

STRIKE THREE: DISCRIMINATION, INCENTIVES AND EVALUATION

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

This Version: August 2009

*Assistant professor of finance, University of North Carolina at Chapel Hill; assistant professor of finance, Southern Methodist University; assistant professor of finance, Auburn University; Sue Killam professor of economics, University of Texas at Austin, professor of labor economics, Maastricht University, 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, participants at seminars at several universities, and the editor and four referees for helpful discussions.

Abstract

Major League Baseball umpires express their racial/ethnic preferences when they evaluate pitchers. Strikes are called less often if the umpire and pitcher do not match race/ethnicity, but mainly where there is little scrutiny of umpires. Pitchers understand the incentives and throw pitches that allow umpires less subjective judgment (e.g., fastballs over home plate) when they anticipate bias. These direct and indirect effects bias performance measures of minorities downward. The results suggest how discrimination alters discriminated groups' behavior generally. They imply that biases in measured productivity must be accounted for in generating measures of wage discrimination.

I. Introduction and Motivation

Tests of labor market discrimination typically compare labor market outcomes (e.g., wages, promotion rates) across groups and, after controlling for worker productivity, assign any residual differences to discrimination. But what if an evaluator who discriminates along the dimension being studied *subjectively* determines a worker's measured productivity, as is true in all but the simplest piece-rate environments? A worker subjected to such biased evaluations might appear less productive, which ordinarily would justify a lower wage. However, in this case the econometrician would underestimate, or perhaps even miss altogether, instances of labor market discrimination when they in fact exist.

A subtler complication is that workers, anticipating biased evaluations, may alter their behavior in ways intended to minimize its impact. For example, consider a police officer who can either: 1) write traffic citations (the number of which can be objectively measured), or 2) investigate crimes (which is subject to performance review by a higher-ranking officer). If the officer has sufficient discretion, a biased evaluation in the second activity would lead the officer to alter the allocation of her time. Presumably, a positive bias would cause the officer to spend more time investigating crimes, and vice versa. Such bias-induced shifts in behavior further complicate the identification problem in assessing the impact of discrimination in labor markets.

This study addresses both of these issues, using detailed data on the evaluation, observed strategies and performances of Major League Baseball (MLB) players. Our focus is on racial/ethnic bias, specifically between the umpire (evaluator) and the pitcher (worker), although the arguments we develop apply to any type of subjective bias.¹ We pay particular attention to

¹Pitches are only subject to the umpire's discretion (are "called") when the batter does not swing, rendering necessary a judgment of whether the pitch was a "ball" or a "strike."

the race/ethnicity “match” of the umpire and pitcher, which occurs when, for example, a Black umpire evaluates a Black pitcher, as opposed to evaluating a White or Hispanic pitcher.

Our first observation is that pitchers who match the race/ethnicity of the home-plate umpire appear to receive slightly favorable treatment, as indicated by a higher probability that a pitch is called a strike, compared to players who do not match. Although this confers an advantage to some players at the expense of others, the effect we document here is small, on average affecting less than a pitch per game. Much more interesting are situations *when* and *where* the effects are strongest. Roughly one-third of the ballparks we study contain a system of computerized cameras (QuesTec) used to evaluate the umpires, comparing their ball/strike calls to a less subjective standard. Umpires have strong incentives to suppress any bias in such situations, as the QuesTec evaluations are important for their own career outcomes. With such *explicit* monitoring, evidence of any race or ethnicity preference vanishes entirely.

We find similar effects with *implicit* monitoring; when a game is well attended (and presumably more closely scrutinized), or when the pitch is pivotal for an at-bat, race/ethnicity matching again plays no role in the umpire’s evaluation. In situations where the umpire is neither explicitly nor implicitly monitored, the effect of the bias is considerable. As an example, a Hispanic pitcher facing a Hispanic umpire in a low-scrutiny setting (e.g., no cameras, poorly attended) receives strikes on 32.5 percent of called pitches, which drops to 30.0 percent if a Black umpire is behind the plate.

However, such direct effects are magnified when pitchers adjust their strategies in response to biased evaluations. Like the multi-tasking police officer mentioned above, a pitcher can alter his behavior to make himself either more immune, or more exposed to, the umpire’s judgment. Specifically, pitches thrown near the borders of the strike zone (e.g., over one of

home plate's corners) are called balls nearly as frequently as they are called strikes. They constitute a "fuzzy" region where the umpire can employ maximum subjectivity. Because such pitches are more difficult for batters to hit than those thrown directly into the strike zone, we would expect pitchers aware of favorable treatment to throw disproportionately to this fuzzy region. We find exactly this. Pitchers who match the umpire's race/ethnicity attempt to "paint the corners," throwing pitches allowing umpires the most discretion. This tendency is much stronger in low-scrutiny situations, when umpires face a lower cost of indulging their preferences.

At the end of both exercises, we are left with two specific conclusions. First, incentives matter. Unless provided strong incentives not to do so, umpires appear to allow the pitcher's race or ethnicity to influence their subjective judgments. This leads to a small, but non-trivial, direct effect on the game, simply by increasing the probability that a pitch is called a strike. Second, pitchers appear to understand these incentive effects, and take measures to protect themselves by avoiding situations requiring high subjectivity when facing a downward bias.

The results also lead to two general conclusions. First, these results show that when worker productivity is measured subjectively, and when such measurements are biased by discrimination, the usual tests for discrimination are biased toward finding nothing. We illustrate the size of this bias in our sample of baseball pitchers. Second, they illustrate the need to be aware of the manner in which discrimination in one facet of evaluation can lead market participants to alter their behavior in other dimensions.

Baseball offers several advantages when studying discrimination. First, 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 games' outcomes significantly. In addition, the very large number of independent pitch-level observations involving the interaction of different races/ethnicities allows us not only to explore umpires' preference for players of their own race/ethnicity, but also to examine preferences toward other races/ethnicities.² 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.³

The most fortunate aspect of the data set is that it allows us to develop several independent proxies for the scrutiny of the umpire's decisions, and in so doing, to test for the existence of price-sensitive discrimination by umpires. The time period that we analyze, 2004-2008, is special, because only during this time were a portion of the ballparks outfitted with computers and cameras to monitor umpires' balls and strikes calls. Because umpires are randomly assigned to venues, observing differences in their behavior between parks with and without monitoring technology makes a convincing case that properly placed incentives can have the desired effect. These results allow us not only to describe how biases can influence subjective performance valuations, but also to offer prescriptive suggestions to minimize their impact.

Several studies (e.g., Luis Garicano *et al*, 2005; Eric W. Zitzewitz, 2006, and Thomas J. Dohmen, 2008) have examined home-team preferences by referees/judges in various sports, and another, Michael A. Stoll *et al* (2004) examines racial match preferences in employment

²The data also include a small number of Asian pitchers, but because there are no Asian umpires, we exclude them in our analysis. Given their trivial numbers however, their inclusion gives nearly identical results in every instance.

³Umpires can be positioned behind home plate or at first, second or third base. The home-plate umpire 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.

generally. Our study most closely resembles Joseph Price and Justin J. Wolfers' (2007) work on NBA officiating crews' racial preferences. Although the first part of our empirical analysis corroborates their findings (but for a different sport), we are mainly interested in *when* or *where* racial/ethnic bias is most likely to be observed. Here, we offer two insights. First, we show that discrimination is price-sensitive, so that making it more costly reduces its expression. Second, we show that, when quantifying how players are affected by biased performance evaluations, the direct effect is only part of the story. Because players will alter their strategies in response, even situations that are seemingly insulated from a biased evaluator (e.g., non-called pitches in baseball games) are affected.

This research adds to a large literature on racial discrimination in sports, specifically in baseball, going back at least to Anthony H. Pascal and Leonard A. Rapping (1972), James D. Gwartney and Charles T. Haworth (1974) and Gerald W. Scully (1974), and recently J.C. Bradbury (2007) generally, with others dealing with particular racial/ethnic issues (Clark Nardinelli and Curtis J. Simon, 1990, David W. Findlay and Clifford E. Reid, 1997, and Rodney D. Fort and Andrew M. Gill, 2000). 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., Lawrence M. 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, the match to the race/ethnicity of their evaluator affects evaluations of workers, 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 difficulty of identifying discrimination pointed out by Stephen G. Donald and Daniel S. Hamermesh (2006).

In the next section we describe the pitch- and game-level data and explain our classification of umpires' and players' races/ethnicities. We analyze individual pitches in Section III, and in Section IV we show that umpires express these preferences strongly only in times of low scrutiny. We examine the indirect impact of discrimination on pitchers' strategies in Section V. Section VI shows the overall effects on pitchers' performances and derives the size of the effects of biased performance evaluation on the measurement of wage discrimination generally and for the example of pitchers' salaries.

II. Data

Pitches. There are 30 teams in Major League Baseball, with each team playing 162 games in each regular season. During a typical game each team's pitchers throw about 150 pitches, so that approximately 700,000 pitches are thrown each season. We collect pitch-by-pitch data from ESPN.com for every regular-season MLB game from 2004-2008.⁴ Our final dataset consists of 3,524,624 total pitches. For each pitch we identify the pitcher, pitcher's team, batter,

⁴The pitch-by-pitch information is from: <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=NNNNNNNN>.

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 other relevant information, including the stadium name, home team, away team, and the identities and positions of all four umpires.

Player and Umpire Race/Ethnicity. 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 and umpire. Players or umpires are classified as Hispanic if they were born in: Colombia, Cuba, Curacao, Dominican Republic, Mexico, Nicaragua, Panama, Puerto Rico or Venezuela. 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.⁵ We also utilize a similar list of past and present Hispanic players in MLB from Answers.com.⁶ All remaining unclassified players and umpires are classified by visual inspection of pictures found in Internet searches.⁷ Three of the race/ethnic groups are represented among umpires (there are no Asian umpires in MLB), and all four are represented among pitchers.

