Report of the Committee Studying
Home Run Rates in Major League Baseball

Jim Albert[1], Jay Bartroff[2], Roger Blandford[3], Dan Brooks[4],
Josh Derenski[2], Larry Goldstein[2], Anette (Peko) Hosoi[5], Gary
Lorden[6], Alan Nathan[7], and Lloyd Smith[8]

[1]Bowling Green State University
[2]University of Southern California
[3]Stanford
[4]BrooksBaseball.net
[5]MIT
[6]Caltech
[7]University of Illinois
[8]Washington State

May 24, 2018

Abstract

A report is given analyzing possible causes of the surge in home
run rate in Major League Baseball in the past several years.

1

Executive Summary

We, the members of the committee, were charged by the Office of the Commissioner of Baseball (BOC) to give the full benefit of our knowledge and expertise and to conduct primary and secondary research in order to identify the potential causes of the increase in the rate at which home runs were hit in 2015, 2016, and 2017.

The committee reported directly to Morgan Sword (Senior Vice President, League Economics and Operations), who, along with Reed MacPhail (Senior Director, League Economics) and Roy Krasik (Senior Director, League Operations) were very supportive of our many requests for resources that were needed to carry out our charge. The committee set its own agenda and arrived at its own conclusions and recommendations, independent of MLB.

Methodology

In order to conduct our investigations, we performed the following analysis and testing:

1. We reviewed and analyzed StatCast statistical data from the 2015, 2016 and 2017 seasons, including data on pitch type, exit velocity, launch angle, spray angle, spin rate, spin axis, and distance traveled. We also reviewed and analyzed HITf/x statistical data from the 2015 and 2016 seasons, including data on exit velocity, launch angle, and spray angle.

2. We reviewed and analyzed StatCast trajectory data, including position and velocity information on batted balls from impact to landing point.

3. We reviewed and analyzed data provided by Rawlings and UMass/Lowell regarding quality control testing that was performed on Major League baseballs during the 2010-2017 period.

4. We reviewed and analyzed Rawlings baseball manufacturing process and conducted an in-person inspection of the Rawlings factory in Turrialba, Costa Rica.

5. We conducted laboratory testing of 15 dozen unused Major League baseballs from 2013 to 2017, and a total of 22 dozen game-used Major League baseballs from the years 2012-2017, to measure properties of

the baseball, including properties that were not measured by Rawlings in the normal course.

6. We devised and conducted novel tests of the baseball at Washington State University to attempt to identify the factors that were causing baseballs manufactured in 2016 and 2017 to have lower drag coefficients than baseballs manufactured prior to 2016.

Our report is organized as follows. We begin with a summary of our findings and recommendations, as well as a brief bio-sketch of the committee members. After an initial Introduction, we present two sections in which we give details of the analysis that form the basis for these findings and recommendations. The primary basis for our findings comes from a detailed analysis of StatCast data, as described in Sec. 2. Supporting evidence for our findings comes from our investigation of properties of the baseball that can affect home run production, as described in Sec. 3, which includes a report on our visit to the baseball manufacturing facility in Costa Rica. Various technical details are given in the appendices.

Findings

The following summarizes the principal findings that result from our research and analysis:

1. The number of home runs per batted ball in MLB has increased steadily over the period 2015-2017.

2. StatCast data show that the increases in home runs are primarily due to better "carry" for given launch conditions (exit velocity, launch angle, spray angle) as opposed to a change in launch conditions. The better carry results in longer fly ball distances for given launch conditions and therefore more home runs. Analysis shows that the better carry is not due to changes in temperature but rather to changes in the aerodynamic properties of the baseball itself, specifically to those properties affecting the drag.

3. There is supporting evidence that the aerodynamic properties of the baseballs have changed, both from laboratory measurements and from

analysis of StatCast/Trackman trajectories, both for pitched and batted balls. Additionally, a physics-based model for the flight of the baseball shows that small changes in the aerodynamic properties of the baseball that are comparable to the measured changes in the drag coefficient since 2015 can explain the observed increase in home run production over the periods studied.

4. There is no evidence that the observed decrease in drag coefficient is attributable to a change in a property of the baseball that is currently tested by Rawlings or UMass/Lowell,including the size, weight, and seam height. While it cannot be ruled out that small year-to-year variations in these properties might be a minor contributing factor to the home run surge, these changes are within normal and expected manufacturing variation. There is no evidence that such variations occur either intentionally or through substandard quality control by Rawlings, but are inherent to the manufacturing process, which relies on substantial"by hand" labor.

5. The committee has not yet succeeded in definitively explaining the cause of the decreased drag coefficient beginning in 2015. Various hypotheses have been proposed and tested, including gradual changes in the manufacturing process affecting the centering of the pill within the baseball or the deformation of the baseball while spinning. There is an ongoing effort to develop more precise measurement techniques to investigate these hypotheses.

6. The increase in home runs between pre-ASG 2015 (2015a) and post-ASG 2015 (2015b) do not follow the pattern of subsequent increases. StatCast data show that the 2015a-2015b increase is primarily due to increased exit velocity rather than reduced drag. Nevertheless, the detailed dependence of exit velocity on launch angle is inconsistent with the pattern that would be expected if the increase were due to an elevated coefficient of restitution (COR) of the ball. Moreover, an increased exit velocity between 2015a and 2015b is not confirmed by HITf/x data, which instead support the conclusion that changes in the home run rate since 2015 are primarily the result of reduced drag rather than higher exit velocity. It appears that the 2015a StatCast data are anomalous. Perhaps that it not surprising, given that 2015 was the

first season for StatCast and contained a number of adjustments and recalibrations.

7. There is no compelling evidence from the laboratory testing of baseballs, whether by Rawlings, UMass/Lowell (UML), or Washington State (WSU), that changes in the COR or CCOR of the ball have played a major contributing role in the home run surge. This conclusion is strongly supported by the StatCast data, which show that the increase in home runs starting with 2015b is due almost entirely to better carry rather than to higher exit velocities, which would be a consequence of higher COR or CCOR.

8. Laboratory data show a strong dependence of the COR and CCOR (and, therefore, of home run production) on the relative humidity at which the balls have been stored, leading to our recommendation (see below) that the storage environments used by the clubs should be monitored by MLB.

9. The yearly reduction in average drag, which accounts for the change in the home run rate, is small compared to the variation in drag among baseballs within a given year, leading to our recommendation (see below) that MLB develop testing procedures and standards for the aerodynamic properties of baseballs.

10. While Rawlings has periodically replaced aging machinery and made minor changes in a few processes since the start of 2014, there is no evidence that these changes have contributed to the home run surge.

11. Suggestions that changes in batter behavior (e.g., "pull hitting" or trying to hit the ball at a higher launch angle) might be contributing to the surge are not borne out by the StatCast data. There has been no significant change in these aspects of batter behavior that correlates to an increase in home run hitting.

12. Analysis of StatCast data shows that the home run surge is "global", affecting players throughout the spectrum of home run hitting ability.

13. Many of the findings from this study have resulted from our analysis of StatCast data. While the conclusions we have reached follow logically from the available data, incomplete or missing data create a potential

source of bias. Nevertheless, the season-to-season comparisons appear to be consistent with our exploration of alternative datasets provided by Retrosheet and Baseball Reference.

Recommendations

The following summarizes the recommendations of the committee to the BOC:

1. Since changes in the carry of a fly ball have played an important role in the home run surge, MLB should work with Rawlings and/or independent test labs to develop methods to measure and monitor parameters of the baseball that affect the carry. This should include using the parameters, with fixed initial conditions, to monitor and control changes in carry distance.

2. MLB should monitor the climate environment (temperature, humidity) at which clubs store their baseballs. Given that this environment can have a significant effect on the COR and CCOR, MLB should require clubs to store the baseballs in a controlled environment that does not change over time.

3. MLB should monitor and attempt to standardize the application of mud on the baseballs, since the surface texture of the baseballs affects drag.

4. MLB should re-evaluate the specifications on parameters of the baseball that affect the game, such as size, weight, COR, and CCOR. They should also relax specifications on features that are unimportant, such as the color of the pill. Finally, they should require Rawlings to do impact testing of the pill at speeds producing deformations similar to those expected under game conditions rather than the low-speed drop test currently done.

5. Rawlings should be encouraged to develop and apply statistical methods to monitor the long- and short-term trends of their test data as part of their quality control procedures.

6. MLB should continue to study the drag properties of baseballs, with the goal of elucidating the reasons for the large variation of these properties among baseballs.

7. MLB should continue their efforts to improve the overall completeness and accuracy of the StatCast data, which were an important resource for the analysis leading to the findings in this report.

The Committee

- **Jim Albert** is Professor of Statistics at Bowling Green State University. He has written many published papers and authored four books on the interface of statistical thinking and baseball. He is active in the Section on Statistics and Sports in the American Statistical Association and is former editor of the Journal of Quantitative Analysis of Sports. He contributes regularly on the blog `https://baseballwithr.wordpress.com` that illustrates the use of the statistical system R in exploring baseball data.

- **Jay Bartroff** is Professor of Mathematics and the Graduate Vice-Chair for Statistics at USC. When he's not writing research papers and books on statistics and probability, he occasionally pontificates on sports outcomes for *ESPN* and its show *Sports Science*.

- **Roger Blandford** grew up (with cricket!) in England and is currently Luke Blossom Professor in the School of Humanities at Stanford University. He is the co-author of a recent graduate textbook Modern Classical Physics (Princeton University Press 2017) which discusses some of the basic physical processes that are relevant to Major League Baseball.

- **Dan Brooks** is a researcher who has long had an interest in baseball analytics. He runs the web site `http://brooksbaseball.net` and is co-organizer of the highly successful summer workshop known as Saberseminar, `http://saberseminar.com`.

- **Josh Derenski** is a statistics Ph.D. student in the USC Marshall School of Business. His research is motivated by the application of statistics to problems in various fields, particularly the field of environmental statistics. Current and past research projects include: Application of Tweedies formula to functional data, modeling water levels in lakes, and analyzing energy use in public schools.

- **Larry Goldstein** is Professor of Mathematics at the University of Southern California in Los Angeles. He specializes in statistics and probability, with his statistical work focusing on the development of new methodology. He has served as a statistical consultant to various industries for over three decades.

- **Anette (Peko) Hosoi** is the Neil and Jane Pappalardo Professor of Mechanical Engineering and professor of Mathematics at MIT. She is a Fellow of the American Physical Society (APS) and co-founder of the MIT Sports Lab. Her research interests include fluid mechanics and biomechanics, particularly the intersection of engineering, applied mathematics, and athletic performance.

- **Gary Lorden** is Professor Emeritus of Mathematics at Caltech. His research specialty is statistics, and he has consulted widely as a statistician for the last forty-five years.

- **Alan Nathan (chair)** is Professor Emeritus of Physics at the University of Illinois at Urbana-Champaign and has spent much of the past two decades doing research in various aspects of the physics of baseball. He has written numerous articles, both for academic journals and for the popular media, and runs an oft-visited website devoted to the topic, `http://baseball.physics.illinois.edu`.

- **Lloyd Smith** is a professor in the School of Mechanical and Materials Engineering at Washington State University. He is Editor in Chief of the journal Sports Engineering and has been testing bats and balls for 20 years. His Sports Science Laboratory (`https://ssl.wsu.edu/`) certifies bats and balls for 10 of 11 federations regulating amateur softball and baseball.

1   Introduction

Much has been written and discussed recently about the home run surge in MLB. As shown in Fig. 1, home runs per batted ball have increased steadily since 2015, so that the rate in 2017 is 35% greater than that in pre-ASG 2015. Speculation about reasons for the surge have settled on two possibilities:

- The properties of the ball have changed, including those that affect how hard a ball can be hit and those that affect the distance that a fly ball will carry. These properties are defined and their connection with home run production are discussed in the Appendices A and B, respectively.

- The batters are changing their approach, including swinging the bat harder to obtain larger exit velocity, swinging with an elevated plane to produce launch angles more conducive to home runs, or pulling the ball more to decrease the distance needed to clear the fence.
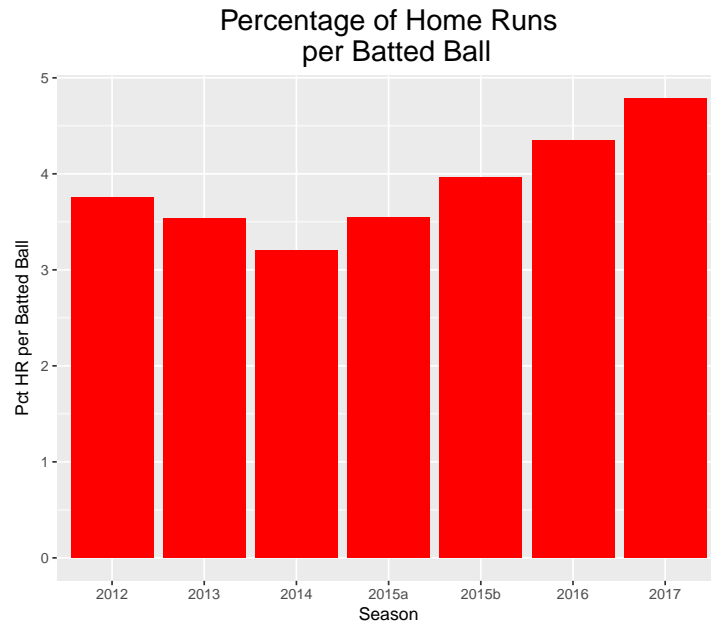

Figure 1: Trends in home runs per batted ball in MLB.

In this report, we use a variety of data sources to investigate these issues, including

- StatCast statistical data on most of the batted balls hit during the 2015, 2016, and 2017 seasons. These data include exit velocity, launch angle, direction, spin rate, spin axis, distance, and outcome.

- HITf/x data from 2015 and 2016, including exit velocity, launch angle, and direction.

- StatCast trajectory data, which include the coordinates and velocity of the batted ball from impact to eventual landing point. The technique for extracting drag coefficients from trajectory data is described in Appendix C.

- Historical laboratory data on various baseball properties from Rawlings and from UMass/Lowell. The latter has been employed by MLB for over a decade to test baseballs.

- New laboratory data from the Sports Sciences Laboratory at Washington State University. These data were commissioned specifically for this committee and include measurements of properties of baseballs that affect both how hard the ball can be hit and how well it carries.

The two principal parts of this report are Section 2, a presentation of our analysis of StatCast data, and Section 3, our study of baseball properties and their effect on home run production. Each of these sections begins with a summary of the principal findings. Various technical details are contained in the appendices.

2   Analysis of StatCast Data

The general task was to gain some understanding about the sudden increase in home run hitting in recent seasons. There is a basic series of events that lead to a home run. First, a ball needs to be put in play. Then the hitter needs the batted ball to have the right range of launch angles and high exit velocity to have an opportunity to hit a home run. Last, given the launch angle and exit velocity, the ball needs to have sufficient distance and height to clear the fence. Are batters hitting balls at a higher exit velocity or launch

angle that might explain the increase in the home run count? Alternatively, are there changes in the characteristics of the flights of the batted balls that might help explain the home run increase? The chance of hitting a home run depends on three inputs, the launch angle, the exit velocity, and the spray angle, and we explored the relationships between these inputs and the occurrences of home runs. We start by summarizing our principal findings and in the subsections that follow provide the detailed analysis leading to those findings.

2.1   Summary of Principal Findings

1. Although the rate of batted balls (per plate appearance) has decreased in recent years, there has been a substantial increase in home run hitting from the second half of the 2015 season through the 2017 season.

2. Practically all home runs occur for launch angles between 15 and 45 degrees and exit velocities between 90 and 115 mph. We will refer to this region as the "red zone" for hitting home runs. We subdivide the red zone (values of launch angle and exit velocity) into $5 \times 5 = 25$ rectangular regions to better understand the pattern of change in home run hitting.

3. There was an increase in the rate of batted balls hit in the red zone from the first to second halves of the 2015 season, but the rate in the red zone has stayed constant from the last half of 2015 through the 2017 season.

4. Within batted balls placed in the red zone, the rate of hitting home runs has dropped a little from the first to second halves of the 2015 season, but the rate of hitting home runs in this zone has exhibited a substantial increase in 2016 and again in 2017.

5. There has been little change in the exit velocities or launch angles of batted balls between the last half of 2015 and 2017 that correlates to an increase in home run hitting.

6. The increase in home run rate is "global", affecting players throughout the spectrum of home run hitting ability rather than dominated by a small group of sluggers.

7. The increase in home run rate between the first and second halves of the 2015 season cannot be explained by a change in drag properties of the ball, according to StatCast data. However, a contradictory conclusion is reached from HITf/x data.

8. The longer distances on fly balls hit with given initial conditions cannot be explained by changes in temperature.

2.2 Historical Perspective

2.2.1 Home run rates

The focus of this report is on the rate of home runs per batted ball. To provide a historical perspective, Figure 2 plots the home run rate for the seasons 1960 through the most recent season 2017. From 1960 through 1980 we saw a decrease in the rate of home run hitting followed by an increase through 2000 and then a gradual decrease until the 2014 season. There has been a dramatic increase in home run hitting the past three seasons and the 2017 rate of 0.048 home runs per batted balls is an all-time high.

Figure 2: Home runs per batted ball for the seasons 1960 through 2017.

Figure 3 displays the rate of home runs per batted ball for each week of the season from 2015 through 2017. There was a substantial increase in the home run rate from the first half to second half in the 2015 season. But there appears to be a general early season effect where the home run rate is low – this early season effect is seen in all three seasons. Also there is a substantial increase in the rate of home runs from 2015 to 2016 and again from 2016 to 2017.
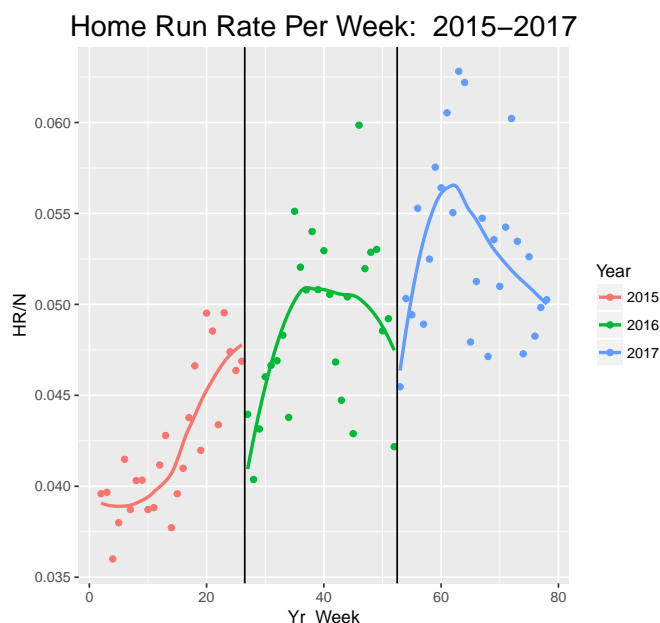

Figure 3: Home runs per batted ball for each week of the 2015 through 2017 seasons.

