COMMENTS WELCOME

STRIKE THREE: UMPIRES' DEMAND FOR DISCRIMINATION

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

This Version: August 7, 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. We thank Jay Hartzell, William Mayew, Paul Tetlock, Sheridan Titman, Deborah White and Justin Wolfers for helpful discussions.

Abstract

We explore umpires' racial/ethnic discrimination in the evaluation of Major League Baseball players. Controlling for umpire, pitcher, and batter fixed effects and other factors, strikes are more likely to be called if the umpire and pitcher match race/ethnicity. This effect exists where there is little scrutiny of umpires' behavior—in ballparks without computerized systems monitoring umpires' calls, at poorly attended games, and when the next 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 earned runs per game and improves his team's chance of winning. The results suggest that attempts to measure salary discrimination generally may be flawed, since the productivity measures can themselves be contaminated by the effects of racial preferences.

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 wage to measured skill 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. Umpires subjectively judge the performance of a pitcher many times within each game, deciding whether pitches are "strikes" or "balls," with the former benefiting the pitcher. Because the pitcher's productivity and performance are so heavily influenced by the umpire's evaluation, discrimination by umpires could conceivably affect both games' outcomes and the labor market, i.e., pitcher compensation and market value.

We compile a rich dataset including individual pitches to explore racial/ethnic discrimination by umpires in the evaluation of baseball players. We collect and analyze every pitch from three complete seasons (2004-2006), paying particular attention to the race/ethnicity of the umpire, pitcher, and batter. Our results indicate that umpires give favorable treatment to pitchers who share their race/ethnicity, as indicated by the probability that a pitch is called a strike rather than a ball. This effect is robust to a wide set of controls, including fixed effects for each pitcher, umpire, and batter, suggesting that differences in umpire or player-specific characteristics are not driving the results.

Our data are particularly well suited to study racial discrimination. First, pitch calls in baseball games are characterized by their high frequency, importance, and subjectivity. Since every pitch is potentially subject to the home-plate umpire's discretion when it is thrown (several hundred times per game), there is both sufficient scope for racial/ethnic discrimination to be expressed as well as for it to affect game outcomes significantly. Second, we have a very large number of independent pitch-level observations involving the interaction of four different race/ethnicities: White, Black, Hispanic, and Asian. Three of these are represented among umpires (there are no Asian umpires in MLB), and all four are represented among players. 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 race/ethnicities other than that of the umpire, 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 preferences in employment generally. Our study most closely resembles Price and Wolfers' (2007) work on NBA officiating crews' racial/ethnic preferences. Our results not only

[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 it 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, but they do so in a particular case where judgments are much less likely to be subjective than in basketball, where (as Price and Wolfers acknowledge) what appears to be racial discrimination might in reality be two racially-defined groups playing a common game under different rules.

This research adds to a large literature on racial discrimination in sports, specifically in baseball. The literature goes 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); and 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, since measured productivity is typically the crucial control variable. We suggest questioning this central assumption, since if officials are themselves subject to racial/ethnic bias, it is difficult to draw meaningful conclusions about discrimination in other areas based on alleged racial differences in the returns to measured productivity.

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, 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 evidence 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 only in times of low-scrutiny, game- and pitch-level situations where monitoring of the umpires 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.

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 on average roughly 150 pitches, so that approximately 730,000 pitches are thrown each season. We collected 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, pitch count, score, inning, and pitch outcome. We classify each pitch into one of seven 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 and the number of allowed hits, walks, strikeouts, homeruns, runs and earned runs. Finally, for each starting pitcher in each game we collect the *game score,* a composite index designed to summarize a starting pitcher's performance.[3]

We next classify each 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 and umpire. Players or umpires are classified as Hispanic if they are born in one of the following

[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, *Game Score* 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.

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]

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 subdivide 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. Our focus for the remainder of the analysis will be on the remaining 53 percent of pitches that result in called strikes and called balls, since it is only these that engender 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 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

[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 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 reduced the strength of any results that we generate.

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

underrepresented among umpires, representing 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, and there are no Asian umpires in our sample.

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 (usually 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 Black umpires judge Hispanic pitchers harshly, relative to how they are judged by White and Hispanic umpires; but 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.

III. Called Pitches and Umpire-Pitcher Matches

The summary statistics in Table 2 ignore, among other effects, the possible different outcomes generated by non-random assignment of pitchers to face different opponents, and of umpires to games played by particular teams. To account for these and other potential difficulties, our most basic test for umpire discrimination is the specification:

(1) $I(\text{Strike} \mid \text{Called Pitch})_i = \alpha + \beta{*}UPM_i + \gamma{*}Controls_i + \varepsilon_i$,

where the dependent variable is an indicator of whether a called pitch is a strike, α, β and 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 most of our tests, we include as control variables: 1) Pitch-count indicators, which record how many balls and strikes have accrued each time a pitcher faces a hitter. Because pitchers alter the location of their pitch based on the ball-strike count, it is important to control for this measure when each pitch is thrown; 2) Inning indicators, which are 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;[8] and 3) Top-of-the- inning indicators**,** which account for whether the home (top=1) or visiting team (top=0) is pitching, thus taking any home-field bias into consideration; and 4) Pitcher's score advantage, 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.[9]

Table 3 presents the results of estimating (1). The first three panels show the analysis for each possible pitcher-umpire race/ethnicity combination rather than using UPM, i.e., they estimate separate equations for each pitcher race/ethnicity and include individual indicators for umpire race/ethnicity. In the final panel we aggregate all pitchers, so that the indicator of interest becomes UPM as in (1). In each panel Columns (a) and (b) present estimates of the marginal effects from probit models, while the equations in Columns (c) are estimated as linear probability models (LPM) with pitcher fixed-effects and robust standard errors.[10] In the first three panels the omitted category for umpire race/ethnicity is White.