Table 1 presents the distributions of the pitch outcomes. The first row of the table summarizes all pitches, while subsequent rows sub-divide pitches based on the race/ethnicity of

⁵The complete list can be found at http://Blackvoices.aol.com/Black_sports/special/a/african-american-players-in-mlb/20070413095009990001.

⁶The complete list can be found at <http://www.answers.com/topic/list-of-hispanic-players-in-major-league-baseball>.

⁷For a few umpires, no pictures were available on the internet. For each of them we watched past games in which the umpires worked to ascertain their 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, this will introduce measurement error into the matches and thus reduce the strength of any results that we generate.

the pitcher, the batter and the home plate umpire, respectively. Approximately 46 percent of pitches elicit a swing from the batter, hit the batter, or are intentionally thrown out of the strike zone. Our pitch-level analysis focuses on the 54 percent of pitches (1.89 million) that result in called strikes or balls, since these alone are subject to evaluation by the home-plate umpire. Of these, about 32 percent are called strikes, and the rest are called balls.

The table also reports the number of pitchers, batters and home-plate umpires in each of the four race/ethnicity categories. The percentages of White pitchers (70 percent) and batters (61 percent) are lower in our sample than the percentage of White umpires (89 percent). On the other hand, Hispanics, comprising 23 percent of pitchers and 26 percent of batters, are under-represented among umpires (only 5 percent). Black pitchers, batters and umpires comprise 3 percent, 10 percent, and 6 percent of the samples, respectively. Asian players comprise 3 percent of pitchers and 2 percent of batters.

Pitch Location. For approximately one-third of the games played in 2007 season and all of those played in 2008, we collected from PITCHf/x several additional variables. PITCHf/x, a computerized technology owned by Sportvision, uses two cameras to record the path of a pitch from the pitcher's hand to home plate.⁸ The parameters measured and calculated using this technology include: 1) the pitch type, determined using MLB's proprietary neural net classification algorithm, 2) the estimated pitch location when it crosses the home plate relative to the center of the front of the home plate, and 3) the top and bottom of the strike zone as determined by the PITCHf/x operator.⁹

⁸The information is provided by the MLB free of charge at: <http://gd2.mlb.com/components/game/mlb/>.

⁹The operator sets a horizontal line at each batter's belt as he settles into the hitting position, and the PITCHf/x software adds four inches up to define the top of the zone. For the bottom of the zone, the PITCHf/x operator sets a horizontal line at the hollow of each batter's knee. More information on PITCHf/x's parameters can be found at: <http://fastballs.wordpress.com/category/pitchfx-glossary/> and <http://webusers.npl.illinois.edu/~a-nathan/pob/tracking.htm>.

Pitcher Performance. For each starting pitcher’s appearance in each game, we collect from box scores the number of innings pitched, the numbers of hits, runs and home runs allowed, walks, strikeouts, and earned runs (downloaded from the ESPN website). We also obtain the final score of the game to identify the winning and losing teams.

III. Called Pitches and Umpire-Pitcher Matches

Table 2 reports for each pitcher/umpire racial/ethnic combination the number of pitches thrown, the number of called pitches, the number of called strikes and the percentage of called pitches that are strikes. About two-thirds of the called pitches in our sample occur when the umpire and pitcher share the same race/ethnicity (mostly White pitcher/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 (53.7 percent), a central difference is that the percentage of called pitches that are strikes is higher when they match (32.0 percent) than when they do not (31.5 percent).

The summary statistics in Table 2 ignore possible differences inherent in the quality or “style” of pitchers by race/ethnicity. They also ignore the possibly different outcomes generated by non-random assignment of pitchers to face different opponents, and of umpires to games played by particular teams.¹⁰ To account for these and other potential difficulties, our central test for umpires’ discrimination in calling strikes is the specification:

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

¹⁰Examination 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.

where the dependent variable is an indicator of whether a called pitch is a strike, the γ are parameters, ε is a well-behaved error term, 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 fixed effects for each pitcher, umpire and batter so that UPM picks up the *marginal* effect of a racial/ethnic match between the home-plate umpire and pitcher. That is, because any player or race-specific effects are swept out by the fixed effects, umpires' bias is identified purely via the interaction term, UPM.

In addition to these, we employ a number of 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 pitches based on the ball-strike count. Inning indicators are also included, because pitchers are usually less fatigued early in games, and because a "relief" pitcher often replaces a pitcher who starts the game in later innings, with a different (often reduced) accuracy.¹¹ Home-field bias is captured by top-of-the-inning indicators, which account for which team is pitching. Lastly, we include the pitcher's score advantage (defined as the number of runs, potentially negative, by which the pitcher's team is ahead).

Table 3 presents the results of estimating equations where the pitcher's and umpire's race/ethnicity are allowed to influence the likelihood of a called strike. All the estimates are based on linear-probability models (but probit estimates present the same picture) with heteroskedasticity-robust standard errors. The first three columns show specifications separately for White, Black and Hispanic pitchers, respectively, controlling for umpire race/ethnicity and pitcher fixed effects. The next three columns show separate equations for White, Black and Hispanic umpires, respectively, controlling for pitcher race/ethnicity and umpire fixed effects.

¹¹With pitcher fixed effects, this second reason for inning indicators is obviously subsumed.

The final three columns include all pitchers and umpires, with each column adding successive vectors of fixed effects, including in the final column pitchers, umpires and batters.

There is some, albeit weak, evidence of favoritism by umpires for pitchers who match their race/ethnicity. For example, Column (1) shows that Hispanic umpires judge White pitchers more harshly than do White umpires (the omitted indicator variable), but that they judge Hispanic pitchers more favorably (Column (3)). Similarly, Column (4) shows that White umpires, the overwhelming majority, judge minority pitchers more harshly than they judge White pitchers. Taking the results in Column (9) with the full sets of control variables and fixed effects as the best description of the underlying behavior, however, it is quite clear that there is no generally significant impact of the match on umpire evaluations ($p=.34$).

Although the results with the broadest sets of fixed effects do not suggest a significant effect of the umpire-pitcher match, the point estimate implies that a given called pitch is approximately 0.16 percentage points more likely to be a strike if the umpire and pitcher match race/ethnicity. The likelihood that a given called 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 one-half percent above the rate when there is no match.¹²

IV. Biased Evaluation When Bias Is Costly

One might examine the results in Table 3 and conclude that, while the point estimates are interesting, their statistical insignificance means that there is very little here. Given an economist's view that agents acting out their preferences will react to the price of an activity, however, it is worthwhile examining the impacts of umpire-pitcher matches as the price of

¹²As 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.

discrimination changes. 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 (Stephanie A. Goodwin *et al*, 2000; Irene V. 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 legal penalties. Here we proxy the price of discrimination by the extent to which an umpire's evaluations of pitchers will be scrutinized. We 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. From 2004-2008, MLB had 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 memorably, pitcher Curt Schilling's assault on a camera after a poor outing), 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

¹³The 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, which can influence his promotion to “crew chief,” assignment to post-season games, or even retention in MLB.¹⁴

Because QuesTec is installed in roughly one-third of ballparks, and because umpiring crews are rotated randomly around the league’s ballparks, virtually every umpire in our data set calls a substantial number of pitches in parks with and without QuesTec.¹⁵ Additionally, both the umpires’ and teams’ schedules change every year, exposing each umpire to a wide cross-section of batters and pitchers in both types of parks. Throughout the analysis we test whether greater scrutiny—the possibly higher cost of bias in subjective evaluation of pitches in QuesTec parks—leads umpires to call strikes “by the book.” Any role that racial/ethnic (or any other) preferences play in influencing pitch calls should be mitigated if costs of being judged substandard are imposed, as through QuesTec. Some pitchers may, however, react differently from others in response to QuesTec.¹⁶ For that reason, in all of the estimates in this part (and hereafter) 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, umpire-QuesTec fixed effects, and batter-umpire-QuesTec fixed effects.

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

¹⁴An umpire’s evaluation is not based solely on 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.

¹⁵The 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.

¹⁶For 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.”

umpires of the matched race/ethnicity, depending on whether the pitch is thrown in a park with QuesTec installed. The difference in the called-strike percentage between QuesTec and non-QuesTec parks is significant for both White and minority pitchers.

Table 4 contains the results of estimating (1) separately for QuesTec and non-QuesTec parks, with controls for inning, pitch count, pitcher score advantage, and top of the inning.¹⁷ The results are striking: In ballparks with the UIS, shown in Column (1), the coefficient on UPM is -0.48 percentage points and is not significantly different from zero. In parks without QuesTec, shown in Column (2), the same coefficient is 0.59 percentage points per pitch ($p=.007$). These differences make clear why UPM is not significant in the aggregate sample. The effects found in Table 3 averaged the statistically significant positive impact of an unscrutinized match (non-QuesTec) with a statistically insignificant negative impact of a scrutinized match (QuesTec) that is nearly as large. Thus, in the presence of price-sensitive discrimination, we should expect the point estimates in Table 3 to be low, since the entire sample consists of a mix of high- and low-scrutiny games. Specifically, QuesTec covers about 37 percent of pitches, so that the average result from Table 3 is easily reconciled: $(.37)(-.48)+(.63)(.59)=.19$, close to the 0.16 estimate obtained with a comparable set of fixed effects.

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 over 1 percentage point, more than offsetting the favoritism shown by umpires when QuesTec does not monitor them. Each effect is highly significant, implying that umpires implicitly allow their apparent preference for matched pitchers to be expressed when the pitches underlying their decisions are not recorded.

¹⁷The direct effect of being in a QuesTec park is, of course, not directly observable, being subsumed in the pitcher-QuesTec fixed effects.