2.2.2 Change in rate of batted balls

While the rate of home runs per batted ball has increased since 2015, the number of balls put into play has declined over this period. Batted balls are the plate appearances that don't result in a strikeout, walk, or hit-by-pitch. As Figure 4 shows, the rate of batted balls has decreased from 73 percent in 2008 to the current 69 percent in the 2017 season. (Actually 69 percent is the lowest season rate of batted balls in baseball history.)
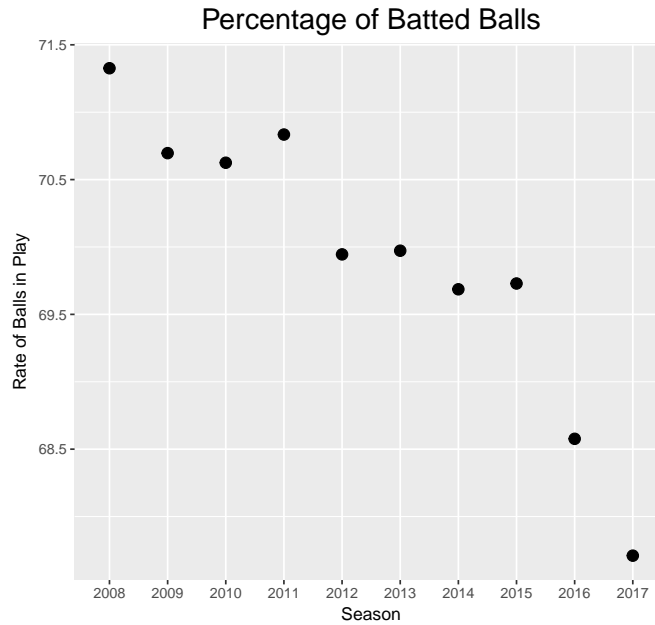
Figure 4: Percentage of plate appearances resulting in batted balls for seasons 2008 through 2017.

2.2.3 Change in home run rates within players

One possible explanation for these changes in home run rates is that particular groups of players are contributing to this home run surge. To control for the player effect, we next looked at the differences in home run rates for the same players in consecutive seasons (we only considered players who had at least 100 batted balls each seasons). For example, for the 1960 season, we looked at paired differences for players for the 1959 and 1960 seasons. Figure 5 displays the median of these paired differences of the home run rates across seasons. Although there is fluctuation in these values, note the large median value for the 2015, reflecting a large average increase for these players and again positive increases in the 2016 and 2017 seasons.
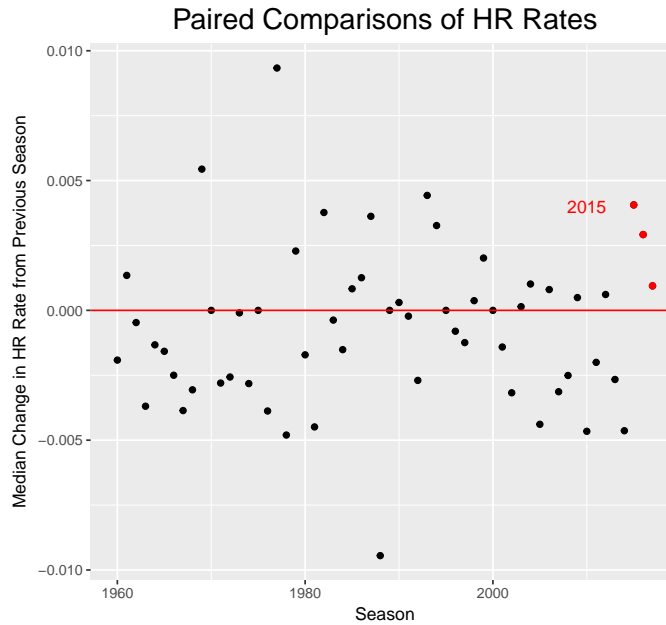
Figure 5: Differences in home runs per batted ball for same players for consecutive seasons.

To look at these increases in home run rates more carefully, Figure 6 displays the 10th, 25th, 50th, 75th and 90th percentiles of the home run rates for all players with at least 100 batted balls for the periods 2015a, 2015b, 2016 and 2017. Note that there has been a substantive increase in the home run rates among the top home run hitters (the 90th percentile), the average home run hitters (the 50th percentile), and the bottom home run hitters (the 10th percentile). This figure shows that there has been a large increase in home run hitting across the entire spectrum of home run talent.
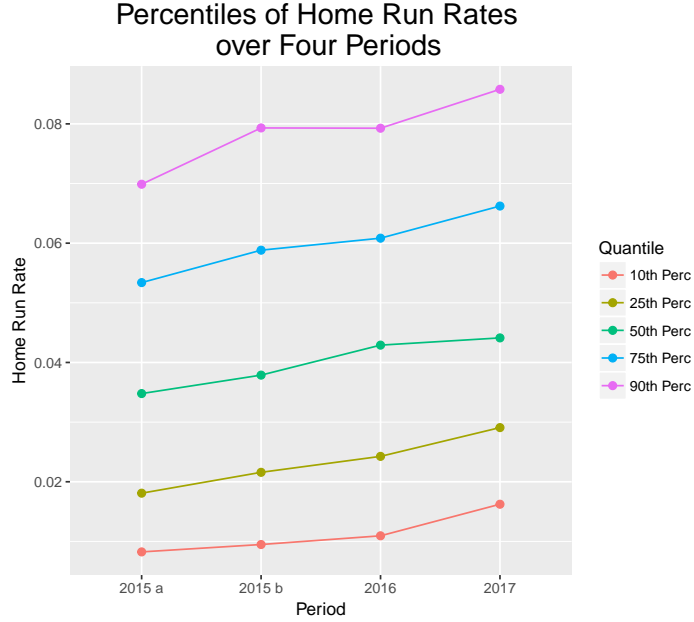
Figure 6: Selected quantiles of home run rates of players with at least 100 batted balls for four periods.

2.2.4 Pitcher effects

It seems plausible that in response to the home run surge pitchers (and catchers) would be tempted to alter the selection and location of pitches to reduce the risk of home runs. Do the pitch-level game data from 2015-17 show any such effect? We analyzed the pitch-by-pitch data in 32 categories: 8 categories of pitch types fastball, cutter, splitter, changeup, curve, sinker, slider, and knuckle- ball, each combined with four quadrants of pitch location — up and in (defined with respect to L or R batters), up and out, down and in, and down and out. The probabilities of home runs for these 32 categories were estimated from the 2015-2017 database of over two million pitches. If indeed the pitching strategies were changing to counter the surge, one would expect to see a tendency for some "high risk" pitch categories to show declining usage frequencies and some "low risk" pitch categories to show increasing frequencies. Neither effect was present in the data.

To get an overall comparison between the four time periods in terms of pitching strategy, we calculated for each period the expected number of home

runs based on the frequency of usage of the 32 categories of pitches. The probability of each category yielding a home run was held fixed at the values estimated from all of 2015-2017. The expected number of home runs per 1000 pitches were 7.76 in 2015a, 7.83 in 2015b, 7.86 in 2016, 7.75 in 2017. These differences are small, indicating that changes in pitching strategy have been insignifcant. In other words, the home run surge has not been attenuated by adjustments in pitching strategy.

2.2.5   Change in home run rates for ballparks

Another possible explanation for changes in home run rates are ballpark effects — perhaps particular ballparks are contributing to the surge in home run hitting. To explore this conjecture, we looked at the changes in the home run rate per batted balls for all ballparks for each change in periods – from the first half of the 2015 season to the end half, from the second half of 2015 to 2016, and 2016 to 2017. Figure 7 displays a dotplot of the changes in home run rates for all ballparks for each of the two-period comparisons. For each comparison, note that most of the ballparks exhibited increased home run rates in the following season. The most dramatic change was between the first and second halves of the 2015 season when 25 out of 30 ballparks exhibited increased rates.

Figure 7: Dotplots of the change in home run rates for all MLB ballparks for each change in periods. A vertical line is drawn at zero representing no change in rates.

2.3 MLB Data

We used a dataset provided by Major League Baseball of 337,881 batted balls collected during the 2015 through 2017 seasons. As Table 1 indicates, this dataset did not include the entire set of batted balls and home runs. This data includes 86-88 percent of the batted balls and 96-98 percent of the home runs.

Table 1: Actual counts of balls in play (BIP) and home runs (HR) and counts of BIP and HR in the dataset for the seasons 2015 through 2017.

Season	BIP	HR	BIP (Data)	Pct	HR (Data)	Pct
2015	130,473	4909	112,287	86	4734	96
2016	128,817	5610	112,744	88	5514	98
2017	127,556	6105	112,850	88	5979	98

The specific variables that will be explored are:

- The launch angle in degrees

- The exit velocity (mph).

- The spray angle (degrees) To adjust for the two batting sides, we defined the *adjusted spray angle* to be minus the observed angle for left-handed hitters. Then a negative adjusted spray angle corresponds to balls that are pulled and a positive value to those hit to the opposite field. In the remainder of this report, "spray angle" will represent the spray angle adjusted by the batting side.

- A 1/0 indicator if a home run was hit or not for the specific batted ball.

2.4   Launch angle and exit velocity

2.4.1   Distribution of launch angle and exit velocity

How has the distribution of launch angles and exit velocities changed during the four periods, 2015 first half, 2015 second half, 2016, and 2017? To address this question, Figure 8 displays a density graph of (launch angle, exit velocity) for each of the four periods. The rectangle in each graph represents the region of launch angle and exit velocity that produces most of the home runs. (See Section 2.5.2.) From this figure, it appears that the distribution of launch angle and exit velocity has been consistent across the four periods.

Looking more carefully, the top graph in Figure 9 shows the lower quartile, the median, and the upper quartile of the launch angles for each period, and the bottom graph shows the same summaries for the exit velocities for the periods. The median launch angle has shown a small increase – from 10.4 degrees in the first half of 2015 to 12.4 degrees in 2017. The median exit velocity has exhibited a small decrease from 91.5 mph in 2016 to 90.6 mph in 2017.
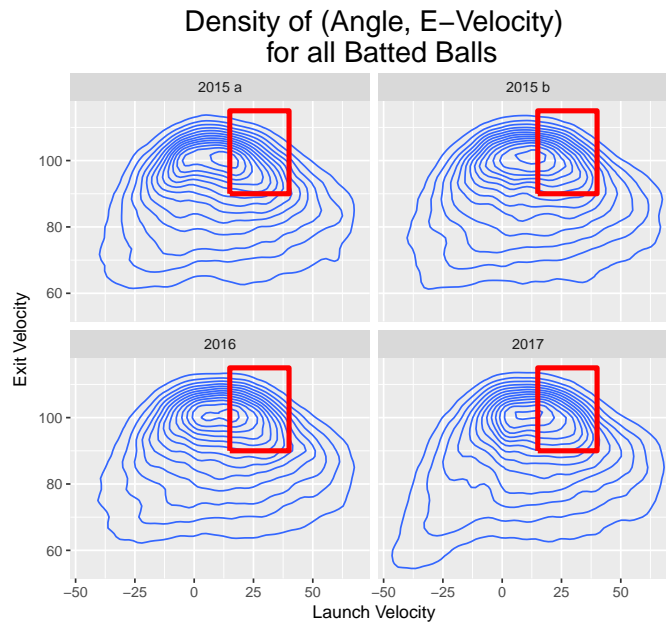
Figure 8: Distribution of launch angle and exit velocity of batted balls for the seasons 2015 through 2017. The rectangle in each graph is the region which produces practically all of the home runs.
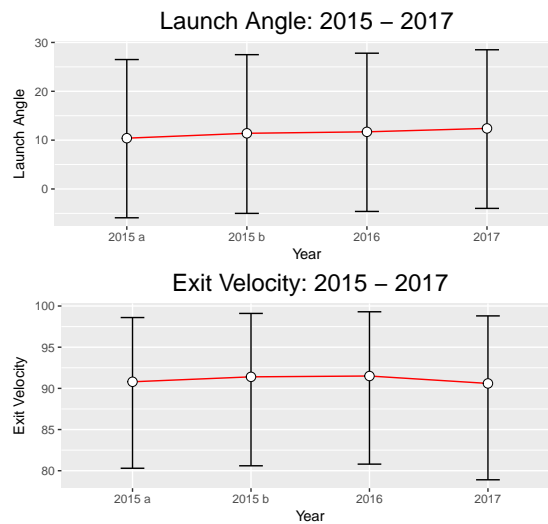

Figure 9: Median and quartiles of launch angle and exit velocity of batted balls for the seasons 2015 through 2017.

20

Figures 10 and 11 focus on the 2016 and 2017 seasons. Figure 10 plots the change in average exit velocity against the home run rate for all players with at least 200 batted balls each season, and Figure 11 plots the change in average launch angle against the home run rate for all players. The takeaways are that the exit velocities have decreased from 2016 to 2017, and the launch angles have exhibited an increase only for the players with small home run talents. These figures confirm that there is little evidence to support the claim that exit velocities or launch angles are contributing to the increase in home run hitting.
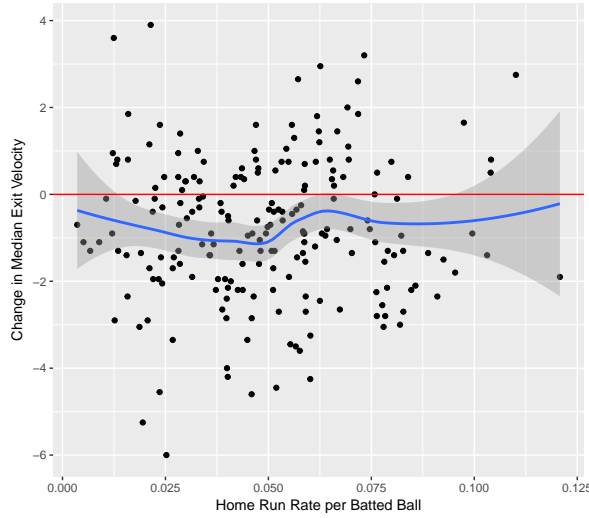

Figure 10: Scatterplot of home run rate and change in average exit velocity from 2016 to 2017 for all batters with at least 200 batted balls each season.

2.4.2 Relationship between launch angle and exit velocity

Since the exit velocity of a batted ball has a strong relationship with the launch angle, we explore how this relationship has changed over the four periods. Figure 12 displays a smoothed fit of this relationship for each of the four periods. Generally, we see that the exit velocity tends to be greatest for launch angles about 10 degrees. For larger launch angles (20 to 40 degrees), the exit velocity is smallest for the first half of the 2015 season. It is interesting that the average exit velocity for angles 0 to 20 degrees tends to be
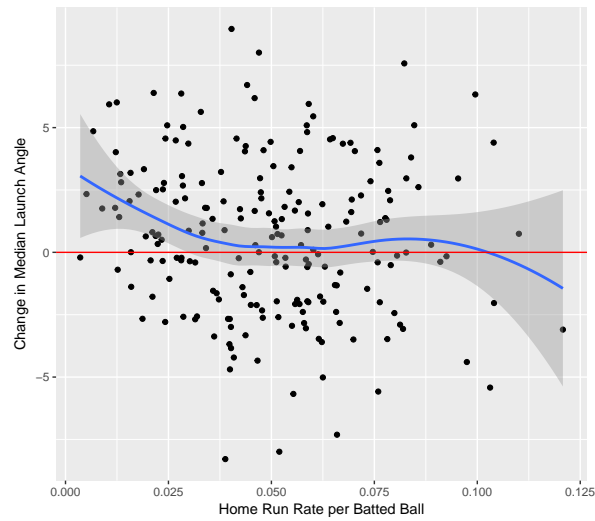
Figure 11: Scatterplot of home run rate and change in average launch angle from 2016 to 2017 for all batters with at least 200 batted balls each season.

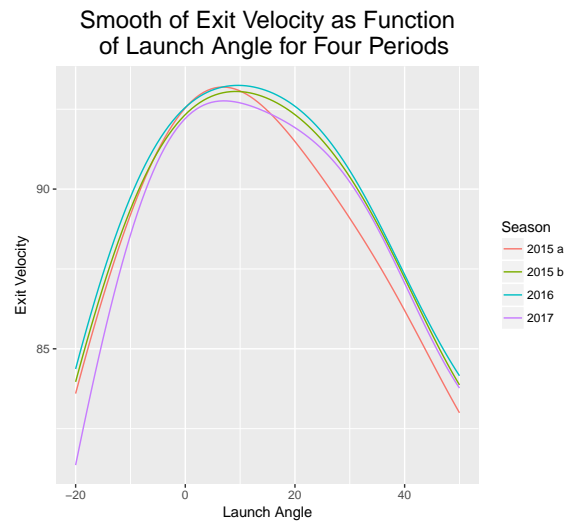smaller in 2017 than in the second half of 2015 and the 2016 seasons.

Figure 12: Smoothed fit of the exit velocity of a batted ball as a function of the launch angle for the four periods.

Suppose we focus on the batted balls in the region where the launch angle is between 15 and 40 degrees and the exit velocity is between 90 and 115 mph where most home runs are hit. Figure 13 displays boxplots of exit velocity of batted balls for the four periods for launch angles in different intervals. Figure 14 turns things around and displays boxplots of the launch angles of the batted balls for the four periods for fixed values of the exit velocities. The message from Figure 13 is that there has been little change in the exit velocities from the last half of 2015 through 2017, and Figure 14 indicates little change in the launch velocities during these last three periods.
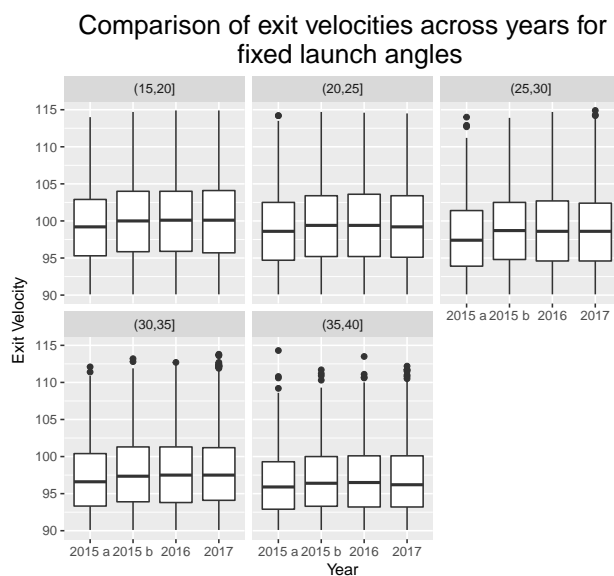

Figure 13: Boxplots of exit velocities of batted balls for four periods for fixed values of launch angles.
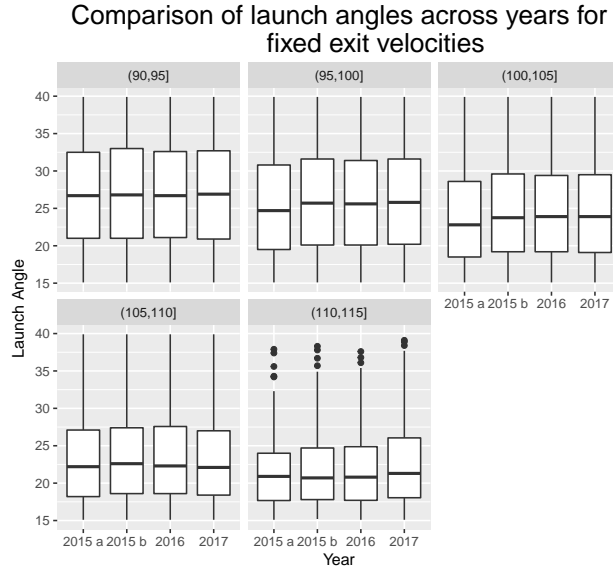
Figure 14: Boxplots of launch angles of batted balls for four periods for fixed values of exit velocities.