Consider first the results from analyzing each pitcher race/ethnicity separately. While none of the coefficients on the umpire race/ethnicity indicators is statistically significant, some suggestive patterns emerge from the first three panels. In Columns 1-3(a), both estimates of the

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

[9]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).

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

coefficients on the Black and Hispanic umpire indicators are negative, suggesting that pitches thrown by White pitchers are less likely to be called strikes by non-White umpires. This pattern is also seen among Hispanic pitchers, for whom results are in Columns 1-3(c). Relative to White and Black umpires, Hispanic umpires appear more likely to call a strike on a pitch thrown by a Hispanic. Only among Black pitchers does this ordering not hold perfectly, although pitches thrown by Black pitchers comprise less than 2.5 percent of the sample.

The final panel considers all pitchers, including the cohort of Asian pitchers.[11] The results from the LPM model with pitcher fixed effects indicate that a given called pitch is approximately 0.34 percentage points more likely to be called a strike if the umpire and pitcher match race/ethnicity, a statistically significant result. Excluding (as we do) pitches where the batter swings, the likelihood that a given pitch is called a strike is 31.8 percent. Thus when the umpire matches the pitcher's race/ethnicity the base rate of called strikes rises by slightly more than 1 percent compared to the result if there is no match.[12]

At first, this effect may seem trivial, affecting on average less than one pitch per game. We later explore this question in more depth, but it is worth mentioning that an umpire's racial bias may affect the game either directly or indirectly. It is obvious that the direct effect of racial bias on pitch calls, such as the potential for a pitcher facing 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. For example, a Black pitcher facing a Hispanic umpire may be awarded fewer close calls at the edges of the strike zone,

---

[11]We include all pitchers in these equations, 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]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.

potentially causing him to throw pitches that are easier for the batter to hit. Because baseball is a relatively low-scoring game, one or two such instances in a game can easily alter the pitcher's performance and the outcome of the game.

IV. Called Pitches When Discrimination Is Costly to the Discriminator

Studies of cognitive biases, such as an umpire's preference for pitchers of the same race/ethnicity, indicate that its expression can be reduced or eliminated. For example, exposing the biased party to counter-examples of the stereotype of interest can reduce the severity and/or frequency of the biased behavior (Goodwin *et al*, 2000; Blair, 2002). Another mitigating mechanism is to increase the visibility of the biased party's behavior, effectively increasing the price of expressing the bias. In this section we employ three different measures of scrutiny of the umpire's evaluations of pitchers to examine whether a higher price of discrimination reduces the extent to which umpires indulge 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.[13] 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. If more than 10 percent of an umpire's calls differ from QuesTec's records, his performance is considered substandard, and that

[13]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.

may influence his promotion to “crew chief,” assignment to post-season games, or even retention in the major leagues.[14]

Because QuesTec is installed in roughly 35 percent of ballparks, and because umpiring crews are rotated around the league’s ballparks, virtually every umpire in our dataset calls a substantial number of pitches in parks both with and without QuesTec. Additionally, that both the umpires’ and teams’ schedules change every year exposes 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. 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.[15] For this reason, in all of our QuesTec regressions, we include fixed effects not only for each pitcher, but also for the presence or absence of QuesTec in each game, i.e., pitcher-QuesTec fixed effects.

Figure 1 graphs the average percentage of pitches that are called strikes in ballparks with and without QuesTec for White and non-White pitchers respectively. The effect of monitoring on umpire behavior is apparent, with both White and non-White pitchers being judged differently by

[14]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 a pitcher meets the QuesTec standards.

[15]Glavine reports (Associated Press, July 9, 2003) 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.”

umpires of the matched race/ethnicity depending on whether the pitch is thrown in a park with QuesTec installed.

Table 4 contains the results of estimating (1) separately by the presence of QuesTec in the ballpark. All of the estimated equations presented in Table 4 include controls for inning, pitch count, pitcher score advantage, and top of the inning. 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. The results are remarkable: In ballparks with the umpire monitoring system (Column 1), the coefficient on UPM is -0.21 percentage points and is not significantly different from zero. In parks without QuesTec (Column 2) the same coefficient is 0.66 percentage points per pitch, significantly different from zero at the 1 percent level.

Columns 3 and 4 present the results when QuesTec is interacted with UPM, first without umpire-QuesTec and batter fixed effects and then with these additional controls. Controls for umpires and batters only strengthen the result, so that when the pitcher and umpire match race/ethnicity, being in a QuesTec park reduces the likelihood that a called pitch is ruled a strike by almost 1 percentage point. Columns 5 and 6 reveal that this effect is nearly twice as strong for minority as for White pitchers. As in the first three sets of results in Table 3, however, separating pitchers by race/ethnicity (or even by minority vs. non-minority) has a large impact on statistical significance, impairing our ability to infer much about the relative sizes of the effects of UPM across pitcher groups.

To extend these results, we employ two additional measures that may influence the scrutiny of umpires. 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.[16] 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 umpire behavior dramatically. Compared to well-attended games, umpires calling poorly-attended games are more charitable to 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 matches non-White pitchers call nearly 1.5 percent more strikes in poorly-attended games, whereas unmatched umpires call fewer strikes.

