**Game of Trades: Using simulation to get an edge in the NBA trade market**

Basketball

ID: 5643

1. **Abstract**

In this paper I present a method to evaluate the impact of trades for NBA teams by simulating different scenarios. Assessing the potential impact of trades is a complex task as it affects not only the composition of the teams involved but also the rest of the league. For instance, the Kyrie Irving – Isaiah Thomas trade not only affects the number of wins for Cleveland and Boston but also for any team playing against these 2 teams.

The main idea is to estimate the offensive and defensive power for each team given a team roster and players’ usage. Assuming a Normal distribution in teams’ scored points, these powers serve as estimated means while the variance for both models is estimated from empirical data. Once the probability distribution is known, I can calculate the probability of any matchup and thus the number of wins in the regular season.

I will show how this model has implications beyond the trade market evaluation.[[1]](#footnote-1)

1. **The Data**

For the main model and further analysis I used the following data variables available from basketballreference.com:

"Age" "G" "GS" "MP" "FG" "FGA" "FG." "X3P"

"X3PA" "X3P." "X2P" "X2PA" "X2P." "eFG." "FT" "FTA" "FT." "ORB" "DRB"

"TRB" "AST" "STL" "BLK" "TOV" "PF" "PTS" "Season"

Once the column stats have been adjusted to per minute stats, I use the following nomenclature:

"Age" "Exp" "FGPer" "FG3Per" "FG2Per"

"effFGPer" "FTPer" "effMin" "effFG" "effFGA" "eff3PM" "eff3PA" "eff2PM" "eff2PA"

"effFTM" "effFTA" "effORB" "effDRB" "effTRB" "effAST" "effSTL" "effBLK" "effTOV"

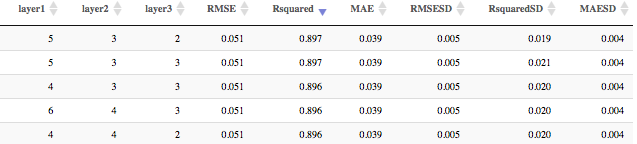
"effPF" "effPTS"

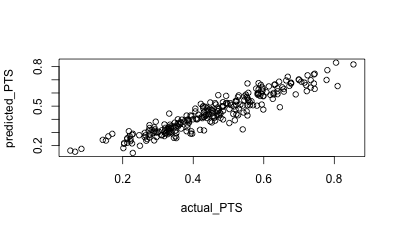
1. **The Model**

The model consists of 2 Neural Networks used to estimate team powers: One for Offense (points scored), one for Defense (points against). Inputs are players' projected per minute stats weighted by their share of minutes of play.

