LEAGUE EXTERNALITIES AND PERFORMANCE IMPROVEMENTS IN MLB[1]

Brian M. Mills

Department of Tourism, Recreation, and Sport Management
College of Health and Human Performance
University of Florida
P.O. Box 118208
Gainesville, FL 32611
Tel: 352-294-1664
E-mail: bmmillsy@hhp.ufl.edu

Abstract: Offense in MLB has decreased substantially since 2006, often attributed to increased testing and punitive action for use of performance enhancing drugs. However, there has been concurrent policy change affecting behavior of other league agents that may have also affected game play. I therefore examine the effect of these agents, MLB umpires, on offensive production in baseball. Estimates reveal a majority of the offensive reduction from 2008 through 2014 can be attributed to changes in the size of the strike zone due to monitoring and evaluation of the league's officials. Implications are discussed in the context of firm-production-relevant externalities.

JEL Codes**:** J44, Z22, Z28, L25

Keywords: Baseball, Performance Improvement, League Policy

[1] ***Acknowledgements***: I would like to thank Charlie Brown, Rodney Fort, Scott Tainsky, Adrian Burgos, Alan Nathan, Jason Winfree, Michael Lopez, and participants at the 2014 Western Economic Association Conference in Denver, the 2014 European Conference on Sports Economics in Antwerp, the 2015 Joint Statistical Meetings in Seattle, and the University of Illinois Center for Critical Approaches to Sport, Culture, and Performance seminar for helpful comments on earlier versions of this manuscript in written and presentation form. Finally, I would like to thank Mike Fast, Jon Roegele, and Aaron Baggett for helpful discussions regarding the data and strike zone definitions in this paper.

LEAGUE EXTERNALITIES AND PERFORMANCE IMPROVEMENTS IN MLB

"If they did get a machine to replace us, you know what would happen to it? Why, the players would bust it to pieces every time it ruled against them. They'd clobber it with a bat."
-Hunter Wendelstedt

I. INTRODUCTION

It is well known that offensive output in Major League Baseball (MLB) has decreased substantially in recent years, from 4.86 runs per team per game in 2006 to 4.07 runs per team per game in 2014. This decline is often attributed to the enforcement of new performance enhancing drug (PED) policies put in place in 2006, an issue not foreign to sports economists interested in productivity, firm output, and demand (Domazlicky & Kerr, 1990; Tainsky & Winfree, 2008; Dinardo & Winfree, 2010; De Vany, 2011). However, while PED policy may have been a contributor to the offensive decline, there were other less salient changes taking place during this time. These other changes could have deleterious effects on any evaluation of the success of PED policy implementation.

Specifically, as noted in Parsons, Sulaeman, Yates, and Hamermesh (2011), it can be misleading to assume performance statistics in sports like baseball are purely a product of objective talent outcomes. Rather, these measures implicitly include influences from subjective evaluation of others, such as game officials. Therefore, this paper investigates the externalities that arose from technological innovations in monitoring and evaluation in MLB that improved the performance (accuracy) of its umpires, originally noted in Mills (2014b). Ultimately, Mills (2014b) notes that the performance of umpires—in this case, accuracy of ball and strike calls—has been starkly improving since 2008.

Based on these accuracy increases specific to balls and strikes, I find that umpires' improvement has come largely at the expense of offense, with the bottom of the called strike zone extending downward an additional three inches between 2008 and 2014. Regression estimations reveal that as much as 65 percent of the drop in scoring can be attributed to improvements in umpire accuracy and expansion of the strike zone to match the rulebook specified zone. Further, as much as 85 percent of the reduction in league walk rate and 27 percent of the increase in league strikeout rate may be attributed to umpire behavioral changes as they relate to ball-strike calls.

These results have important implications for measuring player performance, as well as understanding the success of steroid policies implemented near the same time, often credited as the main driver of offensive decreases across the league. Further, given that past research has found positive demand implications for increased offense, expansion of the strike zone and reduction of PED use could pose a productive efficiency problem for the league. As consumers demand both offense and game integrity, a balance between these league and game characteristics is required for the league to maximize its profits from the product on the field.

II. BACKGROUND AND JUSTIFICATION

Literature

The estimation of professional sports performance has a rich history in the economics literature, beginning with Scully's (1974) seminal contribution to the field. Substantial interest in the measurement of sporting performance in economics lies with estimating players' marginal revenue products (MRP), setting the stage for understanding profit maximization, determining long-term labor contracts, and evaluating league policy effects on the sports labor market (Scott, Long, & Sompii, 1985; Berri & Simmons, 2011; Hakes & Turner, 2011). Further, Kahn (2000) notes the usefulness of sports data to learn lessons about labor markets in general. In particular, many studies have evaluated the stochastic production frontiers and technical efficiency of both players and managers across many sports (Fizel & Ditri, 1996; Haas, 2003; Jewell & Molina, 2004; Kahane, 2005; Fort, Lee, & Berri, 2008).

A pivotal portion of the MRP measurement is, of course, estimating demand for teams and leagues using a number of the characteristics initially proposed by Rottenberg (1956). Most demand estimations include effects of team quality, uncertainty of outcome, and price (Villar & Guerrero, 2009). This estimated impact of wins on revenues allows identification of the proportion of revenue attributable to each player's contribution to fan demand. Ultimately, these estimates can be used to infer player MRP in a competitive free agent labor market and evaluate whether they are being exploited in the context of cooperative professional sports leauges.

More recent work has found evidence of both superstar effects above and beyond performance alone (Hausman & Leonard, 1997; Lucifora & Simmons, 2003; Schmidt, Berri, & Brook, 2004; Berri & Schmidt, 2006; Lawson, Sheehan, & Stephenson, 2008), and a consumer preference for more offensive play, particularly in baseball (Tainsky & Winfree, 2008; and Domazlicky & Kerr, 1990). These findings bring about the question of whether MLB was complicit in the use of PEDs in the 1990s and early 2000s. As PEDs like steroids are oft purported as drivers of offensive prowess—and superstardom for those taking them—it seems likely that baseball benefitted economically from their use.

Tainsky and Winfree (2008) specifically note that performance enhancing drug (PED) use may have positively influenced demand—and relatedly, disparity in player pay—and economists and statisticians have further attempted to measure the impact of steroids on scoring and player performance. Findings on the impacts of PEDs in academic and more casual study settings have been mixed (Groothius, Rotthof, & Strazicich, 2015; Schmotzer, Kilgo &

Switchenko, 2013; Nieswiadomy, Strazicich, & Clayton, 2012; DeVaney, 2011; DiNardo & Winfree, 2010; Bradlow, Jensen, Wolfers, & Wyner, 2008; Tainsky & Winfree, 2008; Cole & Stigler, 2007). During the time of known widespread steroid use in MLB—particularly through the late 1990s—revenues increased dramatically alongside soaring offense and the emergence of a number of superstar hitters that garnered substantial media attention like Mark McGwire, Sammy Sosa, Alex Rodriguez, and Barry Bonds.

However, despite the surge in revenues and offense at the end of the 20th Century, MLB ultimately started PED testing after 2002, and implemented official policies after 2005 under congressional pressure. Since the implementation of these policies, it has been well-documented that the league has experienced substantial decreases in offense, decreases in home runs, and increases in strikeouts (Rymer, 2013; Henderson, 2011). These changes have resulted in the new PED policies being trumpeted as a resounding success (Gaines, 2013). Ultimately, the 2010 season was labeled the “Year of the Pitcher” (Dubner, 2010), and since 2010 offense continued its decline through 2014.

The impact of the use of steroids on increasing offense during this time—and decreasing offense after policies were put in place—is unclear. In particular, it is well documented that both pitchers and hitters were taking PEDs during the time of increased offense. If the use of PEDs increased offense, then there is an implicit assumption that they affected batters substantially more than pitchers. Further, some of the game’s most prodigious home run hitters have been caught continuing to use performance enhancing drugs since 2006, leaving the question as to the actual effectiveness of the policy as a deterrent. Third, and most relevant to this work, there have been other shifts in policy and game play since the implementation of new PED enforcement, which may have negatively impacted offense in baseball. Without accounting for these effects, measurements of the impact of PED testing on offense will be misleading.

