The Cause In Basketball Skill

Daniel Naylor
University of California Santa Barbara

Introduction

Within sports analytics, one of the most important goals is to develop ways to measure skill and performance of players. These statistics are used to inform very important decisions, such as negotiating contracts with current players or forming teams from new talent.

This is especially true in basketball. Anyone who watches NBA games would be able to tell you that LeBron James, Nikola Jokic, and Stephen Curry are among the greatest players. And of course, instead of basing this on subjective and vague ideas, we can turn to statistics to quantify their skill.

Before a proper discussion can be had, it's important to define the following basketball terminology

- **Period**: One of four quarters of a game. Each period lasts 12 minutes (so a full NBA game is 48 minutes).
- **Possession**: A period of time where the ball remains in possession by a single player. Anything that causes the ball to change holders (e.g. passing, foul, scoring, rebound) marks a new possession.
- **Stint**: A period of time where the players on the court remain constant. A stint ends when either a new player enters the court through a substitution, or when the period ends. A game can be understood as a collection of stints.
- **Score Margin**: $\text{Home Score} - \text{Away Score}$. By convention, a positive score margin means the home team is in the lead, while a negative score margin means the away team is in the lead.

The Base Skill Metric: RAPM

The *Regularized Adjusted Plus Minus* (RAPM) is a commonly used metric for reporting the skill of a player. While it isn't often directly reported to decision-makers or fans, it is a very important metric that forms the foundation for basketball analysis.

RAPM comes from a raw player statistic called Plus-Minus. The Plus-Minus is defined as $\text{Points Scored For Player's Team} - \text{Points Scored Against Player's Team}$ **while the player was on the court**. For example, if Kevin Durant's team scored 9 points while he was on the court, and his opponents scored 3 points in the same period, then his Plus-Minus would be 6.

While Plus-Minus only considers points the player could have had an effect on, a major drawback is it's *highly variable* and *not exactly reflective of skill.* A mediocre player will have a high Plus-Minus if their teammates are LeBron James, Stephen Curry, Nikola Jokic, and Kevin Durant. Likewise, an excellent player will have a low Plus-Minus if their 4 teammates are newbies.

This is precisely what RAPM is aiming to address: it *regularizes* each player's Plus-Minus. To explain how RAPM is calculated, define

- s_i: The score margin of stint i. This only includes points scored in the stint, **not** the entire game. s_i is typically standardized to an ideal stint which has 100 possessions.
- β_j: The RAPM of player j.
- p_{ij}: An indicator of player j's presence in stint i. Specifically:
  - $p_{ij} = 1 \iff$ Player j is present on Home in stint i.
  - $p_{ij} = 0 \iff$ Player j is not present in stint i.
  - $p_{ij} = -1 \iff$ Player j is present on Away in stint i.

Then, we set up the linear system

$$\begin{aligned} s_1 &= \beta_1 p_{11} + \beta_2 p_{12} + \cdots + \beta_n p_{1n} \\ s_2 &= \beta_1 p_{21} + \beta_2 p_{22} + \cdots + \beta_n p_{2n} \\ &\vdots \\ s_N &= \beta_1 p_{N1} + \beta_2 p_{N2} + \cdots + \beta_n p_{Nn} \end{aligned}$$

Where there are N stints and n players. From this, we turn to linear regression to find $\vec{\beta}$. RAPM in particular uses Ridge regression to compute $\vec{\beta}$. There are a few important things to note here.

1. Each β_j can be interpreted as the amount of points player j adds to their team's score per (ideal) stint.
2. By using linear regression, RAPM is making a number of statistical assumptions. The important one to highlight is **it assumes similar environments** (i.e. you can only use RAPM to model situations similar to the known data). *Relaxing this assumption is the goal of this project.*
3. RAPM is inherently Bayesian. By using Ridge regression, we shift the coefficients towards a prior belief. In practice, we use cross-validation to choose hyper-parameters (and by extension, choose our prior).

This project uses the 2022–23 NBA season to compute RAPM, which consists of $N = 32384$ stints and $n = 540$ players. We used Leave-One-Out cross-validation and settled on the hyperparameter $\alpha = 5000$. Stints were weighted by number of possessions.

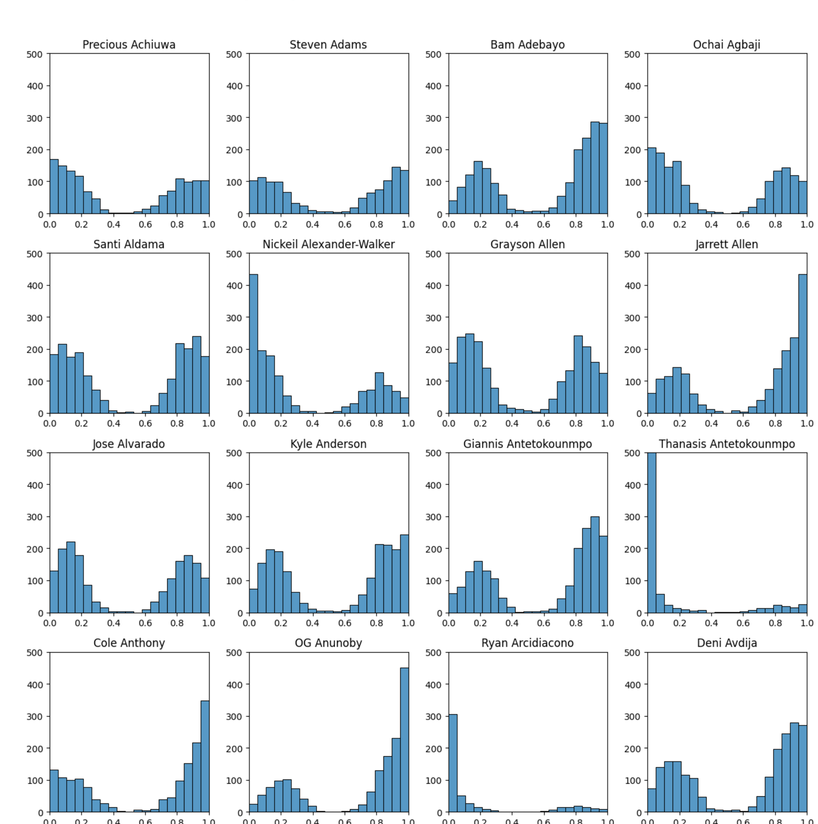
Player Name	RAPM
Nikola Jokic	3.325439
Draymond Green	3.308574
Joel Embiid	3.211865
Jrue Holiday	3.001383
Kawhi Leonard	2.876934

