# Preferences Among NFL Fans: Skin Tone and Merchandise Sales

Kennedy Mutai

April 27, 2023


## Abstract

This paper investigates discrimination against black NFL players by looking at sales of licensed merchandise from the year 2013 to 2021. When controlling for player and team performance, results strongly show that black players are less likely to be among the top merchandise sellers. Using a structural model, I find that for every level of merchandise sales, white players are 84% more likely to achieve sales above that level compared to black players.


## **Introduction**

Racial discrimination is a big ongoing discussion and a lot of research has been done on the topic. A lot of research has been done on racial discrimination in sports and evidence for racial discrimination has been found (Robst et al., 2011). This paper adds to the existing research by investigating racial discrimination against black NFL players by looking at merchandise sales. I also add to the research on customer discrimination. According to O'Reilly et al. (2015), people show their support to sports teams by buying their merchandise, and the most common merchandise bought is the jersey with an image of a particular player. Merchandise sales are increasingly becoming more important since it is a good source of revenue for sports teams.

This paper aims to investigate the presence of racial discrimination against black NFL players by looking at sales of licensed merchandise. I construct panel data for the years 2013-2021 using merchandise sales ranks (from NFLPA), player performance data (from Pro-football-reference.com), and RGB scores, which is a measure of skin tone and a proxy for race. RGB score gives an objective, continuous measure of skin tone. I run a probit regression with the merchandise sales ranks as the dependent variable and the proxy for race as the main independent variable. The probit regression predicts the probability of an NFL player being among the top 50 players by merchandise sales. In addition to the regression, I construct a theoretical model that imputes a continuous measure of the dollar value of merchandise sales using my ordinal rank data. I calibrate this theoretical model to the data. The results support the hypothesis that there is racial discrimination against black in the NFL as I found that moving from the average mean of the RGB scores of white players to the average RGB score of black players decreases the chances of a player being among the top 50 players by merchandise sales by **2.08 %**. Moreover, the model predicts that for every level of sales, black players are 84% less likely to achieve sales above that level compared to white players.

## **Literature Review**

Volz (2017) investigates the existence of discrimination against black NFL players after observing that between 2001 and 2014 only 16-28 % of starting quarterbacks were black (Volz, 2017). The low percentage of starting quarterbacks is incommensurate with the proportion of black players in the league, 67% (Volz, 2017). Volz (2017) uses the survival time analysis technique to investigate the likelihood of players being benched for the next game. He finds evidence of racial discrimination against black QBs who are twice as likely to be benched compared to white players. In order to make sure that player performance is not the main factor for the benching, Volz investigated the performance of the team after benching and discovers that when a white QB is benched the team performs much better than when a black QB is benched. This is strong evidence of discrimination based on race in the NFL.

Robst et al. (2011) use data from player profile pictures to examine the relationship between skin tone and earnings of professional basketball players. The

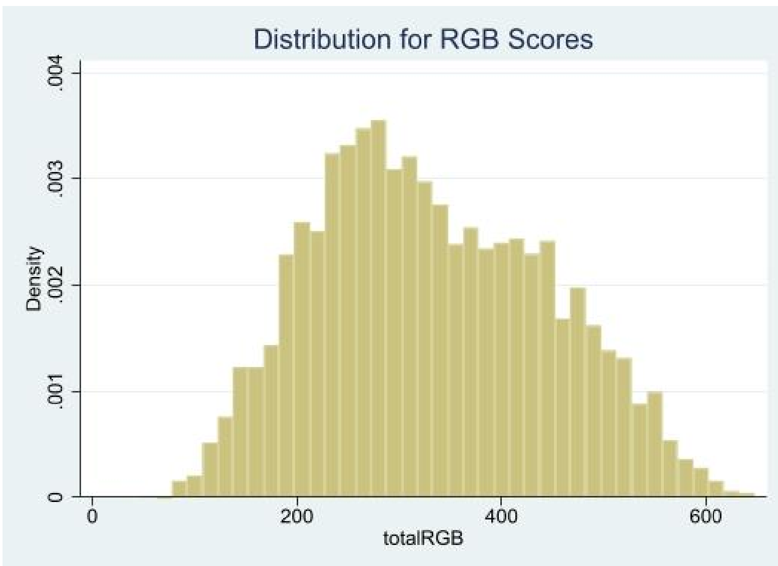
RGB score gives a measure for the red, blue, and green hue in a particular pixel of an image. Higher values indicate a lighter skin tone. Robst et al. were the first to use the RGB color score to measure skin tone. They used Photoshop to get the RGB scores from photos of players from the NBA website. In this paper, I also use the RGB color score as the measure of skin tone. Additionally, I use this measure to construct a continuous measure of race, varying from 0 to 1, analogous to a race dummy. Robst et al. run an OLS with annual salary as the dependent variable and RGB score as the independent variable. In this paper though, I run a probit regression with Merchandise sales ranks as the dependent variable and RGB score as the independent variable. Robst et al. find that there is very little relationship between skin tone and wages.

## **Data**

In order to examine the effect of race/ skin tone on the probability of an NFL player being among the top-ranked players by merchandise sales, I collect annual, player-level data on the National Football League between 2013-2021. I use data from four sources. The first source is the National Football League Players Association (NFLPA). I use NFLPA to gather data on the annual ranking of players by merchandise sales. The NFLPA produces a list of the top 50 players as related to the sale of official player-branded merchandise sold online or in stores. The list has only the top 50 players and provides only the ranking and not the dollar value of sales. The players' sales rank will be our dependent variable.