2.4.3 Distance of batted balls

The StatCast dataset also contains a variable that estimates the distance traveled for each batted ball. In our analysis, we only used batted balls that were tracked to at least 80% of the full distance, to assure the highest accuracy in the inferred distance. To understand the relationship between distance and other relevant variables, we fit a model that predicts the distance as a function of launch angle and exit velocity for the batted balls where the launch angle exceeds 20 degrees. Separate models are fit for batted balls from the second half of the 2015 season, the 2016 season, and the 2017 season. Figure 15 shows the predicted distance as a function of the launch angle for specific exit velocities for the three periods. Generally, for a specific launch angle and exit velocity, we see that the distance traveled is generally greatest for 2017, followed by 2016, and 2015b. The increase in distance from 2016 to 2017 can be as large as five feet for exit velocities near 102 mph and launch angles near 29 degrees. This increased distance in batted balls is consistent with the increase in the home runs in the last two seasons.
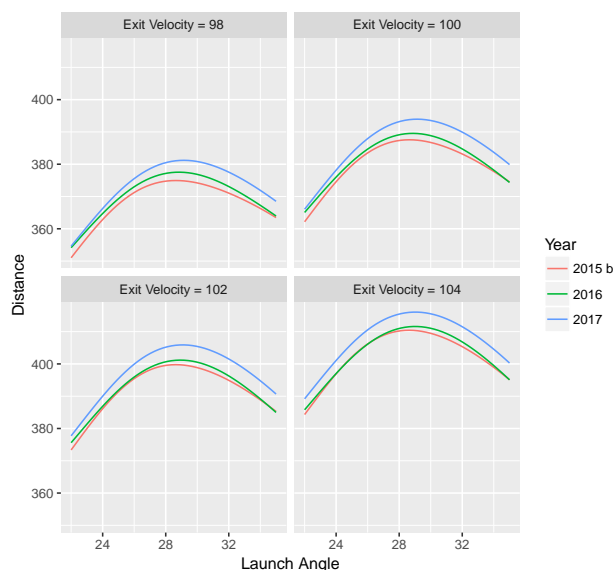
Figure 15: Predicted distances of batted balls as a function of the launch angle for four specific values of the exit velocity. The lines represent predictions using data from the second half of the 2015 season, the 2016 season, and the 2017 season.

2.5 Home runs

2.5.1 Introduction

Sec. 2.4 described the changes in characteristics of batted balls through recent seasons. Now we focus on exploring changes in the characteristics of home runs in recent seasons. We again use the launch angle, exit velocity, and spray angles recorded by StatCast.

2.5.2 Launch angle and exit velocity of home runs

Figure 16 shows the joint distribution of (launch angle, exit velocity) for home runs hit in the first and second halves of the 2015 season, the 2016, and 2017 seasons. We note that 97.5 percent of all home runs are hit in the region where the launch angle lies in the interval (15, 40) degrees and the exit velocity falls in (90, 115) mph. We focus on the home runs hit in this "red zone". Although we earlier observed small changes in the average launch

angle and exit velocity of all batted balls, the distribution of (launch angle, exit velocity) of home runs looks very consistent over these four periods.

Density of (Angle, E−Velocity) of HR

Figure 16: Density of launch angle and exit velocity of home runs for four periods.

2.5.3 Spray angle of home runs

There is a clear pattern of spray angles for home runs and this pattern appears to have remained the same through the four periods. Figure 17 shows density plots of the home run spray angles. A home run is most likely to occur at a spray angle at −20 degrees – that is, a home run is most likely to be pulled 20 degrees from dead center field towards third base. (Recall that we adjusted the definition of spray angle so that negative values always correspond to batted balls that are pulled.)

2.6 Putting a batted ball in the red zone, and hitting a home run

As mentioned earlier, there are two steps to hitting a home run. First, the batted ball needs to have the "right" launch angle and exit velocity, defined
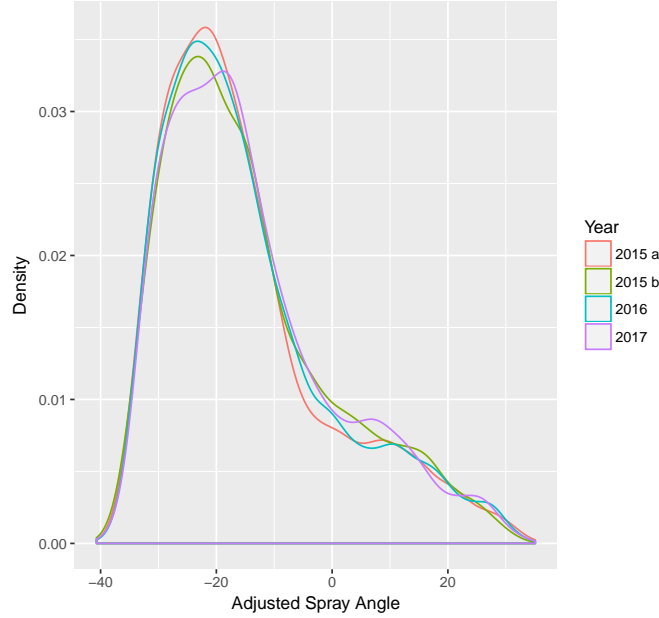
Figure 17: Density of spray angle of home runs for four periods.

by the red zone. Second, given the batted ball is in the red zone, it needs to have sufficient distance and height to be a home run.

There are two probabilities of interest:

1. The probability P_{ZONE} that a batted ball is in the red zone.

2. The probability P_{HR} a red zone ball is a home run.

Using partial Trackman data from some teams where the system was installed (seasons 2011 through 2014) and StatCast data (seasons 2015 through 2017), Table 2 displays the number of batted balls N, the number hit in the red zone Red, the probability of a batted ball in the zone $P_{ZONE} = Red/N$, and the probability a red zone batted ball is a home run P_{HR}=HR / Red.

Figure 18 shows how these two probabilities have changed from the seasons 2011 through 2017. The P_{ZONE} probability in 2011 was relatively high near 0.22, but it has fluctuated and in recent years has stabilized near 0.20. In contrast, the P_{HR} probability was about 0.21 during the seasons 2011 through 2013, dropped to 0.183 in 2014, and it has steadily increased from the first half of the 2015 season through the 2017 season.

27

Table 2: Fraction of batted balls in the red zone and the fraction of red zone batted balls that are home runs for seasons 2011 through 2017. Note that the red zone accounts for 97.5% of all home runs, independent of year.

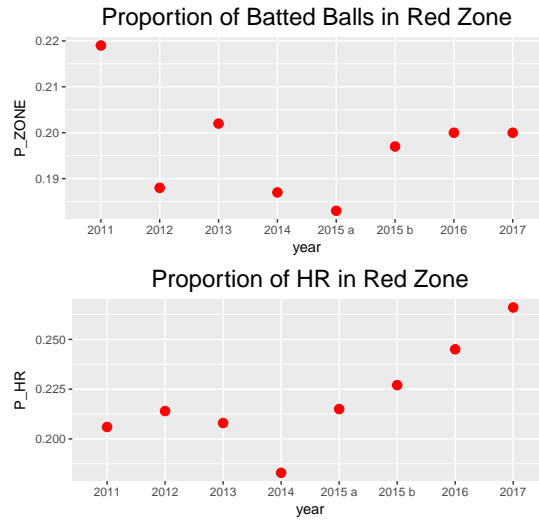| Season	N	Red	P_{ZONE}	HR	P_{HR}
2011	15216	3334	0.219	686	0.206
2012	31350	5881	0.188	1261	0.214
2013	52842	10670	0.202	2219	0.208
2014	74470	13961	0.187	2559	0.183
2015a	53713	9811	0.183	2113	0.215
2015b	58574	11537	0.197	2621	0.227
2016	112744	22541	0.200	5382	0.239
2017	112850	22514	0.200	5831	0.259

Figure 18: Fraction of batted balls that are in the red zone and the fraction of red zone balls that are home runs for seasons 2011 through 2017.

2.6.1 Counts and rates of home runs in the zone

During the last three seasons, the biggest change has not been in the rate of batted balls in the red zone, but rather the substantial increase in the rate of home runs among the balls hit in the zone. To focus on this pattern, we

divided the red zone region into $5 \times 5 = 25$ bins by launch angle and exit velocity.

For each bin in the home run zone of values of (Launch Angle, Exit Velocity) Figure 19 shows the count of batted balls and the number of home runs for the 2016 season. Figure 20 displays similar fractions for the 2017 season. For example, in Figure 19 in the region Launch Angle in (15, 20) and Exit Velocity in (110, 115) (top left portion of the figure) there were 184 batted balls and 60 were home runs.

HR / N for Period y16

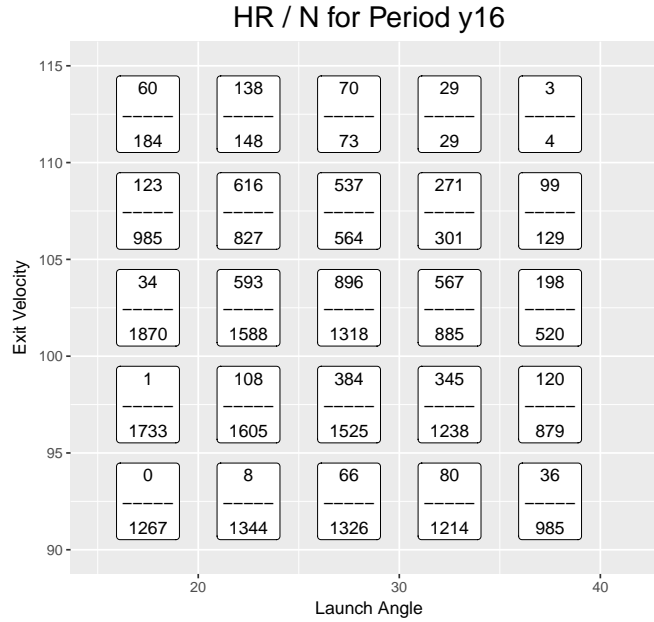| Exit Velocity	20				30				40
115	60 / 184	138 / 148	70 / 73	29 / 29	3 / 4				
110	123 / 985	616 / 827	537 / 564	271 / 301	99 / 129				
105	34 / 1870	593 / 1588	896 / 1318	567 / 885	198 / 520				
100	1 / 1733	108 / 1605	384 / 1525	345 / 1238	120 / 879				
95	0 / 1267	8 / 1344	66 / 1326	80 / 1214	36 / 985				
90									

Launch Angle

Figure 19: Count of home runs and batted balls in different launch angle/exit velocity bins for the 2016 seasons.

To understand the difference between home run hitting in the 2016 and 2017 seasons, look at the third row of Figure 19 and Figure 20 corresponding to an exit velocity between 100 and 105 mph and different launch angles. In the 2016 and 2017 seasons, the number of balls hit in this zone was approximately the same for all launch angles. But the home run counts in these two seasons were substantially different. For example, in the 2017 season, when the launch angle was between 25 and 30 degrees and the exit velocity was between 100 and 105 mph, there were 982 home runs hit (in 1319 batted
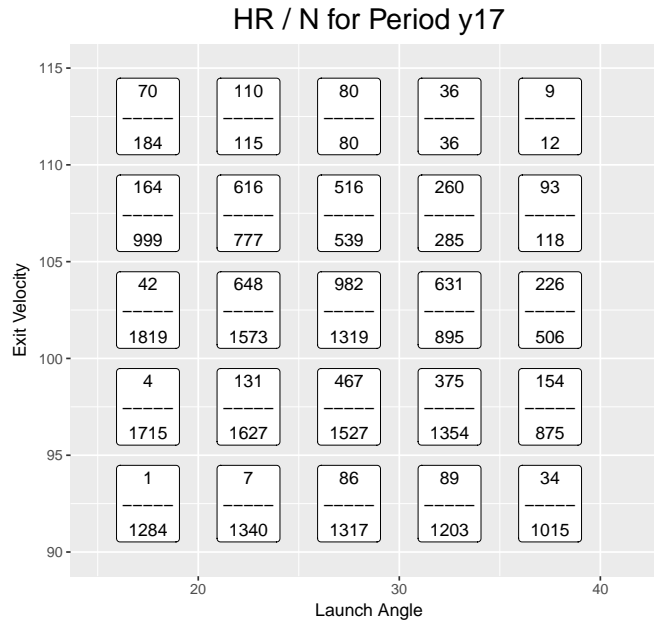
HR / N for Period y17

Exit Velocity \ Launch Angle	~15	~20	~27	~33	~40
~113	70 / 184	110 / 115	80 / 80	36 / 36	9 / 12
~108	164 / 999	616 / 777	516 / 539	260 / 285	93 / 118
~103	42 / 1819	648 / 1573	982 / 1319	631 / 895	226 / 506
~97	4 / 1715	131 / 1627	467 / 1527	375 / 1354	154 / 875
~92	1 / 1284	7 / 1340	86 / 1317	89 / 1203	34 / 1015

Figure 20: Count of home runs and batted balls in different launch angle/exit velocity bins for the 2017 seasons.

balls). In the 2016 season, there were 896 home runs hit (in 1318 batted balls) in the same region.

This point is reinforced in Figure 21 and Figure 22 that display the home run percentages for each of the 25 regions. Again if we focus on the region where the exit velocity is between 100 and 105 mph, the home run percentages in the 2016 season were 1.8, 37.3, 68, 64.1, and 38.1 – for the same launch angles, the home run percentages in 2017 were 2.3, 41.2, 74.5, 70.5, 44.7.
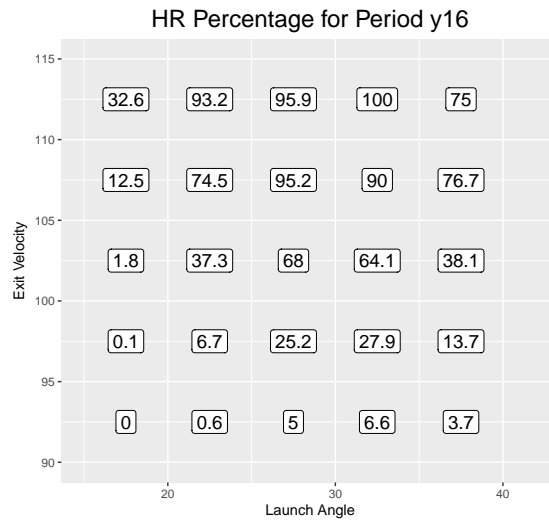
Figure 21: Percentage of batted balls that are home runs in the 2016 season for different launch angle/exit velocity bins.
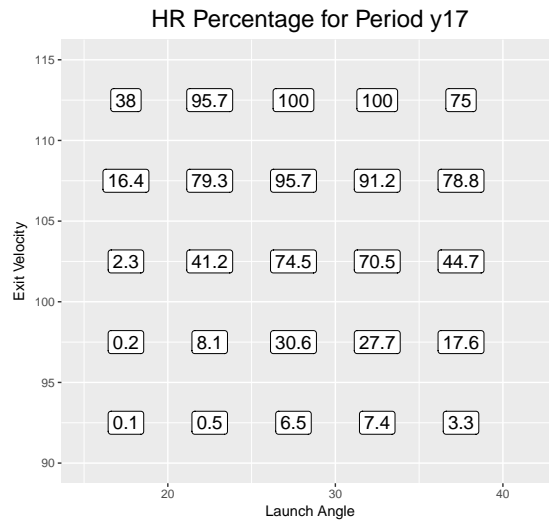

Figure 22: Percentage of batted balls that are home runs in the 2017 season for different launch angle/exit velocity bins.

2.6.2 Increase in home run probability from 2016 to 2017

To better understand the results in Section 2.6.1, a generalized additive model was fit to each of the 2016 and 2017 seasons, where the probability of a home run is represented as a smooth function of the launch angle, exit velocity, and spray angle. By comparing the home run probabilities for the two seasons, one can discover the values of the three variables giving the greatest increase in home run probability from 2016 to 2017. For each of the spray angles of -30, -20, -10, and 0 degrees, Figure 23 displays a spatial map of the increase in home probability from 2016 to 2017. For spray angles of 0 and -10 degrees, there is a region of launch angles between 25 and 35 degrees and exit velocities around 100 mph where there is a 0.10 increase in home run probability and the 2017 home run probability is greater than the 2016 probability for practically everywhere in the total region. For larger spray angles (-20 and -30), the greatest increase in home probability appears for smaller exit velocities and large launch angles, and for large exit velocities and smaller launch angles.

2.6.3 Additional home runs in the zone from 2016 to 2017

The increased home run percentages from 2016 to 2017 for batted balls in the red zone translate to additional home runs hit in the 2017 season. Figure 24 shows the additional number of home runs hit in 2017 compared to 2016 for each region. There is a rectangle around the region of launch angle between 20 and 40 degrees and exit velocity between 95 and 105 mph where 409 additional home runs were hit – this total represents 90% of the 449 additional home runs hit during 2017 in this red region.

2.6.4 Probability of home run given launch angle, exit velocity, and spray angle

Since the probability of a home run is also dependent on the spray angle, a generalized additive model was used to predict the occurrence of a home run given the spray angle, launch angle, and exit velocity. Figure 25 displays curves corresponding to a predicted probability of 0.90 of hitting a home run for the spray angles for 0, -10, -20, and -30 degrees. In each figure, the blue line corresponds to a 90 percent predicted probability using the 2016 data, and the red line to the 90 percent predicted probability using the 2017 data. Generally, we see that the location of the 90 percent contour line is
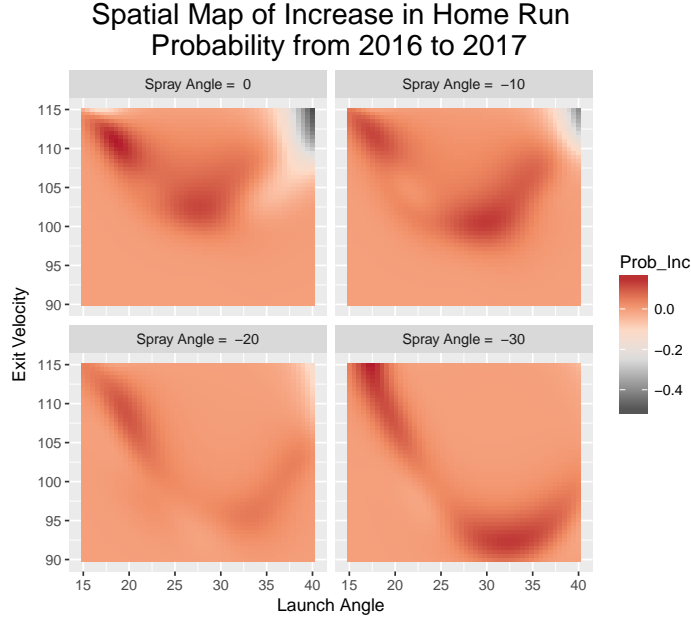
Figure 23: Spatial map of the increase in home run probability from 2016 to 2017 as a function of launch angle and exit velocity for four values of the spray angle.

lower for the 2017 season, especially for launch angles about 30 degrees. This figure confirms the earlier evidence that the home run probability in the red zone has significantly increased for the 2017 season.

