

Racism on Reddit?

Determinants of Sentiment in Online Posting About NBA Players

Michael Patterson*
Microsoft

Matt Goldman †
Facebook

October 13, 2018

INCOMPLETE DRAFT
PLEASE DO NOT CIRCULATE

Abstract

Not too abstract.

1 Introduction

President Trump and Nike express different opinions about Colin Kaepernick. But are these real preferences or just appeals to their audiences? This parallels the literature of discrimination in professional sports, where studies have focused on salary differentials or referee performance, but often could not determine whether disparities were driven by organizational bias or fan preferences. Typically, these studies had difficulty constructing appropriate measures of fan sentiment or isolating the impact of particular characteristics (for example race).

We apply standard techniques in natural language processing to quantify the sentiment of 700,000 player-specific comments from reddit.com/r/nba. We find sentiment is most positive towards young, old, and high-performing players. After controlling for these factors, we estimate that in aggregate whiteness does not statistically significantly predict player sentiment ($t=1.20$). Whiteness is at most worth an additional 2.3 points of PER or an additional 3 years of player youth. However, we do find that sentiment is higher for non-white players in cities where Hillary Clinton performed well in 2016 ($t=2.15$), but that this is not true for white players.

*email here

†mattgoldman5850@gmail.com

2 Methodology

To create a corpus of player related comments, we scraped reddit/r/nba for comments during the 2017-2018 season. For each comment, we used Named Entity Recognition to identify players, and filtered out comments that did not contain exactly one player mention. Next, we used NLTK to calculate the sentiment for each comment and aggregated the sentiment scores at the player level. Our sentiment scores pass the sniff test, as unpopular players like Zaza Pachulia had the lowest scores.

The left panel of Fig. 1 shows, unsurprisingly, that this metric is noisier for players with few comments. The right panel of Fig. 1 shows how sentiment increases for older ($t=1.77$) and younger ($t=4.14$) players.

3 Results

We joined the sentiment scores to performance and demographic data to perform a WLS regression analysis, as shown in Table 2.

4 Conclusion

We provide a framework to test for idiosyncrasies in fan preferences. This may help understand what determines an athlete's brand value or incentives for apparently biased behavior by organizations. It can easily be expanded by including new text sources (Twitter) or be applied to other domains for comparative analysis (like the NFL). Our findings bolster the view that racial bias is not a large determinant of fan sentiment, but that non-white players may be polarizing along the axis of the 2016 election.