2015

The Connection Between Race and Called Strikes and Balls

Jeff Hamrick
University of San Francisco, jhamrick@usfca.edu

John Rasp

Article

The Connection Between Race and Called Strikes and Balls

Jeff Hamrick[1,2] and John Rasp[3]

Abstract
We investigate potential racial bias by Major League Baseball umpires. We do so in the context of the subjective decision as to whether a pitch is called a strike or a ball, using data from the 1989-2010 seasons. We find limited, and sometimes contradictory, evidence that umpires unduly favor or unjustly discriminate against players based on their race. Potential mitigating variables such as attendance, terminal pitch, the absolute score differential, and the presence of monitoring systems do not consistently interact with umpire/pitcher and umpire/hitter racial combinations. Most evidence that would first appear to support racially connected behaviors by umpires appears to vanish in three-way interaction models. Overall, our findings fall well short of convincing evidence for racial bias.

Racial discrimination has long been an issue in American culture. Legacies of slavery and Jim Crow laws have tarnished the national image, though public consciousness has increasingly addressed racial issues in recent decades. Since Jackie Robinson broke baseball's ''color barrier'' in 1947, overt racial prejudice has

[1] Masagung Graduate School of Management, University of San Francisco, San Francisco, CA, USA
[2] Rhodes College, Memphis, TN, USA
[3] Stetson University, School of Business Administration, DeLand, FL, USA

Corresponding Author:
Jeff Hamrick, University of San Francisco, Masagung Graduate School of Management, 101 Howard Street, San Francisco, CA 94105, USA.
Email: jhamrick@usfca.edu

become socially unacceptable in mainstream American society. And yet, subtler patterns of discrimination persist.

Sports provide an excellent context for examining these issues. They are socially prominent and attract the interest of a wide range of people from all walks of life. They have historically been one path for upward socioeconomic mobility. Moreover—and crucially—they feature a large number of subjective calls being made by referees. These judgment calls are made by highly trained professionals. But they represent split-second decisions, and as such may be affected by unconscious biases or presuppositions and generally will not be altered by conscious reference to social (politically correct) norms.

We focus on baseball because of its centrality in the American mythos, its episodic nature, and its huge, readily available databases. In particular, we focus on one subjective decision: the determination, by the home plate umpire, of whether a pitch is a ball or a called strike.

The pitcher in baseball throws a pitch from a distance of 60.5 feet at a pentagonal shaped home plate only 17 inches wide. A pitch is in the "strike zone" if it crosses any point of home plate, at a height between the batter's chest and knees. The batter may choose to swing at any pitch. If he does not, the umpire must then make a split-second subjective decision as to whether the thrown pitch was inside or outside the strike zone, thus leading to a called strike or ball. Umpires in Major League Baseball have trained for years in their craft, and on the whole do a remarkably accurate job in this difficult decision-making task. However, the very nature of the task allows for individual idiosyncrasy and subjectivity while permitting unconscious biases to manifest themselves, at least to some degree.

There exists an extensive literature on the influence of racial discrimination in baseball, in the recent (i.e., post-Jackie Robinson) era. Kahn (1991) reviews earlier work in this area. Several authors have examined racial discrimination in salary structures, including Christiano (1986, 1988), Purdy, Leonard, and Eitzen (1994), Bodvarsson and Pettman (2002), and Palmer and King (2006). In general, they find mixed evidence related to the issue. The influence of race upon the length of a player's career is considered by Groothuis and Hill (2008). The extent to which the race of players affects game attendance is investigated by Breckenridge and Goldsmith (2009) and by Tainsky and Winfree (2010).

The effects of racial discrimination go beyond the basic economic considerations of salary and career length. The extent to which player race influences voting for selection to the Baseball Hall of Fame is considered by Jewell, Brown, and Miles (2002) and by Jewell (2003). The extent to which there is racial positional "stacking" (i.e., the tendency for minority players to be assigned to play particular positions) is considered by Margolis and Piliavin (1999). However, no evidence of customer-based discrimination is found in voting for selection to participate in the All-Star Game, by either Hanssen and Andersen (1999) or Depken and Ford (2006).

Until recently, however, scant attention has been paid to potential racial effects in sports from game referees or umpires. Price and Wolfers (2010) studied the game of

basketball, noting that players were assessed more personal fouls when the officiating crew members were of a different race than the player. Parsons, Sulaeman, Yates, and Hamermesh (2011) examined racial influence on the calling of balls and strikes in baseball. Their study used data from the 2004-2008 baseball seasons, and noted some small but statistically significant discriminatory influence during this time frame. However, their primary focus was not on discrimination, but upon the extent to which an umpire's decision-making behavior was modified under various measures of special scrutiny, such as whether the game was well attended or whether an electronic pitch-monitoring system was in place.

This article adds to the literature on racial discrimination in contemporary American sports in a variety of ways. We examine the largely neglected topic of potential bias by umpires. We employ a massive data set, covering a full 22 seasons (1989-2010) of games. We examine the possible influence of a host of covariate factors and consider a wide variety of statistical models, including models that employ high-level interaction terms to better capture the dynamics of human behavior.

Ultimately, we find mixed evidence supporting claims of discrimination in the umpires' decisions. While racial dynamics are at work, our findings fall well short of convincing evidence for racial bias.

Data, Variables, and Time Periods

This study uses pitch-by-pitch data from all Major League Baseball games for the 1989-2010 regular seasons. For each pitch, we identify the outcome (ball in play, swinging strike, called strike, called ball, etc.). Only those pitches that were called strikes or balls are used in the study, as only these pitches involve a subjective decision by the home plate umpire. Pitch-by-pitch data for all baseball games beginning in 1989 are available from the website Retrosheet (http://www.retrosheet.org), a "baseball encyclopedia" that features box scores for every major league game from 1918 through the present, as well as detailed pitch-by-pitch descriptions of recent games.

Major League Baseball had 30 teams throughout most of the period of the study. Each team is scheduled to play 162 games over the course of a season. Each team's pitchers typically throw approximately 150 pitches per game, for approximately 700,000 pitches per season. Consequently, we use a massive database that should allow us to detect even small levels of racial bias. Over the period of the study, we have data for 13,868,362 total pitches, of which 7,434,903 (53.6% of the total) resulted in a called strike or ball.

For each pitch in the data set, we also identify a collection of potential predictor variables. We first noted the identities of the pitcher who threw the pitch, the hitter to whom the pitch was thrown, and the home plate umpire calling the pitch. The race or ethnicity of these individuals is the primary focus in this study.

