QB or not QB? Measuring Discrimination in Football Labor Markets

Magdalena Bennett The University of Texas at Austin

Abstract

This paper examines the presence of racial discrimination in the labor market of American football, focusing on recruitment for the quarterback (QB) position—a role often regarded as both the strategic leader on the field and the public face of a franchise. Leveraging a novel dataset that integrates high school, college, and professional football statistics, I apply two empirical strategies—a benchmark test and an outcome test—to identify patterns of discrimination. The findings reveal consistent evidence of racial bias in recruitment decisions during the transitions from high school to college and from college to the National Football League (NFL). Importantly, these results are robust to alternative explanations, including player reallocation to other positions, and hold under sensitivity analyses designed to account for potential unobserved confounding. Finally, I estimate the economic and competitive costs of these discriminatory practices, highlighting significant impacts on the team's potential performance.

Keywords Discrimination, Football, Outcome test, Benchmark test

1. Introduction

The identification and measurement of labor market discrimination remains a central challenge in empirical economics. While racial disparities are well-documented across various sectors (Lang & Kahn-Lang Spitzer, 2020), establishing causal evidence of discrimination is complicated due to data limitations and identification challenges (Neumark, 2018). This paper leverages the football labor market's unique features—specifically its rich performance metrics and clear career progression paths—to examine racial discrimination in recruitment, particularly for its main position: The quarterback (QB).

By leveraging a unique dataset that combines players performance and physical characteristics throughout high school, college, and the National Football League (NFL), I use two robust empirical strategies to assess racial discrimination during critical career transitions: a benchmark test to evaluate recruitment probabilities conditional on performance characteristics, and an outcome test to identify disparities in realized performance. Building on the framework developed by Gaebler & Goel (2024), these methods jointly address the identification challenges associated with measuring discrimination in complex labor markets. By incorporating data spanning high school, collegiate, and professional levels, this study provides robust evidence of racial bias in recruitment for the QB position. The findings underscore not only the persistence of these biases but also their economic implications for team performance and franchise valuation.

Working Paper

Last updated Jan 13, 2025

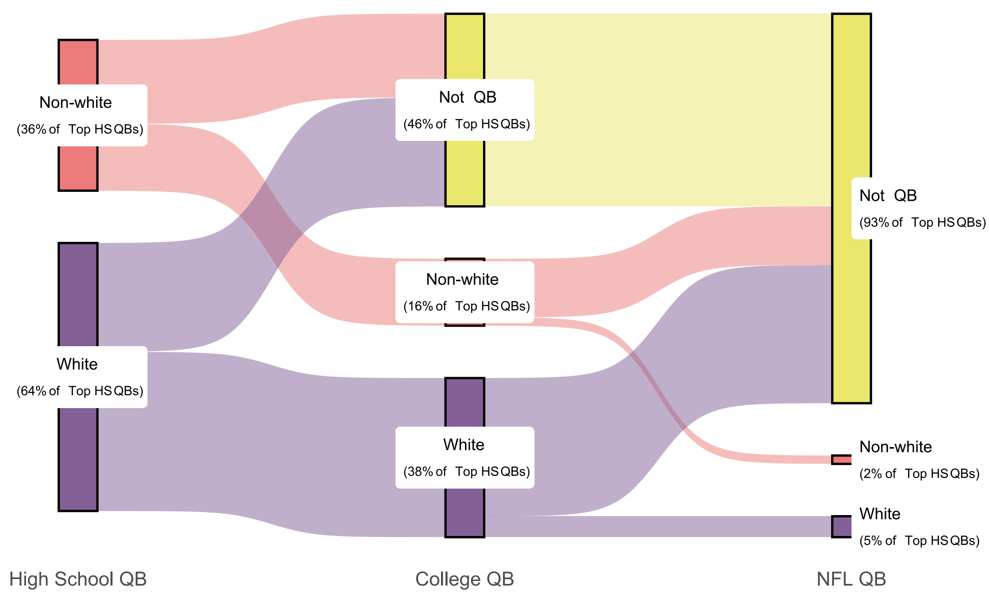
Key Points

- Evidence shows that there are large disparities in football recruitment by race for the QB position.
- Using a benchmark and an outcome test, I show that these results are consistent with racial discrimination.
- Findings are robust to other explanations (e.g. position tracking) and moderately robust to hidden bias.
- Bias in recruitment causes teams to reduce their probabilities of winning by 8%.

Correspondence to
Magdalena Bennett
m.bennett@austin.utexas.edu

Data Availability
Associated code will be made available on GitHub.

Even though American football has been highly regarded as a sport where most players are African-American, this metric is different for the QB position (Bouchet, 2023; The Washington Post, 2021). For example, Figure 1 shows the transition through college and professional football for 2256 top QBs in high school (HS) between the years 2000 and 2018. Within top ranked QBs, only one third are players of color, which already shows potential unbalances in this position from an early start. However, these unbalances are only exacerbated as players transition to upper levels: While 58.6% of white HS QBs in our data are recruited as QBs for college, that number is only 44.2% of non-white QBs. These differences are sustained when advancing to the professional league, where white HS QBs are 40.6% more likely to be drafted as NFL QBs compared to HS QBs of color. The main question that remains after observing these patterns is whether these decisions are due to racial discrimination or other potential explanations, and it is the main focus of my analysis.



Note: The dataset includes 2256 high school quarterbacks ranked as top performers in their respective year (top HS QBs) from 2000 to 2018. Source: ESPN and 247Sports HS QB rankings (HS), ESPN QBR and Sports-Reference (college), and Pro-Football-Reference (NFL). See Data section for more details.

Figure 1: Proportion of top high school quarterbacks from 2000 to 2017 by stage

This paper makes several contributions to the literature on labor discrimination. Firstly, it expands on current applications by analyzing a case study that allows us to inspect not only recruitment processes (Bertrand & Mullainathan, 2004; Jacquemet & Yannelis, 2012; Kline et al., 2022; Pager et al., 2009), but also selection into specific positions, while having comprehensive measures of performance and contributions to the franchise. By collecting rich information on potential recruits (e.g. top high school QBs and college QBs), their qualifications, teams' interests, and players' performance after selection, I can compare results for players that were not selected by specific teams or were not recruited at all.

Another contribution of this paper is that, unlike most discrimination literature that focus either on a benchmark test or an outcome test (Arnold et al., 2018; Becker, 1993; Fryer et al., 2013; Kleinberg et al., 2018), I look at the complementarity of both, providing robust evidence of discriminatory patterns, even under alternative explanations. These results complement other studies that have focused on analyzing discriminatory practices in sports on other fronts, such as referee assessment or salary [Price & Wolfers (2010); parsons2011; berri2022], highlighting the hurdles for minorities to even get to the game.

