Introduction

The Game of Basketball

Assessing Value of a Player
  - Efficient Players
  - Plus/Minus Ratings
  - Other Metrics

In-Game Decisions
  - 2014 NBA Finals
  - End-Game Strategy

References
The data science revolution has impacted almost every industry – sport is no exception.

These are the main reasons it has come to the forefront in this industry:

- There is an increased number of quantitative analysts looking at the data, using it to objectively answer traditional questions.
- New kinds of data are being collected, and more of it is being shared with the public.

In this talk, we try to highlight some of the ways in which our understanding of the game of basketball (as played in the NBA) has been changed due to data science.

Much of the statistics and formulae are elementary, but nonetheless ingenious; they result in great insights to the game.
1 Introduction

2 The Game of Basketball

3 Assessing Value of a Player
   - Efficient Players
   - Plus/Minus Ratings
   - Other Metrics

4 In-Game Decisions
   - 2014 NBA Finals
   - End-Game Strategy

5 References
5 players on a team.

A game is 48 minutes long, broken into four quarters.

Points are scored by throwing the ball through a hoop. Depending on the distance to the hoop and the context, each bucket is worth 1, 2 or 3 points.

At the end of the game, the team with more points wins.
Several websites provide data from each game in the NBA.

As each season progresses, more and more detailed data is being added.

Recently, there's been a spate of attempts to use this data to make more informed decisions in the game.
Where's the Best Place to Shoot From?

Shot chart, 2016 regular season

Field-Goal Percentage by Zones, 2016
Suppose that we are given a choice between two gambles:

1. A game where we win $2 with probability 0.4, or win $0 with probability 0.6.
2. A game where we win $3 with probability 0.38, or win $0 with probability 0.62.

Which game should we choose to play?
Suppose that we are given a choice between two gambles:
1. A game where we win $2 with probability 0.4, or win $0 with probability 0.6.
2. A game where we win $3 with probability 0.38, or win $0 with probability 0.62.
Which game should we choose to play?

- In gamble 1, our **expected gain** is $0.80.
- In gamble 2, our **expected gain** is $1.14.

In the long run, I will gain $1.14 per game with gamble 2.
The chart on the right shows that our expected return is highest for the 3-pt zones.

Strangely, this was a very recent insight. However, it has been quite dramatically taken up in recent years.

Houston Rockets have taken this to the extreme.
Gambling Intelligently in the NBA

- The chart on the right shows that our expected return is highest for the 3-pt zones.

- Strangely, this was a very recent insight. However, it has been quite dramatically taken up in recent years.

- Houston Rockets have taken this to the extreme.
1 Introduction

2 The Game of Basketball

3 Assessing Value of a Player
   • Efficient Players
     • Plus/Minus Ratings
     • Other Metrics

4 In-Game Decisions
   • 2014 NBA Finals
   • End-Game Strategy

5 References
Who’s the Greatest of All Time?

These images were from the player page of http://stats.nba.com/.

Here’s the argument many people begin with.
Traditional Box Scores

Offensive Numbers:
- Points per Game
- Assists
- Offensive Rebounds

Defensive Numbers:
- Blocks
- Steals
- Defensive Rebounds
Let’s consider the PPG first. Here are the top 10 players from 2016 according to this metric:

<table>
<thead>
<tr>
<th>Player</th>
<th>Team</th>
<th>PPG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Russell Westbrook</td>
<td>OKC</td>
<td>31.6</td>
</tr>
<tr>
<td>2 James Harden</td>
<td>HOU</td>
<td>29.1</td>
</tr>
<tr>
<td>3 Isaiah Thomas</td>
<td>BOS</td>
<td>28.9</td>
</tr>
<tr>
<td>4 Anthony Davis</td>
<td>NOP</td>
<td>28.0</td>
</tr>
<tr>
<td>5 DeMar DeRozan</td>
<td>TOR</td>
<td>27.3</td>
</tr>
<tr>
<td>6 DeMarcus Cousins</td>
<td>TOT</td>
<td>27.0</td>
</tr>
<tr>
<td>7 Damian Lillard</td>
<td>POR</td>
<td>27.0</td>
</tr>
<tr>
<td>8 LeBron James</td>
<td>CLE</td>
<td>26.4</td>
</tr>
<tr>
<td>9 Kawhi Leonard</td>
<td>SAS</td>
<td>25.5</td>
</tr>
<tr>
<td>10 Stephen Curry</td>
<td>GSW</td>
<td>25.3</td>
</tr>
</tbody>
</table>

If you had one remaining place on your team, and you could pick any of the above, would you pick Russell Westbrook with probability 1?
The analytics revolution in basketball has led to an emphasis of efficiency over production.

