NBA Game Simulation

Author: Philip Tanofsky
Course: DATA 604, CUNY Summer 2020
Date: July 16, 2020

Problem and Significance

Problem: Model an NBA game based on available team statistics to evaluate probable winner of a game or playoff series.

- Model a team's possession as a discrete event
- Simulate game by repeating possession model to match team's pace
- Simulate a matchup between teams by simulating the game model for each
- Simulate a playoff series between two teams by simulating the matchup

Significance: Ability to predict the winner of an NBA game or playoff series.

- Coach or general manager could identify specific statistics to improve in order to maximize overall team performance
- Influence player personnel decisions and playing time for some players

Statistics Per Team: System definition

- Turnover percent per play (TO%)
- Field Goal Attempt percent per play (FGA%)
- Offensive Rebound percent per play (OReb%)
- 2-Point Field Goal percent of team's Field Goal attempts (2FGA%)
- 2-Point FG shooting percent (2PFG%)
- 3-Point FG shooting percent (3PFG%)
- Free Throw Shooting Percent (FT%)
- Shooting Foul Drawn Percent per play (SFD%)
- Pace (Average possessions per game)
- Offensive Rating
- Defensive Rating

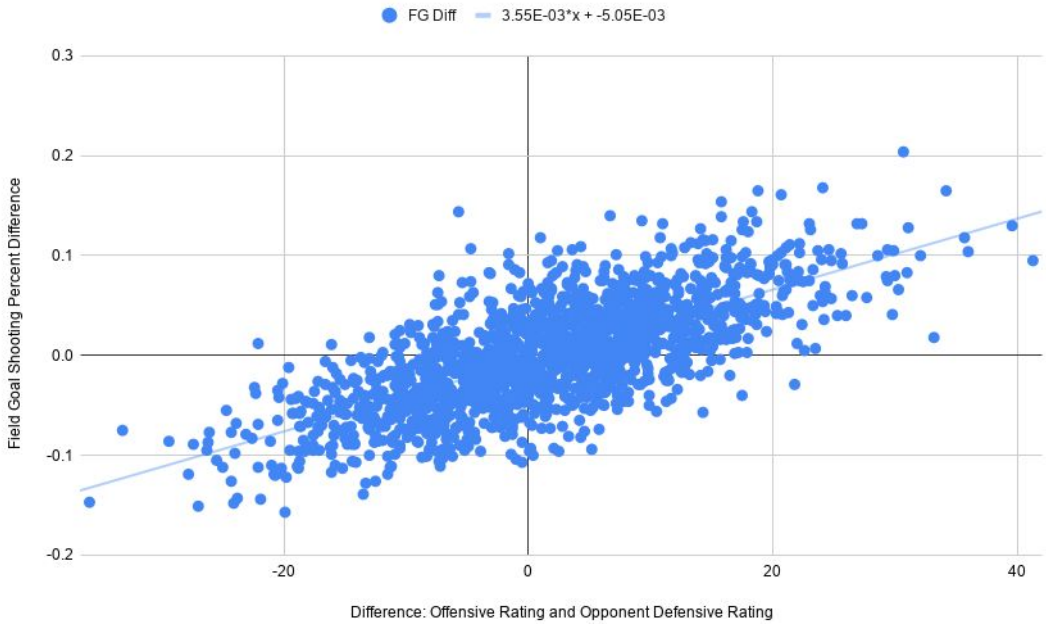
Data sources:
https://stats.nba.com/teams/advanced/
https://www.basketball-reference.com/leagues/NBA_2020.html

Matchup Calculations

Possession calculation relies on team's system variables

To address matchup, calculated linear regression equation in order to adjust below statistics based on difference in offensive rating and opponent's defensive rating