Second, using Pro-football-reference.com, a website that provides different statistics for American football, I get images of all the NFL players who were active from 2013 til 2021 by taking samples from the forehead, right cheek, and left cheek. RGB saturation measures the saturation of the Red, Green, and Blue color pigmentation in the skin of a player. RGB varies with the skin tone of the player. The RGB ranges from a score of 0 to 255 for each of the three colors. Black players have a score close to 0 while white players have a score close to 255 on each dimension (Robst et al., 2011). I sum up the Red, Green, and Blue RGB scores and end up with values that range from 0 to 765. Graph 1 below shows the distribution of the RGB scores. It appears to be similar to a normal distribution. RGB scores will be used as one of our main dependent variables. I verified using principal component analysis( PCA) that equally weighting the three components explains the vast majority of the variance (over 90%) in the measured data.

Graph 1: Histogram of RGB scores

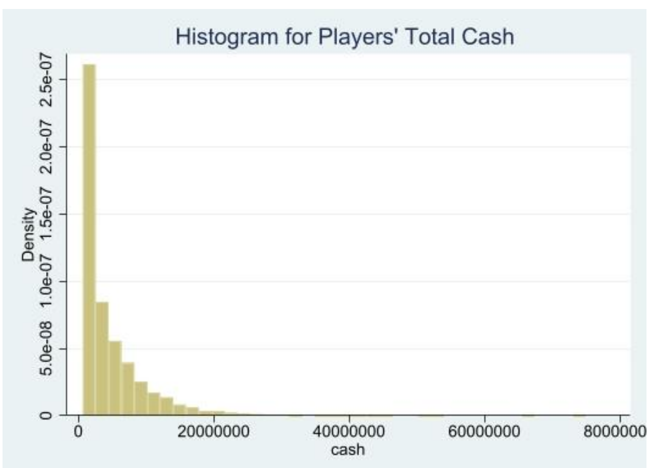



Third, using data from Pro-football-reference.com, I get the top 100 players list for the years 2013 -2021. This is a list generated every year after NFL players cast a vote to choose the best player in the league. The list is a ranking of the top 100 players. I will use this list in our regression to account for player performance.

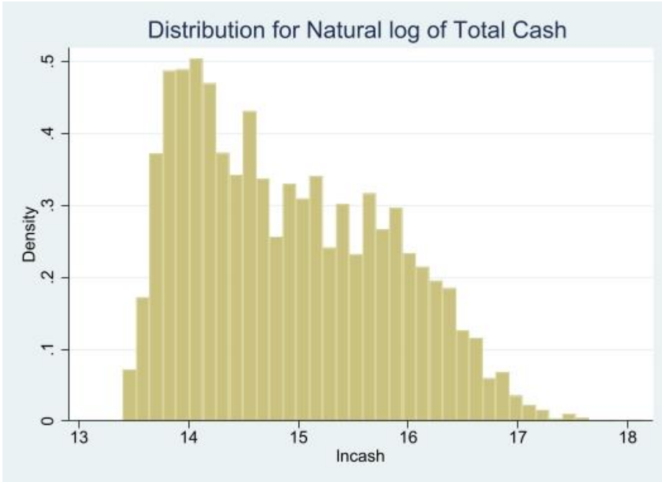
Fourth, I use data from Spotrac.com, an online source for sports players' salaries and contracts, to get the total cash paid to each player every year from 2013 to 2021. Graph 2 below shows the distribution of total cash. The distribution is greatly right-skewed and as a result, a variable that takes the natural logarithm of total cash was created so as to obtain a normal distribution. Graph 3 shows the distribution of the new total cash variable. The distribution is normal and as a result, I use log cash in the regression as one of the independent variables.

I combine the merchandise sales, RGB score, player performance ranking, and total cash to create a panel data set for the years 2013-2021. The observation unit is player-year.

