New predictive model

Derek Corcoran and Nick Watanabe

March 31, 2017

This is a comparison between the new and old model build throught our offensive and deffensive APPS to predict the spread in NBA games. Most of the time that I worked on this I was developing a new dataset, which used the Regular Season resutls to train our models, because even when we were using all of the 980,355 taken through the seasosn 2012 through 2016, we were not using every game to train the model.

Our old algorithm only used results of playoffs in order to train the dataset, that is we calculated our defAPPS and offAPPS for every matchup in the post season given all the shots in that season. That gave us a total of 402 to work with. Further more, we had to divide our dataset in a training and testing set, which gave us 316 games as a training set and 86 as a test set.

The New Dataset

With that in mind I was set to develop a new dataset that made it possible to further train our model and use more available data. For that the first thing I did was to collect all the results of the NBA's regular season from 2013 till March 30th of 2017 (The script is ready to go so we can do this anytime we want to to increase it).

Then we had to calculate the offAPPS and defAPPS for every one of the 6,199regular season games available, but only with the shots taken before that date. Nevertheless it is expected that at the beggining of the season, if we calculate the offAPPS or defAPPS between two teams, the result will be highly unstable though to the low number of shots taken, and that that should become more and more stable throughout the season.

In order to take that into account calculated the defAPPS of the Cavs vs the Warriors for everyday of the season for every season form 2013 to 2017 (Fig 1). I fitted an asymptotic model, the model states that on the day 68.52 is when the data sabilizes, so we calculated the 69th for each season and only use the data after that date.

Table 1: The first day for each season taken into account for further modelling

Season	dates
2013	2013-01-08
2014	2014-01-07
2015	2015-01-06
2016	2016-01-05
2017	2017-01-03

We joined the constructed dataset with the Playoffs dataset that we had previously, which gave us 4,023 games to work with, which were divided into 3,337games as a training set and 686 as a test set. Furthermore adding the 2017 shots to the Dataset, also increased the number of available shots for the model to 1,190,285.

New model

With that dataset, I used the same algorithm as before to train the model and create a new one. The response of that model is shown in figure 2. I will play around with this model over the weekend and extract some

