Investigating the Increase of Home Runs in Major League Baseball Since 2015

Ryan Hamlett
University of California, Riverside

June 13, 2018

Abstract

We do an nonparametric exploration of the effects of launch angle and exit velocity on home run probability in Major League Baseball. We use a linear mixed model approach to generalized additive models to perform a bivariate local logistic regression, perform both univariate and bivariate kernel density estimates for the years 2015 and 2017. We ultimately show the difference in the predicted values of the generalized additive models and conclude that there does not seem to be a significant difference in home run probability when controlling for launch angle and exit velocity. Key words: nonparametric, regression, home run, exit velocity, launch angle,

1 INTRODUCTION

In the late 1990's, home runs skyrocketed across the MLB due to an influx of players using human growth hormone (HGH) and many other performance-enhancing drugs (PED). The record for most home runs in a season league-wide was broken in 1998, again in 1999, and once more in the year 2000. (Baseball Almanac, 2017)[1] This rapid increase in home runs raised eyebrows across the sport and even sparked a Congressional investigation into the possible use of steroids in the MLB. (Mitchell and Waxman, 2008) [6]

As you might imagine, home run levels have continued to be under scrutiny in baseball since the late 1990's. In the years 2013 through 2015, however, home run levels were under scrutiny for an entirely different reason. From the second half of the 2013 MLB season until the first half of the 2015 season, home run rates plummeted. In 2014 alone, scoring was at its lowest since the 1976 MLB season and home runs were at their lowest levels since 1995. (Lindbergh, 2017) [4]

Home run levels spiked after the All-Star Game (ASG) in 2015 in a way that is unprecedented in recent MLB history. The rate of home runs per ball hit (HR/Contact) increased by almost one full percent, from 3.5% to 4.5% from before the ASG to after the ASG, the largest single-season increase in HR/Contact since the 1950 MLB season. Since the 2015 season, home runs have continued to increase, culminating in a record-breaking 2017 season in which there were 6,105 home runs hit, an increase of more than one-thousand home runs from the 2015 season. (Arthur and Lindbergh, 2017) [3]

2 BACKGROUND

In 2017, Ben Lindbergh and Mitchell Lichtman published a landmark article that showed that baseballs from the before the ASG in the 2015 MLB season were significantly different from baseballs after the 2015 ASG. These differences in the ball were theorized to allow the balls from after the ASG to travel further than those from before the ASG for a fixed speed (exit velocity) and angle (launch angle) off of the bat. (Lindbergh and Lichtman, 2017) [5]

The purpose of this paper is to take a nonparametric statistical approach in examining the increase of home runs in baseball since 2015. Building on the research done by Lindbergh and Lichtman (2017), I will examine the relationship of home run probability with exit velocity and launch angle and show that there is some change in the distribution of home run probability with respect to those two variables.

3 DATA COLLECTION AND EXPERIMENTAL DESIGN

There are two distinct datasets in this study. The first dataset consists of $n_{2015} = 30470$ observations of batted balls before the All-Star Game on July 14, 2015. The second dataset consists of $n_{2017} = 31374$ observations of batted balls before the All-Star Game on July 11, 2017. This means I am studying the same amount of games in the same time frame in two different seasons.

It is within the realm of questioning to wonder why I didn't choose to collect the data from the first half of the 2015 season and compare it to that of the last half of the 2015 season. My reasoning for choosing these datasets is as follows: the population of hitters in baseball changes drastically every September, as rosters expand from 25 to 40. Each September, there is an influx of minor-league level hitters now hitting in the major leagues. I worried that this influx of less-talented hitters would possibly skew the data in season. It made the most sense, then, to compare two different seasons, both of which in the first half. The population of MLB hitters does not change much from season to season, and so a two year gap is acceptable. So it stands to reason that we may want to compare the data from the most recent, record-breaking season to the data from just two years ago when baseball was in the midst of its offensive crisis.

The data for this study was collected using Daren Wilman's *Baseball Savant Statcast Search.* Statcast is a system installed in all 30 MLB stadiums that collects information using a series of high-resolution optical cameras. The technology precisely tracks the location and movements of the ball and every player on the field at any given time. (Wilman)

Since our study is retrospective in nature, we have a random design study where we are simply observing random triplets that consist of an indicator of whether a batted ball was a home run (HR), the speed (exit velocity), and angle (launch angle) at which a batted ball leaves the bat.

4 MODELING

In this paper, we consider a bivariate nonparametric regression solution to modeling home run probability against exit velocity and launch angle. In general, since we are only dealing with a bivariate problem, we can perform ordinary bivariate local logistic regression by estimating the bivariate function of exit velocity and launch angle that properly models home run probability in R. In this case, we consider a functional form of Y = η(X) where Y = home run probability, η = the canonical link function for logistic regression, X = (launch speed, exit velocity), and $\eta_i = X_i^T \beta$ In general, however, we are interested in the bivariate regression function as well as the individual functional forms of launch speed and exit velocity with respect to home run probability.

