Explaining the MLB Home Run Record of 2017 with Quality of Pitch (QOP™)[1]

by: Jason Wilson[2], Joseph Lane[2], Jordan Wong[2], Jeremiah Chuang[2], and Wayne Greiner[1]

Summary: The two main explanations currently offered for the MLB home run record of 2017 are suspected changes in the manufacturing of the baseball and the new approach by hitters with increased launch angle and exit velocity. But what about the pitchers? Was there a change in pitch quality? Our Quality of Pitch (QOP™) statistic declined across all pitch types in 2017. We show that the drop in QOP™ average can be traced primarily to a change in two key components: vertical break and horizontal break. It is shown that the switch from SportsVision cameras to the Trackman doppler radar-camera data collection system is not sufficient to explain these changes. We conclude that the change in vertical break and horizontal break is a significant factor in explaining the record number of home runs being allowed by MLB pitchers in 2017.

1. Introduction

In 2017, MLB experienced the all-time record number of home runs (HR, 6105). It was a big jump from 2016 (5610), which was already a spike from 2015 (4909). The record beaten was 5693 HRs in 2000 (see Figure 1 and Table 1). Since our previous work has shown correlation between QOP average[3] (QOPA) and HRs (see Figure 1), we wanted to see if QOP™ could shed light on the much discussed 2017 results.

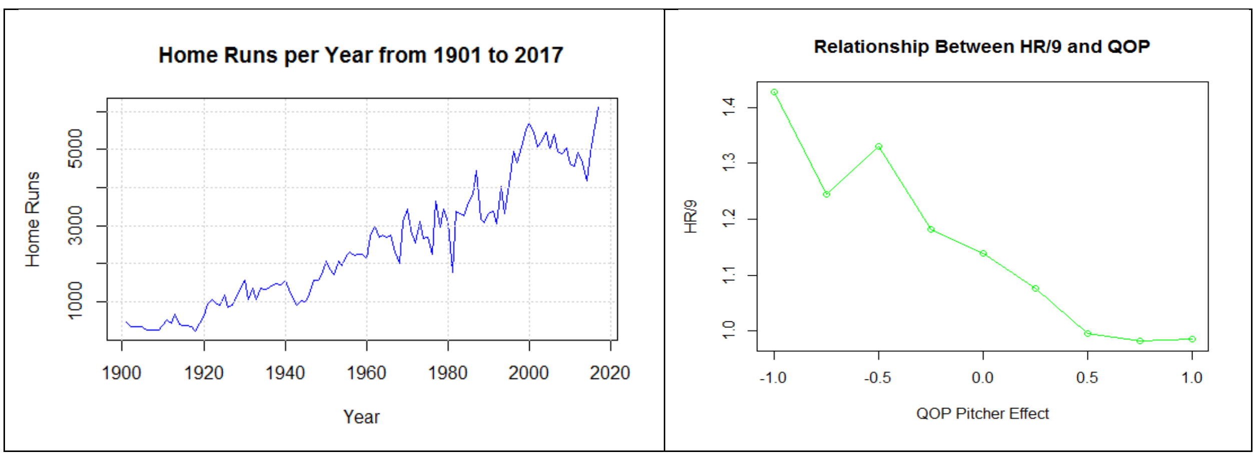


Figure 1. Home runs per year from 1901 to 2017[4] and Relationship between HR/9 and QOP[5].

---

[1] © Greiner Agencies Inc., 1478 Welbourn Drive, Edmonton, Alberta, Canada, T6M 2M1

[2] Biola University, 13800 La Mirada, CA 90639; jason.wilson@biola.edu; joseph.a.lane@biola.edu; jordan.t.wong@biola.edu

[3] The QOP value (QOPV) is calculated from the rise, breaking point, vertical break, horizontal break, location, and speed of a single pitch. The scale is roughly 0 to 10 with the larger the value, the better the pitch. The annual league average (QOPA) is around 4.5, with median median 5. For details see www.qopbaseball.com.

[4] http://www.baseball-almanac.com/hitting/hihr6.shtml

[5] This graph is taken from our QOP applications paper, https://www.fangraphs.com/tht/measuring-the-quality-of-a-pitch/. The x-axis is the QOPA, adjusted by pitch type, pitch count, runners on base, and times through the order. The point is that QOPA is negatively correlated with HR.

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
HR	4878	5042	4613	4552	4934	4661	4186	4909	5610	6105
$HR_{Year}/HR_{PrevYear}$	0.984	1.034	0.915	0.987	1.084	0.945	0.898	1.173	1.143	1.088

Table 1. Home runs (HR) per year during regular season.

The two main explanations in the literature for the HR increase are (i) change in ball manufacturing ("juiced ball")[6] and (ii) change in batter approach[7]. Irrespective of the change in (i) ball and (ii) batter, we propose a third factor influencing the HR increase, namely a reduction in pitch quality as calculated by our QOP™ metric[8]. It should be noted that ball changes may have effects on either the batter (e.g. exit velocity) or pitcher (e.g. grip). Also, batter changes may effect the pitcher (e.g. pitching higher in the zone to combat the upward swing). Irrespective of the precise cause of the pitch quality change, in this paper we will attempt to show that a sharp reduction in vertical break and change in horizontal break from 2016 to 2017 is one significant factor that explains the record number of home runs allowed.

---

[6] We believe there is not sufficient evidence to believe that the juiced ball has been the sole cause for the uptick in homeruns. While the ball has changed little from year to year, the microevolution that it has gone through cannot be proven to have a causational relationship with the amount of homeruns year to year. One of the things that was said by Cork Gaines and Skye Gould is that, "If the baseball's suddenly changed at the 2015 All-Star break, we would expect home runs to spike immediately and then level off, but that is not what we have seen. In fact, home run rates have seemingly been increasing at a steady rate the last three seasons — and possibly longer if we consider 2014 the odd season and not 2015 — and are still going up." (http://www.businessinsider.com/mlb-home-runs-something-is-juiced-2017-7) This idea supports the fact that this change of the ball could not have taken place instantaneously - it must have been progressive. There was also a bump in the drag coefficient for the baseball from the beginning of the season in September to the ball used in the playoffs. However, this increase caused a reduction of distance for fly balls by 11.5 feet. (https://fivethirtyeight.com/features/the-world-series-baseballs-sure-seem-juiced-but-are-they/) So while there may have been minor changes made to the ball over the past couple of years, there is neither sufficient nor significant evidence pointing towards the ball being the only reason why there were so many more homeruns this past year than in previous years. Furthermore, commissioner Manfield's assertion of no change (https://www.freep.com/story/sports/mlb/tigers/2017/08/22/rob-manfred-baseballs-juiced/592015001/), supported by Alan Nathan's affirmative review of the baseball testing report (https://www.theringer.com/2017/5/9/16040456/2017-mlb-home-run-rate-is-the-ball-juiced-report-results-6e1dd0233203) is an important consideration. Pitching is an important factor, as we seek to show in this paper.

[7] The game of baseball has evolved - it has become a game of power, a game of fast arms and even faster swings. The influx of power batters over the past year is attested by the fans, the teams, and even the players themselves. (https://sports.yahoo.com/rob-manfred-fans-say-love-home-runs-strikeouts-041414089.html) The desire to see HRs and accumulate HRs on stat sheets has become a driving force within the Major Leagues. As the batters have become bigger and stronger, some pitching coaches have also been coaching a newer approach: the upward swing - all or nothing. As the number of homeruns has increased, so has the number of strikeouts and doubles. See https://sports.yahoo.com/rob-manfred-fans-say-love-home-runs-strikeouts-041414089.html and https://www.nytimes.com/2017/07/11/sports/baseball/rob-manfred-all-star-game-news-conference.html.

[8] For an overview of QOP™, a good place to begin is https://www.fangraphs.com/tht/measuring-the-quality-of-a-pitch/ or https://en.wikipedia.org/wiki/Quality_of_Pitch. We released QOP to the public with a case study on the Dodgers at the SABR 2015 conference, https://qopbaseball.wordpress.com/. Other information may be found at our website, including searchable QOP averages, www.qopbaseball.com.

There is one confounding factor with QOP, however. In 2017, MLB switched from the SportsVision camera system for data collection which had been used since 2008 to the Trackman Doppler radar system[9]. Both are available via the PITCHf/x data feed. There is no publicly available data with SportsVision and Trackman numbers, in order to perform a comparison. Therefore, any such differences are confounded with the real pitch signal from 2017. In Section (2) we analyze the differences and propose an explanation for how the differences observed in QOP explain the increase in HR. In Section (3) we provide evidence that the differences observed from 2016 to 2017 are not explainable from the switch from SportsVision to Trackman alone; the drop in QOPA is real. We interpret the evidence in Section (4) and conclude in Section (5).