Identifying the races of home plate umpires, pitchers, and hitters was an onerous undertaking. There are no complete and definitive listings. The task was simplest for umpires, due to their relatively small number. Major League Baseball employed a total of 224 individuals who officiated at least one game over the period of the study. Many of these individuals, however, umpired only a small number of games, generally as vacation replacements or emergency substitutes. Major League Baseball's website (http://www.mlb.com) was used as the primary source of information on current umpires. For umpires who worked during the study period but who are currently no longer active, we supplemented this resource with material from two hobbyist websites (Matt's Pen and Paper Heroes, 2008; Out of the Park Developments, 2013). These websites provide extensive information about past umpires, including photographs. We were unable to locate reliable race information for 58 umpires, most of whom worked for only 1 or 2 games and all of whom worked for 33 or fewer games. The number of pitches involved was negligible—a total of 26,643 or less than 1% of the total sample size of 7,434,903 pitches that resulted in called strikes or balls. We were left with 166 umpires, of whom 154 were classified as White, eight as Black, and four as Latin.

Racial identification for pitchers and hitters was considerably more involved, due both to the much larger number of individuals and to the lack of complete databases of player pictures and racial information. We began by writing a computer script that automated the task of extracting pictures of Major League Baseball players from the Internet. Two student research assistants then made independent assessments of player races based upon visual inspection of these photographs. Each research assistant later made a second independent assessment, based upon an independent extraction of Internet photographs, resulting in a total of four independent assessments of player race. There was complete agreement in the four assessments in over 90% of the cases and near-complete agreement (three out of four) in most of the remaining cases. We examined all cases of disagreement individually, using further searches of Internet pictures and biographies, supplemented by partial lists of minority players found on the Internet. While identifying racial and ethnic identity from photographs obtained over the Internet undoubtedly introduces the problem of errors-in-variables into this investigation, we believe that our careful curation and review process has largely limited the scope of the potential problem.

"Race" and "ethnicity" are not well-defined concepts in American (or perhaps broader) social consciousness. To make the task of analysis more manageable, we reduce the initial racial classifications to just three categories: White, Black, and Latin. Latin players are those players born in Latin America or the Caribbean islands. Racially black players born in those countries were classified as Latin for the current analysis. Black players include blacks born in the United States and Canada, while White players are Caucasians born in the United States, Canada, and Europe. Players belonging to various racial groups that are not extensively represented in our data set (such as biracial, Native American, Pacific Islander, etc.) were removed from the analysis. We also dropped from the analysis those players of

Table 1. Number of Pitchers and Hitters in Each Racial Group, 1989-2010.

	Pitchers	Hitters
White	1,854	1,474
Black	89	391
Latin	387	490
Total	2,330	2,355

Asian origin, partially because of their relatively small numbers and partially because of symmetry, as there are no Asian umpires in Major League Baseball. The total number of pitches represented by these deletions was a fraction of 1% of the total. Table 1 gives the resulting number of pitchers and players in each of the three racial categories.

For the models in this article, we experimented with different interpretations of the racial identifications of the umpires, pitchers, and hitters. One alternate approach considered a simple two-fold racial categorization—White and minority. We also considered a scenario in which Blacks born in Latin America were classified as Black rather than Latin, and another in which American-born players of Hispanic origin were grouped with the "Latin" players. In general, experimentation with different racial recodings did not yield materially different results in our models.

All of the models we developed include pitcher-specific and hitter-specific variables, to account for individual player effects. Some pitchers, for example, are known to be particularly precise or erratic in their pitching. Similarly, some hitters have an especially good "batter's eye" view of the strike zone. Likewise, we included umpire-specific variables, to account for individual umpire differences, as some are known to have a particularly generous or stingy strike zone.

In addition to racial codings for pitchers, hitters, and umpires, we identify a host of relevant auxiliary variables for each pitch. Some of these variables are used because they are known to explain some of the variance in the model. These include the inning of the game (fewer strikes tend to be called later in the game, as starting pitchers tire and less-effective relief pitchers are used) and whether the home or visiting team is at bat (there is a small but significant home team advantage).

It is plausible to believe that there may be an "exhaustion effect"—namely, that over the course of the game the pitcher (especially the starting pitcher) tires and becomes less precise in his control. Similar effects may obtain with the batter (as he becomes more accustomed to the pitcher's delivery) and umpire (as the mental task of calling pitches becomes more draining). Accordingly, we have included as predictor variables three within-game "counters," for the number of called pitches involving the pitcher, hitter, and umpire.

Better, or weaker, teams may have reputational effects that potentially explain a portion of the variation of the data, and which may even unconsciously affect an umpire's decision making. Accordingly, we also include measures of the overall

quality of the pitcher's and hitter's team, namely, each team's overall winning percentage for the season.

Other auxiliary variables are included because they may, in fact, subtly influence the call of the pitch. Parsons et al. (2011) have theorized that umpires may well behave differently when the pitch they are calling is subject to particularly high levels of "scrutiny," that is, when conditions in the game may mean that especially close attention is being paid to the outcome of a pitch. Accordingly, we collect data on several scrutiny variables. These variables include game attendance (the more people at the game, the more people watching the umpire's decision making) and the game score (when the game is a blowout, the umpire's call is less material to the game outcome). This last measure is operationalized as the absolute value of the difference between the two team scores.

Also included among these scrutiny variables are the ball and strike count, that is, the number of balls and strikes that the batter already has, at the time the pitch is made. Pitchers behave differently, for example, when the batter has no balls and two strikes (a 0-2 count) than when the count is three balls and no strikes (a 3-0 count). In the former case, the pitcher will typically throw a pitch on the borders of, or even just outside of, the strike zone, figuring that he has little to lose if the batter does not swing, but much to gain if he does. In the latter case, the circumstances are reversed, and the pitcher is highly inclined to throw a pitch that is very likely to be a strike.

The final scrutiny measure is the presence or absence of electronic devices to track the trajectory of a pitch. Major League Baseball began introducing systems designed by the QuesTec Corporation during the 2002 season. These systems included four cameras placed at various locations in the stadium, along with television and computer monitors that reported and recorded game events. The technology was controversial from the start. Its introduction prompted a formal grievance by the umpires' union to the National Labor Relations Board. Moreover, the QuesTec company experienced severe financial difficulties, including sanctions for securities-related improprieties. Presumably as a consequence of the litigatory minefield, it is actually quite difficult to find definitive information on exactly when the QuesTec systems were installed and in which ballparks they were used. The consensus appears to be that the technology was used in eleven stadiums (Anaheim, Arizona, Boston, Chicago White Sox, Cleveland, Houston, Milwaukee, New York Mets, New York Yankees, Oakland, and Tampa Bay) throughout the 2002-2006 seasons. Accordingly, during these years, we identify pitches thrown in these stadiums as "QuesTec" pitches, and those thrown in other stadiums as "non-QuesTec" pitches.