We add controls in Panel A of Table 5 and show the results of estimating (1) separately for both well- and poorly-attended games in Columns 1 and 2 respectively. As with the QuesTec results, the UPM variable is significant at the 1 percent level only in poorly-attended games, with an effect of 0.72 percentage points per pitch. During well-attended games there is no significant effect of an umpire-pitcher racial/ethnic match. Column 3 generalizes the results by aggregating all games, with UPM interacted with an indicator of a well-attended game. We then add umpire and batter fixed effects in Column 4, with a negligible impact on the interaction term. Compared to a well-attended game, a pitch called by an umpire of the same race/ethnicity as the pitcher is 0.44 percentage points more likely to be judged a strike if the game is poorly attended.

[16]There is, of course, 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.

Our final 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 judgment of it can terminate the batter's plate appearance. Thus, for example, a pitch that is thrown with two strikes is potentially terminal, and is likely to be scrutinized more heavily by the catcher, pitcher, batter, managers and fans. The same is true for a count with three balls.

Panel B of Table 5 shows estimates of (1) separately for terminal and non-terminal pitches. In Columns 5 and 6 each type of pitch is considered separately, with the result that the coefficients of UPM have opposite signs. For potentially terminal pitches—where scrutiny of the umpire is likely to be greatest—umpires appear to judge pitchers of their own race/ethnicity more harshly than unmatched pitchers. This result is reversed for pitches of lesser importance, where the estimated coefficient of UPM is 0.62 percentage points, a result significant at the 1 percent level. In Column 7 all pitches are aggregated and UPM is interacted with an indicator for potentially terminal pitches, and umpire and batter-level fixed effects are added in Column 8. With the full complement of control variables, the impact of UPM is a statistically significant 0.73 percentage points.

In Columns 9 and 10 we employ one final proxy for the scrutiny of the umpire—whether the pitch is thrown early in the game or not. 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 "not early." We expect that a pitcher-umpire racial/ethnic match will have a stronger effect in early innings. Comparing the results of Columns 9 and 10, 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).

The results in Tables 4 and 5 strongly suggest a role for racial/ethnic preferences in the evaluation of MLB pitchers. The correlation of the between-game proxies for umpire scrutiny—

QuesTec and attendance percentage—is below 0.05, suggesting that these are independent proxies of scrutiny. Because the type of pitch (terminal or non-terminal) is a within-game measure, it is uncorrelated with either between-game measure. When QuesTec, attendance and terminal counts are all included in (1) (in estimates that are unreported here), the interaction of each indicator with UPM is still statistically significant and negative: Each separate proxy for the scrutiny of umpires reduces measured discrimination against unmatched pitchers. Implicitly each proxy for a higher price of discrimination reduces umpires' demand for discriminatory outcomes.

One might argue that these effects do not reflect the impact of an increased price of umpire discrimination, but instead represent greater care expended by umpires when their decisions are more important. That explanation is consistent with the results on terminal counts, but it does nothing to explain the results on the interactions with QuesTec and attendance. Moreover, why should lack of care generate apparent discrimination rather than simply noisier calls?

V. Impacts on Games' Outcomes

We have focused our analysis on umpires' calls of individual pitches. The potential for racial/ethnic preferences to affect a game directly through called pitches alone is questionable, since, except for bases-loaded walks, in most cases a called pitch does not itself generate offense. If, however, pitchers, hitters or managers alter their strategies because of the umpire's behavior, the potential for racial bias to affect the game is greatly expanded. For example, a pitcher receiving favorable calls from an umpire is afforded the luxury of pitching more aggressively, which could easily alter his and his team's fortunes.

We examine a variety of game-level performance measures for each starting pitcher in our sample: Wins, hits, earned runs, home runs, strikeouts, walks, and game score.[17] Figure 4

[17]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, "game score" is only calculated for starting pitchers.

shows tabulations of each performance measure for the roughly 14,000 pitched games 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, Game Score, 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 give up 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 tiny. 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.

The statistical power of these tests is much lower than in the pitch-level regressions. Consequently, although the vast majority of the pitcher performance measures differ in the directions that we would predict if there are racial/ethnic preferences, most differences are not statistically significant. In aggregate, however, their combined effect can have a significant impact on game outcomes. We analyze the outcomes of 7,124 games from 2004-2006, accounting for approximately 98 percent of all games played. 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 the outcome.

The obvious benchmark is the case when both starting pitchers 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 percent of the time. In the former case, the home team wins

55.6 percent of its games, an improvement of 1.8 percentage points over the benchmark. 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.

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 pitcher and umpire are Hispanic, the home team wins 61.1 percent of its games, compared to 52.0 percent if the pitcher is non-Hispanic.

In Table 6 we present estimates of equations with the dependent variable equaling one if the home team wins. In Column 1 the results of a probit with no controls show a positive but insignificant coefficient on the UPM indicator. Interpreting this coefficient in isolation is difficult: A home-team pitcher receiving favorable treatment from an umpire should give up fewer runs and thus be more likely to win the game conditional on the number of runs his team scores. For this reason, we include the number of runs scored by the pitcher's team in Column 2, where the coefficient on UPM becomes marginally significant with a magnitude of slightly over 3.1 percentage points. When fixed effects are included for both the home and the visiting pitcher (Column 3), the coefficient on UPM increases to 3.4 percent. Adding umpire fixed effects in Column 4 strengthens the results further, indicating a marginal advantage of over 4.2 percentage points for a home team if its starting pitcher matches the umpire's race/ethnicity.

