Investigating the Coefficient of Restitution of Major League Baseballs

by

David Hesslink

Submitted to the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science in Mechanical Engineering

at the

Massachusetts Institute of Technology

June 2017

Signature of Author: Signature redacted
Department of Mechanical Engineering

Certified by: Signature redacted
Anette Hosoi
Associate Department Head for Operations; Professor of Mechanical Engineering
Thesis Supervisor

Certified by: Signature redacted
Rohit Karnik
Associate Professor of Mechanical Engineering
Undergraduate Officer

Investigating the Coefficient of Restitution of Major League Baseballs

by

David Hesslink

Submitted to the Department of Mechanical Engineering
on May 12, 2017 in Partial Fulfillment of the
Requirements for the Degree of

Bachelor of Science in Mechanical Engineering

ABSTRACT

In the middle of the 2015 Major League Baseball season there was a sudden increase in home run rate. Such a sudden league-wide change is difficult to explain, and many baseball pundits suggested that a "juiced" or altered baseball was the culprit. In this thesis, large scale coefficient of restitution (COR) testing is performed on baseballs from before (2014) and after (2016) the alleged change. No evidence is found to suggest that the baseball was altered between 2014 and 2016. The investigation nevertheless revealed telling properties of the coefficient of restitution. First, it was largely found that baseballs from both years fell within the COR specifications of Major League Baseball (between 0.514 and 0.578). Second, the COR is revealed to be a noisy quantity in that the same baseball can exhibit differences in COR on the order of 0.01. Third, most of the variation between baseballs' COR comes from being manufactured on different days in the plant and thus being packaged in different boxes. Additionally, an unexpected fatigue in the strike plate used in the test apparatus is found to have affected the COR measurements for the first 200 trials.

Thesis Supervisor: Anette Hosoi

Title: Associate Department Head for Operations; Professor of Mechanical Engineering

Acknowledgments

The author would like to thank Peko Hosoi for her tireless guidance, enthusiasm, and willingness to lend a listening ear. Thanks also to MIT '17 Ben Eysenbach for his assistance generating a cross-validation video processing algorithm at a crucial moment in the project. Additional thanks to everyone who lent a helping hand, both in and out of MIT.

Table of Contents

List of Figures

1. Introduction

Baseball is one of the most popular American sports. Major League Baseball (MLB) sells over 70 million tickets [1] every year to watch games, and millions more follow games from their home through broadcast television [2]. At the center of the game is a matchup between the pitcher and the hitter. The pitcher, standing sixty feet, six inches away from home plate, throws the ball towards the batter, whose object is to swing a wooden bat to hit the ball. The batter endeavors to hit the ball as hard as he can; when the ball is hit so far as to clear the outfield fence, the batter scores a home run.

In July 2015, the average exit velocity of hit baseballs in the MLB mysteriously increased by approximately 1 MPH with no apparent cause, leading to a burst in home runs [3]. Some baseball fans and academics believe the league tampered with the physical ball itself to make it livelier, increase home runs, make the games more exciting for spectators, and thus increase revenue. Indeed, MLB Commissioner Rob Manfred announced that a top goal of his was to find ways to "inject additional offense into the game" when he took over in January 2015 [4].

The liveliness of a baseball is determined by its *coefficient of restitution (COR)*, which is a measure of the amount of energy the ball loses during impact. The difficulty with testing for changes in coefficient of restitution is that testing is both expensive and unreliable, as a large sample size is needed to overcome the natural variance in ball performance [5, 6].

In this thesis, 499 COR tests were performed: 189 on 12 distinct 2014 major league baseballs from 1 packaged box of baseballs and 380 on 70 distinct 2016 major league baseballs from 10 separately packaged boxes of baseballs. If the theory holds true and the MLB did tamper with the baseball in an attempt to boost run scoring, the measured coefficient of restitution of the 2016 baseballs should be larger than that of the 2014 baseballs.

2. Background

2.1 Juiced Ball Conspiracies

The baseball itself begins to come under scrutiny whenever there are observed changes in player performance that cannot be easily explained by any other league-wide changes. There have been similar conspiracy fervors regarding "juiced" or "loaded" major league baseballs in the 1920s, 1980s, and 2000s, to name a few instances [7]. However, these conspiracies have never been confirmed in Major League Baseball.

The lack of confirmed tampering in the MLB does not completely invalidate the hypothesis of a juiced ball, however. In 2013, there was approximately a 40% home run increase in Nippon Professional Baseball (NPB), the Japanese equivalent of Major League Baseball [8]. Baseball pundits questioned the integrity of the ball, and the league denied any foul play in the construction of the physical baseballs. After a month of denying the allegations, the NPB reversed its stance, admitting to colluding with the baseball manufacturer Mizuno to make the ball livelier and boost offense [9]. The scandal shook the league to its core and led directly to the resignation of the league's commissioner [10].

In March 2016, Rob Arthur and Ben Lindbergh of FiveThirtyEight.com compiled the statistical case for a tampered baseball midway through the 2015 season [3]. They point out that the average exit velocity of batted balls begins to climb significantly beginning in July 2015, increasing by a full MPH before the season ended in September, which led to a parallel spike in the home run rate of batted balls, up from 3.7% midseason to 4.2% by the end of the season. They tested for many factors to account for the difference in exit velocity, controlling for both temperature and player-specific effects. However, they were still unable to explain the uptick in offensive performance, leaving them to speculate about the validity and transparency of the MLB's internal ball performance tests. The aberrantly high exit velocities and home run total continued beyond 2015 into the 2016 MLB season, suggesting a sustained trend of increased offensive performance [5]. The same enhanced levels of offensive output have been evident to date in the 2017 MLB season [11].

2.2 How Baseballs Are Made

Before delving into the performance of major league baseballs, it is important to understand where they come from. Every baseball used in the MLB is manufactured at the same Rawlings manufacturing plant in Turrialba, Costa Rica, where about 2.4 million baseballs are made every year [12]. Each ball is made of six layers: the pill, three separate layers of wound yarn, a finish winding, and the two pieces of cowhide cover, hand-stitched together by 108 stitches [13]. Figure 1 shows the cross section and relative size of each of these layers.