QuesTec is an explicit monitoring technology. Implicit monitoring can have similar effects, suggesting that even subtle incentive mechanisms can have desirable effects on otherwise discriminatory outcomes. The two measures for implicit scrutiny of umpires are crowd attendance (scaled by stadium capacity) and the “importance” of the pitch.¹⁸

The idea for the first is simple. Having many fans close to home plate presumably exposes the umpire to their scrutiny—a badly called pitch is unlikely to go unnoticed.¹⁹ Figure 2 confirms that crowd attendance, like QuesTec, dramatically alters umpire behavior. A game is defined as “well-attended” if the crowd attendance is above the median percentage capacity in this sample, roughly 70 percent. Compared to well-attended games, umpires calling poorly-attended games appear to favor pitchers of matched race/ethnicity. In the case of White pitchers, both minority and White umpires tend to call fewer strikes in poorly-attended games, but the reduction in strikes called by minority umpires is over three times larger. The same effect is seen to an even greater degree among minority pitchers. During well-attended games, matching minority umpires call about 0.8 percent *fewer* strikes. They call 0.7 percent *more* strikes in poorly attended ones, a net effect over 1.5 percentage points.

In Columns (1) and (2) of Panel A in Table 5, we show the results of estimating (1) separately for well- and poorly-attended games respectively. Each equation includes the same battery of controls as in Table 4, i.e., pitcher, umpire and batter fixed effects, pitch counts, and inning indicators. As with the QuesTec results, the UPM variable is significant ($p=.008$) only in

¹⁸We scale by stadium capacity to minimize the impact of differences between stadium sizes. If we assume that stadiums populate relatively uniformly, attendance/capacity is a good proxy for the *number* of fans close enough to judge pitch location. In any case, this scaling makes little difference in our results. If instead we use attendance, all coefficients of interest remain highly significant.

¹⁹Percentage attendance may also proxy the popularity of the participating teams or the importance of a particular game. Thus, not only might the umpire be exposed to more scrutiny from the additional fans present at well-attended games, but he may also face added scrutiny in the form of larger television audiences and increased air-time given to game highlights.

poorly-attended games, with an effect of 0.64 percentage points per pitch. During well-attended games there is no significant effect of an umpire-pitcher racial/ethnic match and, as before, the point estimate is negative. Column (3) generalizes the results by aggregating all games and interacting UPM with the indicator for a game being well attended. 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 0.36 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 for the price of discrimination are qualitatively identical to those obtained for the QuesTec/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. Specifically, a pitch that is thrown with two strikes and/or three balls is potentially terminal, as a third strike or fourth ball terminates the at-bat. In such situations, the umpire’s judgment is likely to be scrutinized more heavily by the pitcher, batter, catcher, managers and fans. An initial glimpse into the effects of this distinction is shown in Figure 3. Here we observe the same contrast as for the previous two proxies for scrutiny, as umpires appear to favor pitchers with whom they match only in non-terminal counts, when scrutiny is likely to be reduced.

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 usual set of control variables. We consider pitches of differing importance separately, with the result that the coefficients of UPM have opposite signs. For pitches that cannot be terminal, the estimated

coefficient of UPM is 0.31 percentage points ($p=.15$)—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), as the coefficient on the interaction term is negative and significant at better than the 1 percent level.

In Columns (7) and (8) we consider another source of within-game variation in implicit scrutiny. We assume that, because umpires' evaluations are more likely to be pivotal late in games, scrutiny in the first few innings is likely to be comparatively less. We thus designate the first third (three innings) of a game as "early," and the remainder "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 over twice as large in early as in late innings (-0.86 vs. -0.38 percentage points).

Our proxies for scrutiny are not redundant. The correlation between QuesTec and attendance percentage is small, and because the type of pitch (terminal or non-terminal) is a within-game measure, it is necessarily uncorrelated with either between-game measure. It is therefore not surprising that, when all three interactions are included simultaneously in Panel C, everything remains significant with nearly identical magnitudes as in Panels A and B.

Before proceeding to issues of robustness, we briefly address whether the UPM effect is due to positive bias for pitchers who match the umpire's race/ethnicity (i.e., favoritism), or to negative bias against those who do not match. Answering this question in our context is

difficult, because ball and strike calls are inherently subjective. (Compare this to tennis, where the definition of a shot being “in” or “out” is completely objective, allowing, for example, computerized instant replay to reverse the judge’s calls.) Absent an objective standard on strike calls, we cannot precisely quantify the bias’ direction; but comparing umpires’ behavior between QuesTec and non-QuesTec ballparks provides some illumination.

If one accepts the premise that umpires exercise special care in QuesTec parks, the strike percentage there, although not perfect, is closer to the desired benchmark of objectivity that would permit the desired calculation. For each of the nine possible race/ethnic combinations, we compare the called strike percentage in QuesTec parks (the quasi-objective benchmark) to that in non-QuesTec parks. First, all three cases of a match (e.g., White-White) show a higher called strike percentage in non-QuesTec parks, which suggests favoritism in less scrutinized situations. Second, five of the six cases of non-match show a lower called strike percentage in non-QuesTec parks, which suggests negative bias. Such a two-sided pattern not only justifies the use of an aggregate UPM variable in Tables 3-5; it also demonstrates that the effect is symmetric and pervasive across nearly every possible combination. However, we do not focus further on the positive/ negative bias distinction, because baseball – and indeed all games with winners and losers – is a zero-sum game. It is *relative* treatment that matters most, just as in labor markets generally it is disparate treatment, not the difficult-to-identify distinction between the absence of favoritism and the presence of negative bias, that underlies so much case law.

Other matches. An umpire influenced by the race of the pitcher may also be influenced by that of the batter or the catcher, especially because in the latter case, the umpire is in continuing close contact. We find little evidence to support this argument. In the same types of regressions as in Tables 3-5, but with new matching variables, there is some very weak evidence

that batters receive the type of preferential treatment experienced by pitchers. But the magnitudes and statistical significance are much smaller in every case. Reliable evidence for similar catcher-umpire dynamics is even weaker. Umpires appear focused on the pitchers they are judging. Their matches with other relevant players do not affect their judgments.

Post-season. The three scrutiny proxies we employ have the advantage of splitting the sample of called pitches into two large groups, generating the statistical power required to detect subtle differences in called strike probabilities. There are many additional cross-sectional tests one could perform, e.g., comparing playoff to regular season games (because the former are likely to be particularly scrutinized), but such thin cross-sectional comparisons contain almost no power. For example, we replicate the analysis in Panel C of Table 5, aggregating playoff and regular season games, and including interaction terms for post-season pitches with the coefficients of interest (unreported). There is only the weakest of evidence that playoff situations reduce further the expression of umpire bias (the interaction of post-season with UPM is negative, as expected, but the p-value is 0.74). We encounter a similar problem when, for example, examining particularly “important” games, such as those pivotal for playoff races late in the season.

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 on each umpire from a variety of sources and include 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.

We find no evidence that the racial composition of an umpire’s birthplace or residence predicts his propensity to penalize non-matching players, but there is some weak evidence that

bias is more likely among younger and less experienced umpires. The coefficient on UPM in a re-specification 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. If (1) is re-estimated separately for the 18 “crew chiefs,” veterans selected for their seniority and performance, the point estimate of the coefficient on UPM is nearly zero. This evidence is consistent with models 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.

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, and also statistically insignificant. Our conclusions are not affected by the racial/ethnic mix of the team’s catchment area.²⁰

Gaming the System. Perhaps managers are implicitly both aware of these preferences and able to act upon them. Because the majority of umpires are White, there is a distinct advantage for a team with one or more minority pitchers (particularly starting pitchers) to have QuesTec in its home park. We found no information about the how teams were awarded QuesTec in their home parks, or whether they could influence this choice. A second possibility is that teams

²⁰The overwhelming majority of minority pitchers are Hispanic. We have aggregated them, but some are White Hispanics, while others are Black Hispanics. 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, we visually inspected the pitchers’ pictures, divided the Hispanic aggregate into White and Black groups, and consequently re-defined UPM. 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. We also investigated whether American-born Hispanic pitchers were treated differently from Hispanic pitchers born outside the U.S., and found no evidence that the pitcher’s birthplace affected expressed racial/ethnic bias by umpires.

receiving QuesTec systems traded for minority pitchers from teams whose parks were not similarly equipped.

Although we have no direct evidence, some simple calculations suggest that either possibility may have merit. For visiting pitchers, the percentage of pitches thrown in QuesTec parks is nearly identical for Whites and minorities (37.4 and 37.9 respectively). This is to be expected, because on average, teams play approximately the same fraction of opponents whose home stadiums contain QuesTec. Thus, there is no evidence that visiting managers adjust their pitching lineups to minimize the exposure of their minority pitchers to the subjective bias of a White umpire.

Home pitchers tell a different story. Minority home pitchers throw 39.2 percent of their pitches in QuesTec parks, compared to only 35.5 percent for White pitchers. Home minority pitchers are more likely to be in QuesTec environments, which can only be the case if their home ballpark has QuesTec. This is consistent with either initial non-random assignment of QuesTec to teams with a disproportionate number of minority pitchers, with transactions that increase the fraction of minority pitchers for teams already equipped with QuesTec, or with game-time lineup juggling by home teams. Although we cannot distinguish among these alternatives, this evidence is interesting by suggesting that biased evaluations in one area (e.g., called strikes) may have unintended consequences in other arenas (e.g., the allocation of Minority pitching talent). Note that none of these possibilities alters the significance or interpretation of the previous results, as all regressions control for player ability, umpire tendencies, and the presence or absence of QuesTec.

V. The Effects of Biased Evaluations on Agents' Strategies

The pitch-level evidence makes very clear that direct effects on pitch outcomes are small. Of course, one can construct specific examples where the estimated direct effect is fairly large: a Black pitcher throwing a *non-terminal* pitch in the *early innings* of *poorly-attended* games in a *non-QuesTec* ballpark gains over 6 percentage points by matching (41.4 vs. 35.2 percent called strikes). But in most situations, the direct impact on called pitches is not large.