Graph 2: Distribution of Total Cash



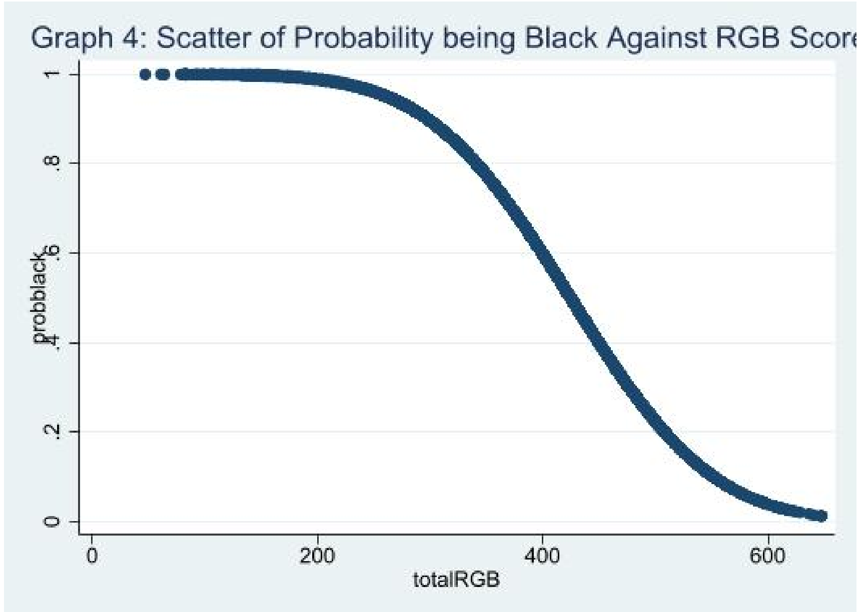

Graph 3: Distribution of Natural Log of Total Cash



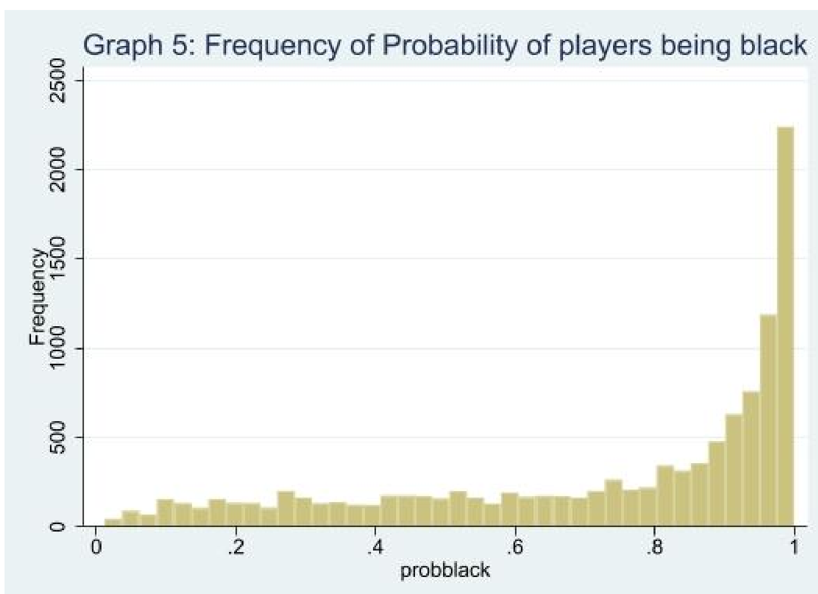

**Constructing Predicted Probability of a player being black given their RGB scores**

In order to get a different measure of skin tone, I hand code the race for 196 players as either black or white or other. I then use the hand-coded race to develop two indicator variables for race: For the first one, the indicator variable is 1 if the race is Black or Other and 0 if the race is White. For the second indicator variable, the variable is 1 if the race is Black and 0 if White. I then use the indicator variable from the subsample to predict the probability of remaining players being black given their RGB score by doing a probit regression of the hand-coded race indicator variable against RGB scores. I then extrapolate this to find the probabilities for all the players by using the formula: The probability of being black= normprob (coefficient of RGB Score *RGB score +constant). The probability will typically be either very close to 1 or very close to 0. As seen in Graph 4 below, as the probability of a player being black gets closer to 1, the RGB score gets closer to 0.

Graph 4:




Graph 5 below shows the frequency distribution of the probabilities of a player being black. It shows that there are more black players in our data sample compared to white players, the ratio of black to white players is 2.65.



## Methodology

In order to determine whether **skin tone** has any effect on the decision of NFL fans to purchase a player's merchandise, I estimate the following equation.

$$Pr(merchtop50|skin\ tone, player\ performance, salary) = F(\beta_0) + \beta_1(Skintone_{it}) + \beta_2(Playerperformance_{it}) + \beta_3(salary_{it}) + u)$$

where the dependent variable, merchtop50, is a dummy variable for players who are in the list of top 50 players by merchandise sales. The variable Skintone, in one set of regressions, is the RGB scores, while in another set, it is the predicted probability of a player being black given their RGB scores. Player performance will be represented by sets of dummy variables for players in the list of top 100 players by player performance.

I create different indicator variables for the different merchandise sales rank positions. Merchtop50 is a dummy variable that takes 1 for players in the top 50, merchtop25 takes 1 for players in the top 25, and merchtop10 takes 1 for players ranked in the top 10.

I also create different indicator variables for different ranges of player performance rank. For example, toprank100 is an indicator variable that takes 1 for players ranked between the 51st and 100th positions, and toprank10 is an indicator variable that takes 1 for players ranked between the 1st and 10th positions.

In the first regression, I run a probit regression with the dummy variables for players in the top 50 by merchandise sales as the dependent variable and the RGB scores as the main independent variable. I also include the natural log of salary and the different indicator variables for player performance rank as independent variables. I also include position and team fixed effects. Position fixed effects to control for the fact that players in certain positions like Quarter Backs, Wide Receivers, and Running Backs are the ones more likely to be ranked among the top 50 players by merchandise sales. Team fixed effects to control for the fact that teams are located in different cities and thus exposed to different numbers of fans. It also controls for the idea that there might be good players in a team that is not performing well. I cluster by player to control for autocorrelation.

In the second regression, I run a probit regression with the dummy variables for players in the top 50 by merchandise sales as the dependent variable and the probability of being black as the main independent variable. Similar to the first regression, I include the natural log of salary and the different indicator variables for player performance ranks as independent variables. I also include position and team fixed effects while clustering by player.

I repeat the above two regressions using the indicator variable for players in the top 10 list by merchandise sales as the dependent variable.