VI. Robustness Checks and Other Considerations

A. Accounting for Batters' 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. We explore this possibility extensively, but find no evidence to support the argument. Estimating (1) substituting UBM, defined as a racial/ethnic match between umpire and batter, for UPM generally yields insignificant results, although the magnitudes are qualitatively consistent with the estimates incorporating UPM. Somewhat surprisingly, the results do not improve if both UBM and UPM are included along with their interaction. Because the benchmark in this case corresponds to the situation when the umpire matches neither the batter nor the pitcher, the coefficient on UBM pins down the marginal effect of changing only the batter's race/ethnicity to match the umpire's. Even in this extreme case, the match between the umpire and batter does not appear to influence the umpire's behavior.

For at least two reasons this finding 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.

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 collected demographic information about 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 obtained 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 (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 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, such 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

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 terminal pitches, when this effect would be most likely to prevail, suggests the argument is invalid.

Our estimates 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 they are implicitly both aware of these preferences and able to act upon them. The racial/ethnic endowments of umpires and starting pitchers in the 7124 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. Conclusions

The analysis of individual pitches and game outcomes suggests 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 becomes significantly stronger when umpire behavior is less well monitored. The evidence also suggests that this bias is strong enough to affect 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.

Several measures enter into the analysis of a pitcher's productivity, but for simplicity consider only victories. Suppose that a non-White starter pitches 32 games per season, 1/5 of his team's games. In a typical season, he can expect to be evaluated by a non-White umpire in at most two games. Let us further assume that half of these games are at home, so that the 4 percent marginal effect from Table 6 is appropriate. Multiplying the number of home games pitched in the presence of a mismatched umpire (14) by the coefficient on UPM (4.2 percentage points), yields approximately 0.6 home losses per season. In 2007 the highest-paid pitcher earned $16 million, and nobody won more than 19 games in 2006. A journeyman starting pitcher playing a full season (starting at least 30 games) will almost surely win 9 games and earn perhaps $1 million. Looking only at average effects, the short-run annual return to a win is about $1.5 million. Multiplying this figure by 0.6 fewer wins, the impact of racial/ethnic discrimination on a non-White starting pitcher may be as high as $1 million per season. Bradbury (2007, p. 193) suggests that the impact on the pitcher's marginal revenue product is roughly the same.

This type of discrimination is particularly 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, then 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 estimates of equations relating earnings to race/ethnicity, even with a large set of variables designed to control for inherent differences in productivity. To the extent that supervisor 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 more likely to be evaluated, 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.[18] Clearly, raising the price of discrimination in the labor market generally through analogous methods is more difficult; but these results from MLB may suggest measures that might have the desired effects.

[18]Whether the installation of a new strike-zone evaluation tool (ZE) in all baseball parks, projected during 2007, will create the same incentives as QuesTec and vitiate 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.

Luis Garicano, Ignacio Palacios-Huerta and Canice Prendergast, "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.

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

*The number of players or umpires in each racial/ethnic group is in parentheses.

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	31.47	30.61	31.97	31.89
Hispanic					
Pitches	45,603	13,737	1,552	1,406	
Called pitches	24,592	7,323	845	805	
Percent called strikes	31.91	31.80	30.77	30.43	31.81
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	30.87	30.76	30.19	31.62
TOTAL percent called strikes	32.05	31.45	30.62	31.84	31.87