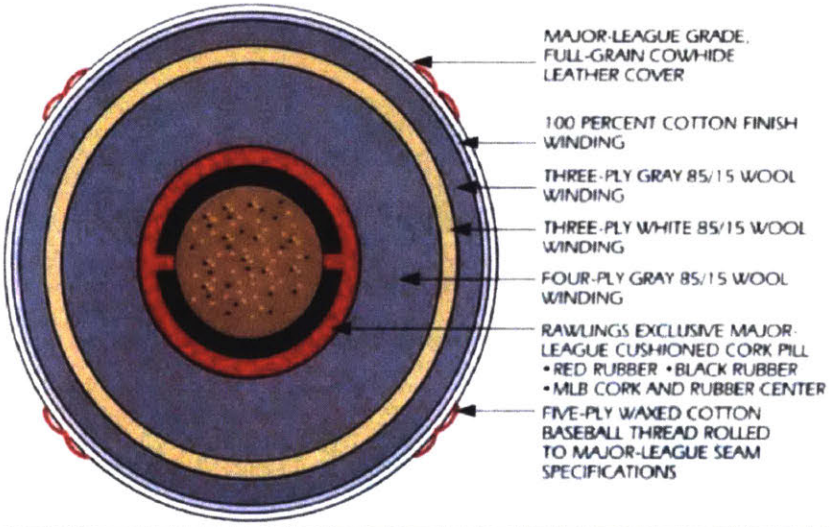

Figure 1: The cross section of a major league baseball. The ball is made of six different layers, and the performance of the baseball can be affected by each layer's composition. Figure credit: *Popular Science* [13].

In 2000, the UMass-Lowell Baseball Research Center published the results of their third party inspection of the manufacturing facility in Costa Rica [14]. During the entire manufacturing process, the temperature and humidity of the half-constructed baseballs are very carefully monitored. First, the cork and rubber pills are glued and allowed to dry for 24 hours before the yarn is machine spun around the pill in a dry environment. The cowhide covering is then hand-

stitched onto the ball in a 75% humid sewing room before it gets rolled and set to dry. COR testing is performed daily for balls that are made by each of the seven distinct winding machines, but these results are private quality control results for internal Rawlings use only.

2.3 Coefficient of Restitution

The coefficient of restitution e is a simple measure of the energy lost during a collision between two objects. It scales as the square root of energy loss:

$$e = \frac{\sqrt{KE_{translational,final}}}{\sqrt{KE_{translational,initial}}}. \tag{1}$$

The coefficient of restitution as quantity is undefined for a singular object; it must be associated with a collision between two separate objects. Thus, to say "the COR of a baseball" is not completely precise. In the industry and throughout the duration of this paper, a baseball's coefficient of restitution is understood to be the coefficient of restitution between a baseball and a fixed wooden plate that is vertically attached to a stationary wall. In this case, the formula for the coefficient of restitution simplifies even further to:

$$e = \left|\frac{v_{x,final}}{v_{x,initial}}\right|. \tag{2}$$

As shown by Equation (2), the coefficient of restitution is by definition a one dimensional quantity; it relates the speed of the ball only in the direction of interest. Thus, the effects of gravity are ignored in our horizontal launch test. In the case of a perfectly elastic collision, kinetic energy would be completely conserved ($e = 1$). In the case of a perfectly inelastic collision, the ball would stick to the wall and all kinetic energy would be dissipated ($e = 0$). In most cases, including those relevant to this experiment, the strike is neither perfectly elastic nor inelastic, giving us a coefficient of restitution $0 < e < 1$.

The MLB requires that the coefficient of restitution of game-used baseballs must lie between 0.514 and 0.578 [15]. When traveling at game speeds, baseballs at the upper and lower end of the legal spectrum exhibit noticeable performance differences of up to 35 feet of travel distance on long fly balls [6].

3. Experimental Design

3.1 Apparatus

The physical testing apparatus was modeled after the COR testing setup of the UMass Lowell Baseball Research Center in 2000. A pitching machine was set up to launch baseballs at a fixed wooden strike plate, and each shot was recorded by a high-speed camera at 240 frames per second, as shown in Figure 2. Leading up to the test, the balls were stored in the same indoor humidity and temperature controlled environment for three days.

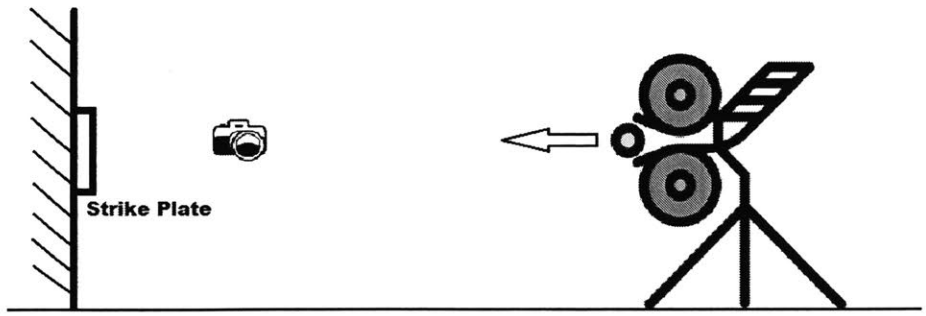

Figure 2: A schematic of the physical testing setup. A three wheeled pitching machine launches the baseball at 60 MPH towards a vertically fixed strike plate. The ball hits the strike plate and rebounds back towards the pitching machine. A high speed camera perpendicular to the flight of the ball tracks the ball's motion.

The pitching machine used was a three wheeled Hack Attack© machine. The left and right wheels were each set to 4, and the bottom wheel was set to 6.5. These wheel settings made the machine shoot the ball straight with a slight backspin at about 60 MPH. The pitching machine was set 10 feet directly perpendicular to the strike plate.

The strike plate was mounted such that the center of the strike plate was at the same height as the release point of the pitching machine. The strike plate was a 12"x12" target made of hickory wood, 4" deep. The plate was comprised of two 2" deep 12"x12" plates epoxied together in a cross grain orientation to preserve the structural integrity of the wood for the duration of testing.

The high speed camera was stationed on a tripod parallel to the strike plate and perpendicular to the direction of motion of the ball, such that the ball did not move into or out of the frame of the camera shot. Additionally, a bright studio light was turned on next to the camera to increase the contrast and resolution of the high speed video being shot. Between every COR test, a whiteboard was updated with the trial number of that particular shot and placed in the lower left corner of the video frame. A key was kept documenting which ball was being used in each trial.

