2-21-2017

Monitoring and Employee Shirking: Evidence from MLB Umpires

John Charles Bradbury
Kennesaw State University

Monitoring and Employee Shirking: Evidence from MLB Umpires

John Charles Bradbury
Kennesaw State University

560 Parliament Garden Way NW
Kennesaw, GA 30144
570-478-2369
jcbradbury@kennesaw.edu

ORCID ID: 0000-0001-9209-569X

Abstract
Standard neoclassical principal-agent theory predicts that stricter monitoring should reduce employee shirking from principal desires; however, recent analyses indicate that social aspects of principal-agent relationships may result in monitoring "crowding out" disciplinary effects. From 2001 to 2008 Major League Baseball (MLB) instituted an automated pitch-tracking system (QuesTec) to assist in monitoring its umpires. The asymmetric implementation of this new monitoring technology allows for the comparison of monitored and unmonitored umpires to identify shirking to placate on-field lobbying pressure. Estimates identify deviations in calls associated with monitoring; however, overall, umpires appeared to be quite sensitive to league directives for changes in the strike zone absent technological monitoring. Thus, while additional monitoring had some effect on umpire behavior, the extreme sensitivity to MLB mandates when unmonitored by the new technology indicates that pre-existing monitoring (which included human oversight and efficiency wages) was effective at limiting shirking by umpires.

Keywords: Principal-Agent Problem, Monitoring, Sportometrics
JEL Classification: D82, Z20

Acknowledgements: The author thanks Dennis Coates and session participants at the 2017 Annual Meeting of the Public Choice Society for helpful comments. The author also wishes to acknowledge the inspiration of Robert D. Tollison.

I try to call the game I would normally call, but I think about QuesTec every once in a while.

– Anonymous umpire (Chass, 2003)

1. Introduction

Interests of principals and agents are rarely in perfect alignment; therefore, it is expected that agents may take actions that are not in the best interest of the principal (Alchian and Demsetz 1972; Jensen and Meckling 1976). In response, principals monitor agents by actively identifying and punishing shirking and/or employing incentive mechanisms to encourage conformity with the principal's interest (e.g., performance bonuses, profit-sharing, efficiency wages, etc.). The necessity of principal-agent relationships in economic activity requires determination of optimal monitoring to balance adherence to principal interests and tolerance of agent shirking.

Recent technological advances provide new means to monitor employees by directly observing and recording specific instances of shirking. Standard principal-agent theory predicts that additional monitoring should reduce agent shirking from principal desires; however, several recent theoretical and empirical studies indicate that monitoring may not have the desired effect. Monitoring can generate distrust among agents that dampens and "crowds out" the shirking deterrence effects of monitoring desired by principals (Frey 1993, Ellingsen and Johannesson 2008, Falk and Kosfeld 2008, Nagin *et al.* 2002). Furthermore, the returns to monitoring are diminishing, so that additional oversight may confer little further benefits beyond existing monitoring.

This analysis uses a new testing ground to examine the sensitivity of employee agents to principal monitoring, where principal-agent preferences and incentives are known and outcomes are easily observed in a controlled setting. Goff and Tollison (1990) suggest sports can serve as a useful economics laboratory for studying human behavior. Sports are ideal for testing economic hypotheses because the relevant variables can be measured with accuracy, which allows for controlled experiments; the number of observations are large; incentives are well defined; and the data are applicable to a wide variety of economic phenomena.[1]

In this study, I use data from Major League Baseball (MLB) regular season games to examine umpire behavior in response to a new monitoring technology. Umpires are highly-trained and experienced experts in applying baseball rules, and thus their performance is difficult to evaluate by a less-knowledgeable principal, resulting in moral hazard. Changes in umpire calls in response to league directives, in light of incentives to shirk, should illuminate our understanding of agent sensitivity to principal mandates as well as the effectiveness of monitoring technology.

Umpires are assigned the task of determining whether a pitch is a ball or a strike based on a rule-book-defined strike zone. Calls are made in real time and are not subject to reversal. Umpires are regularly reviewed by a team of supervisors and may be rewarded with perks, such as participation in post-season games; however, umpires are rarely disciplined or fired for poor performance and turnover rates are low.[2] Though umpires have incentives to serve their chief principal, their incentives are not perfectly aligned with MLB's interests. Umpires face daily scrutiny from players, managers, and fans (directly and through the media). Because baseball is a zero-sum game, pressure from competing factions of interest groups, umpires routinely face lobbying from aggrieved parties. Due to these close and repeat relationships with these parties, umpires likely wish to minimize conflict with effective lobbying constituencies to avoid *ex post*

[1] A prime example of this approach is demonstrated in McCormick and Tollison (1984) and Hutchinson and Yates (2007), which apply an economic model of crime to basketball to examine the behavioral response to increased police presence using the exogenous addition of an extra referee.

[2] DiGiovanna, (2011) reports, "Not one of baseball's 68 major league umpires has been fired in the last 10 years because of poor ratings behind the plate." O'Connell (2007) notes that, on average, there is one umpire opening per year.

disapprobation. If oversight slack is sufficient to permit umpire discretion, then effective lobbying by players or a vocal crowd may persuade umpires to shirk and affect marginal calls. Buraimo, Forrest, and Simmoms (2010) and Buraimo, Simmons, and Maciaszczyk (2012) have identified lobbying effects generating home-team bias by referees in European soccer. The authors note the importance of controlling for in-game incentives for separating bias from strategic considerations that are important at the time calls are issued that are not discernable in aggregate data; thus, they use minute-by-minute observations to assess potential favoritism. The isolated contests between individual hitters and pitchers in baseball allows for the consideration of additional lobbying effects from player constituencies, as well.