How about we assess how well a player does with each attempt to score points, not buckets?

Here's an alternative statistic, that measures Individual Scoring Efficiency (ISE):

\[
ISE = \frac{Pts}{FGA + TOV + 0.44 \times FTA}
\]

We can interpret ISE as the number of points generated by a player each time he tries to score.
Now let’s look at those same players, sorted by ISE rating.

<table>
<thead>
<tr>
<th>Player</th>
<th>Team</th>
<th>ISE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isaiah Thomas</td>
<td>BOS</td>
<td>1.12</td>
</tr>
<tr>
<td>Kawhi Leonard</td>
<td>SAS</td>
<td>1.11</td>
</tr>
<tr>
<td>Stephen Curry</td>
<td>GSW</td>
<td>1.09</td>
</tr>
<tr>
<td>Anthony Davis</td>
<td>NOP</td>
<td>1.05</td>
</tr>
<tr>
<td>Damian Lillard</td>
<td>POR</td>
<td>1.05</td>
</tr>
<tr>
<td>LeBron James</td>
<td>CLE</td>
<td>1.04</td>
</tr>
<tr>
<td>DeMar DeRozan</td>
<td>TOR</td>
<td>1.01</td>
</tr>
<tr>
<td>James Harden</td>
<td>HOU</td>
<td>0.99</td>
</tr>
<tr>
<td>DeMarcus Cousins</td>
<td>TOT</td>
<td>0.97</td>
</tr>
<tr>
<td>Russell Westbrook</td>
<td>OKC</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Here we compare the head-to-head rankings of those 10 players using the two statistics.

![ISE vs PPG Rankings](image_url)

**ISE vs PPG Rankings**

- Russell Westbrook
- James Harden
- Isaiah Thomas
- Anthony Davis
- DeMar DeRozan
- Damian Lillard
- DeMarcus Cousins
- LeBron James
- Kawhi Leonard
- Stephen Curry

- Isaiah Thomas
- Kawhi Leonard
- Stephen Curry
- Anthony Davis
- Damian Lillard
- LeBron James
- DeMar DeRozan
- James Harden
- DeMarcus Cousins
- Russell Westbrook
A quick look at the shooting numbers reveals why Westbrook's ISE score is so low.

<table>
<thead>
<tr>
<th>Player</th>
<th>PPG</th>
<th>FGA</th>
<th>2PA</th>
<th>3PA</th>
<th>FTA</th>
<th>TOV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Isaiah Thomas</td>
<td>28.9</td>
<td>1473</td>
<td>827</td>
<td>646</td>
<td>649</td>
<td>210</td>
</tr>
<tr>
<td>2 Kawhi Leonard</td>
<td>25.5</td>
<td>1311</td>
<td>925</td>
<td>386</td>
<td>533</td>
<td>154</td>
</tr>
<tr>
<td>3 Stephen Curry</td>
<td>25.3</td>
<td>1443</td>
<td>654</td>
<td>789</td>
<td>362</td>
<td>239</td>
</tr>
<tr>
<td>4 Russell Westbrook</td>
<td>31.6</td>
<td>1941</td>
<td>1358</td>
<td>583</td>
<td>840</td>
<td>438</td>
</tr>
</tbody>
</table>

He takes about 30% more attempts, and has more than twice as many turnovers as Isaiah Thomas.

He is an inefficient scorer.
A Team Sport

- We are more likely to put up posters of individuals, not teams.
- However, basketball is a team sport. It is teams that win games, not individuals; everyone had a role to play in the victory or defeat.
- So far, we have focused on the individual.
- Let’s turn our attention to contribution of an individual to the team’s performance.
The first statistic that tied a players’ performance to team performance was invented by Ice Hockey analysts.