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

	White Pitchers			Black Pitchers			Hispanic Pitchers			All Pitchers		
	Probit *(1a)*	**Probit** *(2a)*	**LPM** *(3a)*	**Probit** *(1b)*	**Probit** *(2b)*	**LPM** *(3b)*	**Probit** *(1c)*	**Probit** *(2c)*	**LPM** *(3c)*	**Probit** *(1d)*	**Probit** *(2d)*	**LPM** *(3d)*
Black umpire	-0.0013	-0.0025	-0.00247	0.00157	-0.0018	0.00185	-0.0061	-0.0066	-0.0041			
	(0.0022)	(0.0022)	(0.0021)	(0.0110)	(0.0110)	(0.0110)	(0.0040)	(0.0041)	(0.0039)			
Hispanic umpire	-0.0016	-0.0035	-0.00398	0.00161	0.00375	0.00337	0.00365	0.00737	0.00761			
	(0.0030)	(0.0030)	(0.0029)	(0.0160)	(0.0160)	(0.0160)	(0.0055)	(0.0056)	(0.0053)			
UPM										0.00555	**0.00730**	**0.00341**
										(0.0009)	**(0.0009)**	**(0.0017)**
Pitch count (Balls/strikes)												
0&1		-0.192	-0.227		-0.179	-0.214		-0.182	-0.214		-0.189	-0.224
		(0.0012)	(0.0016)		(0.0065)	(0.0089)		(0.0022)	(0.0029)		(0.0010)	(0.0014)
0&2		-0.280	-0.354		-0.270	-0.345		-0.273	-0.344		-0.278	-0.351
		(0.0010)	(0.0023)		(0.0053)	(0.0120)		(0.0018)	(0.0041)		(0.0009)	(0.0019)
1&0		-0.027	-0.028		-0.031	-0.032		-0.017	-0.018		-0.025	-0.026
		(0.0016)	(0.0017)		(0.0084)	(0.0089)		(0.0028)	(0.0029)		(0.0014)	(0.0014)
1&1		-0.163	-0.192		-0.166	-0.199		-0.157	-0.186		-0.161	-0.190
		(0.0014)	(0.0018)		(0.0074)	(0.0100)		(0.0025)	(0.0033)		(0.0012)	(0.0016)
1&2		-0.266	-0.329		-0.250	-0.314		-0.256	-0.315		-0.263	-0.325
		(0.0011)	(0.0021)		(0.0058)	(0.0110)		(0.0019)	(0.0037)		(0.0009)	(0.0018)
2&0		0.035	0.043		0.005	0.012		0.042	0.051		0.036	0.045
		(0.0026)	(0.0026)		(0.0130)	(0.0130)		(0.0046)	(0.0045)		(0.0022)	(0.0022)
2&1		-0.133	-0.157		-0.157	-0.190		-0.123	-0.144		-0.131	-0.154
		(0.0020)	(0.0026)		(0.0098)	(0.0140)		(0.0036)	(0.0045)		(0.0017)	(0.0022)
2&2		-0.236	-0.294		-0.227	-0.286		-0.234	-0.290		-0.235	-0.292
		(0.0014)	(0.0024)		(0.0072)	(0.0130)		(0.0023)	(0.0042)		(0.0011)	(0.0020)
3&0		0.193	0.206		0.132	0.152		0.197	0.212		0.193	0.206
		(0.0041)	(0.0036)		(0.0190)	(0.0170)		(0.0071)	(0.0062)		(0.0034)	(0.0030)
3&1		-0.060	-0.064		-0.039	-0.038		-0.054	-0.058		-0.057	-0.061
		(0.0033)	(0.0037)		(0.0180)	(0.0190)		(0.0058)	(0.0064)		(0.0028)	(0.0031)
3&2		-0.210	-0.260		-0.206	-0.256		-0.203	-0.251		-0.208	-0.258
		(0.0021)	(0.0035)		(0.0100)	(0.0190)		(0.0035)	(0.0059)		(0.0017)	(0.0029)
Inning												
2		-0.005	-0.006		-0.015	-0.015		-0.005	-0.006		-0.005	-0.006
		(0.0021)	(0.0020)		(0.0120)	(0.0110)		(0.0038)	(0.0037)		(0.0018)	(0.0017)
3		-0.016	-0.016		-0.010	-0.014		-0.014	-0.015		-0.015	-0.016
		(0.0021)	(0.0020)		(0.0120)	(0.0110)		(0.0038)	(0.0037)		(0.0018)	(0.0017)
4		-0.034	-0.034		-0.034	-0.038		-0.027	-0.027		-0.032	-0.032
		(0.0021)	(0.0020)		(0.0110)	(0.0110)		(0.0038)	(0.0037)		(0.0018)	(0.0017)
5		-0.026	-0.026		-0.031	-0.033		-0.024	-0.025		-0.026	-0.026
		(0.0021)	(0.0020)		(0.0110)	(0.0110)		(0.0038)	(0.0038)		(0.0018)	(0.0017)
6		-0.033	-0.033		-0.038	-0.035		-0.030	-0.031		-0.032	-0.033
		(0.0021)	(0.0021)		(0.0110)	(0.0120)		(0.0038)	(0.0038)		(0.0018)	(0.0018)
7		-0.025	-0.026		-0.014	-0.019		-0.023	-0.023		-0.024	-0.025
		(0.0021)	(0.0022)		(0.0120)	(0.0120)		(0.0037)	(0.0040)		(0.0018)	(0.0019)
8		-0.022	-0.025		-0.022	-0.039		-0.021	-0.020		-0.022	-0.025
		(0.0021)	(0.0025)		(0.0110)	(0.0130)		(0.0036)	(0.0043)		(0.0018)	(0.0021)
9+		-0.007	-0.015		0.003	-0.015		-0.017	-0.017		-0.009	-0.016
		(0.0022)	(0.0027)		(0.0120)	(0.0140)		(0.0037)	(0.0046)		(0.0018)	(0.0023)
Pitcher Score Advantage		0.002	0.002		0.004	0.003		0.002	0.002		0.002	0.002
		(0.0002)	(0.0002)		(0.0009)	(0.0009)		(0.0003)	(0.0003)		(0.0001)	(0.0001)
Top of Inning		0.008	0.008		0.016	0.018		0.005	0.005		0.007	0.007
		(0.0011)	(0.0010)		(0.0056)	(0.0054)		(0.0019)	(0.0018)		(0.0009)	(0.0009)
N =	812729	812729	812729	27721	27721	27721	258578	258578	258578	1132145	1132145	1132145
R^2 or pseudo- R^2	.000	.075	0.09	0.000	0.072	0.08	0.000	0.073	0.09	0.000	0.074	0.09
Number of pitcher fixed effects			670			27			221			944

*Standard errors in parentheses here and in Tables 4-6. In the probits the estimates are of the derivatives of the probit function.

Table 4: Explicit Monitoring of Umpires and Racial/Ethnic Discrimination (LPM Estimates, Dependent Variable Indicates a Called Strike)*

Stadium Pitchers	QuesTec All *(1)*	Non-QuesTec All *(2)*	All All *(3)*	All All *(4)*	All White *(5)*	All Minority *(6)*
Umpire-Pitcher Match (UPM)	-0.0021	0.00663	0.00663	0.00617	0.00628	0.0168
	(0.0027)	(0.0021)	(0.0021)	(0.0029)	(0.0096)	(0.0093)
QuesTec*UPM			-0.00874	-0.00961	-0.0136	-0.0290
			(0.0035)	(0.0047)	(0.0150)	(0.0160)
Pitcher Fixed Effects	Yes	Yes				
Pitcher-QuesTec Fixed Effects			Yes	Yes	Yes	Yes
Umpire-QuesTec Fixed Effects				Yes	Yes	Yes
Batter Fixed Effects				Yes	Yes	Yes
Number of Fixed Effects	879	918	1797	3395	2867	2126
N =	420125	712020	1132145	1132145	812729	319416
R^2	0.09	0.09	0.09	0.09	0.08	0.09

*The sample includes only pitches that were called by the umpire. All columns include fixed effects 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. Also included in the equations are the indicators for inning, count, pitcher score advantage, and the top of the inning.