Prior to 2001, umpires were reviewed only through personal subjective observations, and thus there was no mechanism for objective feedback for performance evaluation. In an effort to better monitor umpires, MLB installed the QuesTec computer imaging system in eleven parks over eight years from 2001 to 2008. This new monitoring technology provided objective oversight of umpire performance and thus may have served as an effective deterrent to umpires with incentives to shirk to meet lobbying pressure. Economists have previously used QuesTec to examine the impact of monitoring on umpire behavior. Bradbury (2007) uses outcomes inside and outside QuesTec parks to measure the effects of managerial lobbying for calls, finding scant evidence of managerial influence on strikeouts and walks. Parsons et al. (2011) tests QuesTec monitoring as a deterrent of racial discrimination in umpire calls and finds QuesTec parks to be associated with reduced racial bias in ball/strike calls; however, Tainsky et al. (2015) demonstrates that the findings are sensitive to data and specification choices. Mills (forthcoming A) compares overall differences in total- and called strike rates of individual umpires, using QuesTec implementation and other factors, to examine the role of monitoring and feedback on umpire performance, finding some impact of monitoring and improvement over time.

This analysis uses QuesTec to analyze umpire responsiveness to examine agent responsiveness to new monitoring technology. The empirical strategy of using plate-appearance-level data permits controlling for immediate relevant factors that may affect umpire calls, such as home field advantage, game situation, player abilities, and lobbying influences. Empirical estimates indicate that QuesTec monitoring reduced called strikes, and that umpires were more sensitive to pitcher rather than batter interests. Though the effects are statistically significant, they are small. Relative to the strong observed response of umpires to the MLB directive to increase called strikes in all parks, umpires appeared to be quite responsive to the principal and thus the identified shirking is minimal. Overall, the play of the game on the field was not significantly different inside and outside QuesTec parks; however, QuesTec was associated with shorter games and fewer pitches thrown. Thus, even though additional objective monitoring reduced shirking, existing monitoring, sanctions, training, and compensation were highly effective at influencing umpire behavior.

2. MLB Umpire Monitoring

The game of baseball requires the use of umpires to adjudicate plays on the field in real time. Each play begins with the pitcher pitching the ball to the batter, who may choose to swing at the pitch or let it pass. For pitches at which the batter does not swing, the umpire is responsible for declaring the pitch to be a strike or a ball. The strikes and balls accumulate during the plate appearance, with three strikes resulting in an out and four balls resulting in a safe advance of the batter to first base. Baseball is a zero-sum game in which outs benefit the pitching team and are a cost to the batting team. Thus, on pitches where a batter does not swing, the umpire plays a crucial role in determining the outcome of the game. Pitchers and batters respond to the enforced strike zone as it is called by the umpire, with pitchers placing the ball strategically around the called strike zone and batters responding congruently.

The strike zone is defined *de jure* to be the ball crossing over home plate horizontally, which is 17 inches wide, and vertically between the knees and mid-torso of the batter. While the horizontal strike-zone is objectively defined the same for all batters, the vertical strike-zone varies from batter to batter, and has been adjusted (officially and unofficially) over time. Regardless of the objective definition, the strike zone is open to the interpretation of the umpire, particularly for pitches on the edge of the strike zone.

The umpire is assigned to be a neutral arbitrator of balls and strikes, but he is open to influence from several interested parties. Batters desire balls; pitchers desire strikes; teammates, managers, and team-specific fans want balls for their batters and strikes for their pitchers; and League officials want balls and strikes to be called correctly and fairly in accordance with MLB directives. Due to their proximity, pitchers and batters can lobby umpires directly on the field. Managers also have an opportunity to lobby for their players on the field; however, their interest is team-specific and therefore a manager's interest is not fully in-line with his players. For example, a manager with a weak pitcher may lobby for a bigger strike zone that would benefit his pitchers but be a detriment to his batters. Home team fans may also lobby through praise and punishment by cheering or booing unfavorable calls, while visiting team fans normally lack an effective means for in-game lobbying resulting in an asymmetry of influence between teams. Home team advantage through lobbying of officials is well-documented in baseball and other sports (Moskowitz and Wertheim, 2011).

For much of baseball history, League officials were limited to monitoring umpires through personal observation; however, technological advances in ball-tracking have led to several different types of monitoring technology to identify pitch-placement relative to the strike zone. MLB's strike zone definition was last changed officially in 1996, when the base of the strike zone was extended from the top of a batter's knees to bottom of his knees. Though the strike zone definition has not changed since that time, its enforcement has. As part of an effort to enforce the stated rule-book definition of the strike zone, in 2001 MLB instructed umpires to change the *de facto* strike zone as it was being called by umpires. The changes involved shrinking the typically-called strike zone horizontally and expanding the zone upward approximately seven inches (Chass, 2001b). In conjunction with this rule-enforcement change, MLB introduced the QuesTec monitoring system in three parks in 2001. The camera-based system was expanded to eight parks in 2002, ten parks in 2003, and eleven parks in 2006.[3]

In 2003, the World Umpire Association union filed a grievance, which was later dropped, regarding the use of QuesTec. Umpires felt the monitoring was unfair and claimed that it altered their calls, changing their strike zones from park to park. They asserted that they were told that if more than 90 percent of their calls did not conform to the QuesTec strike zone they would be classified as providing below-standard umpiring (Chass 2003). Thus, there is anecdotal evidence that umpires felt pressure to adjust their ball/strike calls in response to QuesTec monitoring from what their preferred strike zone would be. The overt objection to monitoring may not solely represent a desire to shirk. Frey (1993) argues that increased monitoring may signal principal distrust of the agents. Umpires may have been offended that MLB did not trust their expert judgement and thus altered their calls as an act of defiance.