Finally, this paper also provides evidence of discrimination at different stages of career advancement, by showing results for the transition from high school to a division 1 college team, and from college to the most elite football league, the NFL. Analyzing different stages of recruitment can be seen akin to differences in recruitment according to seniority, allowing us to better understand where are the potential bottle necks in terms of racial disparities.

The rest of this paper is structured in 5 sections, in addition to this introduction. Section 2 details the data sources and provides descriptions for the final datasets used for analysis. In Section 3, I explain the empirical strategy used for assessing discrimination patterns. Results are presented in Section 4, including the analysis of potential alternative mechanisms. Cost analysis for racial bias are shown in Section 5. Finally, Section 6 concludes.

2. Data

2.1. Sources

To analyze the progression of racial disparity in the quarterback position, I collect data and merge several sources:

A. *High School level*

A.1) *ESPN Top Quarterbacks (2006 - 2018)*: Data set from espn.com, which includes the top 100 recruits by position (QBs) each year. This dataset includes information about their college selection (if any), weight, height, high school, grade, and ranking.

A.2) *247 Sports data for top high school quarterbacks (2000 - 2018)*: Information for top high school quarterbacks each year (according to 247 Sports ranking), for both pro-style and dual threat QBs. Data includes physical measures, as well as scouts' ratings.

B. *College level*

B.1) *SR player Stats for College Football (2000 - 2023)*: Dataset that harmonizes information from sports-reference. Includes information about player's position, and passing, rushing, and scoring stats, if available.

B.2) *ESPN College Football Total QBR (2004-2024)*: Data set from espn.com, including 135 QBs per year, their ranking, and other statistics.

B.3) *ESPN College Football Stats (2006-2023)*: Data set from espn.com, with information about passing, rushing, and receiving leader boards for each year. It also includes specific statistics for each board, such as passing/ rushing yards, touchdowns, and interceptions, among others. As a reference, the passing leader board includes around 500 players per year (mostly QBs), while the rushing and receiving leader board include over 1200 players each.

B.4) *College Football Team Ratings (2000-2023)*: Information at the college team level for each year, including the Simple Rating System (SRS), as well as the record of wins and losses for the season.

B.5) *Season games (2000-2023)*: Results for each game in the season for Division I teams. I also include information for the starting QBs in each game for both teams.

C. *NFL level*

C.1) *Draft picks (2000-2024)*: Dataset with all draft picks by year and position.

C.2) *NFL Total QBR (2006-2024)*: Dataset developed by ESPN for starting QBs in the regular season of the NFL. Includes Adjusted Total Quarterback rating, points contributed, and number of plays, among others.

Additionally, for all harmonized datasets from high school QBs and college QBs, I collected race data. This variable was manually gathered and is a *perceived race* classification based on images of the players, with three distinct categories: `White`, `Black`, and `Other`. Even though this variable does not capture the actual race of the player, it will be a proxy for race perception and how this aspect affects their selection at different stages.

2.2. Description

Using the previously mentioned sources, I built two different datasets: (a) high school dataset (HS data) and (b) college dataset.

The high school dataset consists of a panel of 2256 HS quarterbacks, mostly from the 247Sports website (75.6%) from 2000 to 2018. In terms of HS variables, it includes identifying data (e.g. name, high school, city, state), physical characteristics (i.e. height and weight), performance metrics (e.g. scouts' rating and ESPN and/or 247Sports ranking for the year), style of play (pocket passer or dual threat), and recruitment data (e.g. name and total number of schools interested). These individuals are then merged by name[1] to potential candidates within 6 years of their high school ranking.[2] College data for the HS panel includes the player's position in college, ESPN Total Quarterback Rating (QBR) — a score that measures the overall QB contribution to the game —, passing, and receiving statistics. These metrics are available per year and are matched to the HS QB for all years available within the permitted window. Finally, I match this panel to the draft pick dataset using a similar approach as before[3], and includes draft season, round, and pick, in addition to the team that drafted them and the player's position.

Table 1 shows the characteristics for the high school dataset, where 35.6% of the QBs included are non-white and 64.4% are white. In terms of physical characteristics, QBs of color tend to be slightly smaller in high school, but hold higher rankings and ratings compared to their white counterparts. In terms of style of play, most non-white QBs are classified as "Dual Threats" instead of "Pocket Passers", which refers to players who can both run and pass the ball.

In terms of college recruitment, both groups are equally as likely to advance to play collegiate football, but there is a large disparity in their position: While 86.7% of white QBs that are recruited play in that same position in college, only 67.9% of non-white recruits stay as a QB in college. Comparing their college performance, their statistics show a clear difference in playing style, with QBs of color attempting much more rushing plays than white QBs, who have higher metrics in passing statistics. However, it is important to note that when comparing adjusted average passing yards per attempt[4], both groups have comparable performance.

[1] I perform a fuzzy match of the first and last name of the player within the college years permitted, using a maximum distance (*Jaro-Winkler*) of 0.1. If there are multiple matches, the match with the smallest distance is the once that is kept. As a robustness check, I also conducted a manual search for players with no matches using additional data (e.g. matching HS hometown) and changed some spellings accordingly.

[2] Potential matches for HS players come from a pool of college players with a starting year greater or equal to y and a final year of $y + 6$ at most, where y is the year where the player was on the HS ranking. This allows us to include players that are ranked during their junior year, for example. Starting and final year refers to the first and last year the college player appears on the dataset.

[3] Draft data is merged using a fuzzy matching by first and last name, within 3 and 6 years after high school ranking.

[4] Adjusted passing yards per attempt is a formula calculated by *Sports-Reference* to take into account other characteristics of the play, such as interceptions and touchdowns. The metric is calculated as $\text{Adj. avg pass yds per att} = Yds + 20 \times TD - 45 \times Int/Att$ (Sports References, 2024)