A player’s +/- statistic is the number of points scored (per 48 minutes, etc.) minus the number of points conceded while the player is on court. This is known as his points differential.

This seems an intuitive way of capturing the contribution of a player that is not present in box-scores.

Here are the top 10 players, as rated by their +/- rating (per 48 minutes) for the regular season in 2016 – 2017:

<table>
<thead>
<tr>
<th>Rk</th>
<th>Player</th>
<th>Tm</th>
<th>Season</th>
<th>G</th>
<th>MP</th>
<th>PTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>JaVale McGee</td>
<td>GSW</td>
<td>2016-17</td>
<td>77</td>
<td>738.6</td>
<td>+20.5</td>
</tr>
<tr>
<td>2</td>
<td>Stephen Curry</td>
<td>GSW</td>
<td>2016-17</td>
<td>79</td>
<td>2638.5</td>
<td>+18.3</td>
</tr>
<tr>
<td>3</td>
<td>Zaza Pachulia</td>
<td>GSW</td>
<td>2016-17</td>
<td>70</td>
<td>1267.9</td>
<td>+16.2</td>
</tr>
<tr>
<td>4</td>
<td>Kevin Durant</td>
<td>GSW</td>
<td>2016-17</td>
<td>62</td>
<td>2070.7</td>
<td>+16.1</td>
</tr>
<tr>
<td>5</td>
<td>Draymond Green</td>
<td>GSW</td>
<td>2016-17</td>
<td>76</td>
<td>2470.8</td>
<td>+15.9</td>
</tr>
<tr>
<td>6</td>
<td>Klay Thompson</td>
<td>GSW</td>
<td>2016-17</td>
<td>78</td>
<td>2648.7</td>
<td>+14.1</td>
</tr>
<tr>
<td>7</td>
<td>Chris Paul</td>
<td>LAC</td>
<td>2016-17</td>
<td>61</td>
<td>1921.3</td>
<td>+14.0</td>
</tr>
<tr>
<td>8</td>
<td>Andre Iguodala</td>
<td>GSW</td>
<td>2016-17</td>
<td>76</td>
<td>1998.0</td>
<td>+12.3</td>
</tr>
<tr>
<td>9</td>
<td>Patty Mills</td>
<td>SAS</td>
<td>2016-17</td>
<td>80</td>
<td>1754.7</td>
<td>+11.1</td>
</tr>
<tr>
<td>10</td>
<td>Blake Griffin</td>
<td>LAC</td>
<td>2016-17</td>
<td>61</td>
<td>2076.4</td>
<td>+10.2</td>
</tr>
</tbody>
</table>
Notice that 7 of the top ten players are from GSW, who had the league best record of 67-15 (81.7%).

No offence to Zaza, but is it fair to consider him the 3rd best player in the league?

Zaza played 1267.9 minutes in the regular season. Let’s break that down a little more:

- During those minutes, the points differential for GSW was +419.
- 1131.9 of those minutes were played with Steph Curry, during which period the points differential was +441.
- This means that in the remaining 136 minutes, the points differential was -22.

- Raw +/− ratings give too much credit to all players on the floor!
It’s clear that we need to analyse the performance of a team with and without a player in order to know his worth.

That’s what APM does:

- Break up all the sequence of plays within a season where the line-ups do not change - there are typically 60,000 such rows.
- Compute the points differential during each such “stint”.
- Assign a value to each player (APM) such that, when I add up all the home players’ values and subtract all the away players’ values on the court, I get a value equal to (or close to) the true observed points differential.
Suppose that for some particular 5 minute stint when Miami played Boston, these were the following lineups:

<table>
<thead>
<tr>
<th>Boston</th>
<th>Miami</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garnett</td>
<td>Bosh</td>
</tr>
<tr>
<td>Bass</td>
<td>Haslem</td>
</tr>
<tr>
<td>Pierce</td>
<td>James</td>
</tr>
<tr>
<td>Bradley</td>
<td>Wade</td>
</tr>
<tr>
<td>Rondo</td>
<td>Chalmers</td>
</tr>
</tbody>
</table>

The true points differential during the game was +15.