The framework above is a natural fit for an additive local logistic regression model. In this case, we consider

$$Y_i = \eta_i = \eta(X) = \beta_0 + f_1(launchspeed) + f_2(exitvelocity) + \epsilon_i$$

where $f_j(x_j)$ are the unknown univariate smooth functions relate launch speed and exit velocity to the link function. These $f_j(x_j)$ need to be estimated. Ultimately, what we are doing is nonparametrically modeling the relationship between the log odds of home run probability and its relationship to both exit velocity and launch angle. In our analysis later in the paper, we will be using the inverse logit function to transform our predicted values back into simple home run probabilities since that is the actual quantity of interest in this study.

5 ESTIMATION METHODS

The method of estimation we will be using is penalized spline regression. This method is done by minimizing a penalized least squares of the form

$$\sum_{i=1}^{n}(Y_i - \vec{B}(x_i)^T\vec{\gamma})^2 + \lambda\vec{\gamma}^T\Omega\vec{\gamma}$$

where λ = the tuning parameter, $\vec{B}$ is a vector with any spline basis of our choosing, and $\vec{\gamma}^T\Omega\vec{\gamma}$ is a penalty term that we choose. To penalize for roughness, we typically choose $\Omega_{j,j'} = \int B_j^m(x)B_{j'}^m(x)dx$. There are several different options for B, but the most intuitive to the author is the use of the truncated power series basis $\vec{B}_(x) = \{1, x, x^2, \ldots, x^{p-1}, (x-\kappa_1)_+^{p-1}, \ldots, (x-\kappa_j)_+^{p-1}\}$, where we take $(x-\kappa_i)_+ = \begin{cases} x-\kappa_i : & if x-\kappa_i > 0 \\ 0 : & if x-\kappa_i \leq 0 \end{cases}$

If we let X = $\begin{Bmatrix} 1 & x_1 & x_1^2 & \ldots \\ 1 & x_2 & x_2^2 & \ldots \\ \vdots & \vdots & \vdots & \\ 1 & x_n & x_n^2 & \ldots \end{Bmatrix}$ and Z = $\begin{Bmatrix} (x_1-\kappa_1)_+^{p-1} & \ldots & (x_1-\kappa_j)_+^{p-1} \\ \vdots & \vdots & \vdots \\ (x_n-\kappa_1)_+^{p-1} & \ldots & (x_n-\kappa_j)_+^{p-1} \end{Bmatrix}$, then we can use the method established by Ruppert (2003) that allows us to model this as a linear mixed model $Y = X\vec{\beta} + Z\vec{u}$, where $\vec{u}\ N(\vec{0}, \sigma_u^2 I_j)$, $\vec{\epsilon}\ N(\vec{0}, \sigma_\epsilon^2 I_n)$, $\vec{\beta}$ is a vector of the fixed coefficients corresponding to the base polynomial terms and $\vec{u}$ is a vector of the random coefficients corresponding to the spline terms.

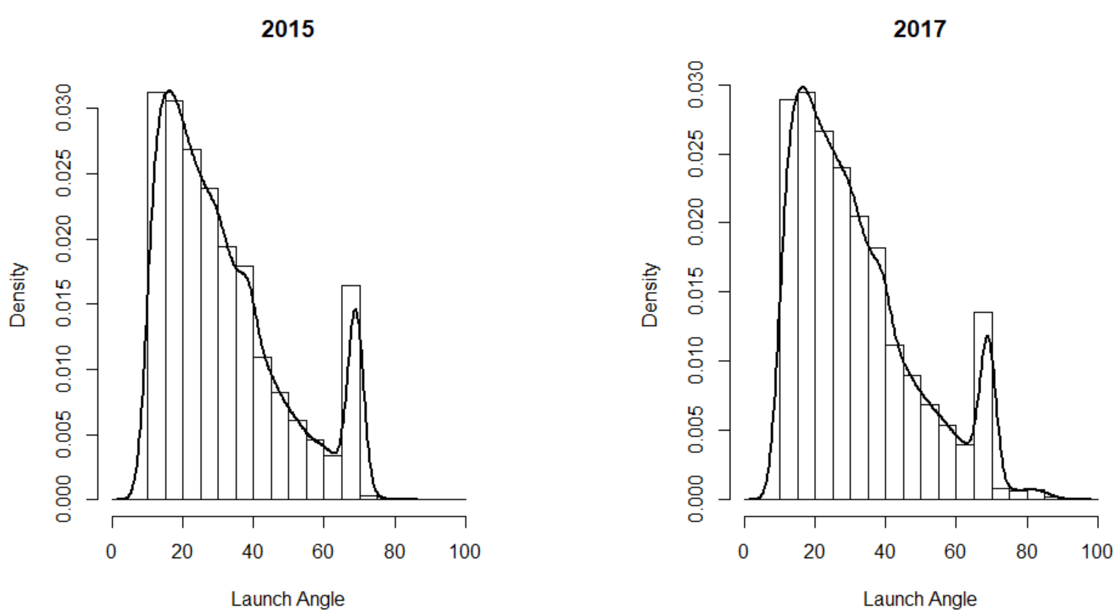
If we maximize the log-likelihood of the above linear mixed model, it is the equivalent of minimizing the penalized least squares we displayed earlier, and we get an estimate of our tuning parameter $\lambda = \frac{\sigma_\epsilon^2}{\sigma_u^2}$ which is generally interpreted as the signal to noise ratio.

