STRIKE THREE:
UMPIRES' DEMAND FOR DISCRIMINATION

Christopher A. Parsons, Johan Sulaeman, Michael C. Yates and Daniel S. Hamermesh*

This Version: December 3, 2007

*Assistant professor of finance, McGill University; graduate student, University of Texas at Austin; assistant professor of finance, Auburn University; Edward Everett Hale Centennial professor of economics, University of Texas at Austin and research associate IZA and NBER. We thank Jason Abrevaya, Jeff Borland, J-C Bradbury, Jay Hartzell, Larry Kahn, William Mayew, Gerald Oettinger, Allen Sanderson, Alan Schwarz, John Siegfried, Paul Tetlock, Sheridan Titman, Deborah White and Justin Wolfers, and participants at seminars at several universities, for helpful discussions.

Abstract

We explore umpires' racial/ethnic preferences in the evaluation of Major League Baseball pitchers. Controlling for umpire, pitcher, batter and catcher fixed effects and many other factors, strikes are more likely to be called if the umpire and pitcher match race/ethnicity. This effect only exists where there is little scrutiny of umpires' behavior—in ballparks without computerized systems monitoring umpires' calls, at poorly attended games, and when the called pitch cannot determine the outcome of the at-bat. If a pitcher shares the home-plate umpire's race/ethnicity, he gives up fewer runs per game and improves his team's chance of winning. The results suggest that standard measures of salary discrimination that adjust for measured productivity may generally be flawed. We derive the magnitude of the bias generally and apply it to several examples.

I. Introduction

Discrimination in the labor market can take many forms, including disparities in wages, promotion, hiring, or performance evaluation, for reasons unassociated with underlying productivity. The last of these is particularly troublesome to economists because of its role as a benchmark: If workers are discriminated against when their performance is evaluated, then the ability to detect discrimination in other areas may be reduced. For example, the observed ratio of wages to measured skills may be identical across racial groups, but this clearly does not insulate workers from discrimination if measurements of skill are themselves influenced by racial bias.

Although the prevalence of subjective performance evaluations implies that discrimination is potentially important, the lack of performance evaluation data in most industries represents an obstacle to its study. A notable exception exists in professional sports—particularly Major League Baseball (MLB), where detailed records of player performance and their evaluators (umpires) are readily available. An umpire subjectively judges the performance of a pitcher many times during each game, deciding whether pitches are "strikes" or "balls," the former benefiting and the latter harming the pitcher. Because the pitcher's productivity and performance are so heavily influenced by the umpire's evaluations, discrimination by umpires could conceivably affect both games' outcomes and the labor market, i.e., pitchers' compensation.

We collect and analyze every pitch from three complete baseball seasons (2004-2006), paying particular attention to the race/ethnicity of the umpire, pitcher, batter and catcher, to explore racial/ethnic discrimination by umpires in the evaluation of pitchers. Our results are consistent with bias by umpires. Pitchers who share the race/ethnicity of the home-plate umpire receive favorable treatment, as indicated by a higher probability that a pitch is called a strike rather than a ball. Stunningly, this effect is *only* present in game and pitch situations where the umpire's actions are not heavily scrutinized. When an umpire's calls are reviewed by a computerized monitoring system, when the game is well attended, or when the pitch is particularly important, race/ethnicity plays no role in the umpire's evaluation. This effect is robust to a wide set of controls, including fixed effects for each pitcher, umpire, batter and catcher, suggesting that differences in umpire or player-specific characteristics are not driving the results.

Our data are particularly well suited to studying discrimination. Because every pitch is potentially subject to the home-plate umpire's discretion when it is thrown (several hundred times per game), there is sufficient scope for racial/ethnic discrimination to be expressed as well as for it to affect game outcomes significantly. In addition, we have a very large number of independent pitch-level observations involving the interaction of four different race/ethnicities: White, Black, Hispanic, and Asian. The data thus allow us not only to explore an umpire's preference for players of his own race/ethnicity, but also to examine preferences between other race/ethnicities , e.g., whether a Black umpire penalizes Hispanic pitchers relative to White pitchers.

An additional feature of baseball data is that, unlike other sports where a group dynamic among officials may alter the expression of individual biases, the home-plate umpire is exclusively responsible for calling every pitch in a typical baseball game.[1] Thus, if the home-plate umpire is biased, the outcomes affected by his bias are more likely to be observed in these data than in those from a sport where there are more interactions among members of an officiating team. Finally, the data allow a variety of tests for the existence of a price-sensitive demand curve for discrimination by umpires, as we develop several proxies for the price of discriminatory behavior.

Several studies (e.g., Garicano *et al*, 2005; Zitzewitz, 2006) have examined home-team preferences by referees/judges in sporting events, and another, Stoll *et al* (2004) examines racial match preferences in employment generally. Our study most closely resembles Price and Wolfers' (2007) work on NBA officiating crews' racial preferences. While we indeed

[1]Umpires can be positioned behind home plate or at first, second or third base. The home-plate umpire (the umpire-in-chief) occasionally appeals to either the first- or third-base umpire, but this is a relatively infrequent occurrence, and in any case is usually initiated by the home-plate umpire himself to help determine if the batter swung at the ball.

corroborate the latter's empirical findings for a different sport, our results speak primarily to the price-sensitivity of discrimination, i.e., whether and to what extent changing an umpire's incentives alters his expression of racial/ethnic bias. The resulting policy implications have relevance not only for baseball, but for the general labor market as well.

This research adds to a large literature on racial discrimination in sports, specifically in baseball, going back at least to Pascal and Rapping (1972) and Gwartney and Haworth (1974), with more recent examples being Nardinelli and Simon (1990), Findlay and Reid (1997) and Bradbury (2007). It includes studies of such outcomes as productivity, wages, customers' approbation of players, selection for honors, and others. There is some evidence of wage disparities among baseball players of different races, but the results are mixed, e.g., Kahn (1991). The conclusions of racial discrimination (or lack thereof) in this literature depend upon each player's productivity being accurately measured, as measured productivity is typically the crucial control variable. We suggest questioning this central assumption: If officials' judgments are themselves subject to racial/ethnic bias, adjusting for differences in the returns to *measured* productivity will not enable us to obtain proper measures of the extent of discrimination.

The results allow us to think about the deeper question of measuring discrimination generally. If, as we show here, evaluations of workers are affected by the match to the race/ethnicity of their evaluator, then the measured productivity of the worker will depend on the nature of that match. This difficulty has serious implications for measuring discrimination and is another manifestation of the problems in identifying discrimination pointed out by Donald and Hamermesh (2006).

In the following section we describe the pitch- and game-level data and explain our classification of umpires' and players' race/ethnicities. We then analyze individual pitches in Section III, presenting some evidence suggesting that umpires evaluate pitchers who match their own race/ethnicity more favorably than pitchers who do not. In Section IV we show that umpires express these preferences strongly only in times of low-scrutiny—game- and pitch-level situations where monitoring of the umpire is less. We examine the impact of discrimination on game outcomes and pitcher performance in Section V and in Section VI consider some other issues and provide a few checks for the robustness of our results. Section VII derives the size of the effects of the bias in performance evaluation on the measurement of wage discrimination and applies the results to salaries of baseball pitchers.

II. Data and Institutions

There are 30 teams in Major League Baseball, with each team playing 162 games in each annual season. During a typical game each team's pitchers throw about 150 pitches, so that approximately 730,000 pitches are thrown each season. We collect pitch-by-pitch data from ESPN.com for every MLB game in the three years 2004-2006.[2] For each pitch we identify the pitcher, pitcher's team, batter, batter's team, catcher, pitch count, score, inning, and pitch outcome. We classify each pitch into one of seven exhaustive and mutually exclusive categories: Called strike, called ball, swinging strike, foul, hit into play, intentional ball or hit by pitch. We supplement each pitch observation with game-level information from ESPN.com box scores, including the stadium name, home team, away team, team standings, and the identities and positions of all four umpires. In addition, for each pitcher's appearance in each game we collect the exact number of innings pitched, the numbers of hits, runs and home runs allowed, walks given up, strikeouts, and earned runs. Finally, for each starting pitcher in each game we collect the *GameScore,* a composite index designed to summarize a starting pitcher's performance.[3]

We next classify each position player, pitcher and umpire who appears in our dataset as White, Hispanic, Black or Asian. To begin this task, we collect country of birth for every player

[2]The URL for the pitch-by-pitch information is:
http://sports.espn.go.com/mlb/playbyplay?gameId=NNNNNNNN&full=1, where NNNNNNNNN represents the nine-digit game ID. The first six digits correspond to the year, month and date of the game. The box score information is from http://sports.espn.go.com/mlb/boxscore?gameId=NNNNNNNNN .

[3]Developed by baseball statistician Bill James, *GameScore* is a composite metric designed to gauge the performance of a starting pitcher. Pitchers are rewarded for recording outs, innings (more points for later innings), and strikeouts, but are penalized for allowing hits, runs, and walks.

and umpire. Players or umpires are classified as Hispanic if they are born in one of the following countries: Colombia, Cuba, Curacao, Dominican Republic, Mexico, Nicaragua, Panama, Puerto Rico or Venezuela. Similarly, players from Japan, South Korea and Taiwan are classified as Asian. We classify an additional 69 players using an AOL Sports article which lists every African-American player on a MLB roster at the beginning of the 2007 season.[4] We also utilize a similar list of past and present Hispanic players in MLB from Answers.com.[5] All remaining unclassified players and umpires are classified by visual inspection of pictures found in internet searches.[6] Three of the four race/ethnic groups are represented among umpires (there are no Asian umpires in MLB), and all four are represented among pitchers.

Our final dataset consists of 2,120,166 total pitches.[7] Table 1 presents their distribution across the seven possible pitch outcomes. The first row of the table summarizes all pitches, while subsequent rows sub-divide pitches based on the race/ethnicity of the pitcher, the batter and the home plate umpire, respectively. As Table 1 demonstrates, approximately 47 percent of pitches elicit a swing from the batter, hit the batter, or are intentionally thrown out of the strike zone. The remaining analysis focuses on the 53 percent of pitches (1.13 million) that result in called strikes and called balls, since these alone are subject to an evaluation by the home-plate umpire. Of these called pitches, about 32 percent are called strikes, and the rest are called balls.