Table 1: High school panel - Data Description

Variable	All	White HS QB (N = 1456) Mean	White HS QB (N = 1456) SD	Non-white HS QB (N = 802) Mean	Non-white HS QB (N = 802) SD	Diff.	p-value
High School							
Height (ft)	6.207	6.227	0.134	6.172	0.144	−0.055	0.000
Weight (lb)	197.601	199.029	14.769	195.020	16.168	−4.009	0.000
Ranking (ESPN or 247sports)	32.568	34.601	23.899	28.897	19.818	−5.704	0.000
Rating (ESPN or 247sports)	82.792	82.661	7.716	83.028	7.492	0.367	0.270
Pocket Passer	0.520	0.694	0.461	0.216	0.412	−0.479	0.000
Dual Threat	0.480	0.306	0.461	0.784	0.412	0.479	0.000
Year HS	2010.363	2010.032	4.696	2010.963	4.816	0.931	0.000
College							
College recruit	0.667	0.676	0.468	0.650	0.477	−0.025	0.228
College QB	0.535	0.586	0.493	0.442	0.497	−0.145	0.000
QBR	57.769	57.050	17.085	59.425	17.752	2.375	0.133
Comp. pass	106.912	111.899	106.186	95.907	95.679	−15.992	0.009
Attempts pass	178.356	185.838	168.437	161.849	153.066	−23.989	0.015
Adj. avg passing yds	5.883	5.799	4.981	6.074	4.039	0.275	0.317
Attempts rush	71.315	63.900	38.483	87.938	52.827	24.038	0.000
Avg. rushing yds	1.362	0.714	2.554	2.854	2.467	2.140	0.000
Points scored	15.977	13.434	17.230	20.810	22.941	7.376	0.000
NFL							
Drafted	0.086	0.093	0.290	0.075	0.263	−0.018	0.127
Drafted as QB	0.076	0.083	0.276	0.062	0.242	−0.021	0.059

Note: College statistics per player correspond to the last year of college available.

For the college data, I built a panel with 7407 observations and 3132 unique QBs (68.5% white). In terms of college performance, and unlike the HS data we previously observed, there is no statistically significant differences in most passing stats between white and non-white QBs in college, but QBs of color still show better results in terms of rushing metrics. With respect to overall performance, it appears that non-white QBs have overall better rankings (in passing and rushing), and a general edge in terms of their QB Total Rating (QBR), the metric created by ESPN to encompass the contribution of the QB to the team.

Regarding college QBs' performance in the NFL, only a few get drafted and play as QBs in the National Football League. However, QBs of color still outperform white QBs in terms of QBR and their ranking in the NFL.

Table 2: College panel - Data Description

Variable	All	White QB (N = 2168)		Non-white QB (N = 1012)		Diff.	p-value
		Mean	SD	Mean	SD		
High School							
HS Top QB	0.434	0.430	0.495	0.440	0.497	0.010	0.611
College							
Games (pass)	6.863	6.621	3.421	7.429	3.339	0.808	0.000
Comp. pass	68.181	69.786	77.286	64.437	69.996	−5.348	0.081
Attempts pass	116.078	118.294	124.726	110.909	114.305	−7.385	0.138
Adj. avg passing yds	5.246	5.155	6.559	5.461	7.048	0.305	0.309
Interceptions	3.541	3.637	3.594	3.317	3.188	−0.320	0.023
Pass. efficiency rating	112.979	112.065	67.216	115.148	68.551	3.083	0.298
Games (rush)	5.070	4.833	3.448	5.454	3.550	0.621	0.049
Attempts rush	53.116	48.426	45.891	62.760	62.730	14.334	0.000
Avg. rushing yds	1.686	1.038	2.567	3.059	2.461	2.020	0.000
Points scored	12.335	10.282	12.182	15.926	17.289	5.644	0.000
Ranking (pass) ESPN	68.141	66.277	33.842	72.565	35.699	6.288	0.011
Ranking (rush) ESPN	675.300	791.751	484.935	398.982	361.937	−392.770	0.000
QBR college	53.889	53.109	15.974	55.608	16.595	2.499	0.022
Ranking (QBR) ESPN	64.989	66.249	32.246	62.212	34.447	−4.037	0.072
NFL							
Drafted	0.097	0.094	0.292	0.103	0.305	0.009	0.421
Drafted as QB	0.078	0.082	0.274	0.068	0.252	−0.014	0.159
Draft round	3.918	4.059	2.164	3.637	2.193	−0.422	0.113
Draft pick	116.974	122.535	81.494	105.961	81.123	−16.574	0.095
QBR NFL	49.006	47.461	12.194	52.140	13.775	4.679	0.093
Ranking (QBR) NFL	20.396	21.183	7.842	18.800	9.116	−2.383	0.191
Play as NFL QB	0.034	0.033	0.179	0.035	0.185	0.002	0.732

Note: College statistics per player correspond to the average of all years available.

3. Empirical Strategy

As an empirical strategy, I employ a method developed by Gaebler & Goel (2024) to test for discrimination in recruitment decisions for quarterbacks across different stages of their careers. The approach leverages two widely used tests in discrimination research: the benchmark test and the outcome test. The benchmark test compares recruitment probabilities across racial groups, adjusting for observable performance indicators such as player's statistics and physical attributes. The outcome test, in contrast, assesses whether differences in realized performance, conditional on recruitment, suggest unequal standards for different groups. While each test has limitations when used in isolation, Gaebler & Goel (2024) demonstrate that their combined application under a monotonicity assumption produces robust conclusions about discrimination.

In the context of this study, this combined method evaluates whether racial disparities in quarterback recruitment can be attributed to different performance thresholds applied to white and non-white players. Specifically, the monotonic likelihood ratio property ensures that at least one of the tests — benchmark or outcome — yields correct conclusions when discrimination exists. By applying this hybrid test to data from high school, college, and NFL recruitment processes, I am able to identify whether observed disparities are consistent with racial discrimination.

4. Results

As previously mentioned, I use two different tests to assess whether recruitment behaviors are consistent with discrimination. The first one is the benchmark test, where I estimate the difference in probability of being recruited *as a QB in college/NFL* for two different groups (i.e. white and non-white QBs) adjusted by multiple potential confounders, such as style of play, scouts' metrics, and physical characteristics for the high school-to-college transition, and game performance for the college-to-NFL stage. Even though these characteristics are relevant determinants of recruitment for the QB position, there could be other unobserved confounders that could skew this test, like leadership skills or other metrics I do not have access to (e.g. hands size and shoulder width).

The second test corresponds to an outcome test, which measures the differences in performance by group *for those that where selected as QBs* in the next stage. As an outcome for high school-to-college transition as well as the college-to-NFL progression, I use the Total QBR rating, as this is the most comprehensive metric in the data for QB performance, unlike the NFL Passer Rating[5] that does not take into account the play level or other intricacies of the game. The Total QBR rating is a complete measure on how the QB "impacts the game on passes, rushes, turnovers, and penalties" (ESPN Stats & Information, 2016). Unlike passing yards statistics, QBR also considers the success or failure at the play level, providing context to the measures.

As it has been shown in the literature, each of these tests have several shortcomings (Canay et al., 2024) for detecting discrimination, as other behaviors might produce the same patterns. For example, if white QBs have better leadership skills than non-white QBs, it could explain the differences in recruitment and the benchmark test would not be capturing racial discrimination, but differences in potential performance that cannot be measured in the data. For the outcome test, a common threat is inframarginality, where the underlying distributions of risk or outcomes differ by race. In the case of QB performance, if non-white players have, on average, higher performance than white QBs because of physical characteristics or speed, then we should observe better results for the former group, even if there is no discrimination in recruitment.