Year	Pitch	All	CH	CU	FF	FT	SI	SL
2017	qop Max	9.91	9.35	8.91	9.91	9.71	9.67	9.13
	qop Avg	4.56	4.35	4.36	4.70	5.03	4.97	4.11
	NP	729,396	73,202	61,708	260,069	102,405	43,931	119,895
2016	qop Max	9.99	9.11	9.08	9.66	9.91	9.99	8.87
	qop Avg	4.59	4.37	4.40	4.82	5.08	5.08	4.24
	NP	715,245	73,372	62,305	258,726	95,847	48,133	108,807
2015	qop Max	9.90	8.80	9.13	9.76	9.66	9.90	8.90
	qop Avg	4.58	4.32	4.36	4.81	5.10	5.05	4.22
	NP	712,273	75,688	54,158	255,565	92,489	57,211	103,159
2014	qop Max	9.75	8.83	9.30	9.71	9.75	9.75	9.09
	qop Avg	4.57	4.34	4.52	4.75	5.08	5.10	4.21
	NP	708,663	73,207	58,169	243,028	94,649	63,655	101,178
2013	qop Max	10.00	8.84	9.35	10.00	9.63	9.69	9.13
	qop Avg	4.57	4.34	4.61	4.74	5.08	5.08	4.24
	NP	720,217	72,968	62,216	253,062	96,194	60,891	110,982
2012	qop Max	10.03	9.36	9.02	9.99	9.73	10.03	8.99
	qop Avg	4.57	4.29	4.65	4.73	5.09	5.10	4.26
	NP	723,185	73,427	65,565	245,513	90,358	73,545	110,055
2011	qop Max	10.21	9.00	9.15	9.62	9.76	10.21	8.89
	qop Avg	4.47	4.24	4.64	4.61	5.01	5.01	4.21
	NP	717,060	73,924	59,068	238,933	83,149	83,855	111,262
2010	qop Max	10.31	8.83	9.24	9.50	9.76	10.31	9.10
	qop Avg	4.46	4.16	4.52	4.63	4.99	5.03	4.12
	NP	737,143	78,678	60,453	241,957	85,912	97,644	107,287
2009	qop Max	9.98	9.49	9.01	9.71	9.84	9.98	9.08
	qop Avg	4.51	4.25	4.58	4.66	5.02	5.10	4.17
	NP	711,945	69,474	58,376	243,332	80,601	94,594	106,801
2008	qop Max	10.07	9.11	9.00	9.78	10.07	9.84	9.15
	qop Avg	4.47	4.26	4.60	4.62	5.00	5.00	4.20
	NP	702,619	69,244	56,390	238,225	73,995	102,170	104,175

Table 2. Historical QOP averages (QOPA). QOPA has dropped for all pitch types in 2017. For the same stats for all pitch types, see www.qopbaseball.com. The differences in QOPA between 2016 and 2017 are all statistically significant with p-values of 10^{-11} or less.

[9] http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8074554

2. The Pitch Characteristics of 2017

Figure 2 shows the means of the six components used to calculate QOP, from 2008 to 2017[10]. Although the exact formula for QOP is proprietary, the formula goes like this:

$$QOP = -Rise + Breakpt + Tot.brk + H.brk2 - Loc + Speed$$

Rise, breaking point (Breakpt), vertical break (Tot.brk), and horizontal break (H.brk2) are measured in feet. Location (Loc) uses a non-liner function of the distance from the corners of the strike zone, where the farther from the corner, the larger the value[11]. Velocity (Speed) is measured in MPH[12]. Rise and Loc are negative coefficients and when increased are considered to decrease pitch quality.

The middle line in Figure 2 is the mean of the data and the upper and lower limits (UCL and LCL) are the mean +/- three standard deviations. Not only is the historic pattern displayed, but these are formal control charts[13]. They show the relationship between the annual means, allowing one to see what is within historic range and what is extreme. Because we have so few points, we have split the data and produced the same graphs for the first and second halves of the season in Appendix C. They confirm the primary observations of Figure 2.

[10] All calculations were done using R, www.r-project.org.

[11] Loc values are roughly on a scale from 0 to 4+.

[12] Velocity is the speed of the pitch at 50', which is what PITCHf/x reported from 2008 to 2016. In 2017, confusingly, that column of data, start_speed, was silently switched to the speed at the release point of the pitch, which is usually around 55', https://www.baseballprospectus.com/news/article/31529/prospectus-feature-estimating-release-point-using-gamedays-new-start_speed/. This change led to unnecessary criticism of Trackman early in 2017, https://www.fangraphs.com/blogs/about-all-these-velocity-spikes/. For this paper, we have adjusted the 2017 start_speed back to 50', for comparability, using the appropriate formula: start_speed = sqrt(vx0^2 + vy0^2 + vz0^2)*3600/5280, where the constant converts feet/sec to miles/hr.

[13] There is a slight adjustment used when calculating standard deviations for control charts. All control charts for this paper used R's *qcc* package with the number of standard deviations set to 3, which is a customary number. Since we were examining potential changes in 2017 from 2008 to 2016, we set it to use the latter as "calibration data" and treat 2017 as "new data", as shown on the graphs. Scrucca, L. (2004). qcc: an R package for quality control charting and statistical process control. *R News* 4/1, 11-17.

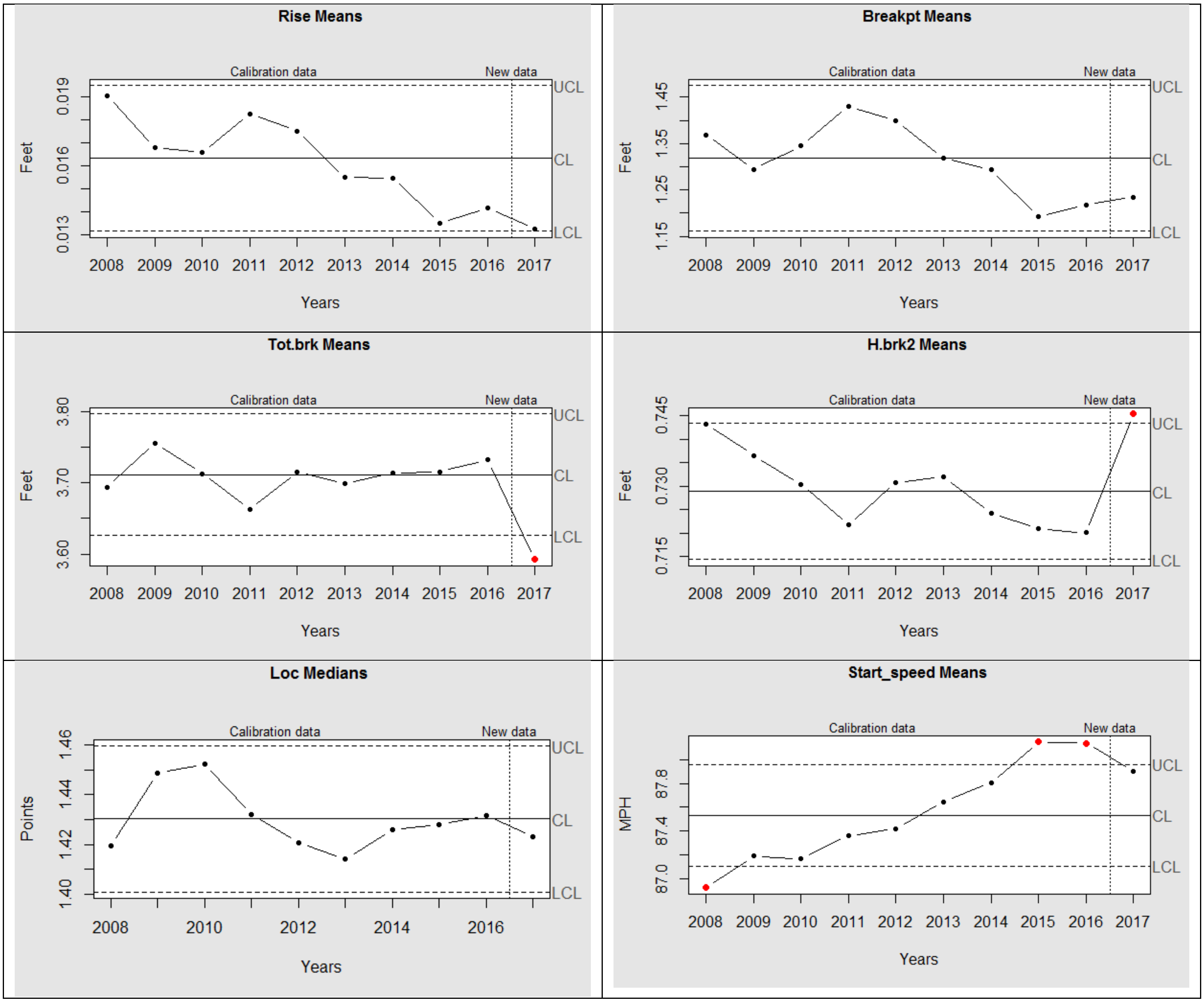


Figure 2. Graphs of the change in the six pitch components from 2008 to 2017. These are formal control charts. There are no error bars/confidence interval bars shown on the graphs because the bars are the size of the dots, due to the enormous number of pitches.

The following table attempts to summarize the salient observation of the trends in the mean[14] of the components.

[14] The mean was used for all statistics except Loc. We used median for Loc, because there are some extreme location values due, for example, to pitch outs and intentional walks.

	Rise	BreakPoint	Tot.brk	H.brk2	Location	Speed
2008-16 Trend	Strong Decline	Moderate Decline	Relatively Flat	Modest Decline	Relatively Flat	Moderate Incline
2017 Feature	On Trend	On Trend	Sharp Drop, Historic Min	Sharp Increase, Historic Max	On Trend	On Trend

Table 3. Primary observations about changes in 2017 pitch components.