In order to estimate the additive model we have discussed in the previous section, we will be using the same linear mixed models method approach to penalized splines, but in a slightly different way. In this process, we will be fitting a model where $Y_i = \eta_i = \beta_0 + f(s_i) + g(t_i) = \beta_s(s_i) + \beta_t(t_i) + \sum_{k=1}^{k^s} u_k^s(s_i - \kappa_k^s)_+ + \sum_{k=1}^{k^t} u_k^t(t_i - \kappa_k^t)_+$, where β terms are modeled as fixed effects while the u_k^s and u_k^t terms are modeled as random coefficients for each spline term, $(x - \kappa_i)_+$. The estimation of these coefficients is done via the backfitting algorithm. For the backfitting algorithm, we start with some current estimate of $\hat{f}_1$ and update the the other $\hat{f}_2 = S_2(\vec{Y} - \vec{\beta} - \hat{f}_1)$ where S_2 is the spline or local polynomial smoother we are using. Once we have an estimate for $\hat{f}_2$, we iterate the process for $\hat{f}_1$ and continue until we reach some convergence criterion.

All of the methodology above is done using the R package "SemiPar" through the "spm" function.

6 DATA ANALYSIS

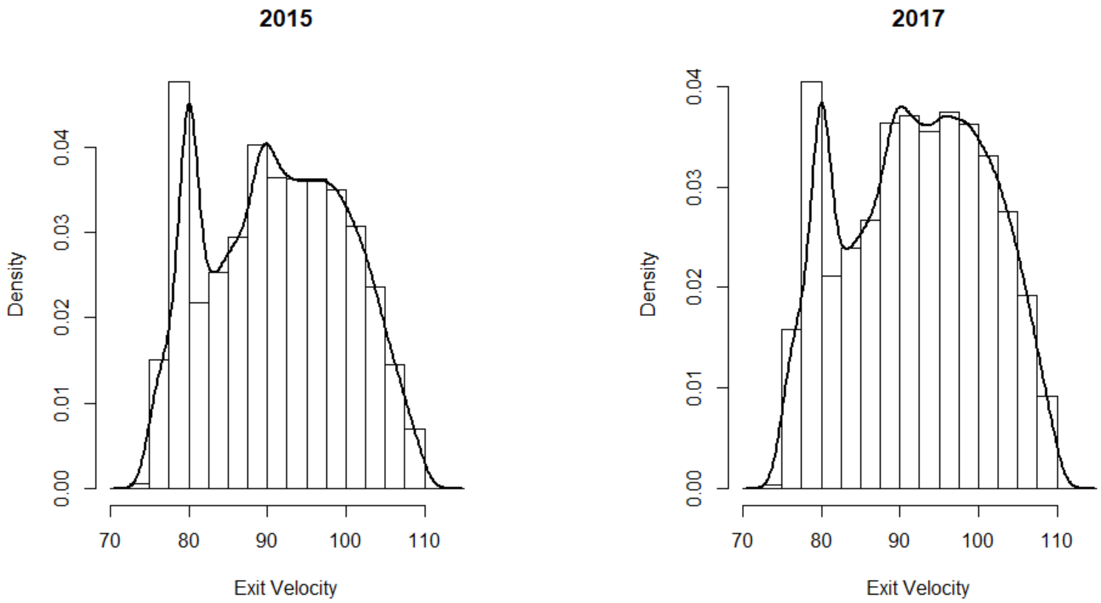
The first thing we should do in our data analysis is explore the shape of the data. It seems intuitive given our nonparametric methodology that we should simply plot histograms and kernel density estimates of each of our covariates as we have done in Figures 1 and 2. For both of the continuous covariates "launch speed" and "exit velocity", we have used the rule-of-thumb bandwidth selector with a Gaussian kernel.



(a) 2015

(b) 2017

Figure 1: Launch Angle Kernel Density Estimates.



(a) 2015

(b) 2017

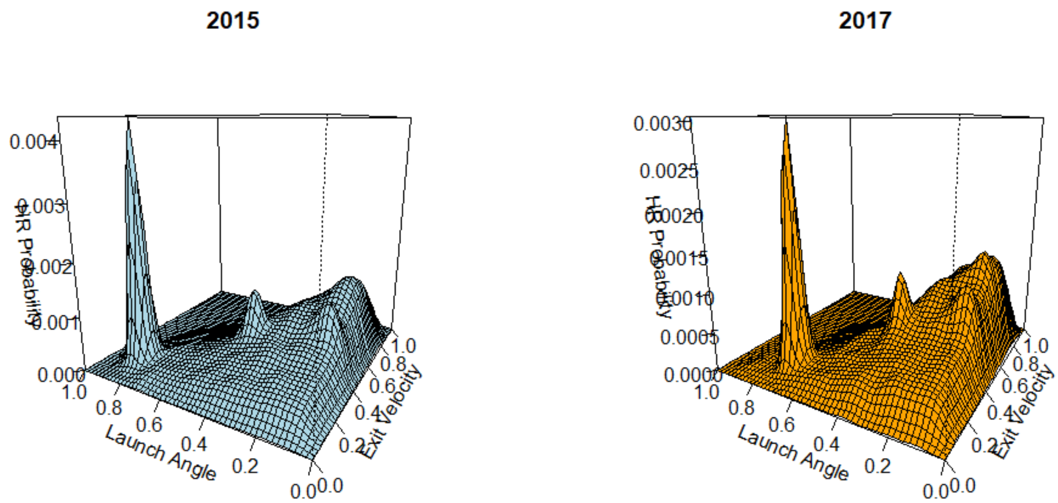
Figure 2: Exit Velocity Kernel Density Estimates.

