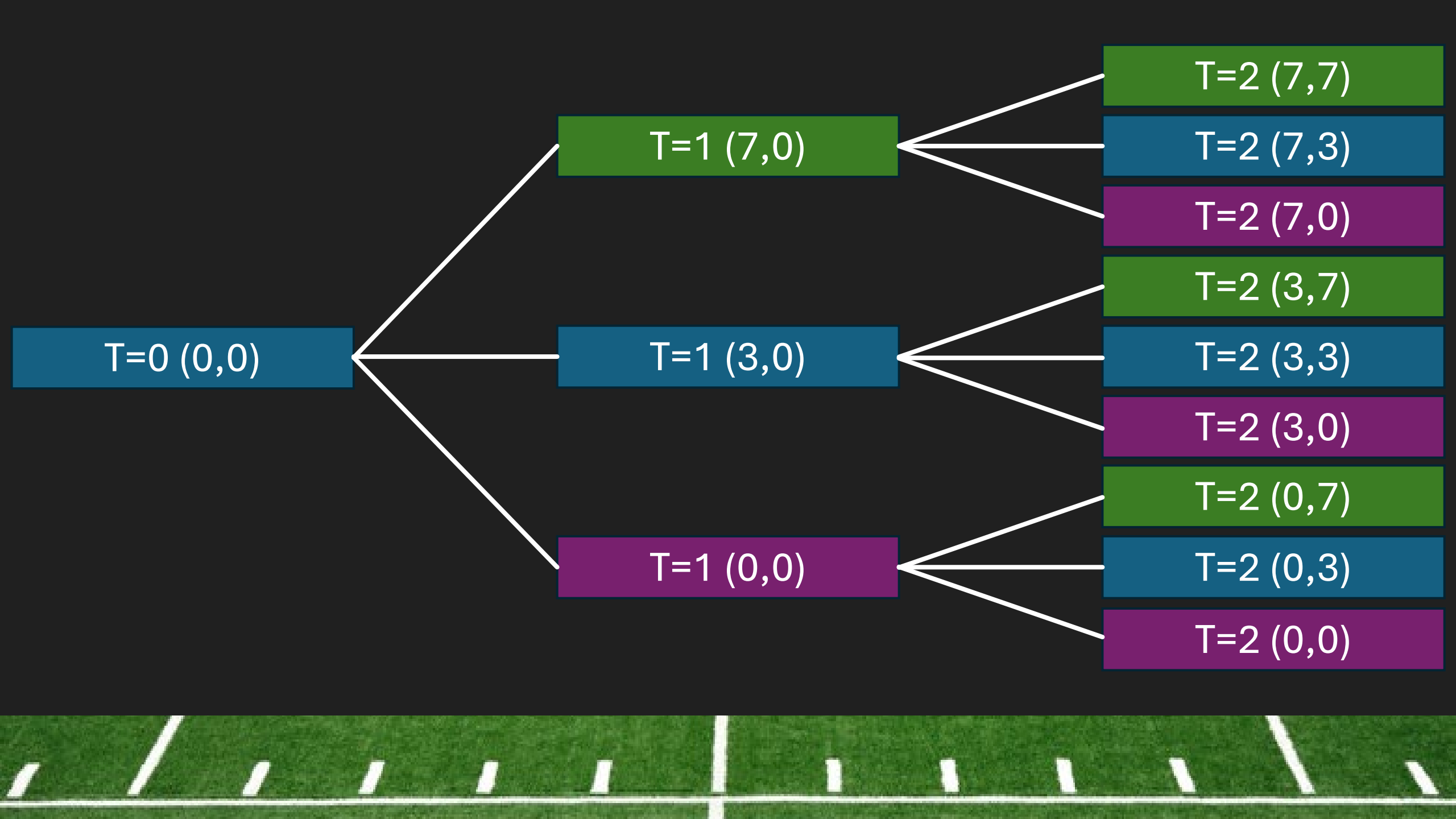



Touchdowns per Team per Game by Season
Touchdowns
2.6
2.4
2.2
2005
2010
2015
2020
Season


