12-2013

# CHIN MUSIC: A STATISTICAL ANALYSIS OF RETALIATION PITCHES IN MAJOR LEAGUE BASEBALL

Peter Jurewicz
*Clemson University*, pjurewi@clemson.edu

# CHIN MUSIC: A STATISTICAL ANALYSIS OF RETALIATION PITCHES IN MAJOR LEAGUE BASEBALL

---

A Thesis
Presented to
the Graduate School of
Clemson University

---

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
Economics

---

By
Peter Jurewicz
December 2013

---

Accepted by:
Dr. Raymond Sauer, Committee Chair
Dr. Scott Baier
Dr. Robert Tollison

## ABSTRACT

This paper is focused on hit batsmen in Major League Baseball from the 2008 season through August 20th of the 2013 season. More specifically, this paper examines the characteristics of retaliation pitches and attempts to determine the intent of the pitcher. The paper also takes into account moral hazard and cost-benefit analysis of hitting an opposing batsman. There has been a vast amount of literature in economics with regard to hit batsmen in Major League Baseball. However, very few of these papers have been able to evaluate economic theories in Major League Baseball using Pitchf/x data. Pitchf/x technology became fully implemented into all thirty Major League ballparks prior to the 2008 season. Pitchf/x provides us with intricate details of every pitch, which include velocity, movement, and pitch location. Using Pitchf/x data, this paper presents a very detailed statistical analysis with regard to hit batsmen under a variety of scenarios. A probit model is used to test for the probability of an intentional retaliation pitch based on a variety of predictor variables. The most important conclusion we are able to draw from the two regression models presented in this paper is that, holding all other factors constant, retaliation pitches that are fastballs increase the probability of being intentional by roughly 17%.

# TABLE OF CONTENTS

LIST OF TABLES

## LIST OF FIGURES

## INTRODUCTION

The term "chin music" has been a slang term in baseball that dates back to the late 1800s. In the late 1800s "chin music" was used to describe heckling fans in the stadium. However, sometime during WWII, it began being used as a synonym for a pitch intentionally thrown at an opposing batsman[1]. The pitcher's goal when throwing "chin music" is to intentionally send a message and strike fear into the opposing batsmen. Baseball is unique in the fact that it has a set of unwritten rules with regard to hit batsmen. Generally, teams perceive the attitude that if you hit one of our guys, then we are going to hit one of your guys. Intentionally throwing at an opposing batter can be very dangerous. As a matter of fact, there has been one Major League player that was killed as a result of being hit by a pitch[2]. In August of 1920, Cleveland Indians shortstop Ray Chapman was a hit by a pitch and died twelve hours later. Shortly after Chapman's unfortunate death, Major League Baseball implemented rules that outlawed intentionally throwing at a batter.

Even though that it is illegal for pitchers to intentionally throw at opposing batsmen, it can be quite difficult for umpires to detect. There are many times, the pitcher exhibits poor control and accidentally hits an opposing batter. However, I believe that there are also a significant number of instances where this is not the case (i.e. Major League Baseball found it necessary to establish an official rule to outlaw intentional retaliation pitches). Umpires can attempt to determine whether a pitcher is intentionally throwing at an opposing batter based on a certain set of conditions. These conditions may include the score of the game, the inning, the type of pitch, the speed of the pitch, the quality of the teams, and whether a player has already

[1] http://www.annarbor.com/lifestyles/love-to-heckle-toast-or-should-we-stop-the-jeering/
[2] http://www.nytimes.com/packages/html/sports/year_in_sports/08.17.html

been hit during the course of a game. The final condition is perhaps the most important and will be referred to as a retaliation pitch in this paper.

Another reason umpires have a difficult time detecting for retaliation pitches is because some guys are more prone to be hit by a pitch during the course of the game. Often times, these players have a natural inclination to lean towards the plate in an attempt to get hit by the pitch and take first base. There was a famous instance of a case like this that occurred just this past season. In a game between the Los Angeles Dodgers and San Diego Padres on April 11th, 2013, Zack Greinke of the Dodgers hit Carlos Quentin of the Padres. Benches cleared, punches were thrown, and several players and coaches were ejected, fined, and suspended. However, there was a great debate because Carlos Quentin happens to be one of those players who is relatively well known for "leaning into" pitches and taking first base. Dave Cameron and Jeff Sullivan of Fangraphs performed a brief investigation of the pitch that hit Carlos Quentin during that April 11th matchup[3]. Cameron and Zimmerman used Pitchf/x data to examine all the pitches that had hit Quentin since the 2008 season. During this span Quentin was hit by 95 pitches, and remarkably, four of these pitches were actually in the strike zone. According to Major League rules, if a pitch is in the strike but hits the batsmen, the umpire is supposed to call that pitch a strike. However, this can be quite deceiving to the naked eye, so umpires will rarely call a pitch that hits a batter a strike. Cameron and Zimmerman then observed data on how many pitches Quentin had been hit by that were between 1.0 and 1.5 horizontal feet from the center of the plate compared to all other Major League players. Most would agree that this range would be called a ball but would likely not hit a batter. Moreover, the data indicates that 0.2% of pitches thrown in this range would result in a hit batsman. For Quentin, this number is equal to 0.4%. The pitch

[3] http://www.fangraphs.com/blogs/carlos-quentins-hbp-zone/

thrown by Greinke on April 11th was 1.504 horizontal feet from the center of the plate. Certainly this pitch would be considered a ball, but the overwhelming majority of batters would not be hit by the pitch.

The purpose of the Grienke/Quentin example is to illustrate how difficult it can be to determine the intent of the pitcher. It is impossible to truly understand Greinke's intentions. In addition, most pitchers will lie to avoid fines and suspensions. Quentin certainly must have believed Greinke was trying to hit him, since he charged the mound and broke Greinke's collarbone. However, we must rely on the data to form our conclusions, and in this case the data indicates Greinke did not intentionally throw at Carlos Quentin.

Essentially, the purpose of my research is to determine important characteristics to help fans and umpires detect for intentional retaliation pitches. In addition to examining these characteristics, I am also interested in testing for the presence of moral hazard. Testing for the intent of the pitcher can be quite difficult, but I will use Pitchf/x data to help try and determine this. I will assume that a pitch that is farther inside is more likely to be a pitch that is intentionally thrown to hit an opposing batter. I will also assume that an intentional pitch is a pitch with little movement (i.e. a fastball) with slightly lower velocity than the average. However, it is nearly impossible to test for a pitch that that is lower in velocity because every pitcher has a different average fastball. Moreover, Aroldis Chapman often throws his fastball over 100 mph, whereas Greg Maddux rarely threw his fastball over 90 mph.

## LITERATURE REVIEW

There has been extensive economic literature with regard to hit batsmen in Major League Baseball. Many economists have disagreed whether retaliation pitches are directed at the opposing pitchers or one of the pitcher's teammates. This is not the focus of my research, but I will certainly touch on this debate. The overwhelming majority of economic papers that focus on hit batsmen approach the topic from a moral hazard prospective or perform some type of cost-benefit analysis.

*Batter up! Moral Hazard and the Effects of the Designated Hitter Rule on Hit Batsmen* by Goff, Shughart, and Tollison take a moral hazard[4] approach when examining hit batsmen. The authors of this paper propose the idea that the introduction of the designated hitter rule in 1973 created the potential for a classic moral hazard problem. Moreover, American League pitchers are not required to appear at the plate; therefore, American League pitchers can throw at opposing hitters with lower costs than National League pitchers, who must take their turn at bat. The authors hypothesize that there will be in an increase in the number of hit batsmen in the American League compared to the National League as a result of the introduction of the designated hitter rule.

To test their hypothesis, Goff, Shughart, and Tollison use the following basic regression model:

$$\mathbf{HBal - HBnl = \beta0 + \beta1DH + \beta2(Zal - Znl) + \varepsilon}$$

[4] Moral hazard is a situation where a party will be more inclined to take risks because that party taking the risks will not have to incur the costs of the risks. For example, people are more likely to drive recklessly if they are forced to wear a seatbelt.

The dependent variable they use is the difference between hit batsmen in the AL compared to the number of hit batsmen in the NL. The variable "DH" is a dummy variable, which is equal to zero before 1973 and equal to 1 from 1973 through 1990 (DH rule introduced in MLB in 1973). The variables "Zal" and "Znl" represent a vector of explanatory variables that control for a variety of factors. Like the data for hit batsmen, these control variables are also a set of yearly data (i.e. total number for a given season). The control variables represented by the variable "Z" that the authors used include: at-bats, slugging average, home runs, bases on balls, strikeouts, saves, std. deviation of winning percentage, and attendance. These variables account to control for the following factors, which include: pitcher control/ability, hitter ability, degree of competiveness of games, the amount of reliance on relief pitching, and the financial rewards of winning. Their results indicate American League batters have been hit by pitches 10% to 15% more their National League counterparts in the seasons that followed the introduction of the designated hitter rule. This result would imply that moral hazard does exist with regard to hit batsmen in Major League Baseball. For the purpose of my research, I will assume that moral hazard exists if American League pitchers throw farther inside than National League pitchers.