Presumably, MLB intended to rid the league of PEDs to assure fans of the fairness and so-called "integrity" of the game, and ensure continued interest in game outcomes. This brings about a (simplified) demand function such that the league must maximize $D(\beta, \theta, \tau)$. In this generalized characterization, demand for sport is a function of vectors of characteristics of games and leagues. β represents a vector of the traditional game, team, fan, and market characteristics like uncertainty of outcome (or excitement and surprise), superstar effects, income, population, loss aversion (Coates, Humphreys, & Zhou, 2014), price, and the goodness of substitutes. θ is a vector of characteristics describing the offensive portion of game play, and τ is consumers' perception of game fairness or integrity as it relates to the stated rules of the game. τ can include both enforcement of on-field rules—such as the correctly called strike zone—or the enforcement of rules against gambling or PEDs. I restrict the discussion to include expansion of the called strike zone and enforcement of PED policy.

I assume both $\frac{\partial D}{\partial \theta} > 0$ and $\frac{\partial D}{\partial \tau} > 0$, or in other words, demand increases with both increases in the perception of game integrity and with increases in game offense (as in Tainsky & Winfree, 2008). During the PED era, the league experienced increases in θ, which eventually led to decreases in τ and Congressional intervention from steroid suspicions. In this vein, MLB can address two issues by giving the perception of decreased run scoring caused by PED policy and increasing the performance of its umpires, and issue often maligned by fans. The decline in offense has presumably given fans confidence that the league has cleaned up the game, or $\frac{\partial D}{\partial \tau} > 0$. But there are likely side effects of this enforcement in that $\frac{\partial \theta}{\partial \tau} < 0$, and as such a ceteris paribus decrease in demand. There are therefore productive efficiency considerations. As the league adjusts performance of one aspect of game production—its umpires—it comes at a cost of another portion of production: offense.

This raises an important question: is it possible for Major League Baseball to have decreased offense using less salient policies that resulted in the perception of decreased steroid use, whether intentionally or unintentionally? While I do not directly measure the production frontier in this work as it relates to offense and perception of game integrity, this paper directly tests the effects of the enforcement of a uniform strike zone among umpires through new technological innovations as exhibited in Mills (2014b). Or, in other words, whether $\frac{\partial\theta}{\partial\tau} < 0$ in the context of τ as strike zone integrity.

Umpire Influences on Performance Measurement

A large body of work has developed an understanding of biases among sports officials as they relate to the performance across various characteristics of MLB players and game or contextual situations (Kim & King, 2014; MacMahon & Starkes, 2008; Mills, 2014a), noting that arbitrators can have clear impacts on the outcomes of games and differential performance of players based on these variables. And a substantial body of work has identified a number of social pressures on referees and umpires in the professional sports setting (Lane, Nevill, Ahmad, & Balmer, 2006; Garicano, Palacios-Huerta, & Predergast, 2005; Nevill, Balmer, & Williams, 2002; Pettersson-Lidbom & Priks, 2010; Price & Wolfers, 2010; Price, Remer, & Stone, 2012). Specifically, Price, Remer, and Stone (2012) note that many of these biases may be incentivized by the league—or at least not heavily punished—given that they could increase league profits at the margin. The (lack of) salience of the biases or rule enforcement changes among game officials are key to the ability to give fans the perception of fairness and game integrity.

Both the start of PED testing and the Year of the Pitcher were approximately concurrent with innovations in monitoring and evaluation of the league's umpires that are known to have improved their accuracy when calling balls and strikes (Mills, 2014b). Given that the strike zone is perhaps the most pivotal strategic interaction on the field, there are likely to be induced impacts on player performance that are not apparent in offensive outcome measures when evaluating decreases in offensive productivity as a measure of PED policy effectiveness. [2] Further, the ability for the average fan to reliably observe small changes in ball-strike calls among umpires is limited, making changes to the enforcement of the rule book strike zone less salient that other rule changes.

Ultimately, in measuring performance, each of the factors involved in producing a players' output, including subjective judgments made by other agents, should be credited accordingly. These subjective judgments are largely unrelated to player skill. There exists a large body of work identifying factors both internal and external to the players themselves that impact measurement of game performance at the individual level (Kim & King, 2014; Mills, 2014a; Parsons et al., 2011; Price & Wolfers, 2010; Tainsky, Mills, & Winfree, 2015). Parsons et al. (2011) make the astute observation that these unobserved effects are particularly important in the context of evaluating discrimination in compensation using standard performance statistics that do not account for impacts of subjective judgment by officials. This is also true at the league level, where offensive declines are credited to other assumed policy changes like PED enforcement.

[2] During this time, there were other relevant changes beyond the scope of this work. First, there were mandated changes in the size of the bat barrel allowed in MLB (MLB Rule 1.10(a)). Secondly, there is speculation as to possible changes in the manufacturing of balls (Jaffe, 2012). From a gameplay standpoint, there have also been substantial improvements in defensive and reliever use strategies with the growth of sabermetric analysis in MLB front offices.

Umpire Tasks and Context

Ball and Strike Calling as the Primary Umpiring Duty. While umpires are given a number of tasks as arbiters throughout a baseball game, one of the most important and salient duties is to judge pitches thrown by the defensive team's pitcher to the offensive team's batter. The batter has the choice of whether or not to swing at a pitch in an attempt to hit it into the field of play. If the batter does not swing, the umpire's task is to judge whether the thrown ball passed through the strike zone, an area where it would be reasonable to expect the batter to be able to hit the ball, over home plate.

If the ball is judged to cross through the zone—an imaginary box—then it is considered a strike. If it does not pass through this zone, then it is considered a ball. Batters are allowed 3 strikes per at bat before they are out, while pitchers are allowed 4 balls before the batter is given first base. Therefore, batters benefit from a smaller strike zone (less pitches considered strikes, conditional on location), while pitchers benefit from a larger strike zone (more pitches considered strikes, conditional on location). It is important to note that the job of the umpire in this scenario is not an easy one: pitches are thrown with a velocity as much as 100 miles per hour or more, and often have various non-traditional trajectories. Further, there is no box drawn such that the umpire can reference the location of the pitch. Therefore, this task requires both snap judgment and rather precise visual acuity to perform well, setting the stage for task performance with room for improvement and substantial variation across individuals.

Umpire Improvements with Monitoring and Training. Mills (2014b) documents clear changes in umpire behavior beginning as far back as 2001, noting that the league directed umpires to call more strikes and reduce pitch counts in games beginning that year. This was the first year in which the league introduced the QuesTec monitoring system to ensure umpires were following this directive. That year, the rate of called strikes grew by more than five percent. Scoring also dropped precipitously, down more than ten percent by the end of 2002. These reductions occurred prior to the league taking any action over the use of PEDs in 2003, with its official policy regarding testing and suspension not arising until 2006.

One year after the implementation of new steroid enforcement policies in 2006, the league also introduced a new system to its ballparks to identify pitch location and other characteristics of each pitch in 2007, colloquially known as Pitch f/x. The data have been publicly available since this time, and can be directly tied to ball-strike calls by the umpires to evaluate accuracy. Subsequently, in 2009, the league introduced an updated evaluation system using this data, known as the Zone Evaluation System, which replaced QuesTec. Mills (2014b) shows that after implementing this system, umpires substantially improved their accuracy on called balls and strikes, an outcome not unexpected in the context of evaluation and feedback among motivated agents such as elite professional umpires (Taylor & Tyler, 2012).

This finding lays the foundation for the inquiry in this work. Here, I use the lessons from prior documentation of behavioral changes to estimate their impact on the net size of the strike zone, changes in pitcher and batter strategic behavior, and offensive production in MLB since just before the implementation of the Zone Evaluation System in 2008.

III. DATA AND ESTIMATION PROCEDURE

Data and Measurement

The data used for measuring umpire accuracy was gleaned from BaseballSavant.com, and consist of Sportvision's Pitch f/x data for every MLB pitch from 2008 through 2014. The data include the location of each pitch as it crosses the front of home plate, allowing for the identification of false positive and false negative rates for each individual umpire in each season. In total, there are over 4.9 million pitches in the data set, with more than 2.5 million subject to judgment by the home plate umpire. Tabulation of accuracy rates at the umpire level were merged with umpire-season level statistical measures recorded from umpire statistical reports at Baseball Prospectus. These were used to evaluate the change in scoring attributable to umpire strike and ball accuracy rates.