By the start of the 2007 Major League Baseball season, a new labor agreement with the umpires' union defused many tensions regarding the use of umpire-monitoring technologies. Moreover, an improved camera-and-computer system was introduced during that year. Data from this new "PITCHf/x" system are posted on the Internet by Major League Baseball. The system has operated in all Major League Baseball parks since the 2007 season.

Analysis

We first investigate the extent to which the race of the umpire (or pitcher, or hitter) affects the probability that a pitch is called a strike rather than a ball, by the home plate umpire, during the 1989-2010 Major League Baseball seasons.

Panel A in Table 2 displays the proportion of called pitches that were deemed strikes by the home plate umpire, given various umpire/pitcher racial combinations. Panel B in the same table performs the same function for various umpire/hitter combinations. Thus, for example, during the entire period of the study, White umpires made calls on 4,983,631 pitches made by White pitchers, and ruled that 30.6% of these pitches were strikes.

Superficially, it appears that there are some grounds for suspicion of a small degree of racial discrimination in pitch calling. Calling a pitch a strike favors the pitcher, and we note that White umpires call strikes at a higher rate for White pitchers (30.6%) than for Black pitchers (30.4%). Black umpires call fewer strikes against Black hitters (29.4%) than against White or Latin hitters (31.0% and 30.9%, respectively). The differences are small but statistically significant, given the huge sample sizes.

Yet even at this basic level of comparison, there are mixed messages regarding racial discrimination. While White umpires call more strikes for White pitchers than Black pitchers, as previously noted, they call even more strikes in favor of Latin pitchers. White umpires actually call more strikes against White hitters (31.0%) than against Blacks (29.4%) or Latins (30.8%). Any rush to conclude that racial discrimination exists in pitch calling must be tempered by these, and other, considerations.

Multiple regression techniques allow us to gain a more complete picture of the underlying pitch-calling dynamics. We used a linear probability model in the present analysis. We also investigated logit and probit alternatives. There are no substantial differences between the results from the various approaches. The linear probability model is computationally more tractable and generates results that are easier to interpret. Moreover, the logistic model is nearly linear for the probabilities under consideration in this article to the data.

Our linear probability model is of the form

$$
\begin{aligned}
Y = \beta_0 + \boldsymbol{\beta}_1 \mathbf{X}_1 + \boldsymbol{\beta}_2 \mathbf{X}_2 + \boldsymbol{\beta}_3 \mathbf{X}_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \\
+ \boldsymbol{\beta}_7 \mathbf{X}_7 + \boldsymbol{\beta}_8 \mathbf{X}_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} \quad (1) \\
+ \boldsymbol{\beta}_{1,2} (\mathbf{X}_1 \times \mathbf{X}_2) + \boldsymbol{\beta}_{1,3} (\mathbf{X}_1 \times \mathbf{X}_3) + \boldsymbol{\gamma} \boldsymbol{Z} + \varepsilon,
\end{aligned}
$$

where

- Y is a dichotomous response variable, coded 1 if the pitch is called a strike and 0 if it is called a ball;
- $X_1, X_2,$ and X_3, respectively, are sets of indicator variables for the races of the umpires, pitchers, and hitters (URACE, PRACE, and HRACE). Every

Table 2. Pitch Counts and Proportions for Umpire/Pitcher Racial Combinations, 1989-2010.

Panel A. Umpire and pitcher racial combinations

Umpires	Outcome	Pitchers			
		White	Black	Latin	Total
White	Proportion called strikes	0.306	0.304	0.308	0.306
	Number of pitches	4,983,631	423,514	1,124,237	6,531,382
Black	Proportion called strikes	0.307	0.304	0.306	0.307
	Number of pitches	280,810	25,276	67,584	373,670
Latin	Proportion called strikes	0.317	0.302	0.313	0.315
	Number of pitches	96,438	8,639	28,113	133,190
Total	Proportion called strikes	0.306	0.304	0.308	0.306
	Number of pitches	5,360,879	457,429	1,219,934	7,038,242

Panel B. Umpire and hitter racial combinations

Umpires	Outcome	Hitters			
		White	Black	Latin	Total
White	Proportion called strikes	0.31	0.294	0.308	0.306
	Number of pitches	3,873,743	1,285,051	1,372,588	6,531,382
Black	Proportion called strikes	0.31	0.294	0.309	0.307
	Number of pitches	224,746	69,793	79,131	373,670
Latin	Proportion called strikes	0.318	0.303	0.317	0.315
	Number of pitches	81,427	21,938	29,825	133,190
Total	Proportion called strikes	0.31	0.295	0.308	0.306
	Number of pitches	4,179,916	1,376,782	1,481,544	7,038,242

umpire, pitcher, and hitter has been designated as White, Black, or Latin. The White racial category is used as the baseline.

- X_4 is 1 if the batter is from the home team and 0 if he is from the visiting team (HOMEVIS);
- X_5 is the attendance at the game, in thousands (ATTEND);
- X_6 is the game score differential, or the amount by which either team is winning (ABSSCODIF);
- X_7 is a set of indicator variables representing the inning of the game (second, third, fourth, etc.) with all innings beyond the ninth aggregated into a single "extra innings" variable and the first inning serving as the baseline case (INNING);
- X_8 is a set of indicator variables for the various possible ball-and-strike counts on the batter when the pitch is thrown. There are 12 possible pitch counts, and therefore 11 indicator variables. The 0-0 count used as the baseline case (PITCHCOUNT);
- X_9 is an indicator variable that equals 1 when there is an electronic pitch monitoring system (either QuesTec or PITCHf/x) in place in the park and 0 if no such system in place (MONITOR);
- X_{10} and X_{11} represent the overall winning percentage (for the season) of the pitcher's and hitter's teams, respectively (PWINPCT and HWINPCT);
- X_{12}, X_{13}, and X_{14} give the cumulative number of called pitches within the game involving the umpire, pitcher, and hitter, respectively (UCTR, PCTR, and HCTR);
- $X_1 \times X_2$ is a set of interaction terms between umpire race and pitcher race, which will account for any tendency that umpires have to treat hitters differently because of their race;
- $X_1 \times X_3$ is similarly a set of interaction terms between umpire race and hitter race, which will account for any tendency that umpires have to treat hitters differently because of their race;
- Z is a set of indicator variables, one for each umpire, pitcher, and hitter in the data set, to account for fixed effects of individual propensities for a called strike; and
- ε is a well-behaved error term.