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

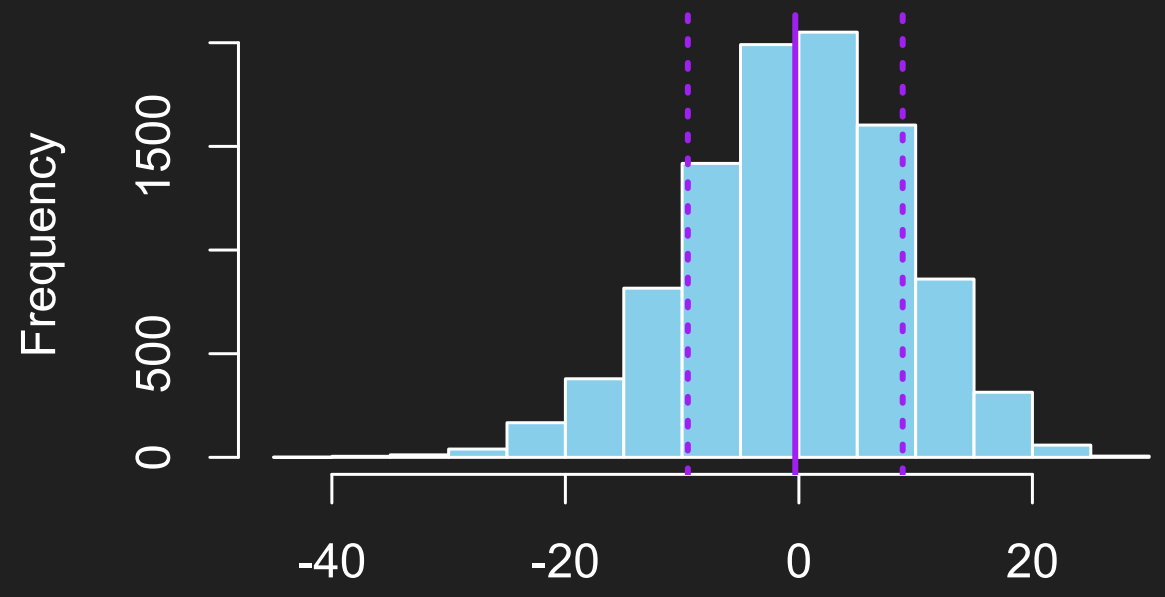

Drives per Team per Game by Season
Drives
12.00
11.75
11.50
11.25
11.00
10.75
2005
2010
2015
2020
Season


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)

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

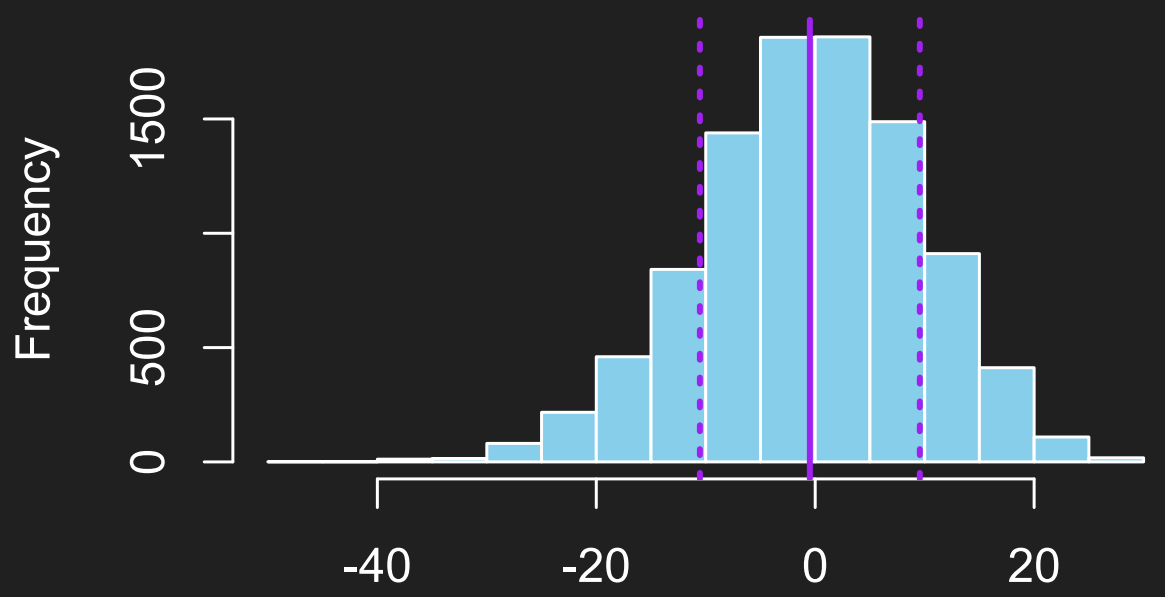
- $td\_per\_drive_t$: Number of touchdowns scored per drive in season *t*
- $playoff_t$: A dummy variable representing whether a team made the playoffs in season *t*
- *season*: The season to account for time trends

Error



($\bar{x}$=-0.31, S_x=9.20, n=9,722)

Errors (Forecasted)



($\bar{x}$=-0.49, S_x=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