The point that Gaebler & Goel (2024) make is that, under mild monotonicity conditions, both tests cannot be wrong at the same time. If non-white QBs have a lower probability of being recruited than white players because they lack certain skills, then the outcome test should show a lower performance on the next stage, not higher. Similarly, if recruiters believe that overall QBs of color perform better than white QBs, they should be recruited at higher rates. Thus, in order to obtain robust evidence of racial discrimination in the recruitment process (i) non-white QBs have to be less likely to be recruited than white QBs, conditional on certain characteristics, and (ii) QBs of color have to show better performance than white players.

For this robust test to provide conclusive evidence, the monotone likelihood ratio property (MLRP) has to hold. The authors show that this assumption is equivalent to assuming that the probability of the lower base group (e.g. QBs of color) given a specific risk rate r is monotonic.

4.1. High school to College

4.1.a. *Benchmark test:* For the benchmark test, I estimate the probability of a top HS QB to be recruited into college *as a QB* by race group, adjusting by relevant characteristics that are of importance in the recruitment process. Some of these characteristics are scouts' HS rating (from 0 to 1), playing style classification (i.e. pocket passer or dual threat), as well as physical characteristics such as height and weight. I also include a binary variable to indicate the source of the data (ESPN or 247Sports).

[5]The NFL passer rating combines passing yards, as well as passing touchdowns and interceptions

Results in Table 3 show that HS QBs of color are, on average, 12 percentage points less likely to be recruited, even after controlling for performance and physical characteristics. These findings hold even after adjusting by playing style, which is one of the main reasons analysts argue that Black QBs are recruited at a lower rate than white QBs (Berri & Simmons, 2009; Bopp & Sagas, 2014).

Table 3: Benchmark Test for QB recruitment from high school to college

	College QB	College QB	College QB	College QB
NonWhiteQB	−0.145***	−0.117***	−0.118***	−0.121***
	(0.022)	(0.024)	(0.023)	(0.023)
Pocket Passer		0.062**	0.083***	0.092***
		(0.023)	(0.023)	(0.024)
HS Rating			0.022***	0.023***
			(0.002)	(0.002)
ESPN			−0.071*	−0.069*
			(0.033)	(0.033)
Height (in)				−0.055
				(0.082)
Weight (lb)				−0.001+
				(0.001)
Avg. Outcome (Control)	0.586	0.586	0.586	0.586
Years	00-18	00-18	00-18	00-18
N	2256	2256	2256	2254

Note: Robust standard errors in parenthesis.

p-value < 0.01: **, <0.05 *, and <0.1 +

If we look at these results by year, before 2011 there is a significant difference in non-white QB presence in colleges, even after adjusting for HS performance and play style. After 2011, however, these differences become smaller, which could be due to a larger presence of successful Black QBs in the NFL.

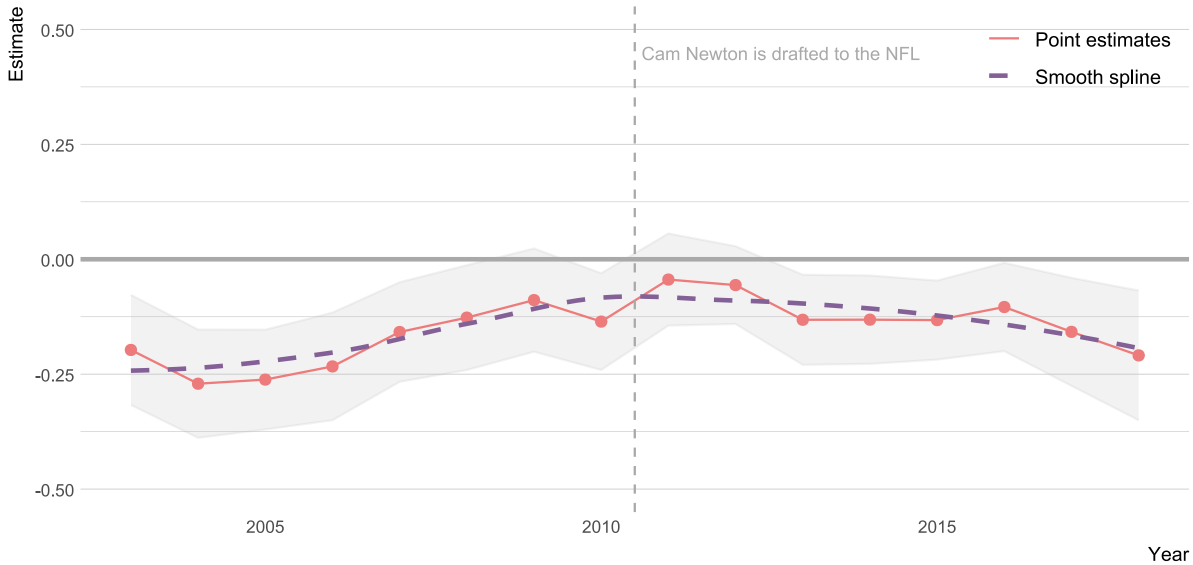


Figure 2: Marginal effects by year for college QB recruitment by race

In order to provide additional evidence for the benchmark test, also conducted a robustness tests to assess the possibility of unobserved confounding by using a sensitivity analysis to hidden bias (Rosenbaum, 1987). Using the `designmatch` package (Zubizarreta et al., 2018), I match a white QB with a non-white QB, matching exactly on play style (e.g. pocket passer or dual threat) and balance them on ranking, rating, height, and weight with a tolerance of 0.025 SD. I also set a fine balance restriction for year of graduation, and exact match players based on data source availability: (i) ESPN only, (ii) 247Sports only, or (iii) both. Table 4 shows the balance between both matched groups, where both white and non-white HS QBs now have comparable recruitment characteristics.

Table 4: Balance table for matching white and non-white HS QBs on HS performance and physical characteristics

	White HS QB (N = 509)		Non-white HS QB (N = 509)			
Variable	Mean	SD	Mean	SD	Diff.	p-value
Height (ft)	6.19	0.14	6.19	0.13	0.00	0.99
Weight (lbs)	196.24	16.24	196.37	14.30	0.13	0.90
Ranking (ESPN)	26.58	19.10	26.81	18.63	0.24	0.88
Grade (ESPN)	76.55	4.72	76.61	4.90	0.06	0.88
Ranking (247Sports)	24.10	14.35	24.14	13.57	0.04	0.97
Rating (247Sports)	0.85	0.06	0.85	0.06	0.00	0.89
Dual Threat (247Sports)	0.70	0.46	0.70	0.46	0.00	1.00
Year	2010.24	4.92	2010.24	4.92	0.00	1.00

Note: Mean restricted balance at 0.25 SD, fine balance for HS graduation year, and exact match on data source.