2.7   Home Run Rates

To summarize our findings, Table 3 displays the number of home runs, number of batted balls, and the rate (as a percentage) for the four periods. We see the clear increase in home run rates per batted ball from a low value of 3.9 percent in the first half of the 2015 season to a value of 5.3 percent for the 2017 season.

Suppose we assume that the batted balls in 2015a, 2015b, and 2016 are hit with the 2017 home run fitted probabilities. That is, for each of the three periods, we sum the predicted probabilities (using the model fit on the 2017 data) over all batted balls for these earlier seasons using the corresponding
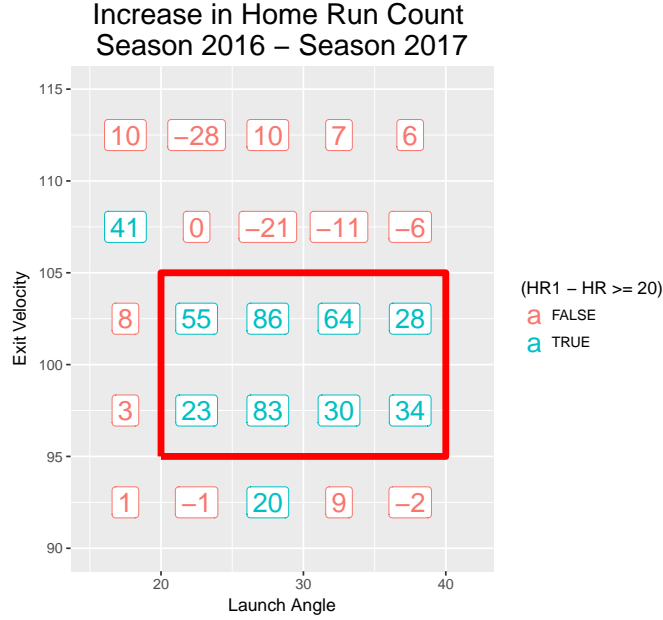
Figure 24: Increase in the count of home runs from 2016 to 2017 in different launch angle/exit velocity bins.

launch angles, exit velocities, and spray angles.

In Table 3, the new column Expected gives the expected number of home runs using the 2017 probabilities. The Adj. Rate variable divides this expected number by the number of batted balls. Figure 26 displays the unadjusted and adjusted home run rates for the four periods. Note that, with this adjustment, 2015b, 2016, and 2017 have similar home run rates. This indicates that the change we see in the home run rate from the second half of the 2015 season through 2017 is primarily due to the change in home run probabilities (for given values of spray angle, launch angle, and exit velocity).

The important takeaway from this analysis is the following: Had the drag properties of the baseballs in 2015b and 2016 been identical to those in 2017, the home run rate would have been the same in all three periods. This result lends further support to the conclusion that the the reason for the increase in home runs 2015b-2017 is due to reduced drag on the baseballs. However, the 2015a season does not fit into the same pattern, given that the adjusted home run rate for that period is 30% below that in the other periods. We
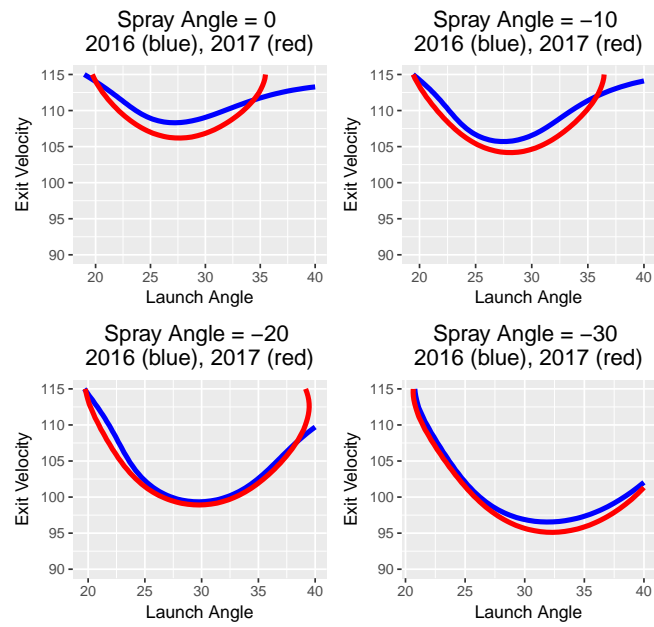
Figure 25: 90 percent quantiles of home run probabilities for 2016 and 2017 seasons for different values of the spray angle.

next examine that situation in more detail.

Table 3: Rate of home runs per batted balls for four periods. The last two columns contain the expected count of home runs and the home run rate adjusted for the 2017 home run probabilities.

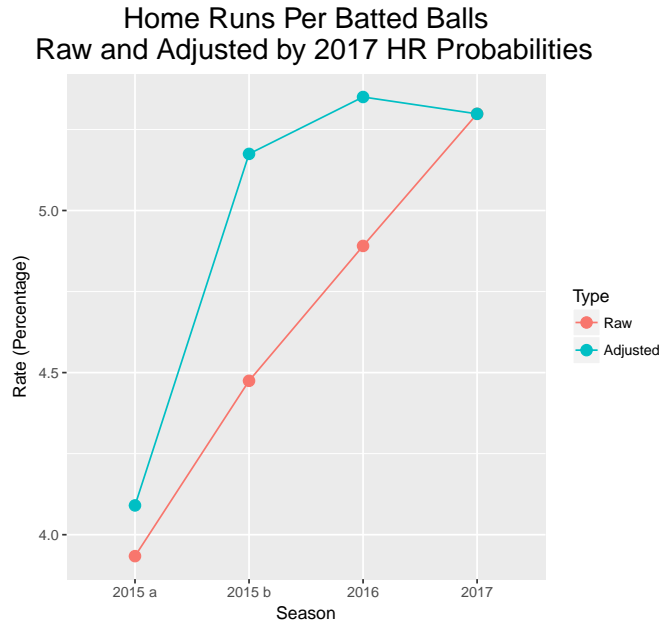| Season	Batted Balls	Observed Home Runs	Home Run Rate	Expected Home Runs	Adjusted HR Rate
2015a	53713	2113	3.9	2197	4.1
2015b	58574	2621	4.5	3031	5.2
2016	112744	5514	4.9	6032	5.4
2017	112850	5979	5.3	5979	5.3

Figure 26: Rates of home runs per batted balls and rates adjusted by the 2017 home run probabilities.

2.8 The Anomalous 2015a Data

As discussed above in the context of Table 3 and Figure 26, the increase in home run rates starting with the 2015b period and extending to the 2016 and 2017 seasons is primarily due to reduced drag on the baseballs, leading

to better carry. However, the 2015a period does not follow the same pattern. We examine this anomaly more fully in this section.

To gain some insight into the reason for the 2015a-2015b increase in home runs, refer to top half of Figure 27, which examines the distribution of exit velocities for the launch angle ranges 25°-30° (left) and 0°-10° (right). The blue curves are from 2015a while the red curves are from the combined 2015b-2016 periods. The data for 25°-30°, the launch angles with the highest home run probabilities, show a significant shift by around 1.5-2.0 mph. Analysis shows that a shift by that amount would be sufficient to account for the 2015a-2015b increase in home runs, a result that has been discussed in the on-line literature. It has been suggested by some that such an increase is evidence for a "juiced ball", meaning an elevated COR or CCOR. However, if such were the case, one would expect a similar increase in exit velocity for line drives (i.e., balls hit with launch angles in the range 0°-10°). No such increase is observed.

To study this anomaly further, we have obtained camera-based HITf/x data from the 2015 and 2016 seasons. Note that that system was discontinuted after the 2016 season, so no data exist for 2017. The bottom half of Figure 27 examines the distributions of exit velocities from HITf/x for the same two ranges of launch angles. No significant shift is seen for either angular range.

We next inquire whether the 2015a-2015b and 2015b-2016 increases in home run rate, as determined exclusively from HITf/x data, are consistent with the hypothesis that reduced drag is primarily responsible for the home run surge. Table 4 shows the same type of analysis as was shown in Table 3, using 2016 home run probabilities to calculate adjusted rates for 2015a and 2015b. While not perfect and keeping in mind that there were many more missing data in the HITf/x set, we find the adjusted rate for 2015a much closer to those of 2015b and 2016, suggesting that reduced drag may explain the 2015a-2015b increase in the home run rate.

We conclude that the 2015a StatCast data are indeed anomalous and not supported by HITf/x data. Perhaps that is not surprising, as 2015 was the first season for the StatCast system, which underwent many adjustments and recalibrations as the season progressed.
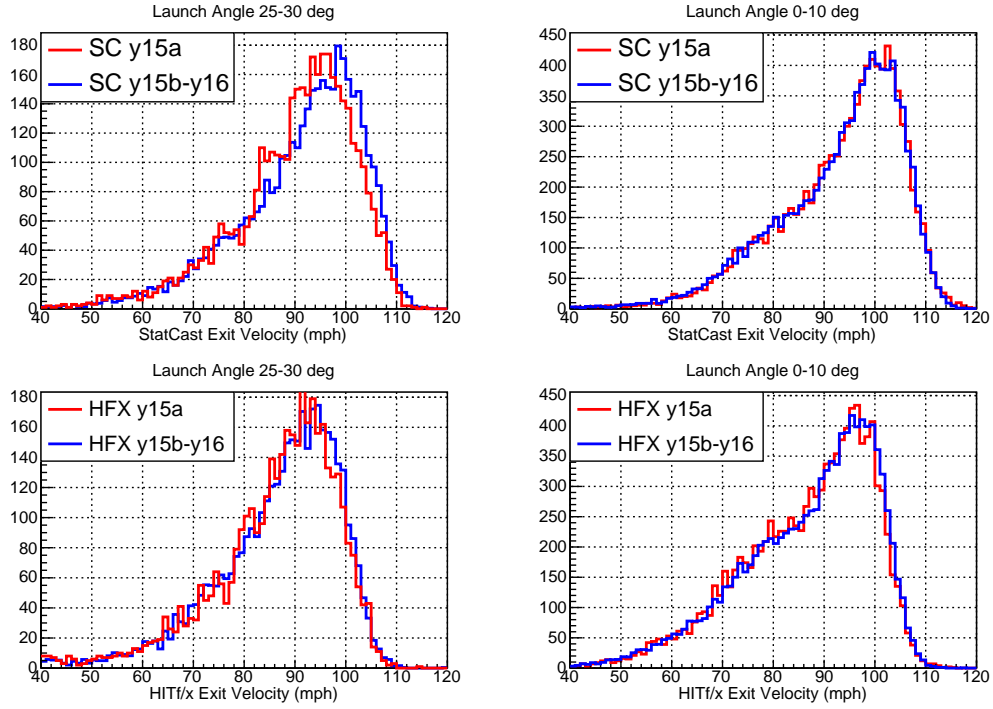
Figure 27: Distribution of StatCast (top, indicated by SC) and HITf/x (bottom, indicated by HFX) exit speeds for 2015a (blue) and 2015b-2016 (red). The left plots are for the launch angle range 25°-30°, the right for 0°-10°. To facilitate comparisons between periods, the red and blue plots are normalized to the same number of batted-ball events. The significant shift in exit speeds shown in the SC plots for 25°-30° is not apparent in the other plots.

Table 4: Rate of home runs per batted balls for three periods, all based on HITf/x data. The last two columns contain the expected count of home runs and the home run rate adjusted for the 2016 home run probabilities.

Season	Batted Balls	Observed Home Runs	Home Run Rate	Expected Home Runs	Adjusted HR Rate
2015a	38184	1617	4.2	1800	4.7
2015b	41575	1946	4.7	2019	4.9
2016	72300	3710	5.1	3721	5.2

2.9   Temperature Effects

The analysis thus far has shown the the primary reason for the increase in home run rate starting with the 2015b period is that the ball carries to longer distances for given initial conditions. One factor that can lead to longer distances is higher temperatures, which result in lower air density and therefore reduced drag. We explore that possibility in Figure 28, which displays a scatterplot of the home run rate against the game-time temperature for open-air stadiums for each of the seasons 2015, 2016, and 2017. Smoothing curves are added to the graph to show the general pattern. We see from Figure 28 that the rate of home runs generally increases for warmer temperatures, as expected, but the rate of increase appears to be the same for each of the seasons. Moreover, the roughly parallel curves are offset from each other, showing that the differences in home run rate persist even at fixed values of the temperature. A similar result is found by examining home run rates in closed stadiums, where the temperature is constant. We conclude that the change in home run rates across periods cannot be explained by changes in temperature.
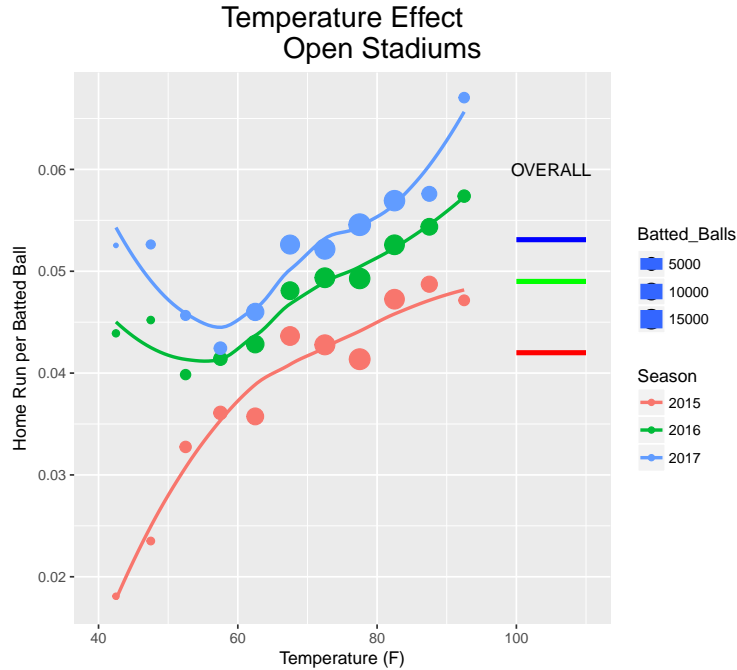
Figure 28: Scatterplot of home run rate against temperature for the seasons 2015, 2016, and 2017. For temperatures between 60 and 85 degrees, the rate of home run hitting is approximately a linear function of temperature, with a slope about the same for each of the three seasons. Moreover, the curves are offset from each other by an amount approximately the same as the differences in overall home run rate among the three seasons.

3   Analysis of Baseball Properties

The goal of this analysis is to explore whether certain properties of the baseball that would lead to an increase in home run production have changed in recent years. We start with a summary of our principal findings.

3.1   Summary of Principal Findings

1. We found no evidence in our visit to the Rawlings facility that ball production practices could be contributing to the recent increase observed in MLB. In particular, we find no compelling evidence that changes in

the properties of the baseball that are currently tested and validated (e.g., weight, size, COR, CCOR) have played a major role in the home run surge.

2. For some of the tested parameters, such as weight, size, and COR, Rawlings achieves much greater precision than allowed by the MLB specifications, leading to our recommendation in the Executive Summary that these specifications be tightened.

3. Laboratory data show a strong dependence of the COR on the relative humidity where the balls have been stored, leading to our recommendation in the Executive Summary that the storage environments used by the clubs should be monitored by MLB.

4. The findings in Sec. 2.1 that the aerodynamic properties of the ball (primarily the drag and lift coefficients) are a major contributing factor to the home run surge receive additional support from laboratory measurements of those properties and from analysis of StatCast trajectories. This finding is further confirmed by a physics-based model for the flight of the baseball, which shows that small changes in the aerodynamic properties of the baseball that are comparable to the measured changes can account for the observed increase in home run production over the periods studied.

5. We find that the ball-to-ball variation in the drag and lift coefficients is large compared to the size of changes of average values that would lead to the home run surge. Various ideas are being pursued to understand better the source of the large variation.

6. The investigations have not succeeded in finding specific properties of the baseball leading to better carry. In particular, the improved carry is not correlated with changes in the size, weight, or seam height of the baseball. Further testing to isolate the reasons for the reduced drag are ongoing.

7. Since changes in the aerodynamic properties of the baseball (primarily the drag and lift coefficients) have been shown to contribute to the home run surge and since these parameters are not currently measured, we recommend in the Executive Summary that methods be developed to measure and monitor these parameters.

3.2 Manufacturing Considerations

Several members of the committee (Lorden, Nathan, and Smith) visited the Rawlings MLB ball manufacturing facility in Turrialba, Costa Rica on December 7, 2017. Prior to the tour we met with engineers from the Turrialba facility and St. Louis. We reviewed the schedule for the plant tour and discussed quality control processes. The following is a summary of the visit and processes used to make MLB baseballs.

Figure 29: Cross sections of the components comprising the MLB baseball.

3.2.1 Pill

The pill forms the center of the ball and is made by Hultec in San Jose, Costa Rica. The rubber pills are coated in a tacky adhesive prior to being wound over by wool yarn.

Figure 30: Pills prepped for winding.

3.2.2 Yarn

The yarn is 85% wool and 15% synthetic. The yarn is conditioned for 24 hours prior to winding into balls. The wool is wound in three separate layers, where the circumference and weight of the ball after each layer is measured to be in spec before the next layer is wound. The first layer is 4-ply grey wool, the second layer is 3-ply white wool, and the third layer is 3-ply gray wool. The winding is finished with a thin layer of cotton, which is given a tacky coating prior to stitching the cover.

Figure 31: Wool in conditioning prior to winding.

Figure 32: Rounds prepped for stitching cover.

3.2.3   Leather Cover

The leather cover is tanned from dairy cows in Tennessee. Panels are cut from top grain leather to avoid areas with visual blemishes and wrinkling. The panels are weighed and skived to meet thickness and/or weight requirements. Prior to sewing, the panels are soaked in cloth of controlled moisture content for 20 minutes.

Figure 33: Cutting of leather panels.

Figure 34: Leather moistening prior to stitching.

3.2.4 Sewing

The 108 stitches of the covers are hand sewn by 300 sewers. They make up the highest skill level of workers in the facility. They have a formal training program to ensure uniformity of the ball covers. Workers have required breaks and a 1 hour lunch break to minimize fatigue effects. They are given minimum and maximum quotas. Sewers can stitch at most 200 balls per week, after which they may leave with full salary. The balls are rolled between grooved wooden platens after stitching (when the balls are moist) and again 24 hours later, to flatten the seams and maintain a spherical shape.

Figure 35: Manual sewing of leather covers.

Figure 36: Ball rolling to flatten seam.

3.2.5 Inspection

After the balls are sewn, the stitches and covers are inspected separately. Production return rates are tracked to encourage quality and quantity. Prior to packaging each ball is individually inspected and weighed. Only 55% of the balls produced meet inspection requirements for use in MLB games.

Figure 37: Ball inspection after stitching.

3.2.6 Testing

The facility includes testing equipment for measuring ball compression, coefficient of restitution (flat plate and cylindrical), wool yarn content, ball durability, and color. Multiple inspections are conducted on each ball throughout the manufacturing process. Prior to labeling, each ball is cleaned, weighed and measured for circumference.