In 2007, the PITCHf/x tacking system was implemented for use by MLB's online game-tracking website, which provided public access to pitch-tracking data. Though PITCHf/x was not explicitly used for evaluating umpires, it allowed fans an objective tool for umpire evaluation—something that was not available from QuesTec's privately held data. Fans, media members, and participants could access this data for *ex post* lobbying purposes. Participants could also observe ball/strike calls and adjust their expectations about the

[3] It is worth noting that there are conflicting media reports regarding the number of QuesTec parks and times of installation. I assigned QuesTec designations based on the most credible and consistent documentation available from multiple sources. QuesTec installation by year: 2001 – Arizona, Boston, and New York Mets; 2002 – Anaheim, Cleveland, Houston, Milwaukee, and Tampa Bay; 2003 – New York Yankees and Oakland; 2006 – Chicago White Sox.

strike zone. In 2009, MLB abandoned QuesTec monitoring and replaced it with the Zone Evaluation system that was installed in all 30 MLB parks, which continues to operate in conjunction with Zone Evaluation. Since this time, objective umpire monitoring has been applied universally across all games, and thus the opportunity to examine cross-sectionals differences in monitoring across parks no longer exists.

3. Estimating the Impact of Umpire Monitoring

Play-by-play data from 2000 to 2009 are available from Retrosheet (2016) event files. Play-by-play level data permits controlling for specific game situations, such as the base-out situation and quality of the players, that may affect pitch selection and umpire calls.[4] The econometric model is represented in Equation 1.

$$(1)\quad P_i = \mu\mathbf{M}_i + \chi\mathbf{C}_i + \psi\mathbf{Q}_{\hat{n}} + \gamma\mathbf{GS}_i + \tau\mathbf{T}_i + \upsilon\mathbf{U}_i + \varphi\mathbf{B}_i + \alpha + \varepsilon_i$$

P is the proportion of called pitches that are declared strikes by the home plate umpire, the proportion of all pitches that resulted in swings by the batter, or the total number of pitches thrown during plate appearance *i*. A proportion is used because the called-pitch duration of each plate appearances is not fixed. It is expected that pitchers and batters will adapt rationally to the umpire's strike zone, and thus pitch or swing at pitches strategically according to the anticipated called zone for each observation. For the former two estimates, I estimate Equation 1 using a generalized linear model that employs a binomial distribution and logit link function, because P is a percentage, bound between zero and one. However, alternate Tobit and Poisson estimators of the dependent variable as a count were estimated and produced similar results. Total pitches are estimated using a Poisson estimator. Only plate appearances in which the game state remained the same for the duration of the appearance are included—for example, plate appearances that experience advances or outs on the bases (e.g. stolen bases or caught stealing) during the pitcher-batter contest are removed. Plate appearances that result in intentional walks are also removed, because they do not involve umpire discretion. In order to ensure that seasonal averages used as control variables approximate batter and pitcher abilities, observations are limited to plate appearances where both the batter and pitcher experienced at least 100 plate appearances or batters faced during the season of analysis.

M is a vector of three monitoring indicator variables: whether or not a park is monitored by the QuesTec, PITCHf/x, or Zone Evaluation monitoring systems. QuesTec varies by park and year, while PITCHf/x and Zone Evaluation monitoring were adopted league-wide, half-way through the 2007 season (which I denote by the All-Star break) for the former and in 2009 for the latter. Because QuesTec differs across parks, it is the primary focus of this analysis. Outcome changes associated with PITCHf/x and Zone Evaluation may be attributable to other unique factors associated with these time periods. Thus, the timeframe of this study evaluates every park for at least one season before and after the implementation of QuesTec.

C is a vector of potential crowd effects on umpires. Larger crowds may provide additional monitoring and proxy fan interest in the game; therefore, I include attendance as an explanatory variable. A home team indicator variable denotes that the batter represents the home team and the pitcher represents the visiting team. Home team bias will benefit the batters of the home team and harm the pitchers of the visiting team. I also include an interaction term, the product of home team batting and attendance, to measure the impact of the additional home team pressure on the umpire.

Q is a vector of player quality variables that are likely to influence the outcome of the plate appearances. I include the average performance in the dependent variable for the batter and pitcher during the season—

[4] Data freely available at http://retrosheet.org. As a condition of use, I include the following statement: The information used here was obtained free of charge from and is copyrighted by Retrosheet. Interested parties may contact Retrosheet at 20 Sunset Rd., Newark, DE 19711.

mean called strikes, swings, and pitches per plate appearance/batter faced—with the expectation that yearly outcomes are associated the plate appearance outcome. I also include the playing experiences of the batter and the pitcher, measured by the number of years since the players' major-league debuts. This may contain some information regarding the players' ability to manage the strike zone gained with experience—and thus impact a batter's decision to swing or how close to the edge of the zone a pitcher may place his pitches—however, this impact should be captured by average annual performance. More important for this analysis, playing experience proxies the ability of players to earn respect or lobby umpires that results in favorable calls. If this influence matters, then pitchers (batters) with longer tenures are expected to experience more (fewer) called strikes.[5] I also include an indicator variable if the batter and pitcher are opposite-handed. Opposite-handed interactions provide a "platoon" advantage to the batter.[6] Mills (2014) finds that umpires are less likely to call a pitch a strike for catchers than non-catchers, because of their close proximity during the game; therefore, I include a catcher indicator variable denoting whether or not the batter plays the position of catcher for the team during the game.

GS is a vector of game-state variables that includes indicator variables for inning (i.e., first through ninth, and one extra-inning indicator variable), runner configurations (e.g., runner on first, runners on first and second, etc.), and outs. I also include the absolute difference in score at the start of the plate appearance. These factors control for strategic considerations of the batter and pitcher during the plate appearance, as well as the umpire's tendency to call balls and strikes in these situations.

T is a vector of year effects which capture changes in the playing environment for that season, excluding the 2000 season. The Zone Evaluation indicator variable serves as the year effect for 2009. **U** is a vector of unique home-plate umpire identifiers to control for individual umpire tendencies in calling balls and strikes. **B** is a vector of ballpark effects to capture unique aspects of home parks in influencing performance. Bradbury (2007) finds little effect of managerial influence on umpires, and preliminary estimates that employed manager indicator variables also indicated little managerial impact; therefore, manager indicators were not included in the final estimates. α is a constant term and ε is an error term, with all errors clustered to each specific umpire. Greek symbols preceding variable vectors denote appropriate vectors of coefficients.