Parameter estimates for the models in this article were obtained in IBM SPSS Statistics (versions 19 and 20) using its univariate general linear modeling functionality. Table 3 gives a summary of parameter estimates and other relevant information such as standard errors, t test statistics, and p values, connected with the main effects of the variables for the basic model, as described by Equation 1. The parameter estimates for γ, representing individual umpire, pitcher, and hitter fixed effects, are omitted. Interaction terms will be displayed and discussed separately. The overall coefficient of determination for the model, R^2, is .099. Our model explains relatively little variation in the data. The large sample sizes we employ will

Table 3. Covariates in the Basic Model (1989-2010).

Variable	Variable Name	Slope	SE	t	p Value
	Intercept	0.415	0.046	9.04	≈ 0***
X_1	URACE				
	Latin	0.007	0.003	2.63	.008***
	Black	−0.002	0.003	−0.59	.557
X_2	PRACE				
	Latin	0.106	0.069	1.54	.123
	Black	−0.151	0.061	−2.47	.014**
X_3	HRACE				
	Latin	−0.068	0.051	−1.32	.186
	Black	−0.219	0.076	−2.89	.004***
X_4	HOMEVIS	−0.008	≈ 0	−22.33	≈ 0***
X_5	ATTEND	0.00003	≈ 0	1.8	.071*
X_6	ABSSCODIF	0.001	≈ 0	13.64	≈ 0***
X_7	INNING				
	Second	−0.004	0.001	−4.82	≈ 0***
	Third	−0.014	0.001	−17.31	≈ 0***
	Fourth	−0.018	0.001	−19.49	≈ 0***
	Fifth	−0.016	0.001	−15.26	≈ 0***
	Sixth	−0.015	0.001	−13.37	≈ 0***
	Seventh	−0.01	0.001	−7.84	≈ 0***
	Eighth	−0.005	0.001	−3.91	≈ 0***
	Ninth	0.004	0.002	5.02	≈ 0***
	Additional	−0.017	0.002	−8.06	≈ 0***
X_8	PITCHCOUNT				
	0-1	−0.198	0.001	−355.17	≈ 0***
	0-2	−0.319	0.001	−405.45	≈ 0***
	1-0	−0.01	0.001	−17.55	≈ 0***
	1-1	−0.161	0.001	−255.23	≈ 0***
	1-2	−0.29	0.001	−405.23	≈ 0***
	2-0	0.055	0.001	63.4	≈ 0***
	2-1	−0.126	0.001	−143.09	≈ 0***
	2-2	−0.258	0.001	−307.96	≈ 0***
	3-0	0.212	0.001	177.27	≈ 0***
	3-1	−0.042	0.001	−33.73	≈ 0***
	3-2	−0.229	0.001	−193	≈ 0***
X_9	MONITOR	0.004	0.001	6.18	≈ 0***
X_{10}	PWINPCT	0.041	0.003	13.55	≈ 0***
X_{11}	HWINPCT	−0.006	0.003	−1.95	.051*
X_{12}	UCTR	0.000006	≈ 0	0.61	.544
X_{13}	PCTR	−0.0004	≈ 0	−23.21	≈ 0***
X_{14}	HCTR	−0.001	≈ 0	−11.83	≈ 0***

Note. This table features the main effects of variables in Equation 1. These same parameters are included in all of the models in this article, and their parameter estimates do not vary widely. In this table, we suppress the display of approximately 6,000 parameters corresponding to every umpire, pitcher, and hitter, that sweep out individual-level propensities to be associated with called strikes.
*p < .10, **p <.05, ***p < .01.

allow us to detect even very small degrees of racial bias in pitch calling. However, as the coefficient of determination underscores, any such effects will be small in magnitude. Coefficients of determination for all models presented in this article are similarly small and are not presented.

In general, race-related main effects are of mixed significance. Latin umpires appear to call slightly more strikes, while Latin and Black hitters appear to have fewer strikes called against them. This finding supports the notion that race-based differences exist in both umpire and player behavior. However, taken by itself, the result is not indicative of racial bias, as these main effects occur regardless of the race of other persons involved with the pitch. Moreover, the effect is small. The observed marginal 0.007 called strike probability for Latin umpires amounts to approximately 1.1 additional strikes per game. (Since each team's pitchers throw approximately 150 pitches per game, and historically 53.6% of pitches thrown during the study's time period have resulted in a called strike or ball, we have $0.007 \times (2 \times 150) \times 0.536 = 1.1256$ additional expected strikes per game.)

The covariates are generally statistically significant, and the signs on the coefficients are sensible. The home team, for example, is slightly less likely to have a strike called on the hitter. Higher game attendance means more strikes are called (plausibly because better teams have higher attendance). After the first inning, the pitch tends to be less likely to be called a strike, although there is no clearly identifiable trend in the inning coefficients. As the score difference increases, the pitch is more likely to be called a strike. Pitchers from better teams tend to get more strikes called for them, while hitters from better teams tend to get fewer strikes called against them. As the pitcher's or hitter's tenure in the game progresses, there is a tiny but statistically significant decrease in the chance of a called strike. (No comparable effect occurs for the umpire.) The small magnitude of these terms, however, indicates that they are generally not of practical interest.

Of special interest is the result indicating that the presence of a pitch monitoring system results in a slight, but statistically significant, increase in the probability that a pitch is called a strike. These systems were controversial when they were introduced, with many pitchers claiming that the strike zone was materially reduced. (Pitcher Curt Schilling, then with the Arizona Diamondbacks, even went so far as to damage a QuesTec camera with a baseball bat, in a fit of anger.) In fact, it appears that the presence of these cameras slightly increases the proportion of called strikes, although the effect is quite small. For example, a probability of 0.004, over approximately 300 total pitches in a game, with approximately half of these pitches resulting in a called strike or ball, leads to an additional 0.6 called strikes per game.

The one class of covariates with large magnitude effects is the ball-and-strike count. This result is not surprising. As we already noted, a pitcher's behavior will be substantially different when the count is 0-2 (and throwing a ball has small downside) than when it is 3-0 (and throwing a fourth ball walks the batter). As expected, the more balls on the batter, the higher the probability that the next pitch will be called a strike, and the more strikes on the batter, the greater the probability that the next pitch will be called a ball.