Table 1 also reports the number of pitchers, batters and home-plate umpires in each of the four race/ethnicity categories. The percentages of White pitchers (71 percent) and batters (59

[4]The complete list can be found at http://Blackvoices.aol.com/Black_sports/special/_a/african-american-players-in-mlb/20070413095009990001.

[5]The complete list can be found at http://www.answers.com/topic/list-of-hispanic-players-in-major-league-baseball.

[6]For a small number of umpires, no pictures were available on the internet. For each of these individuals, we watched past games in which the umpire worked to ascertain his race/ethnicity. Any such classification is necessarily ambiguous in a number of cases. To the extent that we have inadvertently classified pitchers umpires, or batters in ways different from how they might be treated on the field, all we have done is introduce measurement error into the matches and thus reduce the strength of any results that we generate.

[7]Due to their unusual nature, we exclude All-Star and post-season games from the sample.

percent) are lower in our sample than the percentage of White umpires (87 percent). On the other hand, Hispanics, comprising 23 percent of pitchers and 27 percent of batters, are under-represented among umpires (only 3 percent). Black pitchers, batters and umpires comprise 3 percent, 11 percent, and 5 percent of the samples, respectively. Asian players comprise 3 percent each of pitchers and batters.

Table 2 reports the number of pitches thrown, the number of called pitches and the percentage of called pitches that are strikes for each pitcher/umpire racial/ethnic combination. About two-thirds of the called pitches in our sample occur when the umpire and pitcher share the same race/ethnicity (mostly a White pitcher in a game called by a White home-plate umpire). While the percentage of pitches that are called is similar in situations where the umpire's and pitcher's race/ethnicity match and in situations where they do not (53.4 percent), a central difference is that the percentage of called pitches that are strikes is higher when they match (32.1 percent) than when they do not (31.5 percent).

The highest percentage of called strikes occurs when both umpire and pitcher are White, while the lowest percentage is when a White umpire is judging a Black pitcher. What is intriguing is that while Black umpires judge Hispanic pitchers harshly relative to how they are judged by White and Hispanic umpires, Hispanic umpires treat Black pitchers nearly identically to the way Black umpires treat them. Minority umpires treat Asian pitchers far worse than they treat White pitchers and far worse than White umpires treat them.

III. Called Pitches and Umpire-Pitcher Matches

The summary statistics in Table 2 ignore possible inherent differences in the quality or "style" of pitchers by race/ethnicity. They also ignore the possible different outcomes generated by non-random assignment of pitchers to face different opponents, and of umpires to games played by particular teams.[8] To account for these and other potential difficulties, our central test for umpire discrimination is the specification:

(1) $$I(\text{Strike} \mid \text{Called Pitch})_i = \gamma_0 + \gamma_1 \text{UPM}_i + \gamma_2 \text{Controls}_i + \varepsilon_i,$$

where the dependent variable is an indicator of whether a called pitch is a strike, the γ are parameters, ε is a random error, and i indexes pitches. The main explanatory variable of interest is UPM, an indicator of whether the umpire (U) and pitcher (P) match (M) on race/ethnicity. In almost all of our tests, we include pitcher and umpire fixed effects, so that UPM picks up the marginal effect of a racial/ethnic match between the home-plate umpire and pitcher. In addition, we employ a number of important control variables. Pitch-count indicators, which record how many balls and strikes have accrued during a particular at-bat, are crucial because pitchers alter the location of their pitch based on the ball-strike count. Inning indicators are also included, because pitchers are usually less fatigued early in games, and because a pitcher who starts the game is often replaced by a "relief" pitcher in later innings, with a different (often reduced) accuracy.[9] Any home-field bias is captured by top-of-the-inning indicators**,** which account for whether the home (top=1) or visiting team (top=0) is pitching. Lastly, we include the pitcher's score advantage (defined as the number of runs, potentially negative, that the pitcher's team is ahead), since, if a pitcher is ahead in the game, he typically pitches more aggressively and is more likely to throw a pitch in the strike zone.[10]

Table 3 presents the results of estimating equations interacting pitcher and umpire race/ethnicity. All the estimates are based on linear-probability models (but probit estimates

[8]Examination of umpires' schedules indicates that while umpires typically travel as a four-person crew throughout much of the year, crews are randomly assigned across teams, ballparks, geography, and league (American or National). Furthermore, umpires rotate in a specific order, i.e., each serves as the home-plate umpire exactly every fourth game, resulting in random assignment of umpires to starting pitchers.

[9]In models with pitcher fixed effects, this second reason for inning indicators is obviously subsumed.

[10]The reason is that having a lead effectively reduces the pitcher's risk aversion. Relative to throwing a pitch likely to result in a walk, throwing a "hittable" pitch is risky—it increases the probabilities of both a very poor outcome for the pitcher (such as a home run) and a very good one (a fly out).

present the same picture). The first three pairs of columns show the analysis for Whites, Hispanics and Blacks, pitchers and umpires separately. The first column in each pair includes matches of the pitcher with umpires of each race/ethnicity (with pitcher fixed effects), while the second includes all possible matches of umpires with pitchers of each race/ethnicity (with umpire fixed effects). The final three columns include all pitchers and umpires, with each column adding successive vectors of fixed effects.[11] There is some, albeit weak evidence of favoritism by umpires for pitchers who match their race/ethnicity. Taking the results in Column (9) as the most indicative of the underlying behavior, it is quite clear, however, that there is no generally significant impact of the match (p=.12).[12]

Although the results with the broadest sets of fixed effects do not suggest a significant effect of the umpire-pitcher match, the point estimates imply that a given called pitch is approximately 0.27 percentage points more likely to be called a strike if the umpire and pitcher match race/ethnicity. Excluding (as we do) pitches where the batter swings, the likelihood that a given pitch is called a strike is 31.9 percent. Thus when the umpire matches the pitcher's race/ethnicity, the rate of called strikes rises by slightly less than 1 percent above the rate when there is no match.[13]

IV. Called Pitches When Discrimination Is Costly to the Discriminator

One might examine the results in Table 3 and conclude that, while the point estimates are interesting, their statistical insignificance means that there is really little here. Given an economist's view that agents' acting out their preferences will react to the price of an activity,

[11]We include all pitchers in these regressions, although a case could be made that Asian pitchers should be excluded because they are never judged by an umpire of the same race. All the results are nearly identical if they are excluded.

[12]In unreported results, we estimated models with proxies for pitcher accuracy, e.g., earned run average (ERA) or walks/inning, with no qualitative change in the results.

[13]As a check on this issue we re-estimated the model including sequentially the race/ethnic match between the first-, second- and third-base umpire and the pitcher. None of these extensions materially changes our conclusions.

however, it is worthwhile examining the impacts of umpire-pitcher matches as the price of discrimination changes. Our data are particularly well suited to study this question, and it is our primary focus for much of the remaining analysis.

We begin by asking what factors affect the price of expressing racial or ethnic discrimination. Studies of cognitive behavior indicate that presenting the biased party with counter-examples of the stereotype of interest can reduce the severity and/or frequency of the biased behavior (Goodwin *et al*, 2000; Blair, 2002). In other words, simply making conscious a sub-conscious bias imposes a sufficient psychological cost to mitigate its expression. Another mechanism is to increase the visibility of the biased party's behavior, potentially exposing the offender to social or statutory penalties. In this section we proxy the price of discrimination by the extent to which an umpire's evaluations of pitchers will be scrutinized, and employ three different measures to examine whether a higher price of discrimination reduces the extent to which umpires engage in discriminatory behavior.

The first source of scrutiny is QuesTec, a computerized monitoring system intended to evaluate the accuracy and consistency of home-plate umpires' judgments. In 2003 MLB installed QuesTec in 11 of its 30 ballparks.[14] QuesTec's Umpire Information System (UIS) consists of four cameras that track and record the location of each pitch, providing information about the accuracy and precision of each umpire's ball and strike calls. Despite opposition from some umpires and players (perhaps most notably, pitcher Curt Schilling's assault on one of the cameras after a poor performance in 2003), the QuesTec system served as an important tool to evaluate umpires during our sample period. According to the umpires' union's agreement with MLB, QuesTec is the primary mechanism to gauge umpire performance. In particular, if more than 10 percent of an umpire's calls differ from QuesTec's records, his performance is considered

[14]These were the ballparks of the Anaheim Angels, Arizona Diamondbacks, Boston Red Sox, Cleveland Indians, Oakland Athletics, Milwaukee Brewers, Houston Astros, New York Mets, Tampa Bay Devil Rays, Chicago White Sox, and New York Yankees.

substandard, and that may influence his promotion to "crew chief," assignment to post-season games, or even retention in MLB.[15]

Because QuesTec is installed in roughly 35 percent of ballparks, and because umpiring crews are rotated randomly around the league's ballparks, virtually every umpire in our dataset calls a substantial number of pitches in parks both with and without QuesTec.[16] Additionally, both the umpires' and teams' schedules change every year, exposing each umpire to a wide cross-section of batters and pitchers in both QuesTec and non-QuesTec parks. Throughout the analysis we test whether greater scrutiny—the possibly higher cost of indulging in personal discretion in QuesTec parks—leads umpires to call strikes "by the book." Any role that racial/ethnic (or any other) preference plays in influencing pitch calls should be mitigated if the costs of being judged substandard by QuesTec are sufficiently high. Pitchers, however, may act strategically in response to the scrutiny of umpires, altering how they pitch depending on whether the game is in a QuesTec park or not.[17] For this reason, in all of the estimates in this part we include fixed effects not only for each pitcher, umpire and batter, but also for the presence or absence of QuesTec in each game, i.e., pitcher-Questec fixed effects, etc.