Table 1 reports the estimated effects of the variables in vectors **M**, **C**, and **Q** on the proportion of called strikes, called balls, and batter swings, and Table A.1 in the appendix reports the summary statistics for these variables. To ease the interpretation of the estimates, the effects are reported as elasticities calculated at the sample average for continuous variables and as marginal effects for indicator variables.

Plate appearances in QuesTec parks were associated fewer called strikes, and thus more called balls, indicating that pitchers were net-losers from QuesTec. If batters were receiving more favorable calls with additional monitoring, then it may be expected that they would swing less often at pitches; however, this is not the case, because the rate at which batters swung at all pitches was not significantly less in QuesTec parks. QuesTec was also associated with fewer pitches being thrown by approximately 0.016 fewer pitches per plate appearance.

Thus, the estimates indicate that outside of QuesTec parks, where umpires were less monitored, umpires gave more marginal calls to pitchers. This is not surprising, given that some of the most vocal complaints regarding QuesTec were made by pitchers.[7] Pitchers also have a more concentrated interest group than

[5] Player performance and experience characteristics were collected from Lahman (2016).
[6] Bradbury and Drinen (2008) finds support for the platoon effect and that batter-pitcher contests are independent and not meaningfully affected by teammate spillovers.
[7] For example, in 2003, Arizona pitcher Curt Schilling destroyed a QuesTec camera with a bat after a game in which he thought the system was influencing calls against him. Pitcher Tom Glavine also expressed concern that umpires were narrowing their strike zones in QuesTec parks (Chass, 2003).

batters and suffer more individually from a poor umpire in a game. The starting pitcher, who typically pitches most of the pitches for his team during a game, has sustained repeat interactions with the home plate umpire on defense. Relief pitchers also typically face several batters per game. On offense, batting opportunities are dispersed among eight or more batters. An umpire who upsets a batter will have less opportunity for feedback than a pitcher over a course of a game. Also, pitchers have more at stake from each game with an umpire, because an umpire who is missing calls affects a batter around four times per game, while a starting pitcher can expect to have twenty or more interactions during the game. Pitchers have a greater incentive to lobby umpires than batters, because a game with unfavorable calls will have a greater weight on aggregate seasonal performance statistics—used in arbitration and free agent contract negotiations—for pitchers than it will for batters. Therefore, shirking in the direction that benefits pitchers is not surprising.

Table 1. Estimated Impact of Monitoring on Batting Outcomes

	Called Strike	Swing	Pitches
QuesTec	-0.0025168*	-0.0001676	-0.0162808*
	[0.0012029]	[0.0008921]	[0.0067763]
PITCHf/x	-0.0084042**	0.0069358**	0.0185955
	[0.0017275]	[0.0014883]	[0.0102528]
Zone Evaluation	0.0040996	-0.0140204	0.0530349**
	[0.0028759]	[0.0020398]	[0.0144869]
Home Team Batter	-0.007164**	-0.0076385**	0.0544672**
	[0.001711]	[0.001181]	[0.0091524]
Attendance	0.0017388	-0.0019459	0.0126626**
	[0.0040928]	[0.0018008]	[0.0019415]
Home Team Batter * Attendance	0.0050971*	0.0029555*	-0.0055736**
	[0.0023738]	[0.0010568]	[0.001127]
Batter Performance Rate	0.6970562**	0.897872**	-0.054728**
	[0.0055458]	[0.0038676]	[0.0054094]
Pitcher Performance Rate	0.7760705**	0.8289114**	0.0116748**
	[0.0084913]	[0.006513]	[0.0005114]
Batter Experience	-0.0224814**	0.0063417**	0.0012816
	[0.001302]	[0.0007161]	[0.0006636]
Pitcher Experience	0.0046442**	0.0035825**	-0.0074648**
	[0.001302]	[0.0007161]	[0.0006636]
Platoon	-0.0286244**	-0.0020404**	0.0688172**
	[0.000866]	[0.0004078]	[0.0033056]
Catcher Batting	0.0028032**	0.0046399**	-0.0710761**
	[0.0008478]	[0.0006022]	[0.004313]
Observations	1,323,427	1,606,351	1,606,351

** p <.01, * p<.05, standard errors clustered to individual umpires in brackets

The introduction of PITCHf/x was also associated with fewer strike calls, but batters were more likely to swing at pitches. Zone Evaluation was not associated with the called strike or swing rates, but Zone Evaluation resulted in more pitches being thrown. However, the coefficients of the PITCHf/x and Zone Evaluation variables should be interpreted with caution. These monitoring systems were implemented

league-wide, and the lack of cross-section variation makes it difficult to distinguish effects attributable to the addition of these monitoring systems from other league-wide factors that are unrelated to monitoring.

While the results do provide evidence of umpire shirking outside of QuesTec parks, the effect is small in two respects. First, the direct magnitude of the effect is slight, resulting in less than one additional called strike per game (-0.81). The large sample size used to generate the estimates make it likely that standard thresholds of statistical significance are expected even for small relationships. However, the effect was large enough to elicit complaints from pitchers and umpires that were widely reported in the media. Though the measured direct impact is small, the effect on the game is not limited to the direct impact on ball/strike calls. Batter-pitcher matchups are strategic and the parties respond to changing incentives by adjusting pitch location and swing decisions, which I address in Section 4.

Second, though objective computer-image modeling was limited to QuesTec parks, the directive to call the strike zone in a manner more consistent with the rule-book strike zone was league-wide with the intention of increasing called strikes.[8] Though the findings appear to reject the monitoring-deterrence hypothesis, umpires continued to be monitored by human crews who watched games in person and on television as had been done previously. Thus, differences in QuesTec parks must be interpreted in the context of leaguewide changes in strike calls.