Table 5: Implicit Monitoring of Umpires and Discrimination (LPM Estimates, Dependent Variable Indicates a Called Strike)

Panel A. Distinguishing by Games' Attendance

	High Attendance *(1)*	**Low Attendance** *(2)*	**All Games** *(3)*	**All Games** *(4)*
Umpire-Pitcher-Match	-0.00024	0.00753	0.00556	0.00479
(UPM)	(0.0024)	(0.0023)	(0.0019)	(0.0025)
High Attendance			0.00203	0.00194
(>69 percent capacity)			(0.0015)	(0.0015)
High Attendance			-0.00448	-0.00442
* UPM			(0.0019)	(0.0019)
Pitcher Fixed Effects	Yes	Yes	Yes	Yes
Umpire Fixed Effects				Yes
Batter Fixed Effects				Yes
N =	578688	553457	1132145	1132145
R^2	0.09	0.09	0.09	0.09

Panel B. Distinguishing by Terminal Count and Inning

	Non-Terminal *(5)*	**Terminal** *(6)*	**All Pitches** *(7)*	**All Pitches** *(8)*	**Early Inning** *(9)*	**Non-Early Inning** *(10)*
UPM	0.00619	-0.00628	0.00495	0.00420	0.00485	0.00361
	(0.0020)	(0.0027)	(0.0018)	(0.0024)	(0.0042)	(0.0029)
Terminal Count			-0.00719	-0.00730	-0.00982	-0.00561
*UPM			(0.0018)	(0.0018)	(0.0031)	(0.0022)
Pitcher Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Umpire Fixed Effects				Yes	Yes	Yes
Batter Fixed Effects				Yes	Yes	Yes
N =	870656	261489	1132145	1132145	396438	735707
R^2	0.04	0.17	0.09	0.09	0.09	0.08

Table 6: Effect of Umpire and Starting Pitcher Race/Ethnicity on Home Team's Winning Percentage (Probit Estimates, Dependent Variable Indicates the Home Team Wins)

	(1)	*(2)*	*(3)*	*(4)*
UPM	0.0126	0.0313	0.0342	0.0424
	(0.013)	(0.019)	(0.020)	(0.024)
Pitcher's run support		0.128	0.132	0.135
		(0.003)	(0.003)	(0.003)
N =	7124	6983	6983	6979
Home and Visitor Pitcher Fixed Effects		Yes	Yes	Yes
Visitor Team Fixed Effects			Yes	Yes
Umpire Fixed Effects				Yes

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

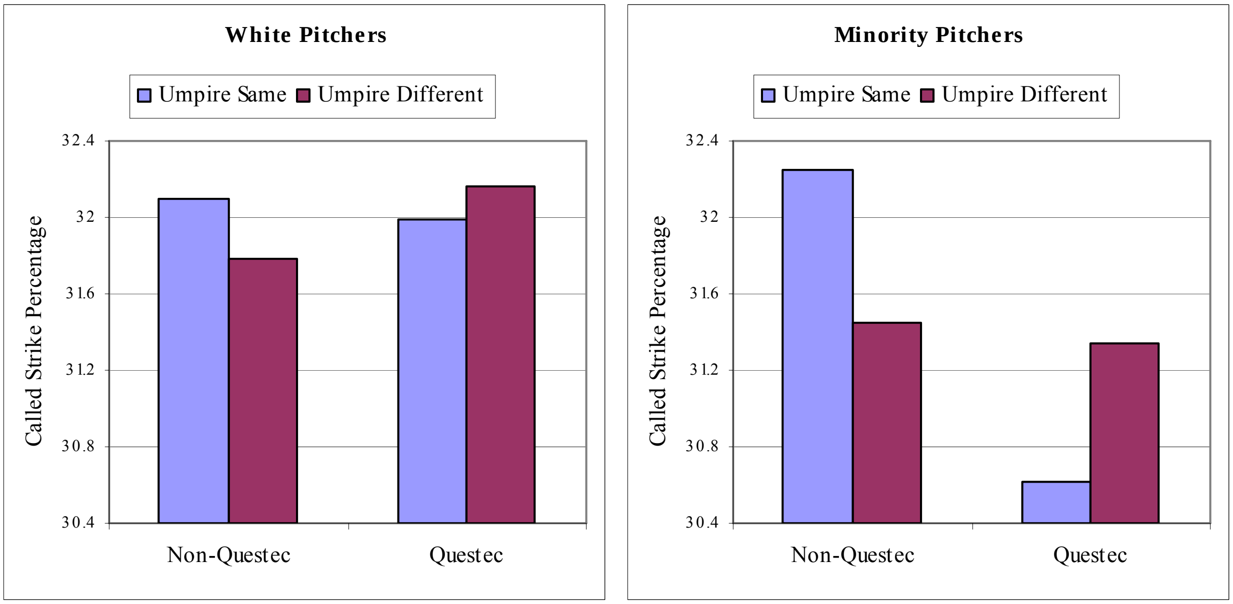
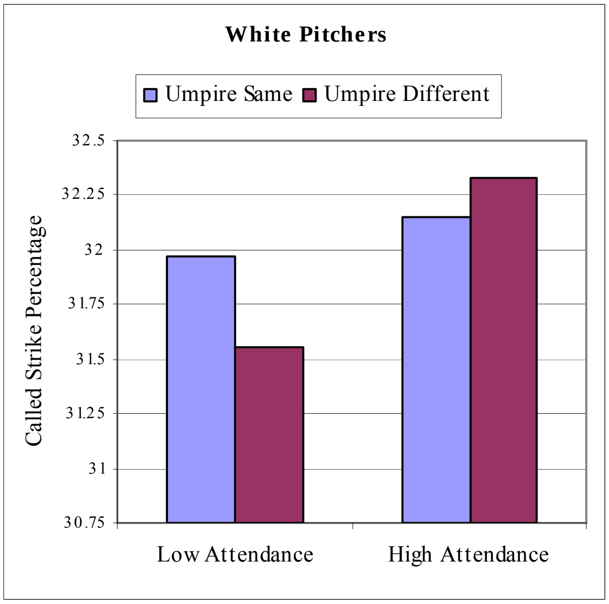