In this case, the difference in probability between matched groups is –0.106 (*95% CI = [–0.166,-0.046]*), which means that QBs of color, after matching by play style classification, performance, physical characteristics, and year, are 10.6 percentage points less likely to be recruited as a college QB than a white high school quarterback. There is also a 0.073 (*95% CI = [0.026,0.12]*) increase in probability for black QBs of being recruited in a different position for college.

The advantage of matching units is that we can directly compare both groups without using extrapolation or relying on parametric functional form assumptions. Additionally, this approach lends itself nicely to simple and straightforward sensitivity analysis to hidden bias (Rosenbaum, 2002). The idea behind this type of sensitivity analysis relies on the fact that, within a matched set in a randomized experiment, each unit has the same probability of assignment to treatment (50% in a two-way matching design). In an observational study, however, there can be unobserved confounding that skews these probabilities. In this case, for a matched pair of QBs of different race, we will assume that one unit is $\Gamma > 1$ times more likely to be recruited as a college QB based on unobserved characteristics, and those characteristics can explain away any significant differences we find.

For this specific design, I find a sensitivity to hidden bias of $\Gamma = 1.33$, which means that an unobserved confounder should change the probability of assignment from 50%-50% to 42.9% - 57.1% in order to qualitatively explain away our findings, which implies that our estimate is moderately robust to hidden bias, according to literature parameters. To put this measure in context of other characteristics, the prior result would be equivalent to the difference in QB recruitment for two matched groups that are balanced in expectation on all characteristics (observed and unobserved), with the exception of an imbalance of 0.32 SD in rating score (247 Sports). This means that an unobserved confounder that has not been accounted for would have to have the same effect on QB recruitment as a 0.32-standard-deviation difference in HS QB performance based on scouts ratings.

4.1.b. *Outcome test:* For the outcome test, as it was previously described, I use ESPN's total QBR as a measure of performance in college. This metric is only available for players with more than 20 plays within a game, so usually only starting QBs are able to be properly assessed.

Table 5 show that non-white college QBs outperform white players in the same position by approximately 1.8 points (3.1%), providing evidence that recruiters could have a higher bar set for players of different race.

Table 5: Outcome Test for QB performance in college based on race

	QBR	QBR
NonWhiteQB	2.093*	1.814*
	(0.911)	(0.891)
Avg. Outcome (Control)	55.73	56.207
Year FE	No	Yes
Years	04-24	04-24
N	2547	2547

Note: Clustered robust standard errors in parenthesis. Clusters at the player level.

p-value < 0.01: **, <0.05 *, and <0.1 +

4.1.c. *Robust test for discrimination:* The results previously presented would provide robust evidence of racial discrimination if and only if the probability of being in the minority group conditional on a risk level — in this case, the performance metric QBR — is monotonic. I choose QBR as a risk measure because given its range (0-100) it can be seen as a probability of success (or a probability of the QBs contribution to game success).

To assess monotonicity, I need to estimate $Pr(NonWhiteQB \mid QBR = q)$ for our entire sample. Given that we only observe the QBR for QBs that are recruited into college *as QBs* (and also are starting players), I use a prediction model to impute the missing values in the distribution. Specifically, I use a random forest approach[6],

[6] In the model, I use as predictors: HS rating and ranking, height, weight, style of play, and race, in addition to adding college year as a factor. After tuning the hyper-parameters using 5-fold cross-validation, the model randomly selects 28 predictors with a minimum node size of 1.

withholding 10% of the sample. The model performs well, with a normalized RMSE of 27.2% with respect to the testing sample.

Figure 3 shows the predicted QBR for the entire sample at different levels of QBR, showing overall a monotonically increasing function.

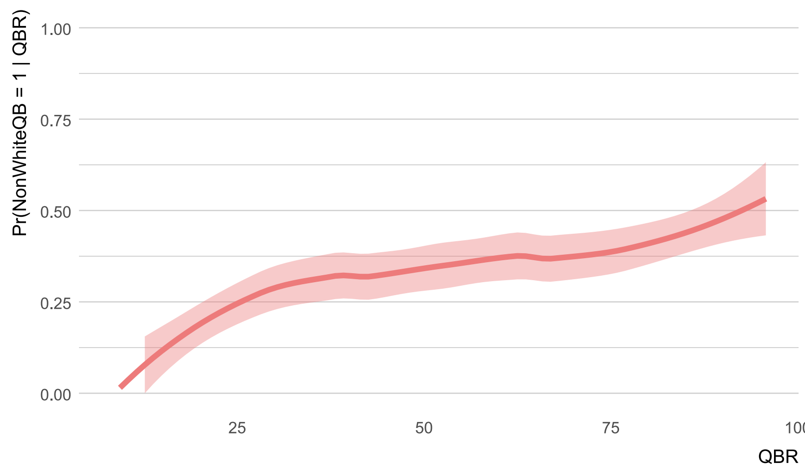


Figure 3: Proportion of HS QBs of color for different levels of predicted QBR

Given the results provided in Table 3 and Table 5, evidence points to the fact that recruitment practices for the QB position appear to be discriminatory towards players of color.

4.2. College to the NFL

4.2.a. *Benchmark test:* Analyzing the transition between college and the NFL, I only consider players that played as QB in college. In total, I observe around 3,200 QBs between the years 2000 and 2023, and around 1,500 of those QBs are present in the ESPN Total QBR dataset[7]. Results in Table 6 show that college QBs of color are 2.2 percentage points less like to be drafted as QBs to the NFL compared to white QBs, even after controlling for their QBR score, when available. This difference increases to 3.7 percentage points when adjusting for other statistics, such as games played and their adjusted average passing yards.

[7] ESPN Total QBR is only available since 2004 and only has records of QBs with at least 20 plays.

Table 6: Benchmark Test for QB NFL draft for College QBs

	Drafted QB	Drafted QB	Drafted QB	Drafted QB
NonWhiteQB	−0.018	−0.022+	−0.029+	−0.037*
	(0.014)	(0.012)	(0.015)	(0.015)
QBR		0.009***	0.009***	0.009***
		(0.001)	(0.001)	(0.001)
QBR (miss)		−0.192***	−0.192***	−0.135***
		(0.013)	(0.014)	(0.016)
Adj. Pass Yards			0.002***	0.002***
			(0.000)	(0.000)
Games				0.010***
				(0.002)
Avg. Outcome (Control)	0.115	0.115	0.136	0.136
Year FE	Yes	Yes	Yes	Yes
Years	00-23	00-23	00-23	00-23
N	7407	7407	5986	5986

Note: Clustered robust standard errors in parenthesis. Clusters at the player level.

p-value < 0.01: **, <0.05 *, and <0.1 +

4.2.b. *Outcome test:* In terms of outcomes in the NFL, I use the ESPN's Total QB Rating for the same reasons why the same rating was used for college QBs. Even though it is not a perfect measure of performance, it is the most comprehensive measure for measuring the contribution of a QB (ESPN Stats & Information, 2016; Stuart, 2014). Table 7 shows that even within a small sample (approximately 100 QBs), QBs of color outperform white NFL QBs by 4.1 points (though this last point estimate is not statistically significant at conventional levels).