If we examine the density estimates of these two covariates, we can see that the distribution of launch angles for the years 2015 and 2017 seem to be nearly identical. It is interesting, however, that the distribution of exit velocity seems to shift a bit to the right, towards larger values, in 2017. It is also important to note that the distributions for launch angle and exit velocity are not normal for either year. One of the key issues in using parametric generalized linear models is that we typically assume normality of errors.

The next thing we may want to do is analyze the joint distribution of exit velocity and launch angle for both 2015 and 2017. This can be seen in figure 3. Once again, we can see that the surface is shifting in the direction of larger exit velocities. We see a similar spike in probability around an exit velocity of 80 mph and a launch angle of about 70 degrees on both perspective plots however, which indicates that the average weakly hit fly ball is still hit about the same way in 2017 as it is in 2015.

Finally, we need to actually fit each of our models and check the functional forms of the smooth covariate functions. In Figure 4 and 5, we have the estimates of the smooth additive functions of exit velocity and launch angle for the 2015 and 2017 seasons respectively. In both cases, it looks like the relationship between the log odds of home run probability and exit velocity is approximately linear and the relationship between the log odds of home run probability and launch angle is roughly quadratic in nature.

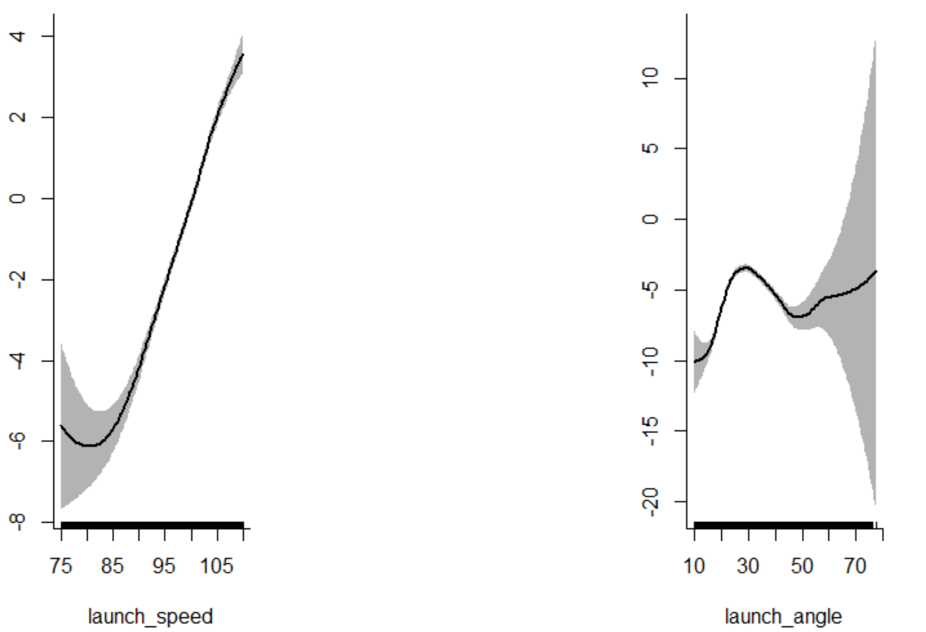
The intuition behind the exit velocity relationship can be understood in this way: we expect that home run probability will increase steadily as exit velocity increases for balls with a launch angle greater than