A New Metric: The Causal Effect

The similar environment assumption of RAPM is too strong: this is violated as soon as a player changes teams. To address this, we developed a new framework motivated from causal inference. First, we fix a player and characterize stints as a population that receives treatment. Our framework says that

- The **treatment** (untreatment) is the presence (absence) of the player. *We only consider stints where the player could have been present.*
- The **outcome** of the treatment (or lack thereof) is the score margin of the stint (suitably standardized to 100 possessions).

Under this framework, we require **propensity scores** to produce metrics. The propensity score of stint i, p_i, is defined as the probability of receiving treatment. Below is a collage of propensity scores for a handful of players.



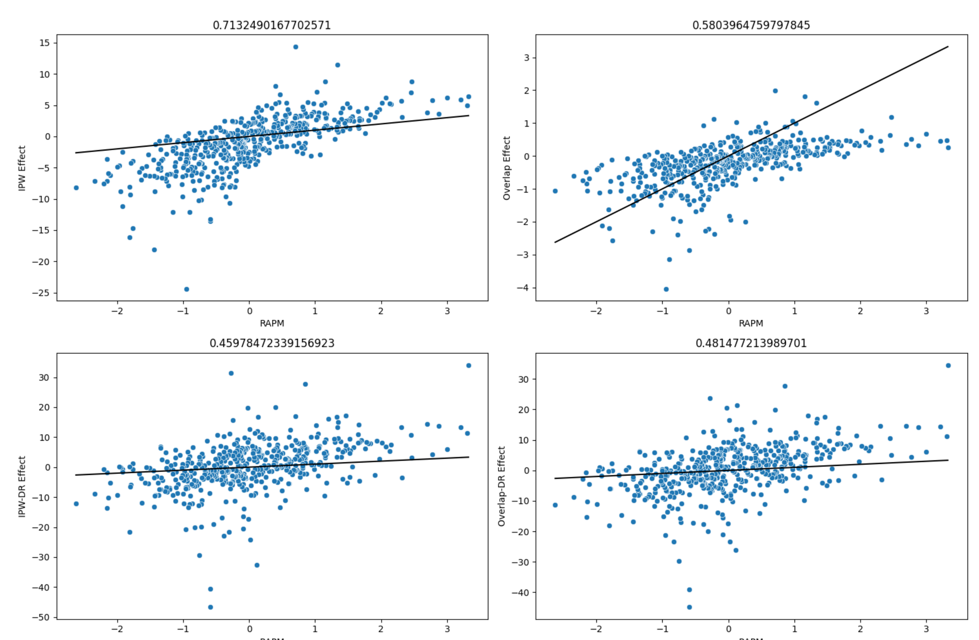
Implementation and Results

To produce metrics, we produced 4 estimators using **propensity weighting** and **doubly robust models** for our estimators. Namely

1. **IPW Effect**: Inverse-propensity weighting. *This up-weighs stints that are "surprising", i.e. weighs stints by the inverse probability of being correctly classified.*
2. **Overlap Effect**: Overlapping weights. *This up-weighs stints that have around 50% probability, i.e. stints that are difficult to classify.*
3. **IPW-DR Effect**: Doubly-robust estimator for IPW Effect.
4. **Overlap-DR Effect**: Doubly-robust estimator for Overlap Effect.

These require models for p_i, score margin given treatment, and score margin given absence. We settled on 100-tree random forests. For information on how these models were trained and validated, see Acknowledgments.

Our metrics are fairly correlated with RAPM. Each dot is a player, and an identity line is included in each figure.



And finally, we summarize our model performances against the 2023–24 season below. We evaluated performance by computing the RMSE of the predictions obtained by the linear model presented in "The Base Skill Metric: RAPM". We obtained an RMSE for players who changed teams, and reported the average across these players (weighted by number of possessions player was present in).

Null RMSE	44.97
RAPM RMSE	44.89
IPW RMSE	46.25
Overlap RMSE	44.91
IPW-DR RMSE	49.36
Overlap-DR RMSE	49.05

Note that this validation scheme is quite ad-hoc, since our metrics have no connection to the linear model. We are still coming up with a scheme that is sound for our metrics.

Acknowledgements and Citations

I would like to thank my mentor, Dr. Alexander Franks, for his guidance in this project. In addition, I would also like to thank my friend, Aarya Kulkarni, for some initial support with pre-processing and model building.

Our work and detailed results can be found at `github.com/DanLeEpicMan/RAPM`. I would like to acknowledge Ron Yurko for his scripts that aided my pre-processing, and `basketball-reference.com` for their publicly available statistics about players.

Thanks to Professor Don Aue and the donors to The Donald Aue Summer Undergraduate Research Fund at CCS for supporting my 2024 Summer Undergraduate Research Fellowship.