3.2 Methods

3.2.1 Video Processing Methods

Videos were manually processed using the publicly available software program Tracker©. First, the videos were trimmed to only include the frames that had the ball in them. Next, the software was calibrated using a black frame of known length 13.876". The frame was arranged perpendicular to the camera at the same depth of field as the ball. Next, a coordinate origin was set near the strike point of the ball. Neither of these measures are necessary to compute the coefficient of restitution. The scale cancels in the coefficient of restitution ratio, and the coordinate shift doesn't affect the position deltas that define velocity. An example of a processed video in Tracker is shown in Figure 3.

Figure 3: A screenshot of a tracked COR test in Tracker. First, the scale is calibrated using the blue calibration stick on the black frame. Next, a point mass is created, and the ball is tracked as it enters the left side of the screen, rebounds off the strike plate, and exits off the left side of the page. The trial number is in the lower left of the frame, and the data in the lower right of the screenshot are transferred to Microsoft Excel.

To transform the video into (t, x, y) coordinates, a "point mass" was created in Tracker. The ball was then manually tracked by clicking the center of the ball in each frame of the high speed video. The timestamp, x position, y position, and Tracker-estimated acceleration were returned from each video and saved into an Excel file. The ball is tracked for six feet as it approaches and rebounds from the strike plate.

3.2.2 COR Computation Methods

The coefficient of restitution is the ratio of two quantities—the horizontal speed before and after impact. Thus, the points need to be classified by whether they occurred before, during, or after the strike. The strike point is identified as the peak of the Tracker-estimated acceleration of the ball. Because the ball's strike doesn't always fall exactly in the frame, the strike was defined as the peak acceleration point plus any neighboring points with an estimated acceleration greater than 2540 m/s^2, denoted by a blue horizontal line in Figure 4. Any points with a timestamp before the collision are then taken as pre-strike data, and points timestamped after the collision constitute post-strike data. The R data processing code is included in Appendix B.

Once the data have been divided into before and after the strike, the velocity is estimated for both parts of the trajectory. To obtain velocity, a linear model is fit to the time and horizontal position data of the form

$$x_{out}(t) = v_{out}t + x_{strike} \tag{3}$$

$$x_{in}(t) = v_{in}t + x_0 \tag{4}$$

as shown by the red fitted lines in Figure 4.

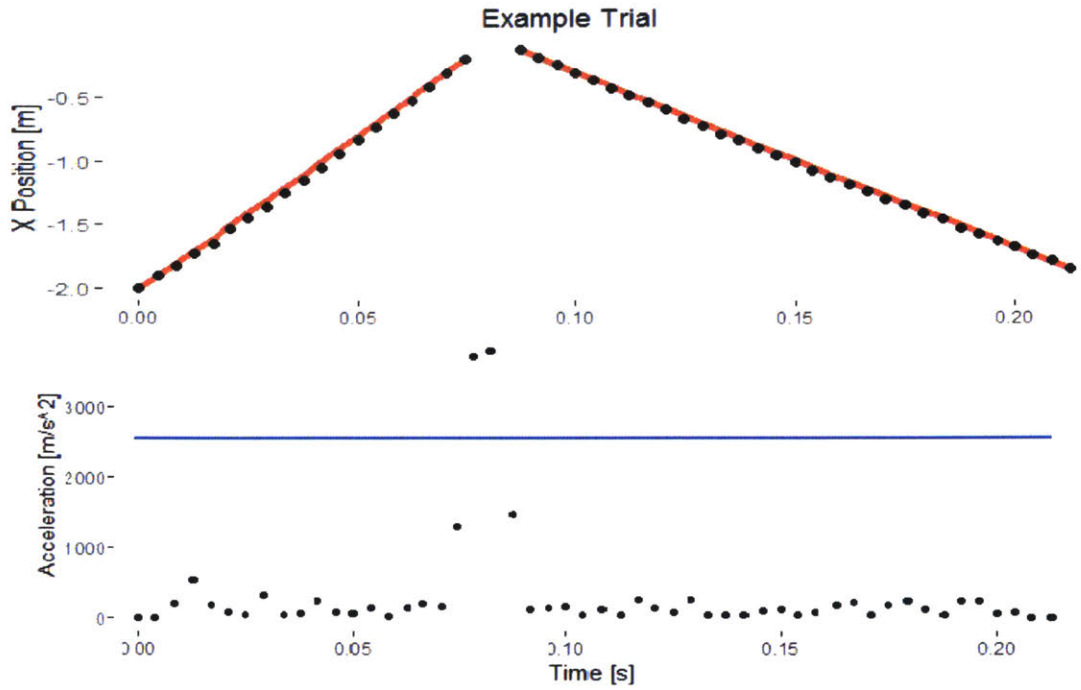

Figure 4: The data from a tracked COR test in Tracker. The upper figure is the horizontal position of the ball as a function of time, calibrated by the fixed-length black frame in Figure 3. The lower figure is the ball's acceleration as a function of time, generated from Tracker's 5-point internal acceleration estimation formula. The acceleration is only used to determine where the ball strike occurs and to eliminate points with an acceleration above the blue line, which corresponds to frames in which the ball is colliding with the strike plate. The remaining position data is then linear in time, and a line is fit through both sets of data. The slope of the red lines is the estimated velocity from before and after the ball hits the strike plate, respectively.

Once the linear regression has been performed, the initial and final velocity of the ball is extracted. The coefficient of restitution can then be computed as

$$e = \left|\frac{v_{out}}{v_{in}}\right| . \tag{5}$$

3.2.3 Precision of Methods

Because of the mediocre resolution of the camera and inconsistent background in the camera's field of view, the trial videos were processed manually instead of via an algorithm. The manual processing method is less repeatable than a defined algorithm, but it is robust in eliminating any algorithmic bias in processing the data, such as a consistent error when the ball crosses in front of a white background.

To understand the uncertainty induced through this manual processing method, a randomly chosen trial was independently processed ten times. After manually clicking to track the ball's location in that same video ten times, the mean coefficient of restitution for that particular trial is

found to be 0.52918. The standard deviation of the different coefficients of restitution obtained from repeatedly processing the same randomly chosen video was .00064. Thus, the resolution of this video processing method is on the order of 0.001. However, this variability is small enough to detect changes that are within the legal limit of a baseball's coefficient of restitution.