## Results

I first run a regression with the RGB scores as the main independent variable. **Table 1** below shows the results. In the first column, I have RGB scores as the only independent variable. On calculating the margins, I find that moving from the average RGB score of black players to the average RGB score of white players increases the probability of being among the top 50 by merchandise sales players by 2.08% as seen in Table 9 below. In the second column, I had the salary to the regression and the percentage increase increased to 5.55%. In the third column, I add player performance dummy variables of the different players' ranks. I find that for players who are ranked between the 51st and 100th positions on performance, represented by toprank100, moving from the average RGB score of black players to the average RGB score of white players increases the probability of being among the top 50 by merchandise sales players by 9.22 %. For those ranked between 1st and 10th positions, the percentage change is lower at 5.72%. In the 4th column, I add position-fixed effects and in the fifth column, I add team-fixed effects. On calculating the margins, after adding the fixed effects, the percentage changes are zero.

Table 1: Using Total RGB scores

| VARIABLES | (1)<br>merchtop50 | (2)<br>merchtop50 | (3)<br>merchtop50 | (4)<br>merchtop50 | (5)<br>merchtop50 |
|---|---|---|---|---|---|
| totalrgb | 0.00166*** | 0.00166*** | 0.00234*** | 0.00221*** | 0.00238*** |
| | (0.000197) | (0.000229) | (0.000278) | (0.000367) | (0.000427) |
| lncash | | 0.546*** | 0.265*** | 0.238*** | 0.279*** |
| | | (0.0292) | (0.0345) | (0.0404) | (0.0444) |
| toprank100 | | | 0.964*** | 1.041*** | 1.196*** |
| | | | (0.0911) | (0.108) | (0.121) |
| toprank50 | | | 1.533*** | 1.629*** | 1.860*** |
| | | | (0.105) | (0.125) | (0.140) |
| toprank25 | | | 2.285*** | 2.369*** | 2.683*** |
| | | | (0.130) | (0.152) | (0.177) |
| toprank10 | | | 3.137*** | 2.980*** | 3.224*** |
| | | | (0.195) | (0.208) | (0.230) |
| Constant | -2.387*** | -10.63*** | -7.025*** | -11.00 | -13.10 |
| | (0.0753) | (0.462) | (0.535) | (158.1) | (130.0) |
| Observations | 11,248 | 8,435 | 8,435 | 6,912 | 6,844 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Next, I run a regression with the probability of a player being black as the main independent variable. **Table 2** below shows the results when I use the probability of a player being black created using a dummy variable black that is 1 if the players I hand-coded their races as “Other” and “Black” and 0 if the race is “White”. **Table 3** shows the results when I use the probability of a player being black created using a dummy variable black that is 1 if the players I hand-coded their races as “Black” and 0 if the race is “White” and “Other”. The two tables have very similar results.

Table 2: Using probability of being black with other as black

| VARIABLES | (1)<br>merchtop50 | (2)<br>merchtop50 | (3)<br>merchtop50 | (4)<br>merchtop50 | (5)<br>merchtop50 |
|---|---|---|---|---|---|
| probblack | -0.653*** | -0.670*** | -0.891*** | -0.862*** | -0.970*** |
| | (0.0732) | (0.0857) | (0.102) | (0.138) | (0.162) |
| lncash | | 0.547*** | 0.267*** | 0.240*** | 0.280*** |
| | | (0.0292) | (0.0345) | (0.0404) | (0.0445) |
| toprank100 | | | 0.966*** | 1.043*** | 1.198*** |
| | | | (0.0912) | (0.108) | (0.121) |
| toprank50 | | | 1.527*** | 1.623*** | 1.859*** |
| | | | (0.105) | (0.125) | (0.140) |
| toprank25 | | | 2.274*** | 2.357*** | 2.666*** |
| | | | (0.130) | (0.151) | (0.176) |
| toprank10 | | | 3.124*** | 2.966*** | 3.213*** |
| | | | (0.194) | (0.207) | (0.229) |
| Constant | -1.363*** | -9.610*** | -5.625*** | -9.680 | -11.54 |
| | (0.0523) | (0.457) | (0.531) | (157.6) | (130.4) |
| Observations | 11,248 | 8,435 | 8,435 | 6,912 | 6,844 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Using probability of being black with other as white

| VARIABLES | (1)<br>merchtop50 | (2)<br>merchtop50 | (3)<br>merchtop50 | (4)<br>merchtop50 | (5)<br>merchtop50 |
|---|---|---|---|---|---|
| probblack_2 | -0.630*** | -0.641*** | -0.881*** | -0.880*** | -0.969*** |
| | (0.0713) | (0.0832) | (0.101) | (0.135) | (0.158) |
| lncash | | 0.547*** | 0.266*** | 0.239*** | 0.279*** |
| | | (0.0292) | (0.0346) | (0.0405) | (0.0445) |
| toprank100 | | | 0.967*** | 1.047*** | 1.202*** |
| | | | (0.0913) | (0.108) | (0.121) |
| toprank50 | | | 1.534*** | 1.632*** | 1.866*** |
| | | | (0.105) | (0.125) | (0.140) |
| toprank25 | | | 2.282*** | 2.369*** | 2.681*** |
| | | | (0.131) | (0.152) | (0.177) |
| toprank10 | | | c*** | 2.987*** | 3.234*** |
| | | | (0.195) | (0.209) | (0.230) |
| Constant | -1.445*** | -9.697*** | -5.724*** | -9.768 | -11.72 |
| | (0.0445) | (0.456) | (0.530) | (157.7) | (130.5) |
| Observations | 11,248 | 8,435 | 8,435 | 6,912 | 6,844 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