Table 4. Interactions in the Basic Model (1989-2010).

Variable	Variable Name	Slope	SE	t	p Value
$\mathbf{X}_1 \times \mathbf{X}_2$	URACE $\times$ PRACE				.083
	Latin $\times$ Latin	0.002	0.003	0.67	.504
	Latin $\times$ Black	−0.014	0.007	−1.93	.054
	Black $\times$ Latin	−0.004	0.002	−1.83	.068
	Black $\times$ Black	−0.003	0.004	−0.84	.401
$\mathbf{X}_1 \times \mathbf{X}_3$	URACE $\times$ HRACE				.574
	Latin $\times$ Latin	−0.002	0.003	−0.81	.42
	Latin $\times$ Black	−0.005	0.003	−1.54	.123
	Black $\times$ Latin	0.001	0.002	0.58	.564
	Black $\times$ Black	≈ 0	0.002	0.1	.992

Note. This table features detailed information about the two-way interaction terms between umpire race and pitcher race and, additionally, umpire race and hitter race as described in Equation 1. Note that both interaction terms are not globally significant; we include individual parameter estimates, standard errors, test statistics, and p values for illustrative purposes only. In general, the URACE $\times$ PRACE and URACE $\times$ HRACE interactions are not globally significant (or are only marginally significant) for the models developed in this article.

Table 4 displays the interaction terms from the model. It is noteworthy that these interaction terms—between umpire and both pitcher and hitter race—are not significant. Given the huge sample size involved in this study (approximately 7 million pitches), and, consequently, the model's ability to detect even small effects, racial discrimination in the calling of balls and strikes appears vanishingly limited.

Parsons et al. (2011) have noted that racial bias may well exist in general, but its presence might be masked by failure to consider whether the umpire's decision is subject to particularly intense observation. Following their reasoning, we augment our analysis by adding terms to the model that assess the interaction between racial interaction and various measures of this phenomenon.

Four of the variables already in the model can be used to proxy this level of scrutiny. These variables include game attendance (since the greater the attendance, the more people there are at the game to observe umpire behavior) and absolute score differential (since people may well pay closer attention to umpire decisions in closer games).

A third, and more explicit, such measure of scrutiny is the presence or absence of an electronic pitch monitoring system at the park. As was previously noted, "QuesTec" monitoring systems were used in some stadiums during the 2002-2006 seasons, and a "PITCHf/x" system has been used in all stadiums beginning with the 2007 season. Since umpires are formally evaluated in part based upon the consistency of their strike calling—as metered by the electronic monitoring systems—there is a clear incentive to alter behavior in the presence of such monitoring.

The fourth scrutiny variable is the pitch count. We have already noted that pitchers behave differently, for example, on a 3-0 count than on a 0-2 count. It is also plausible to believe that audiences, and perhaps even umpires, might pay differing

amounts of attention on more crucial pitch counts. In order to ease the interpretation of our models, for purposes of considering interactions we will reduce the 12 possible pitch count scenarios to simply two binary categorical variables. A "ball critical" variable, BCRIT (X_{15}), is used to indicate when the batter already has three balls. A "strike critical" variable, SCRIT (X_{16}), is employed when the batter already has two strikes. In the case of a 3-2 count, both the BCRIT and SCRIT variables will be equal to 1.

It is conceivable (and such an assumption simplifies the interaction terms in the model greatly) that the subjective effects of unconscious racial discrimination are uniform. That is, what matters is not the specific race of the umpire and pitcher or hitter but whether they are of the same race. Accordingly, we create the following two "matching" variables:

- X_{17} is an indicator variable that is 1 when the umpire and pitcher are of the same race, and 0 otherwise (UPSR); and
- X_{18} is an indicator variable that is 1 when the umpire and hitter are of the same race, and 0 otherwise (UHSR).

These variables will measure the extent to which racial bias plays a role in subjective ball/strike decision making. Interaction between these variables and the scrutiny variables will show whether this form of bias is mitigated when the pitch call in question receives special attention. We modify our model (1) by adding terms for the racial-matching variables (UPSR and UHSR), as well as pairwise interactions between these two matching variables and our five scrutiny measures (ATTEND, ABSSCODIFF, MONITOR, BCRIT, and SCRIT).

As with our first analysis, we find that the covariates are statistically significant, but generally so small in magnitude (save for the pitch count variables) as to be practically meaningless. In fact, there is negligible difference in the estimates of these covariates, when comparing the current model with the model in Equation 1. In the interest of simplicity and for clarity of presentation, we will omit the display of these parameter estimates both here and in future discussion of models. Full output tables are, of course, available from the authors upon request.

The interaction terms, given in Table 5, are the terms on which to concentrate when investigating the extent to which discriminatory practices exist or are moderated in the presence of heightened scrutiny. The model does not indicate any meaningful degree of umpire-on-pitcher discrimination. Neither the main umpire/pitcher "match" variable nor any of the interactions between this variable and the various scrutiny variables is statistically significant at the $\alpha = .05$ level of significance.

However, some of the umpire-on-hitter effects are significant. When the umpire and hitter are of the same race, there is a slightly (0.004) reduced likelihood of calling the pitch a strike. Since called strikes favor the pitcher, this finding implies an advantage to the batter when the umpire is of the same race. But the effect is small. A percentage of 0.004, multiplied over the roughly 150 called strikes or balls thrown in a game, indicates that only approximately 0.6 pitches per game are affected.

**Table 5.** Interactions in the Matching Model (1989-2010).

Variable	Variable Name	Slope	SE	t	p Value
X_{17}	UPSR	−0.002	0.002	−0.74	.458
X_{18}	UHSR	−0.004	0.002	−1.91	.056*
$X_{17} \times X_5$	UPSR × ATTEND	−0.0002	≈ 0	−0.62	.533
$X_{17} \times X_6$	UPSR × ABSSCODIF	≈ 0	≈ 0	−0.43	.669
$X_{17} \times X_9$	UPSR × MONITOR	0.002	0.001	1.92	.055*
$X_{17} \times X_{15}$	UPSR × BCRIT	−0.003	0.002	−1.68	.093*
$X_{17} \times X_{16}$	UPSR × SCRIT	≈ 0	0.001	−0.44	.659
$X_{18} \times X_5$	UHSR × ATTEND	≈ 0	≈ 0	0.08	.938
$X_{18} \times X_6$	UHSR × ABSSCODIF	≈ 0	≈ 0	0.52	.602
$X_{18} \times X_9$	UHSR × MONITOR	−0.002	0.001	−1.76	.078*
$X_{18} \times X_{15}$	UHSR × BCRIT	−0.005	0.001	−3.43	.001***
$X_{18} \times X_{16}$	UHSR × SCRIT	0.003	0.001	3.27	.001***

Note. This table includes the interaction terms in the "matching model" found in Equation 2 for the entire time period (1989-2010). Note that with the exception of the umpire/hitter racial match indicator variable and the ball/strike critical indicator variables, the interactions are generally not significant or are only marginally significant.