Table 7: Outcome Test for NFL Total QB Rating by race

	NFL QBR	NFL QBR
NonWhiteQB	5.747*	4.053
	(2.913)	(3.521)
Avg. Outcome (Control)	47.578	47.578
Year FE	No	Yes
Years	06-24	06-24
N	338	338

Note: Clustered robust standard errors in parenthesis. Clusters at the player level.

p-value < 0.01: **, <0.05 *, and <0.1 +

4.2.c. *Robust test for discrimination:* In order to analyze the monotonicity condition necessary to conclude that there are discriminatory practices in NFL recruitment, I assess the rate of non-white QBs at different levels of QBR. Figure 4 shows a monotonically increasing function of non-white QBs on NFL performance, which, in combination with the benchmark and outcome test for the college-NFL transition, are indicative of racial bias in QB recruitment.

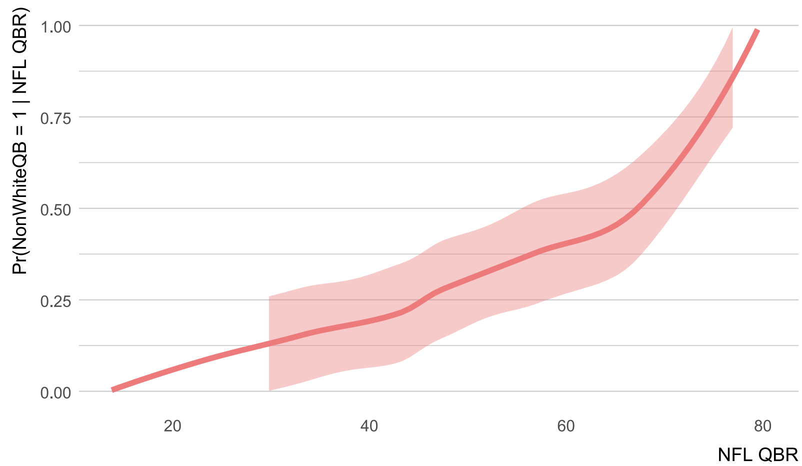


Figure 4: Proportion of College QBs of color for different levels of predicted NFL QBR

4.3. Alternative mechanisms: Switching positions

Previous results show that recruiters have a higher bar for non-white QBs with respect to white QBs, consistent with discriminatory behavior. However, someone could argue that given that football is a team sports, not selecting a player for the QB position but *for another position in the team* could, overall, increase the performance of the team as a whole. This explanation could be plausible, given that there is no significant difference by race in the probability of being recruited into college, but there is a large difference when assessing QB recruitment.

For the sake of testing this hypothesis, let's assume that QBs of color have an advantage on other positions with respect to white players. Given that the data shows the largest transition for non-recruited QBs to the position of wide receiver in college (39.6%), speed could play an important role in this decision.

Table 8 shows the results for the same benchmark test that was used in Table 3, but adding information on the 40-yard dash from high school combines[8], in addition to an accuracy measure also captured in the event. Given that only a few players have combine data, the sample size is reduced significantly and results are no longer statistically significant. However, the point estimate still suggests a lower probability of non-white QBs being recruited to that position for college.

[8]Data on 40-yard dash is not available for HS QBs that are exclusively in the ESPN ranking (and not in 247Sports).

Table 8: Benchmark test for HS to College transition including combine results

	College QB	College QB	College QB
NonWhiteQB	−0.102***	−0.078	−0.127
	(0.027)	(0.070)	(0.146)
Pocket Passer	0.157***	0.082	0.087
	(0.029)	(0.077)	(0.173)
HS Rating	0.014***	0.016	−0.029
	(0.004)	(0.012)	(0.036)
HS Rank	−0.006***	−0.008	−0.022+
	(0.002)	(0.005)	(0.012)
Height (in)	−0.155	−0.139	0.059
	(0.099)	(0.300)	(0.605)
Weight (lb)	−0.001	0.003	0.005
	(0.001)	(0.003)	(0.006)
Forty-yard dash		0.216	−0.012
		(0.208)	(0.341)
Accuracy			0.010
			(0.058)
Avg. Outcome (Control)	0.67	0.682	0.774
Years	00-18	02-17	13-17
N	1703	221	49

Note: Robust standard errors in parenthesis. Observations only for HS QBs in 247Sports

p-value < 0.01: **, <0.05 *, and <0.1 +

In terms of advantages, I also analyze whether HS QBs that are switched to a different position in college perform better than their counterparts. If that is the case, it could make sense to track certain QBs to other positions if they have an overall advantage over players in persistent positions. I once again focus on the wide receiver position, given that it has the largest share of tracked QBs.

Table 9 shows the results for different metrics of performance by a wide receiver, including average receiving yards, touch downs from receptions, and receptions in general. In all metrics, non-white HS QBs that were tracked into this position perform no better (and even worse) than their counterparts, which suggests that they do not stand out in performance compared to non-HS QBs.

Table 9: Tracked QB performance as a wide receiver vs persistent wide receivers

	Avg. Receiving Yards	TD Receptions	Receptions
Top HS QB non-white	−0.049	−1.367***	−9.579***
	(1.721)	(0.168)	(1.958)
Avg. Outcome (Control)	12.506	1.906	21.029
Year FE	Yes	Yes	Yes
Years	02-23	02-23	02-23
N	16785	16843	16829

Note: Clustered robust standard errors in parenthesis. Clusters at the player level.

p-value < 0.01: **, <0.05 *, and <0.1 +

5. How much does discrimination cost?

One natural question that follows the previous analysis is "How much does discrimination cost a team?". In this case, cost can be translated into different measures, such as monetary earnings for the franchise, team wins, or game points.

To analyze the cost of discrimination in terms of team performance, I use the results for each college game between 2005 and 2023. In general, there are around 700 games per year (including playoffs), and each of them can be considered a match-up. I build a dataset with all teams' games and their final results, including each squads' and their opponents' statistics for the previous year[9]. Additionally, to track the specific contribution of the QB, I identify the main QB that played in each game[10] and use their QBR for the prior year as a predictor. To account for the strength of the offensive team outside of the QB, I use the residuals of a regression between the team's offensive score and the QBR for that year's main QB. The rest of the predictors for the model are based on the teams' and their opponents' defensive statistics. Given that not all QBs have QBR metrics, I only use approximately 130 games from each season for this analysis.