I identify correct umpire ball-strike calls by defining the strike zone identically to Mills (2014b). To review, I use information on relative knee, waist, and shoulder position of males from NASA's Human Integration Design Handbook (NASA, 2000) applied in the context of the average height of MLB batters in the data set, or approximately 73.5 inches. The lower and upper boundaries of the strike zone are fixed in place at approximately 18.2 and 41 inches, respectively, scaled from NASA to the average height value. The width of the strike zone is measured as the 17 inch plate width as noted in the official rules plus the radius of the baseball on either side, as a pitch is considered a strike even if a portion of the ball crosses over the plate. Pitch f/x measurements are associated with the center of the baseball, so with the added radius results in a strike zone width of approximately 19.92 inches.

Using these strike zone parameters, I aggregate the total number of pitches crossing through the zone (*CorrectStrikes*), and the total number of pitches failing to cross this strike zone plane that are called balls (*CorrectBalls*). The sum of these two aggregations reveals the total number of *CorrectCalls*. The number of *CorrectStrikes* is divided by the total number of pitches crossing the strike zone plane to calculate the *CorrectStrikeRate*, and the number of *CorrectBalls* is divided by the total number of called pitches that did not cross the strike zone plane to calculate the *CorrectBallRate*. *AccuracyRate* is calculated by dividing *CorrectCalls* by the total number of pitches called by the umpire. This data is summarized in Table 1, and matches closely to that of Mills (2014b). Variability in these rates and their respective changes across individual umpires are also presented graphically in Figure 1. Figure 2 then presents the rate at which pitches were called strikes from 2008 through 2014 at the hollow beneath the knee—an area of the bottom of the strike zone explicitly noted in the MLB Rulebook—exhibiting that many of these additional strikes took place between 18 and 21 inches off of the ground.

Estimation Technique

Strike Zone Size. I estimate a generalized additive model (GAM) using the locational data on called pitches to measure its size and shape across seasons to determine if the net size of the strike zone has changed or shifted. This method has been used in past work investigating bias in umpire ball-strike calls (Mills, 2014a; Tainsky et al., 2013). While a summary of the *AccuracyRate* is useful to some extent, it is important to note that not all correct calls are created equal. For example, a pitch at the very edge of the strike zone is much more difficult to judge than one thrown directly down the middle. Therefore, the generalized additive model allows for a non-parametric estimation of the strike zone surface and associated changes in strike probability and uncertainty over the call as it relates to pitch location. GAMs have the advantage of identifying the more rounded edges of the zone, and fit separate surfaces for batter handedness, without over-fitting the data (Mills, 2014a). A full discussion of GAMs can be found in Wood, (2000; 2003; 2004; 2006; 2011) and Gu and Wahba (1993), but I reprise the basic method developed in Mills (2014a) for pitch location data here. The flexible model allows measurement of changes in the likelihood of a strike call conditional on the location of the pitch when it crosses the plate. This is a binomial logistic type model, but is fully non-parametric. The model is estimated through generalized cross-validation and penalized iteratively re-weighted least squares and is much more flexible than the use of parametric estimation and polynomial representations of the strike zone edges. The GAM takes the general form:

$$g(\mu_{it}) = f_{bt}(Z_{hit}, Z_{vit}) + \varepsilon_{it}$$

Where the response variable, y_i—a variable indicating whether the pitch i thrown in year t was called a strike (strike=1 and ball=0)—has mean μ_i, with $g(\cdot)$ representing the logit link function for the binomial response. The unknown smooth function $f_b(Z_{hit}, Z_{vit})$ is estimated jointly for vertical and horizontal location—indexed by h and v, respectively—using a generalized cross-validation procedure to avoid overfitting, with the right and left handed batter surfaces indexed by b.[3]

From this model, I identify an irregular elliptical contour boundary at which the probability of a called ball is equal to that of a called strike, or alternatively, a 50 percent probability of a strike call. The area inside this boundary is the empirically derived called strike zone, while the area outside the boundary is the called ball zone. This is further used to estimate the surface area inside each probability contour to identify changes to the actual size of the called strike zone.

Induced Behavioral Changes. Following from the findings related to low strike rates, I subsequently evaluate changes in pitcher and batter behavior that may result from changed strike calling patterns of umpires. Specifically, I measure aggregate changes in the propensity for

[3] Smoothing parameter estimation results available from the authors upon request. They are not presented here, as the non-parametric smoothing parameter is not particularly interpretable as a raw estimate of degrees of freedom, and better presented visually as in Figure 3.

pitchers to throw low pitches—defined as pitches thrown below 21 inches—and pitches low in the strike zone—defined as those pitches that are between 18 and 21 inches in height, also known as the "hollow beneath the knee." Further, I measure aggregate changes in the rate at which batters swing in general, at pitches within the strike zone, at pitches outside the strike zone, and specifically at low pitches (below 21 inches in height).

Of course, it is possible that changes following strike zone movement are simply a movement of equilibrium pitch choice and swing rates, without affecting offense. Therefore, , I evaluate the propensity with which hitters are able to make contact with pitches low in the zone (*Contact*), put these pitches into the field of play (*InPlay*), and get a hit when swinging at these pitches (*Hit*). For this portion of the analysis, I estimate a logistic regression model with fixed effects for the season in which each pitch is thrown. Low pitches are measured by a dummy variable indicating that the pitch cross the plate at lower than 21 inches off the ground (*LowPitch*), and alternatively measured by the number of inches below the strike zone height center (*InBelowCenter)*. These regressions take the form:

$$y_{it} = \beta_0 + \beta_1 LowPitch_{it} + \delta_t + \varepsilon_{it}$$

$$y_{it} = \beta_0 + \beta_1 InBelowCenter_{it} + \delta_t + \varepsilon_{it}$$

Here, y_{it} represents one of three dependent variables, *Contact*, *InPlay*, or *Hit*, for pitch *i* in season *t*. β_0 is an intercept for the model, while β_1 represents the estimated impact of the low pitches on the rate of contact, balls put in play, or hits. Lastly, δ_t identifies the fixed effect for each season, *t*, and the pitch specific error term is represented by ε_{it}.

Net Offensive Impact Estimation. As the central part of this investigation, I use individual umpire data to estimate the impact of aggregate ball-strike accuracy at the umpire-season level on Earned Run Average (*ERA*), Strikeouts per Nine Innings Pitched (*K9*), and Walks Per Nine Innings Pitched (*BB9*). In these estimations, I use a panel regression with umpire and season effects. The explanatory variables of interest are individual umpire-season *CorrectBallRate* and *CorrectStrikeRate.* These are used to identify changes strike calling behavior across umpires and across seasons. Observations are restricted to those umpires that called at least 500 pitches in a given season. The model takes the form:

$$y_{it} = \beta_0 + \beta_1 CorrectStrikeRate_{it} + \beta_2 CorrectBallRate_{it} + \delta_i + \tau_t + \varepsilon_{it}$$

Where y_{it} is the *ERA*, *K9,* or *BB9* of umpire *i* in season *t*, δ_i are the umpire-specific fixed effects, τ_t are the season-specific fixed effects, and ε_{it} is the individual umpire-season error term, clustered by umpire. I estimate this model first with fixed effects, and subsequently with the between estimator.[4] Subsequently, the coefficients from these models are applied to the total change from 2008 to 2014 to identify the expected change in each measure that can be attributed to the net changes in umpire ball-strike accuracy rate improvements. This is then taken as a percentage of the true change in each measure over the same time period that is attributable to the umpire behavioral changes as it relates to the strike zone.

The fixed effects and between effects make different assumptions about the data in their respective estimations. First, the fixed effect estimator evaluates the change in each statistic as a function of each umpire's own change in accuracy rate across each season, taking advantage of the variation across both time and umpire. For robustness, I also use the between estimator uses

[4] For the between estimator, standard errors clustered by umpire are not used.

within-season variation across umpires to infer the expected changes in offensive output based on the known changes to overall ball-strike rates across seasons.