Figure 2: Race and Called Strike Percentage by Game Attendance



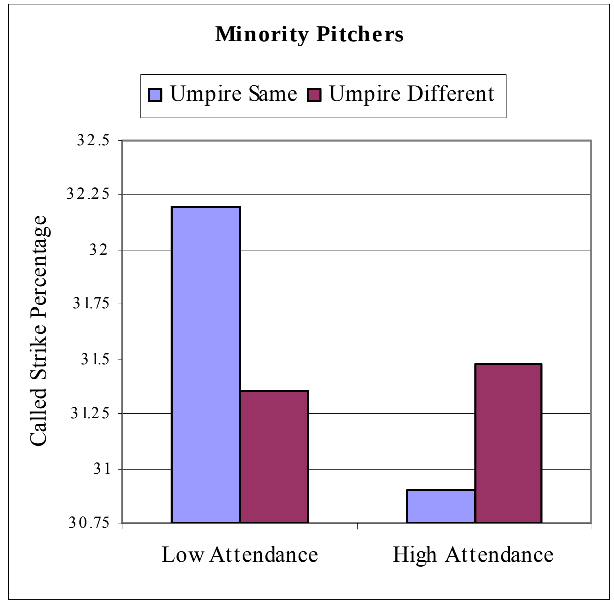


Figure 3: Race and Called-Strike Percentage in Terminal and Non-Terminal Counts

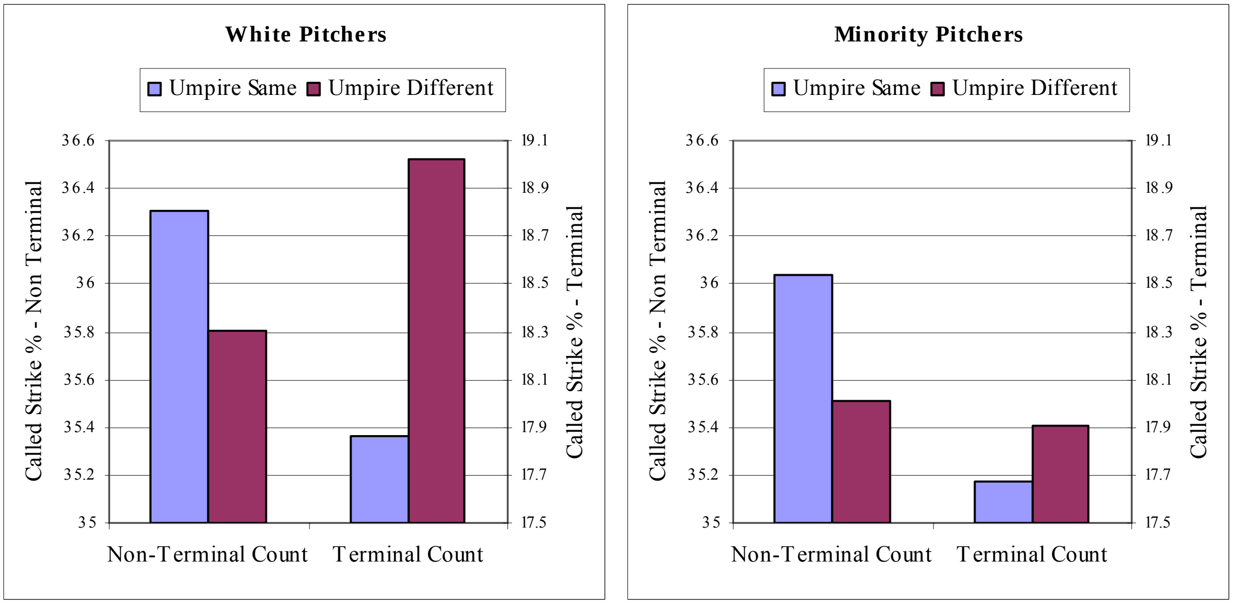
