View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Munich Personal RePEc Archive

MPRA
Munich Personal RePEc Archive

Further Examination of Potential Discrimination Among MLB Umpires

Scott Tainsky and Brian Mills and Jason A. Winfree

University of Illinois, University of Florida, University of Michigan

11. December 2012

Online at https://mpra.ub.uni-muenchen.de/43234/
MPRA Paper No. 43234, posted 12. December 2012 19:16 UTC

Further Examination of Potential Discrimination Among MLB Umpires

Scott Tainsky, Brian Mills, and Jason A. Winfree*

*Tainsky (corresponding author): Assistant Professor, Department of Recreation, Sport and Tourism, University of Illinois, tainsky@illinois.edu; Mills: Assistant Professor, Department of Tourism, Recreation, and Sport Management, University of Florida; Winfree: Associate Professor, Program in Sport Management, University of Michigan.

We would like to thank W. David Allen, Charles Brown, John DiNardo, Rodney Fort, Daniel Hamermesh, Christopher Parsons and Mark Rosentraub for helpful comments and insights. We would also like to thank Christopher Parsons for sharing their data.

Abstract

We address potential racial bias by Major League Baseball umpires with respect to ball-strike calls. We offer a number of econometric specifications to test the robustness of the results, adding the role of implicit and explicit monitoring as well as pitch location. Our analysis shows mixed results regarding the matching of umpire and pitcher race. We conclude that evidence of own race bias is sensitive to specification and methodology. How results can differ based on different data sets, specifications, time periods and race classifications are discussed.

Keywords: Discrimination, Race, Baseball, Monitoring

Introduction

One of the principal functions of any league is to establish the proverbial "level playing field." It seems straightforward that the games themselves ought to be officiated similarly for each competitor and team. As of late, officiating has come under increasing scrutiny in major league sports, with much of this scrutiny coming from the economics literature (e.g., Garicano, Palacios-Huerta, & Prendergast, 2005; Parsons, Sulaeman, Yates, & Hamermesh, 2011; Price & Wolfers 2010; Sutter & Kocher, 2004). While some rules that govern play are highly subjective and therefore difficult to evaluate an official's performance and/or biases, other rules are clear in their definition of how to be enforced.

A basic baseball statute that would seem to be unproblematic to interpret literally is the calling of balls and strikes. According to Major League Baseball (MLB), the strike zone is "that area over home plate the upper limit of which is a horizontal line at the midpoint between the top of the shoulders and the top of the uniform pants, and the lower level is a line at the hollow beneath the knee cap." A ball is simply "a pitch which does not enter the strike zone in flight and is not struck at by the batter," (MLB Official Rules, n.d.) and any pitch that is not struck at by the batter and is not a ball is called a strike. Yet, in spite of this straightforward definition, often announcers and/or pundits will speak to whether the home plate umpire has a wide strike zone, a tight strike zone, or even a high strike zone. These considerations seem to contradict the charge of the umpires, to see that the contest is played under strict adherence to the rules, but do not necessarily present any direct unfairness towards particular players or groups of players. The focus of this study is to gauge whether the calling of balls and strikes has been systematically applied differently for different players. Specifically, if it were the case that there existed a pattern of inequitable application of an objective rule like the calling of balls and strikes according to a player's race, this would not only be a case of direct unfairness, but discrimination.

In the sections following we review the literature and then describe our empirical estimation. We continue by detailing our analysis and the results of regression estimation for interaction effects between monitoring and race-matching of the umpire and pitcher. Finally, we address pitch location using the MLB Gameday Pitch f/x data set. These findings are summarized in the final section, and we conclude with recommendations for future research.

Review of Literature

As Groothuis and Hill (2011) stated, "(p)roclomations of racial discrimination always elicit notoriety. Findings of no discrimination do not procure the same response. Therefore, it is important that any positive findings of racially unequal treatment be particularly robust," (p.2). Research on discrimination in sports has been prevalent for nearly a quarter-century. Although researchers have explored other leagues (e.g., DeBrock, Hendricks, & Koenker, 2003; Kahn, 1992; Jones & Walsh, 1988; Longley, 2000, 2003), the bulk of these studies among North American professional sports leagues examine discrimination issues in the NBA, perhaps due to the unique racial demographics of the league's athletes. The extant research on MLB discrimination has reconsidered many of the same issues first examined in the context of the NBA. The approach has been to apply Becker's (1971) model on labor discrimination as originating from consumer preferences, co-worker discrimination and/or employer prejudice to identify whether differences exist and, if so, the source thereof. Racial sorting or matching are often tested empirically to both explain differences in economic rents and/or relative performance levels between players of different races.

Given the wealth of research on discrimination in the NBA and the conceptual linkages to the narrower line of inquiry on subjective officiating, we begin there. In general the research supports the claim that some discrimination existed during the 1980s and began to disappear during the decade following. Kahn and Sherer (1988) examined salary determination in the NBA and found Blacks were paid roughly 20 percent less than Whites and, moreover, their results were robust to specification and estimation techniques. They also found that White players were associated with higher home attendance, but found no evidence of discrimination in the drafting of players into the league. Others found a lower (9-16 percent), albeit still statistically significant, premium for White NBA players during the same era (Brown, Spiro, & Keenan, 1988; Koch & VanderHill, 1988). Burdekin and Idson (1991) found strong support that demand was positively related to the extent of the team-market racial match (customer discrimination), while others argued that employer discrimination was at the source and was erased with reductions in monopsony power (Bodvarsson & Brastow, 1999). Dey (1997) similarly found that the racial salary gap narrowed in the early 1990s, but attributed the effect to consumers no longer differentiating between White and Black players. This question was later reexamined to find that White stars tend to land in markets with larger White populations (Burdekin, Hossfield, & Smith, 2005). This is consistent with the research finding of no statistical differences between the overall salaries of Whites and Blacks or at the lower end of the distribution, with White players receiving an 18% premium at the upper end of the salary distribution (Hamilton, 1997). Finally, Hill (2004) found no evidence of discrimination during the 1990s once height was entered into the pooled data, while Kahn and Shah (2005) found non-White shortfalls in salary for certain groups of players, but not across the population of NBA players.

Although much of the discrimination-based literature has focused on wages, others have studied the point of discrimination. Hoang and Rascher (1999) examined the role of exit discrimination and found White players faced a lower risk of being cut and therefore enjoyed longer careers and greater career earnings. Subsequent research found that this effect had also disappeared during the 1990s (Goothuis & Hill, 2004). Testing for both wage and exit discrimination, a recent study showed evidence of reverse discrimination as well as a White premium, but neither result was robust to specifications (Groothuis & Hill, 2011).

As stated earlier, most of the wage discrimination and points of discrimination findings reveal that the effects dissipated during the 1990s. With the declining effect, research increasingly turned to other sources of discrimination including coaches. Fort, Lee, and Berri (2008) revisited Hoang and Rascher's (1999) question of exit discrimination within the coaching ranks. They found neither differences in the technical efficiency of coaches, nor in the retention of coaches according to race. Schroffel and Magee (2011) discovered an own-race bias among NBA coaches. They found evidence that NBA coaches allotted greater amounts of playing time to players of their own race during the late 1990s, but that declined in the early 2000s.

Most closely related to the current study, the behavior of officials has been investigated across a range of sports and nations. Much of this research surrounds the question of referee home team bias. The theory goes that home crowd advantage represents a social pressure (Courneya & Carron, 1992). If the crowd can induce a physiological response in players, as was shown by Neave and Wolfson (2003), then it can also influence the decision of referees. Nevill, Newell and Gale (1996) examined the number of penalties awarded in English and Scottish football and found that home teams were awarded significantly more penalties than visitors. Sutter and Kocher (2004) tested this notion further without the assumption of equal probability of being awarded a penalty. They found that referees favored home teams in numerous ways, among them the tendency to award significantly extra time for an equalizer at the end of regulation when home teams trailed by exactly one goal and failure to award a significant number of legitimate penalties to the visiting team. Similarly, Garicano and colleagues (2005) found that referees shortened close soccer games when the home team was ahead and lengthened those where it trailed. Furthermore, they found referee bias was stronger with increased rewards for the home team and unusually high attendance.

Just as with respect to player discrimination issues, the examination of US-based league officiating in the NBA preceded that of MLB. Price and Wolfers (2010) argued that the split-second calls made by NBA referees allow implicit biases to surface that otherwise may go unchecked. In particular, they found that more personal fouls are awarded against players by opposite-race officiating crews than own-race crews. The results were sufficiently large to affect game and seasonal outcomes as well as the relative market value of Black versus White players.

Finally, a similar study to ours was undertaken by Parsons, Sulaeman, Yates, and Hamermesh (2011, hereafter PSYH) using data from 2004-2008. They focused on the presence of discrimination among umpires when matched with own-race and other-race pitchers, finding favorable decisions resulted from umpire-pitcher matches. Further, they showed that the effect vanished under the explicit monitoring conditions of the QuesTec evaluation system. Under implicit monitoring conditions, defined as pivotal pitches, pivotal at-bats or well-attended games, the effect again disappeared. Finally, the researchers contended that pitchers may adjust their strategies as a consequence of fair versus biased umpire treatment.

We believe our study presents several key differences with this paper. For starters, our data is from a different data source and covers more seasons (1997-2008). This can be significant given the clustering required to study the underrepresented groups of pitchers and umpires. Consequently, the relative weight of one outing is more likely to be felt and potentially skew the outcome with respect to underrepresented groups even though the whole data set may seem large. Also, given that race is not always clear, the two studies have different racial classifications. Finally, the two studies have different specifications, and thus different results.