Using this data, I build a prediction model for winning a game (binary). As previously mentioned, I use the aggregated information of the prior year for both the incumbent and the opposing team as predictors, such as Simple Rating System (SRS)[11] score for the defense (Sports Reference, 2024), the offense SRS contribution aside from the QB (i.e. residuals), and I also include as predictors the main QB's Total Quarterback Rating (QBR) for the previous year for both contending QBs in a match-up.

For prediction, I use 75% of the match-up data for training and the rest for testing[12]. The prediction model is a Random Forest using 5-fold cross-validation[13], with an accuracy of 66.4%, a sensitivity of 66.3%, and specificity of 66.5%.

Results in Figure 5 show that the strength of a QB is one of the most important predictors for a win (e.g. 99.6% of relative importance). This finding is relevant considering that non-white QBs have, on average, 2 additional points

[9]Given that current year statistics are affected by the results of each game, I use the prior years' statistics to avoid using post-game information in the prediction process

[10]The main QB for a game is considered the one that had the most attempts.

[11]The Simple Rating System is a measure that takes into account "average point differential and strength of schedule" (Sports Reference, 2024), and it captures points either above or below average

[12]Given that each match-up appears twice (once for the incumbent and once for the opponent as the incumbent), the randomization for the training and testing data is done at the game level to avoid contamination.

[13]Hyper-tuning parameters include 41 randomly chosen variables, minimum node size of 10, and the gini metric is used to create the splits.

in terms of QBR compared to white QBs. Then, if some teams are recruiting sub-optimally due to discriminatory practices, that could affect their probability of winning.

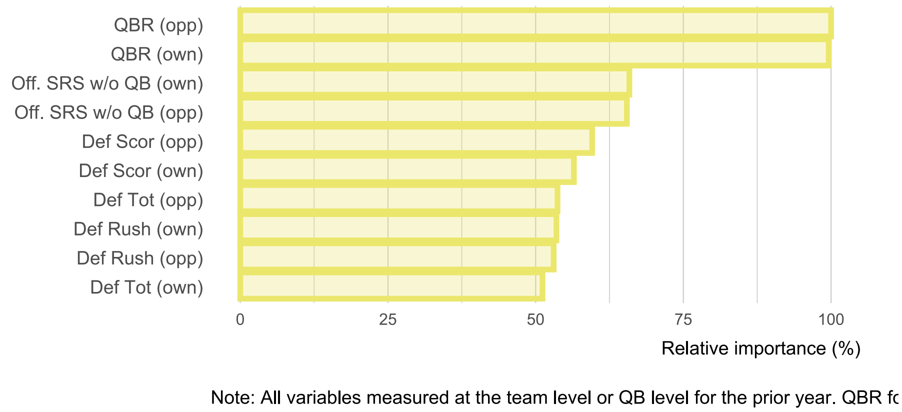


Figure 5: Relative importance of features in prediction model for winning

To assess the cost of bias in terms of wins, I build a simulation scenario where teams would pick a similar QB to their own in term of high school statistics, but without considering race. For this, I match high school data to the previous dataset, and create by-year deciles of HS QBs based solely on their scouts' rating. Then, for each year, a team would randomly select a QB from the same decile as their QB, replacing the QBR stat of their player with the new randomly selected one. With this new team dataset[14], I use the same random forest model previously described to predict the new probabilities of winning.

Figure 6 show the difference in predicted and observed probability of winning based on whether the QB change is from white to non-white or vice-versa[15]. When the simulation randomly selects a non-white QB (instead of the teams' white QB), we can see that there is a slight increase in the probability of winning (2.01 p.p.) compared to the inverse shift.

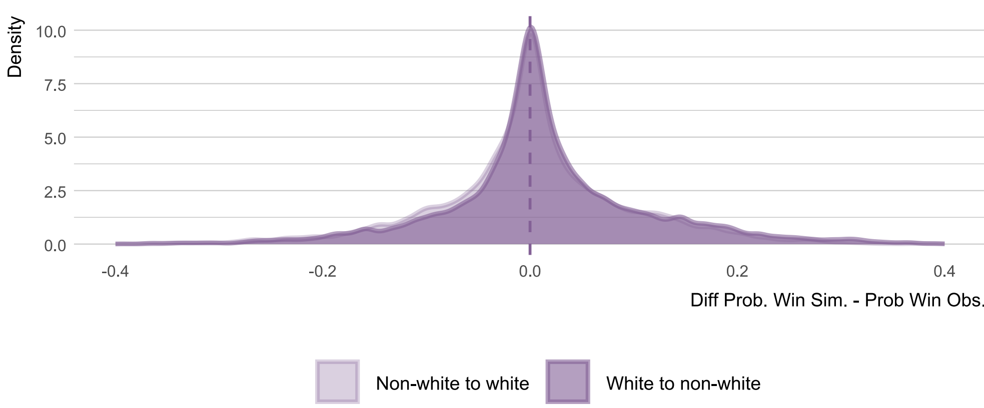


Figure 6: Difference in predicted probability of winning and observed probability of winning for different QB switches

[14]A limitation of this data is that I can only use teams that have a starting QB that is on our HS dataset. Thus, I only use games where I am able to match both datasets.

[15]Changes of QB within the same race/ethnicity are not shown.

I also focus in particular on teams which have a significant lower rate of non-white starting QBs in their games compared to the simulations, and use this as a proxy of over-representation of white QBs in specific teams (`OR white QB`). In this case, if a team has a non-white QB representation that is below the 95% CI of the simulated rate, then `OR white QB` takes the value of 1, and 0 in another case[16].

Table 10 shows the results for additional wins (in number and percentage points) for teams with over-representation of white QBs, comparing their average simulated results and the observed outcome. These teams, who are more likely to have discriminatory patterns of recruitment, would increase their wins by 7.8% if they recruited similar HS QBs in terms of performance, but without considering race in the recruitment. These findings highlight some of the costs that are associated with racial bias in recruiting, specifically due to QB performance[17].

Table 10: Differences in winning between teams with white QB over-representation vs not based on simulations

	Num. extra wins/year	% extra wins/year
OR White QB	0.180**	0.037**
	(0.061)	(0.014)
Avg. Obs. Wins (Control)	2.222	0.47
Year FE	Yes	Yes
Years	05-22	05-22
N	1218	1218

Note: Clustered robust standard errors in parenthesis. Clusters at the team level.