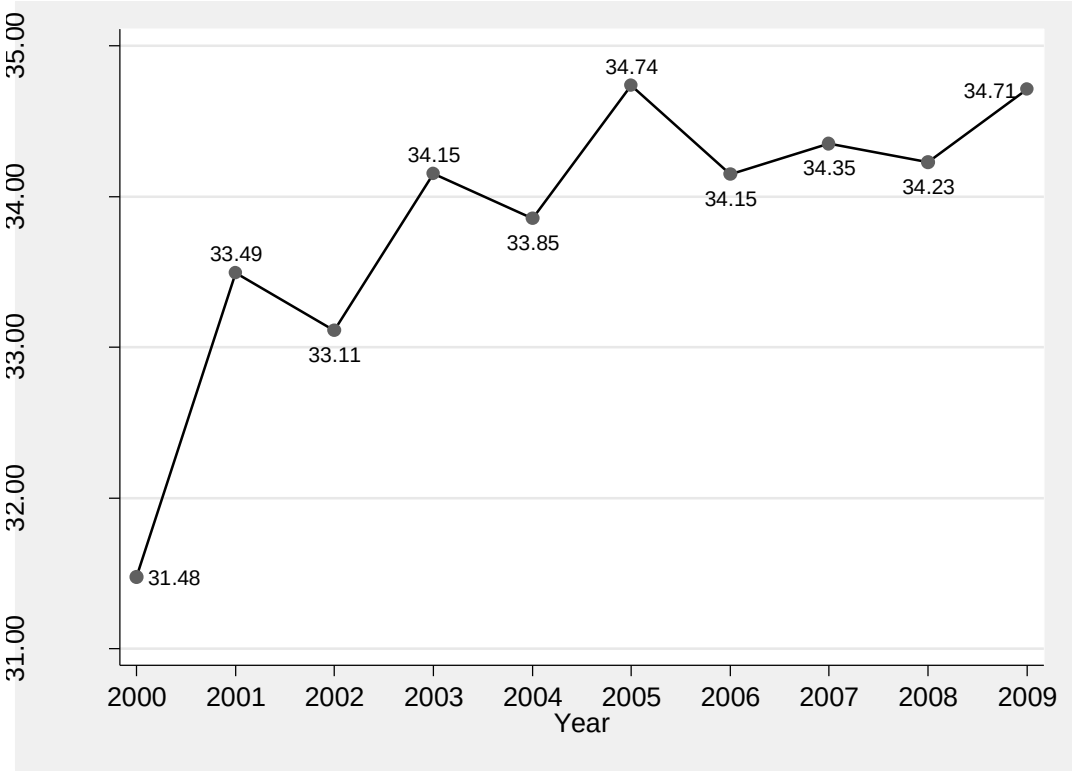
Figure 1 maps the changes in the called strike rate by year for MLB. It shows a dramatic increase in 2001, when MLB directed umpires to call a narrower and taller strike zone. In this context, the 0.02 percentage-point reduction in the called strikes in QuesTec parks is dwarfed by the mean 2.53 percentage-point increase in called strikes during the QuesTec era. This reveals that though umpires were sensitive to QuesTec monitoring and decreased strike calls in QuesTec parks, they were also responsive to the League directive to call a strike zone in a way that made strikes more common both inside and outside QuesTec parks. While the policy change favored pitchers overall, it was minimally dampened by QuesTec monitoring. Thus, umpires may have shirked on the margin to benefit pitchers with less monitoring, but the total effect indicates strong sensitivity to the interests of their principal MLB that are far larger than the shirking identified. The additional monitoring provided little further deterrence from shirking and may explain why MLB discontinued its use of QuesTec after 2008.

Explanatory variables that proxy other interest group influences are largely consistent with expectations and are consistent with shirking on the margin. Home team batters were less likely to receive strike calls, which is consistent with umpires contributing to home team bias. The interaction term of attendance and home team batting is positive, indicating larger crowds reduced home team bias, possibly reflecting the role of visiting fans in the crowd or greater scrutiny associated with the game's popularity. More (less) experienced batters were less (more) likely to have pitches called strikes, and more (less) experienced pitchers were more (less) likely to have pitches called strikes, indicating umpire deference to veterans. Contrary to Mills (2013), catchers were more likely to receive called strikes, which may be a response to the way catchers approach their plate appearances or reflect catchers lobbying for more called strikes for their pitchers. However, the differing effect may be a product of the unit of analysis being each plate appearance rather than each pitch. Estimates for all these variables offer additional evidence of shirking; however, as it was for QuesTec, the estimated impacts are small. The only large estimated effects are the annual batter and pitcher performances, which is consistent with umpires making correct calls. The platoon advantage is negatively related to the called strike and swing rates, which is consistent with umpires correctly calling balls and strikes. Pitchers are

[8] MLB's Executive Vice President for Baseball Operations Sandy Alderson, who issued the directive for closer rule-book enforcement of the strike zone, stated that he expected to see more strikes called: "I expect to see more consistency, and with that greater consistency I think you'll see more strikes called. …That's the direction we're pushing the umpires" Chass (2001b).

expected to be more careful, and pitch on the fringes of the strike zone, when the platoon advantage is in effect.

Figure 1. Called Strike Rate by Year (2000-2009)



4. Impact of QuesTec Monitoring on the Play of the Game

The reason MLB implemented monitoring was to affect the overall play of the game. While the desired consequences of the League are unknown, at the time QuesTec was introduced, baseball was concerned about two particular issues affecting fan interest. First, offense was at a historically high level, with home runs becoming especially more prevalent and noteworthy. Roger Maris's league-record 61 home runs had remained unbroken for 37 years before it was surpassed six times between 1998 and 2001. MLB was facing scrutiny regarding performance-enhancing drugs as the chief contributor to the boost in offense during this era. Mills (forthcoming b) demonstrates that strike zone manipulation after the QuesTec era by MLB was correlated with reductions in offense; perhaps, QuesTec was implemented with similar intentions. Second, baseball games were also growing longer, and MLB openly discussed methods for shortening the games. More accurate umpiring to generate quicker outs may have been another goal.