Indirect effects on players' strategies may, however, have larger impacts on the outcomes of plate appearances and games. The dynamic between a pitcher and batter is clearly affected by each party's beliefs about the umpire's evaluation in the event of a called pitch. If a pitcher expects favoritism, he will incorporate this advantage into his strategy, perhaps throwing pitches that allow the umpire more discretion. This in turn may change the batter's optimal behavior. If the batter expects such pitches to be called strikes, he is forced to swing at "worse" pitches, which reduces the likelihood of getting a hit.²¹

To appreciate more fully such induced changes in strategy, for all the starting pitchers for whom such data are available (over 500,000 pitches), we augment the pitch-level data with the data set on pitch characteristics.²² This level of detail allows addressing the extent to which pitchers alter their strategies (e.g., location and type of pitch), when facing a biased subjective evaluation. Panel A of Table 6 summarizes the two location variables of interest: 1) the horizontal pitch distance, and 2) the pitch height. The first is the distance (in feet) from the

²¹Examining the coefficients on the count indicators in Table 3 illustrates the intuition. When the pitcher has a substantial advantage in the count, he has little incentive to throw a "hittable" pitch, i.e., one near the middle of the plate. Instead, he usually throws pitches near the corners that are both less likely to be hit if the batter swings, and less likely to be called strikes if the batter does not. Such behavior translates into sizeable advantages for pitchers depending on the count. In 2004, batters got a hit 33 percent of the time when the count was 2-1 (two balls and one strike), but dropped to less than 18 percent when the count was 1-2.

²²The number of pitches differs slightly across the panels because of difficulties in classifying by location and type. Jowei Chen (2007) used these data for a single season as controls to examine racial bias in MLB umpires' calls.

center of home plate. (The slightly negative mean value for this variable reflects the tendency to avoid hitting or pitching inside to batters, and most batters are right-handed.) The second is calculated as the pitch's vertical distance from the center of the strike zone, which is set by the computer operator to be between the batter's waist and knee (typically 2.5 feet above the ground). That this region varies among batters is not a problem, as all of the analyses we present include batter fixed effects.

Pitches in certain locations are almost always called one way or the other. This is apparent in Figure 4, which shows the location of all called strikes. A strike generally corresponds to the elliptical region centered around the plate and slightly below the batter's waistline. We define three concentric ellipses corresponding to: 1) the *inside* of the strike zone, 2) the *edge* of the strike zone, located just outside the center region, and 3) the complement to both regions, denoted as *outside*. Figure 4 shows the *inside*, an ellipse with major axis equal to 2 feet, and a minor axis equal to 1.6 feet. The *edge* is bordered by the *inside* and the *outside*, a larger ellipse with major axis 2.6 feet and minor axis 2.2 feet. We experimented with several alternative sizes for these ellipses, and none changes the basic results. Panel B of Table 6 summarizes the distribution of pitches by region. Roughly 40 percent are thrown in each of the *inside* and *outside* regions, with the balance in the *edge*.

Pitches thrown to each region generate different outcomes. A called pitch in the *inside* region will be a strike almost 87 percent of the time. Thus, a pitch thrown in this region is associated with little uncertainty. Similarly, a pitch thrown in the *outside* region has very little chance of being called a strike (3.8 percent), resulting in even less uncertainty about the call. A pitch thrown to the *edge* region, however, is called a strike 44.3 percent of the time, generating

nearly the maximum uncertainty possible for a binomial variable. The *edge* region allows the umpire the greatest discretion.

Given this distinction, it is comforting that the *edge* is where the effects of the previous sections occur. Matches in the *inside* are associated with an increase in the called strike percentage of only 0.3 percentage points, from 86.7 percent (no match) to 87 percent (match). The *outside* shows no difference at all. The percent called strikes in the *edge* is 43.6 absent a match, compared to 44.5 percent with a match. If pitchers understand this advantage, then we can predict that a matching pitcher will throw more pitches to the *edge*, where his advantage (courtesy of a biased umpire) is maximized. This aids the pitcher, because pitches to this region are considerably more difficult for the batter to hit.

Panel C of Table 6 presents the results of regressions similar to (1), except: 1) we include *all* pitches thrown by starting pitchers, not just called pitches, as was required for the previous analysis; and 2) the dependent variable indicates whether a pitch is thrown to the *edge*. As before, we include fixed effects for each pitcher, umpire, and batter, as well as all count and inning indicators. The first column shows the result for pitchers in QuesTec parks, where we see that a race/ethnicity match between the pitcher and umpire has virtually no effect on pitch location. In non-QuesTec parks, the situation changes drastically. Matches lead to a 0.95 percentage-point increase in the probability of throwing to the middle region, representing a 5 percent increase relative to the base non-match rate of 19.7 percent. The third column aggregates all observations, where the magnitude of the interaction term is over 1 percent ($p=.10$). By throwing pitches that can reasonably be called as either balls or strikes, matching pitchers gamble on the fact that this region offers them an advantage.

Panel D of Table 6 shows a related, but distinct, finding. Its interpretation requires some institutional detail. The most common pitch in baseball is the *fastball* (about 58 percent of our sample), which travels in a mostly straight line from the pitcher's hand toward home plate. Skilled pitchers, however, can place spin on pitches, causing them to deviate from a straight trajectory. Pitches with substantial "break" end their flights with dramatic dips that are notoriously difficult to hit solidly. Adding this vertical element also makes these pitches more difficult to judge.²³ As with pitches to the *edge*, judging a curveball requires subjectivity, which is the source of a matching pitcher's advantage. If matching pitchers are aware of a biased umpire, we would expect them to throw more breaking pitches.

The first column in Panel D shows that, in QuesTec parks, a match is associated with a slight preference for breaking balls. In non-QuesTec parks, the magnitude quadruples to 1.28 percent ($p < 0.001$). The aggregation of all pitches in Column (3) tells the same story. Matching pitchers in parks without explicit monitoring select pitches allowing umpires the most discretion, enabling them to maximize their advantage stemming from the umpire's bias. While Panel D only makes the distinction between curveballs and other pitches, the result is nearly identical if we distinguish between all breaking pitches (e.g. sliders, cutters) and fastballs.

The Appendix presents a simple game-theoretic model that formalizes the intuition for the results in Table 6. It shows that, when pitchers expect a racial/ethnic match with the umpire to result in more called strikes, their optimal response is to select pitch locations further from the center of the plate (as shown in Table 6). Intuitively, the umpire's bias reduces the penalty for throwing *edge* pitches that are difficult for the batter to hit.

²³A Google search for "umpire" and "calling a curveball" generates hundreds of links to articles and advice to umpires wishing to master the evaluation of this difficult pitch.

The results and the general theory seem relevant for examining the effect of bias on agents' behavior in a variety of contexts. For example, in the literature on racial profiling (e.g., John Knowles *et al*, 2001; Nicola G. Persico, 2002), while the search data in the empirical literature do not allow examining these indirect effects, the theory demonstrates that they will arise. On the reverse side, the theory of affirmative action (Stephen Coate and Glenn C. Loury, 1993) demonstrates that anti-discriminatory policies will produce indirect effects on agents' behavior.

In the larger labor market the history of occupational segregation is replete with examples of discrimination in occupational choice altering agents' labor-market behavior to their own detriment. The exclusion of Jews from property ownership in the late Middle Ages, the exclusion of African-Americans from most of the railway trades until the 1950s, and perhaps even the "glass ceilings" in corporate hierarchies, all resulted in crowding into occupations (see Barbara R. Bergmann, 1971) that was an indirect effect of bias in other occupations. Our work merely provides a specific example of these effects that allows them to be identified more clearly than in the broader labor-market context.

VI. Measures of Performance and the Measurement of Discrimination

The model in the Appendix and evidence in Section V jointly imply that, conditional on swinging, the batter is less likely to get a hit when the umpire and pitcher match. This implication suggests analyzing a variety of game-level performance measures for each starting pitcher to infer the total of the direct and indirect effects of bias on performance. Table 7 examines each starting pitcher's hits allowed, runs given up, and wins (per game).²⁴ Because the

²⁴We tabulate each starting pitcher's win decisions rather than whether the team actually wins the game. If one considers this second measure instead, the differences are similar, although the overall mean is 0.5 by construction. (The mean for wins is lower in the text table because relief pitchers are frequently awarded decisions.)

sample shrinks by nearly three orders of magnitude compared to the pitch-level results, our ability to detect relatively subtle effects is greatly reduced. Nonetheless, for both groups (non-Hispanic Whites in Panel A, minorities in Panel B), pitchers' outcomes along all three game-level performance measures are superior in matching situations. Non-Hispanic White starting pitchers who match are awarded wins 1.7 percentage points more often in non-QuesTec parks, which reverses to negative 3 percentage points in QuesTec parks. The "QuesTec effect" of 4.6 percentage points is nearly significant ($p=.08$). For minority starting pitchers, the similar gap is even larger, at 12.9 percentage points ($p=.06$), although there are only 74 matches in QuesTec parks.

Several other aggregate performance measures show the same patterns. Both groups give up fewer hits in matching situations in non-QuesTec parks, Whites by about 1 percentage point, minorities by about 2 percentage points. As before, each pattern reverses in QuesTec parks. A similar pattern is seen along additional performance metrics. Figure 5 shows several of them, again for non-QuesTec parks and for White and minority pitchers separately. Presented as percentage changes from their baseline levels (Table 7 presented differences in levels), the vast majority improve in match situations. From the starting pitcher's perspective, a racial/ethnic match with the umpire helps his earned runs (fewer), hits (fewer), walks (fewer), and home runs (fewer). Only strikeouts go in the opposite direction. One might expect little effect for strikeouts, which, at least in the fraction that are called third strikes, require bias on a terminal count, which we have already shown does not occur.

To the extent that pay is based on measured productivity, our findings of small direct and larger indirect effects of racially/ethnically disparate treatment carry important implications for measuring the extent of discrimination in baseball and in labor markets generally. In particular,

they imply that estimates of the extent of discrimination will be understated, even controlling for standard measures of performance.