Data

Data detailing every pitching performance in MLB from 1997-2008 was obtained from baseball-reference.com and Sportvision's public MLB Gameday database. Each observation covered a single pitching outing in our initial model, while data collected from Sportvision (2007 and 2008) included each individual pitch for our locational analysis. The information provided included the pitcher's name, plate appearances (batters faced), total pitches, total strikes, called strikes, strikes swinging, strikes in play (any batted ball in play is tallied as a strike), foul strikes, total balls, intentional balls, and the name of the home-plate umpire.

Player race was then determined by internet investigation. The race identification process began by searching for a player's profile on espn.com, and was completed when the researcher could confidently identify race. The primary considerations were a player's background information, including name origin, place of birth, and photos. Among the sources that figured prominently in these searches were ESPN's list of current African-American players, Wikipedia's list of Hispanic players, baseball-reference.com, mlb.com, and baseball-almanac.com. Two individuals independently researched each pitcher's background and classified the pitcher as White, Black, Hispanic, Asian, or any combination thereof. Players for whom there was not sufficient information or lacked the consensus of both researchers were omitted from the analysis. A similar process was utilized to classify umpire race; however, according to the same process no umpires were excluded in the data used for this analysis. MLB.com's umpire page served as the primary resource for this investigation.

The total number of strikes swinging, foul balls, strikes in play, and intentional balls were tabulated for each race in addition to the number falling into our categories of interest—strikes looking and unintentional balls. Tables 1 and 2 summarize these data over the entire sample and 2004-08 alongside the results from PSYH for comparison. The row percentages speak to what was outlined in the introduction—there is some subjectivity in the strike zone of different umpires. Hispanic umpires had the highest called strike percentage, an increase of 0.53% and 0.98% over our entire sample compared to White and Black umpires, respectively. Taking into account this discrepancy in called strike percentage among umpires, there was some consistency in the match of umpire and pitcher groups. White pitchers received the highest called strike percentage from all three groups of umpires, and two of three called the lowest strike percentage for Black pitchers (Black pitchers received a slightly higher called strike percentage than Asian pitchers, but still lower than White and Hispanic pitchers, from Hispanic umpires). Also of note is that even though the sample has nearly 8.3 million pitches, the number is reduced greatly when examining some combinations of umpires and pitchers only for pitches subject to judgment by umpires. Indeed, from 2004-2008, there were only around 2,550 pitches thrown by Black pitchers requiring the judgment of Black umpires.

During the time period of our data set, MLB implemented an umpire monitoring system known as QuesTec. This system allowed the league to evaluate its umpires' performance by tracking the location of the ball when it crossed the plate. This could explicitly change the cost of acting on any racial bias by the umpires across stadiums. The QuesTec system was not implemented in all stadiums in the league, allowing for comparison of explicitly monitored and unmonitored ball-strike calls for all umpires in our data.

Player-Umpire Interactions by Race

Table 3 uses a difference-in-differences (DID) analysis to estimate discrimination among combinations of umpire and pitcher race. Difference-in-differences actually gives evidence in favor of reverse discrimination. No matter the combination of umpire and pitcher (White/Hispanic, White/Black, Hispanic/Black)[1], the difference-in-differences

[1] There were no Asian umpires.

shows that umpires tended to be nearly neutral or favor pitchers of a different race when using data back through the 1997 season.[2]

Given the importance of monitoring (whether explicit or implicit), we also subdivided Table 2 and Table 3 for those stadiums with and without the QuesTec monitoring system in place (see Table 4). The difference-in-differences outcomes hint toward reverse discrimination both with and without QuesTec present. Finally, Figure 1 presents the estimated bias across explicit monitoring situations for White and minority pitchers separately, aggregating those pitchers who are non-White. We find no reversal pattern in discrimination behavior with White pitchers, but we do find a reversal with respect to minority pitchers. This result calls for more careful analyses, as the overall trend shown in Figure 1 is in some disagreement with the specific difference-in-differences in Table 4.

OLS was then used to regress the percentage of called strikes on different variables. Table 5 presents the results of this regression using the percentage of called pitches being strike as the dependent variable. The unit of observation in this case is a single pitcher outing. For example, if a team used three pitchers over the course of one game, then this would count as three observations. Because there is considerable variation in the duration of outing length, the regression was run for observations that had a minimum of 1, 10 and 50 called pitches. Additionally, the model was run with and without fixed effects for pitcher, umpire and year.[3]

[2] It is important to note that the DID analysis in Table 3, as well as our later analyses, are limited in that it is possible to find evidence of discrimination, it is not possible to tell who is discriminating.

[3] Batter fixed effects were not included, as the observations are at the pitcher-game level in order to account for in-game correlations, and pitchers face a number of different batters throughout the game.

We ran several versions of the model with the independent variable, *Match*, a binary indicator variable representing whether the umpire and pitcher are of the same race. *QuesTec* is an indicator variable representing whether the game was contested in a park fitted with the QuesTec system. This was included to evaluate whether the added scrutiny of the objective strike-gauging device influenced the called strike percentage. *Match*QuesTec* is a dummy variable at the intersection of the two previously described factors. It quantifies whether the outing took place in a park fitted with the QuesTec system and there was a match of umpire and pitcher race. The final indicator variable, *Home,* represents whether the pitcher outing was in his home park. It may be the case that pitchers are more familiar with the surroundings and are therefore better able to accurately locate pitches in their home parks. Whether this is the case is not the subject of this study, nonetheless it has been controlled for in the model. We note that the race of the batter is not available due to the aggregated pitcher-game structure of our data, but could have important implications for model estimations.

The results of the OLS regressions in Table 5 show little evidence of discrimination. The only variables significant for the fixed effects estimation are the *Home* and *QuesTec* variables using the larger sample period of 1997 to 2008, indicating that a higher percentage of pitches are called a strike for the home pitcher and a higher percentage of pitches are called strikes in stadiums fitted with the QuesTec system. The coefficient estimated for the *Match* variable is positive (indicating discrimination) and significant without fixed effects, but any *Match* effect is erased when fixed effects are used. This result arises from inclusion of pitcher fixed effects, likely due to a correlation between pitchers that throw a lot of strikes with *Match*.[4] Additionally, no coefficients within the regression are significant with the reduced sample from 2004-2008 in the fixed effects regressions.

These results are different from PSYH. While Table 1 and Table 2 show similar descriptive statistics to PSYH, Table 3 starts to show differences in the data. In Table 5 we find little evidence of discrimination or change in discrimination, with or without QuesTec, which differs from PSYH. Details of our attempt to reconcile the difference in results are given in the appendices. While data differences and racial classifications do account for part of the difference, the greatest disparity in the results is due to a difference in specification. The result that *Match*QuesTec* is significant relies heavily on using QuesTec-specific fixed effects. In other words, if each pitcher is given one fixed effect, then *Match*QuesTec* is not significant, but if each pitcher has two fixed effects, one in QuesTec stadiums and one in non-QuestTec stadiums, then the variable is significant.

Pitch f/x, Location and Agent Strategies

Recently more detailed pitch data has become available through the MLB Gameday Pitch f/x database. PSYH employ this data in order to evaluate changes in pitcher behavior due to umpire bias, and we follow suit here. Pitch f/x data is able to identify the location of a pitch as well as the velocity. It can also determine what type of pitch (e.g., fastball, curveball) was thrown. We collected pitch f/x locational data for part

[4] The results of all fixed effect combinations are available upon request.

of the 2007 season and all of the 2008 season.[5] An effect posited in previous research is that the cost of discrimination changes due to implicit (e.g., attendance) or explicit (e.g., QuesTec) monitoring and that this should be present both before and after the implementation of Pitch f/x. The argument made for pitcher-umpire race matched observations is the pitcher uses his knowledge of the umpire bias in his own favor, throwing more to the edges of the strike zone. However, Sportvision's pitch f/x system would seem to be a constant explicit monitoring of umpire performance given that the data are publicly available. Presumably the cost of discrimination does not change during the years in which pitch f/x data is available.

In order to estimate the called strike zone from the data, we employ a semi-parametric estimation of the pooled strike zone using a generalized additive model (GAM) and generalized cross-validation for estimation of the smoothing parameter for strike probability, given the pitch location. With this we were able to evaluate the spatial features of the strike zone and identify pitches near the 'edges' of the strike zone using a pooled estimation with all pitches. The smoothing technique allows fitting of a surface dependent on batter handedness, pitch location and batter height. Additionally, the flexible model can account for asymmetrical properties of the called strike zone, as opposed to the symmetrical ellipse used in PSYH. The asymmetry can be seen in Figure 2[6], with lower pitches more likely to be called strikes on the outside corner than the

[5] We restricted data to regular season games in regulation innings (1-9), those pitches which landed beyond the plate or above and within 2 feet on either side of the center of the rulebook strike zone. We also exclude any intentional balls, pitchouts, or unidentified pitches in the data.
[6] Figures are from the view of the umpire in position behind the plate, facing out at the pitcher delivering the ball.

inside corner, and significant shifts in the location of the zone for left and right handed batters.[7]

Using the predicted strike probabilities from the GAM, we defined an indicator variable equal to one if a pitch had a predicted probability of being called a strike between 40 and 60 percent (pooled GAM, see Figure 3).[8] We used this indicator to estimate a linear probability model (LPM) gauging the likelihood of a pitcher to throw to the edges of the strike zone under matched and non-matched pitcher-umpire race/ethnicity while controlling for all other variables within the data. These covariates included speed (in miles per hour), pitch type, inning, year, and the ball-strike count in which each pitch is thrown. We attempted other characterizations of the 'edge' of the zone—for example, between a 30 percent and 60 percent likelihood of a strike call—however, these did not affect the ultimate conclusions of the analysis.