Trandel, White, and Klein have a similar paper but disagree with Goff, Shughart, and Tollison's moral hazard conclusion. Trandel, White, and Klein argue that pitchers are generally poor batters; therefore, pitchers are not going to hit an opposing pitcher to put him on base. Because the National League has weaker hitters than the American League, this explains why fewer batters are hit in the National League. The authors argue that rather than hitting an opposing pitcher to retaliate, teams are more likely to hit a star hitter on the opposing team. They propose a cost-benefit analysis and explain that hitting an opposing batter is costly; therefore, teams are far more likely to put a good hitter on base rather than a poor hitter.

# DATA

The data used in this paper was retrieved primarily from the Pitchf/x database and Retrosheet database. Pitchf/x is a detailed pitch tracking system, created by Sportvision, and is fully implemented into all Major League stadiums since the 2008 season. Pitchf/x uses technology to account for the movement, location, and velocity of every pitch in Major League Baseball. By accounting for these three characteristics, Pitchf/x is able to tell us what type of pitch was thrown. Pitchf/x also provides us with the result of each pitch and each plate appearance. I retrieved this data from baseballsavant.com, which allows sorting the data by the result of the pitch. For the purpose of this paper, I am primarily interested in pitches that resulted in a hit batsman. My data set consists of 8,948 hit batsmen from the beginning of 2008 season through August 20 of the 2013 season (the 2013 season was still in progress during my research). Only regular season games are included in my data set. So far as identifying pitches is concerned, the Pitchf/x system is extremely accurate; however, the system does malfunction at times. Approximately 0.2% of pitches cannot be identified. These pitches were obviously dropped from my sample, since there is no data on these pitches. Pitchf/x classifies the remaining pitches into 10 different types of pitches. These pitch types include fastball, four-seam fastball, two-seam fastball, cutter, sinker, slider, curveball, changeup, knuckle-curve, and knuckleball. For the purpose of my research, I combined the four types of fastballs into one fastball category[5]. I also considered including "sinker" in the fastball category but elected not to due to the increased vertical movement of a sinker. I understand that two-seam fastball and cutter have some horizontal movement, but it is generally pretty small, especially when compared to a slider.

[5] Composed of fastball, four-seam fastball, two-seam fastball, & cutter.

The retrosheet data allows users to download statistics from every box score of every Major League game. Retrosheet does not provide the detailed pitch information like Pitchf/x; however, it does provide us with everything we may find in a typical box score. The retrosheet data is quite extensive and goes back several decades. With the combination of these two datasets, the possibilities for extensive research are quite captivating.

One key component of understanding the data and empirical tests in this is study is understanding what I define to be a retaliation pitch. I was influenced on how to design the retaliation variable based on a paper titled *Reversal of fortune: a statistical analysis of penalty calls in the National Hockey League* by Abrevaya and McCulloch. In the paper the authors examine whether a penalty is more likely to occur on one team if the previous penalty was called against the opposing team. They describe the term *reversal call* to be a penalty call on a team that was not the last team penalized.

This paper defines a retaliation pitch in a similar manner Abrevaya and McCulloch defined reversal calls. A retaliation pitch is a pitch that hits a batter assuming that a batter on the opposing team was hit by a pitch previously in the game. Moreover, let's assume Team A is playing against Team B. Let's also suppose that the hit by pitch sequence during the course of the game is as follows: ABAB. This would imply that there were three retaliation pitches during the course of the game. Another possible hit by pitch sequence might be ABBA. This would also imply three retaliation pitches. However, the hit by pitch sequence AAAB would imply only one retaliation pitch. There is no set rule for the number of retaliation pitches that may occur during the course of game. It is possible that no retaliation pitches occur during a game and also possible that the number of retaliation pitches that occur is very high.

## RELEVANT SUMMARY STATISTICS AND EXPLANATION OF KEY FIGURES/TABLES

Table 1, which can be seen below, shows the average horizontal distance (in feet) of the pitch at the moment when the baseball crosses the plate (or in the case of this paper hits a batter). The Pitchf/x system measures both the vertical and horizontal distance of every pitch at the point at which the ball crosses the plate. The "feet inside" shown in Table 1 is the horizontal distance from the center of the plate. The default of the Pitchf/x system is to act as traditional x and y plot. Furthermore, a pitch to the left of the plate (towards a left handed batter) is negative and a pitch to the right of the plate (towards a right handed batter) is positive. For the purpose of my research, I am primarily interested in analyzing the absolute values of these distances. Figure 1 is provided for a better understanding of the horizontal distance of pitches.

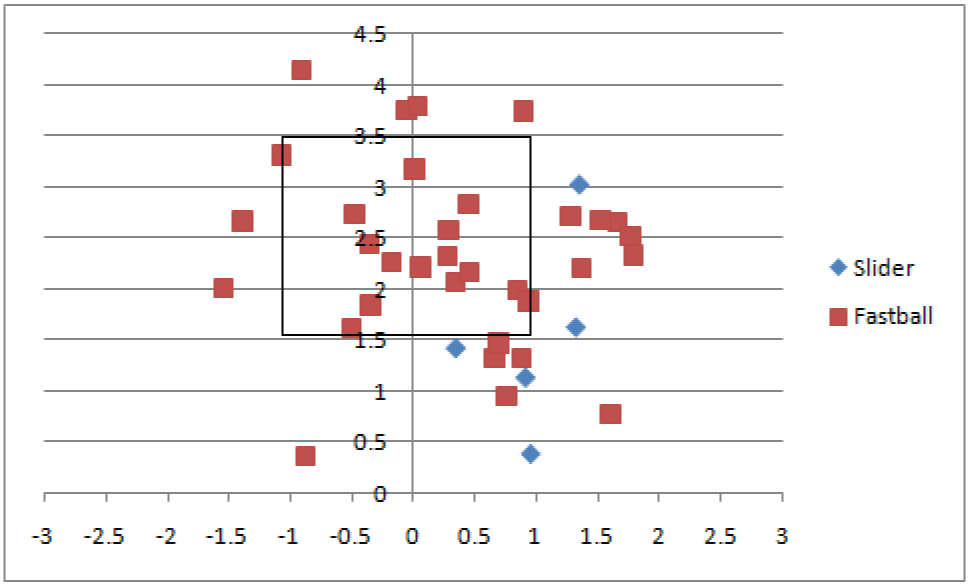


**Figure 1: Pitchf/x strike zone plot**[6]

[6] For a further explanation of Figure 1: http://www.beyondtheboxscore.com/2009/4/17/841366/understanding-pitch-f-x-graphs

| Condition | Average Distance Inside (Feet) | Average Distance Inside (Inches) |
|---|---|---|
| All HBP | 2.065 | 24.780 |
| HBP Top of the 1st | 2.036 | 24.432 |
| HBP Top of the 1st (Fastballs) | 2.018 | 24.216 |
| Retaliation Pitch | 2.101 | 25.212 |
| Retaliation Pitch (Fastballs) | 2.114 | 25.368 |
| AL HBP | 2.074 | 24.888 |
| AL Retaliation HBP | 2.114 | 25.368 |
| AL Retaliation HBP (Fastballs) | 2.113 | 25.356 |
| NL HBP | 2.057 | 24.684 |
| NL Retaliation HBP | 2.090 | 25.080 |
| NL Retaliation HBP (Fastballs) | 2.114 | 25.368 |

**Table 1: Average horizontal distance of HBP**

Table 1 does not tell us whether these results are statistically significant or not; however, we are able to draw few important preliminary results from these estimations. First of all, there were 8,948 hit batsmen in this dataset. I used the top of the first inning as a baseline to show average of feet of a batter hit in the top of the first inning. It can be expected that no pitch in the top of the first inning will be a retaliation pitch. The table shows that retaliation pitches that are only fastballs are farther inside than the average retaliation pitch. This is interesting because fastballs are usually the pitch that pitchers have the best control over. Interestingly, Table 1 also supports Goff Shughart and Tollison's moral hazard theory to a certain extent. Both retaliation pitches and all hit by pitches in the American League are farther inside than in the National League; however, the results show that the average feet inside for a fastball retaliation pitch is nearly identical for both leagues.