- 2-point FG percentage
- 3-point FG percentage
- Turnover percentage



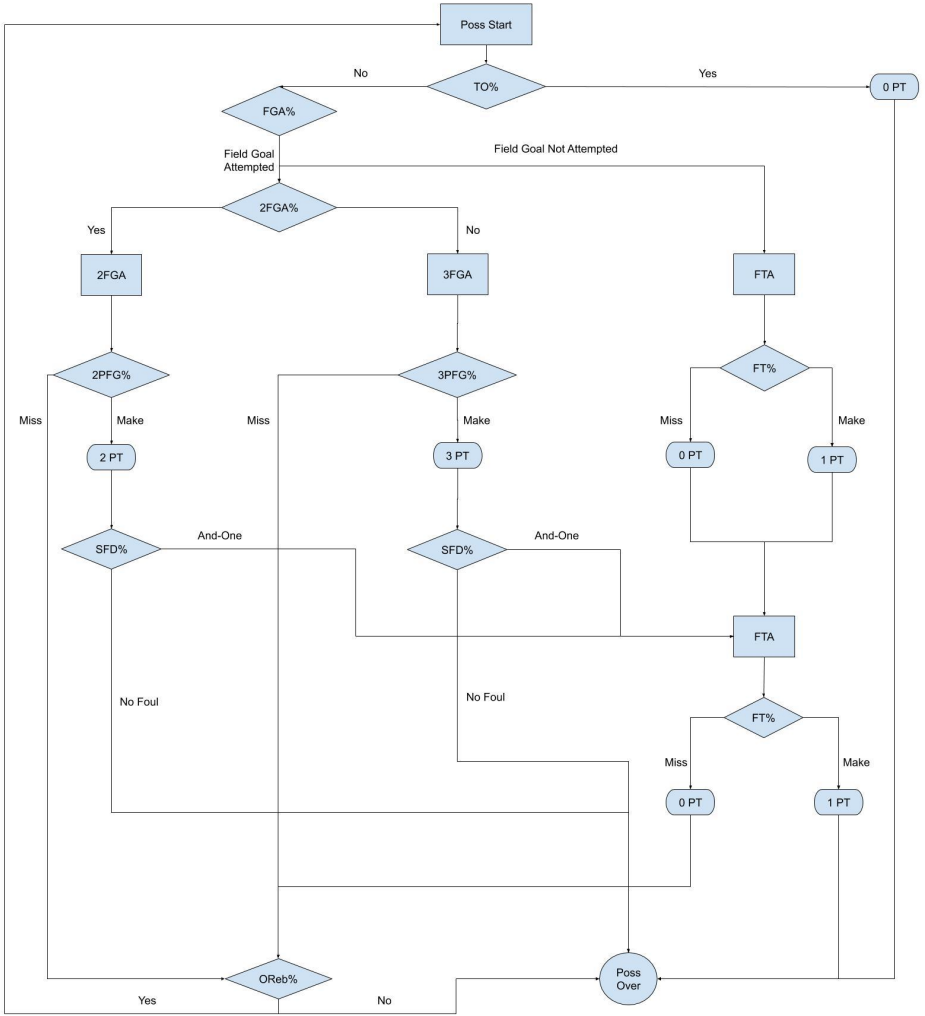
Flow Chart Model

Possession as discrete event

Result of each possession: 0-4 points

Game model repeats possession based on pace of matchup

State of the game tracks the accumulation of the score along with the team statistics



Simulate Process

San Antonio Spurs for the 2019-20 season

- 63 Games
- Average Points Scored: 113.16
- Standard Deviation: 10.72
- Standard Error: 1.35