Consider a simple earnings equation:

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

where W is the logarithm of earnings, M an indicator of minority status, P^* worker i 's true productivity, and v 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:

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

$$P_i = P^*_i, \quad \text{if } M=0,$$

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

$$W_i = [\alpha + \beta\phi]M_i + \beta P_i + v_i, \quad \text{or} \quad (2')$$

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

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

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 5 and estimates of β from three studies of MLB that examined pitchers and used at least some of these outcomes as determinants of salaries. We are essentially estimating the reduction in minority pitchers' salaries as a result of the average amount of bias arising during the 2004-2008 seasons due to umpire-pitcher racial/ethnic matches. Lawrence M. 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 2004-2008, applying the effects in Figure 5, and using his parameter estimates yields an estimated bias of $\beta\phi = 0.034$. Mark P. Gius and Timothy P. Hylan (1996, Table 6.2) use strikeouts/inning, walks/inning 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 $\beta\phi = 0.012$. Finally, using the estimates for starting pitchers by Anthony C. Krautmann *et al* (2003), the estimate of $\beta\phi = 0.074$.²⁵

While we have demonstrated the extent of bias to estimated discrimination in earnings that arises because of biased evaluations of MLB 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 discretely and very frequently—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 disparate outcomes generally.

The general point, that bias will affect measures of productivity, is not new (see, e.g., Glen G. Cain, 1986). It is, however, generally ignored in the scholarly literature measuring the

²⁵We use the means in this sample as the baselines. For the fraction of games won, 0.37; for ERA, 4.44; for strikeouts/inning by starting pitchers, 0.75; and for walks/inning, 0.43. We can take the estimates of the bias as examples here to infer the dollar impacts of this subtle form of discrimination. In 2006, the mid-point of our sample, the average salary of starting pitchers in MLB was \$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.

wage effects of discrimination. In the huge industry of employment litigation, standard practice is to adjust wages using measures of supervisors' evaluations of workers. As we have shown, even in a very controlled and highly scrutinized environment, these can be biased against minorities. Our results suggest that this bias must be accounted for whenever one wishes to measure racial/ethnic disparities in rewards in the workplace.

VII. 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 slightly 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 has substantial effects on pitchers' measured performance and games' outcomes. The link between the small and large effects arises, at least in part, because pitchers alter their behavior in potentially discriminatory situations in ways that ordinarily would disadvantage themselves (such as throwing pitches directly over the plate). As in many other fields, racial/ethnic preferences work in all directions—most people give preference to members of their own group. In MLB, as in so many other fields of endeavor, 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, minority 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 almost 90 percent of the umpires are

White implies that the *measured* productivity of minority 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 expression of 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. Indeed, these findings imply that the particular impacts of racial/ethnic match preferences in baseball may now have been vitiated, since beginning in 2009 all ballparks are equipped with QuesTec or similar technologies.²⁶ Clearly, however, raising the price of discrimination in the labor market generally is more difficult; but our results may suggest analogous measures that might have the desired effects.

²⁶ See "Ball-Strike Monitor May Reopen Wounds" (Alan Schwarz, *New York Times*, March 1, 2009, electronic version available at <http://www.nytimes.com/2009/04/01/sports/baseball/01umpires.html>).

REFERENCES

- Blair, Irene V.** “The Malleability of Automatic Stereotypes and Prejudice.” *Personality and Social Psychology Review*, 2002, 6(3), pp. 242-61.
- Bergmann, Barbara R.** “The Effect on White Incomes of Discrimination in Employment.” *Journal of Political Economy*, 1971, 79(2), pp. 294-313.
- Bradbury, J.C.** *The Baseball Economist: The Real Game Exposed*. New York: Dutton, 2007.
- Cain, Glen G.** “The Economic Analysis of Labor Market Discrimination: A Survey.” In Ashenfelter, Orley C. and Layard, P. Richard G. *Handbook of Labor Economics, Vol. 1*. Amsterdam: North-Holland, 1986.
- Chen, Jowei.** “What Does Baseball Teach Us About Reducing Racial Discrimination? Evidence From Two Natural Experiments.” Stanford University working paper, 2007.
- Coate, Stephen and Loury, Glenn C.** “Will Affirmative-Action Policies Eliminate Negative Stereotypes?” *American Economic Review*, 1993, 83(5), pp. 1220-40.
- Dohmen, Thomas J.** “The Influence of Social Forces; Evidence from the Behavior of Football Referees.” *Economic Inquiry*, 2008, 46(3), pp. 411-24.
- Donald, Stephen G. and Hamermesh, Daniel S.** “What Is Discrimination? Gender in the American Economic Association.” *American Economic Review*, 2006, 96(4), pp. 1283-92.
- Findlay, David W. and Reid, Clifford E.** “Voting Behavior, Discrimination and the National Baseball Hall of Fame.” *Economic Inquiry*, 1997, 35(3), pp. 562-78.
- Fort, Rodney D. and Gill, Andrew M.** “Race and Ethnicity Assessment in Baseball Card Markets.” *Journal of Sports Economics*, 2000, 1(1), pp. 21-38.
- Garicano, Luis; Palacios-Huerta, Ignacio and Prendergast, Canice.** “Favoritism under Social Pressure.” *Review of Economics and Statistics*, 2005, 87(2), pp. 208-16.
- Gius, Mark P. and Hylan, Timothy R.** “An Interperiod Analysis of the Salary Impact of Structural Changes in Major League Baseball: Evidence from Panel Data.” In *Baseball Economics, Current Research*. Westport CT: Greenwood, Praeger, 1996.
- Goodwin, Stephanie A.; Gubin, Alexandra; Fiske, Susan T. and Yzerbyt, Vincent Y.** “Power Can Bias Impression Processes: Stereotyping Subordinates by Default and By Design.” *Group Processes and Intergroup Relations*, 2000, 3(3), pp. 227-56.

- Gwartney, James D. and Haworth, Charles T.** “Employer Costs and Discrimination: The Case of Baseball.” *Journal of Political Economy*, 1974, 82(4), pp. 873-81.
- Kahn, Lawrence M.** “Racial Discrimination in Professional Sports: A Survey of the Literature.” *Industrial and Labor Relations Review*, 1991, 44(3), pp. 395-418.
- , “Managerial Quality, Team Success and Individual Player Performance in Major League Baseball.” *Industrial and Labor Relations Review*, 1993, 46(3), pp. 531-47.
- Knowles, John; Persico, Nicola G. and Todd, Petra E.** “Racial Bias in Motor-Vehicle Searches: Theory and Evidence.” *Journal of Political Economy*, 2001, 109(1), pp. 203-29.
- Krautmann, Anthony C.; Gustafson, Elizabeth F. and Hadley, Lawrence.** “A Note on the Structural Stability of Salary Equations: Major League Baseball Pitchers.” *Journal of Sports Economics*, 2003, 4(1), pp. 56-63.
- Nardinelli, Clark and Simon, Curtis J.** “Customer Racial Discrimination in the Market for Memorabilia: The Case of Baseball.” *Quarterly Journal of Economics*, 1990, 105(3), pp. 575-95.
- Pascal, Anthony H. and Rapping, Leonard A.** “The Economics of Racial Discrimination in Organized Baseball.” In Pascal, Anthony H., ed., *Racial Discrimination in Economic Life*. Lexington, MA: Heath, 1972.
- Persico, Nicola G.** “Racial Profiling, Fairness, and Effectiveness of Policing.” *American Economic Review*, 2002, 92(5), pp. 1472-97.
- Price, Joseph and Wolfers, Justin J.** “Racial Discrimination Among NBA Referees.” National Bureau of Economic Research, Working Paper No. 13206, June 2007.
- Scully, Gerald W.** “Pay and Performance in Major League Baseball.” *American Economic Review*, 1974, 64(6), pp. 915-30.
- Stoll, Michael A.; Raphael, Steven and Holzer, Harry J.** “Black Job Applicants and the Hiring Officer’s Race.” *Industrial and Labor Relations Review*, 2004, 57(2), pp. 267-87.
- Zitzewitz, Eric W.** “Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making.” *Journal of Economics and Management Strategy*, 2006, 15(1), pp. 67-99.

Table 1–Summary Statistics of Pitches

Pitch Outcomes, 2004-2008 (percentage distributions)									
	Total Pitches	Called Strike	Called Ball	Swinging Strike	Foul	In Play	Intentional Ball	Hit by Pitch	
All	3,524,624	17.09	36.56	8.98	17.08	19.41	0.63	0.25	
Pitcher									
White(N=861)	2,544,515	17.19	36.48	8.77	17.10	19.58	0.64	0.24	
Hispanic(N=278)	793,797	16.86	36.77	9.57	17.03	18.86	0.64	0.27	
Black(N=37)	89,355	16.24	36.68	9.71	17.54	19.07	0.52	0.24	
Asian(N=39)	96,957	17.12	36.81	8.87	16.59	19.70	0.64	0.27	
Batter									
White(N=1,147)	1,813,768	17.37	36.90	9.11	16.92	18.84	0.58	0.28	
Hispanic(N=493)	1,061,115	16.81	35.91	8.72	17.35	20.31	0.68	0.22	
Black(N=187)	571,563	16.65	36.67	9.17	17.08	19.50	0.70	0.23	
Asian(N=46)	78,178	17.63	36.80	7.44	17.35	19.88	0.71	0.19	
Umpire									
White(N=91)	3,215,949	17.09	36.55	8.97	17.09	19.41	0.64	0.25	
Hispanic(N=5)	111,524	17.06	36.80	8.87	16.97	19.33	0.70	0.27	
Black(N=6)	197,151	17.13	36.63	9.00	16.99	19.44	0.59	0.22	

Table 2—Summary of Umpires’ Called Pitches by Umpire-Pitcher Racial/Ethnic Match, MLB 2004-2008