This effect plays out even further when the pitch is "critical." When the batter is facing a critical ball count (i.e., with three balls), having an umpire of the same race as the batter benefits the hitter by reducing the probability that the pitch is called a strike by an additional 0.005. Again, the probability is small, but the influence is more significant in game terms, since a called ball results in the batter reaching first base on a walk. However, when the batter is facing a critical strike count (i.e., with two strikes), having an umpire of the same race as the batter harms the hitter by giving an increased (0.003) probability of the pitch being called a strike. Again, the probability is small, but the game effect is larger, because an additional called strike means that the batter will be called out on strikes.

Figure 1 helps to visualize the dynamics at work in the complex interactions here. In ball critical situations, there is a small but real tendency for strikes to be called less often when the umpire is of the same race as the hitter (a "match") and more often when the umpire is of a different race than the hitter ("no match"). The effect is reversed in situations that are not ball critical. A different dynamic is at work regarding strike critical situations. While there is a consistently higher tendency for umpires to call strikes when they are of the same race as the hitter, the effect is more pronounced in strike critical situations.

The significance of several interaction terms leads us to suspect that some level of racial discrimination may take place in the calling of balls and strikes. However, we should not rush too quickly to this conclusion. First of all, the results of Parsons et al. (2011), based upon data from the 2004-2008 seasons, found potential discrimination against pitchers but generally no significant results regarding hitters. Second, the
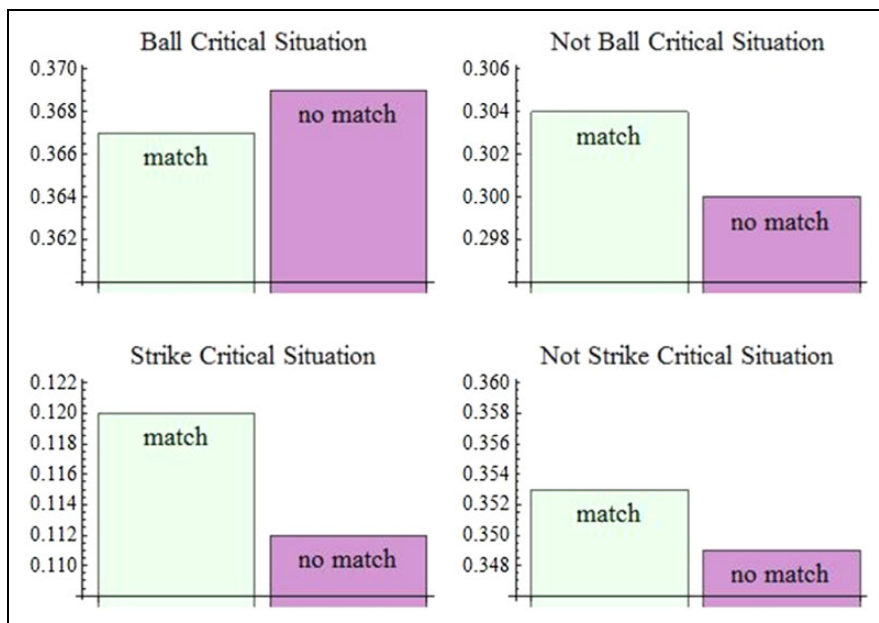
Figure 1. Proportion of time a strike is called, in ball- and strike critical situations.
Note. The charts cover the period from 1989 to 2011. For example, when there are three balls in the count and the races of the umpire and hitter match, strikes are called 36.7% of the time. In contrast, when the races of umpire and hitter are not the same, strikes are called in ball-critical situations 36.9% of the time. Note that the origins of the four graphs are different, reflecting the large differences in overall propensity for the pitcher to throw a strike in various conditions. However, the scalings within each pair of graphs are the same, enabling direct comparison of the magnitude of the interaction effects.

mixed pattern of findings for hitters in "critical pitch" situations indicates that something more subtle than a pure, albeit small, discriminatory effect takes place.

Parsing racial interactions out fully, rather than simply with a binary "match/no match" classification, may give more complete information about the nature of any racial discrimination. Accordingly, we modify our second model by dropping the two "matching" variables X_{17} and X_{18}, and allowing for full expression of the interactions. That is, we create three-way interactions involving umpire race and pitcher race, as well as umpire race and hitter race, with each of the following five variables: ATTEND, ABSSCODIFF, MONITOR, BCRIT, and SCRIT. Furthermore, because the coefficients connected with the three-way interaction terms are not interpretable unless all subsidiary two-way interactions are included in the model, we include all relevant two-way subsidiary interactions. For example, because we add the three-way interaction term URACE × PRACE × ATTEND, we also add URACE × PRACE, URACE × ATTEND, and PRACE × ATTEND.

Table 6. Three-Way Interactions between Umpire Race, Pitcher Race, and Scrutiny Variables.

Variable	Variable Name	p Value
$X_1 \times X_2$	URACE × PRACE	.437
$X_1 \times X_3$	URACE × HRACE	.488
$X_1 \times X_2 \times X_5$	URACE × PRACE × ATTEND	.683
$X_1 \times X_2 \times X_6$	URACE × PRACE × ABSSCODIF	.241
$X_1 \times X_2 \times X_9$	URACE × PRACE × MONITOR	.524
$X_1 \times X_2 \times X_{15}$	URACE × PRACE × BCRIT	.919
$X_1 \times X_2 \times X_{16}$	URACE × PRACE × SCRIT	.764
$X_1 \times X_3 \times X_5$	URACE × HRACE × ATTEND	.587
$X_1 \times X_3 \times X_6$	URACE × HRACE × ABSSCODIF	.695
$X_1 \times X_3 \times X_9$	URACE × HRACE × MONITOR	.325
$X_1 \times X_3 \times X_{15}$	URACE × HRACE × BCRIT	.651
$X_1 \times X_3 \times X_{16}$	URACE × HRACE × SCRIT	.818

Note. The three-way interactions model includes covariates from the previous model and all subsidiary two-way interaction terms (to make the three-way interaction terms interpretable). Note that none of the three-way interaction terms are globally significant, even in the presence of approximately 7,000,000 called pitches.