I use the results in **Table** 2 to analyze the results of using the probability of being black to predict the probability of a player being in the top 50 merchandise sales list. In the first column, I have the probability of being black as the only independent variable. As shown in **Table 9**, I find that black players are 6.52 % less likely to be among the top 50 players by merchandise sales. In the second column, I add the salary to the regression and find that black players are 15.82 % less likely to be among the top 50 players by merchandise sales. In the third column, I add player performance dummy variables of the different players' ranks. I find that for players who are ranked between the 51st and 100th positions by performance, black players are 16.38 % less likely to be among the top 50 players by merchandise sales. For those ranked between 1st and 10th positions, black players are 12.41% less likely to be among the top 50 players by merchandise sales. In the 4th column, I add position-fixed effects and in the fifth column, I add team-fixed effects. After adding the fixed effects, the percentage differences in the probabilities become zero.

The results are very similar when I use the RGB scores and when I use the predicted probabilities of being black as the race variables. As seen in Table 9, the estimates are very similar in the different specifications of the race variable. The estimates are also very similar across the two dummy variables for predicted probabilities of being black. The race variable has the same magnitude when I use the probability of a player being black created using the dummy variable black which is 1 for the players I hand-coded their races as "Other" and "Black" and 0 if the race is "White" and when I use the predicted probability created using the dummy variable black_2 which is 1 for the players that I hand-coded their races as “Black” and 0 for the players that I hand-coded their races as “White” or “Others”. The non-race variables are also robust across all three specifications.

Table 9: Margins for the RGB score and probability variables with Merchtop50 as the dependent variable

| Main Independent Variable | (1) Main Independent Variable Alone | (2) Adding total cash | (3) Adding players' rank | (4) Adding position fixed effects | (5) Adding team-fixed effects |
|---|---|---|---|---|---|
| totalrgb | 2.08 | 5.55 | 9.22 (toprank100)<br>5.72 (toprank10) | 0.0<br>0.1 | 0.0<br>0.0 |
| probblack | 6.52 | 15.82 | 16.38 (toprank100)<br>12.41 (toprank10) | 0.0<br>0.0 | 0.0<br>0.0 |
| Probblack_2 | 5.47 | 13.99 | 22.87 (Toprank100)<br>25.39 (toprank10) | 0.0<br>0.0 | 0.0<br>0.0 |

## **Further Analysis**

As a robustness check, I use the indicator variable of players in the top 10 list of merchandise sales as the dependent variable. In **Table 4,** I use RGB scores as the main independent variable to predict the probability of a player being in the top 10 merchandise ranks. The percentage differences in the probability of being in the top 10 between black and white players are smaller than the probability of being in the top 50 merchandise sales list. **Table 10** below shows the percentage differences between the probability of white players and the probability of black players being in the top 10. When I have only the RGB scores as the independent variable, moving from the average RGB score of black players to the average RGB score of white players increases the probability of being in the top 10 merchandise sales list by 0.6%. When I add salary, black players are 1.87% less likely compared to white players to be in the top 10. When I add player performance, black players are 1.86% less likely to be among those in the top 10 for players ranked between the 51st and 100th positions on performance in player performance, and for those ranked between 1st and 10th positions, black players are 14.21% less likely to be in the top 10n compared to white players. When I had position-fixed effects and team-fixed effects in columns 4 and 5, the coefficients on RGB scores became insignificant.

Table 4: Using Total RGB scores

| VARIABLES | (1) merchtop10 | (2) merchtop10 | (3) merchtop10 | (4) merchtop10 | (5) merchtop10 |
|---|---|---|---|---|---|
| totalrgb | 0.00197*** | 0.00188*** | 0.00243*** | 0.000605 | 0.000104 |
| | (0.000370) | (0.000422) | (0.000538) | (0.000746) | (0.00102) |
| lncash | | 0.526*** | 0.197*** | 0.121* | 0.169* |
| | | (0.0542) | (0.0655) | (0.0725) | (0.0900) |
| toprank100 | | | 0.711*** | 0.791*** | 1.064*** |
| | | | (0.186) | (0.223) | (0.295) |
| toprank50 | | | 0.720*** | 0.639** | 0.826** |
| | | | (0.234) | (0.275) | (0.339) |
| toprank25 | | | 1.855*** | 1.811*** | 2.380*** |
| | | | (0.180) | (0.209) | (0.307) |
| toprank10 | | | 2.413*** | 2.286*** | 3.291*** |
| | | | (0.188) | (0.213) | (0.356) |
| Constant | -3.181*** | -11.18*** | -6.822*** | -8.290 | -15.46 |
| | (0.149) | (0.872) | (1.020) | (234.9) | (348.3) |
| Observations | 11,248 | 8,435 | 8,435 | 4,804 | 3,231 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 5** below shows the results when I use the probability of being black to predict the probability of players being in the top 10 merchandise sales list. In the first column, I have the probability of being black as the only independent variable. As shown in **Table 10**, I find that black players are 2 % less likely to be among the top 50 players by merchandise sales. In the second column, I had the salary to the regression and I find that black players are 5.75 % less likely to be among the top 50 players by merchandise sales. In the third column, I add player performance dummy variables of the different players' ranks. I find that for players who are ranked between the 51st and 100th positions by performance, black players are 5.87 % less likely to be among the top 50 players by merchandise sales. For those ranked between 1st and 10th positions, black players are 33.31%  less likely to be among the top 50 players by merchandise sales. In the 4th and 5th columns, the coefficient on the probability of being black becomes insignificant when I add position-fixed effects and team-fixed effects respectively.