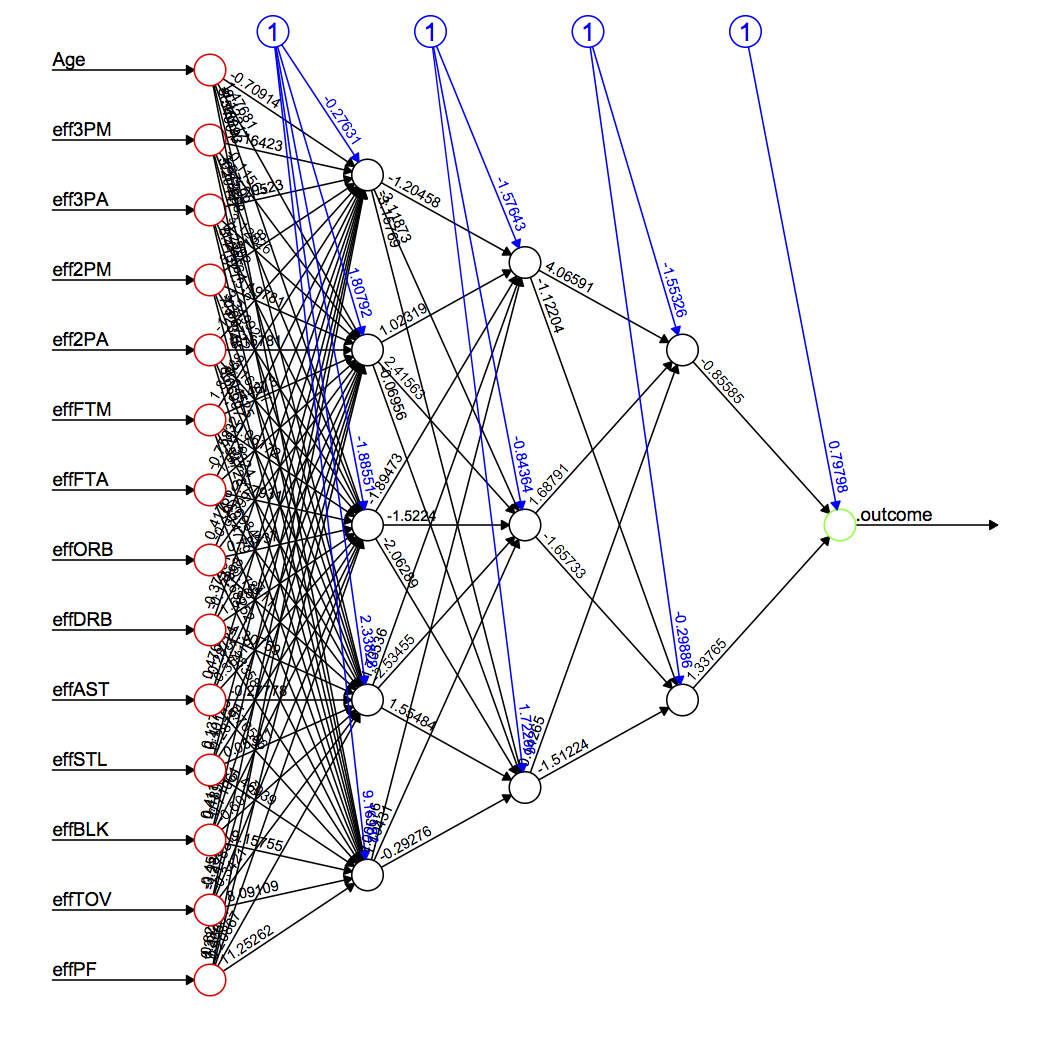
Steps:

* Read historical players: From Season 1979-1980 when the 3 point shoot was established. (write\_playersHist.R) (data: playersHist.csv) . Size: 18927 records.
* Differentiate players with the same name by adding a number after the name in ascending order by decreasing age. (.rename\_PlayerName\_Duplicates.R). Example: Tim Hardaway who played in the 90s and Tim Hardaway 2 (current NYK player) as basketballreference does not differentiate them and I use Player as primary key.
* Calculate stats per minute of play: effStat = Stat/MP. And effMinutes as Minutes played over total possible minutes: 82 games \* 48 minutes per game: effMin = MP/3936 (.team\_prepareAll() in prepare\_rosters.R). Finally, adjust effMin relative to total minutes played by team. effMin = effMin/sum(effMin)
* The input vectors for the neural network is the weighted average of all stats per team per season. Size: 1063 records.
* Remove columns that depend linearly on others: FG, FGA, FG%,3P%,2P%,FT%,effFG%, effPTS.
* Scale the data [0,1] for easier convergence of backpropagation algorithm.
* Split sample in 75-25% training-testing. Used 10-fold cross-validation with 10 repetitions (leave one out).
* Train a neural network with 3 layers: nnetGrid <- expand.grid(layer1 = c(4,5,6), layer2 = c(3,4), layer3 = c(2,3)). Used neuralnet package under caret package in R.
* Run the network twice: Same inputs, different outputs: Average points scored and average point scored ag ainst per team per season.
  1. **Neural Network model**
     1. **Offense**

Best Offense model (5-3-2) based on R-squared: top 3 results

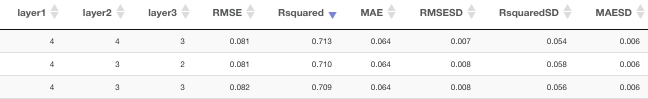
Fit on testing data (25% = 266 observations)