Since parameter estimates for main effects in this model are virtually identical to those from the previous two models, as noted previously, we are suppressing the presentation of these results for sake of clarity and space. We do present the results for the interaction variables, as they are the ones that describe any potential racial discrimination, and because they differ from model to model. Table 6 gives these results for the model with the three-way interaction terms.

First of all, note that none of the three-way interactions are statistically significant. This situation is noteworthy, as it indicates that whatever possible interaction there might be between (say) umpire race and various measures of scrutiny, these effects are essentially the same across pitcher race. Following standard statistical practice, we therefore rerun the model without the insignificant higher-order (three-way) interactions in order to focus just on the two-way interactions.

Parameter estimates for main effects are once again inconsequentially different from those estimates associated with previous models, and so, as noted before, these results are omitted for the sake of brevity. Table 7 gives the results for the interaction terms.

It is interesting to note that most of the two-way interaction terms are statistically significant in the absence of the three-way interaction terms, including several that were not significant in the previous computation (not presented) when higher-order terms were included. This finding suggests that including the three-way interaction terms is obscuring some of the two-way interactions that are present.

While umpire race (but not pitcher race or hitter race) significantly interacts with both the game attendance and the score differential variables, the small magnitude of

Table 7. Relevant Two-Way Interaction Terms From the Reduced Full Model (1989-2010).

Variable	Variable Name	Slope	SE	t	p Value
$X_1 \times X_2$	URACE × PRACE				0.050**
$X_1 \times X_3$	URACE × HRACE				0.419
$X_1 \times X_5$	URACE × ATTEND				0.046**
	Latin	0.00019	≈ 0	1.77	0.077*
	Black	0.00011	≈ 0	1.8	0.072*
$X_1 \times X_6$	URACE × ABSSCODIF				0.022**
	Latin	−0.001	0.001	−2.24	0.025**
	Black	0.001	≈ 0	1.53	0.125
$X_1 \times X_9$	URACE × MONITOR				≈ 0***
	Latin	−0.01	0.003	−3.98	≈ 0***
	Black	0.002	0.002	0.91	0.363
$X_1 \times X_{15}$	URACE × BCRIT				≈ 0***
	Latin	−0.01	0.005	−1.99	0.046**
	Black	0.011	0.003	3.53	≈ 0***
$X_1 \times X_{16}$	URACE × SCRIT				≈ 0***
	Latin	−0.014	0.003	−4.59	≈ 0***
	Black	≈ 0	0.002	0.26	0.799
$X_2 \times X_5$	PRACE × ATTEND				0.691
$X_2 \times X_6$	PRACE × ABSSCODIF				0.607
$X_2 \times X_9$	PRACE × MONITOR				0.001***
	Latin	−0.004	0.001	−3.13	0.002***
	Black	0.006	0.003	1.81	0.072*
$X_2 \times X_{15}$	PRACE × BCRIT				≈ 0***
	Latin	0.005	0.002	2.91	0.004***
	Black	−0.012	0.003	−3.47	0.001***
$X_2 \times X_{16}$	PRACE × SCRIT				≈ 0***
	Latin	−0.003	0.001	−2.23	0.026**
	Black	0.018	0.002	7.99	≈ 0***
$X_3 \times X_5$	HRACE × ATTEND				0.191
$X_3 \times X_6$	HRACE × ABSSCODIF				0.51
$X_3 \times X_9$	HRACE × MONITOR				0.084*
$X_3 \times X_{15}$	HRACE × BCRIT				0.004***
	Latin	0.006	0.002	3.14	0.002***
	Black	0.003	0.002	1.89	0.059*
$X_3 \times X_{16}$	HRACE × SCRIT				≈ 0***
	Latin	−0.01	0.001	−9.61	≈ 0***
	Black	0.004	0.001	3.83	≈ 0***

Note. The two-way interactions linked to the linear probability model in Table 6, with three-way interaction terms removed from that. Note the extensive collection of significant terms, which may suggest that some of the racially connected behaviors detected in some sabermetric studies are functions of (possibly spuriously) race-connected behaviors of, for example, pitchers in various situations that occur independently of the race of the umpire.

the coefficients indicates little practical effect of these results. On the other hand, the interactions involving the monitoring variable and the critical pitch variables are more interesting. The presence of pitch-monitoring technology does differentially affect the probability that the umpire calls a pitch a strike. For example, Latin umpires are significantly less likely to call strikes in the presence of monitoring. But, since the three-way interaction was not significant, this effect is roughly constant across all races of pitchers and of hitters. Thus, while umpires of different races may be differentially altering pitch-calling behaviors in the presence of monitoring technology, this behavior does not depend upon the race of the pitcher or the hitter. There is a racial effect, but not racially discriminatory treatment on the part of umpires.

We likewise see a racial effect for umpires when the ball or strike count is critical. Latin umpires are less likely to call a strike, and Black umpires more likely to call a strike, on ball critical pitches. In this situation, a called ball benefits the hitter, as he reaches base on a walk. Hence, in ball critical situations, there is a small but significant global tendency for Latin umpires to favor hitters and for Black umpires to favor pitchers. In strike critical situations, Latin umpires have a reduced tendency to call a strike (again, benefiting the hitter). However, as the three-way interactions with pitcher race and with hitter race were not significant, these effects are global and not dependent upon the race of the hitter or pitcher.

We find analogous effects for pitcher race. In the presence of monitoring technology, Latin pitchers are significantly less likely, while Black pitchers are significantly more likely, to have a pitch called a strike. If anything, this might imply a reverse discrimination effect against Latin pitchers—in the absence of monitoring scrutiny, they may occasionally get the benefit of the doubt. In ball critical situations, Latin pitchers benefit (more strikes called) while Black pitchers are hurt (less strikes called), while the pattern reverses in strike critical situations. Yet, umpire race is not a factor in this phenomenon: The lack of three-way interaction implies a consistency in these judgments across umpire race.