Figure 1 graphs the average percentage of called pitches that are strikes in ballparks with and without QuesTec for White and non-White pitchers respectively. The effect of monitoring on umpires' behavior is apparent, with both White and non-White pitchers being judged differently by umpires of the matched race/ethnicity depending on whether the pitch is thrown in a park with

[15] An umpire's evaluation is not based solely on his performance as measured by QuesTec. If an umpire falls below the QuesTec standards, his performance is then reviewed by videotape and live observation by other umpires to determine his final evaluation score. No such measures are taken, however, if an umpire meets the QuesTec standards.

[16] The fraction of games in which QuesTec was installed was virtually identical for all umpires in our sample, differing for the few umpires calling only a handful of games.

[17] For example, New York Mets pitcher Tom Glavine, known as a "finesse" pitcher who depends on pitches close to the strike zone border, complained publicly that QuesTec's influence on umpire calls forced him to change his style (Associated Press, July 9, 2003). Glavine reports that he was told, "[umpires do] not call pitches on the corners at Shea [his home ballpark] because they [the umpires] don't want the machine to give them poor grades."

QuesTec installed. The difference in the called-strike percentage between QuesTec and non-QuesTec parks is significant for both White and non-White pitchers.

Table 4 contains the results of estimating (1) separately by the presence of QuesTec in the ballpark. The regressions presented in Table 4 also include controls for inning, pitch count, pitcher score advantage, and top of the inning.[18] The results are remarkable: In ballparks with the umpire monitoring system (Column 1), the coefficient on UPM is -0.35 percentage points and is not significantly different from zero. In parks without QuesTec, shown in Column (2), the same coefficient is 0.63 percentage points per pitch (p=.015).

Column (3) of Table 4 presents the results when the QuesTec indicator is interacted with UPM. When the pitcher and umpire match race/ethnicity, pitching in a QuesTec park reduces the likelihood that a called pitch is ruled a strike by almost 1 percentage point, more than offsetting the favoritism shown by an umpire to a matched pitcher when the former is not scrutinized by QuesTec. Each effect is statistically significant. Implicitly umpires indulge their apparent preference for matched pitchers when the pitches underlying their decisions are not recorded, so that the effects found in the previous section average the statistically significant positive impact of an unscrutinized match in a non-QuesTec park with the statistically insignificant negative impact in a QuesTec park. Thus, in the presence of price-sensitive discrimination, we should *expect* the point estimates in Table 3 to be low, given that the entire sample consists of a mix of high- and low-scrutiny games.

Even though the negative effect of a match in a QuesTec park is not statistically significant, what is intriguing is why umpires' decisions might favor unmatched pitchers when they are scrutinized. One might speculate that umpires feel that they are favoring matched pitchers in other parks and that they sub-consciously overcompensate in instances when they know they are under scrutiny.

[18] The direct effect of being in a QuesTec park is, of course, not directly observable, as it is subsumed in the pitcher-QuesTec fixed-effects terms.

To extend these results, we employ two additional measures that proxy the scrutiny of umpires and thus the price of discrimination. First, we collect each game's crowd attendance and then divide by each ballpark's capacity to arrive at a "percentage of capacity attendance." We scale by the size of each venue for two reasons. First, we are attempting to proxy the number of fans sitting close enough to home plate to judge whether a pitch is a strike or a ball. Although ballparks vary considerably in overall size, the concentration of seats close to home plate is nearly identical. If a stadium populates relatively uniformly based on the interest in each game, then the *number* of fans close to the pitcher, catcher, and umpire will be highly correlated with the percentage of capacity attendance for each game.[19] A second reason is that a game's attendance relative to its capacity may be correlated with the number of viewers watching the game on television. Scaling by ballpark size partly mitigates the possible low correlation between the size of a team's stadium and the size of its television market (compare the Chicago Cubs, Boston Red Sox, Toronto Blue Jays, etc.).

Figure 2 shows that crowd attendance alters umpires' behavior dramatically. A game is defined as "well-attended" if the crowd attendance is above the median percentage capacity for the sample, roughly 70 percent. (Our results are not sensitive to normalizing each game's attendance in this fashion, being nearly identical when we use the percentage attendance.) Compared to well-attended games, umpires calling poorly-attended games favor pitchers of matched race/ethnicity, as evidenced by higher called-strike percentages. In the case of White pitchers, both non-White and White umpires tend to call fewer strikes in poorly-attended games, but the reduction in strikes called by non-White umpires is over three times larger. The same effect is seen to an even greater degree among non-White pitchers. Umpires whose race/ethnicity

[19]There is no way to test this assumption directly. The fact that such premium seats are almost exclusively held by season ticket holders who would have to sell their tickets in a secondary market suggests, however, that many "close-in" seats simply go vacant in games of little interest.

matches non-White pitchers call nearly 1.5 percent more strikes in poorly-attended games, whereas unmatched umpires call fewer strikes.

In Columns (1) and (2) respectively of Panel A in Table 5 we show the results of estimating (1) separately for both well- and poorly-attended games. Each equation includes the same pitcher, umpire and batter fixed effects that are included in Table 4, as well as the control variables included there, e.g., pitch counts, inning indicators, etc. As with the QuesTec results, the UPM variable is significant at the 1 percent level only in poorly-attended games, with an effect of 0.84 percentage points per pitch. During well-attended games there is no significant effect of an umpire-pitcher racial/ethnic match and, indeed, the point estimate is negative. Column (3) generalizes the results by aggregating all games, with UPM interacted with an indicator of a well-attended game. Compared to a pitch in a poorly-attended game when the umpire and pitcher do not match, a pitch called by an umpire of the same race/ethnicity as the pitcher is 0.52 percentage points more likely to be judged a strike. If the game is well attended, a pitch is no more likely to be called a strike if the pitcher and umpire match race/ethnicity. The results for this completely different proxy of the price of discrimination are qualitatively identical to those obtained for the QuesTec vs. non-QuesTec distinction.

A third proxy for the scrutiny of umpires varies many times within each game. We separate pitches into two categories, "terminal" and "non-terminal." A pitch is potentially terminal if the umpire's next judgment can terminate the batter's plate appearance. Thus, for example, a pitch that is thrown with two strikes is potentially terminal, and the umpire's judgment may be scrutinized more heavily by the pitcher, batter, catcher, managers and fans. The same is true for a count with three balls. An initial glimpse into the effects of this distinction is shown in Figure 3. Again here we observe the same contrast between umpires' calls in terminal counts, where umpires are likely to be under more scrutiny, and non-terminal counts, that we saw for the previous two proxies for scrutiny.

Columns (4) and (5) of Panel B of Table 5 show estimates of (1) separately for terminal and non-terminal pitches, with pitcher, umpire and batter fixed effects and the now standard set of control variables. Each type of pitch is considered separately, with the result that the coefficients of UPM have opposite signs. For pitches of lesser importance, where the pitch cannot be terminal, the estimated coefficient of UPM is 0.46 percentage points (p=.03)—umpires favor pitchers who match their own race/ethnicity. For potentially terminal pitches, where scrutiny of the umpire is likely to be greater, umpires appear to judge pitchers of their own race/ethnicity insignificantly more harshly than unmatched pitchers. In Column (6) all pitches are aggregated and UPM is interacted with an indicator for potentially terminal pitches. The results mimic those implicit in the estimates in Columns (4) and (5).

In Columns (7) and (8) we examine how umpire-pitcher matches in terminal and non-terminal counts change when they are observed early in a game, when the decisions may be less well scrutinized, and late in the game, when their impact on the game's outcome may be greater. We designate the first third (three innings) of a game as "early," assuming that the umpire's actions are less closely scrutinized when the game's outcome is far from certain, and the last six (or more) innings as "late." We expect that a terminal count will have a stronger effect on the outcome of a pitcher-umpire racial/ethnic match in early innings. Comparing the results across the two columns, we see that this is the case, with the magnitude of the interaction between terminal count and UPM being nearly twice as large in early as in late innings (0.98 vs. 0.56 percentage points).[20]

One might argue that these effects do not reflect the impact of an increased price of discrimination, but instead represent greater care expended by umpires when their decisions are more important. An alternative way of framing this question is to ask whether umpires

[20]One could imagine still more indicators of the potential extent of the scrutiny of umpires. Indeed, we performed similar analyses for one of them, the closeness of a game, with results very much like those found for non-terminal and terminal pitches.

consciously change their behavior in response to explicit threats of punishment, or whether any effect of sub-conscious bias is mitigated by focusing on important situations. Unfortunately, the perfect correlation between important situations and scrutiny prevents us from addressing this interesting question. Even in QuesTec stadiums, where thousands of comparatively meaningless games are played, these games are highly important for umpires, precisely *because* the associated scrutiny is relevant for their career concerns. Although the inability to distinguish between conscious and sub-conscious bias is likely to present a similar difficulty were one to extend our results to employment relationships more generally, the policy implications between the two are identical: To protect workers from the adverse effects of discrimination, increase the consequences for those who discriminate.

V. Impacts on Games' Outcomes

When viewed in isolation, the pitch-level results presented in Tables 1-4 may seem trivial, affecting at most a pitch or two per game. The importance of the results cannot, however, be appreciated without recognizing that individual at-bats are highly path dependent. Suppose, for example, that the count is 1-1 (one ball and one strike), and that the umpire calls a pitch a strike that should have been called a ball. Rather than a 2-1 count that favors the batter, the resulting count is 1-2, one that significantly advantages the pitcher. Armed with this advantage, the pitcher takes more risks, throwing pitches that are more difficult to hit (as evidenced by the coefficients on the count indicators in Table 3). This translates into batters getting fewer hits – in 2004, the batting average on 2-1 counts was .330, but on 1-2 was merely .176. Given that baseball is a relatively low-scoring game, altering the outcomes of one or two at-bats can be of central importance.

This example highlights the fact that racial bias may affect the game both directly and indirectly. It is obvious that the direct effect of racial bias alone, such as the potential for a pitcher who faces a racially/ethnically unmatched umpire striking out fewer batters or giving up more walks, can alter games, especially close ones. The indirect effect—when players anticipate the effect of a biased umpire and strategically alter their behavior—may, however, have an even larger impact on outcomes.