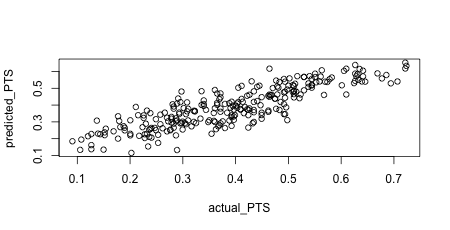
The Network



* + 1. **Defense**

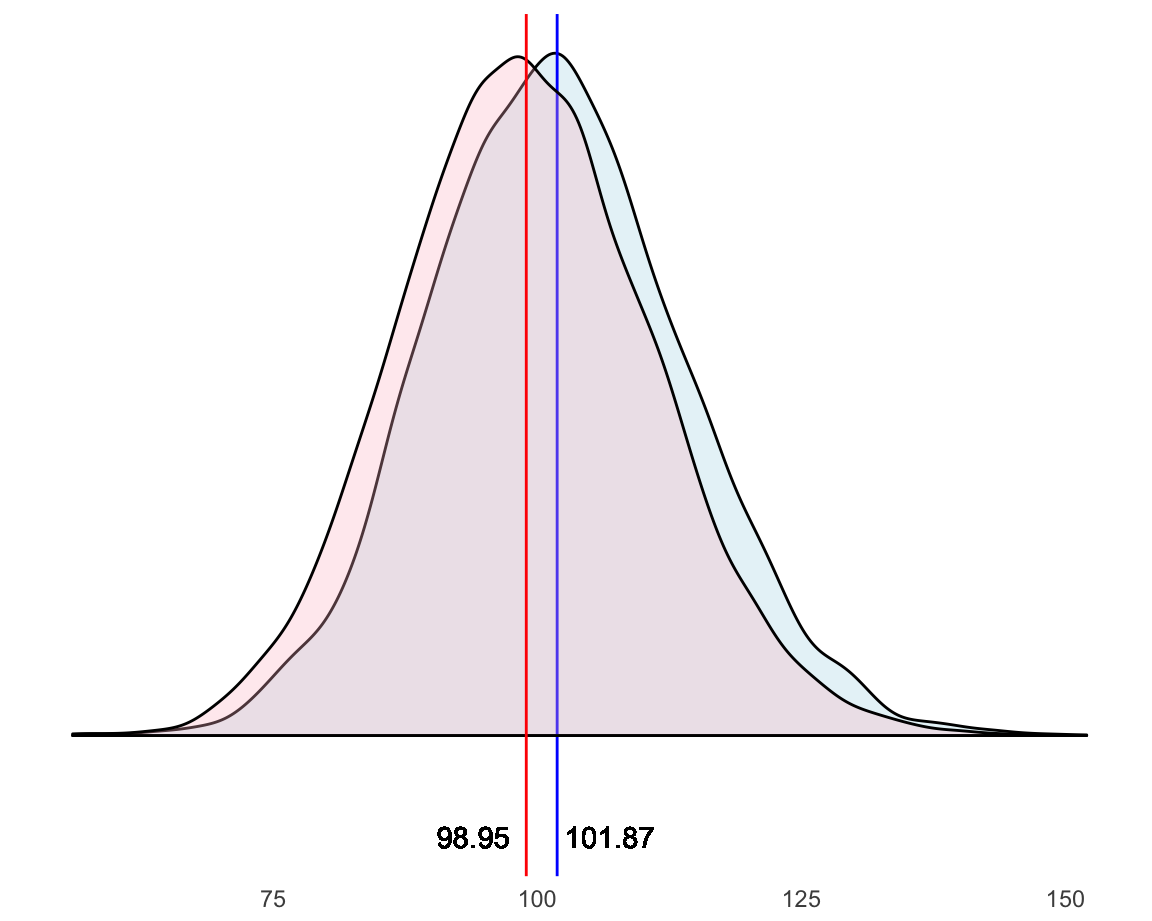
Same drill for the defense network. Here the results:



Clearly not a great fit but still good enough.