In this model, positive coefficients indicate that a pitcher is more likely to throw to the edge of the strike zone, while negative coefficients would indicate that he is more likely to throw either well within or well outside the zone. Table 6 presents the results of an LPM estimation using this data. The only time pitchers change their propensity to throw within this 'edge' region with a high level of statistical significance is when the count is no balls and two strikes, one ball and two strikes, or with certain types of pitches. This makes sense, as pitchers often try to get the hitter to 'chase' an unhittable pitch when the count is no balls and two strikes, or one ball and two strikes. In this model, there is no significant effect of matched race between the umpire and pitcher, the

---

[7] As noted by a reviewer, this asymmetry could occur due to differentiated positioning of the umpire depending on the batter's handedness.

[8] For brevity, we do not go into detail regarding the GAM estimation of the strike zone; however, the computer code for this calculation and a full explanation can be provided upon request. For a full review of generalized additive modeling, the reader is referred to Wood (2000; 2003; 2004; 2011).

presence of the QuesTec system, or any significant interaction of matched race in the presence of QuesTec.

Additionally, the boundaries of the called strike zone—defined as the contour band at which a strike is no longer more likely to be called than a ball—are exhibited in Figure 2 across umpire judgment scenarios from a purely pitch location-based, non-parametric GAM for each scenario. These visuals compare matched and non-matched pitcher-umpire race/ethnicity across stadiums with and without the presence of QuesTec. As the reader can see, there is a relatively ambiguous relationship between the size of the called strike zone from one setting to the other, at least related to discrimination by MLB umpires.

We estimated a final model in order to lend further support to the edge-of-strike-zone model by evaluating the linear distance from the center of the strike zone that pitchers throw their pitches in each situation.[9] We begin with a fixed 2.6 foot height for the center of the strike zone and adjust this by each batter's height (moving the zone center up or down one-fifth of the difference between the batter's height in inches and the average height within the data set). We note that height is an inexact measure, as batters have varying stances; however, this measure is intended as a relatively consistent proxy for the strike zone height center. Using this measure and the horizontal center of home plate as the strike zone center, we calculate the linear distance from the center point of

[9] While PSYH use the upper and lower boundaries of the strike zone provided within Gameday's pitch f/x data (those input by the computer operator—or 'stringer'—at the time of the game) in order to evaluate whether a pitch was within or outside the strike zone, closer inspection reveals these measures are often inaccurate. In many cases, there is a wide (or even bimodal) distribution of 'upper' and 'lower' strike zone limits. Under this scenario, batter fixed effects may not be sufficient in dealing with the true strike zone center, and any correlation in the distribution across stringers at QuesTec and non-QuesTec parks could affect the coefficients in a regression.

the strike zone for each pitch when it crosses the plate, and use it as the dependent variable in the OLS estimation in Table 7.

We again find no evidence of pitchers throwing further from, or closer to, the center of the strike zone when they are matched in race or ethnicity with the umpire. The coefficient estimates indicate that pitchers tend to throw closer to the center of the strike zone in QuesTec parks, but with no significant effect of race matching or its interaction with the presence of QuesTec. This is consistent with the higher strike rates overall in QuesTec parks found in the primary estimations of this paper.

It is important to note that this data set includes a number of pitches that would certainly be called balls. This could affect coefficient estimates if many of these pitches are those that were errantly thrown, and not highly correlated with the pitcher's intended location. Therefore, we reduced the data for ancillary models including only pitches predicted to be called a strike from the pooled GAM estimation with a probability above 10%, 20%, 30%, 40% and 50%, respectively. The results of these models did not significantly impact any of the coefficients regarding race/ethnicity or presence of QuesTec from the initial estimation, and are not presented here.[10]

Conclusion

Findings of this research provide a challenge to the suggestions that there is racial or ethnic discrimination in the calling of balls and strikes by MLB umpires, and that pitchers react to that bias. Although portions of our data do not contain all of the control variables used in previous work, the analyses presented in this study demonstrate that the finding of MLB umpire discrimination is not particularly robust. We ran multiple

[10] The results of these models are available upon request.

estimations with various data sets and measures accounting for factors that could potentially explain variation in called strike percentage. With our data, only when pitcher, umpire and year effects were *not* accounted for was there any support for the notion that there was discrimination. We caution that even with multiple seasons and millions of pitches these data may still be subject to fluctuations in underrepresented umpire-pitcher matches, especially when these small subsets are divided into smaller ones with multiple fixed effects. Furthermore, the results may be sensitive to racial coding and/or different specifications.

Much of the empiricism of this and previous work centers on whether there are systematic differences in the calling of balls and strikes according to race and, in particular, match of umpire and pitcher race. In all of the above scenarios, the evidence for discrimination was mixed, at best, and at times signaled reverse-discrimination. We further evaluated whether any advantage (disadvantage) may have induced pitchers to approach their trade differently depending on umpire race. Again this hypothesis was unsupported.

These findings, of course, do not preclude the lessons of considering underlying discrimination in econometric estimations using supposedly objective performance data. However, we advocate proceeding with caution when categorical groups may be highly influenced by a small number of clustered observations, as in the data presented here. Despite the large sample size overall, the standard caveats of small samples nonetheless apply when even just a few observations have the potential to alone influence the presence or absence of an observed effect among the subsets of groups.

Furthermore, categorical race-coding is an inexact—and oftentimes severely biased—method of evaluating discrimination in this situation. In this case we adopt the elementary coding schema in order to mirror the previous research and replicate accordingly. How the findings of the current study would manifest under a more sophisticated racial coding method is an important area for future exploration. Further research in this area using more robust race classifications or measures, as exhibited by Fort and Gill (2000), would be a welcome addition to the present analysis.

While we stress that previous studies were rather extensive in their analyses, the secondary purpose of this study is to show that replication of the results is not straightforward given a different data set, racial classification, or specification. While some of our results are similar to previous work there are also differences. Investigation into these differences shows that the largest disagreements were due to specification dissimilarities. Other analysts have had trouble finding umpire discrimination as well (Birnbaum 2008). Given the sensitivity of racial discrimination, we argue, straightforwardly, counter evidence should be given equal weight.

References

Becker, G. (1971). *The Economics of Discrimination*, (2^{nd} ed.). Chicago, IL: University of Chicago Press.

Birnbaum, P. (2008). "Is There Really Racism Among MLB Umpires?" Power Point file retrieved from www.philbirnbaum.com.

Bodvarsson, O. B., & Brastow, R. T. (1999). A Test of Employer Discrimination in the NBA. *Contemporary Economic Policy, 17(2),* 243-255.

Brown, E., Spiro, R., & Keenan, D. (1988). Wage and Nonwage Discrimination in Professional Basketball: Do Fans Affect It? *American Journal of Economics and Sociology, 50(3),* 333-345.

Burdekin, R. C. K., Hossfield, R. T., & Smith, J. K. (2005). Are NBA Fans Becoming Indifferent to Race? Evidence from the 1990s. *Journal of Sports Economics, 6(2),* 144-159.

Burdekin, R. C. K., & Idson, T. L. (1991). Customer Preferences, Attendance and the Racial Structure of Professional Basketball Teams. *Applied Economics, 23(1),* 179-186.

Courneya, K. S., & Carron, A. V. (1992). The Home Advantage in Sport Competitions: A Literature Review. *Journal of Sport and Exercise Psychology, 14,* 13-27.

Debrock, L., Hendricks, W. E., & Koenker, R. (2003). Uncertainty, Hiring and Subsequent Performance: The NFL Draft. *Journal of Labor Economics, 21,* 857-886.

Dey, M. S. (1997). Racial Differences in National Basketball Association Players' Salaries: A New Look. *The American Economist, 41(2),* 84-90.

ESPN. http://espn.go.com/mlb/players. Last viewed on 8/27/07.

Fort, R. and Gill, A. (2000). Race and Ethnicity Assessment in Baseball Card Markets. *Journal of Sports Economics 1(1):*21-38.

Fort, R. D., Lee, Y. H., & Berri, D. (2008). Race, Technical Efficiency, and Retention: The Case of NBA Coaches. *International Journal of Sport Finance, 3,* 84-97.

Garicano, L., Palacios-Huerta, P. & Prendergast, C. (2005). Favoritism under social pressure. *The Review of Economics and Statistics, 87,* 208-216.

Groothuis, P. A., & Hill, J. R. (2004). Exit Discrimination in the NBA: A Duration Analysis of Career Length. *Economic Inquiry, 42(2),* 341-349.

Groothuis, P. A., & Hill, J. R. (2011). Pay Discrimination, Exit Discrimination or Both?: Another Look at an Old Issue Using NBA Data. *Journal of Sports Economics,* doi: 10.1177/1527002511418515

Hamilton, B. H. (1997). Racial Discrimination and Professional Basketball Salaries in the 1990s. *Applied Economics, 29,* 287-296.

Hill, J. R. (2004). Pay Discrimination in the NBA Revisited. *Quarterly Journal of Business and Economics, 43,* 81-92.

Hoang, H., & Rascher, D. (1999). The NBA, Exit Discrimination, and Career Earnings. *Industrial Relations: A Journal of Economy and Society, 38(1),* 69-91.

Jones, J. C. H., & Walsh, W. D. (1988). Salary Determination in the National Hockey League: The Effects of Skills, Franchise Characteristics, and Discrimination. *Industrial and Labor Relations Review, 41(4),* 592-604.

Kahn, L. M. (1992). The Effects of Race on Professional Football Players' Compensation. *Industrial and Labor Relations Review, 45(2),* 295-310.

Kahn, L. M., & Shah, M. (2005). Race, Compensation and Contract Length in the NBA: 2001-2002. *Industrial Relations, 44,* 444-457.

Kahn, L. M., & Sherer, P. D. (1988). Differences in Professional Basketball Players' Compensation. *Journal of Labor Economics, 6(1),* 40-61.

Koch, J., & Vander Hill, C. W. (1988). Is there Discrimination in the "Black Man's Game?" *Social Science Quarterly, 69,* 83-94.