It is important to note that there could be structural shifts in the way the game is played that are induced by changes in the umpire strike zones, ultimately resulting in underestimation of the total effect of umpire strike zone changes using the between estimator. The between estimator would not likely include this effect in its coefficient estimates. Specifically, it could be difficult for players to change their strategies at the game-by-game level within season when umpires change every game. However, they may do so across seasons at the league aggregate level. Therefore, structural changes in strategy—such as swinging more at lower pitches, or throwing more low pitches—are likely to result in an underestimation of the induced effects of umpire strike zone changes if they change between seasons across the entire league (and are induced by umpires), but not across umpires within a given season. I present both models, and note that the between estimator is likely to be conservative. However, it may better isolate the umpire-specific contribution to changes in offensive levels in the league.

IV. RESULTS AND DISCUSSION

Accuracy and Strike Zone Size Increases

As noted earlier, summaries of league-level correct strike rate, correct ball rate, and overall accuracy rate from 2008 through 2014 can be found in Table 1, with Figure 1 presenting the changes over time for the league and individual umpires. While I do not reprise the entirety of the analysis of Mills (2014b), I note that there is substantial evidence of umpire improvements over time. Most importantly, the data exhibit larger improvements in correct strike rates than correct ball rates, indicating a likely net expansion of the strike zone called by MLB umpires.

Figure 2 presents the rate at which pitches low in the strike zone—between 18 and 21 inches off of the ground, or the "hollow beneath the knee"—are called strikes by umpires in each year of the data. It is easy to see from this visual that for all umpires, the rate at which strikes are called on these pitches has increased dramatically from about 25 percent of the time in 2008 to more than 60 percent of the time in 2014. These pitches are notoriously more difficult to hit, revealing the possibility that there may be induced changes from calling these pitches strikes substantially more often.

Results of the non-parametric strike zone estimation for 2008 and 2014 are exhibited visually in Figure 3. There is a clear narrowing and downward extension of the strike zone for both right-handed and left-handed batters over the past 7 years. The estimated net change in surface area of the two-dimensional strike zone—calculated from the plane using yearly generalized additive models—is presented in Table 2 for the 50 percent and 90 percent contours. Note that the more certain 90 percent strike zone contour has grown by substantially more than the 50 percent contour in both square inches and as a percentage of its size in 2008, indicating that the edge of certain strikes is moving out toward the edges of the rulebook strike zone more quickly, again indicating an increase in accuracy for umpires as a whole. The size of the increase in surface area is equivalent to approximately 6 to 7 baseballs lined up across the bottom of the zone below the 2008 contour.

Induced Effects: Pitcher and Batter Behavioral Changes

Table 3 presents the rate at which pitchers have thrown pitches to different portions of the strike zone from 2008 through 2014, using all pitches in the Pitch f/x database with locational data included. Note that the proportion of pitches thrown to the rulebook strike zone have not changed over this period, making clear that a large portion of the increase in strike over this period must be attributed to increases in the umpires' propensity to call more strikes—as seen in Figure 3—rather than pitchers throwing nearer to the center of the existing zone. Rather, there is clear evidence that pitchers are throwing *away* from the center of the zone more often. Columns 4-7 in Table 3 show that pitchers have seemingly recognized the lower half of the strike zone is being called more often by umpires, and are therefore more likely to choose to throw pitches to that portion of the zone. The proportion of low pitches—defined as those below 21 inches from the ground—has increased by nearly 25 percent (3.45 percentage points), from a rate of 22.1 percent in 2008 to 26.2 percent in 2014. [5] Additionally, the propensity for pitchers to throw low and within the zone (between 18 and 21 inches high) as a percentage of both all pitches, and as a percentage of pitches within the zone, have each increased by more than 18 percent. Finally, the average pitch height has decreased from nearly 28.9 inches high, to 27.26 inches, or a decrease of 5.55 percent.

It is common rhetoric within baseball that lower pitches are more difficult to hit, indicating that this umpire-induced change in pitch location could be contributing to the decline in offense in the league through this apparent induced behavioral change as well. This hypothesis can, of course, be tested. I test whether batters are becoming more likely to swing at these low pitches—as well as whether the outcomes from pitches is reduced relatively to other pitches—using all pitches which umpires did not subjectively judge during the Pitch f/x era: pitches at which the batter swings. Table 4 presents basic summaries of the swing rates for all pitches, pitches within the rulebook strike zone, pitches outside the rulebook strike zone, and on pitches that are below 21 inches in height. Each of these rates has experienced a statistically significant

[5] Similar results hold for other definitions of a low pitch.

increase since the 2008 season. Most strikingly, the rate at which batters swing at pitches categorized as "low" has experienced the largest change, increasing by 11 percent over this time.

Table 5 presents estimates of the changes in success of batter swings as it relates to the vertical location of using a simple logistic regression with dummy fixed effects for season, ball-strike count, pitch type, and individual umpire. There are large, statistically significant decreases in contact rate, in-play rate, and hit rate on lower pitches, making clear the additional impact of the changes in the size and shape of the zone could also have on balls at which the batter swings. Specifically, when a batter swings at a pitch that is below the low-pitch boundary of 21 inches, the odds of making contact decrease by approximately 73 percent, the odds of putting the ball in play decrease by 48 percent, and the odds of a hit decrease by 26 percent. There are also clear negative impacts for swinging at a pitch as it inches further and further below the strike zone center.

From a more general standpoint, it is clear from this data that changes in subjective evaluation may not only change our measures of performance, but also induce changes in behavior of other relevant agents. Taken together with the increased propensity with which pitchers throw to this area, the induced increases in swings at these low pitches, and the estimated disadvantage to hitters when swinging at these pitches point toward a possible negative effect of the enhanced umpire strike zone on offensive output across the league. Depending on the behavioral change that follows changes in evaluation, there could be positive or negative effects on a firm, as employees prioritize one task over another due to evaluation focus. The next section directly estimates umpire strike zone impacts on relevant measures of batter results.

Net Offensive Impacts

While much of the basic changes to umpire strike zones—in addition to apparent strategic batter and pitcher changes induced by umpires—are interesting in their own right, the final point of this inquiry is the league level impact of these performance changes. To explicitly estimate the impact of umpire behavior on offensive levels in MLB, I exploit the variation in individual umpire accuracy rates within each season and associated changes in runs and other offensive events. Table 6 presents the results of panel regression estimates of the relationship between these individual umpire accuracy rates and *ERA*, *K9*, and *BB9*.

Each of the fixed effects models estimates statistically significant impacts of correct strike and correct ball call rates of umpires in each year in the expected directions, while the between estimator effects models results in similar estimates, with the exception of the impact of *CorrectBallRate* on *ERA*. The general results of these models are as follows: as the rate of correct strikes increases, *ERA* is reduced (fewer runs allowed by pitchers per game), *K9* increases (more batters strike out per game), and *BB9* decreases (fewer bases on balls per game). For changes in the rate of correct balls, the effects are in the opposite, expected direction for each measure. Again, I note that these estimations use accuracy, rather than overall ball and strike rates, as independent variables. Therefore, the effects exhibited here are directly attributable to change in umpire ball-strike calling, rather than some other locational change by the pitcher.

Noting that the rate correct strike calls have increased by substantially more than correct ball calls, the net effect on offensive statistics appears to be negative as they relate to umpire accuracy improvements over the sample period for this empirical analysis (Table 7). I use the coefficients from these regression models as an estimate of impact of the aggregate per-percentage point change for each of the dependent variables. The aggregate change is then calculated by simply multiplying the change from 2008 to 2014 in *CorrectBallRate* and *CorrectStrikeRate* by the associated coefficient estimate. The expression below gives a change in each measure, ΔY, that would be expected based on the changes in umpire ball-strike calling behavior. Note that β_1 and β_2 come directly from the regression expression presented in Section III.

$$\Delta Y = \beta_1 * (\Delta_{2008-2014} CorrectStrikeRate) + \beta_2 (\Delta_{2008-2014} CorrectBallRate)$$

Over this time period, the rate of correct strikes increased by 8.419 percentage points ($\Delta_{2008-2014} CorrectStrikeRate = 8.419$) and correct balls increased by only 1.402 percentage points ($\Delta_{2008-2014} CorrectBallRate = 1.402$) from 2008 to 2014. Using the expression above, *ERA* is estimated to decrease by 0.35 to 0.37 runs per team, per nine innings due to umpire strike zone changes alone. This amounts to more than 60 percent of the total change in *ERA* in MLB during this time for the year trend. Further, approximately 20 to 27 percent of the increase in *K9* and 65 to 85 percent of the reduction in *BB9* can be attributed to umpire accuracy improvements, respectively. The relatively low proportion of *K9* attributable to umpire accuracy changes is interesting; however, it seems likely that a proportion of this increase is a result of strategic changes in the way high strikeout relief pitchers are used, and a movement toward higher velocity pitchers in general across the league.