	Pitcher Race/Ethnicity				Total percent called strikes
	White	Hispanic	Black	Asian	
Umpire Race/Ethnicity					
White					
Pitches	2,319,522	726,137	81,251	89,039	
Called pitches	1,244,523	389,411	42,986	47,973	
Called strikes	398,673	122,441	13,194	15,269	
Percent called strikes	32.03	31.44	30.69	31.83	31.86
Hispanic					
Pitches	80,956	24,844	2,559	3,165	
Called pitches	43,632	13,299	1,374	1,760	
Called strikes	13,857	4,194	429	549	
Percent called strikes	31.76	31.54	31.22	31.19	31.68
Black					
Pitches	144,037	42,816	5,545	4,753	
Called pitches	77,472	23,035	2,922	2,561	
Called strikes	24,900	7,195	886	784	
Percent called strikes	32.14	31.24	30.32	30.61	31.86
Total percent called strikes	32.03	31.43	30.69	31.75	31.86

Table 3—Effects on Called Strikes of the Relationship between Pitcher and Umpire Race/Ethnicity, MLB 2004-2008

A. Main Parameter Estimates									
Pitchers	<i>White</i>	<i>Black</i>	<i>Hispanic</i>	<i>All</i>	<i>All</i>	<i>All</i>	<i>All</i>	<i>All</i>	<i>All</i>
Umpires	<i>All</i>	<i>All</i>	<i>All</i>	<i>White</i>	<i>Black</i>	<i>Hispanic</i>	<i>All</i>	<i>All</i>	<i>All</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Black Umpire	-0.0005 (0.0019)	0.0004 (0.0105)	-0.0010 (0.0031)						
Hispanic Umpire	-0.0045 (0.0024)	0.0097 (0.0127)	0.0079 (0.0049)						
Black Pitcher				-0.0148 (0.0023)	-0.0157 (0.0103)	-0.0027 (0.0125)			
Hispanic Pitcher				-0.0072 (0.0009)	-0.0089 (0.0034)	0.0020 (0.0054)			
UPM							0.0024 (0.0013)	0.0021 (0.0017)	0.0016 (0.0017)
Observations	1,365,660	47,285	425,731	1,676,942	103,429	58,305		1,838,676	
R²	0.031	0.031	0.030	0.028	0.025	0.030	0.031	0.091	0.091
Fixed Effects	P	P	P	U	U	U	P	PU	PUB

B. Coefficients on Pitch Count Indicators in the Specification in Column (9)

<i>0&1</i>	<i>0&2</i>	<i>1&0</i>	<i>1&1</i>	<i>1&2</i>	<i>2&0</i>	<i>2&1</i>	<i>2&2</i>	<i>3&0</i>	<i>3&1</i>	<i>3&2</i>
-0.226 (0.001)	-0.355 (0.001)	-0.023 (0.001)	-0.190 (0.001)	-0.326 (0.001)	0.042 (0.002)	-0.151 (0.002)	-0.289 (0.001)	0.206 (0.003)	-0.060 (0.003)	-0.257 (0.002)

C. Coefficients on Inning Indicators and Pitcher's Score Advantage in the Specification in Col. (9)

<i>2nd</i>	<i>3rd</i>	<i>4th</i>	<i>5th</i>	<i>6th</i>	<i>7th</i>	<i>8th</i>	<i>9th</i>	<i>Top of Inning</i>	<i>Pitcher's Score Advantage</i>
-0.010 (0.001)	-0.024 (0.001)	-0.032 (0.001)	-0.032 (0.001)	-0.034 (0.001)	-0.025 (0.001)	-0.024 (0.002)	-0.018 (0.002)	0.006 (0.001)	0.002 (0.0004)

Notes: All estimates are based on linear-probability models with heteroskedasticity-robust standard errors in parentheses, here and in Tables 4-6. UPM indicates whether the umpire and pitcher match on race/ethnicity. The control variables whose coefficients are reported Panels B and C are included in all the estimates. Pitcher Score Advantage is the number of runs, potentially negative, that the pitcher's team is ahead at the time of the pitch. Top of Inning is an indicator equaling 1 if the home team is pitching. P, U and B represent pitcher, umpire and batter fixed effects, respectively.

Table 4—Effects on Called Strikes of Explicit Monitoring of Umpires and Racial/Ethnic Discrimination, MLB 2004-2008

	<i>QuesTec</i>	<i>Non-QuesTec</i>	<i>All</i>
	(1)	(2)	(3)
Umpire-Pitcher Match (UPM)	-0.0048 (0.0027)	0.0059 (0.0022)	0.0059 (0.0022)
QuesTec*UPM			-0.0107 (0.0035)
Observations	679,979	1,158,697	1,838,676
R²	0.089	0.088	0.088

Notes: UPM indicates whether the umpire and pitcher match on race/ethnicity. All columns here and in Tables 5 and 6 include the control variables shown in Table 3. All columns also include fixed effects: 1) pitcher-QuesTec fixed effects, i.e., two fixed effects for each pitcher who pitched in both a ballpark where QuesTec was and was not installed; 2) umpire-QuesTec fixed effects, and 3) batter-QuesTec fixed effects.

Table 5—Effects on Called Strikes of Implicit Monitoring of Umpires and Racial Discrimination, MLB 2004-2008

A. Distinguishing by Game Attendance						
	<i>High Attendance</i>	<i>Low Attendance</i>	<i>All Games</i>			
	<i>(1)</i>	<i>(2)</i>	<i>(3)</i>			
UPM	-0.0034 (0.0025)	0.0064 (0.0024)	0.0036 (0.0019)			
Well Attended (>69% capacity)			0.0059 (0.0012)			
Well Attended*UPM			-0.0037 (0.0015)			
Observations	902,261	936,415	1,838,676			
R ²	0.089	0.088	0.088			
B. Distinguishing by Terminal Count and Inning						
	<i>Terminal</i>	<i>Non-Terminal</i>	<i>All Pitches</i>	<i>Early Inning</i>	<i>Late Inning</i>	
	<i>(4)</i>	<i>(5)</i>	<i>(6)</i>	<i>(7)</i>	<i>(8)</i>	
UPM	-0.0026 (0.0027)	0.0031 (0.0021)	0.0031 (0.0018)	0.0044 (0.0031)	0.0023 (0.0022)	
Terminal Count *UPM			-0.0058 (0.0014)	-0.0086 (0.0024)	-0.0038 (0.0017)	
Observations	427,136	1,411,540	1,838,676	641,053	1,197,623	
R ²	0.175	0.042	0.088	0.095	0.085	
C. Combining Explicit and Implicit Monitoring Proxies						
	UPM Interacted with				Observations	R²
	UPM	Questec	Well Attended	Terminal Count		
<i>All Pitches</i> <i>(9)</i>	0.0089 (0.0024)	-0.0102 (0.0036)	-0.0035 (0.0015)	-0.0058 (0.0014)	1,838,676	0.088

Note: Low (high) attendance games are defined as games with percentage attendance below (above) the median. A terminal count is defined as a count with three balls and/or two strikes.

Table 6–Pitch Location, Type and the Effects of Pitcher-Umpire Racial/Ethnic Matches, MLB 2007-2008

A. Pitch Locations (Distance from Home-plate Center), 2007-2008, N=538,194							
	Mean	Quantiles:	5 th	25 th	50 th	75 th	95 th
Horizontal Pitch Distance (ft.)	-0.04		-1.53	-0.68	-0.04	0.60	1.44
Pitch Height (ft.)	-0.11		-1.60	-0.69	-0.10	0.48	1.36

B. Percentage Distributions of Pitches by Type, 2007-2008, N=533,150

<i>By Location</i>			<i>By Type</i>				
<i>Inside</i>	<i>Edge</i>	<i>Outside</i>	<i>Change-up</i>	<i>Curveball</i>	<i>Fastball</i>	<i>Slider</i>	<i>Other</i>
39.55	19.98	40.47	13.43	10.88	57.48	13.52	4.69

C. Effects on Probability of Pitch in the Edge of the Strike Zone

	<i>Questec</i> (1)	<i>Non-Questec</i> (2)	<i>All Games</i> (3)
UPM	-0.0005 (0.0047)	0.0095 (0.0042)	0.0095 (0.0042)
Questec*UPM			-0.0102 (0.0063)
Observations	199,085	339,109	538,194
R ²	0.001	0.001	0.001

D. Effects on Probability of a Curve Ball

	<i>Questec</i> (4)	<i>Non-Questec</i> (5)	<i>All Games</i> (6)
UPM	0.0033 (0.0032)	0.0128 (0.0029)	0.0125 (0.0028)
Questec*UPM			-0.0087 (0.0043)
Observations	195,777	337,373	533,150
R ²	0.020	0.021	0.021

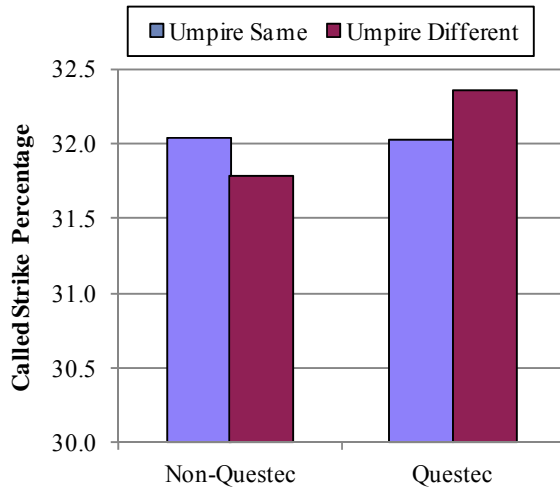
Notes: The sample consists of all pitches (called and non-called, excluding intentional balls) thrown by starting pitchers. In Panel A the pitch location is the Cartesian coordinate, where the origin is the intersection of the vertical line from the center of the home plate and the horizontal line equidistant to the top and the bottom of the strike zone. The information is from PITCHf/x.