To examine this issue, we analyze a variety of game-level performance measures for each starting pitcher in our sample: Wins, hits, earned runs, home runs, strikeouts, walks, and *GameScore*.[21] Figure 4 graphs each performance measure for the roughly 14,000 starting pitchers in the roughly 7,000 games in the three seasons in our sample. As in the previous figures, we display the results for White and non-White pitchers separately to highlight the magnified effect of racial/ethnic preference on non-White pitchers.

For virtually every measure of pitcher performance, the impact of having a matched umpire benefits the pitcher. The composite measure, *GameScore*, is raised for both White and non-White pitchers when the home-plate umpire's race/ethnicity matches theirs. Similarly, both White and non-White pitchers allow fewer home runs (HR), hits, runs and walks, and have lower earned-run-averages (ERA), when a match occurs. Only strikeouts (K) among White pitchers do not accord with the observed racial/ethnic preferences by umpires, although the effect is minuscule. Among these performance measures, most are not solely influenced by the umpire's judgment. Yet many indirect outcomes, such as the number of home runs allowed by the pitcher, are also affected, suggesting that the umpire's behavior may alter the strategies of pitchers and batters.

Looking at the mean outcomes in various instances of umpire-pitcher match, the obvious benchmark is the case when both or neither starting pitcher matches the umpire's race/ethnicity. In that case, the home team wins 53.8 percent of the time, reflecting a slight home-field advantage. In 18.7 percent of the games only the home-team pitcher matches the umpire, while the opposite case, a match only between the visiting-team pitcher and the umpire, occurs 19.0

[21]Although most of our results are similar when we include all pitchers, starting pitchers are of particular interest because of their relative importance and because a team's starting pitcher generally interacts directly with the umpire far more than any other member of the team besides its catcher. In addition, *GameScore* is only calculated for starting pitchers.

percent of the time.[22] In the former case, the home team wins 55.6 percent of its games. In the latter case the home team's winning percentage is unaffected—it remains 53.8 percent. These differences in the means suggest that there is an asymmetry in the impact of racial/ethnic matching: Matches are much more important between the umpire and the home-team's pitcher than between the umpire and the visiting team's pitcher.

The effect of racial/ethnic preferences on winning probabilities is even more striking when we disaggregate by umpire race/ethnicity. With White umpires the home team wins 54.4 percent of the time if its starting pitcher is White, but only 52.9 percent of the time if he is not. In the case of Black umpires, the corresponding percentages are 72.7 percent and 55.1 percent, although there are only 11 games in which a Black starting pitcher is evaluated by a Black umpire. In the 36 games in which both the starting pitcher and the umpire are Hispanic, the home team wins 61.1 percent of its games, compared to 52.0 percent if the pitcher is non-Hispanic.

Our game-level sample comprises 7,124 games, accounting for approximately 98 percent of all games played in these three seasons. For each of these games we compare the race/ethnicity of both starting pitchers to that of the umpire and analyze whether racial/ethnic relationships influence a particular outcome, adjusting for other characteristics. In Column (1) of Table 6 we present estimates with the dependent variable equaling one if the home team wins. We include the number of runs scored by the pitcher's team and specify fixed effects for the pitcher, the umpire and the identity of the opposing team. The coefficient on UPM is marginally significant, with a magnitude of slightly over 4 percentage points.

Columns (2) and (3) examine the effect of an umpire-pitcher match on *GameScore* (with higher values of the dependent variable indicating a better performance) and the number of runs allowed by each starting pitcher (so that both variables are available for both starting pitchers in a

[22] That these are nearly identical is further evidence of random matching between umpire and pitcher races/ethnicities.

game).[23] The results are qualitatively similar to those obtained on the probability of winning—there is a benefit to a pitcher's *GameScore* if he matches the race/ethnicity of the umpire; and there is a marginally significant impact of the pitcher-umpire match, even after adjusting for all the vectors of fixed effects, on the number of runs allowed.

In light of the evidence that the effects of an umpire-pitcher match are seen only when scrutiny is less, we can disaggregate the samples underlying the estimates in Table 6 and estimate the equations separately for games played in QuesTec and non-QuesTec parks. Given the small sub-samples and the tightness of the specification with the inclusion of all three vectors of fixed effects, even the estimated effects of an umpire-pitcher match become only slightly more significant when we use the non-QuesTec sub-samples. Nonetheless, for all three dependent variables the impacts of the match are larger in absolute value in these sub-samples than in the sub-samples for QuesTec parks.[24]

VI. Robustness Checks and Other Considerations

A. Accounting for Matches with Batters' and Catchers' Race/Ethnicity

It is natural to suppose that an umpire influenced by the race of the pitcher may also be influenced by that of the batter or the catcher, especially since in the latter case the umpire is in continuing close contact. We explore this possibility extensively, but find no evidence to support the argument. As shown in Column (1) of Table 7, estimating (1) substituting UBM, defined as a racial/ethnic match between umpire and batter, for UPM yields absolutely no effect rather than the negative effect that one might expect. A catcher-umpire match, indicated by the analogously defined variable UCM, has a small insignificant positive effect on the probability of a called

[23]The increase in sample size is due to the fact that game-level outcomes are analyzed in Column (1), so that the number of observations is the number of games in our sample. In contrast, the remaining columns consider the performances of each starting pitcher as independent observations, roughly doubling the sample size.

[24]In most of the work in this section we have concentrated on the three outcome variables win probability, *GameScore* and runs allowed to avoid duplication. The results are qualitatively the same when we expand the analysis to the other measures depicted in Figure 4.

strike, as shown in Column (2). These results are unchanged when all three match variables—UPM, UBM and UCM, are included simultaneously (Column (3) of Table 7), and, indeed, the coefficients on all three match variables are essentially the same as when each is included separately.

When interactions of UPM, UBM and UCM with QuesTec, high-attendance and terminal counts are sequentially included in Equation (1), as shown in Columns (4)-(6) of Table 7, each indicator UPM is still statistically significantly positive, and each interaction with UPM is negative. Moreover, none of the main effects of UBM and UCM approaches significance, nor do the interaction terms with those indicators. The correlation of the between-game proxies for umpire scrutiny—QuesTec and attendance percentage—is below 0.05, suggesting that these are indeed independent measures of scrutiny. Because the type of pitch (terminal or non-terminal) is a within-game measure, it is uncorrelated with either between-game measure. Thus it is not surprising that the results in Column (7), in which all the main-effect and interaction terms are included, give the same picture as the other results: Implicitly umpires engage in discrimination against unmatched pitchers, and each proxy for a higher price of discrimination reduces umpires' demand for discriminatory outcomes. Umpires appear focused on the pitchers they are judging—there is no evidence whatsoever that matches with other relevant players affect their judgment.

For at least two reasons the absence of any impact of UBM may not be as puzzling at it first appears. First, as suggested above, the per-pitch effect represents racial/ethnic discrimination only relatively infrequently and is concentrated in low-scrutiny situations. Both scrutiny and batters' race/ethnicity change frequently (many times within each game), so any effect may be swamped by the impact of scrutiny. We have no such concerns about statistical power with pitchers, who interact with each umpire over a hundred times within each game under varying degrees of scrutiny. The second possibility is more subtle, owing to the physical proximity of the umpire and batter relative to that of the umpire and pitcher. Psychological studies suggest that, although people may not recognize their own prejudice (Bargh, 1999, Devine and Monteith, 1999), the risk of being confronted reduces the frequency of biased behavior (Czopp *et al*, 2006). If physical proximity to the batter increases the probability of confrontation for an umpire, perhaps it acts as an additional check on the umpire's tendency to express discrimination.[25]

That there is only a small, insignificant positive effect of an umpire-catcher match, given the proximity of the umpire and catcher for half of each game, may be more surprising. It suggests, however, that umpires realize that they are judging the pitcher. A match with the catcher is less important; and additional interactions, of UPM with UCM, do not alter the conclusions about the effect of each first-order interaction individually. One can speculate about the absence of a UCM effect, including the possibility that only the pitcher directly faces the umpire, but we cannot distinguish among possible interpretations.

B. Accounting for Umpire and City Characteristics

It may be that umpires' measurable characteristics (beyond their race/ethnicity) and those of the city where a game is played explain our results. We collect demographic information on each umpire from a variety of sources, data including his age and experience, and in many cases both his state of birth and residence. For each ballpark we also obtain the racial/ethnic breakdown of the surrounding metropolitan statistical area.

While we find no evidence that the racial composition of an umpire's birthplace or residence predicts his propensity to penalize non-matching players, there is somewhat weak evidence that discrimination is more likely among younger and less experienced umpires. The coefficient on UPM in the re-estimation of Equation (1) among the upper half of umpires ranked by experience is less than half its magnitude in estimates for umpires in the lower half of the distribution of experience. In addition, the 18 "crew chiefs," veterans selected for their seniority and performance, do not appear influenced by the race/ethnicity of the pitcher: If (1) is estimated

[25]Batters' confrontations with the umpire are far more common than pitchers', lending support to this interpretation of the evidence.

separately for this group, the point estimate of the coefficient on UPM is nearly zero. This evidence is consistent with either a model of selection or learning. Perhaps discriminating umpires are not promoted and are dropped from the ranks. Alternatively, experience may teach umpires to restrain their own biases, so that highly experienced umpires are not likely to express racial/ethnic bias in their subjective calls.

We also re-estimated the basic equation for Blacks, and for Hispanics, separately, adding in each case main effects and interactions with UPM of the percentage of the minority group in the metropolitan area where the ballpark is located. Among Blacks the interaction was positive, but statistically insignificant; among Hispanics it was negative, but also statistically insignificant. Our conclusions are not affected by the racial/ethnic mix of the team's catchment area.