4. Results and Discussion

4.1 The Effects of Testing One Ball

Generally speaking, it is assumed that the coefficient of restitution is a constant property over time—that is, the previous history of the ball shouldn't affect its coefficient of restitution. To test this assumption, three baseballs were impact tested 20 times in a row to investigate any possible time effects. Time effects could bias the tests performed in this experiment, but they could also create strategic advantages for one team. For instance, if the ball displays noticeable fatigue after repeated impact, then it would be in the pitcher's best interest to keep using the same ball for as long as possible.

No such history-dependent trend is found. For each of the three balls that were tested, no correlation is found between number of previous tests on that ball and the coefficient of restitution; linearly predicting coefficient of restitution as a function of time leads to p-values that are greater than 0.5. See Appendix A for a visualization of the three balls that underwent 20 consecutive tests.

While no trend over time is observed with repeated testing, there remains significant variation across trials for a particular baseball. The noise associated with one ball's coefficient of restitution may arise from minor differences in the spin of the ball as it strikes the wall, which part of the ball is the first to make contact with the wall, and how the internal yarn wrapping and rubber/cork pill compress during the impact. The average standard deviation for a given ball across each of its trials is 0.0148. Thus, one ball could feasibly exhibit behavior towards both the lower and upper end of the legal spectrum of coefficient of restitution (.546 ± .032). Because the uncertainty of one ball across multiple tests is an order of magnitude larger than the uncertainty induced from the manual video processing method, it is safe to conclude that the observed trends are not a byproduct of the video processing method.

4.2 The Effects of Testing Multiple Balls from Multiple Boxes of Baseballs

Baseballs come packaged in boxes containing twelve balls each. While each ball is manufactured at the same Rawlings manufacturing plant in Costa Rica, they are manufactured on different days, with slightly different material, weather conditions, and factory operators. However, balls that are packaged in the same box of baseballs were all produced on the same day, from the same batch of raw material, and by the same operator. The relative variation between each ball in a sample box of baseballs is shown in Figure 5.

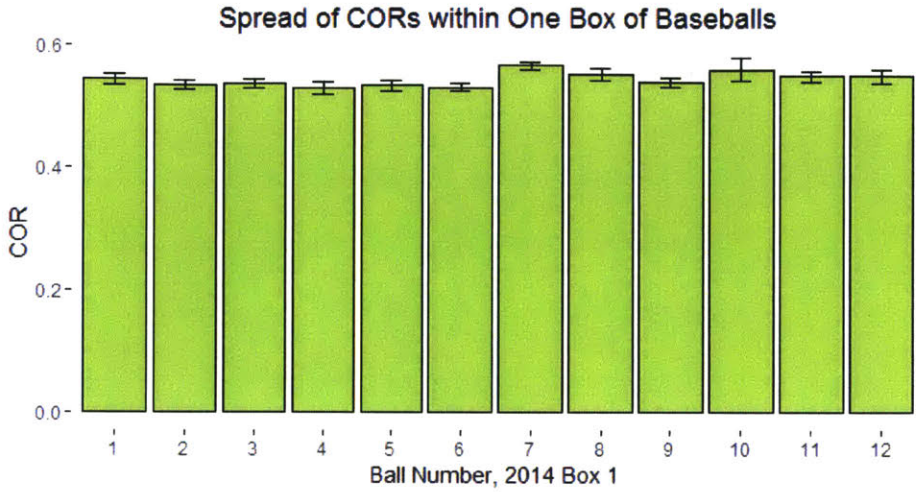

Figure 5: The coefficient of restitution for each ball in the tested box of 2014 major league baseballs. The 95% confidence interval is shown for each ball's coefficient of restitution.

A linear mixed effects model was run in R to control for the natural variance of coefficient of restitution tests and the variation between the boxes. The resulting standard deviation induced from the variability within one box is 0.0066. Figure 6 shows the average coefficient of restitution for the different boxes of baseballs that were sampled.

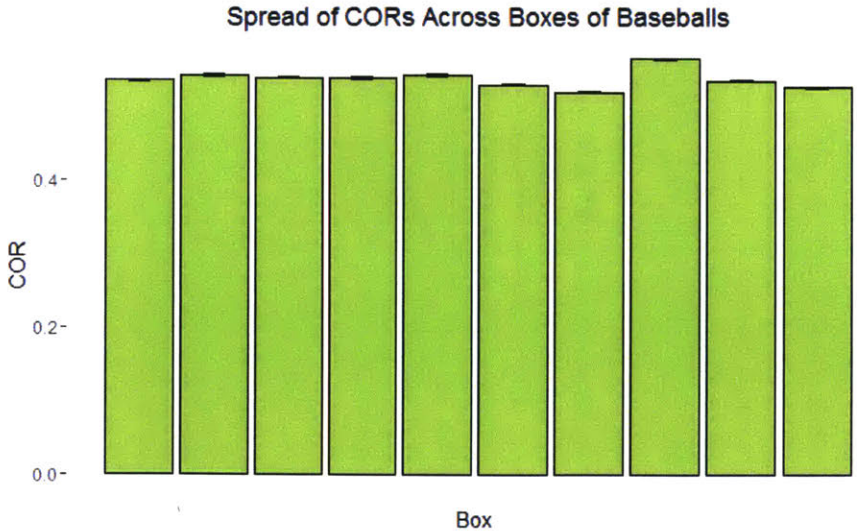

Figure 6: The average coefficient of restitution for each box of 2016 major league baseballs that was tested. The 95% confidence interval is shown for each box's average coefficient of restitution. There is a larger variation in coefficient of restitution between boxes of baseballs than there is within one box of baseballs.

The standard deviation of the coefficient of restitution across the different boxes of baseballs in the ball population is 0.011. This deviation is significantly higher than the deviation induced from intra-box variation, which suggests that there is indeed a noticeable effect between the performances of baseballs that were manufactured in different batches at the MLB baseball manufacturing facility in Costa Rica.

4.3 Difference between 2014 and 2016 Baseballs

With an understanding of the variability associated with baseball's coefficient of restitution, the theory of the baseball becoming juiced can be investigated. The distribution of baseballs' coefficient of restitution for the 189 COR tests of 12 distinct 2014 major league baseballs and the 380 tests of 70 distinct 2016 major league baseballs is shown in Figure 7.