Table 5: Using probability of being black with other as black

| VARIABLES | (1)<br>merchtop10 | (2)<br>merchtop10 | (3)<br>merchtop10 | (4)<br>merchtop10 | (5)<br>merchtop10 |
|---|---|---|---|---|---|
| probblack | -0.733*** | -0.723*** | -0.863*** | -0.218 | -0.0873 |
| | (0.131) | (0.152) | (0.190) | (0.270) | (0.376) |
| lncash | | 0.527*** | 0.201*** | 0.122* | 0.168* |
| | | (0.0543) | (0.0654) | (0.0724) | (0.0900) |
| toprank100 | | | c*** | 0.793*** | 1.063*** |
| | | | (0.185) | (0.223) | (0.295) |
| toprank50 | | | 0.711*** | 0.637** | 0.826** |
| | | | (0.234) | (0.276) | (0.339) |
| toprank25 | | | 1.838*** | 1.810*** | 2.378*** |
| | | | (0.179) | (0.209) | (0.307) |
| toprank10 | | | 2.391*** | 2.283*** | 3.289*** |
| | | | (0.188) | (0.213) | (0.357) |
| Constant | -1.991*** | -10.05*** | -5.433*** | -7.940 | -15.34 |
| | (0.0872) | (0.864) | (1.010) | (235.2) | (349.0) |
| Observations | 11,248 | 8,435 | 8,435 | 4,804 | 3,231 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Margins for the RGB score and race variables with Merchtop10 as the dependent variable

| Main Independent Variable | (1) Main Independent Variable Alone | (2) Adding total cash | (3) Adding players' rank | (4) Adding position fixed effects | (5) Adding team-fixed effects |
|---|---|---|---|---|---|
| totalrgb | 0.6 | 1.87 | 1.86 (toprank100) 14.21 (toprank10) | ns ns | ns ns |
| probblack | 2 | 5.75 | 5.87 (toprank100) 33.31 (toprank10) | ns ns | ns ns |
| Probblack_2 | 1.66 | 4.96 | 4.79 (toprank100) 32.79 (toprank10) | ns ns | ns ns |

**Model**

In our data I cannot observe the actual dollar value of the merchandise sales; I only have the ranks of the top 50 players and their races. I, therefore, use a theoretical model to convert the ranks into actual dollar values.

I assume that there are two groups of players B and W. There are $n_g$ players in group g for $g \in \{B, W\}$ and denote $n = n_B + n_W$ . Let $X_{gi}$ denote the dollar value of merchandise sales for player i in group g. Suppose each $X_{gi}$ is an independent draw from a distribution $F_g(x)$, so that $F_g$ is the CDF of merchandise sales for players in each group. I do not observe the dollar value of merchandise sales (I only see the ranks by sales), and as a result, I need to map the ranks to the dollar values.

I let $t$ be some threshold level of sales. The probability that any given player from group g has sales exceeding this level is $1-F_g(t)$. Hence, the expected number of players from group $g$ with sales exceeding this level is $n_g(1 - F_g(t))$, and the expected number of players across all groups exceeding this level is $n_B(1 - F_B(t)) + n_W (1 - F_W (t))$.

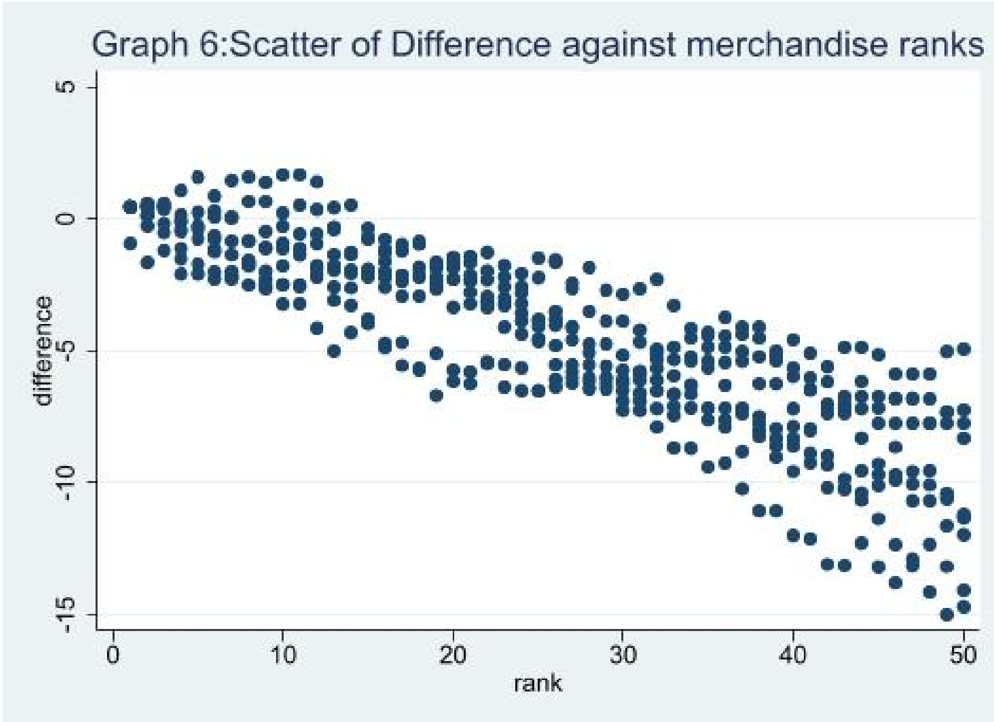
For each $k \in \{1, 2, . . . , 50\}$, I observe the race of the players who are in the top k players by merchandise sales. As a result,for each $k \in \{1, . . . , 50\}$, I choose the threshold $t(k)$, which is the cut-off sales level between players in the top $k$ by rank, and the players outside this set, to ensure that the expected number of players exceeding the threshold is $k$. $t(k)$ is given by:

$$n_B(1 - FB(t(k))) + n_W (1 - F_W (t(k))) = k$$

The left-hand side is decreasing in $t$ while the right-hand side is constant in $t$. Therefore, there is a unique $t(k)$ that causes the equation to hold. $t(k)$ is different for each $k$; as $k$ increases, $t(k)$ decreases. Given these thresholds, we can compute the expected difference between the number of white and black players who are in the top $k$ players by sales:

$$d(k) = n_W (1 - F_W (t(k))) - n_B(1 - F_B(t(k)))$$

I show in the Appendix that if $X_{ig}$ follows a Pareto distribution, with possibly different group means, but a common shape parameter, then $d(k)$ will be linear in k. As shown in Graph 6 below, when I plot $d(k)$ against $k$ I get roughly a straight line, suggesting that the Pareto assumption is appropriate. Indeed, the Pareto distribution has been shown to capture the tail distributions of many phenomena, including the salaries of CEOs, sales by artists, sizes of big cities, etc - so it is a natural candidate for modeling the distribution of merchandise sales. As I show in the Appendix, the slope of the linear plot of d(k) against k can be directly mapped onto a meaningful property of the Pareto distribution - formally we can use it to capture the standardized difference in means between the two groups. Moreover, we can interpret this quantity as the relative probability of agents from one group having sales above an arbitrary threshold relative to players from the other group (The Pareto distribution has the special property that this relative probability is the same for any arbitrary threshold.) I find that for every level of sales white players are 84% more likely to achieve sales above or at that level compared to black players. This is evidence of the existence of racial discrimination in the NFL.




**Discussion  & Conclusion**

The paper involves the analysis of merchandise sales data and the construction of a model that converts the merchandise rank into a continuous measure of merchandise sales. The results presented support the hypothesis that black NFL players face discrimination from fans when it comes to purchasing merchandise. Merchandise belonging to black players is less likely to be bought compared to those of white players. As the skin tone of players become lighter, players sell more merchandise. This supports the hypothesis on the existence of consumer discrimination among NFL fans.  This paper contributes to the study of racial discrimination in sports. and also to the study of consumer biases by using sports data.

This study can be applied to other sports to study discrimination. Further analysis can be done to see if other factors such as social media presence and the on-pitch and off-pitch behavior of players affect merchandise sales.

The results of this study can be useful to NFL players when negotiating contracts especially when it comes to commissions from merchandise sales.

## **References**


O'Reilly, N., Foster, G., Murray, R., & Shimizu, C. (2015). Merchandise sales rank in professional sport: Purchase drivers and implications for National Hockey League clubs. *Sport, Business and Management: An International Journal*, *5*(4), 307–324. https://doi.org/10.1108/SBM-10-2012-0044

Robst, J., VanGilder, J., Coates, C. E., & Berri, D. J. (2011). Skin Tone and Wages: Evidence From NBA Free Agents. *Journal of Sports Economics*, *12*(2), 143–156. https://doi.org/10.1177/1527002510378825

Volz, B. D. (2017). Race and Quarterback Survival in the National Football League. *Journal of Sports Economics*, *18*(8), 850–866. https://doi.org/10.1177/1527002515609659

## Appendix

**Pareto Distribution**

The Pareto distribution is commonly used to model the distribution of tail events (e.g. size of the largest cities, sales of the most popular artists etc). The Pareto distribution has two parameters: *x* which is the lowest value of the distribution, and α which is a shape parameter (with α > 2). The expected number of sales is: (α / α−1)* *x*. Suppose $X_{gi}$ ~ Pareto($x_{gi}$ , α) — i.e. sales have different means in each group, but the same shape parameter. The (counter)-CDF of a Pareto distribution is given by:

$$1 - F(x; x, \alpha) = \Pr[X > x] = (x // x)^{\alpha}$$

Then, the condition that defines t(k) becomes:

$$n_B(1 - F_B(t(k))) + n_W (1 - F_W (t(k))) = k$$

$$n_B \ (x_B / t(k))^{\alpha} + n_W \ (xW / t(k))^{\alpha} = k$$

$$t(k)^{\alpha} = (n_B x_B^{\alpha} + n_W x_W^{\alpha}) / k$$

$$t(k) = (n/ k)^{1/\alpha} \left[(n_B/ n)\, x_B^{\alpha} + (n_W/n)\, x_W^{\alpha}\right]^{1/\alpha}$$