I estimate the effect of QuesTec on several performance areas using Equation 2. Y is the total number of runs, home runs, strikeouts, or walks in the game (g), or total length of the game in minutes. The vector variables are the same as those included in Equation 1, except they are aggregated at the game level. In place of individual performance characteristics, I include overall average team performance in the same performance area as the dependent variable for the playing teams. I also include the total number of outs in the game to control for game length. I estimate Equation 2 using Ordinary Least Squares, and standard errors are clustered to individual umpires. Table 2 reports the coefficient estimates, and summary statistics are available in Table A.2 in the appendix.

(2) $Y_g = \mu \mathbf{M}_g + \beta_1 \text{Home Team Performance}_g + \beta_2 \text{Visiting Team Performance}_g + \beta_3 \text{Outs}_g + \tau \mathbf{T}_g + \upsilon \mathbf{U}_g + \varphi \mathbf{B}_g + \alpha + \varepsilon_g$

Table 2. Effect of QuesTec on Play of Game

	Runs	Home Runs	Stikeouts	Walks	Time of Game
QuesTec	-0.1802954	-0.0489682	-0.0237386	-0.023212	-1.3759853
	[0.1296994]	[0.0509151]	[0.1196751]	[0.0849533]	[0.5122588]**
PITCHf/x	0.1755923	0.0629618	0.1486324	0.0308878	0.9751861
	[0.1294803]	[0.0408866]	[0.0995800]	[0.0882538]	[0.5238900]
Zone Evaluation	-0.1587898	-0.0567999	0.0201309	-0.0208632	0.0604884
	[0.1236161]	[0.0396464]	[0.0964337]	[0.0903194]	[0.5038562]
Attendance	-0.0000047	0.0000012	0.0000242	-0.0000097	0.0001357
	[0.0000040]	[0.0000012]	[0.0000025]**	[0.0000021]**	[0.0000165]**
Mean Home Team Performance	1.1745377	1.197726	1.0528676	0.9833334	4.6832674
	[0.0674629]**	[0.0581967]**	[0.0415092]**	[0.0544191]**	[0.3003518]**
Mean Visiting Team Performance	0.8125342	0.9552287	0.8786769	0.8371833	4.8212822
	[0.0642260]**	[0.0527568]**	[0.0411365]**	[0.0417322]**	[0.2965928]**
Outs	0.0232616	0.0113668	0.2631054	0.1476173	3.4050818
	[0.0063422]**	[0.0023247]**	[0.0066727]**	[0.0050507]**	[0.0334740]**
R^2	0.06	0.07	0.21	0.12	0.45
Observations	24,156	24,156	24,156	24,156	24,156

** p <.01, * p<.05, standard errors clustered to individual umpires in brackets

None of the monitoring variables are associated with game-level performances. Given the small estimated effects estimated in Table 1, this is not surprising. However, the effect on the length of the game is statistically significant, and the effect is not trivial. QuesTec was associated with reduction in game time of one minute and 23 seconds. While this is small compared to the average length of MLB games during the sample period (174 minutes), it demonstrates that the pace of the game can be increased through monitoring. Using the average number of plate appearances per game during the sample (78.36), the total reduction in pitches was just over one pitch per game (-0.0163 * 78.36 = - 1.28). The average number of pitches per game was approximately 301; thus, on average, each pitch reduction lowered the game length by approximately 35 seconds. The reduction in pitches explains only one-third of the reduction in game length directly. The rest of the reduction may reflect monitoring effects that spillover on the other aspects of the game in QuesTec parks. The act of increased monitoring in one area (strike zone enforcement) may have encouraged umpires to maintain the pace of the game in other areas (limiting the duration of common delays such as pitching changes, time-outs, warm-ups, etc.). This response is akin to the "broken windows" effect of criminal enforcement, where an emphasis on enforcing low-level criminal violations reduces crime in other areas (Corman and Mocan 2005).

5. Discussion and Conclusion

This paper uses baseball data to evaluate the effectiveness of monitoring to deter shirking in a principal-agent framework. Umpire agent responsiveness to principal MLB oversight following the asymmetric introduction of monitoring technology indicates small changes in umpire behavior. The estimates provide evidence that umpires were more likely to favor pitchers absent monitoring, as well as responding to fan preferences and player experience. Overall, umpires appeared to be quite sensitive to League mandates to change the strike zone whether or not they were being surveilled by objective computer monitoring. This responsiveness

indicates that that existing monitoring which involved subjective human review and efficiency wages was sufficient to effect agent conformity to League.

Umpires calling fewer strikes in QuesTec parks may reflect umpire shirking absent external monitoring. However, the fact that umpires were highly responsive to the League directive to increase called strikes in all parks indicates that shirking may not be the only explanation. Frey (1993) posits that tighter monitoring can create distrust in principal-agent relationships and result in "crowding out" of the disciplinary effects of monitoring in response. Falk and Kosfeld (2006) further demonstrates that imposing monitoring can create hidden costs through reduced performance in response to the perceived slight. However, the monopsony status of MLB in hiring umpires and the high wages that MLB umpires receive gives umpires little bargaining power to shirk directly in response to what might be viewed as an act of distrust.[9] At the time of the analysis, the annual salaries of MLB umpires ranged from $120,000 to $350,000, well above the salaries minor-league umpires of less than $20,000 per year (O'Connell 2007, MiLB 2017). Cappelli and Chauvin (1991) demonstrate that the wage premium relative to alternate employment options is negatively related to shirking; thus, the fear of losing a highly-coveted position with a significant wage-premium likely encourages umpire conformity with MLB dictates. While umpire punishment and replacement were rare during the timeframe of analysis, MLB had previously replaced umpires who had been uncooperative in a labor dispute (not due to performance). A disobedient umpire would face significant financial penalties for shirking. Thus, the overall incentive structure and the small reduction in called strikes in QuesTec parks, relative to the much larger increase across all parks (including the parks monitored by QuesTec) indicates that crowding out is an unlikely explanation for the differences in calls inside and outside of QuesTec parks.