Suppose my initial guess for the individual players’ APM is as follows:

<table>
<thead>
<tr>
<th>Boston</th>
<th>APM</th>
<th>Miami</th>
<th>APM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garnett</td>
<td>10</td>
<td>Bosh</td>
<td>5</td>
</tr>
<tr>
<td>Bass</td>
<td>0</td>
<td>Haslem</td>
<td>0</td>
</tr>
<tr>
<td>Pierce</td>
<td>5</td>
<td>James</td>
<td>10</td>
</tr>
<tr>
<td>Bradley</td>
<td>0</td>
<td>Wade</td>
<td>10</td>
</tr>
<tr>
<td>Rondo</td>
<td>5</td>
<td>Chalmers</td>
<td>5</td>
</tr>
</tbody>
</table>

This yields a predicted differential of $30 - 20 = +10$ in favour of Miami; a residual error of +5.
In general, we vary the APM for each player until the total error (over all stints) is minimised.

The error that is used is known as squared error, and minimising this is in fact what we do when we fit a linear regression with Gaussian errors.

However, there are some issues:

- There are some players who play too often with the same player (cf. Zaza).
- If there are 13 players on an NBA team, there are \( \binom{15}{2} = 3003 \) possible lineups. Ideally we want to observe all these lineups for an equal amount of time.
- Not being able to do so means that we are unable to estimate the APM of each player with the same reliability.

An intermediate statistical technique known as ridge regression is used to overcome this issue.
A few years ago, a couple of analysts created a metric known as Real APM. It utilises the regularisation technique, and also adds other variables into the mix. Here’s a list from 2016 – 2017 of the top 10 players:

<table>
<thead>
<tr>
<th>RK</th>
<th>NAME</th>
<th>TEAM</th>
<th>GP</th>
<th>MPG</th>
<th>ORPM</th>
<th>DRPM</th>
<th>RPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chris Paul, PG</td>
<td>LAC</td>
<td>61</td>
<td>31.5</td>
<td>5.20</td>
<td>2.82</td>
<td>8.02</td>
</tr>
<tr>
<td>2</td>
<td>LeBron James, SF</td>
<td>CLE</td>
<td>74</td>
<td>37.8</td>
<td>6.05</td>
<td>1.37</td>
<td>7.42</td>
</tr>
<tr>
<td>3</td>
<td>Stephen Curry, PG</td>
<td>GS</td>
<td>79</td>
<td>33.4</td>
<td>6.74</td>
<td>0.53</td>
<td>7.27</td>
</tr>
<tr>
<td>4</td>
<td>Jimmy Butler, SF</td>
<td>CHI</td>
<td>76</td>
<td>37.0</td>
<td>4.87</td>
<td>1.91</td>
<td>6.78</td>
</tr>
<tr>
<td>5</td>
<td>Nikola Jokic, PF</td>
<td>DEN</td>
<td>73</td>
<td>27.9</td>
<td>4.45</td>
<td>2.27</td>
<td>6.72</td>
</tr>
<tr>
<td>6</td>
<td>Rudy Gobert, C</td>
<td>UTAH</td>
<td>81</td>
<td>33.9</td>
<td>0.54</td>
<td>6.08</td>
<td>6.62</td>
</tr>
<tr>
<td>7</td>
<td>Draymond Green, PF</td>
<td>GS</td>
<td>76</td>
<td>32.5</td>
<td>1.58</td>
<td>4.92</td>
<td>6.50</td>
</tr>
<tr>
<td>8</td>
<td>Kawhi Leonard, SF</td>
<td>SA</td>
<td>74</td>
<td>33.4</td>
<td>5.55</td>
<td>0.87</td>
<td>6.42</td>
</tr>
<tr>
<td>9</td>
<td>Russell Westbrook, PG</td>
<td>OKC</td>
<td>81</td>
<td>34.6</td>
<td>6.76</td>
<td>-0.47</td>
<td>6.29</td>
</tr>
<tr>
<td>10</td>
<td>Kyle Lowry, PG</td>
<td>TOR</td>
<td>60</td>
<td>37.4</td>
<td>4.73</td>
<td>1.28</td>
<td>6.01</td>
</tr>
</tbody>
</table>