Taken together, there is clear evidence that umpire behavior changes have strongly influenced offensive output in MLB. The simple changing of a ball to a strike has direct impacts on the game, but also induces other behaviors among pitchers and hitters. Pitchers continue to increase the rate at which they throw pitches low in the zone—where they know they are now more likely to receive strike calls from the umpire—and batters are induced to swing at these pitches more often, which are shown to have negative effects on the likelihood of making contact or getting a hit. Ultimately, there are stark and substantive effects of arbiter performance changes on the production of the baseball game.

V. SUMMARY AND CONCLUSIONS

This paper identifies the impact of improvements in the performance of MLB umpires on the overall offensive output in the league. While enforcement of penalties for PED use are often credited with a strong decline in offense in MLB since 2006, umpires have likely had a larger influence than any other factor in this decline. This highlights an important consideration for sports organizations and their PED policies specifically—that investigators must account for other structural changes in the game that are otherwise unobserved in, or implicit in, traditional performance measures.

Recently, MLB has announced that they will evaluate ways to increase offense (Passan, 2015; Rosenthal, 2015), possibly in response to a stated lack of interest in such a low scoring version of the game (Tainsky & Winfree, 2008). Though, this effort is seemingly more preemptive, as recent contract negotiations for the league have not indicated a slowdown in consumer interest. Nevertheless, this seems to indicate that the league recognizes its consumer base's interest in higher scoring contests.

From an industrial organization standpoint, policies impacting offense in baseball could affect league competitive balance. In particular given that contests would necessarily be closer in lower scoring environments, games would necessarily be closer in terms of discrete run scoring. The ultimate balance impact depends largely on the distribution of run scoring and relative changes in mean and variance across teams. Ultimately, manipulation in less salient areas of league policy could affect competitive balance in a sports league positively or negatively as it relates to consumer demand. Further, from the perspective of sports labor, changes in the strike zone could asymmetrically affect the attractiveness of certain players over others in the labor market. For example, if some hitters are better at hitting low pitches, their relative value to teams could increase and net them larger contracts than they would otherwise receive. The emergence of historically great low-pitch hitters during this time—such as Mike Trout (Sullivan, 2014)—may provide an interesting new context for testing asymmetric effects that follow policy changes, rather than innovation shocks as in Hakes and Sauer (2006).

Lastly, the findings from the present empirical evaluation point to more general implications for organizational performance and measurements of production that implicitly include subjective judgment. First, the performance improvement of a small number of agents within a firm that hold considerable judgmental power could have impacts on other forms of production within the firm. For example, if managers are monitored to ensure more accurate evaluation of subordinates, then there could be behavioral changes among these subordinates that prioritize certain portions of production over others. Secondly, measured changes in the quality of performance from evaluation should be considered carefully, as these measures implicitly include changes in the evaluation procedure itself, rather than changes in the actual production of employees. As a whole, it seems clear that unexpected productive efficiency issues can arise, even in the face of improvement in performance of a given employee.

REFERENCES

Berri, D. J. and Schmidt, M. B. "On the road with the National Basketball Association's superstar externality." *Journal of Sports Economics 7*, 2006, 347-358.

Berri, D. J., Schmidt, M. B., and Brook, S. L. "Stars at the gate: The impact of star power on NBA gate revenues." *Journal of Sports Economics 5*, 2004, 33-50.

Berri, D. J. and Simmons, R. "Catching a draft: On the process of selecting quarterbacks in the National Football League Amateur draft." *Journal of Productivity Analysis 35*, 2011, 37-49.

Bradlow, E., Jensen, S., Wolfers, J., and Wyner, A. "Report backing Clemens chooses its facts carefully." *New York Times*. Retrieved July 24, 2015 from: http://www.nytimes.com /2008 /02/10/sports/baseball/10score.html, 2008.

Coates, D., Humphreys, B. R., and Zhou, L. "Reference-dependent preferences, loss aversion, and live game attendance." *Economic Inquiry, 52*, 2014, 959-973.

Cole, J. R. and Stigler, S. M. "More juice, less punch." *New York Times*. Retrieved July 24, 2015 from: http://www.nytimes.com/2007/12/22/opinion/22cole.html, 2007.

De Vany, A. "Steroids and home runs." *Economic Inquiry 49*, 2011, 489-511.

DiNardo, J. and Winfree, J. "The law of genius and home runs refuted." *Economic Inquiry 48*, 2010, 51-64.

Fizel, J. L. and Ditri, M. P. "Estimating managerial efficiency: The case of college basketball coaches." *Journal of Sport Management 10*, 1996, 435-445.

Fort, R. D., Lee, Y. H., and Berri, D. "Race, technical efficiency, and retention: The case of NBA Coaches." *International Journal of Sport Finance, 3*, 2008, 84-97.

Gaines, C. "Major League Baseball has gotten very good at getting rid of steroids." *Business Insider*. Retrieved September 1, 2015 from: http://www.businessinsider.com /major-league-baseball-home-runs-steroids-2014-9, 2013.

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

Groothuis, P. A., Rotthoff, K. W., and Strazicich, M. C. "Structural breaks in the game: The case of Major League Baseball." *Journal of Sports Economics* DOI: 10.1177/1527002515593113, 2015.

Haas, D.J. "Technical efficiency in Major League Soccer." *Journal of Sports Economics 4*, 2003, 21-38.

Hakes, J. K. and Saur, R. D. “An economic evaluation of the Moneyball hypothesis.” *The Journal of Economic Perspectives, 20*, 2006, 173-186.

Hakes, J. K. and Turner, C. “Pay, productivity and aging in Major League Baseball.” *Journal of Productivity Analysis 35*, 2011, 61-74.

Hausman, J. A. and Leonard, G. K. “Superstars in the National Basketball Association: Economic value and policy.” *Journal of Labor Economics 15*, 1997, 586-624.

Jewell, R. T. and Molina, D. J. “Productive efficiency and salary distribution: The case of U.S. Major League Baseball.” *Scottish Journal of Political Economy 15*, 2004, 413-419.

Kahane, L. H. “Production efficiency and discriminatory hiring practices in the National Hockey League: A stochastic frontier approach.” *Review of Industrial Organization 27*, 2005, 47-71.

Kim, J. W. and King, B. G. “Seeing stars: Matthew effects and status bias in Major League Baseball umpiring.” *Management Science 60*, 2014, 2619-2644.

Lane, A. M., Nevill, A. M., Ahmad, N. S., and Balmer, N. “Soccer referee decision-making: Shall I blow the whistle?” *Journal of Sports Science and Medicine 5*, 2006, 243-253.

Lawson, R. A., Sheehan, K., and Stephenson, E. F. “Vend it like Beckham: David Beckham’s effect on MLS ticket sales.” *International Journal of Sport Finance 3*, 2008, 189-195.

Lucifora, C. and Simmons, R. “Superstar effects in sport: Evidence from Italian soccer.” *Journal of Sports Economics 4*, 2003, 35-55.

MacMahon, C. and Starkes, J. L. “Contextual influences on baseball ball-strike decisions in umpires, players, and controls.” *Journal of Sports Sciences 26*, 2008, 751-760.

Mills, B. M. “Social pressure at the plate: Inequality aversion, status, and mere exposure.” *Managerial and Decision Economics 35*, 2014a, 387-403.

Mills, B. M. “Expert workers, performance standards, and on-the-job training: Evaluating Major League Baseball Umpires.” *SSRN Working Paper #2478447*. Accessed August 5, 2015 from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2478447, 2014b.

Nevill, A. M., Balmer, N. J., and Williams, A. M. “The influence of crowd noise and experience upon refereeing decisions in football.” *Psychology of Sport and Exercise 3*, 2002, 261-272.

Nieswiadomy, M. L., Strazicich, M. C., and Clayton, S. “Was there a structural break in Barry Bond’s bat?” *Journal of Quantitative Analysis in Sports 8*(3), 2012, 1-17.