C. Other Issues

The overwhelming majority of minority pitchers are Hispanic. In our main tests, we aggregated them, but some are Hispanic Whites, others are Black Hispanics. We inspected their pictures, divided the Hispanic aggregate into these two groups and re-defined UPM to allow for the possibility that the two different groups of minority umpires might treat Hispanic pitchers who match their own characteristics differently from other Hispanic pitchers. This reclassification had almost no effect on the estimates produced in Tables 3-5. Implicitly, Hispanic and other umpires treat Hispanic pitchers the same regardless of the pitcher's racial identity.[26]

As the discussion has made clear, there is no objective measure of the quality of a pitch. We only have information on whether it is called and, conditional on that, if it is called as a ball or strike. It might, for example, be that pitchers, assuming that they will be treated worse if there is a racial/ethnic mismatch, are "rattled" and less likely to pitch strikes. We cannot refute this possibility with certainty; but one might argue that the absence of any mismatch effect on

[26] In addition, we investigated whether American-born Hispanic pitchers were treated differently from Hispanic pitchers born outside the U.S. We find no evidence that the Hispanic pitcher's birthplace affects the expression of any racial/ethnic bias by umpires.

terminal pitches, when this effect would be most likely to prevail, suggests the argument is invalid. Even if it were valid, such a finding would still support the main result, although we would classify it as an “indirect” effect, similar to the effect if a pitcher intentionally altered his strategy in expectation of the umpire’s bias.

Our estimates in Table 6 would still be unbiased if managers were able to alter their starting pitchers’ assignments to take advantage of the umpires’ preferences that we have demonstrated exist. Nonetheless, it is interesting to inquire whether managers are implicitly both aware of these preferences and able to act upon them. The racial/ethnic endowments of umpires and starting pitchers in the 7,124 games in our sample would lead one to expect matches in 0.680 of the games. In fact, matches occur in only 0.677 of the games. The difference, aside from being in the unexpected direction, is statistically insignificant (t=-0.69). Quite clearly there is no evidence in our sample of non-random matching of umpires and starting pitchers.

VII. Biases in Measuring Wage Discrimination

In the previous sections we generate some evidence that presumably objective measures of a worker’s (in this case, baseball pitcher’s) activities can be subtly affected by his evaluator’s racial/ethnic preferences, and that this effect in turn leads to reductions in his measured productivity (the game outcomes discussed in Section V). To the extent that pay is based on measured productivity, this finding carries important implications for measuring the extent of discrimination in baseball and in labor markets generally. In particular, it implies that estimates of the extent of discrimination will be understated.

Consider a simple earnings equation:

(2) $W_i = \alpha M_i + \beta P^*_i + \upsilon_i$,

where W is the logarithm of earnings, M an indicator of minority status, P* worker i’s true productivity, and υ a random error in the determination of earnings. The parameter α is the true effect of minority status on earnings when productivity measurements are free of bias. Assume that the majority workers’ productivity is measured without bias, but that minority workers are subject to a negative bias in their assessment by evaluators, which leads to a shortfall of their measured productivity P below their true productivity:

$$(3) \qquad P_i = P^*_i - \varphi, \text{ if } M=1; \qquad P_i = P^*_i, \text{ if } M=0,$$

$\varphi>0$. Then we can rewrite (2) to obtain an estimating equation in observables:

$$(2') \qquad W_i = [\alpha+\beta\varphi]M_i + \beta P_i + \upsilon_i \text{, or}$$

$$W_i = \alpha' M_i + \beta P_i + \upsilon_i .$$

The standard estimate of earnings discrimination adjusted for productivity differences, α', has a positive bias in the amount $\beta\varphi$.

To obtain some feel for the size of this bias in the particular case that we have examined we can simulate the wage effects using the estimates of φ underlying Figure 4 and estimates of β from studies of salary determination in Major League Baseball. We are essentially estimating the reduction in minority pitchers' salaries as a result of the average amount of discrimination arising during the 2004-2006 seasons due to umpire-pitcher racial/ethnic matches. Kahn (1993, Table A2) estimates equations like (2') using a set of outcome measures that can be conformed to ours by including the percentage of games won and ERA. Making reasonable assumptions about the means of these outcomes for starting pitchers in 2006, applying the effects in Figure 4, and using his parameter estimates yields an estimated bias of $\beta\varphi = 0.039$. Gius and Hylan (1996, Table 6.2) use strikeouts/innings, walks/innings and winning percentage, all of which are also conformable with our outcome measures. The same method based on their parameter estimates produces an estimate of $\beta\varphi = 0.014$. Finally, using the estimates for starting pitchers by Krautmann *et al* (2003), the estimate of $\beta\varphi = 0.084$.[27]

[27]For the percentage of games won we use 0.5. The mean ERA is around 4.3, the mean strikeouts/inning by starting pitchers is around 0.7, and the mean walks/inning by them is around 0.3. We can take the estimates of the bias that we have produced as examples here to infer the dollar impacts of this subtle form of discrimination. In 2006 the average salaries of starting pitchers in MLB were $4.8 million. A bias to the estimated effect of minority status on compensation of starting pitchers of between 1 and 8 percent suggests that those pitchers are underpaid relative to White pitchers by between $50,000 and $400,000 per year.

While we have demonstrated the extent of bias to estimated discrimination in earnings that arises because of biased evaluations of Major League Baseball pitchers, this effect is probably smaller than would be observed for workers generally. The scope for the expression of racial/ethnic preferences of umpires for/against pitchers is almost surely far less than in most workplaces. Evaluations of pitchers are made at discrete and very frequent times—when a pitch is thrown. These are not one-shot comments made at most monthly at the evaluator's leisure. Also, as our demonstrations of reduced bias when there is greater scrutiny suggest, there are quite stringent external limits on the expression of bias against unmatched pitchers. The relative lack of such limits in the general workplace suggests that the example here may provide a lower bound on the extent of bias to estimates of discrimination generally.

VIII. Conclusions

The analyses of individual pitches and game outcomes suggest that baseball umpires express racial/ethnic preferences in their decisions about players' performances. Pitches are more likely to be called strikes when the umpire shares the race/ethnicity of the starting pitcher, an effect that only is observable when umpires' behavior is not well monitored. The evidence also suggests that this bias is strong enough to affect pitchers' measured performance and games' outcomes. As in many other fields, racial/ethnic preferences work in all directions—most people give preference to members of their own group. The difference in MLB, as in so many other fields of endeavor, is that power belongs disproportionately to members of the majority—White—group.

The type of discrimination that we have demonstrated is disturbing because of its implications for the sports labor market. In particular, non-White pitchers are at a significant disadvantage relative to their White peers, even in the absence of explicit wage discrimination by teams. Although some evidence suggests such explicit discrimination exists, i.e., there is a wage gap among baseball players of different races, the fact that nearly 90 percent of the umpires are White implies that the *measured* productivity of non-White pitchers may be downward biased.

Implicitly, estimates of wage discrimination in baseball that hold measured productivity (at least of pitchers) constant will understate its true size.

More generally, our results suggest caution in interpreting any estimates of wage discrimination stemming from equations relating earnings to race/ethnicity, even with a large set of variables designed to control for differences in productivity. To the extent that supervisors' evaluations are among the control variables included in estimates of wage discrimination, or even if they only indirectly alter workers' objective performances, their inclusion or their mere existence contaminates attempts to infer discrimination from adjusted racial/ethnic differences in wages. If racial/ethnic preferences in evaluator-worker matches are important, standard econometric estimates will generally understate the magnitude of racial/ethnic discrimination in labor markets.

While the specific evidence of racial/ethnic match preferences is disturbing, our novel analysis of the demand for discrimination should be encouraging: When their decisions matter more, and when evaluators are themselves more likely to be evaluated by others, our results suggest that these preferences no longer manifest themselves. These findings imply that it should not be difficult for MLB to devise methods to eliminate the impacts of racial/ethnic match preferences.[28] Clearly, raising the price of discrimination in the labor market generally through analogous methods is more difficult; but these results may suggest measures that might have the desired effects.

[28]Whether the installation of a new strike-zone evaluation tool (ZE) in all baseball parks, projected during 2007, created the same incentives as QuesTec and vitiated apparent umpire discrimination is not clear.

References

Bargh, John, "The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effects," in Shelley Chaiken and Yaacov Trope, eds., *Dual-Process Theories in Social Psychology*. New York: Guilford Press, 1999.

Blair, Irene, "The Malleability of Automatic Stereotypes and Prejudice," *Personality and Social Psychology Review*, 6 (2002): 242-261.

Bradbury, J.C., *The Baseball Economist: The Real Game Exposed*. New York: Dutton, 2007.

Czopp, Alexander, Monteith, Margo, and Mark, Aimee, "Standing Up for a Change: Reducing Bias Through Interpersonal Communication," *Journal of Personality and Social Psychology*, 90 (May 2006), 784-803.

Devine, Patricia., and Monteith, Margo, "Automaticity and Control in Stereotyping," in Shelley Chaiken and Yaacov Trope, eds., *Dual-Process Theories in Social Psychology*. New York: Guilford Press, 1999.

Donald, Stephen, and Hamermesh, Daniel, "What Is Discrimination? Gender in the American Economic Association," *American Economic Review*, 96 (September 2006): 1283-1292.

Findlay, David, and Reid, Clifford, "Voting Behavior, Discrimination and the National Baseball Hall of Fame," *Economic Inquiry*, 35 (July 1997): 562-578.

Garicano, Luis, Palacios-Huerta, Ignacio and Prendergast, Canice "Favoritism under Social Pressure," *Review of Economics and Statistics*, 87 (May 2005): 208-216.

Goodwin, Stephanie, Gubin, Alexandra, Fiske, Susan, and Yzerbyt, Vincent, "Power Can Bias Impression Processes: Stereotyping Subordinates by Default and By Design," *Group Processes and Intergroup Relations*, 3 (2000): 227-256.

Gwartney, James, and Haworth, Charles, "Employer Costs and Discrimination: The Case of Baseball," *Journal of Political Economy*, 82 (July-August 1974): 873-881.

Kahn, Lawrence, "Racial Discrimination in Professional Sports: A Survey of the Literature," *Industrial and Labor Relations Review*, 44 (April 1991): 395-418.

Krautmann, Anthony, Gustafson, Elizabeth, and Hadley, Lawrence, "A Note on the Structural Stability of Salary Equations: Major League Baseball Pitchers," *Journal of Sports Economics*, 4 (February 2003): 56-63.

Nardinelli, Clark, and Simon, Curtis, "Customer Racial Discrimination in the Market for Memorabilia: The Case of Baseball," *Quarterly Journal of Economics*, 105 (August 1990): 575-595.