Longley, N. (2000). The Underrepresentation of French Canadians on English Canadian NHL Teams: Evidence from 1943 to 1998. *Journal of Sports Economics, 1(3),* 236-256.

Longley, N. (2003). Measuring Employer-Based Discrimination Versus Customer-Based Discrimination: The Case of French Canadians in the National Hockey League. *American Journal of Economics and Sociology, 62(2),* 365-381.

MLB Official Rules (n.d.). Retrieved from http://mlb.mlb.com/mlb/official_info/official_rules/definition_terms_2.jsp

Neave, N., & Wolfson, S. (2003). Testosterone, Territoriality, and the Home Advantage. *Physiology and Behaviour, 78,* 269-275.

Nevill, A. M., Newell, S. M., & Gale, S. (1996). Factors Associated with Home Advantage in English and Scottish Soccer Matches. *Journal of Sports Sciences, 14,* 181-186.

Parsons, C.A., Sulaeman, J., Yates, M.C., & Hamermesh, D.S. (2011). Strike Three: Discrimination, Incentives, and Evaluation American Economic Review 101(4): 1410-1435.

Price, J. & Wolfers, J. (2010). Racial Discrimination Among NBA Referees. *Quarterly Journal of Economics, 125(4),* 1859-1887.

Schroffel, J. L., & Magee, C. S. P. (forthcoming). Own-Race Bias Among NBA Coaches. *Journal of Sports Economics,* doi: 10.1177/1527002511402758

Sutter, M. & Kocher, M.G. (2004). Favoritism of agents: The case of referees' home bias. *Journal of Economic Psychology, 25*, 461-469.

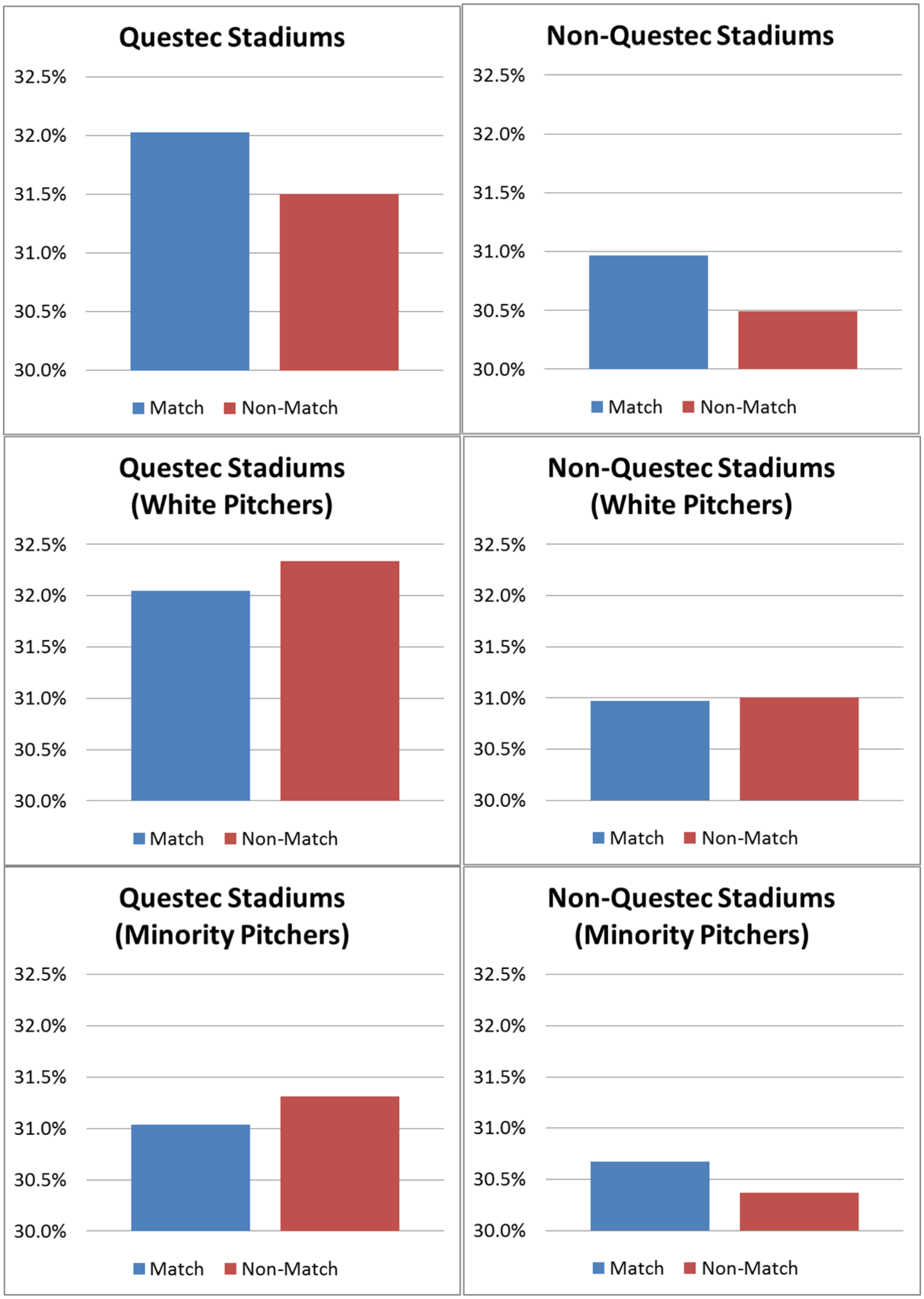
Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B), 73*, 3-36.

Wood, S.N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association, 99*, 673-686.

Wood, S.N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B), 65*, 95-114.

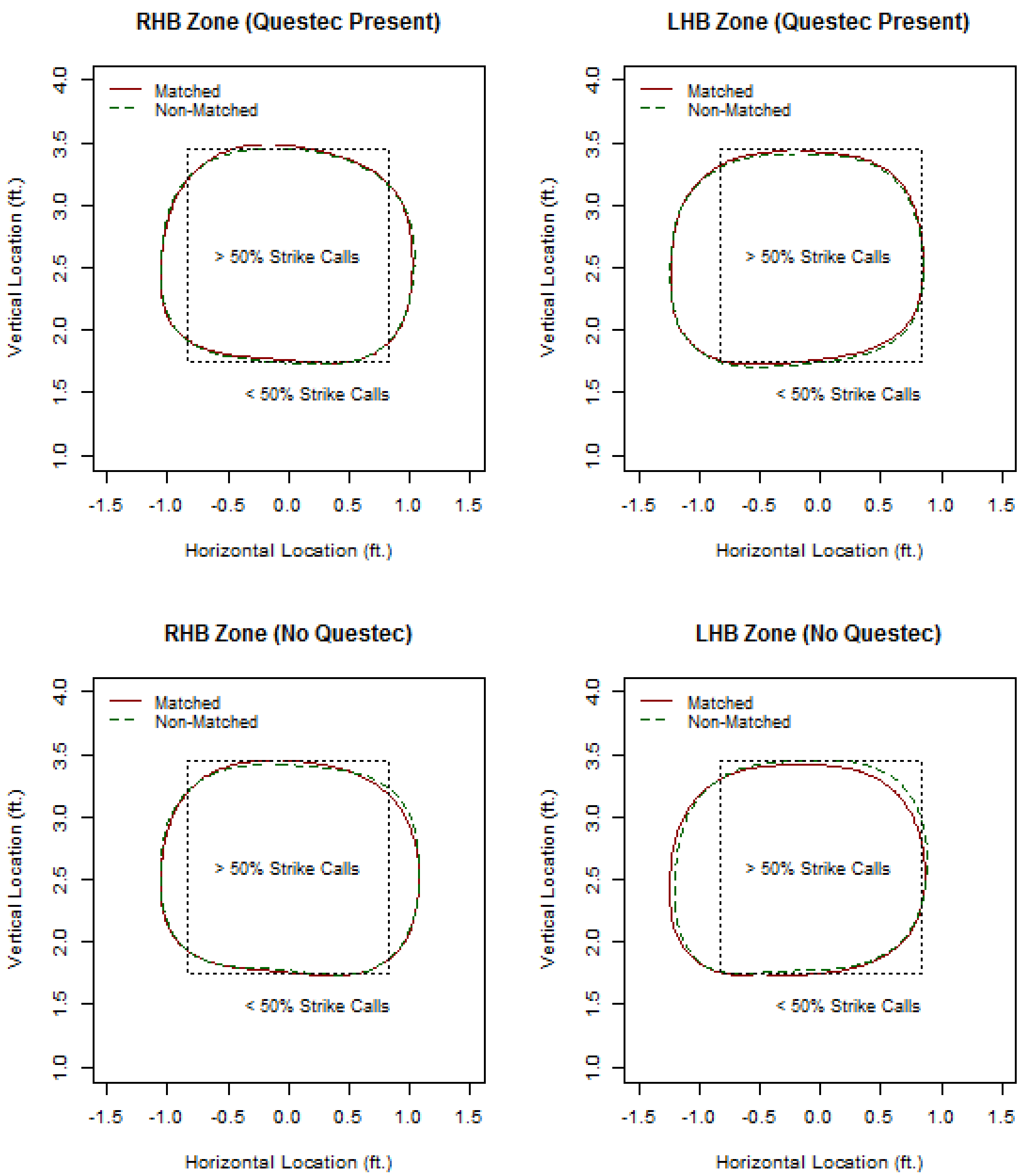
Wood, S.N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society (B), 62*, 413-428.

Figure 1
Comparison of QuesTec and Non-QuesTec Parks



Note: Vertical axes indicate the percentage of umpire-called pitches that were called strikes in the given scenario.

Figure 2
Strike Zone Comparison for Matched Race and QuesTec Presence



Note: These graphs are from the umpire's perspective.