Passan, J. “MLB could alter strike zone as response to declining offense.” *Yahoo! Sports*. Retrieved September 25, 2015 from: http://sports.yahoo.com/news/sources--mlb-could-alter-strike-zone-as-response-to-declining-offense-232940947.html, 2015.

Parsons, C. A., Sulaeman, J., Yates, M. C., and Hamermesh, D. S. “Strike three: Discrimination, incentives, and evaluation.” *American Economic Review 101*, 2011, 1410-1435.

Pettersson-Lidbom, P. and Priks, M. “Behavior under social pressure: Empty Italian stadiums and referee bias.” *Economics Letters 108*, 2010, 212-214.

Price, J. and Wolfers, J. “Racial discrimination among NBA referees.” *Quarterly Journal of Economics 125*, 2010, 1859-1887.

Price, J., Remer, M., and Stone, D. F. “Subperfect game: Profitable biases of NBA referees.” *Journal of Economics and Management Strategy 21*, 2012, 271-300.

Rosenthal, K. “Just a bit outside: Open to change, Manfred the breath of fresh air MLB needs.” *FOX Sports*. Retrieved September 25, 2015 from: http://www.foxsports.com/mlb /just-a-bit-outside/story/mlb-commissioner-rob-manfred-recognizes-boring-needs-change-more-offense-adapt-quickly-012515, 2015.

Schmotzer, B., Kilgo, P., and Switchenko, J. “The natural? The effect of steroids on offensive performance in baseball.” *Chance 22*, 2009, 21-32.

Scott, F., Long, J., and Sompii, K. “Salary vs. marginal revenue product under monopsony and competition: The case of professional basketball.” *Atlantic Economic Journal 13*, 1985, 50-59.

Sullivan, J. “Why Mike Trout – and the rest of the league – is having trouble with the high stuff.” *Just a Bit Outside*. Retrieved December 18, 2015 from: http://www.foxsports.com/mlb/ just-a-bit-outside/story/why-mike-trout-and-the-rest-of-the-league-is-having-trouble-with-the-high-stuff-082914, 2015.

Tainsky, S. and Winfree, J. A. “Financial incentives and league policy: The example of Major League Baseball’s steroid policy.” *European Sport Management Quarterly 8*, 2008, 67-81.

Tainsky, S., Mills, B. M., and Winfree, J. A. (2015). “Further examination of potential discrimination among MLB umpires.” *Journal of Sports Economics 16*, 2015, 353-374.

Taylor, E. S. & Tyler, J. H. “The effect of evaluation on teacher performance.” *American Economic Review 102*, 2012, 3628-3651.

Villar, J. G., Guerrero, P. R. “Sports attendance: A survey of the literature 1973-2007.” *Rivista di Diritto Ed Economia Dello Sport 5*, 2009, 111-151.

TABLE 1 – Umpire Accuracy Summary

Year	Called Pitches	Called Pitches Within Zone	Called Strikes Within Zone	Called Pitches Outside Zone	Called Balls Outside Zone	Correct Strike Rate	Correct Ball Rate	Accuracy Rate
2008	357,195	113,071	89,052	244,124	215,922	78.758	88.448	85.380
2009	372,791	122,108	96,085	250,683	222,781	78.689	88.870	85.535
2010	369,805	122,479	97,991	247,326	220,005	80.006	88.953	85.990
2011	364,211	119,091	97,016	245,120	218,485	81.464	89.134	86.626
2012	357,619	116,436	97,393	241,183	215,646	83.645	89.412	87.534
2013	362,795	116,279	99,652	246,516	221,900	85.701	90.014	88.632
2014	355,478	113,208	98,691	242,270	217,679	87.177	89.850	88.998

Note: Mills (2014b) presents a full treatment of the changes across this period, noting that the improvements in accuracy are both statistically and practically significant from the perspective of umpire strike zone behavior.

FIGURE 1 – Change in Overall Umpire *AccuracyRate* (top), *CorrectStrikeRate* (bottom left), and *CorrectBallRate* (bottom right)