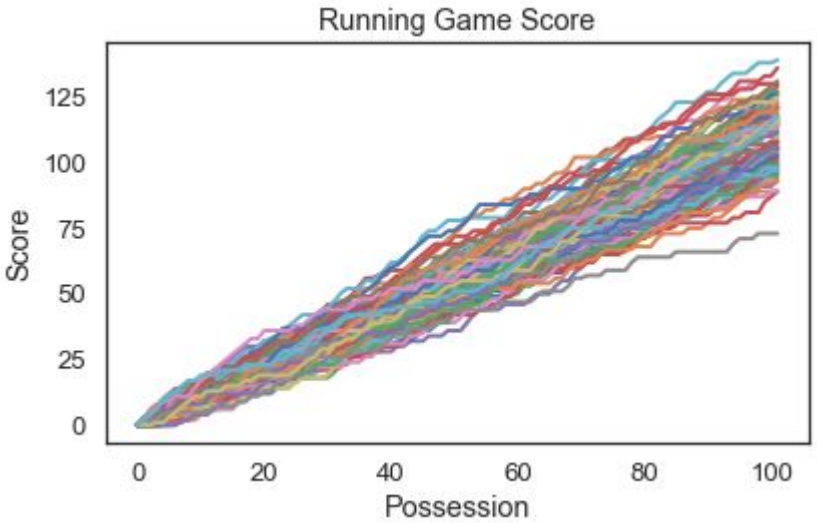
Simulation 100

- 100 Games
- Average Points Scored: 113.26
- Standard Deviation: 11.735
- Standard Error: 1.1735

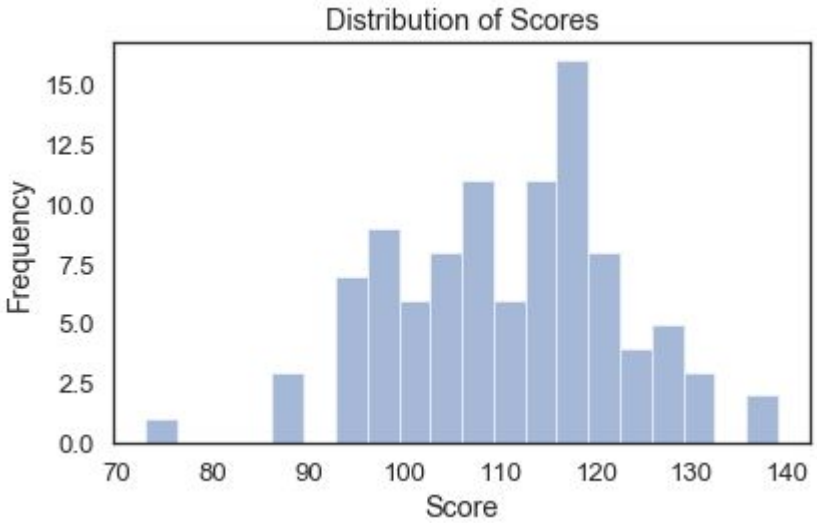
Simulation 500

- 500 Games
- Average Points Scored: 111.73
- Standard Deviation: 11.5464
- Standard Error: 0.516

Game Simulation: Spurs



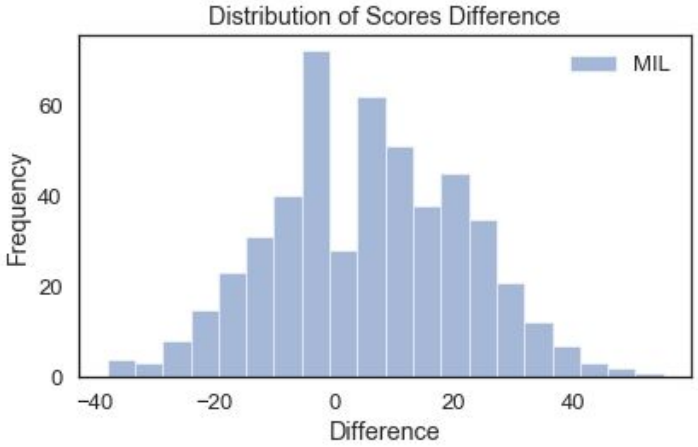
Single team (Spurs) game simulation
Running score for 100 simulations