The most consistent trend is a steady increase in pitch velocity from 2008 to 2017, although this tapers the past two seasons. A close look reveals that rise simultaneously has a decreasing trend. Two components have flagged extrema for 2017: Tot.brk and H.brk2. Our thesis is that in the wake of the increase in home runs allowed in 2016, possibly due at least in part to the ball or batter approach, pitchers made adjustments in 2017 in an attempt to reduce launch angle and exit velocity. The result, however, is a loss of some vertical break and horizontal positioning. In the QOP formula, increased Speed and H.brk2 add to QOP while decreased Tot.brk subtracts. In these competing formula components, the most dominant change is Tot.brk, resulting in an overall reduction in QOP. The fact of the relationship between pitch components and HR will be demonstrated using an explanatory logistic regression model in Section 4. We propose that with less vertical movement batters have had a narrower range within which to successfully connect with the ball, resulting in more home runs.

The evidence of a change in vertical and horizontal break in 2017 is very strong. We propose that it is this change that results in lower average QOP values for 2017. This raises the question, "Why did the vertical and horizontal break change?"

3. SportsVision vs. Trackman

In Section 1, we mentioned the common theory that changes to the baseball and the approach by hitters has resulted in the increase in home runs. In Section 2, we showed that additional factors that were substantially different in 2017, namely the vertical and horizontal break of the pitches. Our thesis is that the overall quality of MLB pitches decreased and this was a contributing factor to the increase in home runs in 2017. Unfortunately for our thesis, in 2017 MLB Advanced Media switched from the old SportsVision camera only pitch tracking system (2008 to 2016) to the newer Trackman system which uses a combination of cameras and doppler radar. This raises the possibility that the observed changes in vertical and horizontal break may be merely an artifact of the different tracking system. Therefore, in order for our thesis to be believable, we need to be able to know that changes in 2017 are due, at least in part, to the actual pitches thrown, and not only differences between systems. In this section, we offer five different lines of evidence.

3.1 Source of Data

Trackman has been fully operational in MLB stadiums since 2015[15]. The ideal way to settle the issue would simply be to look at the SportsVision and Trackman measurements from 2015 and 2016 and directly compare them to see if the vertical and horizontal break measurements match. However, the Trackman data has not been publicly available, for whatever reason.

[15] http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8074554

We do not believe that MLB Advanced Media would make the change of system unless they were confident that the results would be comparable. They had two full years of testing and their practice of making data publicly available as well as permitting their analysts to transparently interact with the public[16] provides assurance. Furthermore, our informal conversations with analysts who have access to this data have found they have general confidence in the reliability of the data. This is an important difference from the roll-out of SportsVision's system in 2007-2008. Indeed, the credibility of both MLB and Trackman is on the line over the reliability of their data. While no company is perfect, they have a good record and one should accept the data unless the evidence indicates otherwise.

3.2 Results inconsistent with bias or difference in variation

If there was a significant change in Trackman measurements from the SportsVision measurements, it should be a change with certain properties. In particular, it may be biased, or have a difference in variation. Bias is a systematic mismeasurement in one direction, so bias would appear, e.g. as usually over (or under) estimating speed, or usually measuring the trajectory shifted in a particular direction (e.g. high or low, left or right). The other property a change may be observed in is variation, which is the amount of spread in the data (If you were to have five Trackman systems in place and take five measurements of the same pitch, how close would they be to one another?).

The pitch characteristics in the previous section do not show either bias or a difference in variation across parameters. Tot.brk has a noticeable decrease, whereas H.brk2 has a noticeable increase. If the doppler radar were over- or under-estimating break, then we would expect to observe Tot.brk and H.brk2 to be either both increasing or both decreasing. Furthermore, Breakpt follows the trend. Regarding variation, all of the parameters in 2017 have similar variation as 2008 to 2016 except H.brk2, which is noticeably larger. However, even that change is consistent with its signal to noise ratio and should be expected (see Figure 3). If there really was a substantial change in bias or variation due to Trackman, it is hard to conceive why the differences observed would obtain. Furthermore, this point is seen even more strongly in the horizontal break graphs in Appendix D where the horizontal break of right-handed versus left-handed pitchers clearly has different properties for different pitch types.

What about the analysts who have written on differences between SportsVision and Trackman? Most of these articles complain about gross mismeasurements of individual pitches[17]. This is not a problem because they are isolated incidents and many of these are corrected in the final version of the data[18]. Like the SportsVision data of previous years, there are undoubtedly errors[19].

The strongest article alleging unreliability of the Trackman data was by Rob Arthur in FiveThirtyEight[20]. His first main argument was from Kyle Boddy, the President and Founder of DrivelineBB, who reported an increase in horizontal and vertical break error of about 0.2 and 0.3 inches, respectively[21]. While these

[16] MLB analyst Tom Tango's blog is quite candid http://tangotiger.com/index.php/site/comments/pitch-velocity-new-measurement-process-new-data-points#36
[17] E.g. https://www.fangraphs.com/blogs/about-all-these-velocity-spikes/
[18] The way we know there are corrections in the data is because there were some slight changes in our QOPAs using data accumulated daily in 2017 compared to the one re-downloaded in December.
[19] The worst one we noticed this year was an anomaly recorded at 104.4 MPH whereas the xyz velocity components put it at 66.2 MPH.
[20] https://fivethirtyeight.com/features/baseballs-new-pitch-tracking-system-is-just-a-bit-outside/
[21] Ibid. 0.2 and 0.3 are read from the Arthur's graph. Kyle Boddy's site is: https://twitter.com/drivelinebases

are meaningful, even if the differences observed in Figure 1 were shifted by these measurements, they would not be enough to overcome the differences observed in the data, which are 0.36 and -1.68 inches, respectively:

$$H.brk2_{diff} = (0.75 - 0.72)ft.\times\frac{12\ in.}{1\ ft.} = 0.36\ in$$

$$Tot.brk_{diff} = (3.59 - 3.73)ft.\times\frac{12\ in.}{1\ ft.} = -1.68\ in$$

The differences would reduce to 0.36-0.20 = 0.16 and -1.68+0.30=-1.38 inches, which are still substantial, particularly Tot.brk. The main arguments of this paper would still hold even with these adjusted differences. The reason is that the observed differences are substantially larger than the alleged measurement errors. Furthermore, if the measurements were completely biased, as the above calculations assumed, then this bias due to error would be removable, leading to more accurate results[22]. Finally, even if these errors were granted, it is more likely that they would be random (i.e. less precision in Trackman doppler radar than SportsVision camera measurements). If so, then they would average out and the league and player means would still be accurately measured, leaving the main arguments of this paper intact.

Arthur's second main argument was that the vertical break error varied by stadium. He writes, "…some ballparks show much larger errors than others. So far this season, Atlanta's brand-new SunTrust Park appears to have the most accurate vertical break numbers, only off by two-tenths of an inch on average. Meanwhile, Cincinnati's Great American Ball Park shows the worst errors, missing by an average of 2.4 inches per pitch. So not only are the errors bigger than in the days of PITCHf/x, they're also more inconsistent: Last year, every park's errors ranged from 0.04 to 1.4 inches[23]." First, the same remarks about whether the error is random or biased from the first argument apply here. There is not enough information provided to know the overall average error. Second, the dates reported are for April only, the first month of the season. This does not reflect calibration during the year, or data correction, both of which occurred. Third, it is not clear whether some numbers reported are from SportsVision or Trackman.

3.3 Signal to Noise Ratio

One way to detect change in a process is to look at the signal to noise ratio. If the signal improves, or the noise decreases, while the other stays the same, then this ratio will increase. Conversely, if the signal worsens, or the noise increases, while the other stays the same, the ratio will decrease. In our case, the signal is the mean and the noise is the standard deviation. The quotient of mean and standard deviation is the coefficient of variation, which we will use to measure the signal to noise ratio (see Figure 3).

[22] The way to remove bias is to add the biased amount to the individual values, which is only appropriate when the bias is known. We appreciate the Kyle Boddys and Alan Nathans out there who alert us public analysts to these issues. Given the graphs in this paper, it is plausible that some of the Tot.brk and H.Break differences could turn out to be Trackman bias. We stress, however, that the observed differences are still substantially different than the alleged bias, leaving the arguments of this paper in play.