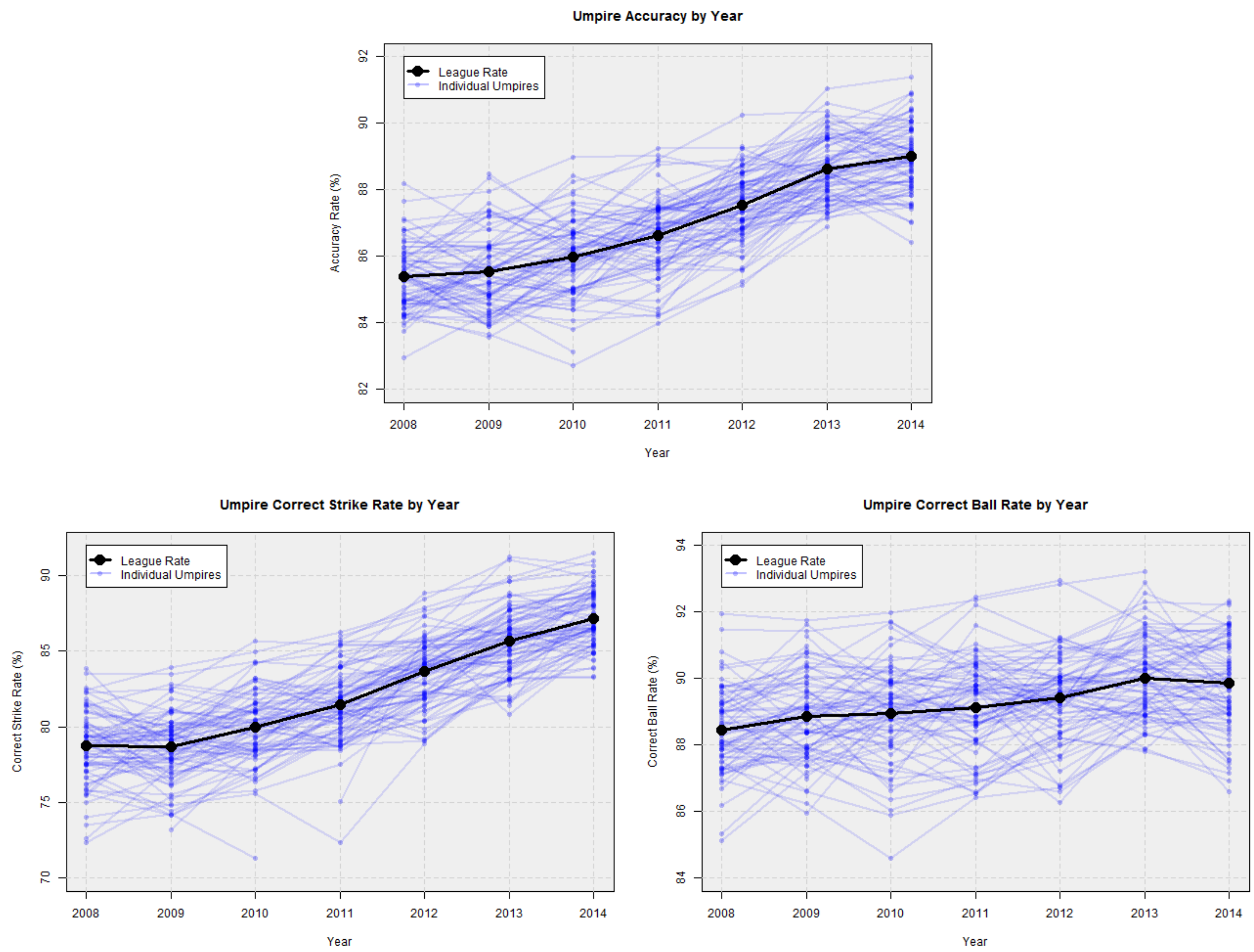

FIGURE 2 – Low Strike Rate Changes by Year

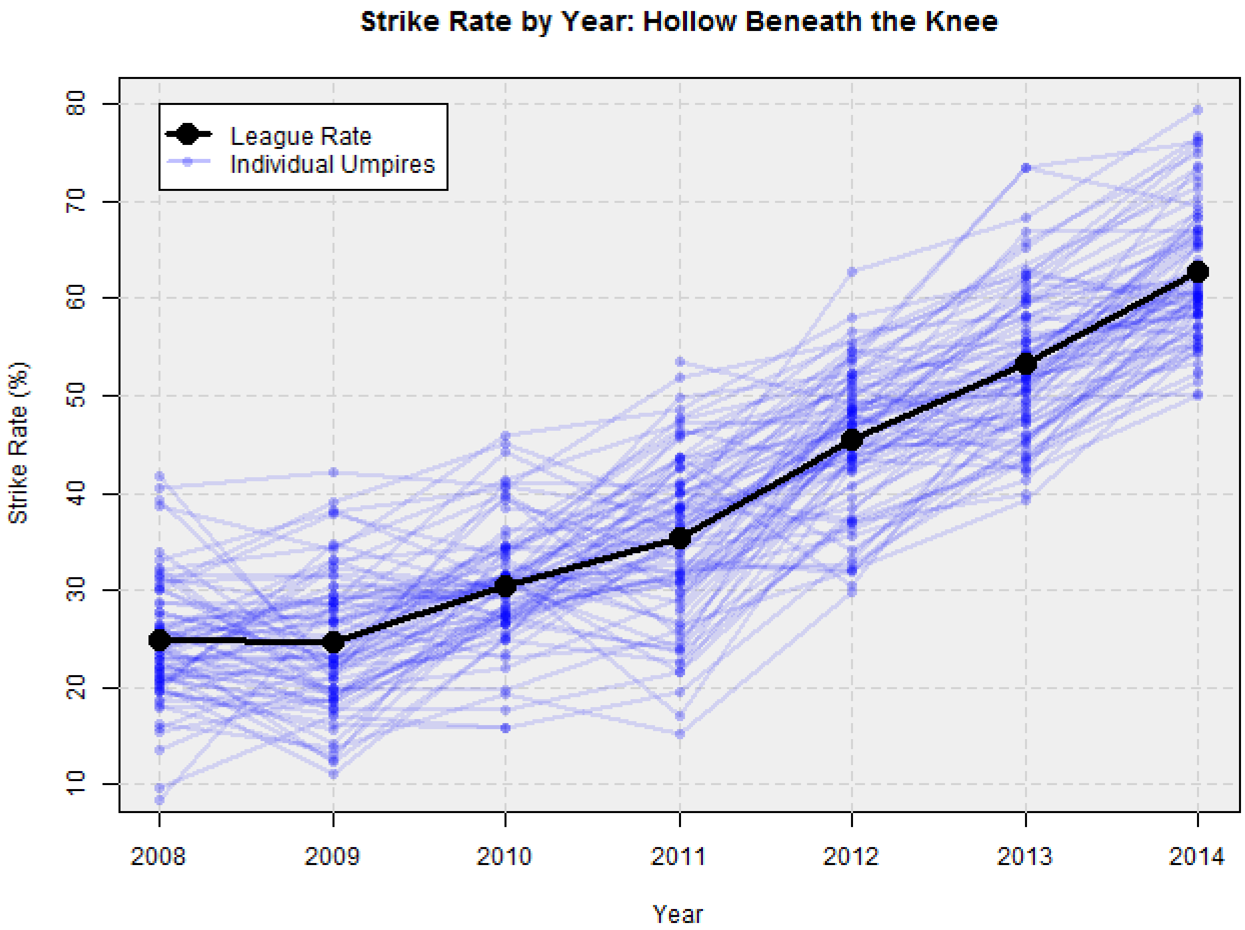

FIGURE 3 – Visualization of Strike Zone Contour

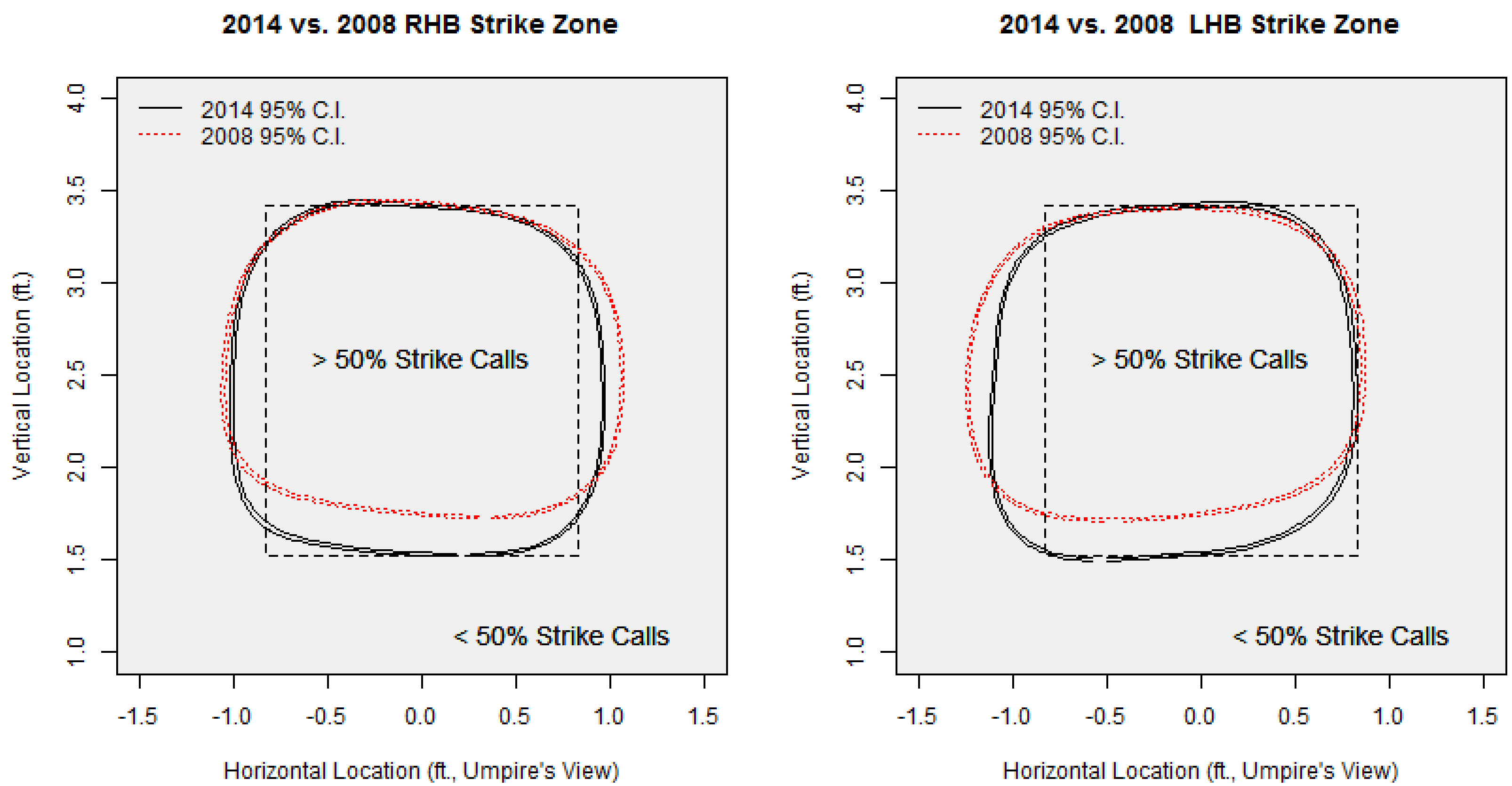

TABLE 2 – Strike Zone Surface Area Change (Square Inches)

Year	50% Probability Boundary		90% Probability Boundary	
	RHB	**LHB**	**RHB**	**LHB**
2008	444.8 *(437.2, 452.5)*	438.0 *(429.3, 446.8)*	236.8 *(228.6, 245.4)*	235.6 *(226.3, 245.4)*
2009	438.5 *(430.7, 446.3)*	436.9 *(428.6, 445.4)*	232.6 *(224.8, 240.9)*	233.6 *(224.7, 242.9)*
2010	443.4 *(434.8, 452.0)*	434.8 *(425.0, 444.7)*	239.1 *(231.3, 247.5)*	233.5 *(224.8, 242.8)*
2011	452.4 *(444.2, 460.7)*	447.1 *(437.5, 456.9)*	245.5 *(237.4, 253.9)*	246.8 *(237.6, 256.4)*
2012	462.2 *(453.6, 470.9)*	453.0 *(443.2, 463.0)*	262.5 *(254.0, 271.5)*	256.7 *(247.5, 266.4)*
2013	467.9 *(459.4, 476.5)*	458.3 *(449.0, 467.7)*	274.3 *(265.8, 283.2)*	268.5 *(247.5, 266.4)*
2014	478.4 *(470.3, 486.6)*	468.0 *(458.7, 477.4)*	282.9 *(274.3, 291.9)*	277.3 *(267.7, 287.5)*
Sq. In. Change	**33.6**	**30.0**	**46.1**	**41.7**
% Change	**+ 7.55**	**+ 6.85**	**+ 19.46**	**+ 17.72**

Notes: 95% confidence intervals in parentheses.