Figure 3
Strike Zone Band for Table 8 (40% to 60% Strike Probability)

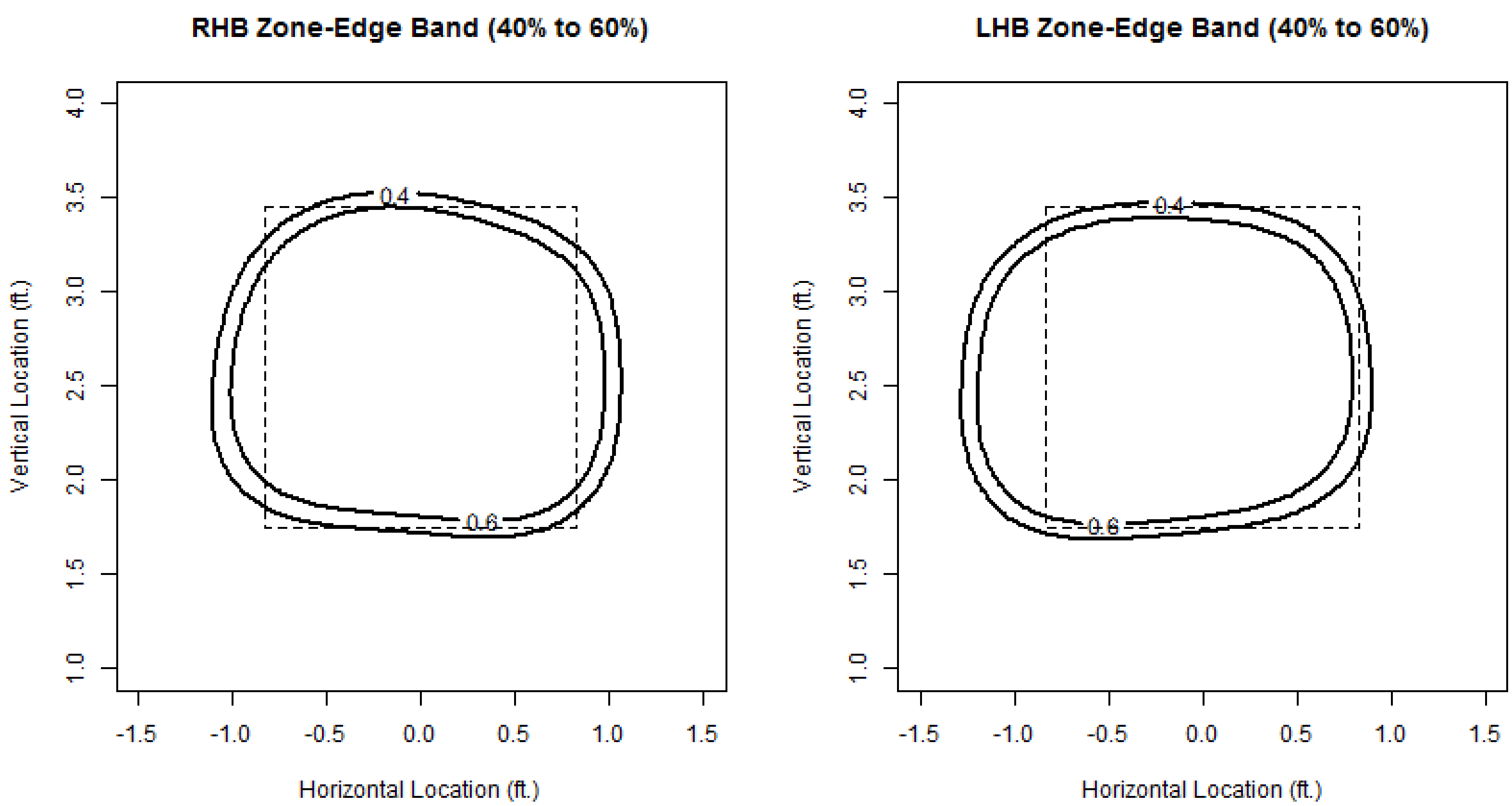


Table 1
Summary Statistics

1997-2008									
	N	Total pitches	Called strike	Called ball	Swinging strike	Foul	In play	Intentional ball	Hit by pitch
All		8,290,073	16.68	37.10	8.96	16.94	19.63	0.63	
Pitcher									
White	1,357	5,994,819	16.77	37.03	8.80	16.93	19.79	0.63	
Hispanic	416	1,810,392	16.45	37.18	9.41	17.00	19.25	0.66	
Black	65	260,153	16.05	37.73	9.26	17.27	19.10	0.55	
Asian	45	225,396	16.52	37.65	9.29	16.67	19.25	0.57	
Umpire									
White	134	7,560,068	16.67	37.10	8.96	16.95	19.64	0.63	
Hispanic	7	339,438	16.96	36.78	8.92	16.99	19.64	0.66	
Black	8	390,567	16.48	37.44	8.98	16.83	19.59	0.63	
2004-08									
	N	Total pitches	Called strike	Called ball	Swinging strike	Foul	In play	Intentional ball	Hit by pitch
All		3,527,750	17.02	36.59	8.77	17.32	19.55	0.66	
Pitcher									
White	830	2,489,818	17.10	36.49	8.57	17.34	19.73	0.66	
Hispanic	285	837,598	16.80	36.85	9.32	17.24	19.01	0.68	
Black	32	88,806	16.18	36.67	9.50	17.78	19.24	0.53	
Asian	34	93,491	17.07	36.90	8.68	16.84	19.81	0.59	
Umpire									
White	91	3,204,574	17.02	36.57	8.78	17.32	19.55	0.66	
Hispanic	6	159,623	17.12	36.52	8.71	17.40	19.46	0.69	
Black	6	163,553	16.92	36.91	8.77	17.13	19.55	0.63	
Parsons et al. 2004-08									
	N	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	861	2,544,515	17.19	36.48	8.77	17.1	19.58	0.64	0.24
Hispanic	278	793,797	16.86	36.77	9.57	17.03	18.86	0.64	0.27
Black	37	89,355	16.24	36.68	9.71	17.54	19.07	0.52	0.24
Asian	39	96,957	17.12	36.81	8.87	16.59	19.7	0.64	0.27
Umpire									
White	91	3,215,949	17.09	36.55	8.97	17.09	19.41	0.64	0.25
Hispanic	5	111,524	17.06	36.8	8.87	16.97	19.33	0.7	0.27
Black	6	197,151	17.13	36.63	9.00	16.99	19.44	0.59	0.22

Table 2
Summary of Called Pitches by Umpire-Pitcher Racial/Ethnic Match

1997-2008					
	Pitcher race/ethnicity				
	White	Hispanic	Black	Asian	Total percent called strikes
Umpire race/ethnicity					
White					
Pitches	5,467,744	1,650,047	238,329	204,379	
Called pitches	2,941,402	885,174	128,169	110,587	
Called strikes	916,583	271,605	38,326	33,744	
Percent called strikes	***31.16***	***30.68***	***29.90***	***30.51***	***31.00***
Hispanic					
Pitches	242,596	77,285	9,233	9,876	
Called pitches	130,415	41,368	4,987	5,409	
Called strikes	41,394	12,891	1,521	1,643	
Percent called strikes	***31.74***	***31.16***	***30.50***	***30.38***	***31.53***
Black					
Pitches	284,479	83,060	12,591	11,141	
Called pitches	153,667	44,409	6,765	6,110	
Called strikes	47,379	13,297	1,907	1,855	
Percent called strikes	***30.83***	***29.94***	***28.19***	***30.36***	***30.55***
Total percent called strikes	***31.17***	***30.67***	***29.84***	***30.50***	***31.00***
2004-08					
	Pitcher race/ethnicity				
	White	Hispanic	Black	Asian	Total percent called strikes
Umpire race/ethnicity					
White					
Pitches	2,259,295	762,949	80,841	85,737	
Called pitches	1,210,519	409,267	42,706	46,223	
Called strikes	386,391	128,262	13,082	14,647	
Percent called strikes	***31.92***	***31.34***	***30.63***	***31.69***	***31.74***
Hispanic					
Pitches	113,945	36,896	3,116	4,168	
Called pitches	61,086	19,762	1,682	2,279	
Called strikes	19,528	6,241	550	714	
Percent called strikes	***31.97***	***31.58***	***32.70***	***31.33***	***31.88***
Black					
Pitches	116,578	37,753	4,849	3,586	
Called pitches	62,759	20,335	2,550	1,960	
Called strikes	19,936	6,232	741	600	
Percent called strikes	***31.77***	***30.65***	***29.06***	***30.61***	***31.40***
Total percent called strikes	***31.91***	***31.32***	***30.62***	***31.63***	***31.73***

Parsons et al. 2004-08					
	Pitcher race/ethnicity				
	White	Hispanic	Black	Asian	Total percent called strikes
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
Difference-in-differences
(positive values denote bias in favor of umpires' own race or ethnicity)

1997-2008	
Races/Ethnicities compared	Difference-in-differences
White/Hispanic[a]	-0.10
White/Black	-1.38
Hispanic/Black	-0.09
2004-2008	
Races/Ethnicities compared	Difference-in-differences
White/Hispanic	0.19
White/Black	-1.42
Hispanic/Black	-2.71
Parsons et al. 2004-2008	
Races/Ethnicities compared	Difference-in-differences
White/Hispanic	0.37
White/Black	-0.48
Hispanic/Black	-0.60

a. This DID estimate, for example, is calculated by taking the difference between the change in White umpire strike rate going from a White pitcher (match) to a Hispanic pitcher, and the change in Hispanic umpire strike rate going from a Hispanic pitcher (match) to a White pitcher. In this respect, it measures the *difference* in the changes across pitcher race/ethnicity for White and Hispanic umpires, but does not allow for evaluation of *which* (or if both) race/ethnicity is discriminating. All other calculations for race/ethnicity combinations follow this strategy.