[23] https://fivethirtyeight.com/features/baseballs-new-pitch-tracking-system-is-just-a-bit-outside/

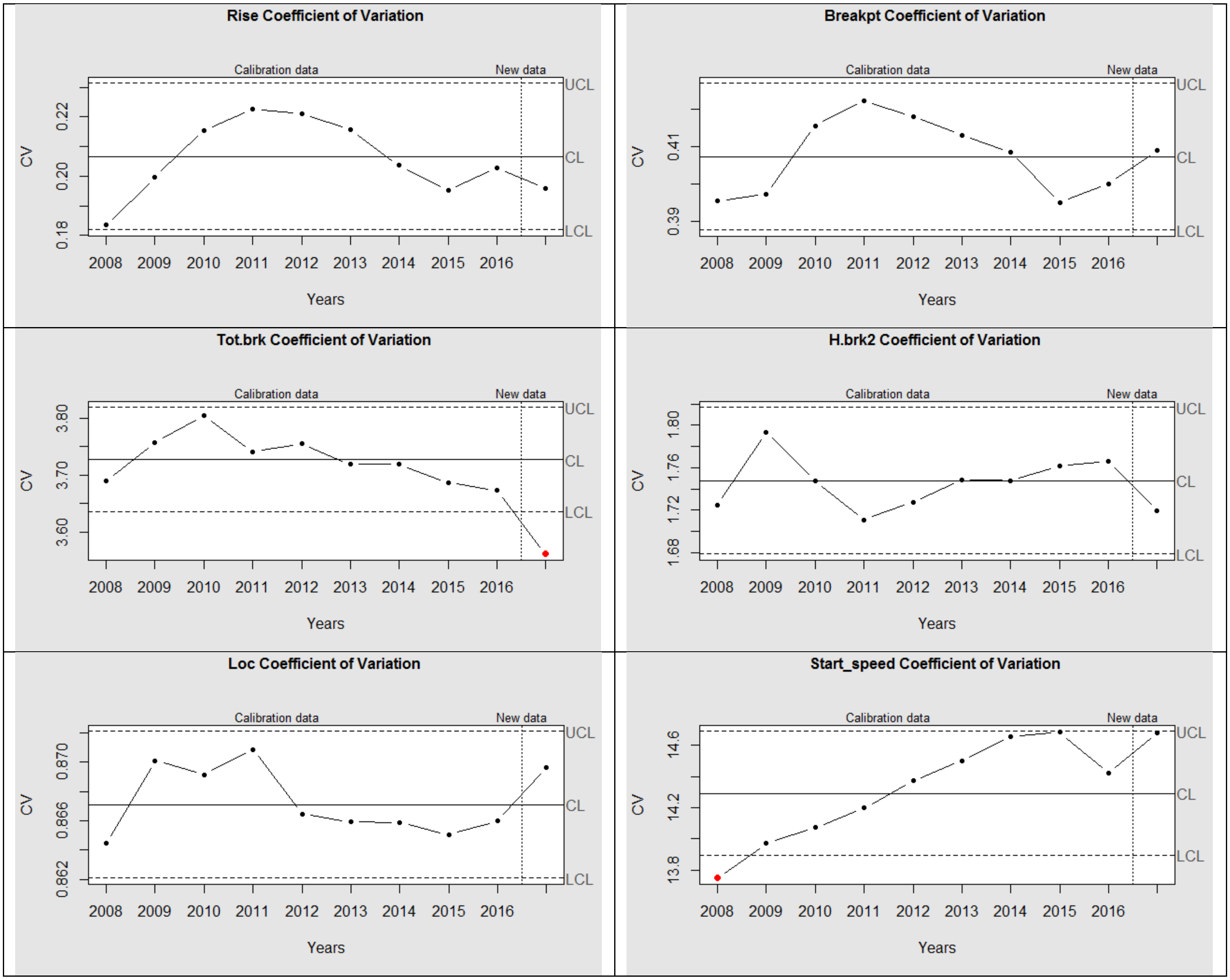


Figure 3. Signal to noise graphs: coefficient of variation for the 6 pitch components[24]

In Figure 3 it is clearly seen that the signal to noise ratio has remained relatively constant, with two exceptions. Vertical break stayed in a nice channel except in 2017 the signal to noise ratio took a sharp drop. If Trackman induced greater error or variation, we would expect the opposite – that it would increase. All components remain within their channel except for velocity, whose signal to noise ratio has increased. Since we know that velocity has increased, this tells us that the variation in speed has

[24] We used the coefficient of variation for all statistics, except location, where we used the median to the median absolute deviation (MAD), due to some extreme values in the location data.

remained relatively constant while the mean velocity increased. Taken together, this implies that if there was a change due to Trackman, it is not due to an increase in the variability, or loss of signal.

3.4 Regression

If the Trackman system increased the variability in the measurements, this should be recorded in the sums of squares of a suitable regression model. In order to test this, we built a model that explains the proportion of hits which result in home runs in a season based on batter and pitcher variables, along with team. In particular, the model is

HR% ~ Team + Pitch.Type + Pitch.Type% + Swinging.Strike% + Batter.Height + Rise + Breakpoint + Vertical.Break+ Location + Horizontal.Break + MPH

This model was built for each year for the seven pitch types with the most pitches, which resulted in 30 teams * 7 pitch types = 210 rows of data per year[25]. The key regression statistics for each model are contained in Table 2.

	Year	RSE	R^2-Adj	p-value	SSE
1	2008	0.00202	0.23491	0.00002	0.00068
2	2009	0.00245	0.17180	0.00106	0.00099
3	2010	0.00187	0.15254	0.00289	0.00058
4	2011	0.00264	0.34158	0.00000	0.00115
5	2012	0.00185	0.28584	0.00000	0.00056
6	2013	0.00189	0.27063	0.00000	0.00059
7	2014	0.00168	0.22113	0.00006	0.00047
8	2015	0.00214	0.10571	0.02412	0.00076
9	2016	0.00234	0.19798	0.00024	0.00090
10	2017	0.00238	0.26702	0.00000	0.00094

Table 4. Regression Statistics. RSE = residual standard deviation. R^2-Adj is the usual R^2 proportion of total variability explained but penalized for adding model parameters. p-value is the overall model p-value. SSE is the sum of squared error.

All of the models have some explanatory power. Frankly, they are not great, but the point is they use a diverse set of potentially relevant statistics and apply the miracle of linear algebra to extract the optimal fitting model, with a measurement of the variation. The result is that the variation (sum of the squared error, SSE=0.00094) of the 2017 Trackman model is small and it is well within the range of the variation of the other models (mean = 0.00076, 25^{th}% = 0.00058, 75^{th}% = 0.0093). Furthermore, its other statistics are all comfortably within the range of the statistics of the other models. There is no indication of a jump in variability of the data.

3.5 Pitcher subset study

The previous analyses in this section looked at league-wide analyses. In this section, we look at 13 individual MLB pitchers and examine their pitch components from 2015 to 2017. The goal is to identify differences and determine whether they are more likely due to Trackman measurement errors or pitching behavior changes.

[25] Throughout the paper, we used the six pitch types which account for approximately 90% of the MLB pitches: CH, CU, FF, FT, SI, and SL. This regression model was done at an early stage, and we had included a seventh pitch type, FC. Whether FC is included, or not, would not significantly change the results.

The manner in which the list of pitchers was chosen is as follows: Jason (statistician) asked Wayne (MLB pitcher expert) for a list of pitchers who met the following criteria:

1. "Steady" - no news of their having changed styles, are fighting injury, added pitch types, or otherwise may have different pitching behavior from 2015 to 2017
2. Name recognition
3. Starting pitchers[26]

We did not look at any other pitchers for the analysis in this section. We used the pitch classification given by the PITCHf/x data and we only selected pitch types of pitchers that had around 30 or more pitches of each type, per year, for 2015 to 2017. The analysis is extensive. Please take a few minutes to glance through the 49 pages of graphs linked in Appendix B to get a feel for it. Consider the following observations:

1. There are multiple instances of each combination of possible relationships between each distribution. For example, Marco Estrada's Change-up:
    a. Rise, BreakPoint, and Location are extremely close
    b. Vertical Break: Same shapes, but centers are ordered 2015, 2017, 2016
    c. Horizontal Break: Same Shapes, but centers are ordered 2016, 2015, 2017
    d. Speed: Same Shapes, but centers are ordered 2016, 2017, 2015

    → This is not consistent with added Trackman bias or increased variation
2. For vertical break, flipping through the graphs reveals that the distribution is pretty similar for each pitcher-pitch type.

    → This is not consistent with added Trackman bias or increased variation
3. For horizontal break, most pitchers-pitch types are the same. When they vary, some have 2017 below 2015 and 2016 (e.g. Christopher Archer - SL) and some have it above (e.g. Christopher Archer - FF).

    → This is not consistent with added Trackman bias or increased variation
4. Kyle Gibson's Slider (SL) Rise graph is clearly bimodal, and captured by all three years, with approximately the same center for each mode. By contrast, there are no stark bimodal graphs that are only for one year, but not the others, particularly 2017.

    → This is not consistent with added Trackman bias or increased variation
5. We conducted a Kolmogorov-Smirnov test of equality of distributions for each graph[27]. There are 25/294 = 0.085 of the graphs where the Kolmogorov-Smirnov tests finds 2017 statistically significantly different from 2016, and 2015, and where 2015 and 2016 are not statistically

[26] Wayne provided Jason with an initial list of 18 pitchers including different types from fastball throwers to knuckleball pitchers, those who rely on high pitch quality (high QOPA) and those who rely on deception (lower QOPA). Jason eliminated five pitchers for the following reasons: (1) Jose Quintana: added Sinker in 2017 (according to the PITCHf/x classification algorithm), (2) Miguel Angel Gonzalez: name didn't show in database, (3) Chris Sale: 495 FT's in 2017 with 0 in 2016, (4) Jeff Samardzija: Introduced KC in 2017, (5) Jeremy Hellickson: Introduced KC in 2017.