(a) 2015

(b) 2017

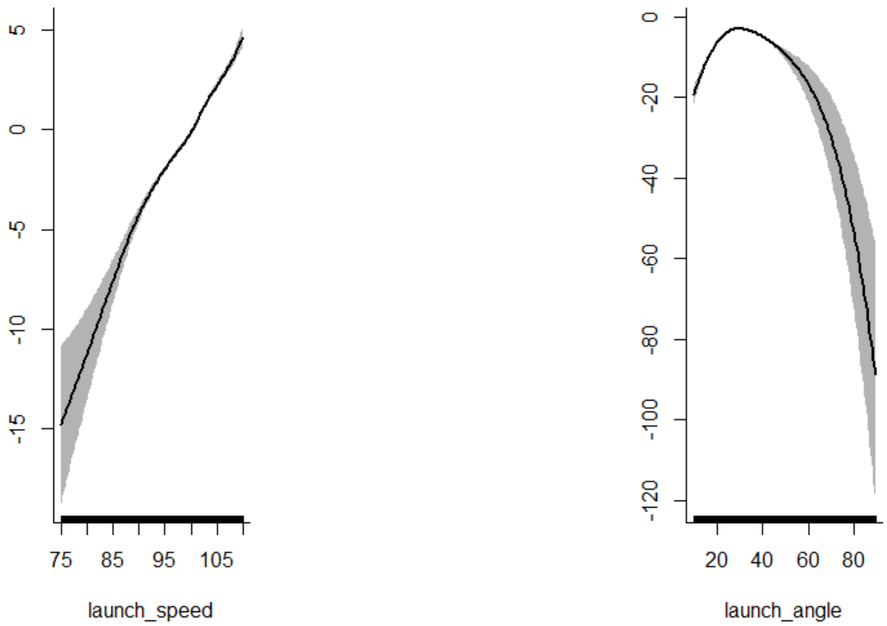
Figure 3: Two-Dimensional Density Estimates



(a) Exit Velocity

(b) Launch Angle

Figure 4: Estimates of Smooth Additive Functions for 2015



(a) Exit Velocity

(b) Launch Angle

Figure 5: Estimates of Smooth Additive Functions for 2017

10 degrees. Since we are only considering balls hit at a launch angle of 10 degrees and above, we might expect that home run probability will increase until the probability is extremely close to 1 for some large exit velocity outside of our current observed range. Nonetheless, in terms of log odds, we see that the probability of a home run divided by the probability of not hitting a home run increases linearly with exit velocity.

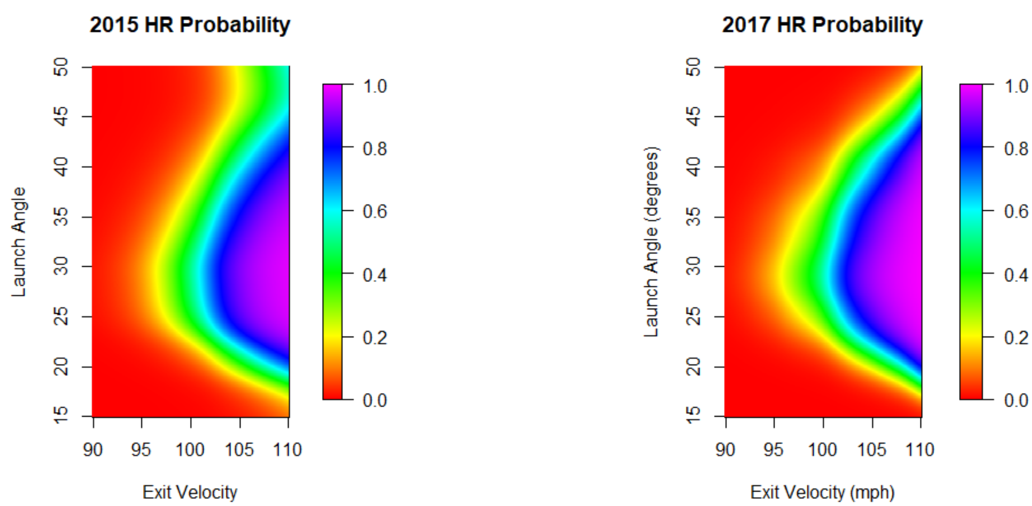
Similar logic can be applied to the intuition behind the launch angle relationship. We expect that balls hit at a lower launch angle may not have the proper height to be able to make it out of the stadium with high probability, but as launch angle increases, we should expect the probability to increase as well. This logic only extends up until some threshold value, since at a certain point, the hitter is hitting the ball at such a steep angle that no matter how hard he hits the ball, it will never be a home run. Thus, a negative quadratic relationship between log odds of home run probability and launch angle seems reasonable.

7 RESULTS

Now that we see that the relationships in our model match our general intuition, we can move on to observing the bivariate local logistic regression of home run probability on launch angle and exit velocity.

Figure 6 shows us the estimated regression function for both 2015 and 2017. We can see that in both years, a ball hit with exit velocity greater than or equal to 105 mph and with a launch angle between 25 and 35 degrees has a high probability of being a home run ($p > 0.9$). We can see that in 2017, the range of values where exit velocity is between 97 and 102 mph with launch angle between 25-35 degrees has a home run probability in the range of 0.35 to 0.45. In 2015, however, the range an exit velocity in the range of about 98 to 101 mph with launch angle between 25-35 degrees has a similar probability. In fact, for many different regions in these plots, the probability is shifting upward from 2015 to 2017. In 2015, the estimated regression function looks quite symmetric, while the regression function in 2017 starts to look a little bit warped. The key differences in the plots are in the places where home run probability was between 0.35 and 0.45 in the plot of 2015. In 2017, these regions either condense while the $p = (0.5, 0.8)$ region expands in its place or the $p = (0.35, 0.45)$ region expands in the place where there was only $p = (0.15, 0.25)$ before. If something had systematically changed in baseball, this would probably be where you would expect to see a difference in probability.