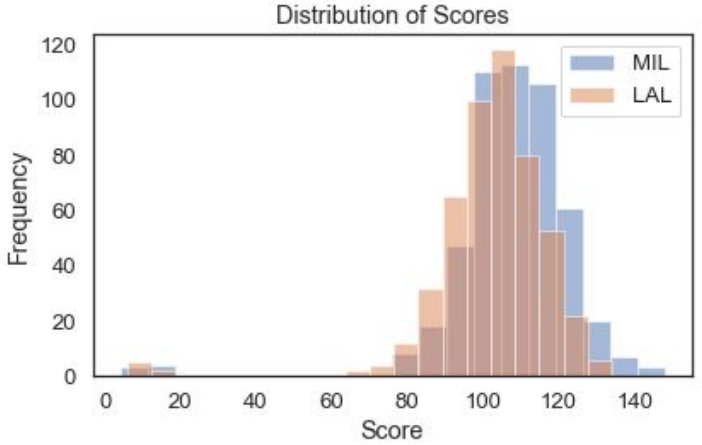


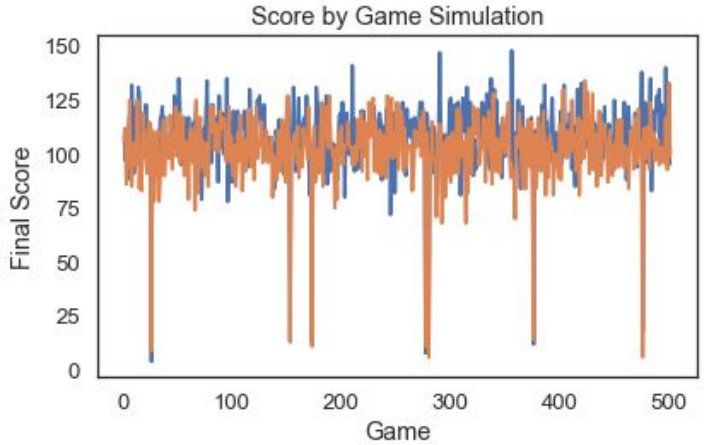
Distribution of simulation final scores
for the 100 game simulations plotted to the left

Matchup: Bucks vs. Lakers

- Games: 501
- MIL wins: 305
- LAL wins: 196
- Overtime games: 7
- Score differential mean: 5.55







Playoff Simulation: 501 games per matchup

MIL	**332**									**300**	**LAL**
ORL	169	**MIL**	**326**					**285**	**LAL**	201	MEM
MIA	225	IND	175					216	OKC	240	UTA
IND	**276**			**MIL**	**311**	**257**	**LAL**			**261**	**OKC**
TOR	**284**			TOR	190	244	LAC			**281**	**LAC**
BKN	217	**TOR**	**277**					**268**	**LAC**	220	DAL
BOS	**254**	BOS	224		**MIL**	LAL		233	DEN	**273**	**DEN**
PHI	247				**323**	178				228	HOU

Total run time: 3682.48 seconds (61.4 minutes)

Verification and Validation

Verification: Performed through coding and testing practices. Small functions tested independently for correct behavior.
Validation: Compared the Spurs 100 game simulation to current season stats

Season averages per game:

- Points: 113.16
- Turnovers: 12.3
- 2P FGAs: 60.8
- 2P FGMs: 31.3
- 3P FGAs: 28.7
- 3P FGMs: 10.7
- FTAs: 22.8
- FTMs: 18.4
- Off Rebounds: 8.8

Simulation averages per game:

- Points: 113.26
- Turnovers: 11.33
- 2P FGAs: 59.18
- 2P FGMs: 31.03
- 3P FGAs: 29.18
- 3P FGMs: 11.2
- FTAs: 23.74
- FTMs: 17.6
- Off Rebounds: 9.52

Difference:

- Points: 0.10
- Turnovers: -0.97
- 2P FGAs: -1.62
- 2P FGMs: -0.27
- 3P FGAs: 0.48
- 3P FGMs: 0.5
- FTAs: 0.94
- FTMs: -0.8
- Off Rebounds: 0.72

Conclusion

- Accurate model for a single team
- Incremental model allows for easy-to-run playoff simulation
- Model relies primarily on offensive statistics
  - Defensive rating of opposing team factored into three statistics
- Matchup simulation identifies better statistical team
- Gaps
  - Outlier scores seem problematic
    - Model does not account for flow of game, particularly blowout
  - Team level statistics instead of player level statistics
  - Would not account for star player not playing
  - Game location not addressed
- NBA statistics are a good predictor of team success