Table 4
Comparison of QuesTec and Non-QuesTec Parks by Race/Ethnicity Match (1997-2008)

QUESTEC

Pitcher R/E:	**White**	**Hispanic**	**Black**	**Asian**
Umpire R/E				
White				
Pitches	979,795	323,712	42,838	37,153
Called Pitches	526,223	174,138	22,752	20,108
Called Strikes	168,643	54,626	7,089	6,355
Strike Rate	***32.05%***	***31.37%***	***31.16%***	***31.60%***
Hispanic				
Pitches	47,675	15,775	1,269	1,511
Called Pitches	25,542	8,463	700	825
Called Strikes	8,286	2,657	223	268
Strike Rate	***32.44%***	***31.40%***	***31.86%***	***32.48%***
Black				
Pitches	49,963	15,863	2,485	1,843
Called Pitches	26,951	8,539	1,326	1,012
Called Strikes	8,689	2,537	381	323
Strike Rate	***32.24%***	***29.71%***	***28.73%***	***31.92%***

Difference-in-Differences

White/Hispanic	***-0.37%***
White/Black	***-2.62%***
Hispanic/Black	***-1.44%***

NO QUESTEC

Pitcher R/E:	**White**	**Hispanic**	**Black**	**Asian**
Umpire R/E				
White				
Pitches	4,487,949	1,326,335	195,491	167,226
Called Pitches	2,415,179	711,036	105,417	90,479
Called Strikes	747,940	216,979	31,237	27,389
Strike Rate	***30.97%***	***30.52%***	***29.63%***	***30.27%***
Hispanic				
Pitches	194,921	61,510	7,964	8,365
Called Pitches	104,873	32,905	4,287	4,584
Called Strikes	33,108	10,234	1,298	1,375
Strike Rate	***31.57%***	***31.10%***	***30.28%***	***30.00%***
Black				
Pitches	234,516	67,197	10,106	9,298
Called Pitches	126,716	35,870	5,439	5,098
Called Strikes	38,690	10,760	1,526	1,532
Strike Rate	***30.53%***	***30.00%***	***28.06%***	***30.05%***

Difference-in-Differences

White/Hispanic	***-0.02%***
White/Black	***-1.14%***
Hispanic/Black	***-1.12%***

Table 5
OLS Estimates Using Pitcher outing as Unit of Observation[a]

1997-2008	Minimum of 1 pitch		Minimum of 10 pitches		Minimum of 50 pitches	
Constant	0.305***	-----	0.298***	-----	0.307***	-----
	(0.00075)	-----	*(0.00056)*	-----	*(0.00070)*	-----
Match	0.00622***	-0.00145	0.00531***	-0.00099	0.00332***	-0.00300*
	(0.00081)	*(0.00176)*	*(0.00061)*	*(0.00128)*	*(0.00076)*	*(0.00156)*
QuesTec	0.0131***	0.00432***	0.00973***	0.00331***	0.00967***	0.00518***
	(0.00149)	*(0.00160)*	*(0.00113)*	*(0.00117)*	*(0.00138)*	*(0.00142)*
Match*QuesTec	-0.00254	0.00006	-0.00030	0.00126	0.00171	0.00005
	(0.00186)	*(0.00197)*	*(0.00139)*	*(0.00143)*	*(0.00169)*	*(0.00171)*
Home	0.00596***	0.00591***	0.00404***	0.00418***	0.00410***	0.00393***
	(0.00069)	*(0.00068)*	*(0.00051)*	*(0.00049)*	*(0.00063)*	*(0.00058)*
Pitcher FE	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**
Umpire FE	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**
N	208,266	208,266	123,792	123,792	33,984	33,984
R^2	0.001	0.045	0.003	0.105	0.007	0.186

2004-2008	Minimum of 1 pitch		Minimum of 10 pitches		Minimum of 50 pitches	
Constant	0.318***	-----	0.308***	-----	0.317***	-----
	(0.00122)	-----	*(0.00093)*	-----	*(0.00117)*	-----
Match	0.00478***	0.00009	0.00506***	0.00230	0.00518***	-0.00210
	(0.00137)	*(0.00265)*	*(0.00104)*	*(0.00195)*	*(0.00132)*	*(0.00245)*
QuesTec	0.00288	-0.00050	0.00214	-0.00098	0.00189	-0.00014
	(0.00180)	*(0.00189)*	*(0.00137)*	*(0.00139)*	*(0.00170)*	*(0.00168)*
Match*QuesTec	-0.00081	-0.00064	-0.00056	0.000100	-0.00142	-0.00194
	(0.00227)	*(0.00235)*	*(0.00170)*	*(0.00171)*	*(0.00209)*	*(0.00204)*
Home	0.00146	0.00151	0.00056	0.00038	0.00037	0.00056
	(0.00105)	*(0.00103)*	*(0.00079)*	*(0.00074)*	*(0.00096)*	*(0.00088)*
Pitcher FE	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**
Umpire FE	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**
N	92,012	92,012	52,958	52,958	14,269	14,269
R^2	0.000	0.049	0.001	0.115	0.002	0.187

***, **, and * denote statistical significance at the 99%, 95% and 90% level, respectively. a. We also have pitcher-only and umpire-only fixed effects models available upon request. These discerned that the change in significance of the Match variable originates from inclusion of pitcher fixed effects, likely due to the significant weight of White-White matches, with White pitchers throwing more strikes on average than pitchers of other races.

Table 6
LPM Estimates for Propensity to Throw to Edge of Strike Zone (2007-2008)[a]

	40% to 60% Pooled[b]	
Predictor Variable	**Coef. Est.[c]**	**Std. Error**
Constant	-----	-----
Right Handed Pitcher	0.00611	0.00743
Pitch Starting Speed	0.00019	0.00011
Curveball[d]	-0.00710***	0.00173
Changeup	-0.00317***	0.00121
Cutter	0.00122	0.00217
Four-Seam Fastball	0.00221	0.00372
Splitter	-0.0101***	0.00404
Two-Seam Fastball	0.01370	0.02060
Knuckleball	-0.00186	0.00951
Sinker	-0.00507*	0.00297
Slider	-0.00475***	0.00109
Bases Loaded[e]	-0.00150	0.00158
First and Second	-0.00089	0.00099
First and Third	-0.00138	0.00146
On First	-0.00115*	0.00067
On Second	0.00045	0.00095
On Third	0.00011	0.00157
Second and Third	0.00031	0.00185
0-1 Count[f]	-0.00092	0.00086
0-2 Count	-0.0107***	0.00110
1-0 Count	0.00068	0.00091
1-1 Count	0.00063	0.00092
1-2 Count	-0.00713***	0.00095
2-0 Count	-0.00021	0.00139
2-1 Count	0.00159	0.00118
2-2 Count	-0.00166	0.00102
3-0 Count	0.00039	0.00240
3-1 Count	-0.00034	0.00171
3-2 Count	-0.00078	0.00126
Home Team at Bat	0.00030	0.00052
Race Match	**-0.00204**	**0.00124**
QuesTec	**0.00101**	**0.00108**
Match*QuesTec	**0.00056**	**0.00125**
N:	952,375	
R^2:	0.0026	
Pitch Location Distribution	**Pitches**	**Proportion**
Inside	438,446	46.04%
Edge	60,051	6.31%
Outside	453,878	47.66%

***, **, and * denote statistical significance at the 99%, 95% and 90% level, respectively. a. Dependent variable is probability of throwing a pitch to the edge of the strike zone and model includes umpire, pitcher, and batter fixed effects, with controls for inning and year the pitch is thrown (2007 or 2008). b. Denotes definition of the "edge" of the strike zone using pooled GAM model for likelihood of a strike call. c. Batter handedness and height excluded, as it was used to create "Edge of Strike Zone" variable. d. Pitch types compared to generic "Fastball" classification as the base level. e. Base level of empty bases. f. Base level of 0-0 count.

Table 7
LPM Estimates for Pitch Linear Distance from Center (2007-2008)[a]

Predictor Variable[b]	**Coef. Est.[b]**	**Std. Error**
Constant	-----	-----
Right Handed Batter	-0.41495***	0.04050
Right Handed Pitcher	-0.08194	0.20624
Batter Height	-0.29870***	0.03860
Pitch Starting Speed	-0.01382***	0.00306
Changeup	1.28787***	0.03423
Curveball	0.66980***	0.05115
Cutter	0.39509***	0.05695
Four-Seam Fastball	-0.50544***	0.09490
Splitter	1.67355***	0.12683
Two-Seam Fastball	0.65175	0.46254
Knuckleball	1.34863***	0.29798
Sinker	1.02514***	0.08179
Slider	0.84845***	0.03164
Bases Loaded	0.15286***	0.04408
First and Second	0.25096***	0.02739
First and Third	0.47609***	0.04094
On First	0.10561***	0.01844
On Second	0.42359***	0.02626
On Third	0.71462***	0.04399
Second and Third	0.63786***	0.05288
Inning	0.01060***	0.00380
2008 Season	-0.05534***	0.01872
0-1 Count	1.20321***	0.02380
0-2 Count	3.30847***	0.03559
1-0 Count	-0.35783***	0.02370
1-1 Count	0.39924***	0.02489
1-2 Count	2.07670***	0.02854
2-0 Count	-0.79961***	0.03556
2-1 Count	-0.50632***	0.03057
2-2 Count	0.75013***	0.02876
3-0 Count	-1.06530***	0.06000
3-1 Count	-1.09705***	0.04296
3-2 Count	-0.74376***	0.03385
Home Team at Bat	-0.00849	0.01418
Race/Ethnicity Match	**-0.00527**	**0.03478**
QuesTec	**-0.05992****	**0.02994**
Match*QuesTec	**0.04091**	**0.03439**
N: 952,375		
R^2: 0.0752		

***, **, and * denote statistical significance at the 99%, 95% and 90% level, respectively. a. Dependent variable is the pitch's distance from center point of the strike zone in inches. Model includes umpire, pitcher, and batter fixed effects. b. Base levels for factor variables are the same as in Table 6. c. Negative coefficient implies pitch is *closer* to the center point of the strike zone.