Figure 7 is the final result of this study. This plot was made by calculating the predicted values of the regression function for both years and for each point on the prediction grid, then subtracting the 2015 value from the 2017 value. In this plot, we can see that in the heart of the plot, there seems to be a systematic difference in terms of larger probabilities, especially on the fringes of ”extreme” launch angles (angles ¿ 35 degrees) and balls hit under 105 mph.



(a) 2015 (b) 2017

Figure 6: Heatmaps of 2015/2017 Regression Functions

The large swath of purple in the top right corner, up around 110 mph and a launch angle of 50 mph, can generally be attributed to a lack of data in this area. In the 2015 dataset, there were only 13 batted balls with a launch angle greater than 45 degrees and an exit velocity larger than 105 mph.

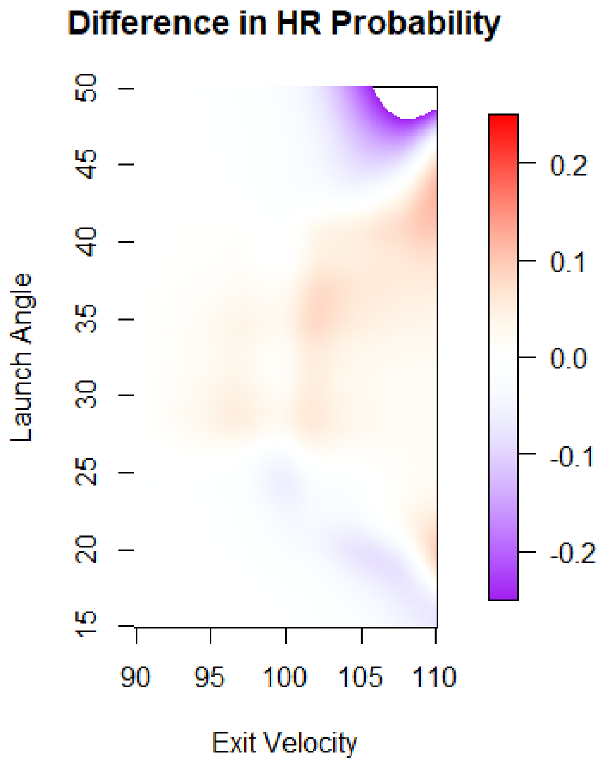


Figure 7: Change in HR Probability from 2015 to 2017

Possibly the most interesting segment of Figure 7 is the patch of negative values down around launch angles of 20 degrees and at exit velocities over 105 mph. This is not at all intuitive since we would expect that hard hit balls will be especially likely to be home runs if there has been a change in the dynamics of the ball. We cannot simply attribute this negative difference to limited data either, as there are 635 batted balls with launch angle between 15 and 20 degrees and an exit velocity larger than 105 mph in 2015 alone.

The plot of the difference in regression functions seems to show that there is a mild, but systematic difference in the regression functions from 2015 to 2017. In the heart of the plot, where most of the data lie, there is a clear trend of increased probability of home runs relative to exit velocity and launch angle.

8 DISCUSSION

The ultimate goal of our study is to analyze if there is some difference in the home run probability with respect to exit velocity and launch angle in the hopes of pointing towards a change in the dynamics of the ball as a reason for the increase in home run probability from 2017 to 2015. A shift in exit velocity could potentially be attributed to a change in the ball as well, since if the ball is more "bouncy", it would theoretically be hit harder given an equally strong swing and pitch speed.

In this study, we have shown that the probability of home runs given exit velocity and launch angle has increased systemically within the range of "usual values" of the data. Since this is a nonparametric analysis, finding if this is a statistically significant increase is not exactly intuitive and is avoided for now due to the nature of this study.

We cannot pretend to assign a reason to this change in home run probability, but by controlling for the two factors hitting that a player can control, namely launch angle and exit velocity, we hope to have found a direct way to point towards a theorized change in the ball from 2015 to 2017. An increased home run probability for an equal launch angle and exit velocity leaves very few alternative covariates that may affect home run probability on a scale as large as half of a season of MLB data.

We also notice that the distribution of exit velocities in 2017 has shifted notably to the right, towards larger values, as compared to 2015. This may also be indicative of a change in the make-up of the ball as described above. This result is slightly less clear, however, since there are many other covariates that may affect exit velocity by itself, namely the speed at which the pitcher throws the ball, the overall strength of hitters, etc. One could potentially point to this as possible evidence that players are either swinging harder, using performance enhancing drugs, or that players are simply getting stronger naturally. In conjunction with the local logistic regression results, however, it seems that this increase in exit velocity may be a point of interest if future researchers are to determine if there is a notable change in the make-up of the ball.

