Team on My Back Index: a new metric for NBA Player Value

Evan Mickas, emickas@stanford.edu

Introduction

- There are many reasons why measuring the value of NBA basketball players is useful, from making salary-cap decisions to selecting All-Star teams and award winners.
- Existing metrics fail to identify players who can serve as team centerpieces, because they estimate how much more successful a player's team is than they would be with a league average player filling in.
- The proposed Team on My Back Index instead examines hypothetical lineups consisting of 1 real-life player paired with 4 imaginary average players.

Organizing the Data

- A lineup consists of five players on the court for a given team. To measure lineup success, we use point differential per 48 minutes, because the objective of a basketball game is to score more points than the opponent within 48 minutes.
- We record 43 statistics per player, ranging from conventional points/assists/rebounds to advanced metrics.
- Our task, put more specifically, is to **use these individual statistics, concatenated together, to predict point differential.**
- For input/output data, we find the point differential (and constituent players) of the 600 most-played lineups from the 2019 NBA season.
- Before training a model, a few preprocessing steps were required:
    1. Correction for players who shared the same first initial and last name for the sake of pairing lineup numbers to individual numbers.
    2. De-meaning individual statistics by subtracting out the statistics of the average player (weighted by minutes played)
    3. Adding every possible permutation of the lineup to our set of input/output pairs. This is to prevent the ordering of the players from having an effect on the model's learning.

Defining TOMBI

We define a player's TOMBI as follows:

$$TOMBI(P) = \frac{\sum_{j=0}^{4} f((\mathbf{0}_{43j}, P - A, \mathbf{0}_{43(4-j)}))}{5}$$

Where P is the 43-vector of the player's stats, A is the average player, $\mathbf{0}_i$ denotes a zeros vector of length j, and f is our yet-unspecified predictor.

- TOMBI represents the mean predicted point differential of all lineups pairing the query player with 4 average players.

Finding a Model

- Our ideal predictor should have a few characteristics:
    - It should be relatively simple. Many of the statistics used are closely related to each other to the point of redundancy, so our model should identify the most important variables.
    - It should deal well with noise. Factors like the opposing lineup, home-court advantage, and refereeing all contribute noise, making it difficult to predict lineup success.
- I partitioned the lineup data 80%-20% into training and validation sets and trained *scikit-learn*'s Linear Regression, Multi-Layer Perceptron, and Lasso models.
- I chose the Lasso model. It performs basic variable selection and makes reasonable predictions (despite noise) because of its linear form.
- I set alpha equal to 0.1 for feature selection
- One additional constraint I made was that each coefficient should be positive, because almost every individual statistic I recorded correlates positively with team success.
- This model didn't perform spectacularly well in terms of lineup predictions, with validation mean squared error of 254.01, but it is robust enough for calculating TOMBI.

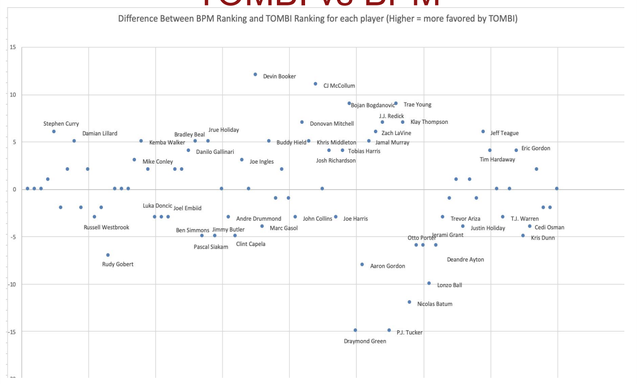
Results

- Results (calculated for players averaging >30 minutes in >41 games) correspond very well with intuition.
- The top two TOMBI scorers were the frontrunners for the 2019 MVP award.
- Each of top 8 players is a franchise centerpiece.
- Players at the bottom of the rankings are low-tier role players.
- TOMBI correlates well with Box Plus-Minus (one of the leading existing metrics for player value), with a correlation coefficient of 0.98.
- The differences between TOMBI and BPM rankings reflect TOMBI's goal of assessing a player's aptitude for team leadership roles.
- For example, Devin Booker's TOMBI is 35th-highest, while his BPM ranking is 47th. This makes sense because he plays a leadership role for his team.
- On the other hand, Draymond Green's TOMBI ranking is 15 spots lower than his BPM ranking because he plays more of a support role for his team.

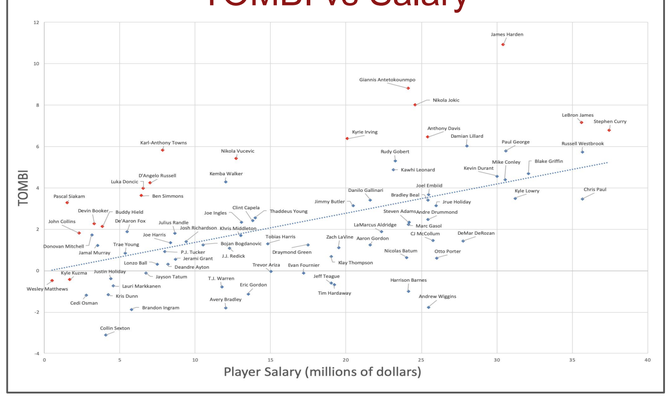
TOMBI Leaders (>30 minutes, >41 games)

1) James Harden (10.911)
2) Giannis Antetokounmpo (8.803)
3) Nikola Jokić (8.001)
4) LeBron James (7.144)
5) Stephen Curry (6.789)
6) Anthony Davis (6.454)
7) Kyrie Irving (6.386)
8) Damian Lillard (6.051)

TOMBI vs BPM



TOMBI vs Salary



An Application: Finding Optimal Lineups

- Having calculated TOMBI values for each player, we can consider the problem of fielding the best possible lineup given a budgetary constraint.
- This is a CSP problem in which we maximize the summed TOMBI of our players under the constraint of their summed salaries.

TOMBI-Optimal Lineups At Different Salary Caps