Appendix 1: Direct Replication of PSYH

We first try to replicate the results of PSYH with our own data, reducing our sample to only the 2004-2008 seasons. However, as noted earlier, there are a number of differences in our data sets. PSYH collected their data from espn.com and Sportvision's Gameday database; our sample comes from baseball-reference.com for 2004-2006, while we use the same data source as PSYH for 2007-2008. This may result in some variation, for example a small number of pitches from baseball-reference.com are categorized as "unknown strikes." Also, data from rained out games might be handled differently since not all games become official.

One important factor in our sample differences is related to race classification. There is considerable subjectivity when classifying umpires or pitchers by race. In Table 2 we report a comparison of the number of called pitches and the percentage of called strikes according to umpire race and pitcher race using our data and PSYH. First, the totals are different, indicating that there are small discrepancies between the data sources. Second, and notably, in some cases the pitcher-umpire intersection figures vary a great deal. This is largely attributable to different coding of umpire race by PSYH and in the current study. Specifically, in our analysis we coded umpire Laz Diaz as Hispanic based on his background and personal correspondence. However, PSYH coded him as Black, perhaps due to his skin tone and the fact that he is from the United States. The conflation of race and ethnicity can cause significant problems. As highlighted by Fort and Gill (2000), if race is the variable of interest, then categorical measures may be inappropriate altogether. When Laz Diaz is coded as Black, he accounts for 24.7% of pitches called by

Black umpires. When Laz Diaz is coded as Hispanic, he accounts for 32.5% of called pitches by Hispanic umpires. Thus, even though we have over 3.5 million pitches in the 2004-2008 data base, misclassification of one individual alone can change the analysis, altering the results or at the very least making the results less robust.

Returning to the difference-in-differences in Table 3 from the paper, both our data and PSYH have positive values for White/Hispanic matchups in the 2004-2008 sample, indicating discrimination. Like PSYH, our data reveal negative values for White/Black and Hispanic/Black. However, our comparable 2004-2008 data find that reverse discrimination effect for the latter two groups is much larger than the favoritism given by White umpires to White pitchers compared to Hispanic pitchers. At minimum, the results in Table 3 of the paper show that the difference-in-differences results vary according to the time period.

Central to the contribution of PSYH, however, is that discrimination does not always exist—the effect dissipated in stadiums that used QuesTec technology and according to game situation. The implication was that discrimination occurred when umpires were not being monitored. The data from espn.com contained variables not included on baseball-reference.com. Specifically, the game situation data—for example, the count, inning and attendance—were not made available to us for all of 2004-2008. Consequently, we focused the replication of PSYH's explicit monitoring on the presence or absence of QuesTec (see Tables A-1 and A-2). PSYH uses multiple fixed effects for each pitcher, batter, and umpire—one for QuesTec stadiums and one for stadiums without QuesTec. In order to make an accurate comparison to our estimates, we remove control variables and fit both single and QuesTec-unique fixed effects for the estimates from the PSYH data and our own. Here, we replicate our aggregated data in order to fit a model using each pitch as an individual observation (in contrast to Table 5, where we use pitcher-umpire-game events). As evidenced by the data in Table A-1, any significant effects of the race matching variable again disappear when including pitcher, umpire fixed effects. Due to the structure of our data, we were not able to include batter fixed effects in the estimation; but, this further model exhibits that standard errors were not simply reduced by aggregating our data at the game-pitcher-umpire level. Additionally, the PSYH data set shows neither evidence of discrimination, nor effects of monitoring in this simplified form with single fixed effects. There are mixed results, however, when using QuesTec and non-QuesTec unique fixed effects for each pitcher and umpire. This specification, originally used by PSYH, will be expanded upon later.

Next we attempt to further reconcile our results with PSYH through corrections in the data, and from here on use the data and code provided by PSYH modified as exhibited in Appendix 2. Table A-2 shows estimates using the data from PSYH in its original form as well as after rectifying various discrepancies within the identification of players and their race or ethnicity. Panel A (of the top row) represents the original estimation of PSYH with their data and code.

Panel B consolidates pitcher identifications, as 34 pitchers had two or more identification numbers in the original PSYH data. Next, Panel C eliminates five pitchers whose names could not be identified as a pitcher who pitched in MLB over the time period. Panel D then deletes pitchers that appear to have inconsistencies as to their pitch count. For example, in the PSYH data, Edgar Gonzalez had 1201 called pitches and Enrique Gonzalez had 1730 called pitches. In our data set, over the same time period, Edgar Gonzalez had 1976 called pitches and Enrique Gonzalez had 975 called pitches. Given that from 2004 to 2008 Edgar Gonzalez had over twice as many innings pitched, it seems likely that our data is more accurate, as the pitches-per-inning ratios align more closely. Therefore, pitchers with these types of discrepancies were eliminated from the sample. Panel E then changes the race classification for certain pitchers. Changes were made to reflect the race classifications in our original classifications. Also, pitchers that we originally classified as "other" were removed from the sample. The estimations were run with Laz Diaz as Black and then Hispanic. We also ran the estimation to include Asian pitchers while Laz Diaz is classified as Hispanic (Panel F).

Lastly, while the specification of QuesTec-unique fixed effects does not make the model incorrect, one could argue that each pitcher should only have one fixed effect. The selective use of two fixed effects assumes a differential change in behavior across QuesTec conditions at the individual pitcher level. Even if pitchers are aware of the presence of QuesTec and its possible impact on race-based bias by the umpire, treating a single pitcher as a completely different player in each park would seem to lose important information about that player. This is essentially the treatment given by PSYH when creating two fixed effects for each player independent of one another. Additionally, this choice is rather selective as individual pitchers change their behavior in different ways depending on the ball-strike count or whether they are home or away. If the assumption is that multiple fixed effects are necessary for individual level QuesTec changes, then these should also be specified for other conditions that are more directly apparent to players such as home park and ball-strike count. Therefore, we estimate each of the panels in Table A-2 using only a single fixed effect for each pitcher, batter and umpire.

This change ultimately results in large changes in the magnitude and significance of coefficients in the model. Table A-2 shows that the coefficient estimate for matching umpire and pitcher race is reduced by nearly 60%, while those coefficient estimates associated with both explicit (i.e., QuesTec) and implicit (i.e., high attendance) monitoring are no longer statistically significant within the regression. Much of this difference is due to the consolidation of fixed effects.

Of course, the lesson from PSYH regarding impact of discrimination implicit in measurement through subjective evaluation is well taken. However, we show here that—while the point made is valid and important—the data from Major League Baseball is sensitive to various specifications and samples.

Appendix 2: Explanation of Table A-2

Preliminary Description
All models in Table A-2 are versions of Model 9 from Parsons et al., Page 1422, Table 5, Panel C. This is the most complete model and includes all variables. We use the code from PSYH, with an adjustment for the consolidation of pitcher codes in order to produce *new* unique identifiers for non-QuesTec and QuesTec stadiums where necessary.

QuesTec-Unique vs. Single Fixed Effects
Table A-2 exhibits all models both with a single fixed effect for each batter, pitcher and umpire, and the originally specified QuesTec-unique fixed effects for each in PSYH.

Diaz Black vs. Diaz Hisp.
Each model is fit with Laz Diaz as the originally classified Black umpire that PSYH included in their data. Those with Laz Diaz coded as Hispanic are the exact same models as described below, but include a newly classified Laz Diaz (as Hispanic). The race-match variables are adjusted to include this change as well.

MODELS:
Panel A: Includes model for original data from the paper with no adjustment unless Laz Diaz or Fixed Effects changes are specified at the table heading.

Panel B: Consolidates pitcher codes that had multiple entries and multiple codes. This is in order to remove the additional fixed effects created by this. In the code, we recreate the QuesTec-unique identification codes for all pitchers to account for this. The consolidated codes are described below:

Player Name	Duplicate ID	Changed to (Original ID)
Brian Anderson	397	17
Alberto Arias	2006	2005
Cha Seung Baek	765	36
Brian Bass	2016	2015
Billy Buckner	20070003	2027
Valerio De Los Santos	943	748
Chris Demania	944	207
Matt DeSalvo	20070012	2047
Lenny Dinardo	216	212
Geno Espineli	2055	2054
Sean Gallagher	2060	2059
Edgar Gonzalez	307	306
Enrique Gonzalez	309	308
Geremi Gonzalez	311	310
Joel Hanrahan	20070021	2070
Philip Humber	20070024	394
Wil Ledezma	2099 & 2100	455

Dan Meyer	2122	544
Josh Newman	2135	2134
Ross Ohlendorf	2140	2139
Chan Ho Park	623	380
Bobby Parnell	2148	2147
Scott Patterson	2152	2151
Jae Kuk Ryu	730	448
Bobby Seay	945	759
Paul Shuey	20070044	20070043
Joakim Soria	2190	2189
Levale Speigner	20070045	2191
John Van Benschoten	856	65
Jermaine Van Buren	855	115
Tod Van Poppel	857	656
Rick Vanden Hurk	20070025	2202
Jared Wells	2208	2207
Randy Wells	2210	2209

Panel C: Has same characteristics of previous file, but deletes 5 unidentifiable pitchers from the analysis. These include the following pitcher codes and reduces the data trivially from 1,838,676 to 1,838,487:

Original "pid"	**Pitcher Name (Parsons et al. Key)**
135	Sil Campusano
417	b Jones
492	Bobby M
853	g valera
2045	Frankie de