In our analysis, we have also analyzed the smooth additive regression functions of log odds of home run probability on exit velocity and launch angle. In both the 2015 and 2017 datasets, we can see that the smooth exit velocity function seems to be approximately linear, while the smooth launch angle function seems to be approximately quadratic. The purpose of this study was primarily to explore this dataset using nonparametric regression techniques. One of the limitations of this methodology is that inference is both difficult and lacks interpretability in many cases. Future studies may be able to use the fact the approximate functional forms of the covariates to model home run probability on launch angle and exit velocity using parametric generalized linear models. One of the main advantages of using parametric techniques is in gaining the ability to perform more straightforward inference on the relationships between the covariates and the data. In fact, the main interest in this study is if there is a significant difference between the bivariate regression function in 2015 as compared to that of 2017. This would all be possible under parametric models and should likely be the primary interest of studies going forward.

The ultimate goal of this study is to inspire further research into the physiological differences in baseballs from 2015 to 2017. Lindbergh and Lichtman (2017) did some very basic and preliminary research into possible physiological differences and found that the balls from after the ASG in 2015 were bouncier and had smaller seams than those from before the ASG in 2015. Arthur and Dix (2018) in conjunction with the Keck School of Medicine at University of Southern California as well as the Department of Chemistry and Biology at Kent State University have done some follow-up research using x-rays to examine the inside of the baseballs. Their research has determined that there are some minor differences in the density of the cores of baseballs from 2015 to 2017. Nathan (2017) presented an argument for how these changes could affect the flight of a baseball enough to truly change the likelihood that a well-hit fly ball may turn into a home run. My hope in this research is to further the direction of the study and show that accounting for variables that hitters may typically be able to control, we still see an increase in home run probability across Major League Baseball from 2015 to 2017.

REFERENCES

[1] "League by League Totals for Home Runs." *The Baseball Almanac*. Retrieved March 16, 2018. (http://www.baseball-almanac.com/hitting/hihr6.shtml)

[2] Arthur, Rob and Tim Dix. 2018. "We X-Rayed Some Baseballs. Here's What We Found." *538*. Retrieved March 16, 2018. (https://fivethirtyeight.com/features/juiced-baseballs/)

[3] Arthur, Rob and Ben Lindbergh. 2016. "A Baseball Mystery: The Home Run Is Back, And No One Knows Why." *538*. Retrieved March 16, 2018 (https://fivethirtyeight.com/features/a-baseball-mystery-the-home-run-is-back-and-no-one-knows-why/)

[4] Lindbergh, Ben. 2017. "As Home Run Rates Rise, MLB Offers Evidence That the Ball Isnt Juiced" *The Ringer*. Retrieved March 16, 2018. (https://www.theringer.com/2017/5/9/16040456/2017-mlb-home-run-rate-is-the-ball-juiced-report-results-6e1dd0233203)

[5] Lindbergh, Ben and Mitchell Lichtman. 2017. "The Juiced Ball is Back". *The Ringer*. Retrieved March 16, 2018. (https://www.theringer.com/2017/6/14/16044264/2017-mlb-home-run-spike-juiced-ball-testing-reveal-155cd21108bc)

[6] Mitchell, George and Waxman. 2008. "'The Steroids Era': George Mitchell on Drugs in Baseball." *The New Atlantis*. 19:pp. 139-141

[7] Nathan, Alan. 2017. "Fly Ball Carry And The Home Run Surge." *The Hardball Times*. Retrieved March 19, 2018. (https://www.fangraphs.com/tht/fly-ball-carry-and-the-home-run-surge/)

[8] Wilman, Daren. "Statcast Search." *Baseball Savant*. (https://baseballsavant.mlb.com/statcast_search)

9 APPENDIX A: R CODE

require(KernSmooth) require(fields) require(SemiPar) require(mgcv) require(plot3D)

#all data is with launch angle ¿ 10, exit velocity ¿ 75 mph because no HR occurred at less than 75 mph #or less than 10 degrees

data.2015.hr = read.csv(”C:/Users/ryanh/Desktop/255T datasets/2015 Before ASG HR.csv”) data.2015.hr = cbind(data.2015.hr, ”HR” = as.integer(rep(1, length(data.2015.hr$launch_speed))))

data.2015.nohr = read.csv(”C:/Users/ryanh/Desktop/255T datasets/2015 Before ASG No HR.csv”) data.2015.nohr = cbind(data.2015.nohr, ”HR” = as.integer(rep(0, length(data.2015.nohr$launch_speed))))

data.2015 = rbind(data.2015.hr, data.2015.nohr)

data.2015 = data.2015[which(data.2015$launch_speed ¡= 110),]

attach(data.2015)

#histogram of launch_speed h.speed.2015 = bw.nrd(launch_speed) f.speed.2015 = bkde(launch_speed, bandwidth = h.speed.2015, kernel = ”normal”) hist(launch_speed, freq = FALSE, breaks = seq(from = 70, to = 115, by = 2.5), xlab = ”Exit Velocity”, main = ”2015”) lines(f.speed.2015$x, f.speed.2015$y, lwd = 2)

#histogram of launch_angle h.angle.2015 = bw.nrd(launch_angle) f.angle.2015 = bkde(launch_angle, bandwidth = h.angle.2015, kernel = ”normal”) hist(launch_angle, freq = FALSE, breaks = seq(from = 0, to = 100, by = 5), xlim = c(0,100), xlab = ”Launch Angle”, main = ”2015”) lines(f.angle.2015$x, f.angle.2015$y, lwd = 2)