* 1. **Probability model**

Now that I have a way to estimate team’s offensive and defensive powers (average points), I can plug those into a Normal distribution probability. The choice for the Normal distribution is obvious if we explore the density of points scored by teams in the last 8 seasons: (since 2009-2010)



Where the blue density corresponds to points scored by home teams and red to away teams. Empirically, we can then assume that the average points scored are 100.41 with a standard deviation of: 12.16 pretty much consistent for home and away teams (sigmaH = 12.13, sigma = 12.0). For the sake of simplicity I will assume the sd is constant across the board which allows me to fully determine the probability model for both Offense and Defense for a given team:

$\Large{X \sim \mathrm{N}(\hat{\mu}\_O,\hat{\sigma})}$

* 1. **Wins**

Now that I can compute how many points does a team score or get scored on average, I will calculate how many wins those offensive and defensive powers award. Let’s suppose teamA plays against teamB. Let’s also suppose teamA is the home team. I know how many points teamA scores on average, call it: ptsA. Empirically we saw in 3.2. that home teams score 1.46 more points than average, thus, ptsA\* = ptsA + 1.46. Next, I need to consider teamB defensive power, call it: pts\_agB. The higher the defensive power is, the worst the defense is (as opposed to offense). So, I will add this number to teamA’s offensive power: ptsA\*\* = ptsA\* + pts\_agB. This will give me an astronomical number of points scored or a basketball game, which will be cancelled out when I calculate teamB’s offense. But in order to not just calculate the win probability but also simulate a plausible score between these 2 teams, I will subtract the overall average points (100.41) from that number. So, here’s how I calculate the points teamA will score against teamB when teamA is the home team:

ptsA\_final = ptsA + home\_court\_coeff + pts\_agB – overallPtsAvg

And for teamB as the away team:

ptsB\_final = ptsB - home\_court\_coeff + pts\_agA – overallPtsAvg

Now I can plug these 2 measures as the means of Normal distributions with the same sd = 12.16 and can calculate the probability of teamA beating teamB:

Let X = N(ptsA\_final,sd) and Y = N(ptsB\_final,sd) random variables. Then:

P(teamA beats teamB) = P(X > Y) = P(X-Y > 0) = P(N(ptsA\_final- ptsB\_final,sqrt(2)\*sd) > 0)

* 1. **Building the roster**

So far, I used historical data to build a model that computes win probabilities based on team’s offense and defense. But, for this to be useful I need to be able to predict these offensive and defensive powers based on the players that make each team at a given point and their skills. In this section I explain the methodology I use to build a team’s roster.

• A Normal distribution is centered around each team’s estimated power (offense and defense) and a fixed common variance (based on empirical data). With the full probability distribution I can simulate any matchup.

• Player similarity by age is computed using t-SNE algorithm which also allows for 2-D visualization of the data..

As conference policy, we do not support LaTeX, so we ask that you use this template instead. We understand that math typesetting can be more cumbersome in MS Word, but we suggest using MS Word’s equation editor. Equations will look like this:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

This is a reference to equation (1) that updates after fields are updated. Notice that each equation is contained in its own table, and the equation numbers are inserted using fields. See [this tutorial](https://www.youtube.com/watch?v=wM57WvO20KA) for more information on this technique.

We do not insist on any specific conventions related to figures, tables, and captions.

1. **Section**

This is the body of text under the third main section.

1. **Section**
   1. **Subsection (Cambria, Bold, 12pt)**

This is the first paragraph of the body of text under the first subsection of the first main section.

This is the second paragraph to give you sense of the spacing.

* + 1. **Sub-subsection (Cambria, Bold, 12pt)**

This is the first paragraph of the body of text under the first subsection of the first main section. Subsections can be nested as far as you want, though the font for the subsection headers remain the same (Cambria, 11pt).

* 1. **Subsection**

This is the first paragraph of the body of text under the first subsection of the first main sectionThis is the body of text under the fourth main section.

**References**

[1] Reference #1 cited using any mainstream citation style (e.g. APA, MLA).

[2] Reference #2

.

.

.

[n] Reference #n

**Appendix**

An appendix is not required, but if you have one please include it here.

1. Footnotes are permitted and should be formatted as shown here (Cambria, 11pt). [↑](#footnote-ref-1)