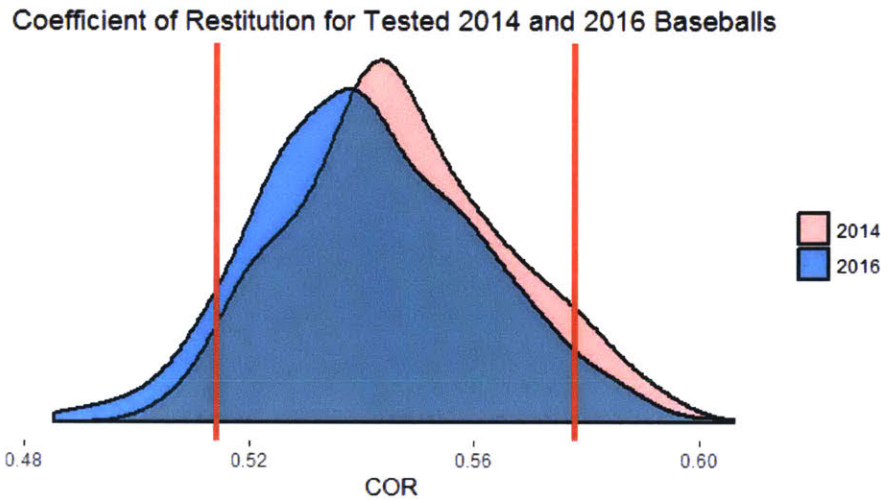

Figure 7: The distribution of baseballs' coefficient of restitution Contrary to expectation, the 2014 baseballs actually exhibit a slightly higher coefficient of restitution than their 2016 counterparts. Most of the measured coefficients of restitution are within MLB specifications, denoted by the vertical red bars. The tests that lay outside these specifications are within about a standard deviation of a ball's performance on a given shot, making it entirely feasible that these tests were from balls whose true coefficient of restitution is within the legal specifications.

The 2014 baseballs exhibit a slightly *higher* coefficient of restitution than the 2016 baseballs, which is opposite of the alleged change in the baseballs. However, this difference is statistically insignificant, as it has a p-value of 0.549.

The box of 2014 baseballs has a mean coefficient of restitution that is 0.0067 higher than the population mean. The standard deviation from box to box is 0.011, which means that the 2014 box is only 0.61 standard deviations higher than the mean. Thus, it is impossible to tell if the difference in 2014 and 2016 baseballs comes from the difference in year or is just a result of the natural variation between boxes of baseballs. More than one box of 2014 baseballs would need to be tested to overcome the box-to-box effect and conclude that the observed 2014 difference is from the year difference of the boxes.

4.4 Fatigue of the Testing Apparatus

Section 4.1 demonstrated that the coefficient of restitution is independent of the number of times that particular ball has undergone impact testing. However, a curious trend was observed over the course of an entire day of testing baseballs' coefficients of restitution, as shown in Figure 8.

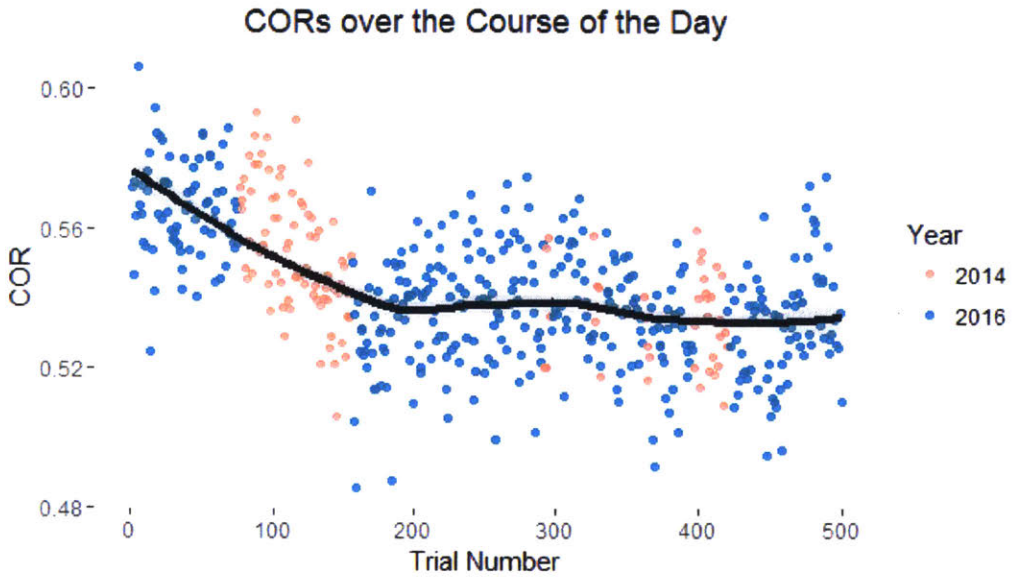

Figure 8: The trend in coefficient of restitution as the number of total trials taken increased. The average coefficient of restitution is constant from about the 200th trial onwards, but it decreases from the first trial until the 200th. The temporal trend seems to be independent of the year the ball was manufactured in, suggesting that this trend is likely derived from something in the testing apparatus.

It is not immediately apparent what could lead to such a trend; there was no change in experimental methodology that should have led to a different average coefficient of restitution for the first 200 trials. There are four possible explanations for such a trend: 1) that the pitching machine changed launch speeds over of the course of the day, 2) that something unwittingly changed in the manual video processing from the first month of processing to the final month, 3) that the strike plate exhibited some form of fatigue, or 4) that the baseballs happened to have been tested in an order that led to this result. We explore these four hypotheses in the following sections.

4.4.1 Pitching Machine Trends over Time

The Hack Attack pitching machine was not moved during the experiment, except to occasionally re-aim the machine at the center of the plate when the reaction forces of the launches accumulated to slightly deflect the ball of its original trajectory. Additionally, the machine was plugged into the same outlet for every trial. The machine was powered off once for about one hour after trial number 155. The impact speeds over the course of the day of testing are shown in Figure 9.