Pascal, Anthony, and Rapping, Leonard. "The Economics of Racial Discrimination in Organized Baseball," in Anthony Pascal, ed., *Racial Discrimination in Economic Life*. Lexington, MA: Heath, 1972.

Price, Joseph, and Wolfers, Justin, "Racial Discrimination Among NBA Referees," National Bureau of Economic Research, Working Paper No. 13206, June 2007.

Scully, Gerald, "Pay and Performance in Major League Baseball," *American Economic Review,* 64 (December 1974): 915-930.

Stoll, Michael, Raphael, Steven and Holzer, Harry, "Black Job Applicants and the Hiring Officer's Race," *Industrial and Labor Relations Review*, 57 (January 2004): 267-287.

Eric Zitzewitz, "Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making," *Journal of Economics and Management Strategy*, 15 (Spring 2006): 67-99.

Table 1. Pitch Summary, Major League Baseball, 2004-2006

	Called Strike	Called Ball	Swinging Strike	Foul	In Play	Intentional Ball	Hit by Pitch
All	360,809	771,314	188,989	362,381	417,211	13,956	5,506
Pitcher							
White (N=669)	260,601	552,545	132,574	259,752	301,718	10,018	3,883
Hispanic (N=219)	81,175	176,967	46,219	83,184	92,805	3,222	1,326
Black (N=27)	8,489	19,229	5,014	9,357	10,215	288	134
Asian (N=29)	10,544	22,573	5,182	10,088	12,473	428	163
Batter							
White (N=833)	189,239	401,755	98,314	185,183	208,976	6,601	3,156
Hispanic (N=385)	107,219	228,911	56,167	111,248	131,292	4,537	1,430
Black (N=154)	57,208	125,956	31,352	58,794	68,651	2,472	838
Asian (N=31)	7,143	14,692	3,156	7,156	8,292	346	82
Umpire							
White (N=85)	329,826	704,531	172,858	331,463	381,534	12,829	5,047
Hispanic (N=3)	10,681	22,884	5,471	10,488	12,198	402	174
Black (N=5)	20,302	43,899	10,660	20,430	23,479	725	285

Table 2. Summary of Umpires' Calls by Umpire-Pitcher Racial/Ethnic Match*

	Pitcher Race/Ethnicity				
	White	**Hispanic**	**Black**	**Asian**	**Total percent called strikes**
Umpire Race/Ethnicity					
White					
Pitches	1,388,318	445,107	47,797	56,866	
Called pitches	741,729	236,937	25,108	30,583	
Percent called strikes	32.06 (0.05)	31.47 (0.10)	30.61 (0.29)	31.97 (0.27)	31.89 (0.05)
Hispanic					
Pitches	45,603	13,737	1,552	1,406	
Called pitches	24,592	7,323	845	805	
Percent called strikes	31.91 (0.30)	31.80 (0.54)	30.77 (1.56)	30.43 (1.62)	31.81 (0.25)
Black					
Pitches	87,170	26,054	3,377	3,179	
Called pitches	46,825	13,882	1,765	1,729	
Percent called strikes	31.93 (0.22)	30.87 (0.39)	30.76 (1.10)	30.19 (1.10)	31.62 (0.18)
TOTAL percent called strikes	32.05 (0.05)	31.45 (0.09)	30.62 (0.28)	31.84 (0.26)	31.87 (0.04)

*Standard errors of means in parentheses.

Table 3. Effects of the Relationship Between Pitcher and Umpire Race/Ethnicity (Dependent Variable Indicates Called Strike)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Umpires	**All**	**White**	**All**	**Black**	**All**	**Hispanic**	**All**	**All**	**All**
Pitchers	**White**	**All**	**Black**	**All**	**Hispanic**	**All**	**All**	**All**	**All**
Black umpire	-0.0025		0.0019		-0.0041				
	(0.0021)		(0.0110)		(0.0039)				
Hispanic umpire	-0.0040		0.0034		0.0076				
	(0.0029)		(0.0160)		(0.0053)				
Black Pitcher		-0.0187		-0.0196		-0.0124			
		(0.0029)		(0.0110)		(0.0160)			
Hispanic Pitcher		-0.0069		-0.0110		0.0040			
		(0.0011)		(0.0043)		(0.0060)			
Asian Pitcher		-0.0056		-0.0151		-0.0307			
		(0.0026)		(0.0110)		(0.0160)			
UPM							0.0034	0.0028	0.0027
							(0.0017)	(0.0023)	(0.0023)
Pitch Count (Balls/Strikes)									
0&1	-0.2270	-0.2230	-0.2140	-0.2180	-0.2140	-0.1990	-0.2240	-0.2240	-0.2240
	(0.0016)	(0.0014)	(0.0089)	(0.0059)	(0.0029)	(0.0081)	(0.0014)	(0.0014)	(0.0014)
0&2	-0.3540	-0.3490	-0.3450	-0.3340	-0.3440	-0.3500	-0.3510	-0.3510	-0.3530
	(0.0023)	(0.0020)	(0.0120)	(0.0082)	(0.0041)	(0.0110)	(0.0019)	(0.0019)	(0.0019)
1&0	-0.0282	-0.0274	-0.0324	-0.0256	-0.0173	-0.0385	-0.0258	-0.0255	-0.0245
	(0.0017)	(0.0015)	(0.0089)	(0.0059)	(0.0029)	(0.0082)	(0.0014)	(0.0014)	(0.0014)
1&1	-0.1920	-0.1910	-0.1990	-0.1860	-0.1860	-0.1810	-0.1900	-0.1900	-0.1890
	(0.0018)	(0.0016)	(0.0100)	(0.0066)	(0.0033)	(0.0092)	(0.0016)	(0.0016)	(0.0016)
1&2	-0.3200	-0.3250	-0.3140	-0.3080	-0.3150	-0.3220	-0.3250	-0.3250	-0.3240
	(0.0021)	(0.0018)	(0.0110)	(0.0074)	(0.0037)	(0.0100)	(0.0017)	(0.0017)	(0.0017)
2&0	0.0430	0.0407	0.0122	0.0498	0.0514	0.0303	0.0447	0.0452	0.0461
	(0.0026)	(0.0023)	(0.0130)	(0.0089)	(0.0045)	(0.0120)	(0.0022)	(0.0022)	(0.0022)
2&1	-0.1570	-0.1580	-0.1900	-0.1380	-0.1440	-0.1540	-0.1540	-0.1540	-0.1500
	(0.0026)	(0.0023)	(0.0140)	(0.0091)	(0.0045)	(0.0130)	(0.0022)	(0.0022)	(0.0022)
2&2	-0.2940	-0.2940	-0.2860	-0.2730	-0.2900	-0.2950	-0.2920	-0.2930	-0.2890
	(0.0024)	(0.0021)	(0.0130)	(0.0085)	(0.0042)	(0.0120)	(0.0020)	(0.0020)	(0.0020)
3&0	0.2060	0.1980	0.1520	0.2410	0.2130	0.1860	0.2060	0.2070	0.2110
	(0.0036)	(0.0032)	(0.0170)	(0.0120)	(0.0062)	(0.0170)	(0.0030)	(0.0030)	(0.0030)
3&1	-0.0644	-0.0669	-0.0376	-0.0379	-0.0567	-0.0726	-0.0611	-0.0605	-0.0586
	(0.0037)	(0.0032)	(0.0190)	(0.0130)	(0.0064)	(0.0180)	(0.0031)	(0.0031)	(0.0031)
3&2	-0.2620	-0.2610	-0.2560	-0.2350	-0.2500	-0.2680	-0.2580	-0.2570	-0.2520
	(0.0035)	(0.0030)	(0.0190)	(0.0120)	(0.0059)	(0.0170)	(0.0029)	(0.0029)	(0.0029)
Inning									
2	-0.0058	-0.0048	-0.0150	-0.0129	-0.0060	-0.0156	-0.0057	-0.0057	-0.0114
	(0.0020)	(0.0018)	(0.0110)	(0.0072)	(0.0037)	(0.0100)	(0.0017)	(0.0017)	(0.0018)
3	-0.0163	-0.0154	-0.0136	-0.0155	-0.0152	-0.0193	-0.0156	-0.0155	-0.0262
	(0.0020)	(0.0018)	(0.0110)	(0.0072)	(0.0037)	(0.0100)	(0.0017)	(0.0017)	(0.0017)
4	-0.0341	-0.0308	-0.0375	-0.0353	-0.0269	-0.0525	-0.0317	-0.0317	-0.0339
	(0.0020)	(0.0018)	(0.0110)	(0.0073)	(0.0037)	(0.0100)	(0.0017)	(0.0017)	(0.0017)
5	-0.0262	-0.0262	-0.0329	-0.0172	-0.0254	-0.0344	-0.0258	-0.0259	-0.0349
	(0.0020)	(0.0018)	(0.0110)	(0.0072)	-0.0038	(0.0100)	(0.0017)	(0.0017)	(0.0018)
6	-0.0332	-0.0318	-0.0351	-0.0305	-0.0308	-0.0548	-0.0329	-0.0330	-0.0361
	(0.0021)	(0.0018)	(0.0120)	(0.0073)	(0.0038)	(0.0100)	(0.0018)	(0.0018)	(0.0018)
7	-0.0256	-0.0237	-0.0189	-0.0184	-0.0232	-0.0391	-0.0249	-0.0251	-0.0294
	(0.0022)	(0.0018)	(0.0120)	(0.0072)	(0.0040)	(0.0100)	(0.0019)	(0.0019)	(0.0019)
8	-0.0254	-0.0216	-0.0387	-0.0153	-0.0202	-0.0382	-0.0245	-0.0249	-0.0284
	(0.0025)	(0.0018)	(0.0130)	(0.0072)	(0.0043)	(0.0100)	(0.0021)	(0.0021)	(0.0021)
9+	-0.0152	-0.0094	-0.0151	0.0042	-0.0172	-0.0305	-0.0157	-0.0163	-0.0203
	(0.0027)	(0.0019)	(0.0140)	(0.0075)	(0.0046)	(0.0100)	(0.0023)	(0.0023)	(0.0023)
Pitcher Score Advantage	0.0018	0.0024	0.0028	0.0017	0.0017	0.0009	0.0018	0.0018	0.0018
	(0.0002)	(0.0001)	(0.0009)	(0.0005)	(0.0003)	(0.0007)	(0.0001)	(0.0001)	(0.0001)
Top of Inning	0.0077	0.0066	0.0175	0.0070	0.0048	0.0073	0.0071	0.0071	0.0065
	(0.0010)	(0.0009)	(0.0054)	(0.0035)	(0.0018)	(0.0049)	(0.0009)	(0.0009)	(0.0008)
Observations	812745	1034379	27721	64201	258562	33565	1132145	1132145	1132145
R-squared	0.09	0.09	0.08	0.08	0.09	0.08	0.09	0.09	0.09
Fixed Effects	P	U	P	U	P	U	P	PU	PUB

*Standard errors in parentheses. All estimates are LPM with robust standard errors. P=pitcher, U=umpire, B=batter.