| All HBP | | Retaliation HBP | | Non-Retaliation HBP | |
|---|---|---|---|---|---|
| **Pitch Type** | **Percentage** | **Pitch Type** | **Percentage** | **Pitch Type** | **Percentage** |
| Changeup | 4.29% | Changeup | 3.97% | Changeup | 4.34% |
| Curveball | 10.16% | Curveball | 9.75% | Curveball | 10.22% |
| Fastball | 56.15% | Fastball | 58.26% | Fastball | 55.82% |
| Knuckle-Curve | 0.45% | Knuckle-Curve | 0.25% | Knuckle-Curve | 0.48% |
| Knuckleball | 0.75% | Knuckleball | 0.83% | Knuckleball | 0.74% |
| Sinker | 14.08% | Sinker | 15.62% | Sinker | 13.84% |
| Slider | 14.13% | Slider | 11.32% | Slider | 14.56% |

**Table 2: Pitch type breakdown**

Table 2, shown above, provides us with a breakdown of the types of pitches thrown that hit opposing batsmen. This table presents us with the difference scenarios, which include all hit batsmen, hit batsmen hit by a retaliation pitch, and hit batsmen hit by a non-retaliation. Not surprising, the overwhelming majority of pitches thrown are fastballs. We also do not observe a great deal of variation among the three scenarios. Interestingly, there are more fastballs thrown for retaliation pitches, which would certainly support my hypothesis.

| Inning | Percent |
|---|---|
| 0 | 12.43% |
| 1 | 24.95% |
| 2 | 21.15% |
| 3 | 15.58% |
| 4 | 8.91% |
| 5 | 7.14% |
| 6 | 5.19% |
| 7 | 2.97% |
| 8 | 1.30% |

**Table 3: Interval of retaliation pitches**

Table 3, which can be seen above, represents the number of innings Team A takes to retaliate and hit a batter on Team B (assuming a player on Team A has already been hit). Similar to the retaliation pitch definition there is no limit on how many times this statistic can occur during the course of a game.  On average it takes a team approximately 2.517 innings to throw a retaliation pitch when their own player has already been hit by a pitch.  Table 3 and Figure 2 clearly illustrate that teams do not waste time when throwing retaliation pitches.  Moreover, approximately 74.12% of teams throw a retaliation pitch between 0 & 3 innings.

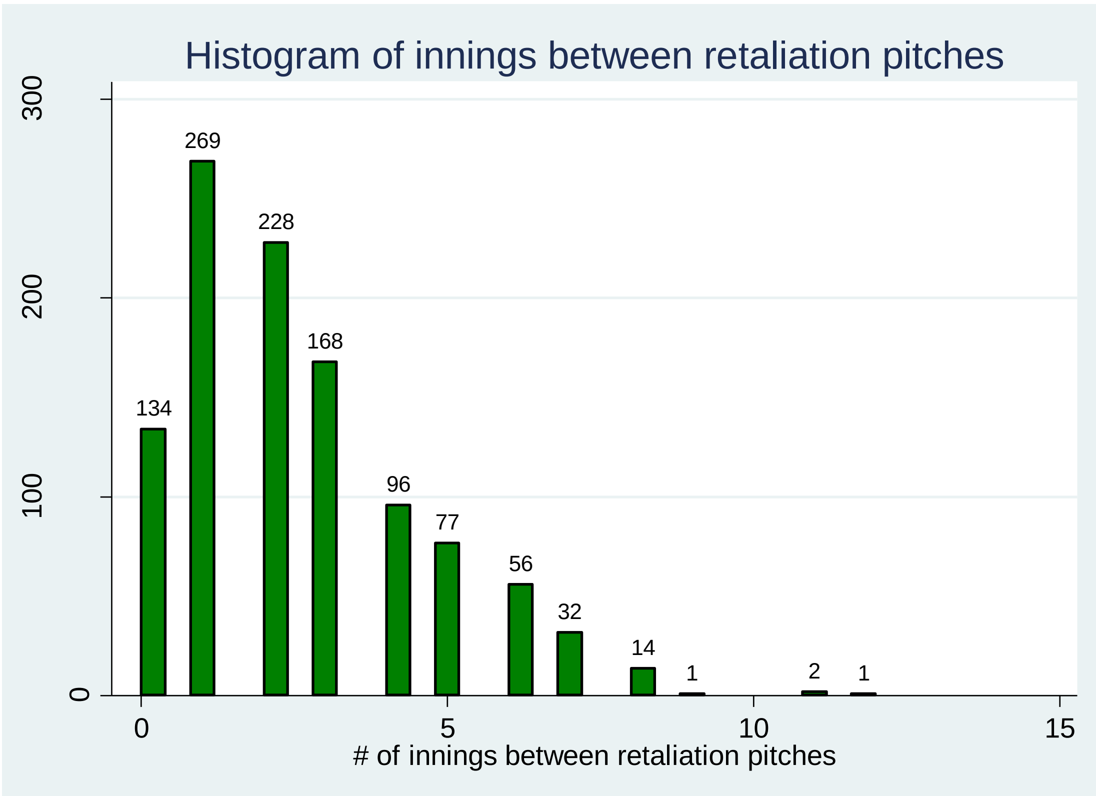


**Figure 2: Interval of innings between retaliation pitches**

The following two histograms (Figure 3 and Figure 4) illustrate the number of retaliation pitches based on the inning.  The histograms only include innings 1-8 because there are many instances where the home team will not take at-bats in the ninth inning due to the circumstances of the game.  However, as expected, the results show an increasing trend as the game progresses. This is rational because there are more pitches thrown as the game progresses; therefore, greater opportunity for retaliation pitches in the later innings.  The one exception is that the results show a clear decline in the number of retalition pitches in the fifth inning.  One possible explanation as to the decline we see in the fifth inning could be that starting pitchers place additional emphasis on the fifth inning in order to qualify for the win. As archaic as the win statistic is in Major League Baseball, there is no question that starting pitchers are motivated to earn a win for themselves. The overwhelming majority of front offices in Major League Baseball do not analyze pitcher win/loss record when evaluating players; however, many fans and writers consider pitcher win/loss record when voting for all-stars, Cy Young award, most valuable player award, etc. Essentially starting pitchers trying to qualify for the win during a game may choose not to throw an intentional retaliation pitch during the fifth inning in an effort to not put an opposing player on base.