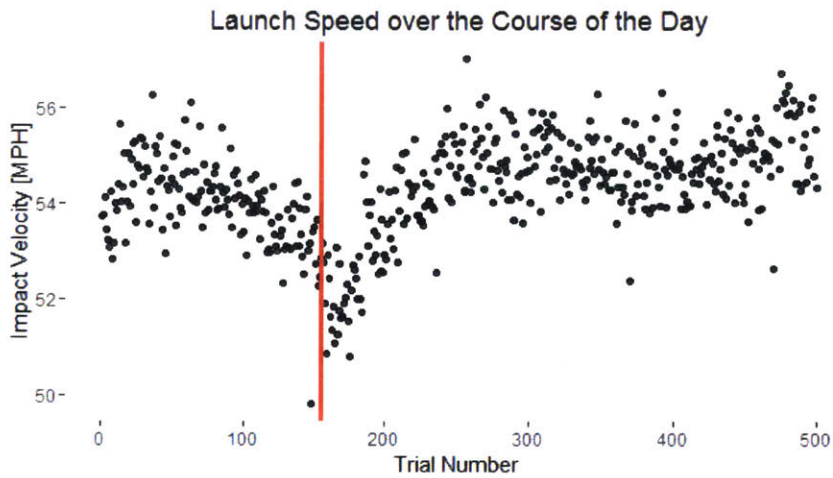

Figure 9: The trend in impact velocity over the course of the testing day. There appears to be a temporal effect on the testing equipment, as the pitching machine exhibits a dip in impact velocity right around 55 MPH, which is slightly slower than the launch velocity of 60 MPH due to drag as the ball approaches the strike plate.

While there was a substantial change in the pitching machine's performance over the course of the day, there should still have not been any effect on the coefficient of restitution as a result of this trend; studies have shown that there is no correlation between coefficient of restitution and impact velocity [6]. That is, the speed of a pitch should have no effect on the ball's performance when it hits the bat, and this assumption is affirmed by Figure 10.

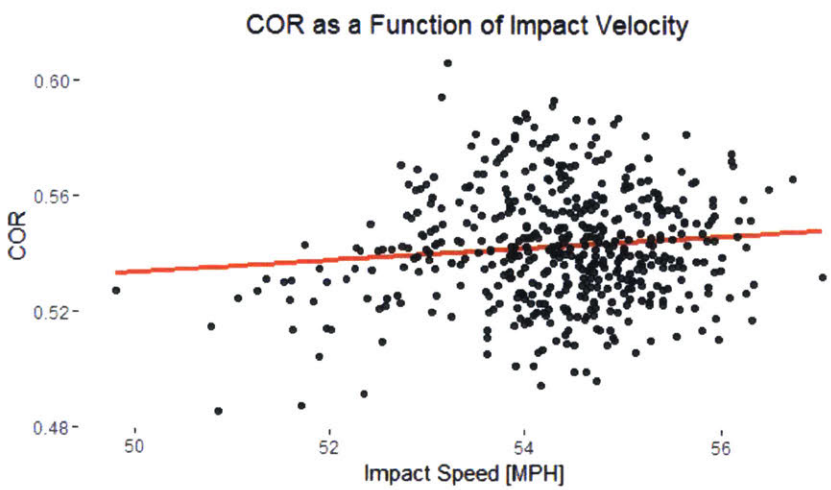

Figure 10: Predicting the coefficient of restitution based on the impact speed from the pitching machine. There is no correlation between the impact speed and the coefficient of restitution ($R^2 = 0.01$), so it is safe to conclude that the inconsistency in the pitching machine's launch speed did not cause the observed trend in coefficient of restitution.

This thesis affirms what Nathan and others have found [6]; the coefficient of restitution is independent of impact velocity, meaning the unusual trend in coefficient of restitution is not a result of the pitching machine's inconsistency over the course of the day.

4.4.2 Video Processing Method Comparison

The videos were manually processed over the course of about three months, so it is possible that the trend came not from any changes on the physical day of the testing, but rather from a subtle change in video processing methods over time. To investigate this possible confounding variable, a video processing algorithm was created that automatically tracked the location of the ball in each frame. The algorithm-generated coefficient of restitution results are displayed in Figure 11.

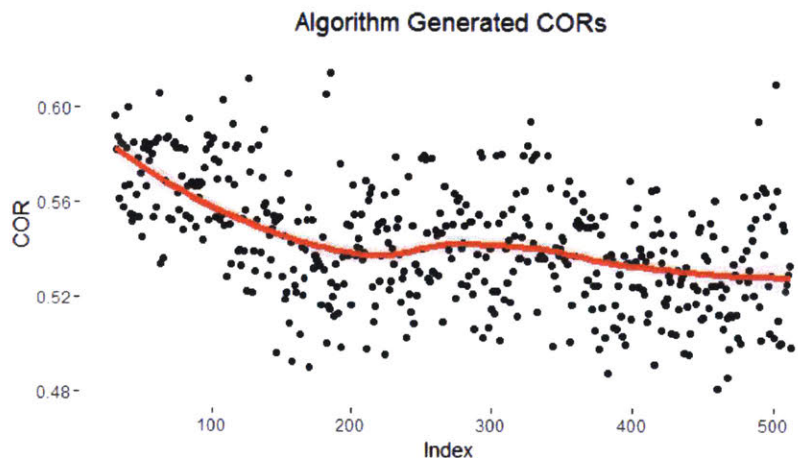

Figure 11: The results from the automatic video processing algorithm. The algorithm's computed coefficients of restitution follow the same trend as the manually processed coefficients of restitution. The consistency between the manual processing method and the automatic processing method suggest that the trend in average coefficient of restitution derives from the physical test setup and not the video processing method.

The automatic processing algorithm is designed to be an estimate of the coefficient of restitution. The algorithm induces outliers and fails to collect data from some of the trials. Additionally, it's unclear if there is an algorithmic bias that applies to every video, such as difficulty tracking the ball when it crosses in front of a white background. The coefficient of restitution is computed in a slightly different method in the algorithm. All in all, the algorithm is designed to be an estimate to validate the original methodology of video processing.

While the algorithm doesn't exactly replicate the results of the manual video processing method, it still exhibits the same trend in coefficient of restitution over time—namely, that the first two hundred trials display a decreasing average coefficient of restitution over time. Thus, the unusual trend in coefficient of restitution is not a result of any inconsistency in the manual video processing method over time.

4.4.3 Strike Plate Fatigue

There is a very slim chance that the baseballs were tested in exactly the order that leads to the observed trend in coefficient of restitution. However, it seems almost certain that an external factor biased the tests. The effects of the pitching machine were shown to not be the cause of such a trend. The camera was never repositioned, and no settings on the camera were changed during the course of the experiment. The video processing methods were consistent over time, as

validated by a rough video processing algorithm. The only component of the experiment left to explain the trend of the coefficient of restitution is the strike plate itself.