While the estimates identify real effects of external objective monitoring, the practical effects are quite small, especially in comparison to the observed changes in umpire behavior regarding strikeouts that occurred in monitored and unmonitored ballparks at a magnitude that was much greater than identified shirking. Overall, QuesTec was not associated with changes in on-field performance, though it was associated with shorter game lengths. In total, the findings indicate that despite the lack of technological oversight, MLB umpires correctly managed games in accordance with MLB's directives. Though umpires are rarely publicly disciplined or fired, they are responsive to League preferences. The results mirror those of Weingast and Moran (1983) which notes that effective oversight is consistent with the absence of obvious monitoring by the principal. Thus, the threat of termination from a high-compensation position by a monopsony employer was effective in deterring employee shirking prior to objective monitoring.

QuesTec's implementation also may have served a public relations purpose. Automated monitoring provided a signal that the game was being called fairly—umpires were not manipulating calls to favor teams, players, or assist gamblers—which may have boosted consumer interest in MLB's product. QuesTec was introduced at a time when other leagues were also introducing technological review of officials. In 1999, the National Football League instituted instant replay review—following a beleaguered experiment from 1986 to 1991—in an effort to reverse incorrect calls on the field that were obvious to the viewing audience. In the 2000s, college football conferences began to implement replay review as well. MLB would later institute limited replay review in 2008 before adopting a more comprehensive replay review system in 2014. QuesTec was abandoned after 2008 and replaced universal ball/strike monitoring systems in all ballparks that are accessible by fans and have become regularly displayed during television broadcasts. The use of this technology as an agent monitoring system by the principal has become secondary to the fan experience.

[9] MLB umpires are represented by a union, but it would be incorrect to view this labor market as a bilateral monopoly similar to the relationship that MLB has with players' union. While MLB umpires may have exceptional skill, they are much more easily replaced than the players and thus have weak bargaining power. A labor battle following the 1999 season resulted in the replacement of 22 long-time umpires.

Analysis of more recent data in the Zone Evaluation era by Mills (forthcoming a) shows umpires have continued to improve the accuracy rates from 85 percent, at the conclusion of the QuesTec era in 2008, to 89 percent by 2014. Even with more advanced technology, the gains have been small. Mills (forthcoming b) also demonstrates that league-mandated changes regarding the enforcement of the strike zone had a significant impact on offense. This analysis shows that umpires were highly responsive to principal edicts, and thus offense changes could have been achieved through strike zone manipulation with pre-QuesTec oversight. Improvement in pitching technology has likely aided umpires through self-evaluation and feedback that were not previously available, in addition to the threat of punishment for missing calls as objectively determined by computer monitoring.

The overall response to League dictates, and the paltry estimates of shirking where monitoring was implemented, indicate that MLB's monitoring strategy of human observation combined with relatively high compensation was sufficient to align agent behavior with principal goals. Thus, additional monitoring may not necessarily improve performance in certain principal-agent relationships even when crowding out is an unlikely explanation.

Appendix

Table. A.1. Summary Statistics, Play-By-Play Variables

	Observations	Mean	Standard Deviation	Minimum	Maximum
Called Strikes Per Pitches Thrown	1,323,427	0.3422318	0.3411171	0	1
Swings Per Pitches Thrown	1,606,351	0.5278087	0.2743856	0	1
Pitches Per Plate Appearance	1,606,351	3.730344	1.879353	1	18
QuesTec	1,606,351	0.2484258	0.4321001	0	1
PITCHf/x	1,606,351	0.244634	0.4298701	0	1
Zone Evaluation	1,606,351	0.0993108	0.299079	0	1
Home Team Batter	1,606,351	0.4912675	0.4999239	0	1
Attendance	1,606,351	30,505	11,369	2,134	61,707
Batter Called Strikes Rate	1,606,351	0.3709482	0.0329289	0.2365039	0.4814952
Pitcher Called Strike Rate	1,606,351	0.367597	0.0276025	0.2491103	0.4934579
Batter Swing Percentage	1,606,351	0.4508216	0.050161	0.2791075	0.6786633
Pitcher Swing Percentage	1,606,351	0.4522806	0.0284973	0.3271028	0.584344
Batter Pitches per Plate Appearance	1,606,351	1.72647	0.1393338	1.115073	2.273504
Pitcher Pitches per Batter Faced	1,606,351	2.108997	1.697285	0.1468599	15.24
Batter Experience	1,606,351	6.285266	4.201766	0	25.40178
Pitcher Experience	1,606,351	5.540736	4.413035	0	24.47912
Platoon	1,606,351	0.544231	0.4980399	0	1
Catcher Batting	1,606,351	0.1241827	0.32979	0	1