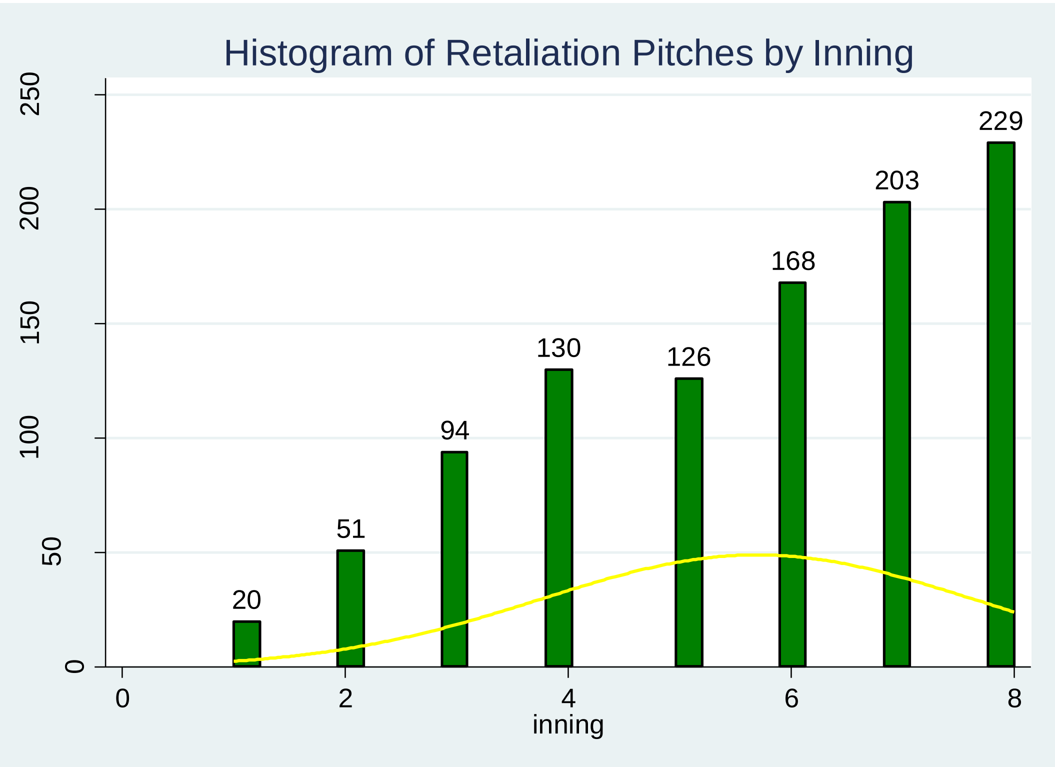


**Figure 3: Retaliation pitches by inning**

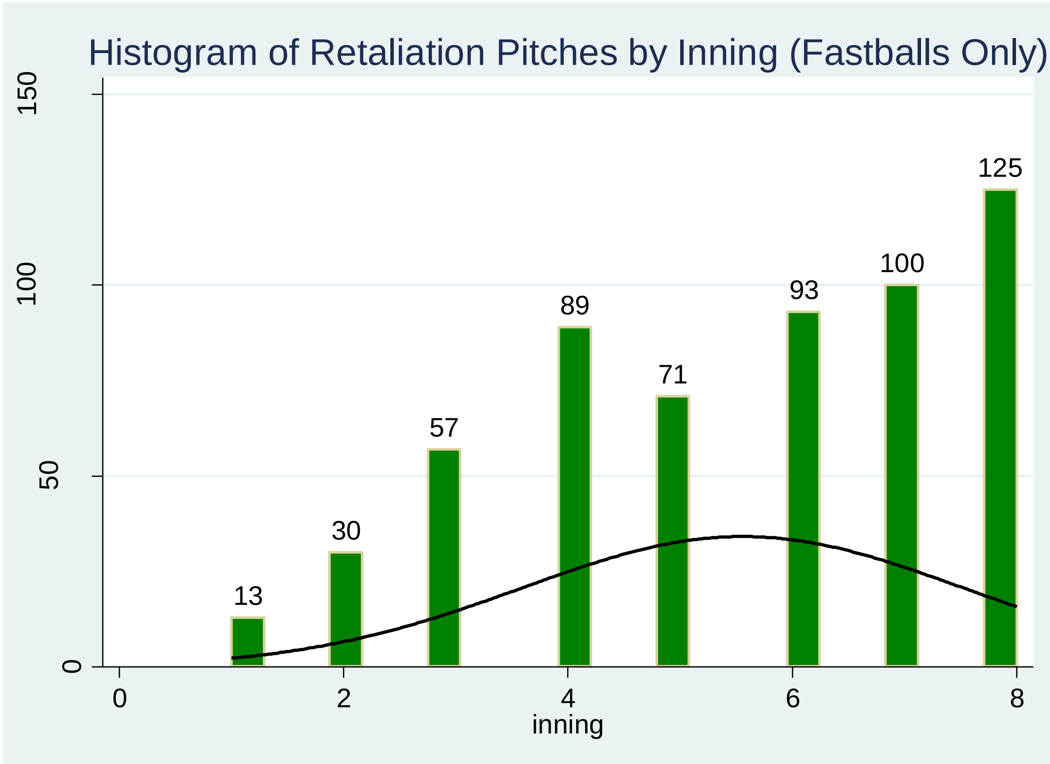


**Figure 4: Retaliation pitches by inning (fastballs only)**

The next set of figures and tables show how retalation pitches are related to the winning percentage of the team throwing the retaliation pitch. I chose to divide teams into three equal groups based on winning percentage. The first group is the top one-third of teams since 2008. These teams have a winning percentage greater than .540. The second group is the middle-third of teams. These teams have a winning percentage between .457 and .540. The final group is the bottom-third of Major League teams. These teams have a winning percentage less than .457. Interestingly, these results show a positive correlation between winning percentage and retaliation pitches. In other words, good teams are far more likely to throw a retalitation ptich than bad teams. This is interesting because one would believe that intentionally putting an additional man on base would decrese your chances of winning. While I believe there is no question that is certainly true, there also must be a rational explaination as to why winning teams throw more reataliation pitches.

| Winning Percentage | Total Retaliation Pitches |
|---|---|
| Top one-third | 375 |
| Middle one-third | 358 |
| Bottom one-third | 288 |
| Top one-third (Fastballs Only) | 209 |
| Middle one-third (Fastballs Only) | 201 |
| Bottom one-third (Fastballs Only) | 168 |