For hitters, minority players (both Latin and black) have a slightly increased chance in a ball critical situation of having a pitch called a strike. In strike critical situations, Latin hitters are less likely to have a strike called while Black hitters are more likely to have a strike called. The net effect will result in marginally fewer walks for minority players (since they draw more strikes in ball critical situations) and marginally more strikeouts for Black players (since they get more strikes in strike critical situations). However, once again, umpire race does not factor into these observations since they are based on the two-way interaction terms.

Conclusion

Racial discrimination has been part of American political and economic history, and contemporary studies of racial bias continue to demonstrate its lingering effects. So,

it is natural to believe that racially motivated behaviors may well be exhibited in many different realms of human activity. Modern baseball is one of those realms.

Thus, we might expect to find some level of racial bias in the calling of balls and strikes. Even though the umpires in question are highly trained professionals, the very nature of the split-second subjective decisions they make means that their actions are prone to influence by unconscious biases, including those motivated by race.

What we find, in fact, is a smaller and nuanced picture of the effect of race has on called balls and strikes. Our models did find statistically significant effects. Racial bias does play a role in pitch calling. However, the evidence for this role is both sparse and small. It is sparse insofar as it manifests itself only in selected situations in certain models. We found racially biased effects, for example, in critical pitch situations when the umpire's decision involved a hitter of a different race, but not when the decision involved a pitcher of a different race. Further, it is small in that the effects, while statistically significant, are generally of limited magnitude, often influencing one pitch or less per game on average.

Moreover, modeling choices affect whether model terms indicating discrimination are statistically significant. In examining the entire 1989-2010 time period, for example, a "matching" model that considered simply whether umpire and hitter were of the same race finds a discriminatory effect in critical-pitch situations, whereas a "full interaction" model that parsed out the various racial combinations, did not. This sensitivity of results to model choice suggests a phenomenon that is more complex than a straightforward discriminatory effect.

We do not mean to suggest that racial dynamics are not at work in the calling of balls and strikes. It is noteworthy that many main effects in the models are statistically significant. As can be seen from Table 2, overall Latin umpires are slightly more aggressive about calling strikes than White umpires (31.5% vs. 30.6%). Analysis of two-way interactions (see Table 7) shows, for example, that umpires of different races react differently to the presence of pitch-monitoring technologies.

There are presumably reasons for these observed racial differences. However, we are not in a position to speculate, a posteriori, as to what these reasons might be. The general paucity of statistical significance for interaction terms between umpire and pitcher/hitter race, however, does indicate that while there may be racial differences in behavior, there is considerably less evidence for racially discriminatory actions on the part of umpires.

Moreover, these findings come from analysis of a truly massive database, involving 22 full seasons of data and more than 7,000,000 called pitches. The sample size is sufficiently large that any meaningful discriminatory effect should show up abundantly clearly in the analysis. The absence of such findings would indicate that even though race and racism are still real phenomena in American society, there is, in general, not a problem of racial discrimination by Major League Baseball umpires, at least with regard to the calling of balls and strikes.

Acknowledgment

Declaration of Conflicting Interests

Funding

References

Bodvarsson, O. B., & Pettman, S. P. (2002). Racial wage discrimination in Major League Baseball: Do free agency and league size matter? *Applied Economics Letters*, *9*, 791-796.

Breckenridge, R. S., & Goldsmith, P. R. (2009). Spectacle, distance, and threat: Attendance and integration of Major League Baseball, 1930-1961. *Sociology of Sport Journal*, *26*, 296-319.

Christiano, K. J. (1986). Salary discrimination in Major League Baseball: The effect of race. *Sociology of Sport Journal*, *3*, 144-153.

Christiano, K. J. (1988). Salaries and race in professional baseball: Discrimination 10 years later. *Sociology of Sport Journal*, *5*, 136-149.

Depken, C. A., & Ford, J. M. (2006). Customer-based discrimination against Major League Baseball players: Additional evidence from all-star ballots. *Journal of Socio-Economics*, *35*, 1061-1077.

Groothuis, P. A., & Hill, J. R. (2008). Exit discrimination in Major League Baseball: 1990-2004. *Southern Economic Journal*, *75*, 574-590.

Hanssen, F. A., & Andersen, T. (1999). Has discrimination lessened over time? A test using baseball's all-star vote. *Economic Inquiry*, *37*, 326-352.

Jewell, R. T. (2003). Voting for the baseball hall of fame: The effect of race on election date. *Industrial Relations*, *42*, 87-99.

Jewell, R. T., Brown, R. W., & Miles, S. E. (2002). Measuring discrimination in major league baseball: Evidence from the baseball hall of fame. *Applied Economics*, *34*, 167-177.

Kahn, L. M. (1991). Discrimination in professional sports: A survey of the literature. *Industrial and Labor Relations Review*, *44*, 395-418.

Margolis, B., & Piliavin, J. A. (1999). "Stacking" in Major League Baseball: A multivariate analysis. *Sociology of Sport Journal*, *63*, 16-34.

Matt's Pen and Paper Heroes. (2008). *MLB Umpires TTM Success*. Retrieved from http://www.mattspenandpaperheroes.com/Umpires.html

Out of the Park Developments. (2013). *Major League Baseball Umpires*. Retrieved from http://www.ootpdevelopments.com/board/ootp-mods-rosters-photos-quick-starts/206835-major-league-baseball-umpires.html

Palmer, M. C., & King, R. H. (2006). Has salary discrimination really disappeared from major league baseball? *Eastern Economic Journal*, *32*, 285-297.

Parsons, C. A., Sulaeman, J., Yates, M. C., & Hamermesh, D. S. (2011). Strike three: Discrimination, incentives, and evaluation. *The American Economic Review*, *101*, 1410-1435.

Price, J., & Wolfers, J. (2010). Racial discrimination among NBA Referees. *Quarterly Journal of Economics*, *125*, 1859-1877.

Purdy, D. A., Leonard, W. M., & Eitzen, D. S. (1994). A reexamination of salary discrimination in major league baseball by race/ethnicity. *Sociology of Sport Journal*, *11*, 60-69.

Tainsky, S., & Winfree, J. (2010). Discrimination and demand: The effect of international players on attendance in major league baseball. *Social Science Quarterly*, *91*, 117-128.

Author Biographies

Jeff Hamrick is Associate Professor of Finance and Analytics at the University of San Francisco. He is interested in empirical finance, stable distributions, financial contagion, sabermetrics, and nonparametric estimation techniques for stochastic processes.

John Rasp is Associate Professor of Statistics at Stetson University. His research interests are eclectic, and he has published papers on statistical analysis in a variety of fields, including accounting, agriculture, literary analysis, and sabermetrics.