As discussed in Section 3.1, the strike plate is a 12"x12" square of hickory wood 4" deep. The strike plate was mounted to a large piece of plywood through back of the strike plate, such that no screws were on the strike surface of the strike plate. The large piece of plywood was then mounted to a cement wall with the strike plate on it, such that the plywood was flush with the wall. There was no apparent pattern in the location of the ball's strike on the strike plate, and there was no visible deflection of the strike plate during testing.

It is hypothesized that the brand new strike plate wore down over the first 200 experiments before becoming sufficiently worn in and yielding consistent average coefficient of restitution results after about the 200th trial. The mechanisms that would lead to such a change in the strike plate are uncertain, but this seems to be the only rational explanation for the observed pattern in average coefficient of restitution.

5. Conclusion and Future Work

This thesis is an independent investigation into the performance of major league baseballs before and after a noticeable home run spike in the MLB. While many have postulated that the league altered the ball to boost offensive performance and make the games more interesting, no evidence is found to support this theory. There is no statistically significant difference between the performance of 2014 and 2016 baseballs. These results are consistent with Major League Baseball's internal testing, which also found no change in ball performance in recent years [11].

The performance of baseballs' coefficients of restitution is better understood through this thesis. A particular ball can display a coefficient of restitution that is ± 3% of its true value during any given collision. This begins to explain why the MLB specification for coefficient of restitution is so wide to create performance gaps of up to 35 feet in batted ball flight [6]. It's also found that most of the variation between coefficients of restitution comes from different boxes of baseballs displaying different CORs, suggesting inconsistencies in the manufacturing process that make every batch of baseballs unique. This experiment also yielded an unexpected result in the behavior of the strike plate, which is suspected to wear down over the course of time before reaching a steady state after about 200 impacts.

If the baseballs haven't been juiced, how can the spike in exit velocity, home run rate, and offensive performance in Major League Baseball be explained? This remains an open question that puzzles baseball fans, academics, and executives everywhere. There are two leading theories to explain the growth that don't involve tampering with the baseball [16]. The first is that the pool of players may be morphing due to an influx of hitters that are better home run hitters, as a wave of young sluggers like Mike Trout, Bryce Harper, and Kris Bryant have revolutionized the game and brought their prestigious skill to the league. The second has to do with hitters' awareness of their launch angle and exit velocity. These quantities that define the quality of a batted ball began being measured in 2015, and the patterns associated with these quantities were understood as the sample size of measurements grew. Thus, midway through the 2015 season seems like a reasonable time for the effects of exit velocity and launch angle to stabilize. If

batters were to learn from these statistics and alter their batting approach to emphasize hitting the ball hard and at a home-run inducing launch angle, then an offensive trend like this one may be observed. Indeed, the average launch angle of batted balls has increased recently in Major League Baseball [16].

An altered hitting approach seems to be the most reasonable explanation of the league-wide offensive improvement in baseball, given the recent evidence that the baseball itself has not been juiced in recent years. If the offensive spike is indeed a result of a changed approach, then it's only a matter of time before the pitching begins to adjust league-wide to counteract the new approach and bring offensive production back to its normal level.

This thesis provides a valuable baseline for continued independent testing of coefficient of restitution of future baseballs. To make the process more tractable, it's recommended that a dark backdrop be placed behind the ball in the camera shot. With a consistent dark background and well-lit testing area, a video processing algorithm should be a reliable and efficient way to process the videos.

6. Appendices

6.1 Appendix A: Ball Fatigue

Figure 12 shows the results of the three baseballs that were tested twenty times consecutively to investigate the effect of past history on the coefficient of restitution. The results are plotted as three independent experiments on the same axes.

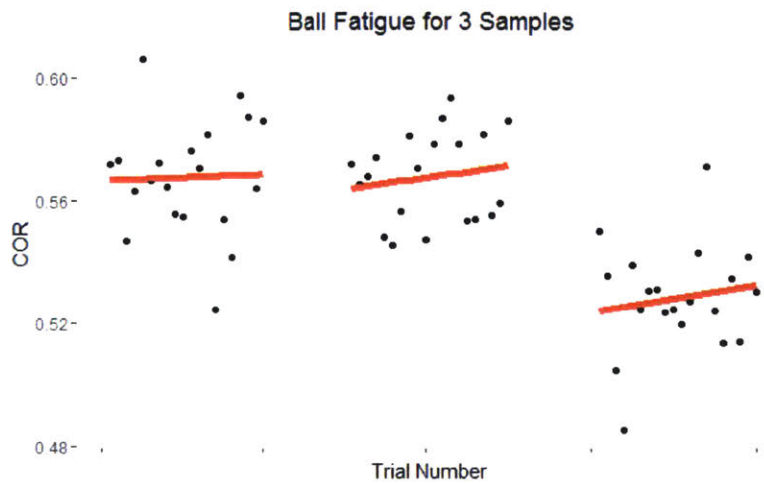

Figure 12: Three baseballs were each tested twenty times consecutively to test for ball fatigue effects.

If the ball were to have fatigued with repetitive impacts, the coefficient of restitution should have decreased with trial number. However, no such trend is observed. There is a slight counterintuitive upward trend for the second and third balls, but these trends are found to be statistically insignificant with p-values greater than 0.5.