**Table 4: Retaliation pitches based on winning percentage**

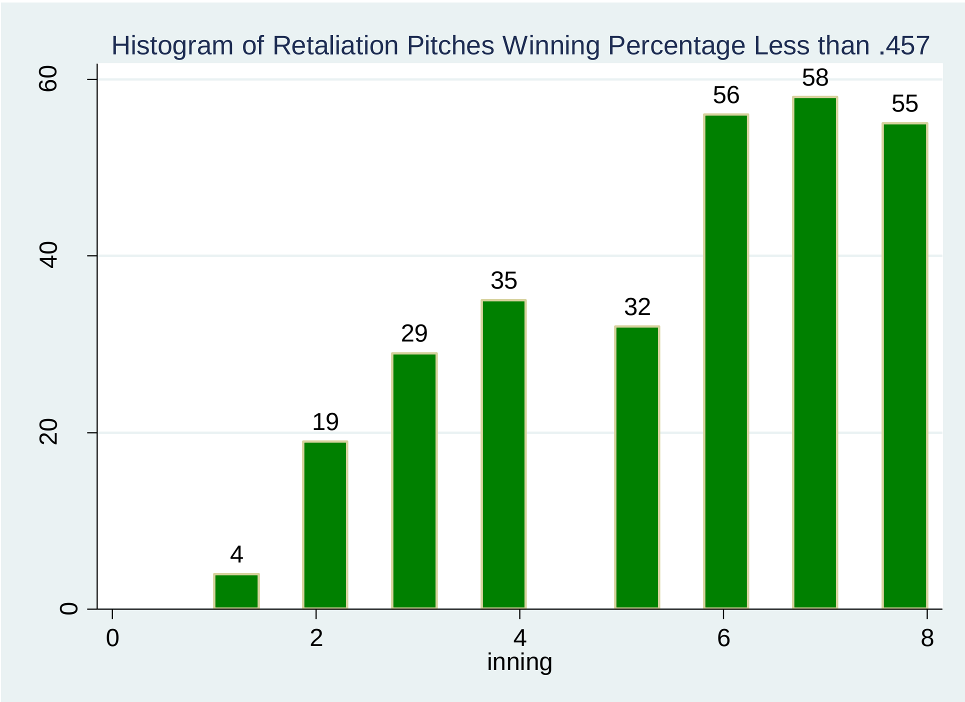


**Figure 5: Retaliation pitches by winning percentage <.457**

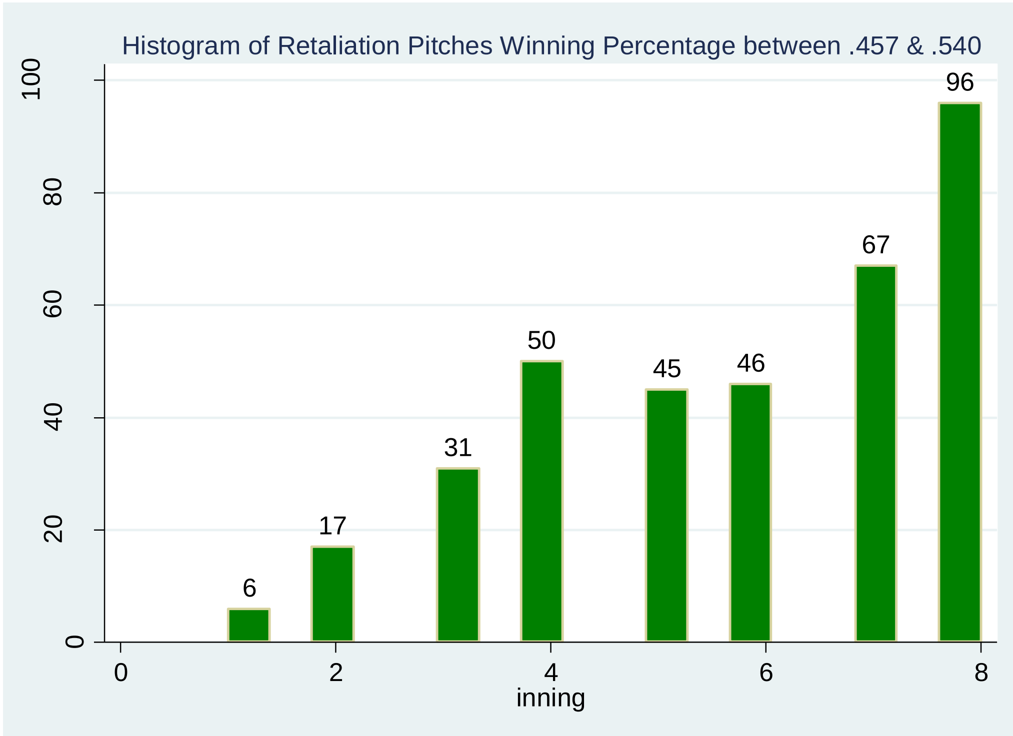


**Figure 6: Retaliation pitches by winning percentage between .457 & .540**

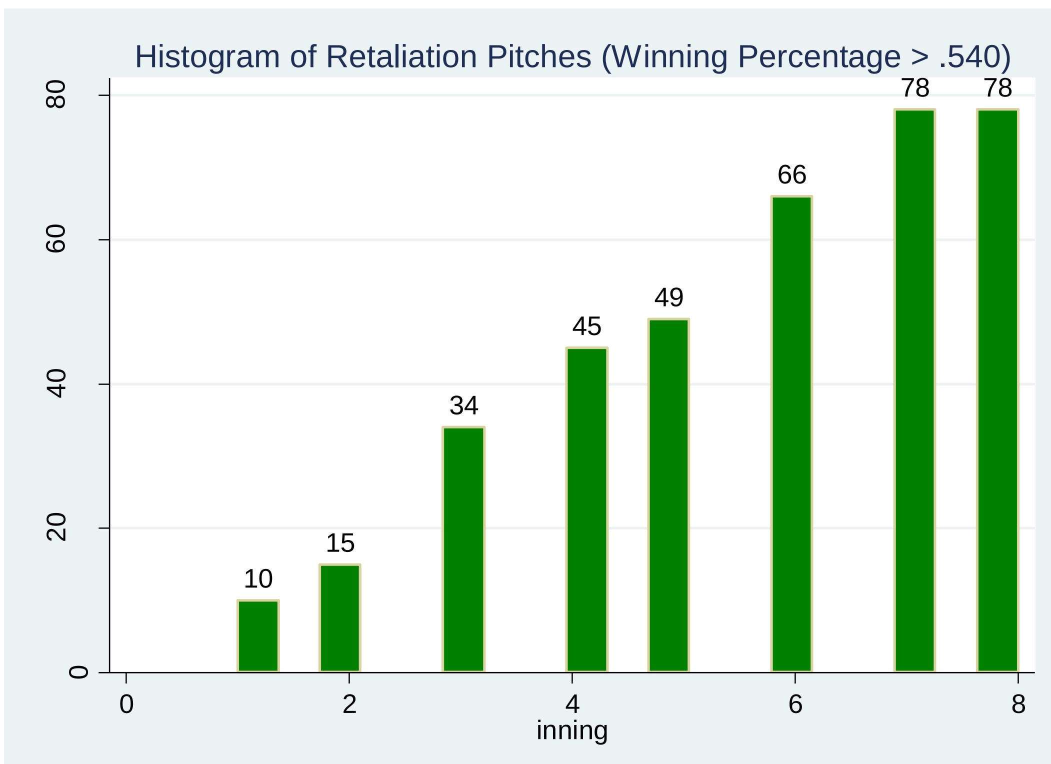


**Figure 7: Retaliation pitches by winning percentage > .540**

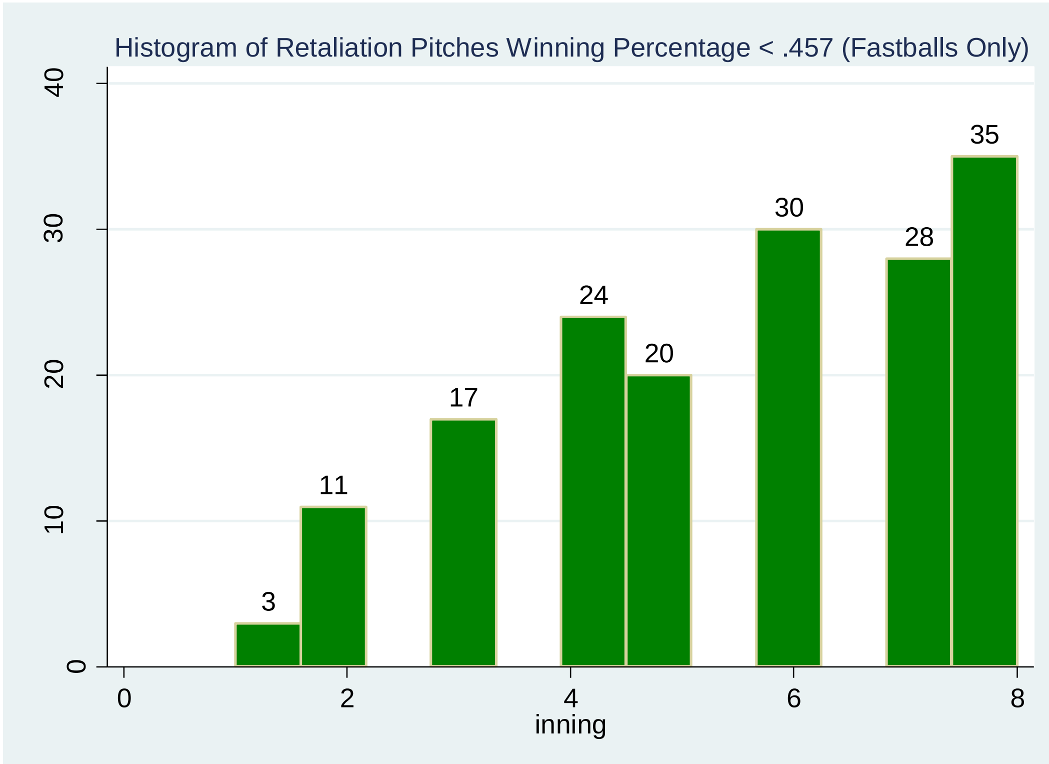


**Figure 8: Retaliation pitches by winning percentage < .457 (Fastballs)**

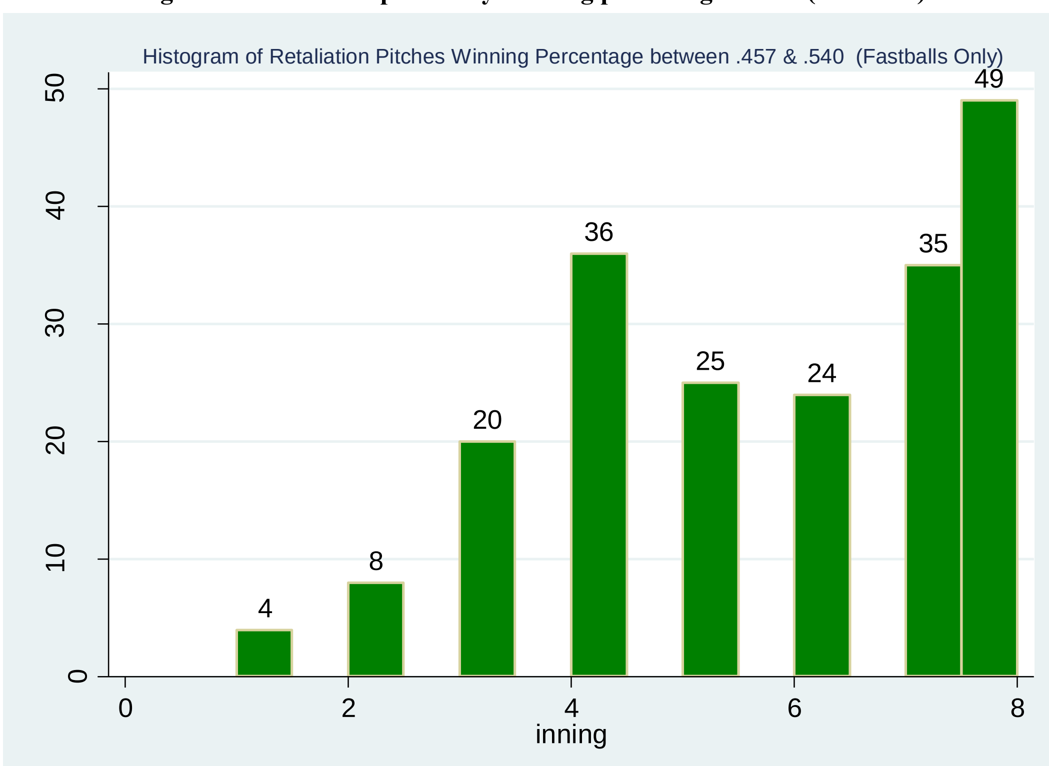


**Figure 9: Retaliation pitches by winning percentage between .457 & .540 (Fastballs)**

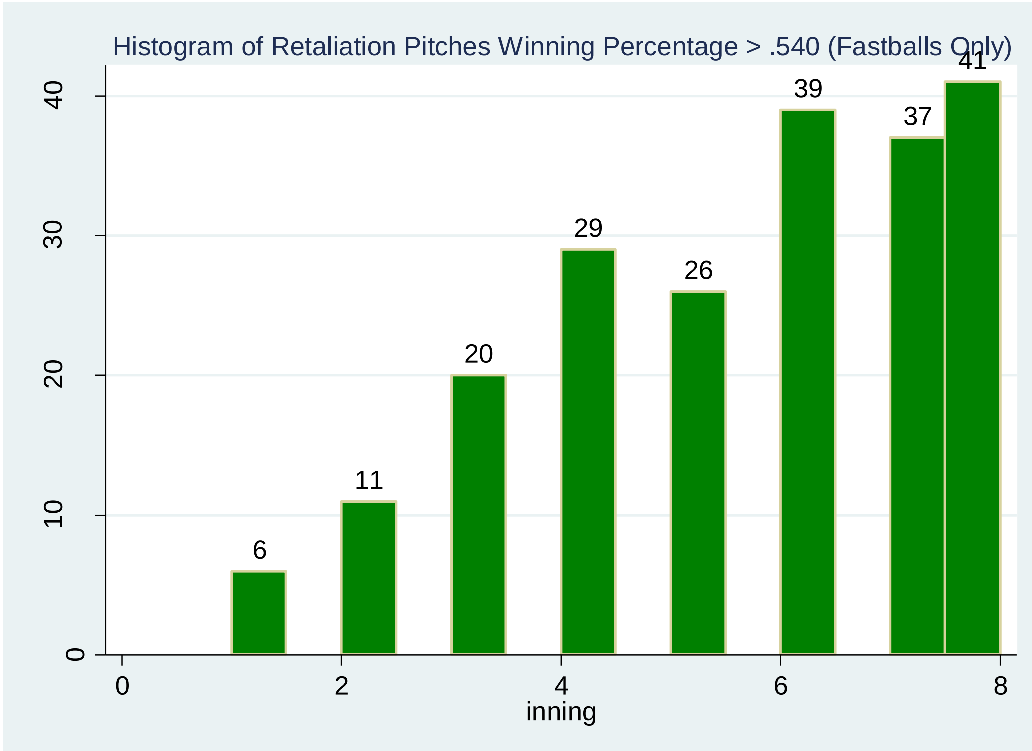


**Figure 10: Retaliation pitches by winning percentage > .540**

One possible explanation as to why better teams throw significantly more retaliation pitches than poor teams could be that the better teams are more invested in the outcome of the season. Moreover, I believe that if a star player (or any starting player) on a good team was hit by a pitch then the good team would be more likely to take offense than if a star player on a poor team was hit by a pitch. The winning teams are fully invested in the season and are strongly concerned with potential post-season implications; therefore, they are incentivized to retaliate in an effort to ensure their player(s) are not hit by a pitch by opposing teams in the future.

The next set of figures and tables illustrate the first hit batter during the course of a game based on the inning when a retaliation pitch occurs during the course of the game. Essentially these results represent the pitch that motivated the opposing pitcher to throw a retaliation pitch. As one would expect, the results indicate a downward trend. Obviously, it is more likely for the first hit batter to occur during the beginning of the game. This is the completely opposite effect that occurred for reataliation pitches, which can be seen in Figures 3 and 4 (i.e. retaliation pitches were more likely to occur as the game progressed). It is clear that when there is a hit batsmen in a game, this event occurs very early. Moreover, almost half (45.38%) of the first hit batsmen occurs in the first two innings.

| Inning of 1st HBP | Percent |
|---|---|
| 1 | 24.29% |
| 2 | 21.09% |
| 3 | 14.56% |
| 4 | 13.56% |
| 5 | 9.60% |
| 6 | 8.31% |
| 7 | 4.85% |
| 8 | 2.27% |

**Table 5: First HBP during course of game**

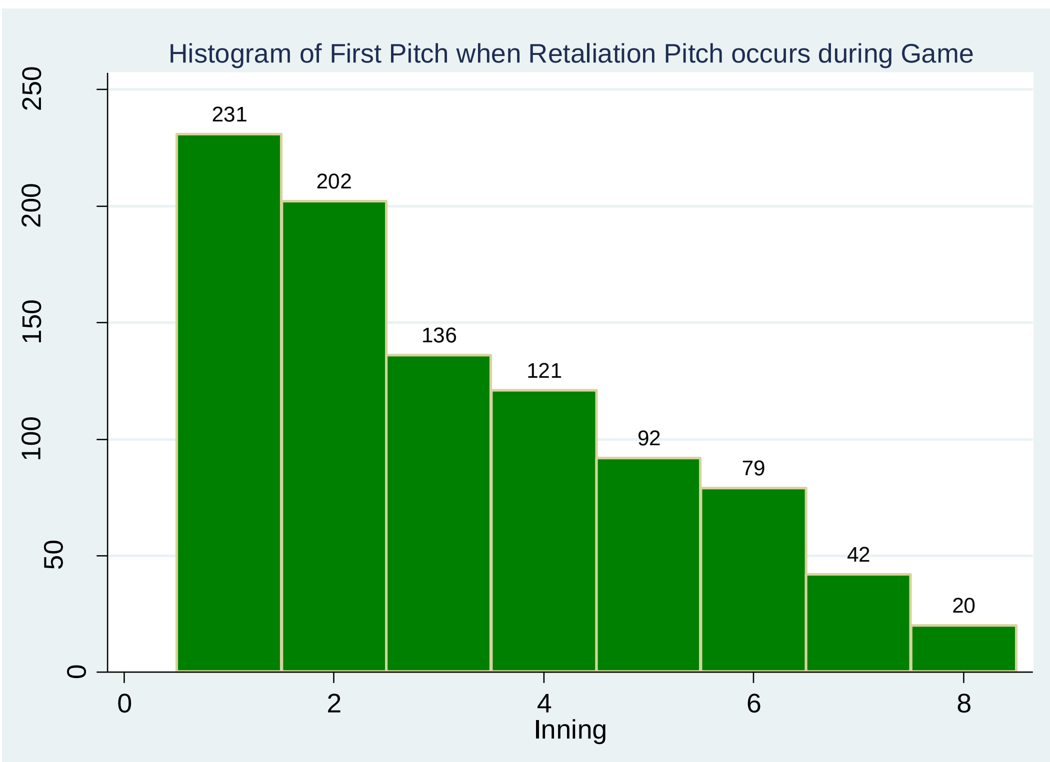


**Figure 11: Inning of first HBP**

Table 6, which can be seen below, essentially illustrates how often chaos ensues during the course of a baseball game. Chaos would imply that there is a high number of hit batsmen in a single game.

| HBP per Game | Percent |
|---|---|
| 1 | 51.17% |
| 2 | 33.48% |
| 3 | 10.97% |
| 4 | 3.59% |
| 5+ | 0.79% |

**Table 6: Frequency of number of HBP per game**

Table 6 only includes games where there is at least one hit batsmen. The results show that there are very few games with a high number of HBPs. Interestingly, over half of games with at least one hit batsman only have one HBP. This would imply that retaliation pitches might be somewhat of rare phenomenon. One detail not provided in the table, but I believe to be relevant is the number of hit batsmen per game (when there is at least one HBP during game). This number is equal to approximately 1.696 HBP per game.

