Benjamin Xiao

MSDS 456 Assignment 2

Team: San Antonio Spurs

Question 1: Pythagorean Wins Formula

The data used in my calculation starts from the 2015-16 season up until the current season up to

February 1st, 2020. The adjusted formula for expected wins is:

$$Expected\ Wins\ = Games\ Played\ \times\ \frac{Offensive\ Rating^{10.8}}{Off\ Rating^{10.8} - Def\ Rating^{10.8}}$$

A team's offensive rating is the total number of points scored divided by a rate of 100 possessions. A

team's defensive rating is the total number of points scored by the opponent divided by a rate of 100

possessions. Rate adjusted measures are used in this context because pace differs per team. Finally, the

expected wins are rounded down to an integer.

      The 10.8 value was calculated by looking for the smallest mean squared error when assessing

the past 5 seasons worth of rounded expected wins value and the actuals win total that season. Daryl

Morey's and NBAStuffer's calculations result in exponents, 13.91 and 16.5, respectively.

      By this measure, the Spurs are expected to have won 23 games so far and are underperforming

with a record of 21-26. Using Morey's and NBAStuffer's formulas, we are still underperforming this

season. Those formulas predict 22 games won instead of 23.

Question 2: Player Rank by Position

	Luka Doncic	Bradley Beal	Zach LaVine	Trae Young
Usage %	36.1	32.4	30.8	33.8
Catch & Shoot 3P%	30.2	35.8	41.8	47.8
Assist %	45.9	28.4	21	42
Deflections per G	1.6	1.8	2.1	1.9
FG Diff %	1.3	2.4	-4.6	3.5
Pick & Roll PPP	1.06	0.97	0.9	0.96

These four players are in the league's top 15 players in usage percentage. Using their catch and shoot 3-point percentage, assist percentage, deflections per game, defended difference in field goal percentage, and pick and roll points per possession, we have a good cross-section into their effectiveness on both offense and defense. The data comes directly from NBA.com.

These players' catch & shoot 3-point percentage may be the lowest importance given they have such high usage rates. However, given the Spurs' ball movement heavy system, these players should have a larger portion of their 3-pointers become catch and shoot shots. Deflections and FG difference are measures of defensive ability and activity. Assist percentage and pick and roll PPP help assess each player's ability to create points for himself and for the team given the high usage rates.

Usage rate, deflections, assist percentage, and 3-point percentage are all seasonal data. FG difference and pick and roll PPP are play-by-play statistics.

They are all known as poor defenders confirmed by a positive difference in field goal percentage difference. The only negative difference comes from Zach Lavine which may signify he is an above average defender for his position. One caveat is that none of these players primarily defend the team's best players.

Looking at shooting metrics, Trae Young and Zach Lavine are by far the two best shooters of the group. Doncic's catch and shoot percentage is worrying because the possession will barely end in an expected value of 1 point. Beal's rate still signals he's a good shooter, just not as great as Young and Lavine. A good follow-up would be to look at how many catch and shoot opportunities these players actually generate for themselves.

Doncic adds tremendous value in his playmaking ability with a significant advantage

in pick and roll PPP. He also holds a large assist percentage difference over Beal and Lavine. Playmaking ability is the one area where a lot of the Spurs' roster lacks, so a primary creator is more important to this team.

Ranking:

1. Luka Doncic
2. Trae Young
3. Bradley Beal
4. Zach Lavine

Question 3: Updated Simulation

In this simulation, Luka Doncic will replace DeMar DeRozan within the Spurs roster. The new offensive and defense metrics are calculated first by subtracting DeRozan's rates from the team's rates then adding Doncic's.

Results:

New Offensive Rating: 111.44 (-1.0)

New Defensive Rating: 108.30 (-5.0)

Using the formula in the first section, the new expected wins for this current season is up to 27 wins which equates to a **6-game improvement**.

Question 4: Caveats and Extensions

The simulation used is simple and subject to different levels of expansion. The crux of the calculations is developed on total team statistics that ignore a number of factors. Some include player injury, competition level, and in-season changes. Basketball is a synchronous game among 10 players on both ends of the floor, and the calculations should reflect that. For example, Dwight Powell, a frequent screener for Doncic, was reason injured, and this may affect Doncic's overall metrics. Most of all, these

metrics and simulations completely ignore team composition. They are more top down calculations based off the aggregated team results thus far.

To start, I would investigate more on the player level to explore out what DeRozan does that Donic does not and vice versa. This will tell us more about what can be missing, at what rates, and by what intensity. This will also tell us what can be added to the on-court product and allows us to investigate how the surrounding players may interact differently.

Next, I only used the past 5 years as a result of the ever-changing NBA landscape and for ease of data collection. Ideally, I would scan when the scoring rates stabilized after the zone defense rule change was put in effect after 2000-01 season.

These ratings do not directly adjust for minutes played. The aggregated number of points is one way to think about it, but it does not quantify how many minutes each player is directly involved. The simulation assumes that each player was involved in every possession and every game of the current season. No team has their players play entire games, and this should be included in the model simulation. The second-unit minutes are a significant portion of the total points scored and/or allowed by each team.

The most difficult part to then quantify and value wins. Wins numbers 40-50 differ from 50 and on. The NBA season is long, and the win totals do not scale linearly as evidenced by the Pythagorean formula being an exponential function.

References:

https://en.wikipedia.org/wiki/Pythagorean_expectation

https://www.nbastuffer.com/analytics101/pythagorean-winning-percentage/

https://stats.nba.com/

https://www.basketball-reference.com/leagues/NBA_2020.html