3.2.7 Changes to Manufacturing Process

As part of the committee investigations, we have requested and Rawlings has supplied a list of all changes to the baseball manufacturing process since the start of 2014. These changes, listed in Table 5, were made with the prior

approval of MLB. They are largely technical in nature and very unlikely to be in any way related to the home run increase.

As an example, consider the change to the Pill Mold in May 2015, necessitated by wear on the previous mold. The latter had resulted in the finished diameter of the pill slightly smaller (1.366 inches) than the target diameter (1.375 inches). The new mold restored the pill to its target diameter, keeping the weight unchanged. There is no evidence that this change to the pill affected the size, weight, COR, or CCOR of the finished baseball, as discussed in subsequent sections of this report.

Table 5: Changes in the baseball manufacturing process since 2014.

Change	Date	Reason for Change
Yarn Double Blending	February 2014	Process Improvement
Pill Ring	March 2014	Process Improvement
Leather Dampening	June 2014	Process Improvement
Pill Mold	May 2015	Maintenance
Toggle Dryer	March 2016	Maintenance
Eliminate Wet Shaving	February 2017	Process Improvement
Leather Cutting Die	August 2017	Maintenance
Set-Out Machine	August 2017	Mainentance
Crusting Dryer	February 2018	Maintenance

3.2.8   Summary

The MLB baseball manufacturing facility in Costa Rica is labor intensive and showed evidence of systems producing a reproducible high quality product. While many of the processes could be automated, it is unlikely automation would be an improvement over existing processes. Since some processes cannot be automated (stitching) locating production in a region with lower labor costs is a justifiable business decision. The employees appear to be well trained, proficient in their assigned tasks, and diligent in record keeping and specification verification. The raw materials are inspected, monitored and conditioned prior to ball production. We found no evidence in our visit that ball production practices could be contributing to the recent increase observed in MLB.

3.3 Quality Assurance Testing by Rawlings

Rawlings does extensive, weekly QA/QC testing on baseballs shipped to MLB. The committee was provided with four time/date matched values from that Rawlings QA testing, including weight, circumference, COR, and compression, the latter being a measure of the "stiffness" of the ball. We investigated these features to address the following:

- Have any features changed over the time period provided (2010-2017)?

- Are the changes correlated with the change in home run rate during those years?

Regardless of whether the ball provided by Rawlings falls within the specified range for any given parameter, it is possible that there would be yearly shifts in those parameters. That is, because the specified range in each parameter value is quite wide, there is still room to observe possible systematic changes in those values. The yearly distributions of the four parameters is shown in Fig. 38.
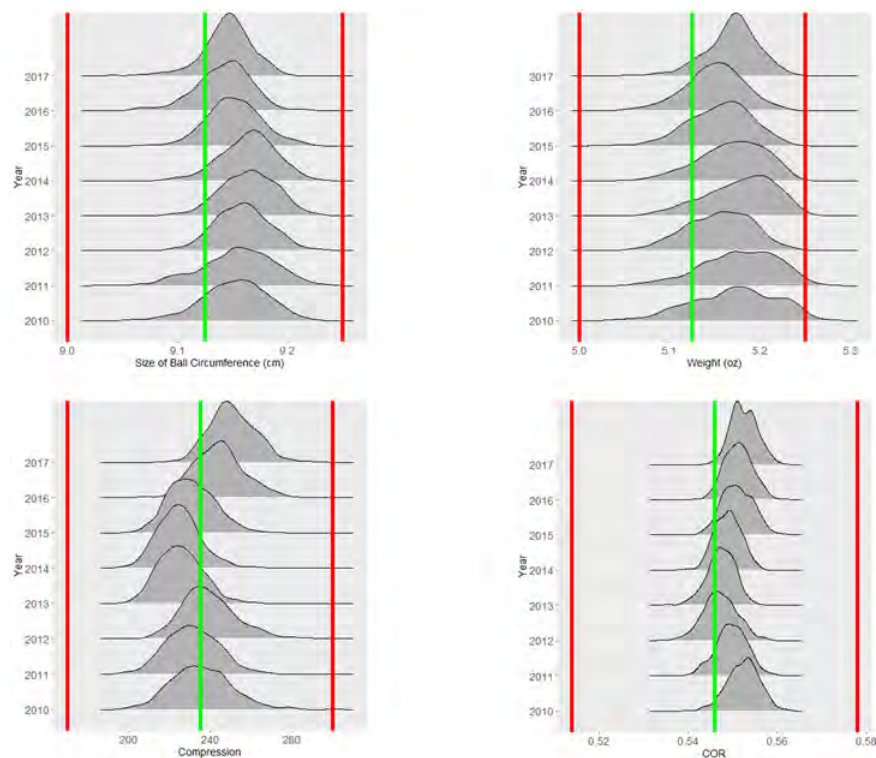
Figure 38: Yearly distributions of size, weight, COR, and compression, as measured by Rawlings, for the years 2010-2017. The red lines show the upper and lower bounds allowed by the MLB specifications and the green lines show the target value.

These plots show that there is substantial year-to-year variation in each of the features. For example, COR has a left shift (i.e., toward lower values) through 2012-2013, followed by a right shift through 2017. Size follows nearly the opposite trajectory. Compression shifts right in 2016 and 2017. These yearly changes are apparent not only from the plots but also when subjected to rigorous statistical testing. So, there appear to be rather clear year-over-year changes that do not match a "static production" model, suggesting that although the ball is consistently within specifications, there are drifts within the bounds of those specifications over time.

The committee also investigated the possibility that changes in the ball might happen at timepoints other than yearly, to see if any discrete produc-

tion runs could be identified. While possible epochs could be identified, as suggested by the graphs above, there was no clear discrete shift that occurred in a physical parameter during 2015-2017 ball production that would account for the home run surge. In fact, the clearest outlier timeframe identified by such analysis is the 2013-2014 ball, which is characterized by relatively lower compression and COR and relatively higher size and weight, though there is likely no meaningfully large change in any of those features observable (see below).

Several of these features have strong year-over-year correlations with the home run per batted ball rate. The strongest of these correlations is compression. However, as detailed elsewhere in the report, although these changes correlate with the change in home runs, there is no evidence that the physical parameters measured by the ball would produce changes either in the way suggested by this report or on the order of the changes observed.

To the first point, analysis by other sections of this report strongly implicate carry effects, rather than initial launch condition, as being changed during the period characterized by the home run surge. Changes in COR or compression would almost certainly impact those initial launch conditions but not the carry.

To the second point, as also detailed elsewhere in the report, the range of changes in each of these parameters is relatively small on the practical scale, and the mere correlation of changes between physical parameters and the home run rate is not enough to implicate any particular ball change that we observe in the Rawlings data. So, while compression might increase directionally with the change in the home run rate over some period, the amount of change in compression is extremely small relative to the physically modeled impact on home run production of that amount of compression change.

Finally, much has been written in the media about the ball specifications being so large as to render meaningless claims that "the ball is being made within specs." For example, with the current COR specifications, a ball at the upper limit (0.578) would have nearly 36 more feet of projected distance than a ball at the low limit (0.514). Clearly, this standard is unreasonably large, given that Rawlings achieves far greater precision in COR than is communicated by that standard. For this reason, we have recommended in the Executive Summary that MLB tighten this and other parameters of the baseball that affect the game.

3.4 Baseball Testing at Washington State

3.4.1 Introduction

Many have speculated that the home run surge observed in Major League Baseball since 2015 is due to changes in the ball. The following presents experimental tests of the Rawlings MLB baseball to quantify its performance and determine how measured properties of the ball compare before and after the 2015 All Star Game. These tests were all done at the Sports Sciences Laboratory (SSL) at WSU.

A challenge in comparing baseballs over multiple years is obtaining balls from prior years. Balls were taken from two sources for this study. First were balls supplied to UML by MLB for testing three times each year, at the beginning, the middle, and near the end of each season . Each batch consists of eight dozen balls, of which four dozen are tested for compliance purposes while the remainder are set aside for future testing. One dozen of the set-aside balls from each batch were provided for this study, for a total of 36 balls for each of the 2013-2017 seasons. In the following, we refer to these as the "new" balls.

Second were so-called "authentication" balls provided by MLB. These are balls that have been used in actual MLB games and removed from play after one or two hits. MLB has developed an authentication program where balls taken from play are given a record of the game they were used and subsequently made available for sale. Twenty-two dozen balls from the 2012 through 2017 seasons were taken from the authentication program for use in this study. The balls were selected to provide as even a distribution as possible for each year. Unfortunately only two authentication balls from the 2015 season were available. The yearly distribution of both the new and authentication balls are listed in Table 6.

Table 6: Quantities of balls from each year that were tested at WSU.

Year	new	authentication
2012	-	50
2013	36	53
2014	36	27
2015	36	2
2016	36	34
2017	36	98
Total	15 dz	22 dz

In this and several subsequent plots, the error bar represents the standard deviation of each yearly sample. The number next to the arrow is an estimate of the change in fly ball distance that would result from a change in the plotted parameter equal to the length of the arrow, assuming an exit speed of 100 mph, a launch angle of 30°, and 2500 rpm of backspin.

3.4.2 Size and weight

We first measured the size and weight of each ball, with the results given in Figs. 39 and 40. While the authentication balls tended to be lighter than the new balls, the difference was small. The difference in circumference between the UML and SSL balls was also small and is likely due to measurement technique. At SSL ball diameter was measured with calipers, while circumference was measured at UML with a pi tape. The annual trends of weight and circumference do not correlate with the recent MLB home run increase.
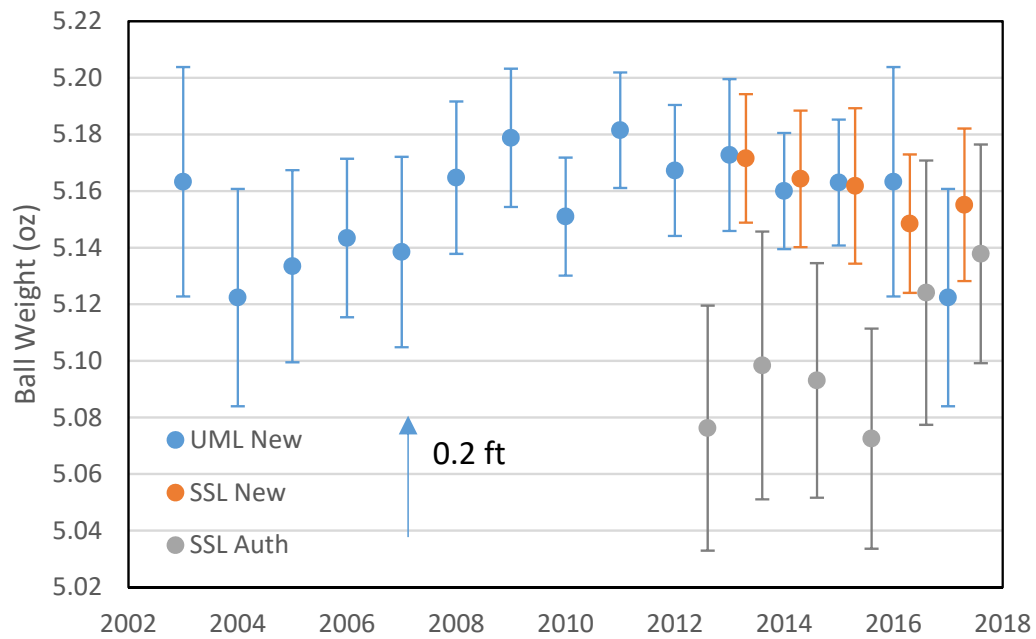
Figure 39: Average weight of the 15 dozen new balls and the 22 dozen authentication balls. In this and subsequent plots, UML refers to measurements done at UMass/Lowell in their compliance testing, while SSL refers to the new measurements done specifically for this study.
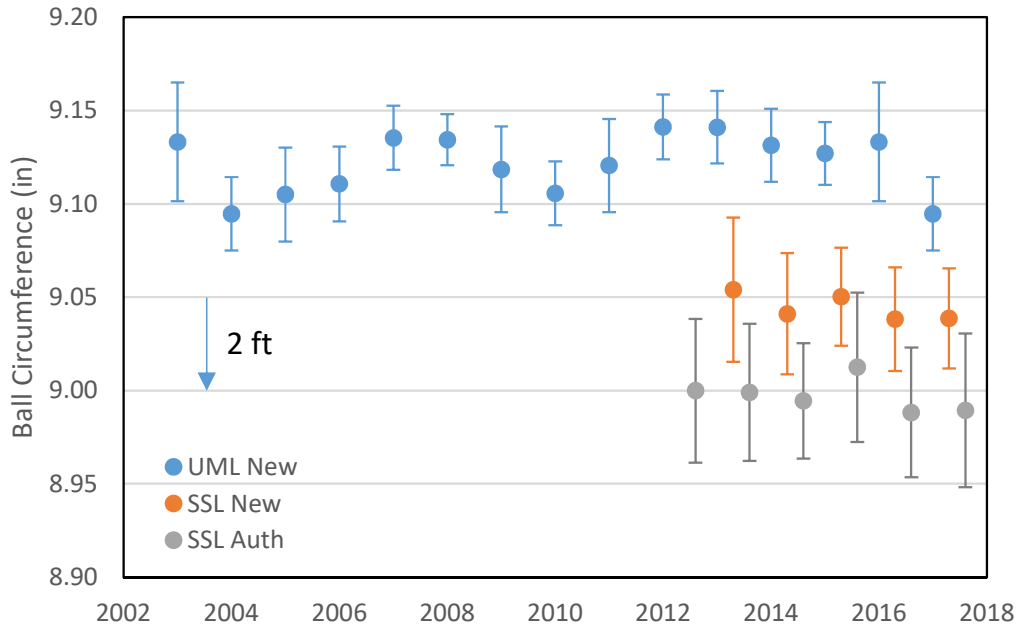
Figure 40: Average circumference of balls.

3.4.3 Impact Testing

The recent MLB home run increase is anecdotally said to be due to a "juiced ball." Since a large fraction of the impact energy is lost during ball-bat collision, small changes in energy dissipation can have a large effect on ball performance. The parameter most often associated with a "juiced ball" is it's coefficient of restitution. This can be measured by impacting a rigid flat surface at 60 mph and is commonly denoted COR. The COR speed and impact surface induce less ball deformation than occurs in play. Increasing the ball speed to 120 mph and impacting a rigid cylinder (Fig. 41) results in ball deformation much closer to play conditions and is denoted CCOR. The relationship between COR/CCOR and exit velocity is given in Appendix B.

Figure 41: Image of a softball impacting a rigid cylinder to measure its cylindrical coefficient of restitution or CCOR.

The COR and CCOR of new and authentication balls is presented in Figs. 42 and 43, respectively. While the hit distance associated with COR changes (Fig. 42) are larger than for weight and circumference, they are not substantiated by the more representative CCOR results of Fig. 43. The CCOR results of new balls from UML and SSL are in good agreement, and suggest that the MLB home run surge is not due to a "juiced ball".
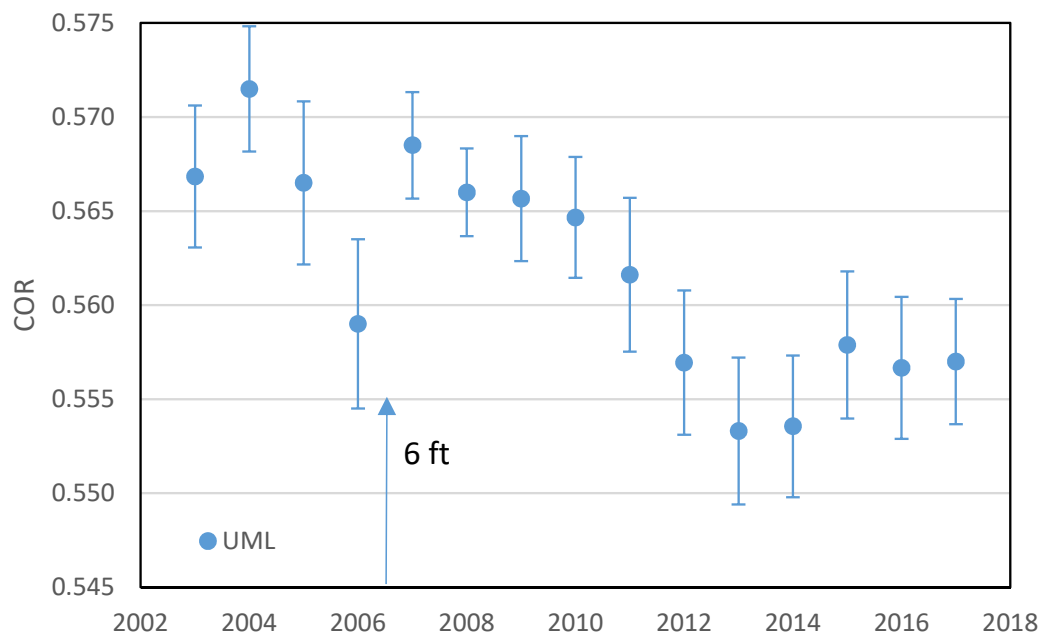
Figure 42: Coefficient of restitution of new MLB balls, 6 dozen per year, as measured at UML.
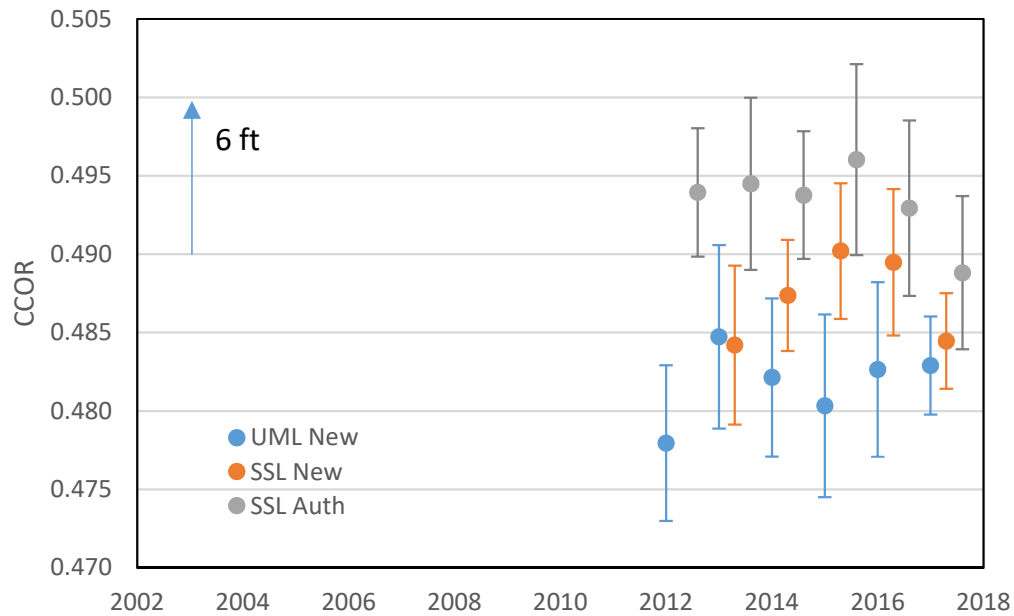
Figure 43: Cylindrical coefficient of restitution of 7 dozen authentication MLB balls.