Table 7—Estimated Effects on Performance of Umpire and Starting Pitcher Racial/Ethnic Match, N=12,127 Games, MLB 2004-2008

A. White Pitchers					
	<i>Umpire-Pitcher Racial Match</i>	<i>N</i>	<i>Win</i>	<i>Hits Allowed</i>	<i>Runs Allowed</i>
Questec	Match	5,953	0.347	6.190	3.215
	Non-Match	605	0.377	6.109	3.179
	Diff		-0.030 (0.021)	0.081 (0.102)	0.036 (0.092)
Non-Questec	Match	10,491	0.351	6.174	3.154
	Non-Match	1,003	0.334	6.240	3.234
	Diff		0.017 (0.016)	-0.066 (0.073)	-0.080 (0.069)
	Diff-in-Diff		-0.046 (0.026)	0.147 (0.126)	0.116 (0.115)
B. Minority Pitchers					
	<i>Umpire-Pitcher Racial Match</i>	<i>N</i>	<i>Win</i>	<i>Hits Allowed</i>	<i>Runs Allowed</i>
Questec	Match	74	0.257	6.284	3.581
	Non-Match	2,313	0.356	6.006	3.179
	Diff		-0.099 (0.052)	0.278 (0.297)	0.402 (0.276)
Non-Questec	Match	119	0.370	5.891	3.185
	Non-Match	3,696	0.340	6.080	3.223
	Diff		0.030 (0.045)	-0.189 (0.226)	-0.038 (0.214)
	Diff-in-Diff		-0.129 (0.069)	0.466 (0.373)	0.440 (0.349)

Notes: Standard errors in parentheses.

A. White Pitchers



B. Minority Pitchers

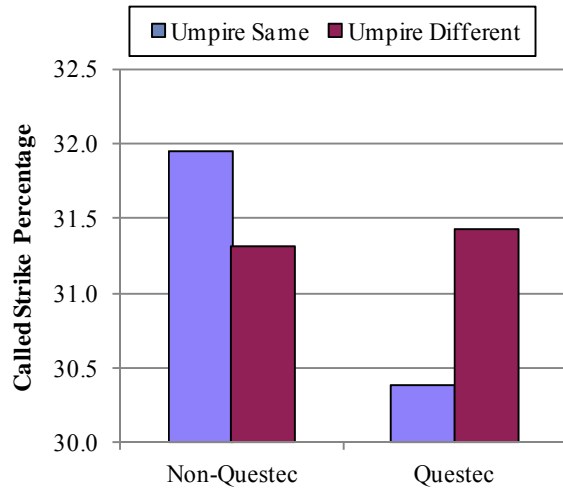
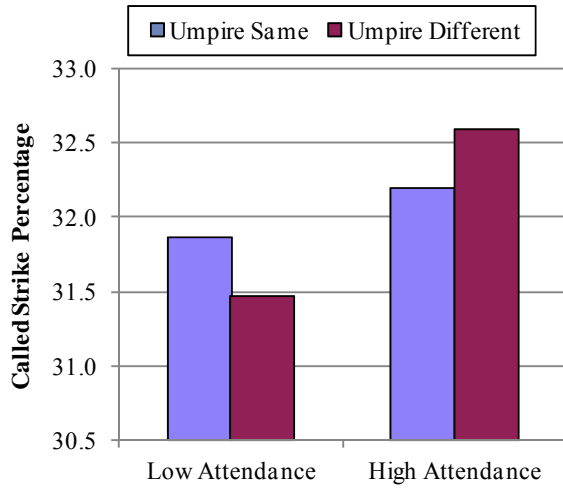


Figure 1–Race and Called Strike Percentages in QuesTec and Non-QuesTec Ballparks

A. White Pitchers



B. Minority Pitchers

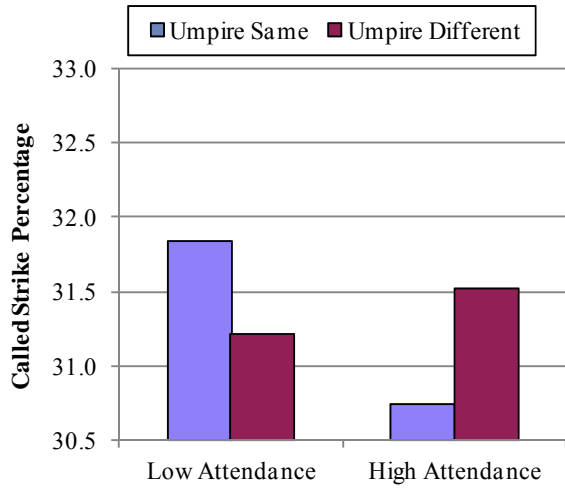
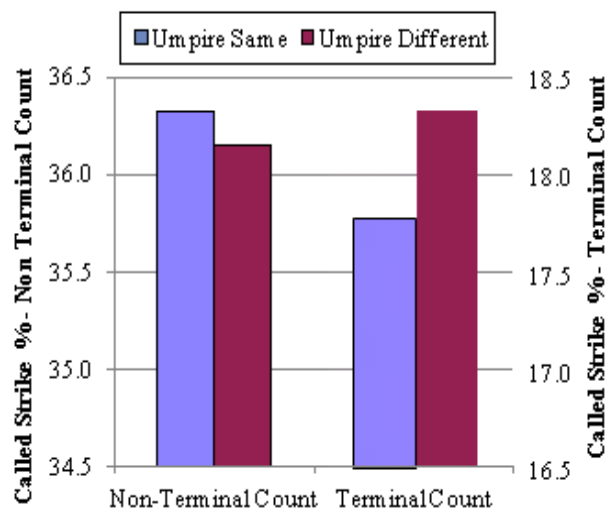


Figure 2—Race and Called Strike Percentages by Game Attendance

Note: Low (high) attendance games are defined as games with percentage attendance below (above) the median.

A. White Pitchers



B. Minority Pitchers

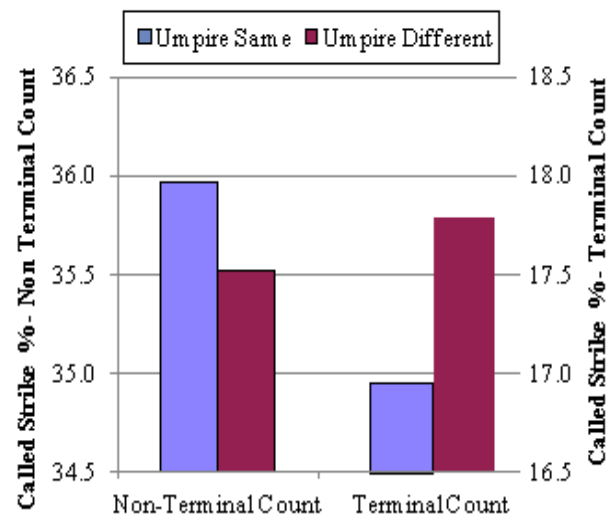


Figure 3–Race and Called Strike Percentages in Terminal and Non-Terminal Counts

Note: A terminal count is defined as a count with three balls and/or two strikes.

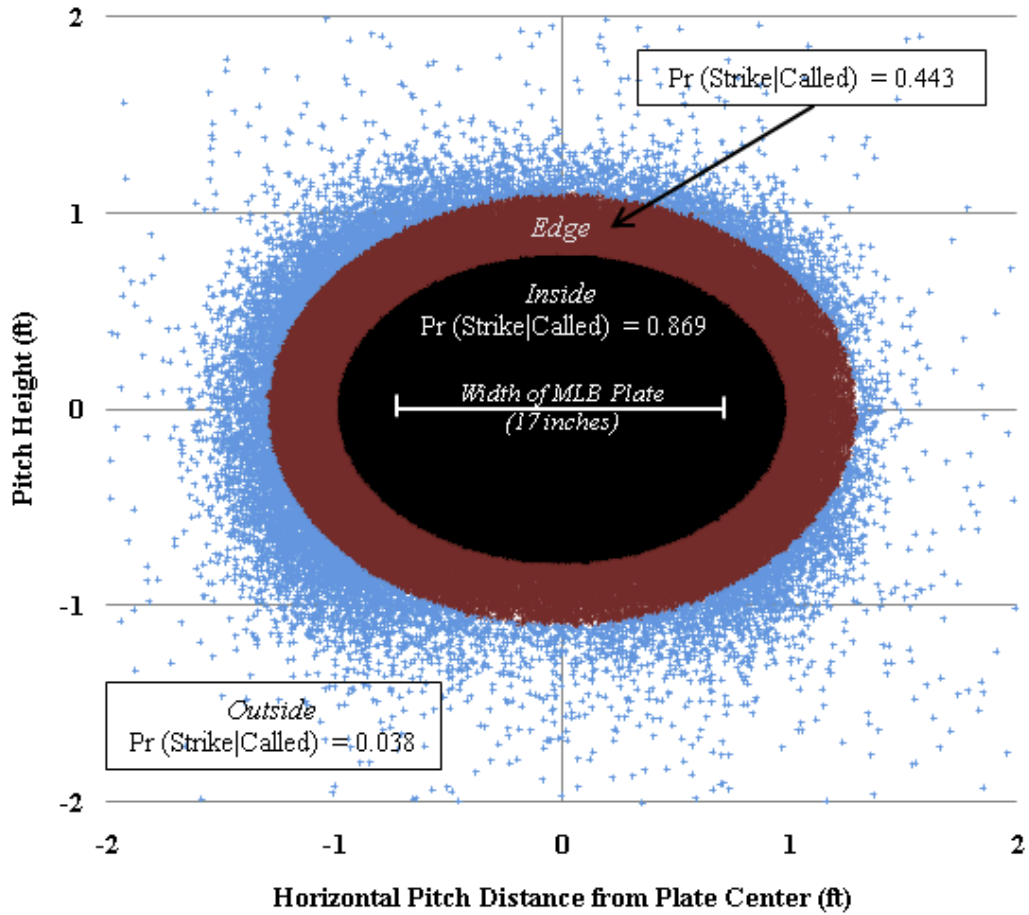


Figure 4—Called Strikes by Distance from Home-Plate Center, 2007-2008 (N=144,990)