Panel D: This is a second round of consolidation and deletion of pitcher codes that have serious anomalies or overlap between two players. For example, the confusion in "de paula", "en gonzalez", and "ed gonzalez". This is a larger task than the last and deletes 20 players, reducing the observations from 1,838,487 to 1,788,126:

Original "pid"	**First/Last Name (Parsons et al. Key)**
205	jorge depaula
262	randy flores
263	ron flores
306	ed gonzalez
308	en gonzalez
410	jason johnson
411	jim johnson
412	josh johnson
548	justin miller
610	ramon rrtiz
611	russ ortiz

629	de paula
635	tony pena
675	ramon ramirez
691	al reyes
692	anthony reyes
890	jeff weaver
891	jered weaver
928	jamey wright
929	jaret wright

Panel E: This step changes the race of a number of pitchers from White to Hispanic, Hispanic to White, and deletes those classified as "other" in our data. Pitcher-umpire race match variables are adjusted to accommodate new classifications. The sample size is reduced from 1,788,126 to 1,779,041. These changes include:

Original "pid"	**Pitcher Name**	**Race Changed To:**
26	Bronson Arroyo	DELETED
2036	Joba Chamberlain	DELETED
738	Brian Sanches	DELETED
149	Frank Castillo	White → Hispanic
151	Jaime Cerda	White → Hispanic
175	Chad Cordero	White → Hispanic
181	Nate Cornejo	White → Hispanic
259	Nelson Figueroa	White → Hispanic
281	Brian Fuentes	White → Hispanic
2058	Armando Gallaraga	White → Hispanic
320	Danny Graves	DELETED
2109	Warner Madrigal	White → Hispanic
2112	Justin Masterson	Hispanic → White
518	Thomas Mastny	DELETED
576	Rodney Myers	Black →White
2150	Manny Parra	White → Hispanic
674	Horacio Ramirez	White → Hispanic
2159	Clay Rapada	White → Hispanic
2162	Jojo Reyes	White → Hispanic
734	Chris Seanz	White → Hispanic
2199	Erick Threets	White → Hispanic

Panel F: This data set includes all of the previous changes to the data, but then goes on to include Asian pitchers within the data. Here, the sample size increases from 1,779,041 to 1,829,482. This combined with all of the other changes reduces the coefficients the most toward zero, but most are still statistically significant.

Table A-1
LPM Estimates Using Individual Pitches as Unit of Observation (2004-2008)

	Current Analysis			**Parsons et al. Data**		
	No Fixed Effects	**Single FE**	**QuesTec Unique FEs**	**No Fixed Effects**	**Single FE**	**QuesTec Unique FEs**
Constant	0.313***	-----	-----	0.315***	-----	-----
	(0.00076)	-----	-----	*(0.00074)*	-----	-----
Match	0.00602***	0.00009	0.00322	0.00572***	0.00119	0.00409*
	(0.00093)	*(0.00179)*	*(0.00216)*	*(0.00091)*	*(0.00194)*	*(0.00237)*
QuesTec	0.00286**	0.00026	-----	0.00130	-0.00088	-----
	(0.00124)	*(0.00134)*	-----	*(0.00120)*	*(0.00130)*	-----
Match*QuesTec	-0.00169	-0.00166	-0.00898**	-0.00161	-0.00246	-0.00927**
	(0.00152)	*(0.00162*	*(0.00351)*	*(0.00148)*	*(0.00159)*	*(0.00380)*
Pitcher FE	**No**	**Yes**	**Yes**	**No**	**Yes**	**Yes**
Umpire FE	**No**	**Yes**	**Yes**	**No**	**Yes**	**Yes**
N	1,825,680	1,825,680	1,825,680	1,890,970	1,890,970	1,890,970
R^2	0.0000	0.0037	0.0042	0.0000	0.0037	0.0041

Table A-2
Parsons et al. (2011) Data Manipulations[a]

QuesTec-Unique Fixed Effects						
	Diaz Black					
	A	***B***	***C***	***D***	***E***	***F***
UPM	0.00888***	0.00877***	0.00883***	0.00900***	0.00953***	0.00830***
	(0.00239)	*(0.00238)*	*(0.00238)*	*(0.00242)*	*(0.00238)*	*(0.00235)*
UPM*QuesTec	-0.0103***	-0.0102***	-0.0102***	-0.0104***	-0.0108***	-0.00955***
	(0.00358)	*(0.00357)*	*(0.00357)*	*(0.00362)*	*(0.00360)*	*(0.00356)*
High Attendance	0.00573***	0.00574***	0.00572***	0.00562***	0.00569***	0.00530***
	(0.00129)	*(0.00129)*	*(0.00129)*	*(0.00131)*	*(0.00131)*	*(0.00126)*
Attend *UPM	-0.00359**	-0.00359**	-0.00358**	-0.00322**	-0.00347**	-0.00306**
	(0.00151)	*(0.00151)*	*(0.00151)*	*(0.00153)*	*(0.00153)*	*(0.00149)*
UPM*Terminal	-0.00588***	-0.00589***	-0.00591***	-0.00600***	-0.00581***	-0.00583
	(0.00143)	*(0.00143)*	*(0.00143)*	*(0.00145)*	*(0.00145)*	*(0.00141)*

	Diaz Hisp.					
	A	***B***	***C***	***D***	***E***	***F***
UPM	0.00681***	0.00670***	0.00675***	0.00691***	0.00691***	0.00583***
	(0.00215)	*(0.00214)*	*(0.00214)*	*(0.00218)*	*(0.00218)*	*(0.00216)*
UPM*QuesTec	-0.00864**	-0.00855**	-0.00861**	-0.00871**	-0.00869**	-0.00757**
	(0.00349)	*(0.00348)*	*(0.00348)*	*(0.00354)*	*(0.00354)*	*(0.00350)*
High Attendance	0.00553***	0.00554***	0.00552***	0.00541***	0.00543***	0.00506***
	(0.00129)	*(0.00129)*	*(0.00129)*	*(0.00131)*	*(0.00131)*	*(0.00126)*
Attend *UPM	-0.00329**	-0.00329**	-0.00328**	-0.00292*	-0.00308**	-0.00269*
	(0.00150)	*(0.00150)*	*(0.00150)*	*(0.00153)*	*(0.00153)*	*(0.00149)*
UPM*Terminal	-0.00568***	-0.00569***	-0.00571***	-0.00580	-0.00556***	-0.00559***
	(0.00143)	*(0.00143)*	*(0.00143)*	*(0.00145)*	*(0.00145)*	*(0.00141)*

Single Fixed Effect						
	Diaz Black					
	A	***B***	***C***	***D***	***E***	***F***
QuesTec	-0.00104	-0.00105	-0.00105	-0.00055	-0.00054	-0.00044
	(0.00135)	*(0.00135)*	*(0.00135)*	*(0.00138)*	*(0.00138)*	*(0.00133)*
UPM	0.00543***	0.00543***	0.00547***	0.00564***	0.00613***	0.00538***
	(0.00202)	*(0.00202)*	*(0.00202)*	*(0.00205)*	*(0.00204)*	*(0.00201)*
UPM*QuesTec	-0.00182	-0.00182	-0.00181	-0.00191	-0.00201	-0.00210
	(0.00157)	*(0.00157)*	*(0.00157)*	*(0.00160)*	*(0.00160)*	*(0.00156)*
High Attendance	0.00582***	0.00586***	0.00586***	0.00573***	0.00582***	0.00529***
	(0.00125)	*(0.00125)*	*(0.00125)*	*(0.00127)*	*(0.00128)*	*(0.00123)*
Attend *UPM	-0.00323**	-0.00328***	-0.00328***	-0.00286*	-0.00314**	-0.00262*
	(0.00148)	*(0.00148)*	*(0.00148)*	*(0.00150)*	*(0.00150)*	*(0.00147)*
UPM*Terminal	-0.00583***	-0.00584***	-0.00586***	-0.00593***	-0.00576***	-0.00578***
	(0.00143)	*(0.00143)*	*(0.00143)*	*(0.00145)*	*(0.00145)*	*(0.00141)*

	Diaz Hisp.					
	A	***B***	***C***	***D***	***E***	***F***
QuesTec	-0.00122	-0.00119	-0.00119	-0.00686	-0.00074	-0.00063
	(0.00135)	*(0.00135)*	*(0.00135)*	*(0.00137)*	*(0.00138)*	*(0.00132)*
UPM	0.00393**	0.00421**	0.00421**	0.00436**	0.00438**	0.00372**
	(0.00184)	*(0.00185)*	*(0.00185)*	*(0.00188)*	*(0.00189)*	*(0.00186)*
UPM*QuesTec	-0.00155	-0.00160	-0.00160	-0.00170	-0.00171	-0.00180
	(0.00156)	*(0.00156)*	*(0.00156)*	*(0.00159)*	*(0.00159)*	*(0.00155)*
High Attendance	0.00565***	0.00569***	0.00569***	0.00557***	0.00558***	0.00508***
	(0.00125)	*(0.00125)*	*(0.00125)*	*(0.00127)*	*(0.00127)*	*(0.00123)*
Attend *UPM	-0.00300**	-0.00303**	-0.00304**	-0.00261*	-0.00279*	-0.00229
	(0.00147)	*(0.00147)*	*(0.00147)*	*(0.00150)*	*(0.00150)*	*(0.00146)*
UPM*Terminal	-0.00566***	-0.00567***	-0.00569***	-0.00576***	-0.00552***	-0.00556***
	(0.00142)	*(0.00142)*	*(0.00142)*	*(0.00145)*	*(0.00145)*	*(0.00141)*

a. Data Manipulations of Panels A through F are provided in the Appendix. Each is a variation of that in Parsons et al. (2011), Table 5, Panel C, Equation 9 (pp. 1422) and includes all control variables (not presented here) originally in that estimation.