3.4.4 Effect of Humidity

The ball coefficient of restitution is not constant, and can depend on many factors. An important factor affecting ball COR is the relative humidity (RH). Figure 44 illustrates the dependence of ball COR on the relative humidity. In extreme cases (going from 100% to 10% RH), the flight distance can increase by 56 ft.
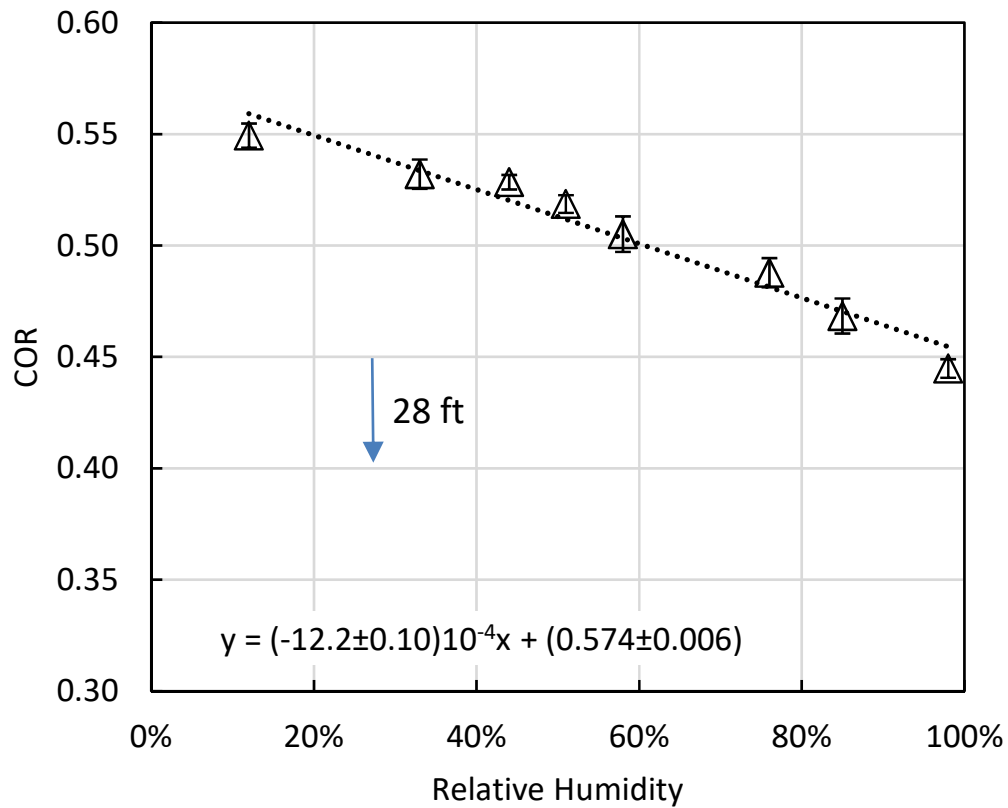
Figure 44: MLB baseball coefficient of restitution as a function of relative humidity, from American Journal of Physics, vol. 70, pp 575-580, 2011.

The sensitivity of the MLB baseball to humidity should also be considered in light of the April 1, 2015 letter from MLB to the clubs. The clubs were instructed to store balls at 70F and 50% RH. While uniform ball storage among the clubs can help reduce field to field variation, it can also increase ball performance if the majority of clubs had previously stored their balls at more than 50% RH. This very large effect is the basis of our recommendation in the Executive Summary that the storage environment employed by MLB clubs be monitored.

3.4.5   Aerodynamic Testing

Aerodynamic performance is often described by a drag coefficient (how a ball slows along its path) and a lift coefficient (how it moves normal to its path). Lift and drag are important for the flight of a baseball, but are rarely measured. The irregular shape of the baseball results in relatively large variation in lift and drag, complicating the study of aerodynamic effects. Wind tunnels can measure lift and drag, but require attachment to the ball to keep it stationary in the airflow. Attachment schemes can interrupt the airflow and its effects on lift and drag, and constrain the ball from its response traveling freely through air. Additional details about properties of a baseball that affect its carry distance are given in Appendix B.

In the following, lift and drag were found by measuring the change in speed and flight path of balls projected through still air. Pitching machines with spinning wheels are often used to project balls with controlled spin. The wheels can roughen the ball surface, however, and alter its aerodynamic response. In this work balls were projected using a linear accelerator, as shown in Fig. 45. Spin was induced by supporting the ball between parallel clamps as it was accelerated. The clamps had opposing high and low friction surfaces, which induced spin as the piston decelerated and released the ball.
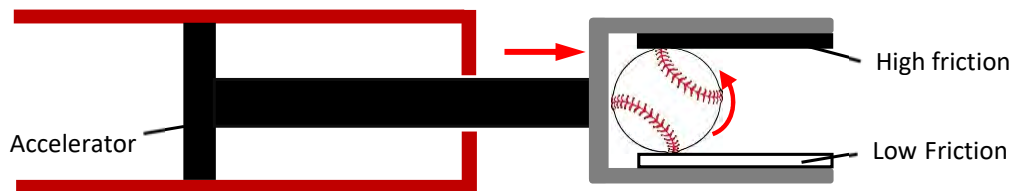

Figure 45: Linear accelerator used to project baseballs with spin.

The balls were projected through three successive speed sensors as shown in Figs. 46 and 47. Drag was found from the change in speed between sensors 2 and 3 (12.5 ft apart). The ball location was recorded at the three stations (over 17.5 ft). Lift was found from the changes in ball path, after subtracting the effect of gravity.
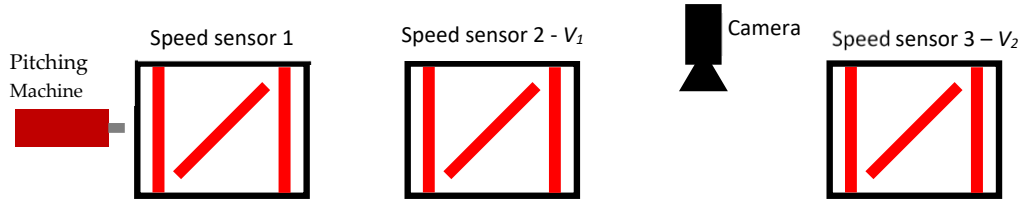
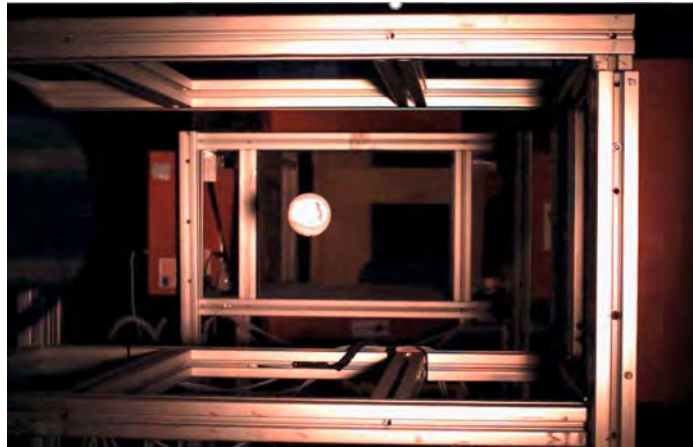Figure 46: Setup used to measure lift and drag of baseballs traveling through still air.

Figure 47: Image of ball traveling through the light sensors to measure its lift and drag.

3.4.6 Aerodynamic Results

Home runs are often hit with exit velocities close to 100 mph and launch angles around 30°. Balls hit this way will slow to near 40 mph at their apex, before accelerating under gravity to near 55 mph as they approach the ground. Both drag and lift are sensitive to ball speed. In comparison to amateur baseballs the MLB baseball is relatively smooth, resulting in a so-called "drag crisis", where the drag decreases with increasing speed. Lift is primarily sensitive to the ratio of spin to linear speed. Since the ball changes

more in linear speed than spin, the effect of lift also changes over the ball's flight path.

To account for the dependence of lift and drag on speed, both parameters were measured at 55 and 80 mph, describing the low speed and high speed regimes of the ball's flight path. Drag and lift are presented in Figs. 48 and 49, respectively. Note that drag has a larger effect on hit distance than comparable changes in lift, and the balls show decreased drag after 2015 at 55 and 80 mph for both the authentication and new balls. Systematic changes in lift are not as apparent, and contribute less to hit distance given its lower sensitivity. The authentication balls had higher drag than the new balls at both 55 mph (0.011) and 80 mph (0.026), which may be a result of the mud that is rubbed on balls prior to play.
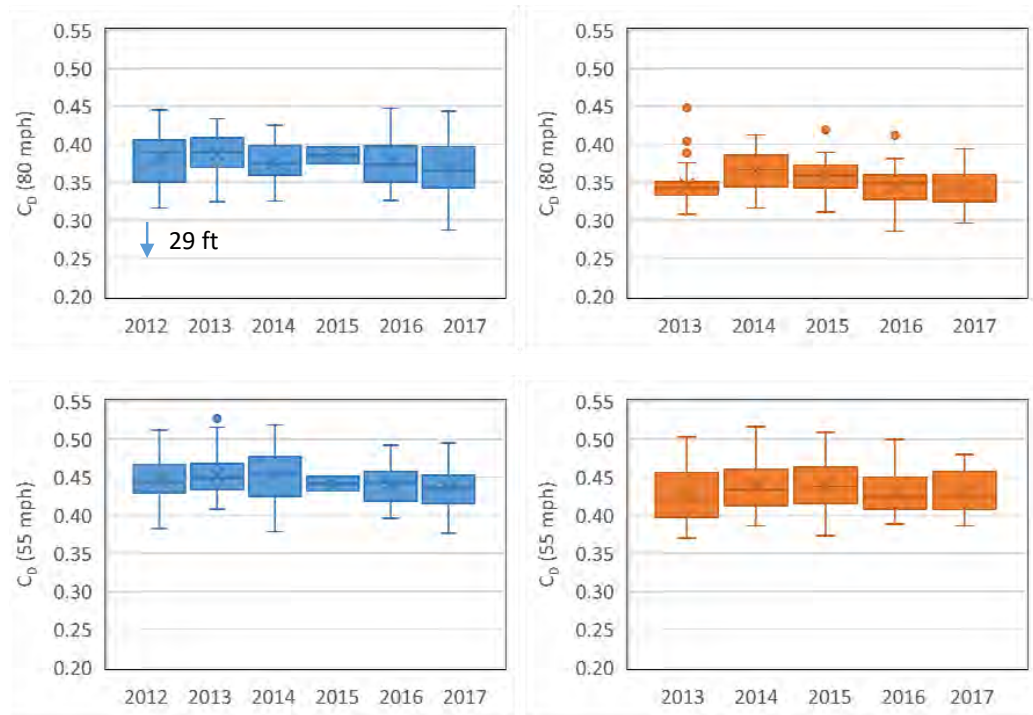

Figure 48: Drag of authentication (blue) and new (orange) balls, projected through still air. Note that only two balls are in the authentication sample for 2015.
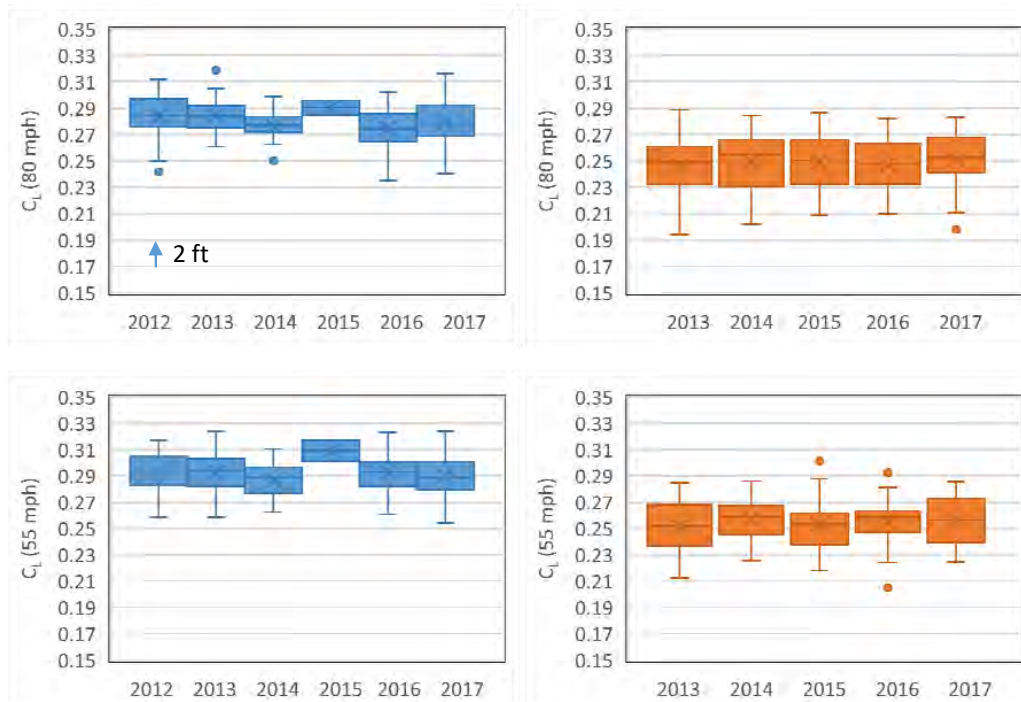
Figure 49: Lift of authentication (blue) and new (orange) balls, projected still air. Note that only two balls are in the authentication sample for 2015.

Each of the parameters measured (size, weight, CCOR, C_d, and C_l) has an effect on the distance of a fly ball. By taking the yearly averages of these quantities as input to models of both the ball-bat collision and the trajectory, one can estimate the distance for an "ideal" batted ball for each year. The ideal hit was given initial conditions of 100 mph exit velocity, 30° launch angle, and 2500 rpm backspin, which travels a nominal distance of around 390 ft. The results are presented in Fig. 50, where it is clear that for every year except 2015 (for which only two balls were tested), the primary factor changing the distance is the drag. Using 2014 as a reference, Fig. 50 suggests that changes in the ball may account, on average, for a nearly 6 ft increase in batted ball distance. This technique of using all the relevant measured properties of the baseball to estimate the batted ball distance for given launch conditions is one that MLB might adopt as a method of monitoring and regulating the carry of a baseball.
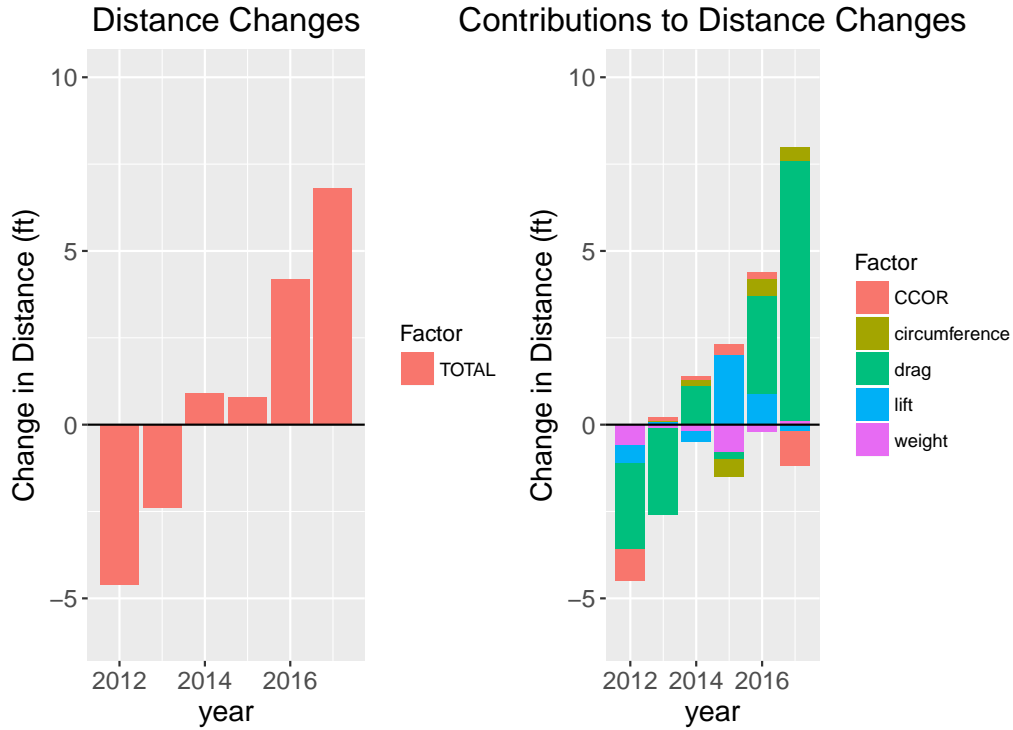
Figure 50: Estimated changes in distance of an ideal batted ball, using annual average ball properties as input to collision and trajectory models. The left plot shows the total change while the right plot shows the contribution of each factor to the change.

3.5   Analysis of Statcast Trajectories

3.5.1   Introduction

To complement the measurements performed at WSU, we extracted drag coefficients (C_d) directly from StatCast trajectory data. A random sample of StatCast trajectories was selected with the following characteristics for our analysis:

1. To minimize environmental effects such as wind and temperature, all trajectories were taken from Tropicana Field, which is always closed and climate controlled.

2. Selected trajectories were tracked for at least 90% of the distance traveled.

3. Preference was given to trajectories with no gaps in the tracking.

Approximately 200 trajectories for each period – 2015a, 2015b, 2016, and 2017 – were chosen. Each sample includes the trajectory of both the batted ball and the associated pitched ball. A distribution of the events included in our dataset is shown in Figure 51.
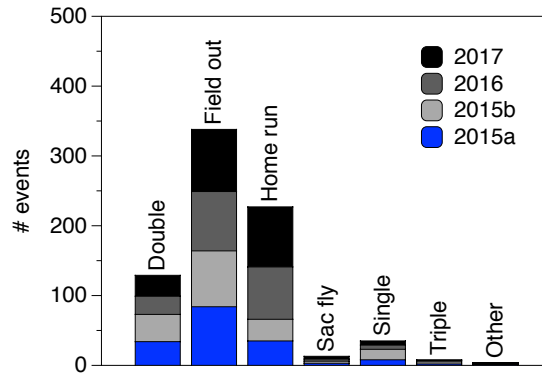

Figure 51: Distribution of events used in the trajectory dataset. (Other = field error or double play).

3.5.2 Drag Coefficient Analysis

To determine whether these trajectories exhibit a change in the mean drag coefficient relative to 2015a, we consider all samples in which the ball velocity lies in the range 70 – 93 mph, with the spin rate approximately 2500 rpm. Both velocity and spin were chosen to facilitate comparision with the WSU measurements. This range of velocities falls above the drag crisis, where C_d is expected to be roughly constant. Drag coefficients are computed using the change in ball speed along the trajectory, as discussed in Appendix C. Results from this approach are summarized by year in the box plot in Figure 52.