[27] The Kolmogorov-Smirnov test is the standard hypothesis test for testing whether two different distributions come from the same population.

significantly different at the 5% significance level. This means that 2017 is different from 2016 and 2015, but 2015 and 2016 are the same only 8.5% of the time. Of these, some would not be considered different by eye. For example, Justin Verlander's curveball vertical break is one of the differences.

6. Of the 25 pitcher-pitch type combinations where there was a statistically significant difference between 2017 against 2015 and 2016, their distribution is shown in Table 5:

Component	Rise	Breakpt	Tot.brk	Location	H.brk2	Speed	Sum
Count	1	3	8	1	7	5	25

Table 5. Distribution of the statistically significant differences between 2017 over against 2015 and 2016 using the Kolmogorov-Smirnov test of equality of distributions.

It is noteworthy that both the Tot.brk and H.brk2 are the highest frequency. Of the 8 Tot.brks, one is up, five are down, and two differ in shape. Of the 7 H.brk2s, four are up, two are down, and one differs in shape.

If there were no changes, using the 5% level of significance, we would expect

$$294*0.05*0.05*0.95 = 0.698$$

or about one pitcher-pitch type change. We have twenty-five changes, so something is happening. If the changes were due to a miscalibration of Trackman, we would expect a systematic effect, e.g. most vertical breaks down (or up). However, this is not what we see. There is at least one change for each component. While the most changes are for Tot.brk, H.brk2, and Speed, each component has some increasing and some decreasing. This does not rule out a Trackman effect, but it does suggest there is at least more going on than merely Trackman. That is – there is a real change in pitcher performance in 2017. Of those changes, the most predominant is a drop in vertical break and increase in horizontal break, but this is not across all pitchers nor across all pitch types for a particular pitcher.

To summarize, we have presented five lines of evidence for why the 2017 doppler radar measurements reported in the PITCHf/x data may be considered reasonably accurate and consistent with preceding years of 2008 to 2016: the data source is reliable; the results are inconsistent with bias or difference in variation; the signal to noise ratio is consistent; there is no variation increase in HR% regression model; and individual pitchers have opposing pitch characteristics.

4. Explaining the vertical break drop and horizontal break increase

Having argued that the primary pitching changes in 2017 are a drop in vertical break and increase in horizontal break (Section 2), and that these cannot be explained by the switch from SportsVision to Trackman measurement (Section 3), it remains to interpret the meaning of these changes. In this section, we address vertical and horizontal break, followed by the results of a model which successfully explains the number of home runs in terms of pitch components.

4.1 Vertical Break

It has been observed that pitchers were pitching higher in the zone in 2017[28]. According to QOP™ (see Table 2), there was a substantial drop in the quality of pitching in 2017. The drop in vertical break shown in Section 2 is the primary reason why QOP averages (QOPA) have dropped. Therefore, since QOPA and HR are negatively correlated (see Figure 1), we conclude that the drop in vertical break is one factor the increase in HR.

4.2 Horizontal Break

Horizontal break is not nearly as easy as vertical break. In our QOP™ model (Section 2), horizontal break adds to QOP. Since it increased in 2017, if everything else stayed the same, QOPA would go up. It turns out that the vertical break decrease outweighs the horizontal break increase.

But there is more going on. It may be that an increase in horizontal break for RR and LL matchups would result in the ball going farther up the barrel of the bat, resulting in better contact. It turns out that the splits of right-handed and left-handed pitchers in Appendix D clearly show an overall increase in h.brk2 by right handers, with a sharp jump in 4 of the 6 main pitch types and a drop in 2 of the main 6 pitch types[29]. The left handers is the opposite, though. Another way to see it is the ratio of the mean h.brk2 of one season to the next. This is shown in Table 6 where it can be seen that the largest percent increase in h.brk2 is by right handed pitchers in 2017/2016 and the largest percent decrease in h.brk2 is by left handed pitchers in 2017/2016. The next largest percent changes are both in 2011/2010, but the changes are opposite! Clearly there is something happening with horizontal break in 2017. In addition, according to Table 7, there is an increase in both home runs and non-home runs for the RR matchups. In conclusion, in 2017, some of the h.brk2 increase, since it has the highest proportion in the RR matchup, may actually contribute to better batter contact.

	17/16		16/15		15/14		14/13		13/12	
Pitcher	**Batter**									
	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left
Right	1.0739	1.0960	1.0016	0.9818	0.9922	1.0025	0.9852	0.9893	0.9936	0.9925
Left	0.9086	0.9505	1.0080	1.0574	1.0042	0.9679	1.0036	1.0078	1.0228	1.0225

Pitcher	12/11		11/10		10/09		09/08	
	Batter							
	Right	Left	Right	Left	Right	Left	Right	Left
Right	1.0205	1.0222	0.9688	0.9555	0.9915	0.9933	0.9996	1.0019
Left	0.9798	0.9811	1.0528	1.0697	1.0007	0.9733	0.9661	0.9704

[28] For example, pitching higher in the zone not working: https://www.fangraphs.com/blogs/pitchers-went-up-in-2017-and-it-didnt-work/. Pitching the four-seam fastball higher in the strike zone and throwing fewer sinking fastballs: https://sabr.org/latest/trueblood-sinker-doesnt-play-well-others. Cubs preparing for Dodger higher zone pitching: http://www.chicagotribune.com/sports/baseball/cubs/ct-spt-cubs-dodgers-high-pitch-strategy-20180221-story.html.

[29] Right-handed pitchers h.brk2 sharply jumped for CH, FF, FT, and SI, but dropped for CU and SL.

Table 6. H.brk2 ratios of handedness match-up for pitchers vs. batters. Cell entries are the mean h.brk2 of the first year divided by the second year. For example, for 17/16, 1.0739 means that the mean RR h.brk2 in 2017 was 1.0739 times what it was in 2016.

The work to tease out the nuances begins in the next subsection by addressing the relationship directly.

Handedness	2017		2016		2015	
	Non-HR	HR	Non-HR	HR	Non-HR	HR
LL	7.19% 52445	0.04% 322	7.20% 51514	0.04% 276	8.11% 57769	0.04% 293
LR	18.45% 134540	0.16% 1181	18.59% 133002	0.16% 1162	18.75% 133564	0.14% 991
RL	33.87% 247040	0.30% 2169	34.16% 244286	0.27% 1912	34.65% 246752	0.25% 1791
RR	39.65% 289183	0.34% 2516	39.26% 280802	0.32% 2255	37.79% 269144	0.27% 1922

Table 7. Proportion of home runs by pitcher-batter handedness matchup.

4.3 Model

The most difficult part of this analysis was separating the relationship between home runs, pitch type (and the changes in percentage of pitch type by year), pitch components, and handedness match-up. In the process of doing it, pitcher-batter handedness emerged as sometimes significant, but still the underlying relationships remained elusive. Finally, one reviewer gave us the breakthrough idea. He said what would be convincing to him was if we were able to construct a model that used the pitching components to successfully model home runs. Furthermore, he suggested a logistic regression model with Home Runs {Yes, No} as the response variable and the pitching components, “and anything else you want” as the explanatory variables. After some experimentation, we arrived at the following model[30]:

$$HR\% = intercept + rise + breakpt + tot.brk + h.brk2 + loc + start.speed + height + hand$$

where *height*[31] is the batter height and *hand* is the pitcher-batter match-up {RR, RL, LR, LL}. Algorithm 1 was used to generate the results.

1. Remove pitches with QOPV < 0
2. For each pitch type, randomly divide the pitches of one season into 1/2 test and 1/2 validation datasets
3. Construct a logistic regression model from the test data and apply the model to the validation data [for each pitch type]
4. Sum the explained HRs across pitch types
5. Construct a confidence interval for actual number of HRs

[30] We used an *intercept* and *hand* in order to obtain as much explanatory power as possible. Based on preliminary results, it is likely that results would be similar by dropping the *intercept* and *hand*. For some pitch types hand entered as statistically significant and others it did not.

[31] We used *height* because the taller the batter, the higher the strike zone, and therefore the higher the pitch, for the same height of pitcher. This turned out to be a significant variable in the models.

6. Set flag=1 if prediction is in CI, flag=0 if outside 7. Repeat steps 1-6 1000 times 8. Validation % = sum of flag divided by 1000
Algorithm 1. Logistic regression model, run 1000 times with 50% of 2017 data as test sample to construct model and 50% as validation sample.

The reason that more than one model is constructed is that if all of the data is used to build the model, it can fit the data perfectly. Therefore, we used one standard approach, which is to use half of the data to build a model which is used to explain the results in the other half of the data[32]. We built models from 1000 different random test and validation sets. The reader may view a set of sample models in Appendix G. The final result was that 83.7% of our test set models accurately explained their validation set model. By accurately explained, we mean that the predicted number of home runs fell within a 95% confidence interval of the validation home runs[33]. For comparison, we show the results for 90%, 95%, and 99% confidence intervals in Table 8.

Confidence Level	**90%**	**95%**	**99%**
Validation Rate	0.768	0.837	0.933

Table 8. Validation rates for home run models.

What this means is that the pitch components, along with batter height and pitcher-batter handedness match-up, are sufficient to explain the record number of home runs in 2017. This does not rule out other factors, because as with any model there is error (the projected number of home runs is not exact). However, it does provide very strong evidence that the pitch components explain home runs. The interpretation of the models is given in Appendix H. So how about tot.brk and h.brk?