Table 4. Explicit Monitoring of Umpires and Racial Discrimination (LPM Estimates, Dependent Variable Is an Indicator of a Called Strike)*

Stadium **Pitchers**	**QuesTec** **All** ***(1)***	**Non-QuesTec** **All** ***(2)***	**All** **All** ***(3)***
Umpire-Pitcher Match (UPM)	-0.0035	0.0063	0.0064
	(0.0038)	(0.0029)	(0.0029)
QuesTec*UPM			-0.0098
			(0.0047)
N =	420,125	712,020	1,132,145
R^2	0.09	0.09	0.09

*The sample includes only pitches that were called by the umpire. All columns include fixed effects: 1) For each pitcher interacted with whether he pitched in a QuesTec ballpark, i.e., two fixed effects for each pitcher who pitched in both a ballpark where QuesTec was and was not installed; 2) For each umpire, i.e., umpire-QuesTec fixed effects, and 3) For each batter. Also included in the equations is the same set of control variables shown in Table 3, the indicators for inning, count, pitcher score advantage and the top of the inning.

Table 5. Implicit Monitoring of Umpires and Discrimination (Dependent Variable Indicated a Called Strike)*

Panel A. Distinguishing by Game Attendance

	High Attendance *(1)*	**Low Attendance** *(2)*	**All Games** *(3)*
Umpire-Pitcher-Match	-0.0028	0.0084	0.0052
(UPM)	(0.0034)	(0.0031)	(0.0025)
Well Attended			0.0062
(>69 percent capacity)			(0.0016)
Well Attended*UPM			-0.0050
			(0.0019)
N =	546,855	585,290	1,132,145
R^2	0.09	0.09	0.09

Panel B. Distinguishing by Terminal Count and Inning

	Terminal *(4)*	**Non-Terminal** *(5)*	**All Pitches** *(6)*	**Early Inning** *(7)*	**Late Inning** *(8)*
UPM	-0.0028	0.0046	0.0042	0.0049	0.0036
	(0.0037)	(0.0024)	(0.0024)	(0.0042)	(0.0029)
Terminal Count			-0.0073	-0.0098	-0.0056
*UPM			(0.0018)	(0.0031)	(0.0022)
N =	261,670	870,475	1,132,145	396,438	735,707
R^2	0.17	0.04	0.09	0.09	0.08

*All columns are LPM estimates with pitcher, umpire, and batter fixed effects. Included (but not shown) are controls for inning, pitch count, pitcher score advantage, and top of the inning.

Table 6. Effect of Umpire and Starting Pitcher Race/Ethnicity on Performance*

	(1)	*(2)*	*(3)*
Dependent Variable	Win	*GameScore*™	Runs Allowed
UPM	0.0424	1.0238	-0.1635
	(0.0236)	(0.8219)	(0.1026)
Pitcher's run support	0.1345	-0.0728	0.0029
	(0.0033)	(0.0454)	(0.0056)
N =	6979	14,229	14,229

*Each equation contains controls for the home team's runs scored, as well as fixed effects for both the home and away starting pitchers, the opposing team, and the home plate umpire. The dependent variable in Column (1) is an indicator of a win by the home team, in Column (2) it is *GameScore*™, a composite metric calculated only for staring pitchers. The dependent variable in Column (3) is the number of runs allowed by each starting pitcher.

Table 7. Batter-Umpire and Catcher-Umpire Race/Ethnicity Matches, N=1,132,145 (Dependent Variable Indicates Called Strike)*

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Umpire Pitcher Match			0.0027	0.0064	0.0054	0.0044	0.0107
(UPM)			(0.0023)	(0.0029)	(0.0025)	(0.0024)	(0.0031)
Umpire Batter Match	0.0000		-0.0001	0.0005	-0.0001	-0.0008	-0.0003
(UBM)	(0.0018)		(0.0018)	(0.0023)	(0.0020)	(0.0019)	(0.0025)
Umpire Catcher Match		0.0008	0.0008	0.0002	-0.0008	0.0013	-0.0009
(UCM)		(0.0010)	(0.0010)	(0.0013)	(0.0014)	(0.0011)	(0.0016)
Questec*UPM				-0.0099			-0.0097
				(0.0048)			(0.0048)
Questec*UBM				-0.0014			-0.0014
				(0.0038)			(0.0038)
Questec*UCM				0.0018			0.0017
				(0.0022)			(0.0022)
High attendance					0.0046		0.0046
					(0.0020)		(0.0020)
High attendance*UPM					-0.0056		-0.0056
					(0.0019)		(0.0019)
High attendance*UBM					0.0001		0.0001
					(0.0018)		(0.0018)
High attendance*UCM					0.0032		0.0032
					(0.0018)		(0.0018)
Terminal count						-0.145	-0.1450
						(0.0330)	(0.0330)
Terminal count*UPM						-0.0074	-0.0074
						(0.0021)	(0.0021)
Terminal count*UBM						0.0033	0.0033
						(0.0020)	(0.0020)
Terminal count*UCM						-0.0021	-0.0021
						(0.0020)	(0.0020)
R^2	0.09	0.09	0.09	0.09	0.09	0.09	0.09

* All estimates are LPM and include controls for inning, score, top of the inning, as well as fixed effects for each pitcher-QuesTec, umpire-QuesTec, and batter-QuesTec combination.