Three points are worth noting at this stage. First, there is considerable ball-to-ball variation within each period. Second, the average StatCast drag coefficients are consistent with the independent measurements performed at
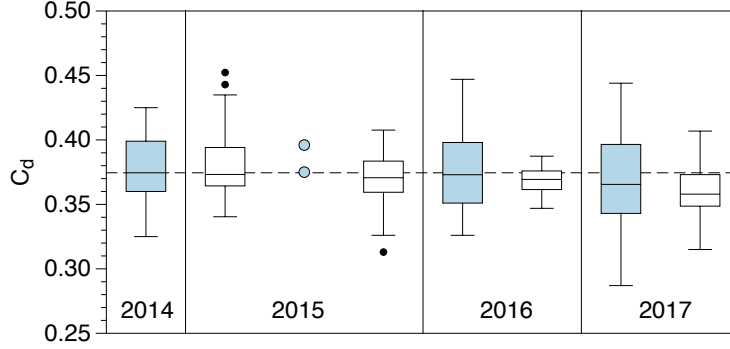
Figure 52: Box and whisker plots showing drag coefficients from 80 mph WSU authenticated ball tests (light blue) with a spin of 2500 rpm, and StatCast data (white) with similar velocities and spins as described in the text. Blue points in 2015 indicate the WSU measured drag coefficients for the two available authentication balls from that period.

WSU on authentication balls. Third, for the range of velocities under consideration, we found that the drag coefficient is well approximated by a linear function of spin:

$$C_d \approx C_{d0} + C_{d1} \times \text{spin}. \tag{1}$$

Despite the variation in the drag coefficient across different balls, we have a sufficient number of measurements in the StatCast and WSU new ball datasets to determine the mean C_d in each period to an accuracy that allows us to detect changes in C_d across multiple years. The measured change in the drag coefficient by year is shown in Figure 53. In the top plot, error bars indicate the standard deviation which is dominated by ball variation rather than by error in the measurements. In the bottom plot, error bars indicate the standard error of the mean. For the StatCast data, we use the average of 2015a and 2015b as a baseline to compute differences. Since we only have one set of baseballs which span all of 2015 in the WSU data, we take all of 2015 as the baseline for the WSU new ball data. Changes in C_d from the WSU authentication ball data are consistent with the other measurements but we have omitted those points in the plot due to the large uncertainty associated with the authentication 2015 baseline, for which there were only two authentication baseballs from 2015.

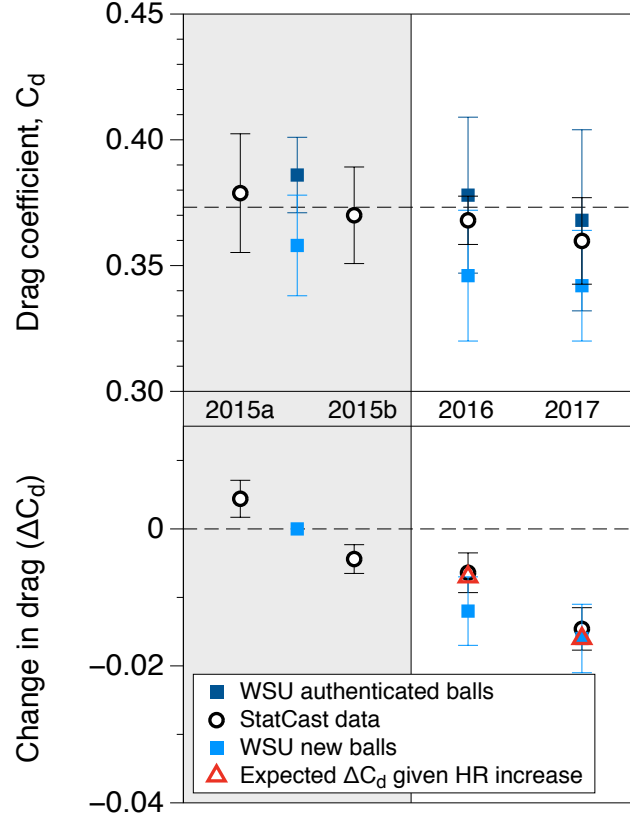All three analyses – StatCast, WSU authentication balls, and WSU new

Figure 53: Top: Mean drag coefficients calculated using StatCast data (white circles), WSU measurements with authentication balls (dark blue squares), and WSU measurements with new balls (light blue squares). Error bars in the top plot represent the standard deviation in the measurements. For the StatCast data, 2015 was divided into pre- and post-All Star Game periods (2015a and 2015b). The WSU measurements span all of 2015 (indicated by the blue points between the 2015a and 2015b circles). Bottom: Measured change in the drag coefficient C_d. StatCast measurements use the Trackman computed average of 2015a and 2015b as a baseline; WSU measurements use WSU 2015 measurements as a baseline (as indicated by the grey box). Error bars in the bottom plot represent the standard error of the mean. Red triangles indicate the expected change in drag coefficient given the observed change in home runs.

balls – indicate that the drag coefficient of Major League baseballs has changed since 2015. To determine whether this change could feasibly account for the increase in home runs, we next estimate the change in drag coefficient one would expect to measure, if the increase in home runs resulted entirely from changes in drag. Our strategy for computing this estimate is summarized in Figure 54, with details in Appendix D. Results of this calculation are summarized in Figure 55.
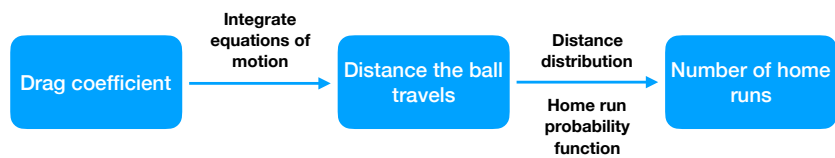

Figure 54: Strategy for estimating the number of home runs as a function of drag coefficient.
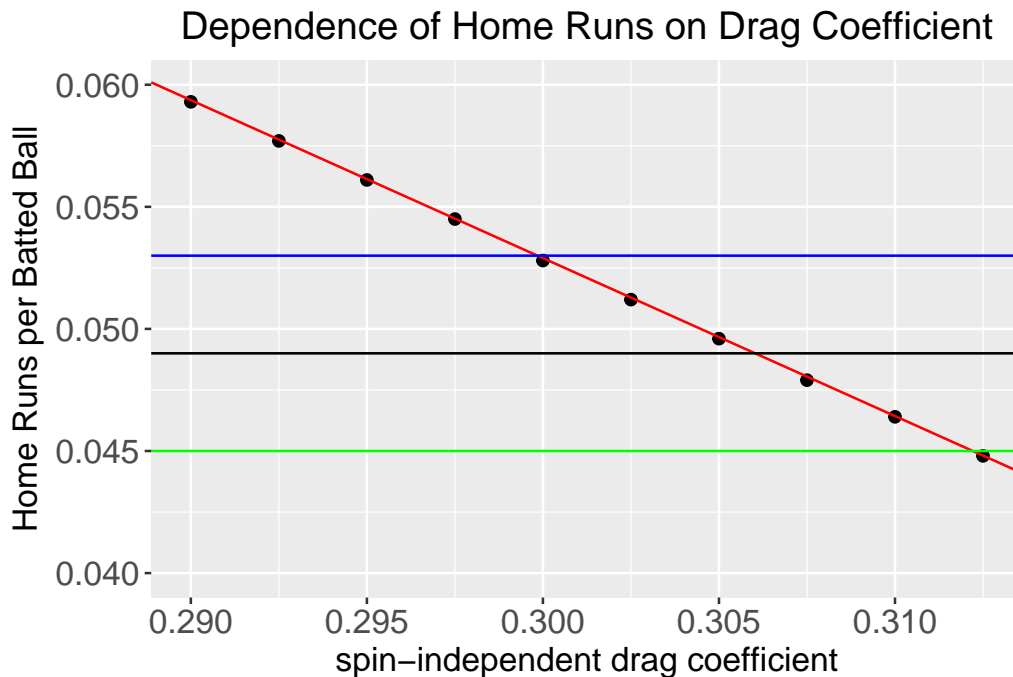
Figure 55: Expected number of home runs per batted ball as a function of the spin-independent component of the drag coefficient, C_{d0}, indicated by the points and red line. The horizontal lines are the actual values for 2015b (green), 2016 (black), and 2017 (blue).

If the change in the home run rate from 2015b to 2016 were due entirely to changes in drag, we would expect the drag coefficient to decrease by approximately 0.006. Similarly, the change in the home run rate from 2015b to 2017 corresponds to a change in the drag coefficient of approximately 0.012. These two numbers are indicated by the red triangles in Figure 53 along with the measured changes in drag coefficient from the WSU group and the TM data.

3.5.3 Conclusions

1. Independent estimates of the drag coefficient from StatCast data and from WSU measurements on authentication balls are consistent with one another.

2. Both analyses – StatCast and WSU – indicate that the drag coefficient

has changed by approximately 0.0153 since 2015. (The estimate 0.0153 represents the average of the StatCast and WSU data.)

3. From 2015b to 2017, the expected change in the home runs rate resulting from the observed change in the drag coefficient is about 25% larger than the observed change in home run rate over the same period (see Figure 53).

3.6 Exploring Other Properties Affecting Baseball Aerodynamics

Empirical evidence from laboratory results (above) and play (presented in Sec. 2) suggests that the recent MLB home run surge is primarily due to small changes in the ball's aerodynamic behavior. While these changes in average behavior are relatively small, the variability in aerodynamic response among the balls is not. As an example, Fig. 56 shows the distribution of C_d and C_l at 80 mph for all 96 authentication balls from 2017. The range for C_d is 0.15, from a low of 0.29 to a high of 0.44, and is about 8 times larger than the average change in C_d of 0.018 between 2014 and 2017. If we are to understand why there are small changes in the average behavior of C_d, we must first understand the factors leading to the large range of values among a given sample of baseballs. This leads us to pose two questions:

- What physical features of the ball lead to the broad range of C_d values?

- Why has the average of these features changed?

In this section, we address the first of these questions by investigating three possible features: seam height, center of gravity, and surface roughnness.
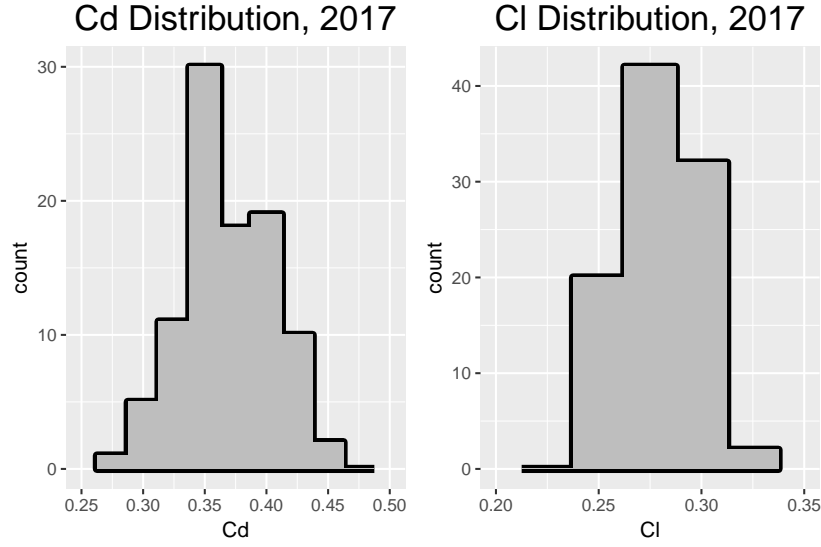
Figure 56: Distributions of drag and lift coefficients of the 2017 authentication balls at 80 mph.

3.6.1 Seam Height

The drag and lift coefficients of a baseball are almost surely affected by the seams, which are still sewn manually for all professional and amateur baseballs. The results of a recent experiment showing the relationship between drag coefficient and seam height is shown in Fig. 57. These data show that large changes in seam height (i.e., comparing raised- and flat-seam balls) correlate well with drag effects, while small changes in seam height (i.e., variation within a ball model) do not. For example, the trend line in the figure, which primarily comes from differences in C_d values from balls with large differences in seam height, predicts a 0.026 increase in C_d for a 0.005-inch increase in seam height. This increase likely overestimates the effect of seam height on C_d for MLB baseballs, where the variation in seam height is relatively small.
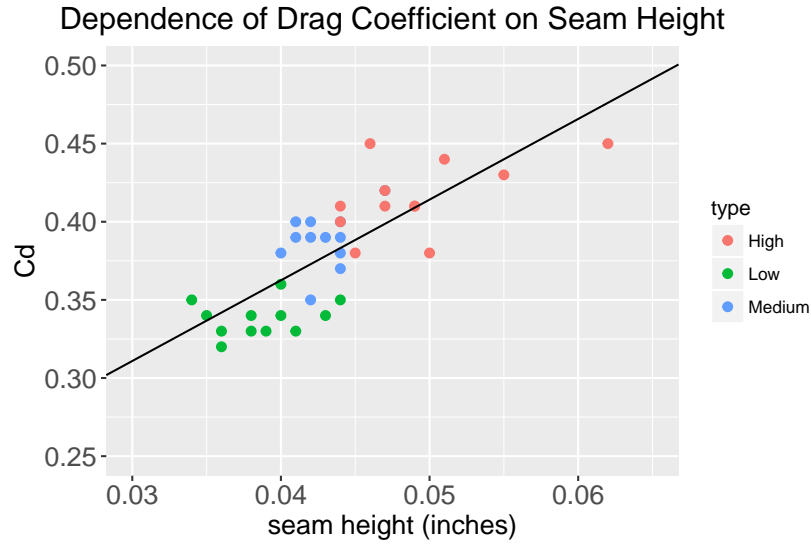
Figure 57: Relationship between C_d and seam height among different types of baseballs, where Low refers to MLB balls, Medium to MiLB balls, and High to balls formerly used by the NCAA. These data were published in Proceedings of the 7th Asia-Pacific Congress on Sports Technology APCST 2015.

By comparison, the seam height measured at UML for new balls is presented in Fig. 58. The average seam height in these data does not appear to correlate with the trends of yearly averages of C_d values. For example, the 2017 seam height is near a 15-year high whereas the mean C_d (Fig. 52) is at a recent low.
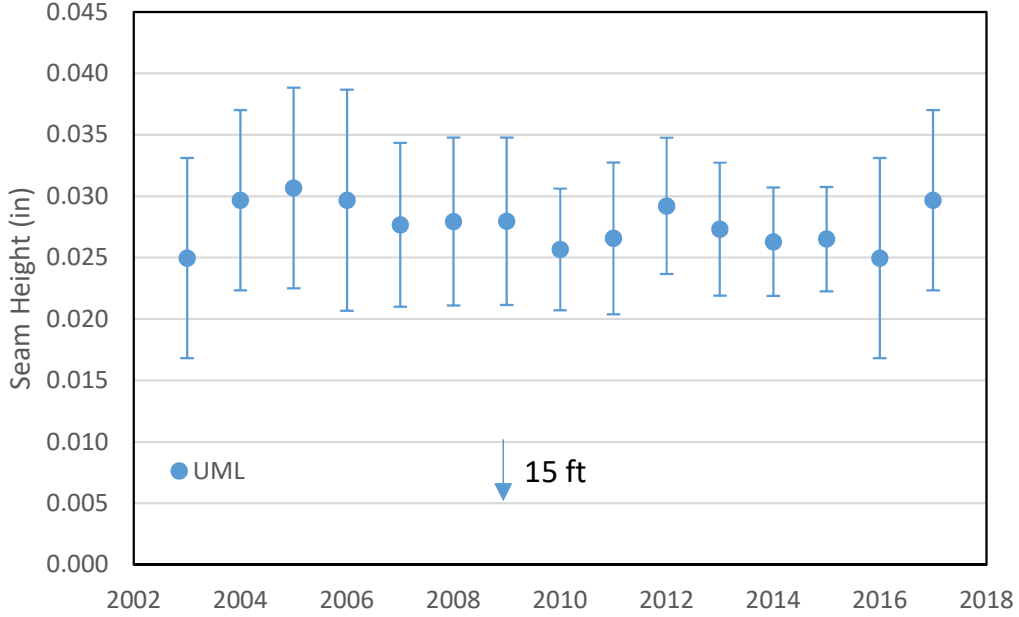
Figure 58: Annual seam height measurements from new balls at UML.

To further investigate this point, we have measured the seam height of 24 authentication baseballs. One group of 12 were from balls with the highest drag (average C_d=0.439); the other group from balls with the lowest drag (average C_d=0.310). The results are shown in Fig. 59. The average seam height of the high-C_d balls is 0.003 inches greater than that of the low-C_d group. Based on the average dependence of C_d on seam height from Fig. 57, this difference in seam height would results in a 0.016 difference in C_d, which is much smaller than the observed difference of 0.129. We conclude that seam height does not account for the wide spread of C_d values among the authentication balls.
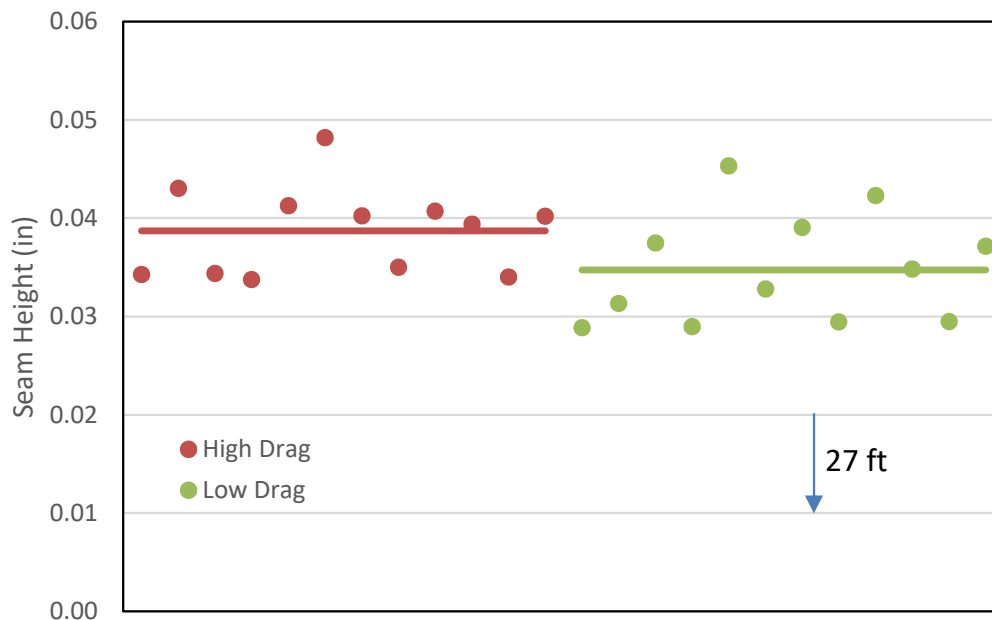
Figure 59: Comparison of seam height of the authentication balls; 12 with the highest drag, and 12 with the lowest drag.