For tot.brk, it did turn out to be the most statistically significant variable in the models. This confirms our prior observations. See Appendix H for more information.

For h.brk2, it turns out that the model components generally lined up with what was expected, with the main exception that the greater h.brk2 increases HR% for the four seam fastball (it decreases HR% for all other pitch types). See Appendix H for graphs depicting this surprise. We do not currently have a good physical explanation for this exception and are currently investigating it. Nevertheless, it does help interpret the sharp increase in right-handed pitcher h.brk2 in 2017 that went unexplained in Section 4.2. In particular, since FF is the highest proportion pitch type (36%, see Appendix E), and the right-handed pitchers are the highest proportion (RR=40%, RL=34%), that the real increase in h.brk2 for this case increases HR%, while at the same time the real decrease in h.brk2 for the left-handed pitchers (LR=19%,

[32] This is an explanatory model, in that it explains the results of the season, and it is statistically valid. Another approach is to build a model using all of the data from one season and use that model to predict the home runs of the subsequent season. We tried that, but the only two successful predictions, i.e. predicted number of home runs within the 95% confidence interval of the subsequent season's home runs, was 2010-2011 and 2016-2017. Thus, while the full model for 2016 does successfully predict 2017 home runs, we did not consider it to be validated statistically because the same technique only worked for 2 out of 9 season pairs.

[33] The confidence interval was generated using R's prop.test() function, and multiplying it by the number of pitches in the validation set. The prop.test() function uses a score test for its confidence interval, which is close to the common Wald or Agresti-Coull confidence interval for proportions.

LL=7%) increases HR% for other pitch types (see Table 6). Combining these opposing phenomena with the other cases helps explain the simultaneous increase in h.brk2 and increase in HR%. See Appendix H for more information.

5. Conclusion

In 2017, there was a spike in home runs. Many have thought it was due to changes in ball manufacturing or a change in the hitter's approach. While we see some evidence of the new approach by hitters, we propose that is only one side of the equation. The other side of the equation is a drop in pitch quality. There may be different reasons leading to the change in pitch quality, perhaps pitcher reaction to batters, or a change in the manufactured seams of the ball. Whatever these reasons may be, our point is that league-wide analyses by pitch type show that quality of pitch average (QOPA) dropped in 2017.

Although there were some concerns about the reliability of the new Trackman data, as changed from SportsVision in previous years, we presented five lines of evidence in support that any alleged error increases from Trackman are not enough to drown the true signal of decreased vertical break. A league-wide multiple regression analysis showing the variation present in the 2017 data was consistent with the variation of previous years. An individual pitcher analysis revealed no systematic trend of change in pitch components in 2017, including vertical break and horizontal break. Sometimes 2017 was lower, sometimes higher, usually in the same way as 2015 and 2016. The bottom line is: the overall MLB pitch quality was lower in 2017 and there were more home runs.

Examining the 2017 pitches by pitch type shows the predominant change in these pitch types from previous years is a decrease in vertical break and an increase in horizontal break. While a larger horizontal break is expected to decrease home runs, the high proportion four seam fastball turned out to correlate high horizontal break with an increase in home run %. A possible explanation may be that the increase in horizontal break for RR matchups could move the ball up the barrel of the bat and result in better contact. Whatever the explanation, the decrease in vertical break has been shown in our validated logistic regression model to be the most significant explanatory variable for home runs. Since the model is validated, and the primary changes from 2016 to 2017 are a decrease in vertical break and an increase in horizontal break in the manner described, we conclude that these changes in pitch quality are a significant factor in the home run increase.

For further research, we could further investigate the exact nature of the interactions between the number of home runs, pitch sequencing, the proportion of pitch types thrown in a season, and handedness. This could be done by deeper study of interaction terms in the logistic regression model. Another is that if pre-2017 Trackman data were made available, we would be able to confirm or refute the claims of Section 3. Finally, we will monitor pitch quality and its relationship to home runs allowed in 2018.

Acknowledgement: *We would like to thank statistician Don Lewis, Ph.D, who read the paper and provided feedback that was instrumental in improving the quality of our analysis. We would also like to thank another statistician who provided valuable feedback on the problem of analyzing the relationship between home runs, pitch sequencing, the proportion of pitch types thrown in a season, and handedness.*

Appendix A: Average Pitch Components for Home Runs in 2017

See file AppendixA_Components_HR_2017.pdf. This file contains plots of the average components, by year, for the most frequent six pitch types: CH, CU, FF, FT, SI, SL. Note that these graphs are produced from only the home runs for the regular + post season. The number of home runs is shown at the bottom of the graphs, by year. The graph in the upper left corner contains the name of the pitch type for the page.

For example, the first page has the six pitch components for the CH (Change-Up). The blue dots are the means. The blue line shows the change over time. The red bars are 95% confidence intervals. E.g. For the mean horizontal break in 2008, the error bar is about 0.87 to 0.95. The 530 below the bar means there were 530 home runs off of change-ups in 2008.

Appendix B: Comparison of the Distribution of Pitch Components for 13 MLB Pitchers from 2015 to 2017

See file AppendixB_SportsVision-Trackman-Graphs02.pdf. This file contains plots of the average components, by year, for the primary pitch types used by 13 different MLB pitchers. The graph in the upper left corner contains the name of the pitcher and his pitch type for the page. The subtitle states that 2015 is blue, 2016 is black, and 2017 is red.

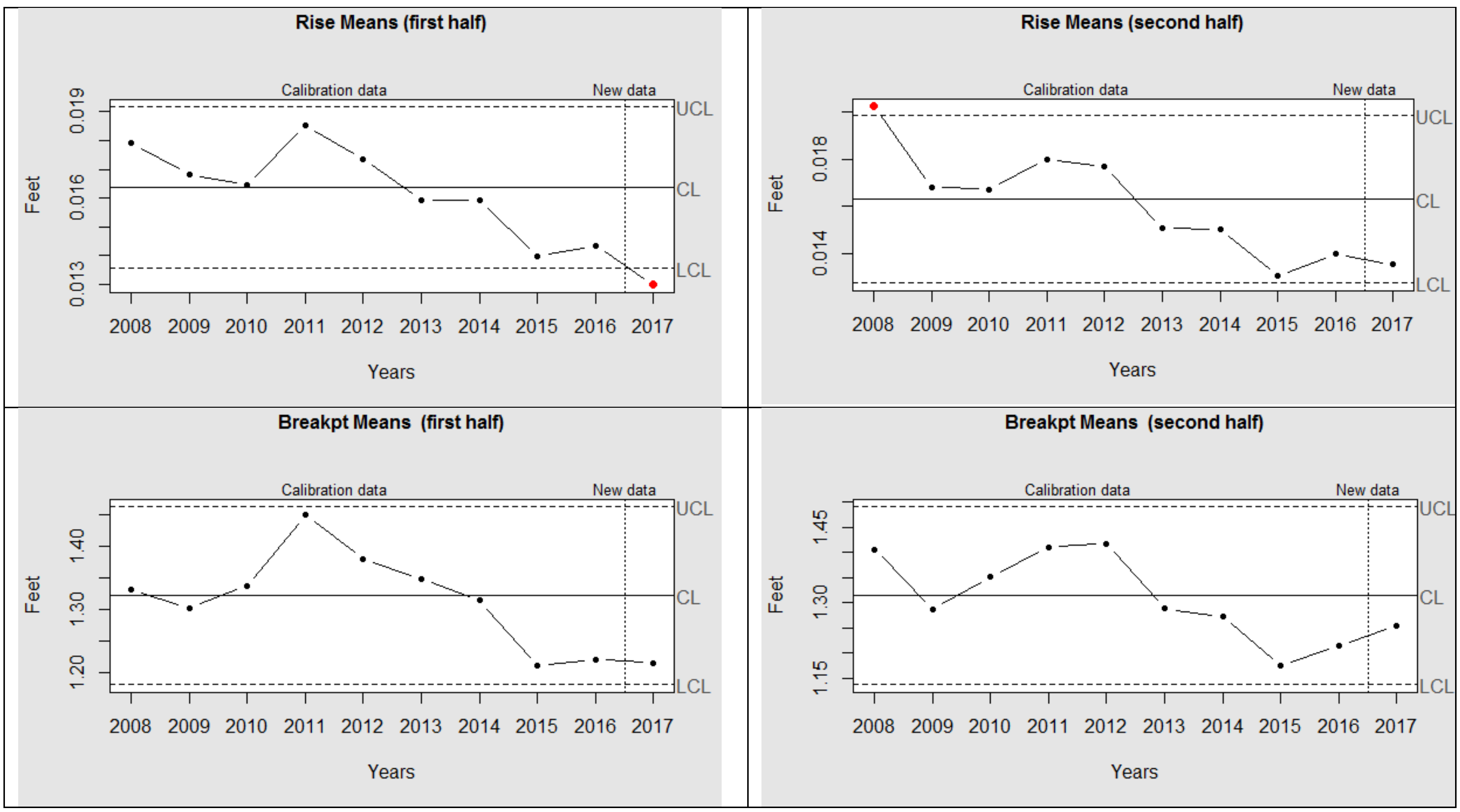
For example, the first page is for Christopher Archer's CH (Change-Up). His change-ups have two different breakpoints, consistent across all three years. His vertical break is also consistent across all three years. By contrast, his horizontal break went down a bit in 2016 and increased in 2017 past what it was in 2016.