Figure 1. Race and Called Strike Percentage in QuesTec and Non-QuesTec Ballparks

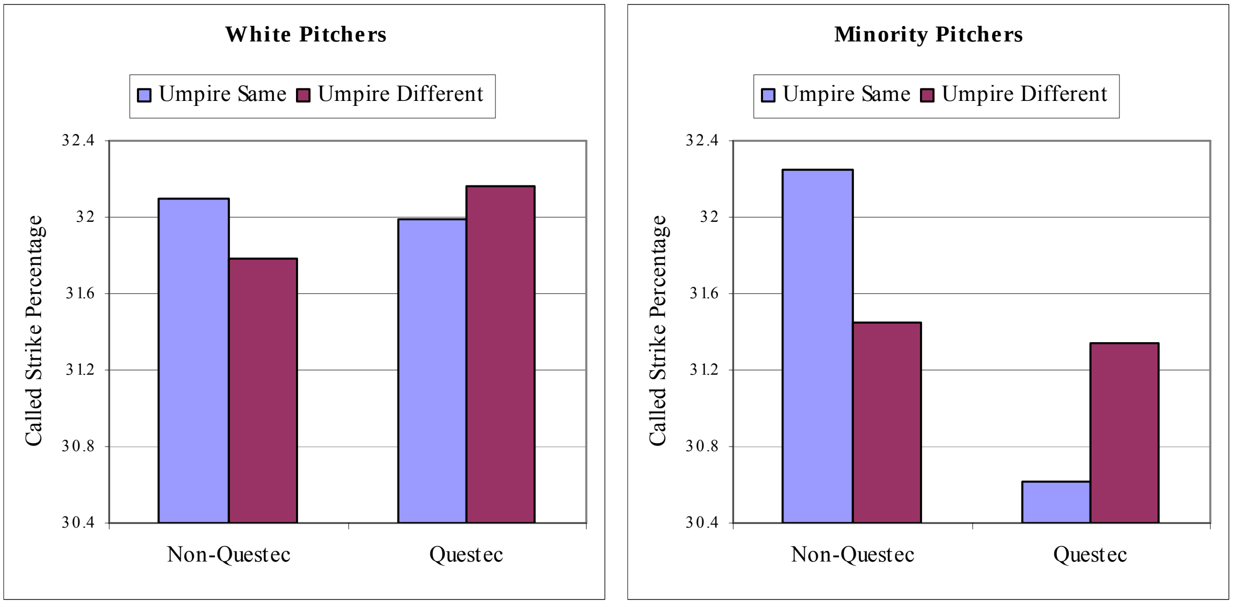


Figure 2. Race and Called Strike Percentage by Game Attendance

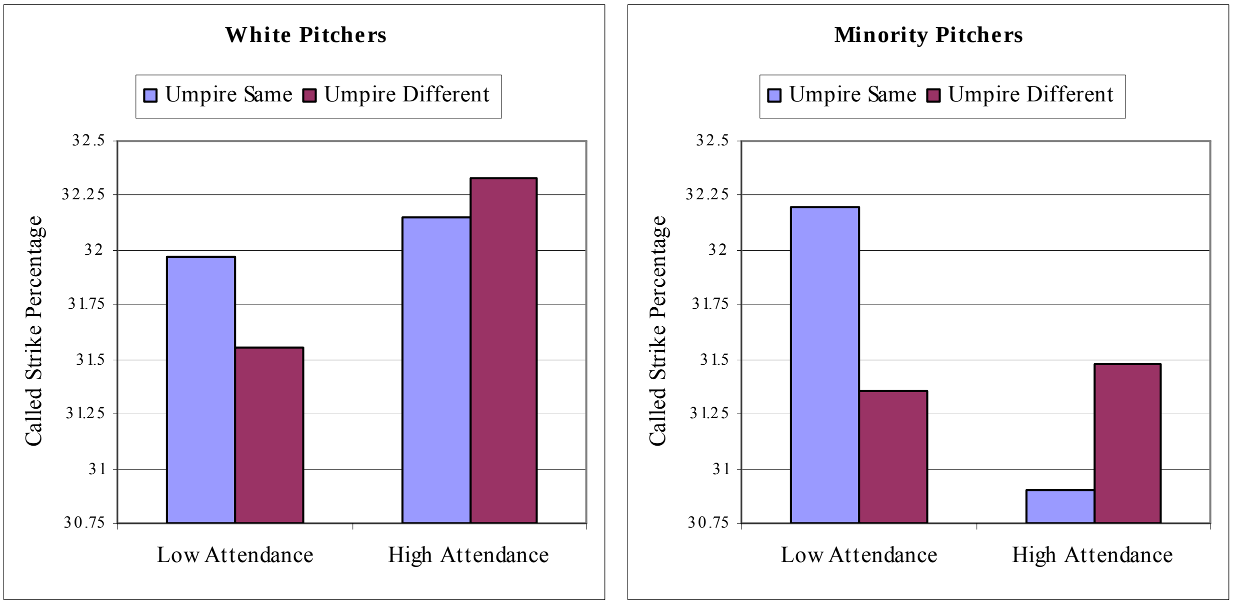


Figure 3. Race and Called-Strike Percentage in Terminal and Non-Terminal Counts

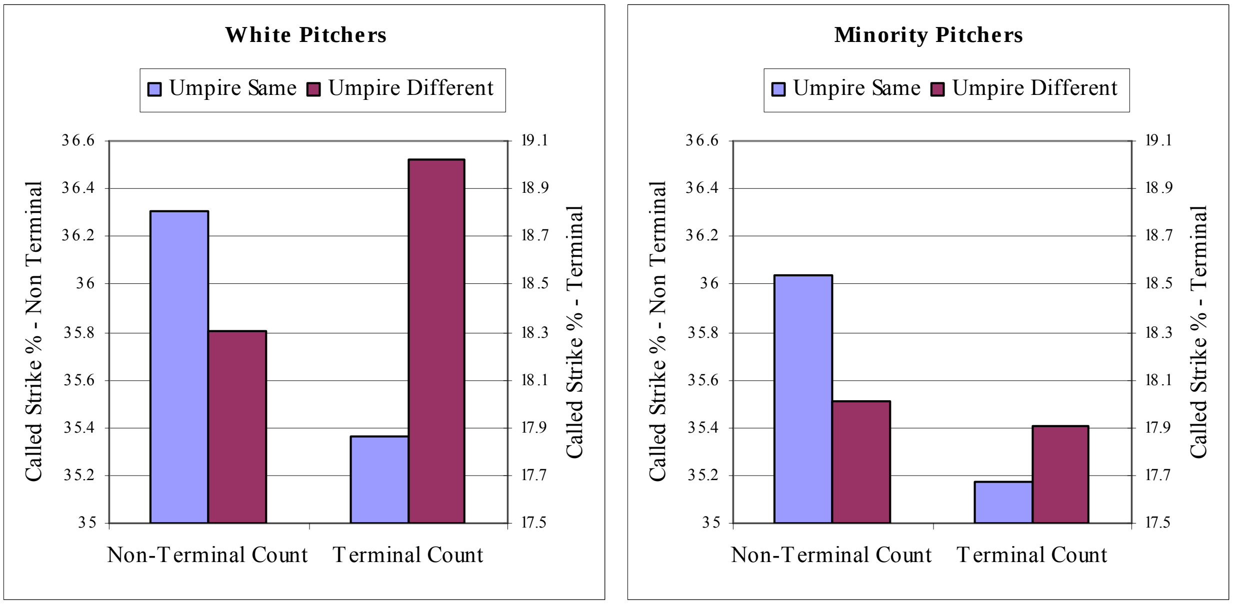
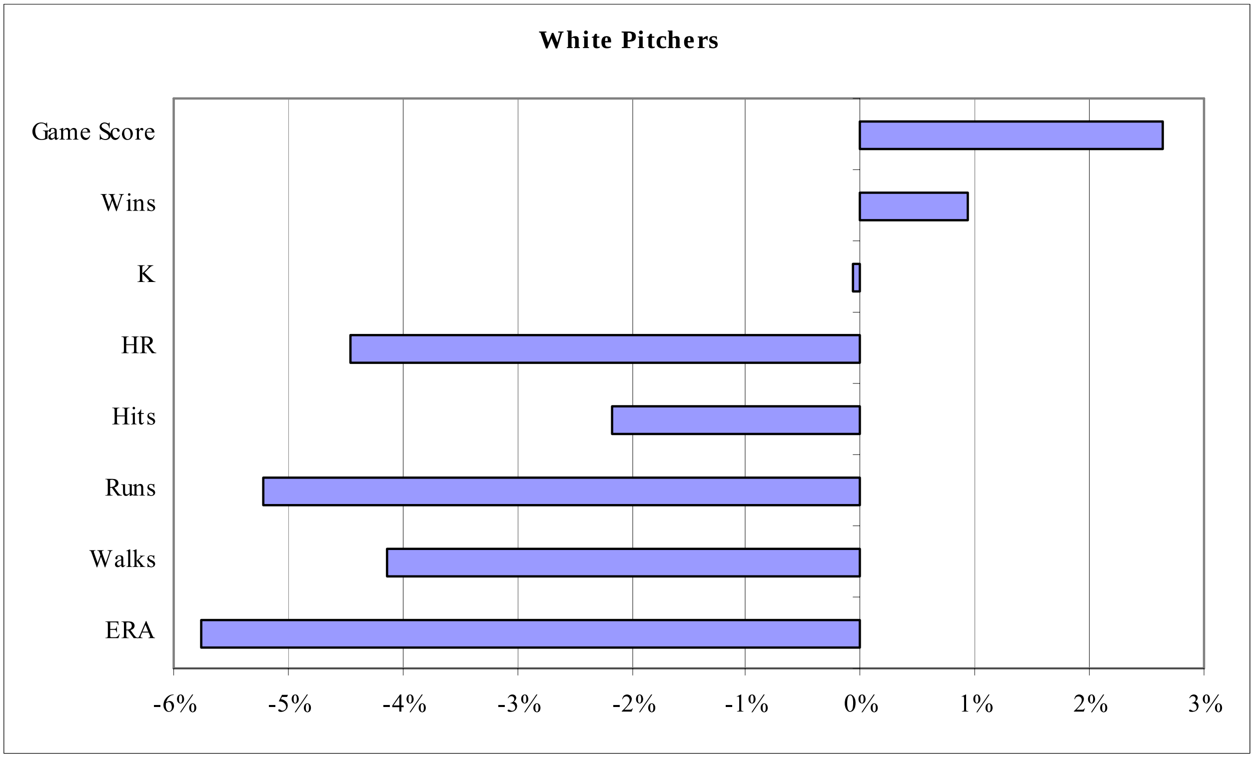
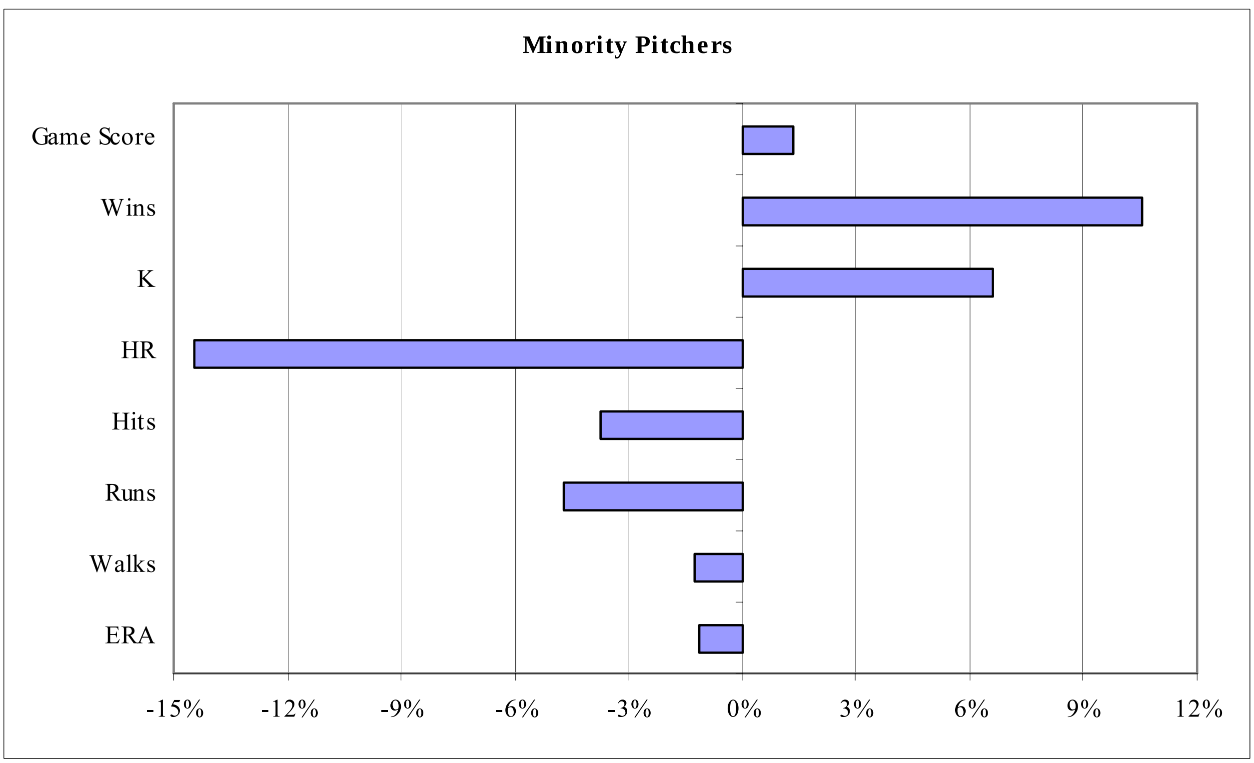


Figure 4: Change in Pitcher Performance When Umpire Matches Race/Ethnicity





*Baseline is mismatch of race/ethnicity of umpire and pitcher.