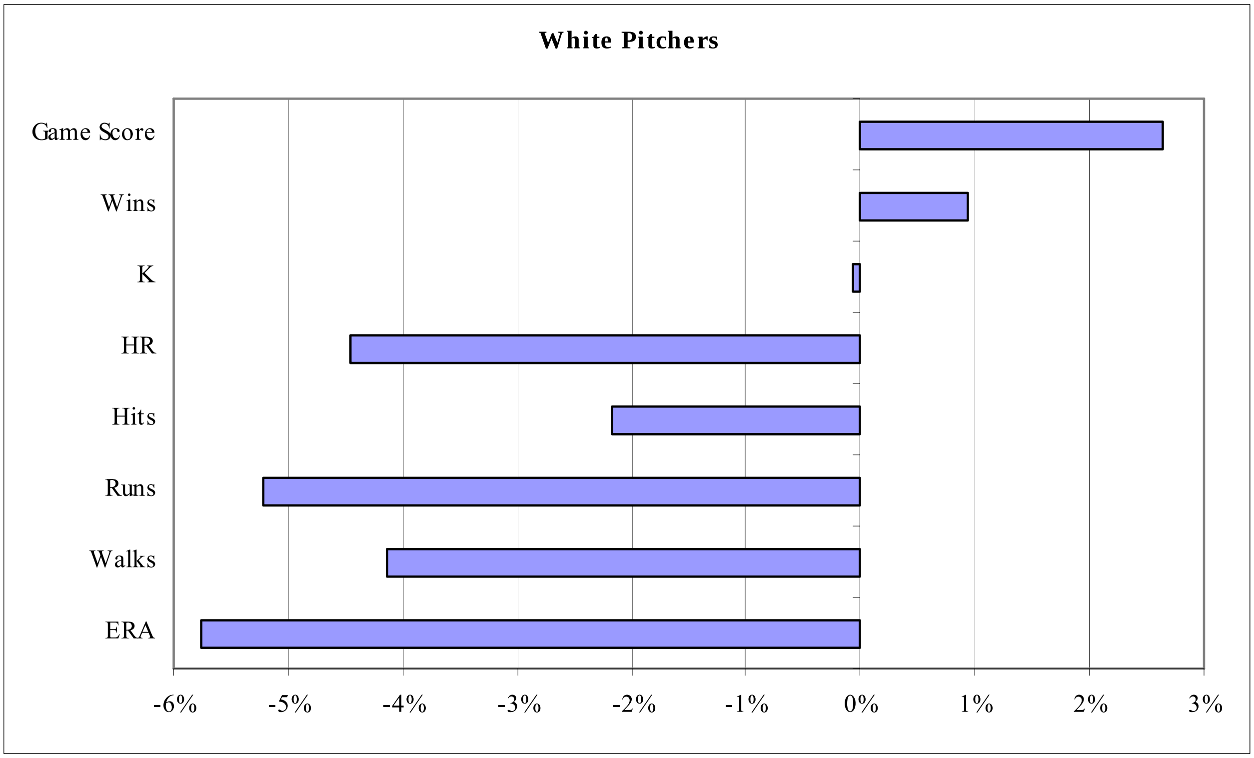
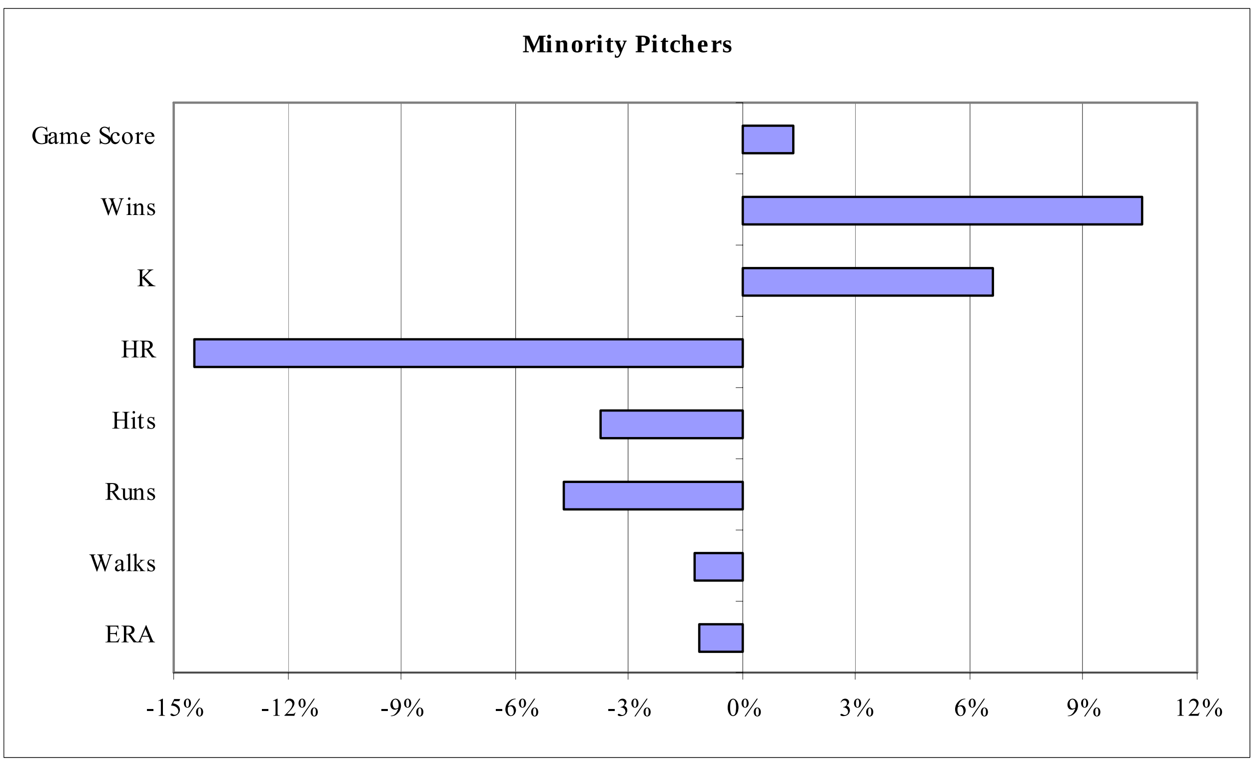


Figure 4: Change in Pitcher Performance When Umpire Matches Race/Ethnicity





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