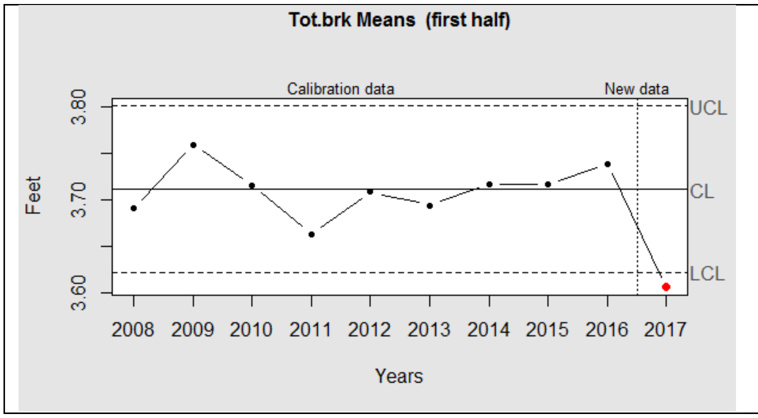
Appendix C: Control Charts for Means for Comparison

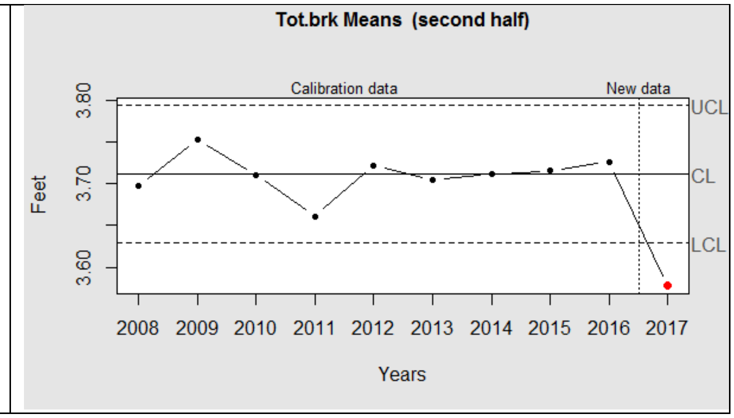
The purpose of this appendix is to provide a deeper inspection of the annual variation of the pitch components of section 2. Since there are only ten years of data, it is difficult to reliably spot patterns. Below is the same data, except the left column of graphs is the means of the first half of the season and the right column the second half. If the pattern seen in the full data is also seen in the two halves, it is confirmatory.

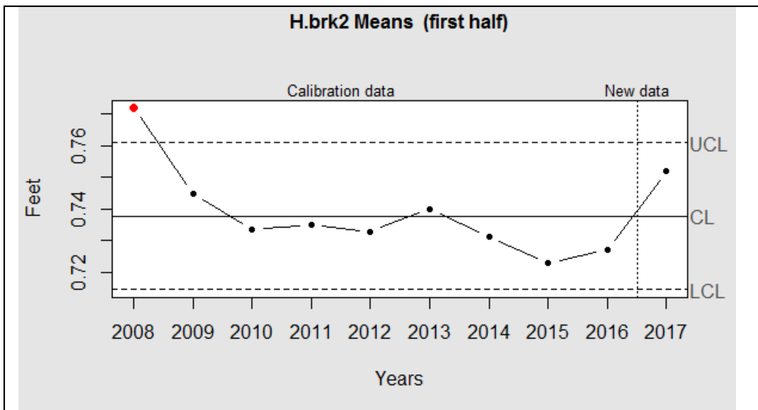
Inspection reveals the patterns do hold for all six pitch components, with the exception of horizontal break (h.brk2). There is a major decrease in h.brk2 during the second half of 2008. Since it does not effect our thesis, we will not pursue it further. The rest of the h.brk2 pattern generally comes through in both halves, although this component appears to vary most between the first and second halves. Although the neither half shows 2017 passing the three sigma boundary, as the full shows, both do show a clear jump in h.brk2.

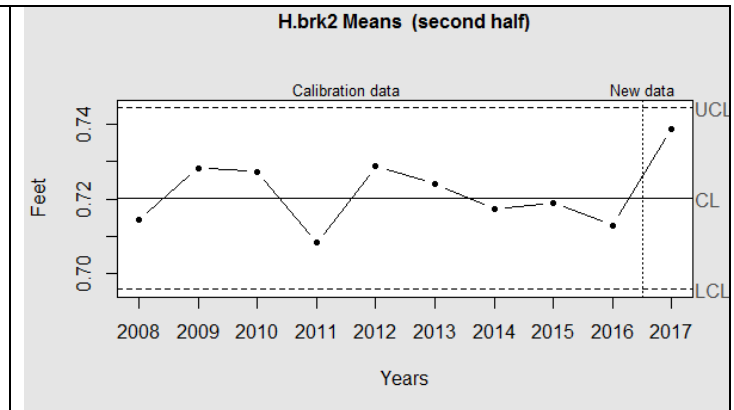
Lastly, the full rise graph shows no extreme, whereas the first half does for 2008 while the second half does for 2017. Similarly, whether speed in 2017 crosses the three sigma boundary, or not, it is still at the upper end of historic MPH. These patterns imply that the data is trending (speed increasing, rise decreasing), as opposed to staying within a channel with historic variation.

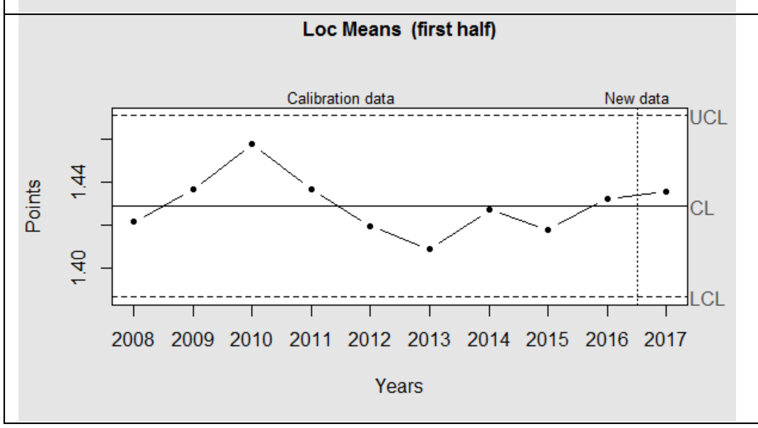


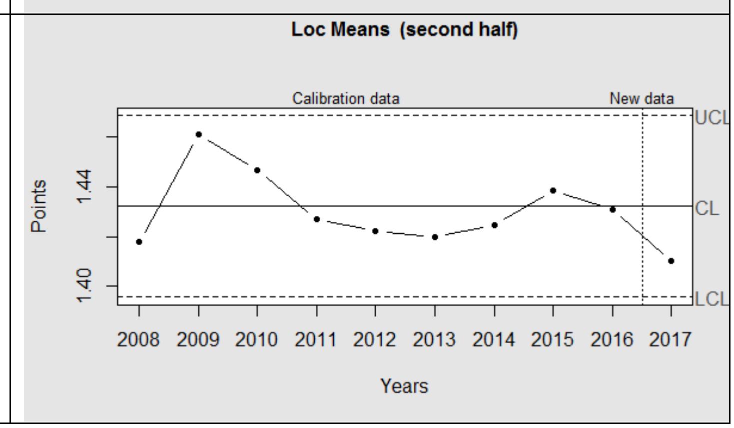

Tot.brk Means (first half)
Calibration data
New data
UCL
CL
LCL
Feet
3.60
3.70
3.80
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
Years

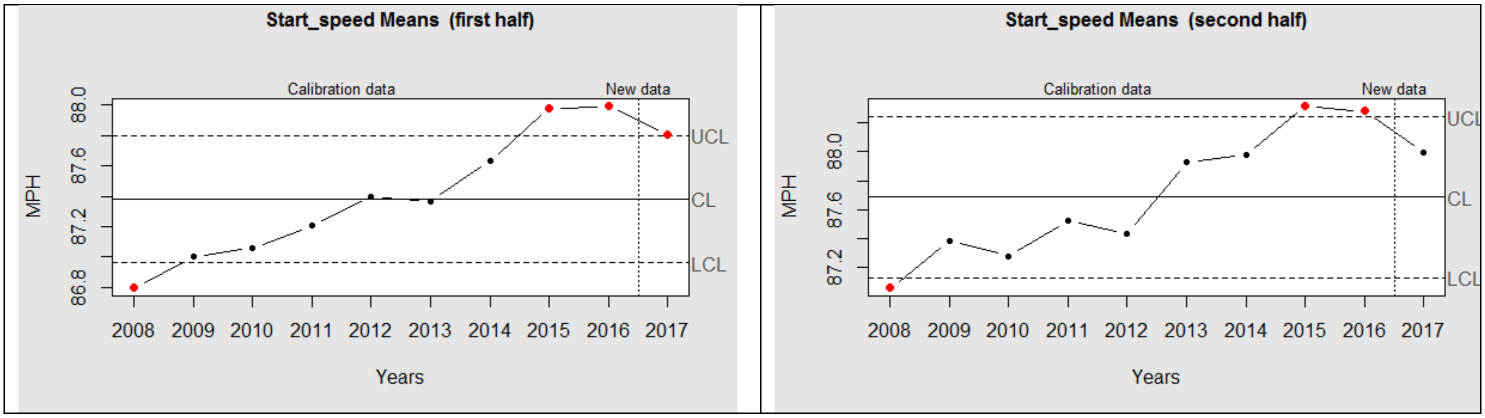

Tot.brk Means (second half)
Calibration data
New data
UCL
CL
LCL
Feet
3.60
3.70
3.80
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
Years