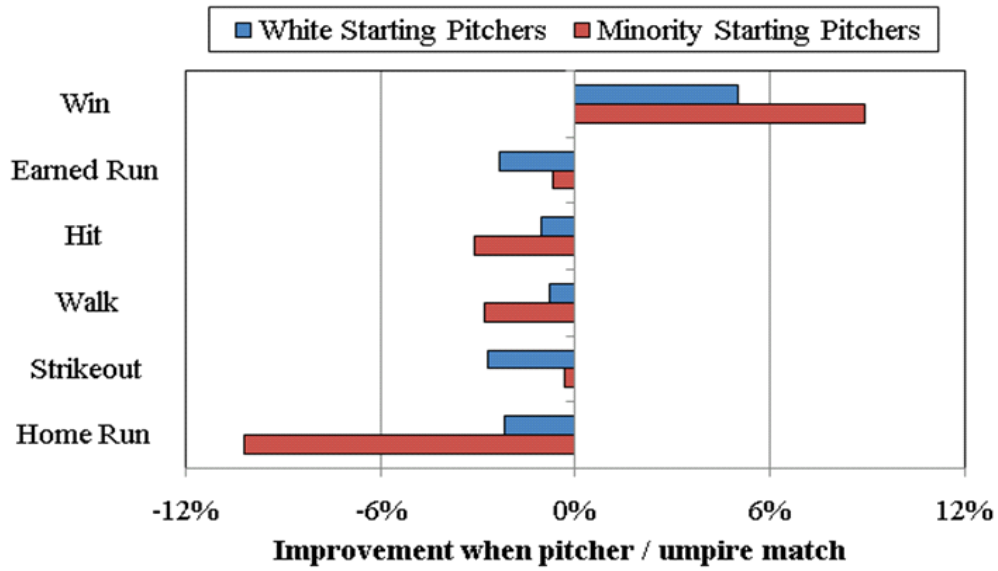


Figure 5—Effects of Umpire-Pitcher Racial/Ethnic Match on Pitcher Performance, non-Questec Ballparks, MLB 2004-2008 (N=15,308)

Appendix– A Model of Bias-Induced Changes in Pitcher Strategies

Consider the following simple representation of the interaction between the pitcher and hitter. Denote the horizontal distance from the center of the plate \tilde{y} . Assume for simplicity that the pitcher can control the width of pitches (i.e., the horizontal dimension), but not their height. Further suppose that the batter is left-handed, and that the pitcher never aims left of center, i.e., $\mu = 0$. This simplifying assumption is little more than a normalization, although a realistic one, as pitchers are usually cautious to avoid hitting the batter.

The game unravels as follows.

1. The pitcher moves first. He can select his aim, $\mu \geq 0$, but not the final pitch location, \tilde{y} , which is random. On average, the pitcher's aim is correct, i.e., $E(\tilde{y}) = \mu$.
2. The batter moves next. A batter must decide whether to swing or not soon after a pitch is thrown, but before it reaches its final location \tilde{y} . To capture this timing, the batter's swing decision is made immediately after observing μ .²⁷
3. If the batter does not swing, two outcomes are possible. For a given value of μ , with probability $s(\mu)$, the pitch is called a strike, and confers the batter a payoff S . With probability $1 - s(\mu)$, the pitch is called a ball, with payoff $B > S$. We assume $s' < 0$, $s'' < 0$, i.e., that pitches aimed closer to the plate are more likely to be called strikes, and at an increasing rate.
4. If the batter swings, two additional outcomes are possible. With probability $h(\mu)$, the batter gets a hit, and enjoys a payoff H . With probability $1 - h(\mu)$, the batter does not get a hit, with payoff $N < H$.²⁸ Similar to the assumptions for s , we assume $h' < 0$, $h'' < 0$.

The Batter's Problem:

To determine whether he swings at a pitch with expected location μ , the batter compares his expected payoff from swinging,

$$\pi(\text{swing}|\mu) = h(\mu)H + [1 - h(\mu)]N \tag{1}$$

²⁷This strict timing assumption is not crucial. Instead, it is a simplified way of modeling that the batter makes his swing decision under imperfect information. For example, the batter could instead observe a noisy signal of \tilde{y} without changing the results.

²⁸ Here, N captures the average payoff of swinging and missing, (S), and hitting into an out.

with that from not swinging,

$$\pi(\text{no swing}|\mu) = s(\mu)S + [1 - s(\mu)]B. \quad (2)$$

Lemma 1. *Assume $\pi(\text{swing} | \mu=0) > \pi(\text{no swing} | \mu=0)$, so that a batter always prefers to swing at a pitch aimed down the center of the plate. Then there exists a unique cutoff M whereby if: i) $\mu < \hat{\mu}$, the batter strictly prefers to swing, ii) $\mu > \hat{\mu}$, the batter strictly prefers to not swing, and iii) $\mu = \hat{\mu}$, the batter is indifferent between swinging and not.*

Proof. $\partial(\pi(\text{swing}|\mu)) / \partial\mu < 0$, which follows from the assumptions that: i) called strikes are assumed to be more likely when thrown closer to the plate, $s' < 0$, ii) the batter's expected payoff from called balls is higher than that from called strikes, $B > S$. By similar logic, $\partial(\pi(\text{no swing}|\mu)) / \partial\mu > 0$. The convexity assumptions s'' , $h'' < 0$ then guarantee a single crossing for $(\pi(\text{swing} | \mu))$ and $(\pi(\text{no swing} | \mu))$, which we denote $\hat{\mu}$.

The intuition for Lemma 1 is straightforward. Batters will not attempt to hit pitches that have very little chance of being called a strike should they not swing, i.e., for sufficiently low values of μ . Moreover, the cutoff for swinging $\hat{\mu}$ is a function of the payoffs S , B , H , and N that correspond to the possible outcomes of the plate appearance. Generally, these payoffs will depend on game conditions, such as the score, the count, runners on base, or other factors that determine the payoffs to each outcome. For example, with runners on second and third base but no outs, the benefit of a hit (H) is substantial, where the cost of hitting into an out (N) is relatively small. In this situation, the batter will be less selective at the plate, which increases the swinging cutoff $\hat{\mu}$. We do not model differences in these payoffs across plate appearances, although the present set-up easily allows for this extension.

Our main interest is in how changes in the conditional strike function, $s(\mu)$, influence the batter's optimal behavior. Specifically, assume that the race/ethnicity match of the umpire and pitcher influences the probability that a pitch aimed at location μ will be called a strike. If the pitcher and umpire match (M), denote the conditional called strike probability $s_M(\mu)$. If they are different (D), the conditional strike probability becomes $s_D(\mu)$. To capture the idea that similar race or ethnicity helps the pitcher, we assume:

$$s_M(\mu) > s_D(\mu), \quad \forall \mu \quad (3)$$

In other words, the same pitch has a different probability of being called a strike, conditional on whether the umpire and pitcher have the same or different races or ethnicities.

Lemma 2. *When the pitcher and umpire share the same race/ethnicity, the batter swings at pitches further from the center of the plate. That is, the cutoff location under a match is strictly greater than the cutoff location otherwise, i.e., $\hat{\mu}_M > \hat{\mu}_D$.*

Proof. *Denote $\hat{\mu}_M$ as the cutoff swinging location when $s(\mu) = s_M(\mu)$ and $\hat{\mu}_D$ as that when $s(\mu) = s_D(\mu)$. Suppose $s(\mu) = s_M(\mu)$ and $\mu = \hat{\mu}_M$. From equation (2), when $s(\mu)$ changes to $s_D(\mu)$ the expected payoff of not swinging declines by $[s_M(\hat{\mu}_M) - s_D(\hat{\mu}_M)](S - B) > 0$, while the payoff from swinging is unchanged. We can now use the proof for Lemma 1. Because $\partial(\pi(\text{swing} | \mu)) / \partial\mu < 0$ and $\partial(\pi(\text{no swing} | \mu)) / \partial\mu > 0$, the new cutoff $\hat{\mu}_D$ is strictly less than $\hat{\mu}_M$.*

Lemma 2 indicates that when the batter anticipates judgments that favor the pitcher, his optimal strategy changes. Expecting the umpire's bias to reduce his payoff from not swinging, the batter takes matters into his own hands by swinging at pitches that he would otherwise let pass. Empirically, this implies a distinct advantage to the pitcher, not only for pitches that are called, but also for pitches that are hit. We complete this exercise by extending consideration to the pitcher's optimal strategy.

The Pitcher's Problem:

The pitcher's choice variable is μ , the expected location of the pitch. His expected payoff is the inverse of the batter's. If the batter swings, then the pitcher's expected payoff is $-h(\mu)H - [1 - h(\mu)]N$. If the batter does not swing, then his expected payoff is $-s(\mu)S - [1 - s(\mu)]B$.

Lemma 3. *The pitcher's optimal pitch location is $\hat{\mu}$, so that the batter is indifferent between swinging and not.*

Proof. *The batter will swing at any pitch aimed at $\mu < \hat{\mu}$, but because $\partial(\pi(\text{swing} | \mu)) / \partial\mu < 0$, the pitcher is always strictly better off increasing μ given that the batter will swing. The batter will not swing at any pitch aimed at $\mu > \hat{\mu}$, but because $\partial(\pi(\text{no swing} | \mu)) / \partial\mu > 0$, the pitcher will always decrease μ given that the batter will not swing. It follows then that the optimal pitch location must be $\hat{\mu}$.*

The main prediction is that the umpire's bias influences not only called strikes and balls, but also pitches where the umpire's judgment plays no direct role. Lemma 3 shows that the umpire's judgment influences the choice of pitch location, which in turn influences the batter's incentive to swing at the ball. It follows that *conditional on swinging*, the batter is less likely to hit the ball when the umpire and pitcher share race or ethnicity. As indicated by the model, this is because pitches are, on average, more difficult to hit in these situations.