Additional wins are estimated as the difference between simulated wins and observed wins.

p-value < 0.01: **, <0.05 *, and <0.1 +

6. Conclusions

This paper contributes to the growing literature on labor market discrimination by providing robust evidence of racial bias in quarterback recruitment across the high school, college, and professional levels of American football. Using a combination of benchmark and outcome tests, the analysis reveals clear disparities in recruitment probabilities and performance expectations, with non-white quarterbacks facing higher recruitment thresholds despite demonstrating comparable or superior on-field performance. These findings are consistent with theoretical models of discrimination and lend empirical support to the hypothesis that racial bias persists in decision-making processes within this labor market.

The results highlight several important implications for economic theory and policy. First, the findings underscore the role of implicit biases in labor market sorting, even in contexts where performance metrics are readily observable. The disproportionate representation of white quarterbacks at the collegiate and professional levels, despite the comparable performance of non-white quarterbacks, suggests that decision-makers may be overrelying on heuristics or outdated stereotypes rather than objective measures of ability.

Second, this paper emphasizes the economic costs of discrimination. By demonstrating that teams with racially biased recruitment practices underperform relative to their potential, the analysis highlights the inefficiency introduced by discriminatory decision-making. This inefficiency not only reduces the competitiveness of individual

[16] This measure is created at the team level, aggregating their white QB rate by year. In that case, if teams systematically select white QBs over non-white QBs over several years, they are more likely to be categorized as over-representing white QBs.

[17] Simulations only shifts in QB performance measured as their QBR metric, but other costs can also arise if there are positive associations between QB performance and other offensive team performance as well.

teams but may also distort broader market dynamics in the industry, such as salary structures and franchise valuations.

While the results are robust and address many potential confounders, there are limitations that warrant further exploration. For instance, unobservable characteristics, such as leadership skills or off-field behavior, could partially explain some of the observed disparities in recruitment. Although the outcome tests mitigate this concern by showing that non-white quarterbacks outperform their white counterparts when recruited, additional data on non-performance-related attributes could strengthen the analysis, providing additional evidence on the decision-making process of recruitment.

Finally, these findings open several avenues for future research. The use of football as a case study provides a controlled setting with objective performance metrics and clear career paths, but similar methodologies could be applied to other labor markets with hierarchical structures and performance-based promotions. Additionally, exploring how the observed biases interact with other characteristics of the team (e.g. location and how liberal colleges campus are), could yield a more comprehensive understanding of the barriers faced by minority groups in this labor markets.

In conclusion, this study highlights the persistence of racial bias in a high-profile labor market and its implications for both equity and efficiency. The findings call for further promotion of transparency and accountability in recruitment decisions, ensuring that talent allocation aligns more closely with performance metrics. By addressing these biases, not only can we improve fairness and representation, but we can also enhance performance outcomes in ways that benefit all stakeholders in the industry.

References

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial Bias in Bail Decisions. *The Quarterly Journal of Economics*, *133*(4), 1885–1932. https://doi.org/10.1093/qje/qjy012

Becker, G. S. (1993). Nobel Lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy*, *101*(3), 385–409. https://doi.org/10.1086/261880

Berri, D. J., & Simmons, R. (2009). Race and the Evaluation of Signal Callers in the National Football League. *Journal of Sports Economics*, *10*(1), 23–43. https://doi.org/10.1177/1527002508327383

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, *94*(4), 991–1013. https://doi.org/10.1257/0002828042002561

Bopp, T., & Sagas, M. (2014). Racial Tasking and the College Quarterback: Redefining the Stacking Phenomenon. *Journal of Sport Management*, *28*(2), 136–142. https://doi.org/10.1123/jsm.2012-0296

Bouchet, A. (2023). The 2023 Racial and Gender Report Card: National Football League. *The Institute for Diversity and Ethics in Sport*. https://www.tidesport.org/nfl

Canay, I. A., Mogstad, M., & Mountjoy, J. (2024). On the Use of Outcome Tests for Detecting Bias in Decision Making. *Review of Economic Studies*, *91*(4), 2135–2167. https://doi.org/10.1093/restud/rdad082

ESPN Stats & Information. (2016). *How is Total QBR calculated? We explain our quarterback rating*. https://www.espn.com/blog/statsinfo/post/_/id/123701/how-is-total-qbr-calculated-we-explain-our-quarterback-rating

Fryer, R. G. J., Pager, D., & Spenkuch, J. L. (2013). Racial Disparities in Job Finding and Offered Wages. *The Journal of Law and Economics*, *56*(3), 633–689. https://doi.org/10.1086/673323

Gaebler, J. D., & Goel, S. (2024). A Simple, Statistically Robust Test of Discrimination. *Arxiv Preprint Arxiv:2407.06539*. https://arxiv.org/abs/2407.06539

Jacquemet, N., & Yannelis, C. (2012). Indiscriminate Discrimination: A Correspondence Test for Ethnic Homophily in the Chicago Labor Market. *Labour Economics*, *19*(6), 824–832. https://doi.org/10.1016/j.labeco.2012.08.004

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, *133*(1), 237–293. https://doi.org/10.1093/qje/qjx032

Kline, P. M., Rose, E. K., & Walters, C. R. (2022). Systemic Discrimination Among Large U.S. Employers. *The Quarterly Journal of Economics*, *137*(4), 1963–2036. https://doi.org/10.1093/qje/qjac024

Lang, K., & Kahn-Lang Spitzer, A. (2020). Race Discrimination: An Economic Perspective. *Journal of Economic Perspectives*, *34*(2), 68–89. https://doi.org/10.1257/jep.34.2.68

Neumark, D. (2018). Experimental Research on Labor Market Discrimination. *Journal of Economic Literature*, *56*(3), 799–866. https://doi.org/10.1257/jel.20161309

Pager, D., Western, B., & Bonikowski, B. (2009). Discrimination in a Low-Wage Labor Market: A Field Experiment. *American Sociological Review*, *74*(5), 777–799. https://doi.org/10.1177/000312240907400505

Price, J., & Wolfers, J. (2010). Racial Discrimination Among NBA Referees. *The Quarterly Journal of Economics*, *125*(4), 1859–1887. https://doi.org/10.1162/qjec.2010.125.4.1859

Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, *74*, 13–26.

Rosenbaum, P. R. (2002). *Observational studies*. Springer.

Sports Reference. (2024). *College Football Stats Glossary*. https://www.sports-reference.com/cfb/about/glossary.html#srs

Stuart, C. (2014). *Is ESPN's QBR the best measure of quarterback play?*. https://www.footballperspective.com/is-espns-qbr-the-best-measure-of-quarterback-play/

The Washington Post. (2021). *Black quarterbacks are taking over the NFL. But there is more ground to be gained*. https://www.washingtonpost.com/sports/2021/09/10/black-quarterbacks-nfl-2021/

Zubizarreta, J. R., Kilcioglu, C., & Vielma, J. P. (2018). designmatch: Matched samples that are balanced and representative by design. *R Package Version 0.3*, *0*.