Tables 7 and 8 break down retaliation pitches by team and league, respectively. These tables show us whether there is a team or league that throws an exceptionally high number of retaliation pitches. I find the results in table 7 exceptionally interesting. Perhaps, the biggest rivalry in all of sports is between the New York Yankees and Boston Red Sox. As one might expect, the Yankees and Red Sox throw the highest percent of retaliation pitches. There is a perception that the Yankees and Red Sox do not like each other very much, so it is interesting that these two teams lead the MLB in retaliation pitches.

| Team | Retaliation Pitches (%) | ALL HBP (%) |
|---|---|---|
| ARI | 4.05 | 3.69 |
| ATL | 1.82 | 3.11 |
| BAL | 2.56 | 3.64 |
| BOS | 5.54 | 4.26 |
| CHC | 3.39 | 3.84 |
| CIN | 4.05 | 3.67 |
| CLE | 4.96 | 3.54 |
| COL | 2.23 | 3.06 |
| CWS | 2.98 | 3.05 |
| DET | 3.72 | 3.34 |
| HOU | 2.64 | 3.15 |
| KC | 2.07 | 3.27 |
| LAA | 3.06 | 3.05 |
| LAD | 3.06 | 3.13 |
| MIA | 3.14 | 3.05 |
| MIL | 3.72 | 2.74 |
| MIN | 2.56 | 3.15 |
| NYM | 3.64 | 3.58 |
| NYY | 5.04 | 3.64 |
| OAK | 2.98 | 2.69 |
| PHI | 3.97 | 3.60 |
| PIT | 3.64 | 3.61 |
| SD | 2.81 | 2.46 |
| SEA | 1.74 | 2.89 |
| SF | 3.88 | 3.34 |
| STL | 3.55 | 3.33 |
| TB | 3.47 | 3.15 |
| TEX | 3.14 | 3.64 |
| TOR | 3.22 | 4.14 |
| WSH | 3.39 | 3.17 |

**Table 7: HBP by team**

Based on the theory of moral hazard, we would expect the American League to throw more retaliation pitches than the National League. However, table 8 presents us with the opposite result. Moreover, it is evident that National League pitchers actually hit more batters than American League pitchers. One possible explanation for this result could be that National League pitchers are more inclined to hit an opposing batsman because they know they have the opportunity to face the opposing pitcher. It is possible that National League pitchers expect the opposing pitcher to make an out when he takes his turn at bat. Essentially, it is less costly for National League pitchers to put a man on base (i.e. hit an opposing batter) than American League pitchers.

| League | Retaliation Pitches (%) | ALL HBP (%) |
|---|---|---|
| AL | 47.77 | 47.98 |
| NL | 52.23 | 52.02 |

**Table 8: HBP by league**

Finally, it is important to consider how pitchers behave depending on the situation of the game. Sabermetercians[7] have actually developed a leverage index that allows fo us to quantify a situation during the course of a baseball game. Essentially, there are some points during a game that are more suspenseful than other, and a leverage index allows us to observe differences in performance at these different levels. There are three levels, which are broken down as high, medium, and low. These levels are calculated based on the inning, score of the game, number of outs, and number of men on base. A tie ball game in the bottom of the ninth with two outs and the bases loaded would imply a high leverage situation. Conversely, if it is the bottom of the first inning with the home team down six runs with nobody out and nobody on base, this would illustrate a low leverage situation. The vast majority of situations during the game are low leverage[8]. Table 9 listed below shows HBP based on leverage. The data is comprised of all games from 2008 through the 2013 seasons.