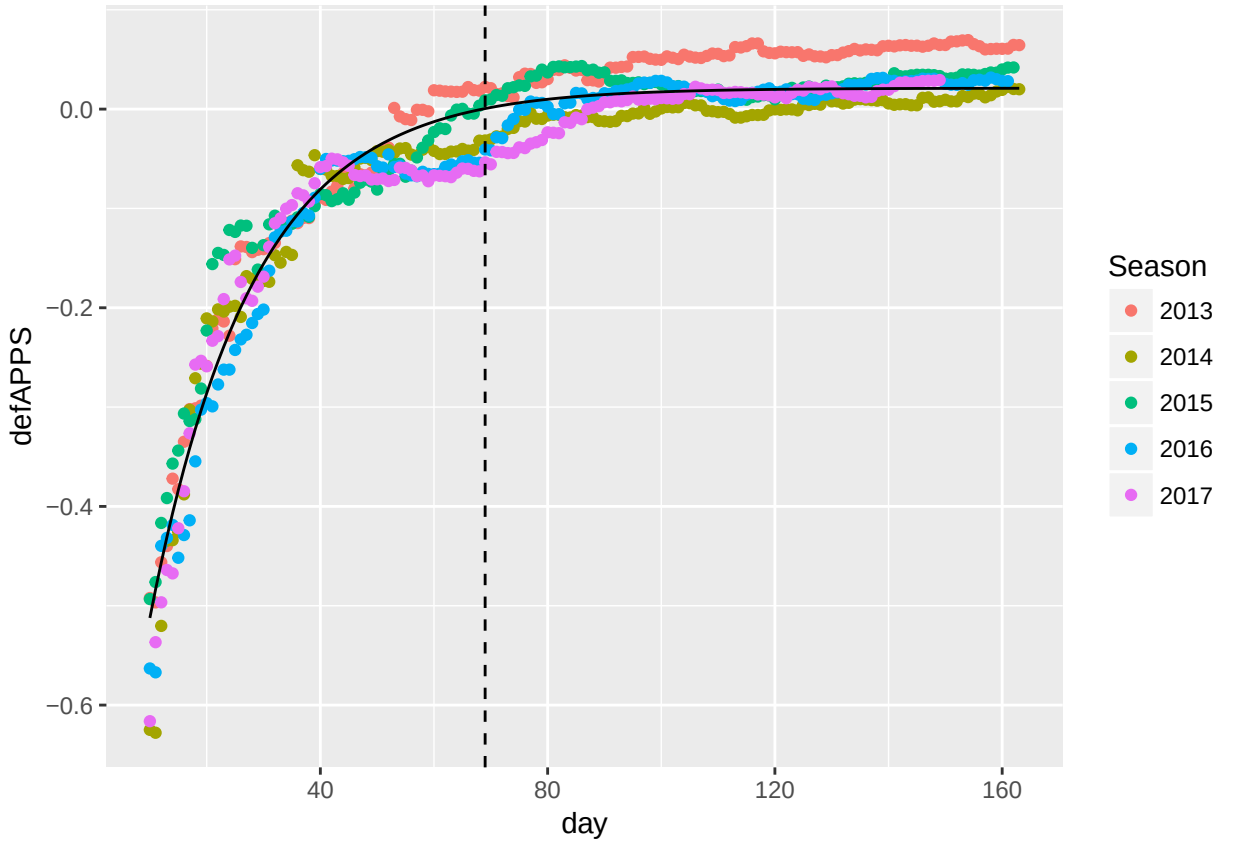

Figure 1: defAPPS for every day of the season from 2013 to 2017 between the Cavs and the Warriors, the solid line represents the fitted asymptotic model, and the dashed line, the day of the season when the defAPPS get stable.

performances values, but other than that we should be set.

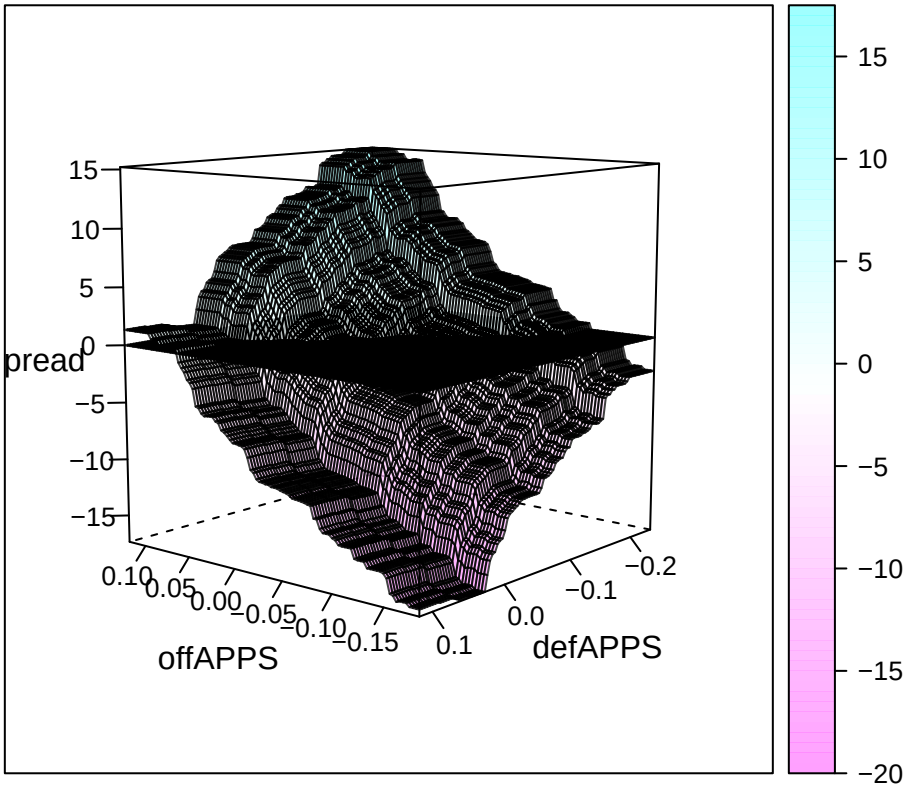

Figure 2: Predicted response of the predicted spread to defAPPS and offAPPS