This is more believable than raw plus/minus ratings.
1 Introduction

2 The Game of Basketball

3 Assessing Value of a Player
   • Efficient Players
   • Plus/Minus Ratings
   • Other Metrics

4 In-Game Decisions
   • 2014 NBA Finals
   • End-Game Strategy

5 References
Defensive Stops

- The numbers and events tracked have traditionally been biased towards offensive ones.
- Blocks and steals are spectacular, but they are not good summaries of whether a player is a good defender.
- Many players defend well by even preventing their man from shooting or forcing him to shoot without a good look at the basket.
- Defensive Stops Gained (DSG) is a more accurate reflection of a player’s defensive capabilities. It combines the change in
  - Opponent points
  - Opponent rebounds
  - Opponent turnovers

when the player is on/off the court.

- With spatial tracking data, it is even possible to measure the opponent field goal percentage, when the player is within 5 feet of the hoop, and the shot is taken within the paint.
Miscellaneous Metrics

- **Individual Playmaking Efficiency**: This is similar to ISE, but it tracks the points generated and the number of assists, 2nd assists, FTA and TOV.

- **Player Versatility Index**: How does each player score? Is Curry more likely to execute a pull-up shot or drive (46% to 18%)? How about LeBron? (17% to 22%).

- **Point Balance**: This assesses how the points scored by a team are distributed. Over the period 2006 to 2013, the most unbalanced team was Miami - the one with the big three. The next off-kilter teams were OKC with Durant and Westbrook.
1 Introduction

2 The Game of Basketball

3 Assessing Value of a Player
   - Efficient Players
   - Plus/Minus Ratings
   - Other Metrics

4 In-Game Decisions
   - 2014 NBA Finals
     - End-Game Strategy

5 References
In 2013, MIA played SAS for the NBA title in a best-of-seven series.

After three games, SAS was leading the series 2-1. The margin of victory in the third game was 36 points.

Yet MIA went on to win the title. What was significantly different between the first three games and the last three?
It is clear that LeBron contributed more in the last four games, but why was he so poor in the first three, and how did his team help to turn it around?
It is clear that LeBron contributed more in the last four games, but why was he so poor in the first three, and how did his team help to turn it around?
Primary Changes That MIA Made

- SAS had been crowding the paint in Games 1 to 3, preventing James from driving to the basket efficiently.
- In Game 4, MIA replaced Udonis Haslem with Mike Miller, who shot 0.417 on 3FGA.
- At other times, Ray Allen, the all-time leader of 3FG, replaced Dwayne Wade.
Effect of More Floor Spacers

<table>
<thead>
<tr>
<th>Floor Spacers</th>
<th>CHAO</th>
<th>CHAD</th>
<th>PPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>43.80</td>
<td>14.70</td>
<td>0.39</td>
</tr>
<tr>
<td>3</td>
<td>43.40</td>
<td>16.70</td>
<td>0.77</td>
</tr>
<tr>
<td>4</td>
<td>50.20</td>
<td>20.70</td>
<td>1.30</td>
</tr>
</tbody>
</table>

- CHAD: Convex Hull Area of Defence
- CHAO: Convex Hull Area of Offence
- PPP: Points Per Possession
1 Introduction

2 The Game of Basketball

3 Assessing Value of a Player
   • Efficient Players
   • Plus/Minus Ratings
   • Other Metrics

4 In-Game Decisions
   • 2014 NBA Finals
     • End-Game Strategy

5 References
Go For the Win or Draw?

- Suppose we are two points down, and we have the ball for the last possession. The game will end after our shot.
- Should we try for a 2-pointer that ties the game (sending it into overtime), or should we try for a 3-pointer?
Suppose we are two points down, and we have the ball for the last possession. The game will end after our shot.

Should we try for a 2-pointer that ties the game (sending it into overtime), or should we try for a 3-pointer?

In general, the probabilities of scoring a two-pointer and three-pointer are 0.45 and 0.33.

If we assume that over-time could go either way with equal probability, then the probability winning in overtime is

$$0.45 \times 0.5 = 0.225 < 0.33$$

Go for the Win!
Books

Websites