TABLE 3 – Pitcher Behavioral Changes

Year	Pitches	Within Zone (%)	Low (%)	Low Zone (%)	Low Zone (% of Within)	Pitch Height (in.)
2008	685,129	46.703	22.130	4.390	9.399	28.864
2009	708,859	46.933	22.273	4.421	9.420	28.860
2010	705,758	47.168	23.197	4.596	9.745	28.509
2011	705,959	47.017	23.737	4.609	9.803	28.410
2012	693,646	46.841	25.519	4.940	10.547	27.796
2013	709,163	46.655	26.273	4.991	10.698	27.618
2014	700,959	46.652	27.575	5.187	11.119	27.262
% Change	**----------**	**- 0.11**	**+ 24.6**	**+ 18.2**	**+ 18.3**	**- 5.55**
χ^2	----------	*0.3611*	*5494.6****	*482.74****	*518.95****	----------
t	----------	----------	----------	----------	----------	*85.518****

Note: % Change and test statistics refer to comparison of 2008 season to 2014 season only.

TABLE 4 – Batter Behavioral Changes

Year	Swing (%)	Swing Within (%)	Swing Outside (%)	Swing Low (%)
2008	45.469	64.650	28.662	31.207
2009	44.949	63.282	28.735	31.052
2010	45.078	63.203	28.896	31.638
2011	45.712	64.080	29.412	32.610
2012	45.749	64.159	29.527	33.329
2013	46.054	64.850	29.615	33.930
2014	46.381	65.377	29.769	34.650
% Change	**+ 2.00**	**+ 1.12**	**+ 3.86**	**+ 11.03**
χ^2	*115.58******	*37.501******	*109.114******	*454.05******

Note: % Change and test statistics refer to comparison of 2008 season to 2014 season only.

TABLE 5 – Pitch Height Impacts on Contact, In-Play, and Hit Rates on Batter Swings

	Contact		***InPlay***		***Hit***	
Intercept	1.27712*** *(0.00779)*	0.75639*** *(0.00745)*	-0.18134*** *(0.00634)*	-0.36135*** *(0.00623)*	-1.02781*** *(0.00738)*	-1.11179*** *(0.00730)*
2009	-0.00776 *(0.00671)*	-0.00916 *(0.00657)*	-0.00410 *(0.00516)*	-0.00474 *(0.00513)*	-0.00720 *(0.00602)*	-0.00750 *(0.00601)*
2010	-0.01572** *(0.00669)*	-0.01665** *(0.00654)*	-0.00348 *(0.00516)*	-0.00718 *(0.00513)*	-0.02755*** *(0.00604)*	-0.02884*** *(0.00604)*
2011	0.00923 *(0.00668)*	0.00442 *(0.00652)*	-0.00663 *(0.00514)*	-0.01323*** *(0.00512)*	-0.03496*** *(0.00603)*	-0.03739*** *(0.00603)*
2012	-0.02978*** *(0.00666)*	-0.03549*** *(0.00650)*	-0.02286*** *(0.00518)*	-0.03695*** *(0.00515)*	-0.03081*** *(0.00606)*	-0.03615*** *(0.00606)*
2013	-0.02637*** *(0.00662)*	-0.03631*** *(0.00646)*	-0.03058*** *(0.00515)*	-0.04960*** *(0.00512)*	-0.03524*** *(0.00603)*	-0.04258*** *(0.00603)*
2014	-0.01581** *(0.00664)*	-0.03030*** *(0.00647)*	-0.03888*** *(0.00517)*	-0.06373*** *(0.00514)*	-0.04534*** *(0.00606)*	-0.05499*** *(0.00606)*
LowPitch	-1.31648*** *(0.00400)*	---------- ----------	-0.65756*** *(0.00410)*	---------- ----------	-0.29856*** *(0.00483)*	---------- ----------
InBelowCenter	---------- ----------	-0.04582*** *(0.00027)*	---------- ----------	-0.00362*** *(0.00017)*	---------- ----------	-0.00278*** *(0.00020)*

TABLE 6 – Accuracy Impact on Relevant Statistical Measures

Measure	**ERA**	**ERA**	**K/9**	**K/9**	**BB/9**	**BB/9**
Umpire Effects	*F.E.*	*B.E.*	*F.E.*	*B.E.*	*F.E.*	*B.E.*
Intercept	4.00592	5.97173	11.63678***	11.58192***	2.57238	0.96639
	(2.44383)	*(4.09241)*	*(1.84577)*	*(2.77863)*	*(1.88035)*	*(2.00050)*
2009	-0.02031	-1.81767***	0.18180***	-0.21630	0.05015	-1.26265***
	(0.07742)	*(0.57880)*	*(0.06216)*	*(0.39299)*	*(0.05081)*	*(0.28294)*
2010	-0.23128***	-0.68910	0.25571***	0.78685	-0.06033	-0.45249
	(0.07175)	*(0.70883)*	*(0.05823)*	*(0.48127)*	*(0.05160)*	*(0.34650)*
2011	-0.30410***	-0.45064	0.23548***	0.55826	-0.14903**	-1.19107***
	(0.06802)	*(0.63582)*	*(0.05848)*	*(0.43171)*	*(0.06039)*	*(0.31081)*
2012	-0.16698*	-0.95411	0.56636***	-0.02602	-0.09314	-0.75579**
	(0.09076)	*(0.69589)*	*(0.07378)*	*(0.47249)*	*(0.08864)*	*(0.34017)*
2013	-0.23187*	-0.96794*	0.55529***	0.57213	-0.02450	-0.63877**
	(0.12095)	*(0.52400)*	*(0.09304)*	*(0.35578)*	*(0.10836)*	*(0.25615)*
2014	-0.25511*	-0.93462**	0.62336***	0.81570***	-0.07962	-0.70495***
	(0.13596)	*(0.42464)*	*(0.10392)*	*(0.28832)*	*(0.12673)*	*(0.20758)*
CorrectBallRate	0.04827**	0.03368	-0.09368***	-0.08519***	0.06314***	0.08089***
	(0.02372)	*(0.03708)*	*(0.01817)*	*(0.02518)*	*(0.01583)*	*(0.01813)*
CorrectStrikeRate	-0.05005***	-0.05005**	0.04439***	0.03566**	-0.06068***	-0.05222***
	(0.01630)	*(0.02155)*	*(0.01014)*	*(0.01463)*	*(0.01344)*	*(0.01054)*
R^2	0.226	0.212	0.474	0.580	0.391	0.546
Obs.	569	569	569	569	569	569

Notes: Minimum 500 pitches called by umpire during season. Standard errors clustered by umpire.

TABLE 7 – Changes in Measures Attributed to Umpires (2008 – 2014)

Year	*ERA*	*K/9*	*BB/9*
2008	4.32	6.83	3.39
2009	4.32	6.99	3.46
2010	4.08	7.13	3.28
2011	3.94	7.13	3.11
2012	4.01	7.56	3.05
2013	3.87	7.57	3.02
2014	3.74	7.73	2.89
Δ from 2008 to 2013	**- 0.58**	**0.90**	**- 0.50**
Δ from Umpire Strikes			
F.E.	- 0.42	0.37	- 0.51
B.E.	- 0.42	0.30	- 0.44
Δ from Umpire Balls			
F.E.	0.07	-0.13	0.09
B.E.	0.05	-0.12	0.11
Net Δ from Umpire Accuracy			
F.E.	- 0.35	0.24	- 0.42
B.E.	- 0.37	0.18	- 0.33
% Attributable to Δ in Accuracy			
F.E.	**61.0**	**26.9**	**84.5**
B.E.	**64.5**	**20.1**	**65.2**

Based on FE models from Table 6. Strike accuracy increased by 8.419 percentage points, while ball accuracy increased by 1.402 percentage points.