#do a bivariate kernel densiy estimate f.2d.2015 = bkde2D(cbind(launch_speed, launch_angle), bandwidth = c(h.speed.2015, h.angle.2015)) contour(f.2d.2015$x1, f.2d.2015$x2, f.2d.2015$fhat) persp(x = f.2d.2015$x1, y = f.2d.2015$x2, z= f.2d.2015$fhat, xlab = ”Exit Velocity”, ylab = ”Launch Angle”, zlab = ”HR Probability”, main = ”2015”, phi = 15, theta = -60, col = ”light blue”,ticktype = ”detailed”)

#2015

#additive model for 2015 fit.spm.2015 = spm(HR f(launch_speed) + f(launch_angle), family = ”binomial”) plot(fit.spm.2015)

#creating grid for to predict on angle.grid = seq(from = 15, to = 50, length.out = 500) speed.grid = seq(from = 90, to = 110, length.out = 500) grid = data.frame(cbind(”launch_angle” = angle.grid, ”launch_speed” = speed.grid))

heatmap.2015 = matrix(nrow = 500, ncol = 500)

#creating a matrix of predicted values

for(i in 1:length(angle.grid))

pred = predict(fit.spm.2015, newdata = data.frame(cbind(”launch_speed” = rep(grid[i,2], length(speed.grid)), "launch_angle" = grid[, 1]))) heatmap.2015[i,] = 1/(1+exp(-(pred)))

#image plot of additive model

image.plot(speed.grid, angle.grid, heatmap.2015, main = "2015 HR Probability", xlab = "Exit Velocity", ylab = "Launch Angle", col = rainbow(1000, start = 0, end = 5/6), breaks = seq(from = 0, to = 1, by = 0.001))

detach(data.2015)

#2017

data.2017.hr = read.csv("C:/Users/ryanh/Desktop/255T datasets/2017 Before ASG HR.csv") data.2017.hr = cbind(data.2017.hr, "HR" = as.integer(rep(1, length(data.2017.hr$launch_speed))))

data.2017.nohr = read.csv("C:/Users/ryanh/Desktop/255T datasets/2017 Before ASG No HR.csv") data.2017.nohr = cbind(data.2017.nohr, "HR" = as.integer(rep(0, length(data.2017.nohr$launch_speed))))

data.2017 = rbind(data.2017.hr, data.2017.nohr) data.2017 = data.2017[which(data.2017$launch_speed ¡= 110),]

attach(data.2017)

#histogram and kernel density estimate for exit velocity in 2017

h.speed.2017 = bw.nrd(launch_speed) f.speed.2017 = bkde(launch_speed, bandwidth = h.speed.2017, kernel = "normal") hist(launch_speed, freq = FALSE, breaks = seq(from = 70, to = 115, by = 2.5), xlab = "Exit Velocity", main = "2017") lines(f.speed.2017$x, f.speed.2017$y, lwd = 2)

#histogram and kernel density estimate for launch angle in 2017

h.angle.2017 = bw.nrd(launch_angle) f.angle.2017 = bkde(launch_angle, bandwidth = h.angle.2017, kernel = "normal") hist(launch_angle, freq = FALSE, breaks = seq(from = 0, to = 100, by = 5), xlim = c(0,100), xlab = "Launch Angle", main = "2017") lines(f.angle.2017$x, f.angle.2017$y, lwd = 2)

#joint density for 2017 f.2d.2017 = bkde2D(cbind(launch_speed, launch_angle), bandwidth = c(h.speed.2017, h.angle.2017)) persp(f.2d.2017$x1, f.2d.2017$x2, f.2d.2017$fhat, xlab = "Exit Velocity", ylab = "Launch Angle", zlab = "HR Probability", main = "2017", phi = 15, theta = -60, col = "orange",ticktype = "detailed")

#additive model for 2017 fit.spm.2017 = spm(HR f(launch_speed) + f(launch_angle), family = "binomial") plot(fit.spm.2017)

heatmap.2017 = matrix(nrow = 500, ncol = 500)

#matrix of predicted values for 2017 for(i in 1:length(angle.grid))

pred = predict(fit.spm.2017, newdata = data.frame(cbind("launch_speed" = rep(grid[i,2], length(speed.grid)), "launch_angle" = grid[, 1])))

heatmap.2017[i,] = 1/(1+exp(-(pred)))

image.plot(speed.grid, angle.grid, heatmap.2017, main = ”2017 HR Probability”, xlab = ”Exit Velocity (mph)”, ylab = ”Launch Angle (degrees)”, col = rainbow(1000, start = 0, end = 5/6), breaks = seq(from = 0, to = 1, by = 0.001))

#difference

hmap = heatmap.2017 - heatmap.2015

image.plot(speed.grid, angle.grid, hmap, col = designer.colors(n=1000, col = c(”purple”, ”white”, ”red”)), breaks = seq(from = -0.25, to = 0.25, length.out = 1001), main = ”Difference in HR Probability”, xlab = ”Exit Velocity”, ylab = ”Launch Angle”)