[7] A sabermetrician is one is studies advanced baseball statistics (i.e. sabermetrics).
[8] Approximately 60% of situations are low leverage. The remaining 40% include medium and high leverage situations, with majority of that being medium leverage situations.

| HBP by Leverage | | | |
|---|---|---|---|
| Leverage | Percent | Total PA | Total HBP |
| Low | 0.80% | 542737 | 4359 |
| Medium | 0.86% | 456934 | 3917 |
| High | 0.97% | 114885 | 1115 |

**Table 9: HBP by leverage**

The percentage seen in the table implies the percentage of hit batsmen per plate appearance. Interestingly, we observe that as the situation in the game becomes more important, there is a positive correlation with the number of hit batsmen. I would have expected to observe the opposite effect because I would have expected pitchers to focus more, which should result in better control. However, the results indicate that pitchers may become more nervous in high leverage situations, which causes poor control, which ultimately leads to a higher percentage of hit batsmen. Another possibility could be a problem of a small sample size for the high leverage situations. Over 100,000 plate appearance certainly seems to be a large amount; however, it is relatively small compared to the number of plate appearances observed with the low leverage situation.

The leverage situations can be quite difficult to quantify, and I have yet to determine how to analyze leverage with regard to retaliation pitches. However, I would not expect pitchers to throw retaliation pitches in a high leverage situation. I believe that teams value winning games more than they value following the unwritten rules in baseball (i.e. throwing a retaliation pitch).

FORMAL REGRESSION AND RESULTS

Table 10, which can be seen below, provides a description of the key variables used in the two regression models.

| Variable | Description |
|---|---|
| Intent (Dep. Variable) | equal to 1 if pitch is more than 2.583 feet inside, 0 otherwise |
| pitch | equal to 1 if retaliation pitch, 0 otherwise |
| inning | inning HBP occurs |
| winperc | winning percentage of pitcher's team |
| AL | equal to 1 if pitcher's team is in American League, 0 otherwise |

**Table 10: Description of key variables**

There are two regression models used to estimate the intent of the pitch, and both models are fairly similar. Probit models require a binary (i.e. coded with zeros and ones) dependent variable but allow us to predict for the probability for the intent of the pitch. In order to generate this binary variable, I calculated the average horizontal distance of a pitch that hits a batter. This distance was equal to 2.065 feet. I then added one standard deviation to the mean, which equaled approximately 2.583 feet. A pitch with a horizontal distance greater than 2.583 feet was equal to one, and I assume to be intentionally thrown at a batsmen. A pitch less than 2.583 feet was equal to zero, and I assume that pitch did not intentionally hit the batter.

The binary dependent variable is referred to as "intent" and predicts the probability that a particular pitch was intentional based on a variety of circumstances. There are also a number of predictor variables, which include "pitch," "inning," "winperc," and "AL." "Pitch" is perhaps the most important variable and is a dummy variable equal to one if the pitch is considered a retaliation pitch and equal to zero otherwise. The variable "inning" is equal the inning in which the hit batsman occurred. "Winperc" is equal to the winning percentage of the team of the pitch who hit the opposing batsmen. Finally, "AL" is a dummy variable equal to one if the pitcher is on an American League team and zero if the pitcher is on a National League team.

| Variable | Regression 1 | Regression 2 (Fastballs) |
|---|---|---|
| Intercept | -0.825<br>(-6.98)*** | -0.907<br>(-5.76)*** |
| pitch | 0.051<br>(1.10) | 0.170<br>(2.84)*** |
| inning | 0.014<br>(2.37)** | 0.011<br>(1.33) |
| winperc | -0.533<br>(-2.34)** | -0.397<br>(-1.31) |
| AL | 0.041<br>(1.28) | 0.038<br>(0.89) |

Note: z-statistics are in parentheses. Asterisks indicate significance at the 1% (***), 5%(**), and 10%(*) levels

**Table 11: Regression results**

The only difference between the two probit regression models is that the first model accounts for all hit batsmen, whereas the second model only includes hit batsmen that were hit by fastballs. Interestingly, both regression models show a positive correlation in that a retaliation pitch increases the probability of that pitch being a pitch that was intentionally thrown at a batter (i.e. greatest horizontal distance from the center of the plate). However, only the second regression, which only includes fastballs, is statistically significant. A retaliation pitch is actually statistically significant at the 1% level for the fastballs model. The model tells us that assuming all other factors are equal to zero, a retaliation pitch that is a fastball is roughly 17% more likely to be intentional. This result falls right in line with my hypothesis, as I suggested pitchers are likely to have the best control of their fastball. Since there is a positive and statistically significant relationship with throwing a retaliation pitch farther inside, this would suggest that perhaps pitchers are intentionally trying to hit an opposing batsman with this pitch.

Furthermore, since the "pitch" variable is not significant in the first model, this would imply that the off-speed pitches that hit batters are likely mistakes.

The two models also allow us to consider for moral hazard with regard to the differences between the American and National Leagues and throwing a pitch intentionally at an opposing batsmen. While the results initially indicate that there is a positive correlation with regard to American League pitchers throwing farther inside than National League pitchers, this result is not statistically significant; therefore, I am unable to conclude that moral hazard is a factor when pitchers are intentionally retaliating. While I believe it is certainly possible and perhaps even likely that moral hazard is a factor with regard to hit batsmen, my results indicate that it is not relevant with regard to intentionally throwing at an opposing batsmen. The paper by Goff, Shughart, and Tollison found that American League pitchers are more likely to throw inside than National League pitchers; however, maybe there is no significant difference between the two leagues when pitchers are intentionally throwing at an opposing batter. Perhaps, that when pitchers find the need to throw a retaliation pitch, they do not consider any sort of consequences. This would suggest that pitchers simply follow the unwritten rule in Major League Baseball, which was referred to previously: "if you hit one of our guys, then we are going to hit one of your guys."

So far as the final two explanatory variables ("inning" and "winperc") are concerned in the two models, they do not provide us with strong and useful conclusions. As a matter of fact, these two variables are only statistically significant (at the 5% level) in the first regression. The negative coefficient on the "winperc" variable implies an inverse relationship between winning percentage and throwing inside. Moreover, better teams will throw inside less. This is the opposite effect that was observed in summary statistics section of this paper. However, we see no statistical significance with regard to fastballs.  The inning coefficient also provides us with very little useful information; however it is statistically significant at the 5% level in the first regression.  Previously we observed the drop in hit batsmen in the fifth inning, so this could explain the inconsistency in the "inning" variable in the model.

## FUTURE WORK

The Pitchf/x database is a fantastic resource that has allowed baseball enthusiasts to explore many of our important questions regarding baseball. However, the technology is still fairly new and is constantly improving; therefore, in some respect, we are still limited by what we can study.

One aspect that is very important to a pitcher's control is the pitcher's pitching mechanics. Often times, poor or inconsistent pitching mechanics will result in wildness, which could likely lead to a hit batsmen. For decades, it has proved to be a struggle to evaluate pitching mechanics over a large sample. Many researchers have observed film and photographs to evaluate pitching mechanics, but this method usually proves to result in a relatively small sample size. We are still learning how we can use the Pitchf/x system to evaluate pitching mechanics. Mike Fast of Baseball Prospectus provides an excellent analysis of how pitching mechanics can be evaluated using Pitchf/x[9]. Pitchf/x does not do a very good job explaining to us the pitcher's release point. Pitchf/x has proved to be quite accurate with the information and data of a pitch when the baseball is in the vicinity of home plate and even as the baseball travels towards the plate. However, the area in the vicinity of the pitcher's mound (i.e. pitcher's release point) could still be improved. The reasoning for this is that Pitchf/x cameras that are installed in each of the thirty Major League ballparks are calibrated based on certain "landmarks" within the ball's trajectory. Home plate is closer to more "landmarks" within the ballpark, whereas the pitcher's mound is essentially in the middle of the baseball diamond. Many researches have attempted to estimate a fixed release point for all pitchers. However, the issue with this method is that pitchers are different heights with different wingspans, which will certainly result in a

[9] http://www.baseballprospectus.com/article.php?articleid=12432

variation of release points. This method would likely only be effective if evaluating individual pitching mechanics rather than a large sample size.