6.2 Appendix B: Data Processing Code

Below is the R code that takes in the Tracker outputted .csv with the columns [step, t, x, a] and computes the coefficient of restitution based on the time and position data from Tracker. These data were collected manually, but the data collection process could be automated with better lighting and a consistent background in the camera shot.

```
data = read.csv(file.choose(), header = TRUE, stringsAsFactors =
FALSE)

#Replace the NA's in the dataframe with 0's
data[is.na(data)] = 0

#Function that computes the COR from a given trial
COR_fit = function(df, accel_cap = 100000) {
  #'COR_fit takes in a dataframe of time, position, and
  #'acceleration data from Tracker and computes the COR
  #'from that trial. Note that the dataframe input needs
  #'to be for just one trial, so the full dataframe needs
  #'to be trimmed before inputting here. Any point with an
```

```
  #'acceleration greater than accel_cap is thrown out as a
  #'noise from being near the wall. COR_fit fits a linear
  #'model x(t) = v*t + x0 to the time/position data and
  #'outputs speed before impact, speed after impact, and
  #'the COR from the one trial represented by df.

  #Determine where the ball strikes the plate
  #by looking for the spike in acceleration
  strike_step = which(df$a == max(df$a)) - 1

  #split the dataframe in two to separate pre & post strike
  before = subset(df, step < strike_step)
  after = subset(df, step > strike_step)

  #Remove the datapoints from near the strike by requiring points
  #to fall below a certain acceleration
  before = subset(before, a < accel_cap)
  after = subset(after, a < accel_cap)

  #Fit a simple x(t) = v*t + x0 for both before
  #and after the strike
  pre_strike_fit = lm(before$x ~ before$t)
  post_strike_fit = lm(after$x ~ after$t)

  #Extract the velocities from before and after the strike
  #from the linear model
  speed_in = pre_strike_fit$coef[2]
  speed_out = abs(post_strike_fit$coef[2])

  #Compute COR, remove the named number
  COR = as.numeric(speed_out/speed_in)

  #Return speed in, speed out, and COR
  to_return = as.data.frame(matrix(nrow = 1, ncol = 0))
  to_return$v1 = speed_in
  to_return$v2 = speed_out
  to_return$COR = COR

  return(to_return)
}

#Applies above COR function to each video in dataframe
generate_CORs = function(data) {
  #'data contains all the video files tracked information,
  #'indexed by a column called trial. Outputs a dataframe
  #'with one row per trial, containing index/COR pairs.

  #find number of videos in this dataframe
  rows = unique(data$trial)

  #initialize dataframe
  CORs = as.data.frame(matrix(nrow = length(rows), ncol = 0))

  count = 0

  #go through one video at a time
  for (index in rows) {
    count = count + 1
```

```
    #trim to just one video
    this_video = subset(data, trial == index)

    #save the trial number
    CORs$index[count] = index

    #call COR function to populate dataframe
    temp = COR_fit(this_video, accel_cap = 100000)
    CORs$COR[count] = temp$COR
    CORs$v_in[count] = .0568*temp$v1
    CORs$v_out[count] = .0568*temp$v2

  }

  #return the dataframe with 1 row per video
  return(CORs)
}

#Call function to create data, save in dataframe called 'CORs'
CORs = generate_CORs(data)

#Pull in the .csv with the ball number to trial index map
key = as.data.frame(matrix(nrow = nrow(key_import), ncol = 0))

#Rename columns to a more tractable form
key$index = key_import$Experiment.Number
key$year = sapply(strsplit(key_import$Box, "\\-"), '[', 1)
key$box = sapply(strsplit(key_import$Box, "\\-"), '[', 2)
key$ball = key_import$Ball
key$notes = key_import$Notes

#Join the COR data  with the key mapping data into a full dataframe
#called 'full_df'; note this requires the package 'dplyr'
library(dplyr)
full_df = inner_join(CORs, key, by = "index")
```

7. References

[1] "Major League Baseball (MLB) attendance 2006-2016 | Statistic," *Statista*. [Online]. Available: https://www.statista.com/statistics/193421/regular-season-attendance-in-the-mlb-since-2006/. [Accessed: 8-May-2017].

[2] "Viewers (on broadcast TV) of a Major League Baseball event in the U.S. 2016 | Statistic," *Statista*. [Online]. Available: https://www.statista.com/statistics/231834/mlb-events-on-broadcast-tv-usa/. [Accessed: 8-May-2017].

[3] R. Arthur and B. Lindbergh, "A Baseball Mystery: The Home Run Is Back, And No One Knows Why," *FiveThirtyEight*, 30-Mar-2016.

[4] "New commissioner Rob Manfred open to defensive-shift ban." [Online]. Available: https://sports.yahoo.com/blogs/mlb-big-league-stew/new-commissioner-rob-manfred-open-to-banning-defensive-shifts-172529801.html. [Accessed: 8-May-2017].

[5] R. Arthur and B. Lindbergh, "Are Juiced Balls The New Steroids?," *FiveThirtyEight*, 20-Jul-2016.

[6] A. Nathan, L. Smith, W. Faber, and D. Russell, "Corked bats, juiced balls, and humidors: The physics of cheating in baseball," *American Journal of Physics*, vol. 79, no. 6, pp. 575–580, Jun. 2011.

[7] J. Moore, "A Brief History of Juiced Ball Conspiracies," *The Hardball Times*, 05-Aug-2016. [Online]. Available: http://www.hardballtimes.com/a-brief-history-of-juiced-ball-conspiracies/. [Accessed: 9-May-2017].

[8] L. Granillo, "The rabbit ball: Did it happen here?," *SBNation.com*, 13-Jun-2013. [Online]. Available: http://www.sbnation.com/2013/6/13/4426478/japan-juiced-baseball-power-mlb-1987-rabbit-ball. [Accessed: 9-May-2017].

[9] J. Jaffe, "A juiced baseball scandal in Japan," *SI.com*, 12-Jun-2013. [Online]. Available: https://www.si.com/mlb/strike-zone/2013/06/12/japan-juiced-baseballs. [Accessed: 9-May-2017].

[10] "Japan commissioner resigns over juiced ball," *ESPN.com*, 19-Sep-2013. [Online]. Available: http://www.espn.com/mlb/story/_/id/9692461. [Accessed: 9-May-2017].

[11] B. Lindbergh, "Are Juiced Baseballs Behind Rising Home Run Rates?," *The Ringer*, 09-May-2017. [Online]. Available: https://theringer.com/2017-mlb-home-run-rate-is-the-ball-juiced-report-results-6e1dd0233203. [Accessed: 11-May-2017].

[12] S. Blaskey, "Costa Rica's major league concern -," *The Tico Times*, 12-Nov-2014. [Online]. Available: http://www.ticotimes.net/2014/11/12/costa-ricas-major-league-concern. [Accessed: 9-May-2017].

[13] J. Kaat, "Baseball's New Baseball," *Popular Mechanics*, pp. 62–67, Oct-2000.

[14] L. Fallon and J. Sherwood, "Performance Comparison of the 1999 and 2000 Major League Baseballs," UMass Lowell Baseball Research Center, Investigation.

[15] D. Kagan and D. Atkinson, "The Coefficient of Restitution of Baseballs as a Function of Relative Humidity," *The Physics Teacher*, vol. 42, pp. 89–92, Sep. 2004.

[16] T. Sawchik, "If the Ball Isn't Juiced, Then What Explains the Homer Surge?," *FanGraphs Baseball*, 12-May-2017.