This implies that:

$$d(k) = n_W (1 - F_W (t(k))) - n_B(1 - F_B(t(k)))$$

$$= n_W \left(x_W / t(k)\right)^{\alpha} - n_B \left(x_B / t(k)\right)^{\alpha}$$

$$= n_W \left(( x_W / ((n/ k)^{1/\alpha} [(n_B/n)\, x^{\alpha}_B + (n_W/n)\, x^{\alpha}_W]^{1/\alpha})\right)^{\alpha} - n_B \left(( x_B / ((n/ k)^{1/\alpha} [(n_B/n)\, x^{\alpha}_B + (n_W/n)\, x^{\alpha}_W]^{1/\alpha})\right)^{\alpha}$$

$$= k. \cdot ((n_B/n)\, x^{\alpha}_W - (n_W/n)\, x^{\alpha}_B) / ((n_B/n)\, x^{\alpha}_W + (n_W/n)\, x^{\alpha}_B)$$

This expression again depends on 3 distributional parameters: $x_W$ , $x_B$ and α, and we would need to conjecture all 3 to compute it. But we can simplify. Dividing top and bottom by $(n_B/n)\ x^{\alpha}{}_B$ gives:

$$d(k)= \;= k \cdot \Big(((nW/nB)\ \Delta - 1) / ((n_W/n_B)\Delta + 1)\Big)$$

Hence, it suffices to conjecture $\Delta = (x_W / x_B)^{\alpha}$ . We don't need to know each of xW , xB, and α separately. For any $x > x_W$:

$$Pr[X_{iW} > x] / Pr[X_{iB} > x] =(x_W/x)^{\alpha} / (x_B / x)^{\alpha} = (x_W / x_B)^{\alpha} = \Delta$$

Hence, Δ is meaningful in that for any level of sales, it measures the probability that white players achieve that level (or higher) relative to black players.

The expression above is in two parts. The first part is *k*. The second part only involves *n*'s and Δ which are independent of *k*. So for a Pareto distribution, the difference between the number of whites and blacks in the top *k* scales with *k*.

For each *k* in each year, I find the number of white and black players in the top *k* by sales, and take the difference. I find the number of black players in the top *k* by adding together ProbBlack for all players in top *k* each year; likewise, I find the number of white players in the top *k* by adding together 1-ProbBlack for all players in top *k* each year. I find that $n_B/n_W$ is relatively constant across time periods and thus I do a linear regression of d(K) on k, pooling data across years. I find the estimated coefficient $\hat{\beta}$ and I use it to estimate $\Delta = (1+ \hat{\beta})/ (1- \hat{\beta})\ .\ (n_B/n_W)$

**Regression results: side by side comparison of the six regression results**

Table 8 below shows side by side comparison of the results for the six regressions I run. The results are from the last columns of the tables in the results section. This means that the regressions results presented included the team and position fixed effects. For each independent variable i.e the RGB scores and the two predicted probability of a player being black, I have presented the results for when merchtop50 and merchtop10 are the dependent variables.

Table 8: side by side comparison of merchandise ranks results

| VARIABLES | (1) merchtop50 | (2) merchtop10 | (3) merchtop50 | (4) merchtop10 | (5) merchtop50 | (6) merchtop10 |
|---|---|---|---|---|---|---|
| totalrgb | 0.00238*** | 0.000104 | | | | |
| | (0.000427) | (0.00102) | | | | |
| lncash | 0.279*** | 0.169* | 0.280*** | 0.168* | 0.279*** | 0.169* |
| | (0.0444) | (0.0900) | (0.0445) | (0.0900) | (0.0445) | (0.0901) |
| toprank100 | 1.196*** | 1.064*** | 1.198*** | 1.063*** | 1.202*** | 1.064*** |
| | (0.121) | (0.295) | (0.121) | (0.295) | (0.121) | (0.295) |
| toprank50 | 1.860*** | 0.826** | 1.859*** | 0.826** | 1.866*** | 0.826** |
| | (0.140) | (0.339) | (0.140) | (0.339) | (0.140) | (0.339) |
| toprank25 | 2.683*** | 2.380*** | 2.666*** | 2.378*** | 2.681*** | 2.380*** |
| | (0.177) | (0.307) | (0.176) | (0.307) | (0.177) | (0.307) |
| toprank10 | 3.224*** | 3.291*** | 3.213*** | 3.289*** | 3.234*** | 3.291*** |
| | (0.230) | (0.356) | (0.229) | (0.357) | (0.230) | (0.356) |
| probblack | | | -0.970*** | -0.0873 | | |
| | | | (0.162) | (0.376) | | |
| probblack_2 | | | | | -0.969*** | -0.0436 |
| | | | | | (0.158) | (0.372) |
| Constant | -13.10 | -15.46 | -11.54 | -15.34 | -11.72 | -15.39 |
| | (130.0) | (348.3) | (130.4) | (349.0) | (130.5) | (348.1) |
| Observations | 6,844 | 3,231 | 6,844 | 3,231 | 6,844 | 3,231 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1