The Pitchf/x system does provide us with the release speed of every pitch. This velocity is calculated at the point in which the system believes the pitcher releases the ball; however, it is likely this point occurs shortly after the pitcher actually releases the ball. This point is measured in horizontal feet from the center of the pitching rubber to vertical feet from the ground. Figure 12 should provide a better understanding of this idea. Despite the fact that this is not the true moment that the pitcher actually releases the baseball, we could still use this point because it is expected to be the same for each individual pitcher. In order to evaluate pitching mechanics using this point, one would only include fastballs in the dataset because pitchers are likely to change their arm angles in an effort to produce movement on the baseball. It would also be necessary to construct a sample of pitchers that pitch a sufficient amount of innings so that you could derive an accurate sample mean of the release point for each individual pitcher. You would then flag pitches within your dataset that deviate too far from each pitcher's given mean. I would hypothesize that retaliation pitches with a release point closer to the pitcher's average release point would more likely to be thrown intentionally at an opposing batsmen.

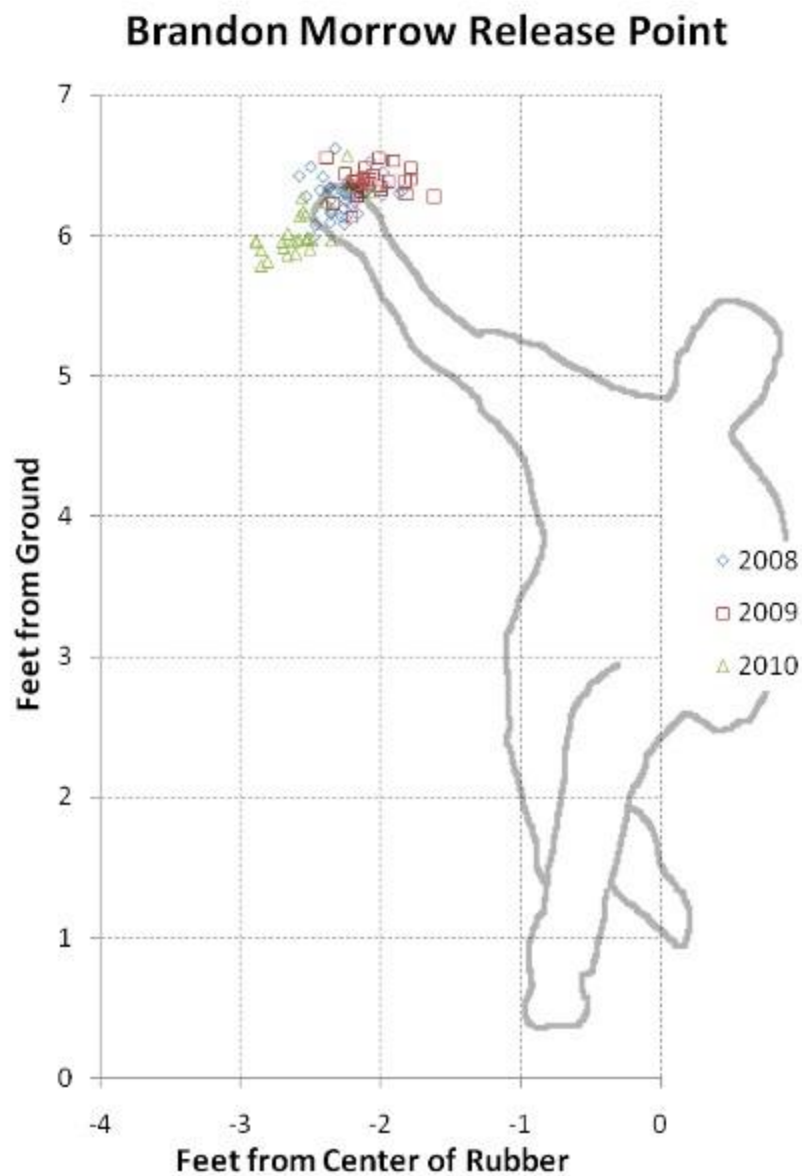


**Figure 12: Release point graphic**

# CONCLUSION

Attempting to detect for intentional retaliation pitches has certainly proved to be a difficult task due to the fact that is impossible to fully understand the pitcher's thoughts and motives as he throws a pitch from the mound. However, I am certain that the method I have used is a strong start in an effort to better understand this phenomenon in Major League Baseball.

Sports provide economists with an abundance of accurate data to test economic theory, and I am excited with the direction of statistical analysis in Major League Baseball. Baseball has taken the lead in statistical analysis in sports. Because of this, many economists and sabermetricians are incentivized to test not only economic theory but pursue their own personal interests in baseball. For some reason, baseball has always been a subject that has interested various individuals to study and with the implementation of Pitchf/x, this will only intensify future research. As the Pitchf/x system continues to advance and become more accurate, I am certain that our knowledge of baseball will also become more advanced. The Pitchf/x data has only existed since the 2008 season, so we will continue to learn how we can evaluate advanced baseball statistics in a more riveting and innovative manner.

WORKS CITED


Abrevaya, Jason, and Robert E. McCulloch. "Reversal of Fortune: A Statistical Analysis of Penalty Calls in the National Hockey League." (2004): 1-50. Print.

Baker, Wayne. "Love to Heckle? Toast? Or Should We Stop the Jeering?" *The Ann Arbor News* 22 June 2010: n. pag. Web. 30 Oct. 2013. <http://www.annarbor.com/lifestyles/love-to-heckle-toast-or-should-we-stop-the-jeering/>.

"Beaned by a Pitch, Ray Chapman Dies." *New York Times* 17 Aug. 1920: n. pag. Web. 29 Oct. 2013. <http://www.nytimes.com/packages/html/sports/year_in_sports/08.17.html>.

Cameron, Dave. "Carlos Quentin's HBP Zone." *Baseball Statistics and Analysis | FanGraphs Baseball*. N.p., 12 Apr. 2013. Web. 1 Oct. 2013. <http://www.fangraphs.com/blogs/carlos-quentins-hbp-zone/>.

Cameron, Dave. "What Is PITCHF/x? | FanGraphs Sabermetrics Library." *Baseball Statistics and Analysis | FanGraphs Baseball*. N.p., n.d. Web. 1 Sept. 2013. <http://www.fangraphs.com/library/misc/pitch-fx/>.

Fast, Mike. "Spinning Yarn: Pitcher Release Points." *Baseball Prospectus*. N.p., 11 Nov. 2011. Web. 20 Oct. 2013. <http://www.baseballprospectus.com/article.php?articleid=12432>.

Goff, Brian L., William F. Shughart, and Robert D. Tollison. "Batter Up! Moral Hazard and the Effects of the Designated Hitter Rule on Hit Batsmen." *Economic Inquiry* XXXV (1997): 555-61. Print.

Kalkman, Sky. "Understanding Pitch F/x Graphs: Location vs. Movement." *Beyond the Box Score*. N.p., 17 Apr. 2009. Web. 30 Aug. 2013. <http://www.beyondtheboxscore.com/2009/4/17/841366/understanding-pitch-f-x-graphs>.

Levitt, Steven D. "The Hazards of Moral Hazard: Comment on Goff, Shughart, and Tollison." *Economic Inquiry* 36 (1998): 685-87. Print.

Pankin, Mark. *Retrosheet*. N.p., n.d. Web. <http://www.retrosheet.org/>.

Trandel, Gregory A. "Hit By Pitches: Moral Hazard, Cost-Benefit, Retaliation, or Lack of Evidence?" *Journal of Sports Economics* 5 (2004): 87-92. Print.

Trandel, Gregory A., Lawrence H. White, and Peter G. Klein. "The Effect of the Designated Hitter Rule on Hit Batsmen: Pitcher's Moral Hazard or the Team's Cost-benefit Calculation." *Economic Inquiry* 36 (1998): 679-84. Print.

Willman, Daren. "Baseball Savant: Advanced Baseball Statistics." *Baseball Savant: Advanced Baseball Statistics*. Baseball Savant, 20 Aug. 2013. Web. 20 Aug. 2013.

Wyers, Colin. "Databases for Sabermetricians, Part One." *The Hardball Times*. N.p., 24 Nov. 2009. Web. 29 Aug. 2013. <http://www.hardballtimes.com/main/article/databases-for-sabermetricians-part-one/>.