3.6.2 Center of Gravity Effects

It is possible that differences in the way a ball spins affects its aerodynamic response. If a ball has a center of gravity that is offset from its geometric center, the ball will wobble as it spins, increasing its effective surface roughness and potentially altering its drag. To test this hypothesis, the balls from Fig. 59 were projected through still air with and without rotation. The essential line of reasoning is as follows. It is known that drag increases with spin. If part of that increase is due to the center of gravity effect and if there is a large variation in those effects, for whatever reason, then the variation in drag among these balls at low spin will be less than that at high spin. That is exactly what is observed in the data, as shown in Fig. 60, where the difference between the high- and low-drag balls at a spin rate of 2500 rpm was 0.085, while the difference with zero spin was only 0.017. These results are consistent with the aforementioned line of reasoning and suggest that differences in the ball center of gravity may be contributing to changes in ball drag. Additional work to study this effect is ongoing.
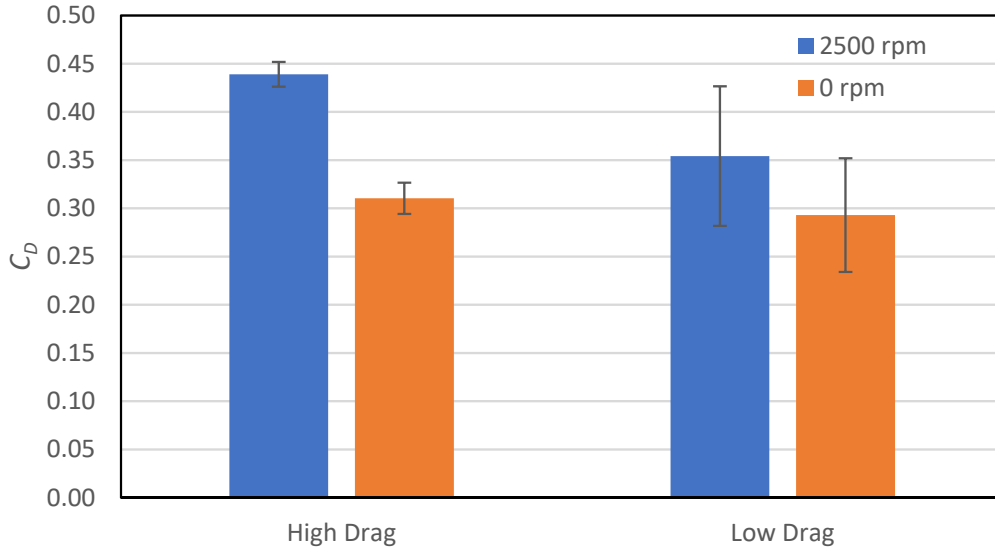
Figure 60: Drag coefficients of low-drag and high-drag baseballs at spin rates of 0 and 2500 rpm.

While the results of Fig. 60 suggest that ball center of gravity may play an important role in ball aerodynamic response, it does not answer the question of why the ball center of gravity may have changed. One possible cause could be more uniform winding of the ball or more uniform yarn used to make MLB balls. Improvements in either the yarn, or the way it is wound, could lead to improved alignment of the geometric and mass centers of the ball.

3.6.3   Surface Roughness

It is known from both wind tunnel studies and trajectory analysis that the drag coefficient of a baseball depends on speed. At low speeds, up to about 50 mph, C_d is relatively large and independent of velocity, $\approx$0.50. At some speed beyond 50 mph, it starts to decrease, reaching a minimum at around 70 mph. This drop in C_d is known as the "drag crisis". We can characterize the size of the drag crisis as the difference between the low-speed value of C_d and the minimum. It is also known that size of the drag crisis depends on the surface roughness of the ball, being smaller for a rougher surface.

One speculative idea about the large ball-to-ball variation in C_d is that it is due to variation in the surface roughness. The WSU data on authentication

balls can be used to investigate this idea by looking at how the difference in C_d between 55 and 80 mph depends on C_d at 80 mph. The expectation based on this idea is that the difference (a measure of the size of the drag crisis) is anti-correlated with the C_d value at 80 mph.

The results of this analysis are presented in in Fig. 61 and seem to be consistent with this idea: Balls with higher drag are observed to have a lower drag crisis, lending some credence to the notion that surface roughness plays a role in baseball drag. This is another line of research that will be pursued in the near future.
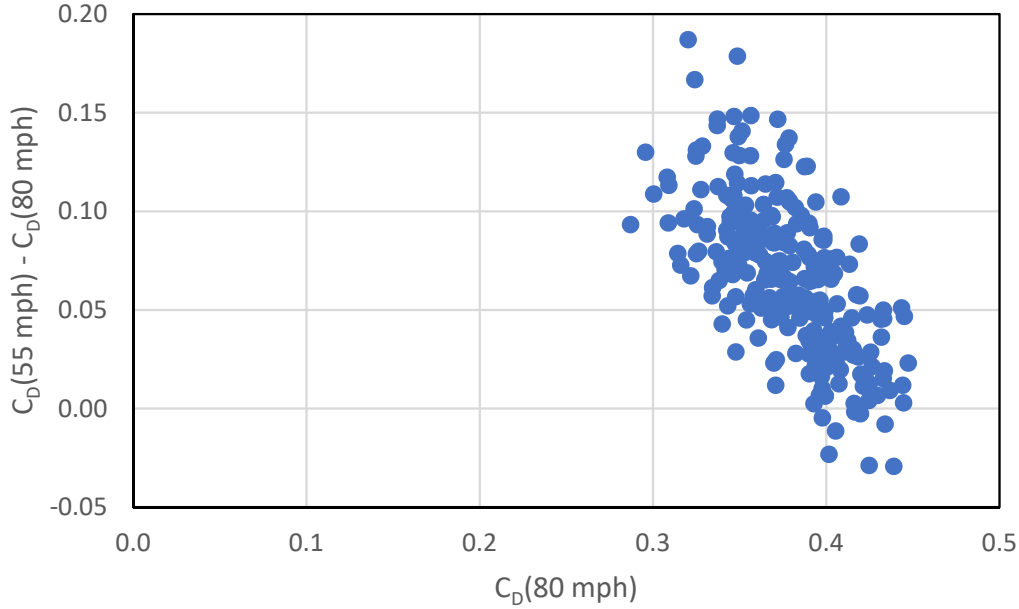

Figure 61: The dependence of MLB baseball drag crisis on drag magnitude.

Appendix A Factors Affecting Exit Velocity

For a squared-up collision between baseball and bat, the exit velocity v_{exit} is related to the pitch speed and swing speed by

$$v_{exit} = qv_{pitch} + (1+q)v_{swing}, \tag{2}$$

where the collision efficiency q is related to the ball-bat coefficient of restitution, denoted here by the symbol e, by

$$q = \frac{e - m/M}{1 + m/M},\tag{3}$$

where m and M are the ball and effective bat mass, respectively. For speeds typical of a high-speed ball-bat collison in MLB, with an impact at the sweet spot of the bat, $e \approx 0.45$ and $m/M \approx 0.25$ (m=5 oz, M=20 oz), resulting in $q \approx 0.160$. Taking $v_{pitch} = 85$ mph and $v_{swing} = 77$ mph (total impact speed of 162 mph) results in an exit speed of 103 mph, which is very typical of home runs in MLB. As a reminder, v_{pitch} is the speed at home plate and corresponds to a release speed of about 94 mph.

The coefficient of restitution e can be related to the quantities that are measured in the laboratory. One of these is the quantity called the COR, which is measured at an impact speed of 60 mph on a fixed flat steel plate. The other is the quantity called the CCOR, which is measured at an impact speed of 120 mph on a fixed cylinder. Ignoring the small difference between a flat and cylindrical surface, these quantities can be related to the quantity of interest, e, which corresponds to an impact speed at 162 mph on a free bat. We make the assumption that the fractional changes at the two speeds are identical. For example, suppose the CCOR at 120 mph, nominally 0.485, is higher by 0.008. This corresponds to a fractional change of 0.0165. Applying the same fractional change at 162 mph, then δe=0.0074. A similar analysis can be applied to the COR, nominally 0.555, to find that the same change in e is obtained with a change in COR of 0.009.

Given the formulas, it is straightforward to investigate how a specified change in any given parameter affects exit speed. In the table below, we show how the indicated change in various parameters changes the exit speed Δv_{exit}. Assuming small changes, these effects are all linear, so the net effect of changing several parameters is simply to add the results.

Table 7: Effect of changes in parameters from their nominal value exit speed, assuming all other parameters are kept fixed. The last column is an estimate of the resulting change in distance, assuming that each 1 mph of exit velocity adds 5 extra feet of carry for balls hit at launch angles typical of long fly balls.

parameter	change	Δv_{exit} (mph)	Δd (ft)
m	+0.125 oz	-0.94	-4.7
v_{pitch}	+1 mph	+0.16	0.8
v_{swing}	+1 mph	+1.16	5.8
COR	+0.009	+1.00	5.0
CCOR	+0.008	+1.00	5.0

Appendix B Factors Affecting Baseball Carry

When a spinning baseball travels through the air, it experiences three forces. Gravity pulls the ball downward; drag slows it down; and lift or the Magnus force (assuming the ball is spinning) changes its direction. If baseball were played in a vacuum, it would only experience gravity. Under such conditions, all balls projected with the same initial velocity and launch angle wou,d carry the same distance. However, our world is complicated by the fact that as the ball moves through the air, it has to push air molecules out of the way. As it does so, the air pushes back on the ball, slowing it down. As a simple example, a typical home run hit with an exit velocity of 100 mph at the optimum launch angle of 27.5° travels just about 400 ft. In a vacuum that ball would travel 555 ft. And if the ball were hit with the same exit velocity but at the optimum vacuum launch angle of 45°, it would travel 673 ft. So, the drag and lift on the baseball have an enormous effect on the carry of a fly ball. Generally speaking, drag reduces the carry and lift (assuming it is the result of backspin) increases the carry. But the effect of drag on carry is far greater than the effect of lift. Therefore we focus primarily on drag.

The following simple equation, in words, shows the properties of the baseball that contribute to the effect of drag on the carry of a fly ball:

$$\text{effect of drag} \propto \frac{\text{air density} * \text{diameter}^2 * \text{drag coefficient}}{\text{mass}} . \tag{4}$$

This equation shows that the air density, size, the mass, and the drag coefficient, which we denote by C_d, all contribute to the carry. The larger the ball, the more air it has to push out of the way, resulting in a larger drag force and less carry. The drag force causes the ball to slow down (i.e., to decelerate), where the rate of deceleration is inversely proportional to the mass. Therefore, a heavier ball has less deceleration, resulting in greater carry. The drag coefficient C_d is a factor reflecting the fact that ball doesnt have to push the air completely out of the way, since the air can sort of slide around the ball. The drag force is directly proportional to C_d, so that a larger value of C_d means greater drag and less carry. Generally C_d depends on the surface roughness of the ball, possibly including the height of the seams. It depends on both the speed and spin of the ball and is typically in the range 0.25-0.50.

The air density is not a property of the ball but is an important effect nevertheless, depending on altitude, temperature, and to a smaller extent on humidity. It shows up most dramatically in mile-high Denver, where the density is only about 82% of that at sea level, leading to significantly reduced drag, longer fly balls, and more home runs. It also is affected by temperature. For example, the air density at 95° is about 91% that at 50°, which is half of the Denver effect.

In the table below, we show how variation of mass, circumference, and drag coefficient affect distance for fly balls on a typical home run trajectory.

Table 8: Estimate of the effect of changes in mass m and circumference C (both in the full range allowed by MLB), and drag coefficient on fly ball distance for exit velocity=102.5 mph and launch angle=27.5°), for which the nominal fly ball distance is 405 ft. A change in distance of 5 ft increases home run probability by about 12%.

parameter	change	Δd (ft)
m	$\pm$ 0.125 oz	$\pm$ 4.6 ft
C	$\pm$ 0.125 in	$\pm$ 5.2 ft
C_d	$\pm$ 0.01	$\pm$ 5.2 ft

Appendix C Extracting Drag Coefficients from Trackman Trajectories

A baseball flying through the air experiences three forces: the drag force, $\mathbf{F}_D$, the Magnus – or lift – force, $\mathbf{F}_M$, and gravity. The drag force and the Magnus force can be expressed as:

$$\mathbf{F}_D = -\frac{1}{2}\rho A C_D V^2 \,\hat{\mathbf{t}} \tag{5}$$

$$\mathbf{F}_M = \frac{1}{2}\rho A C_L V^2 (\omega \times \hat{\mathbf{t}}). \tag{6}$$

Here ρ is the density of air, A is the cross-sectional area of the baseball, V is the magnitude of the velocity of the ball, $\mathbf{t}$ is the unit vector tangent to the trajectory, and ω is a vector representing the spin of the ball. Values of these parameters that were used in this study are summarized in Table 9. The coefficients C_D and C_L are the drag and lift coefficients, respectively; both are considered unknown in the following analysis.

Baseball mass	m	0.146 kg
Density of air	ρ	1.225 kg/m^3
Gravitational acceleration	g	9.8 m/s^2
Baseball radius	a	3.69 cm
Dynamic viscosity of air	μ	1.85×10^{-5} Pa $\cdot$ s

Table 9: Parameters used to extract C_d from Statcast data.

Putting all three forces together leads to the equations of motion for the baseball:

$$m\frac{d\mathbf{V}}{dt} = \mathbf{F}_D + \mathbf{F}_M + m\mathbf{g} \tag{7}$$

where $\mathbf{g} = (0, 0, -g)$ represents the vector associated with gravitational acceleration.

To isolate C_d we first project equation (7) onto the tangent vector $\mathbf{t}$ which leads to a scalar differential equation for the evolution of the magnitude of the velocity

$$m\frac{dV(t)}{dt} = -\frac{1}{2}\rho A C_d V(t)^2 - mg\sin\theta(t). \tag{8}$$

Since the Magnus effect is perpendicular to the trajectory, this step has the advantage of removing one unknown, C_L from the formulation.

Statcast provides data for the individual components of velocity as a function of time which can be used to compute the instantaneous magnitude of the velocity at the ith time step, V_i:

$$V_i = (v_{x,i}^2 + v_{y,i}^2 + v_{z,i}^2)^{1/2}$$

and the instantaneous acceleration:

$$dV_i/dt = (V_{i+1} - V_{i-1})/(t_{i+1} - t_{i-1})$$

where $v_{x,i}$, $v_{y,i}$, and $v_{z,i}$ are the components of the ith measured velocity in the trajectory. The last piece we need to compute instantaneous drag coefficients $C_{d,i}$ is the instantaneous angle of the trajectory, θ_i which can be obtained from the change in position from time i to $i + 1$:

$$\theta_{i+1/2} = \arctan \frac{(z_{i+1} - z_i)}{[(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2]^{1/2}}$$

$$\theta_i = \frac{\theta_{i+1/2} + \theta_{i-1/2}}{2}.$$

The instantaneous velocities, accelerations, and trajectory angles can now be combined to compute the instantaneous drag coefficients for our randomly selected trajectories.

$$C_{D,i} = -\frac{2m}{\rho A V_i^2} \left(\frac{dV_i}{dt} + g \sin \theta_i \right). \tag{9}$$

Each trajectory has an associated vector of drag coefficients which represents the evolution in time as the velocity and spin decrease.

We observed that the first $\approx$10 ms of Trackman data often contain unphysical velocities (e.g. velocities that increase in the absence of accelerating forces). Hence we remove these points from both pitched and batted trajectories. A typical pitch takes approximately 40 ms to go from the pitchers mound to the plate. To avoid complex interactions near the bat, we only consider the first 30 ms (minus the first 10 ms) of the pitch trajectory. For batted balls, we only consider points up to the trajectory peak.

Using equation (9), we compute the C_d's, for a fixed range of spins and bin them by velocity. Figure 62 shows C_d's computed for all four time periods (2015a, 2015b, 2016, and 2017). For each period we show two spin populations, one centered about 1500 rpm and one centered about 2500 rpm. Spin

varied by approximately ±150 rpm and the exact interval was selected such that the mean of the first population was 1500 rpm and the mean of the second 2500 rpm. Trackman data for spin is truncated at 3500 rpm in the early datasets; to remove this bias we only considered spins below 3500 for all years. In both cases – 1500 rpm and 2500 rpm – we bin the points by velocity (in the Figure we take 50 points per bin); points in the plot represent bin averages.[1]
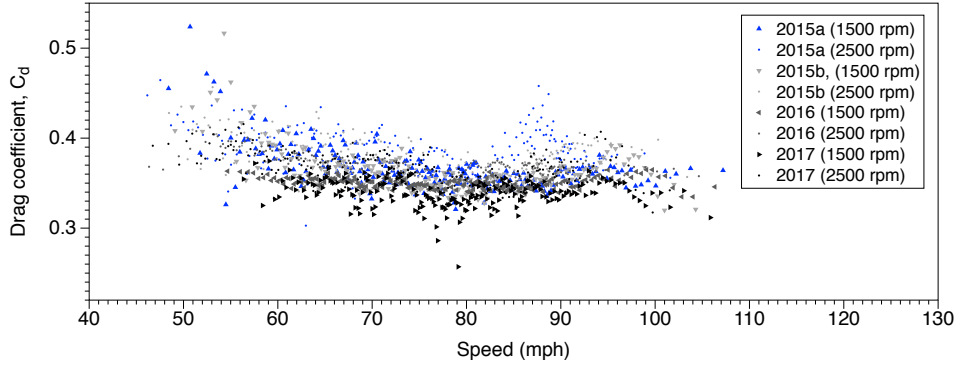

Figure 62: Drag coefficients, C_d calculated from Trackman data using both batted and pitched trajectories in Tropicana Field. Note that for speeds above approximately 65 mph, the drag coefficient does not exhibit a strong dependence on speed.

Appendix D   Estimating expected increase in home runs

Using a logistic smoothing mechanism with StatCast data from 2017, we compute the probability of a home run as a function of both distance and spray angle. The latter is necessary since the fence distances (and therefore the home run probability) is a strong function of that angle. These probability distribution functions are shown in Figure 63. As expected, the

---

[1]As an alternate method of smoothing the data, we also computed C_d's using a high order polynomial to fit the Trackman velocities. Results from this method did not vary significantly from the binning and averaging technique presented here.

probability is zero for small distances and unity for large distance, with a relative sharp transition at a distance that depends on the spray angle, getting larger for hits to CF and smaller as spray angles move closer to the foul line. The steepness near the transition means that small changes in the distance traveled result in significant changes in home runs for distances near this transition.

Next we consider all batted balls in with exit speeds $\geq$ 90 mph, and launch angles that range between 15–45 degrees, a range that encompasses over 98% of home runs. Using a well-developed model for lift and drag, along with the actual batted ball parameters (including spin and spin axis), the equations of motion are integrated using these initial conditions to find the distance traveled. The C_d utilized both the spin-independent and spin-dependent parts (see Eq. 1, where the latter was fixed from a global analysis of the data. These computed distances can then be combined with the home run probability function to calculate the expected number of home runs per batted ball for a given drag coefficient. We then repeat this exercise with a range of drag coefficients to calculate the expected number of home runs per batted ball as a function of the spin-independent portion of the drag coefficient, C_{d0}, with the spin-dependent part C_{d1} fixed. Results of this calculation are summarized in Fig. 55.
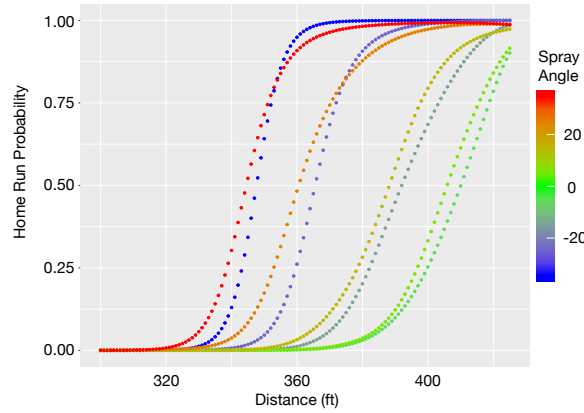

Figure 63: Home run probability as a function of distance for a variety of spray angles.