Table A.2. Summary Statistics, Game-Level Variables

	Observations	Mean	Standard Deviation	Minimum	Maximum
Runs Scored	24,156	9.513951	4.63529	1	36
Home Team Runs per Game	24,156	4.834299	0.6547064	3.345679	7.814815
Visiting Team Runs per Game	24,156	4.68368	0.5220422	3.395062	6.037037
Home Runs	24,156	2.147127	1.595547	0	12
Home Team Home Runs per Game	24,156	1.080429	0.2607399	0.5555556	1.888889
Visiting Team Home Runs per Game	24,156	1.066482	0.2007517	0.5679013	1.703704
Strikeouts	24,156	13.12328	4.071992	2	40
Home Team Strikeouts per Game	24,156	6.24371	0.7310756	4.493827	8.888889
Visiting Team Strikeouts per Game	24,156	6.877698	0.711014	5.15	9.074074
Walks	24,156	6.688939	3.106158	0	24
Home Team Walks per Game	24,156	3.402914	0.4752322	2.175	5.08642
Visiting Team Walks per Game	24,156	3.287262	0.4690266	2.2375	4.679012
Length of Game (minutes)	24,156	174.4818	26.45195	79	395
QuesTec	24,156	0.2472678	0.4314326	0	1
PITCHf/x	24,156	0.2462328	0.4308247	0	1
Zone Evaluation	24,156	0.1001408	0.3001938	0	1
Attendance	24,156	30,490	11,368	2,134	61,707
Outs	24156	53.48882	4.83664	27	132

References

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

Bradbury, John C. and Drinen, Douglas J. (2008). "Pigou at the Plate: Externalities in Major League Baseball." *Journal of Sports Economics*, 9: 221-224.

Buraimo, Babatunde; Forrest, David; and Simmons, Robert (2010). The 12th Man?: Refereeing Bias in English and German Soccer." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173: 431-49.

Buraimo, Babatunde; Simmons, Robert; and Maciaszczyk, Marek (2012). "Favoritism and Referee Bias in European Soccer: Evidence from the Spanish League and the UEFA Champions League." *Contemporary Economic Policy*, 30: 329-343.

Cappelli, Peter and Chauvin, Keith (1991). "An Interplant Test of the Efficiency Wage Hypothesis". *Quarterly Journal of Economics*, 106:. 769-787.

Chass, Murray (2001a). "Pitchers Regaining Upper Hand." New York Times, July 21. http://www.nytimes.com/2001/07/01/sports/baseball-pitchers-regaining-upper-hand.html (accessed February 17, 2017).

Chass, Murray (2001b). "Calling Pitches By the Book Won't Be Easy." New York Times, March 30. http://www.nytimes.com/2001/03/30/sports/baseball-calling-pitches-by-the-book-won-t-be-easy.html (accessed February 17, 2017).

Chass, Murray (2003). "Umpires Say System Has Them Alter Calls." New York Times, May 29. http://www.nytimes.com/2003/05/29/sports/baseball-umpires-say-system-has-them-alter-calls.html (accessed February 17, 2017).

Corman, Hope, and Mocan, Nancy (2005). "Carrots, Sticks, and Broken Windows." *Journal of Law and Economics*, 48: 235-266.

DiGiovanna, Mike (2011). "In Baseball, Balls and Strikes Still Are in Eye of the Beholder." *Los Angeles Times*, October 8. http://articles.latimes.com/2011/oct/08/sports/la-sp-mlb-strike-zone-20111009 (accessed April 28, 2017).

Ellingsen, Tore and Johannesson, Magnus (2008). "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review*, 98: 990-1008.

Falk, Armin and Kosfeld, Michael (2006). "The Hidden Costs of Control." *American Economic Review*, 96: 1611-1630.

Frey, Bruno S. (1993). "Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty." *Economic Inquiry*, 31: 663-670.

Goff, Brian L. and Tollison, Robert D. (1990). *Sportometrics*. College Station: Texas A&M University Press.

Hutchinson, K. P., and Yates, A. J. (2007). "Crime on the Court: A Correction." *Journal of Political Economy*, 115: 515-519.

Jensen, Michael C. and Meckling, William H. (1976). "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, 3: 305-360

Lahman, Sean (2016). The Lahman Baseball Database, 2015 Version. http://www.seanlahman.com/baseball-archive/statistics/ (accessed (accessed January 27, 2017).

McCormick, Robert E. and Tollison, Robert D. (1984). "Crime on the Court." *Journal of Political Economy*, 92: 223-235.

MiLB.com. (2017). "Umpire Salaries." http://www.milb.com/milb/info/umpires.jsp?mc=_ump_salaries (accessed April 28, 2017).

Mills, Brian M. (2014). "Social Pressure at the Plate: Inequality, Aversion, Status, and Mere Exposure." *Managerial and Decision Economics*, 35: 387-403

Mills, Brian M. (forthcoming a). "Technological Innovations in Monitoring and Evaluation: Evidence of Performance Impacts among Major League Baseball Umpires." *Labour Economics*.

Mills, Brian M. (forthcoming b). "Policy Changes in Major League Baseball: Improved Behavior and Ancillary Productivity Outcomes." *Economic Inquiry*.

Moskowitz, Tobias J. and Wertheim, L. J. (2011). *Scorecasting: The Hidden Influences Behind How Sports Are Played and Games are Won*. New York: Crown Archetype.

Nagin, Daniel S., Rebitzer, James B., Sanders, Seth, and Taylor, Lowell J. (2002). "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment." *American Economic Review*, 92: 850-873.

O'Connell, Jack (2007). "Much Required To Become MLB Umpire." MLB.com, August 28. http://m.mlb.com/news/article/2173765/ (accessed April 28, 2017).

Parsons, Christopher A., Sulaeman, Johan, Yates, Michael C., and Hamermesh, Daniel S. (2011). Strike Three: Discrimination, Incentives, and Evaluation. *American Economic Review*, 101: 1410-1435.

Retrosheet (2016). "Play-by-Play Data Files (Event Files)." http://www.retrosheet.org/game.htm (January 27, 2017).

Tainsky, Scott, Mills, Brian M., and Winfree, Jason A. (2015). Further Examination of Potential Discrimination Among MLB Umpires. *Journal of Sports Economics*, 16: 353-374.

Weingast, Barry R. and Moran, Mark J. (1983). "Bureaucratic Discretion or Congressional Control? Regulatory Policymaking by the Federal Trade Commission." *Journal of Political Economy*, 91: 765-800.