H.brk2 Means (first half)
Calibration data
New data
UCL
CL
LCL
Feet
0.72
0.74
0.76
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
Years


H.brk2 Means (second half)
Calibration data
New data
UCL
CL
LCL
Feet
0.70
0.72
0.74
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
Years


Loc Means (first half)
Calibration data
New data
UCL
CL
LCL
Points
1.40
1.44
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
Years

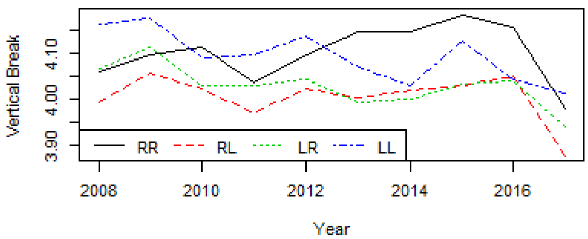

Loc Means (second half)
Calibration data
New data
UCL
CL
LCL
Points
1.40
1.44
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
Years



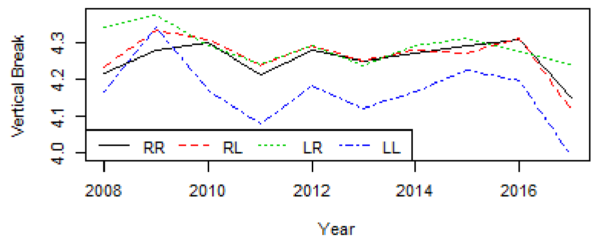
Appendix D: Effect of Handedness on Vertical Break, Horizontal Break, and HR

In this appendix are graphs of the mean vertical and horizontal break, broken out by the four different pitcher-batter match-ups (RR, RL, LR, LL) and separated by pitch type. From vertical break, we see that the pattern and magnitude is about the same – not much going on. However, for horizontal break, there is a significant discovery: the RH and LH pitchers behave differently. The final graph shows HRs by handedness, for completeness.

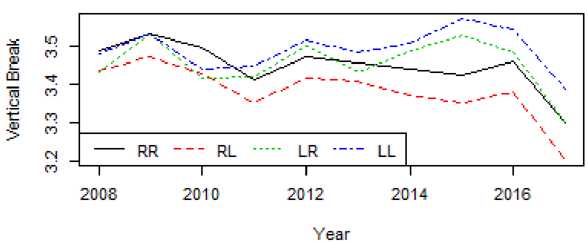
Change-Up Vertical Break



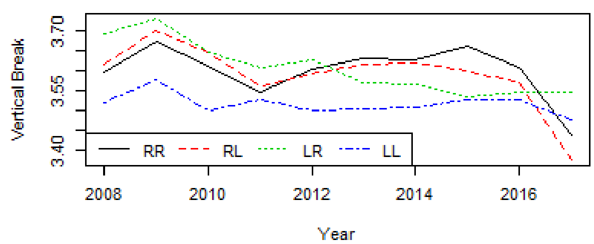
Curveball Vertical Break



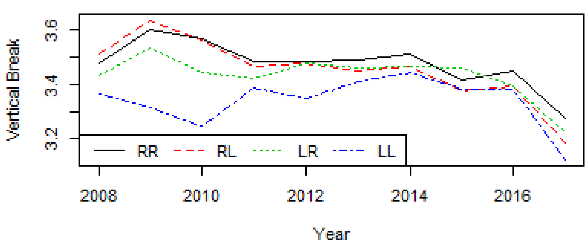
Four Seam Fastball Vertical Break



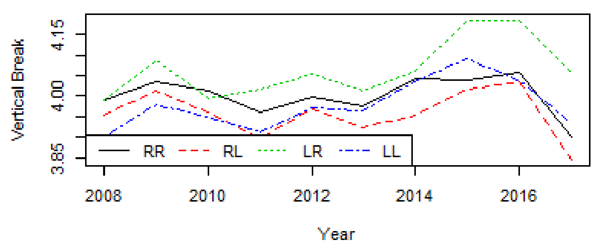
Two Seam Fastball Vertical Break

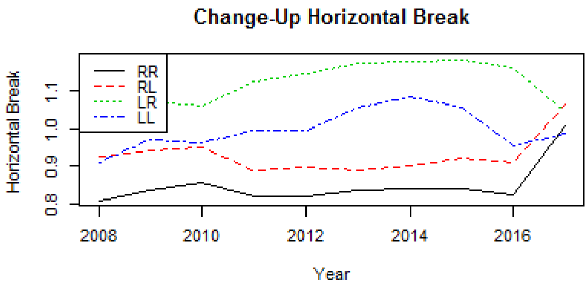


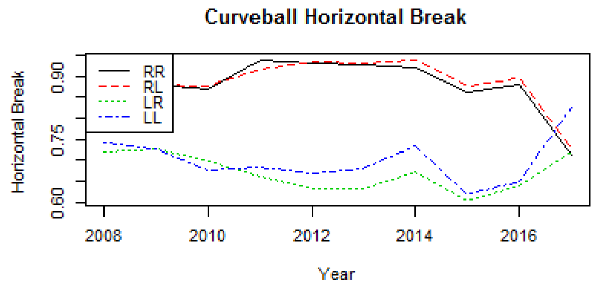
Sinker Vertical Break

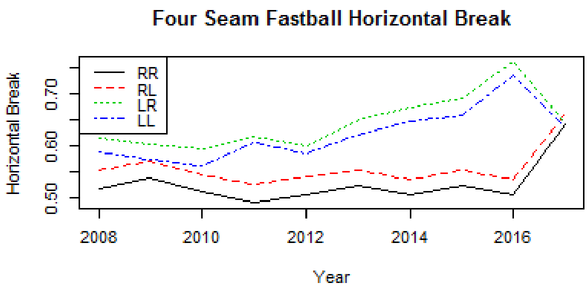


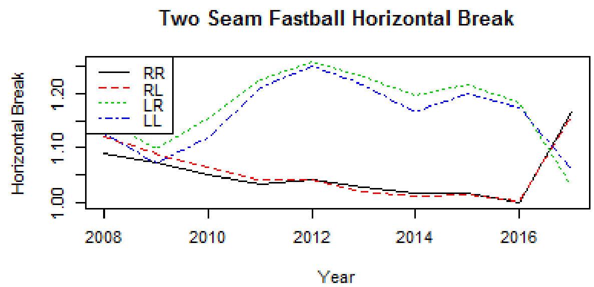
Slider Vertical Break

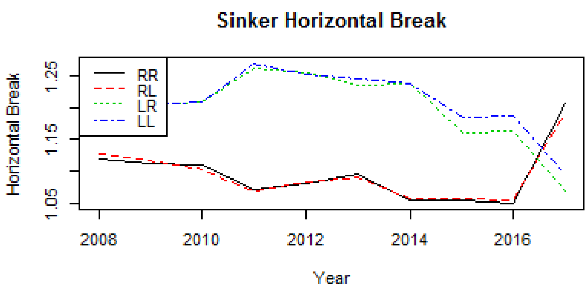


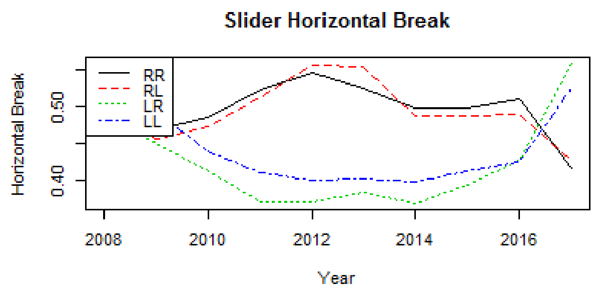

Change-Up Horizontal Break
RR
RL
LR
LL
Horizontal Break
0.8
0.9
1.0
1.1
2008
2010
2012
2014
2016
Year


Curveball Horizontal Break
RR
RL
LR
LL
Horizontal Break
0.60
0.75
0.90
2008
2010
2012
2014
2016
Year

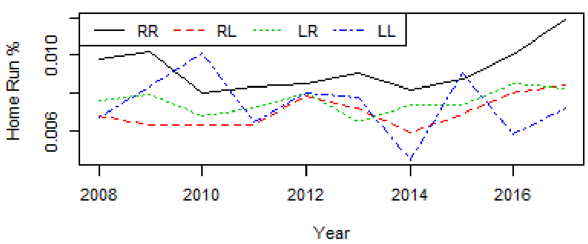

Four Seam Fastball Horizontal Break
RR
RL
LR
LL
Horizontal Break
0.50
0.60
0.70
2008
2010
2012
2014
2016
Year


Two Seam Fastball Horizontal Break
RR
RL
LR
LL
Horizontal Break
1.00
1.10
1.20
2008
2010
2012
2014
2016
Year

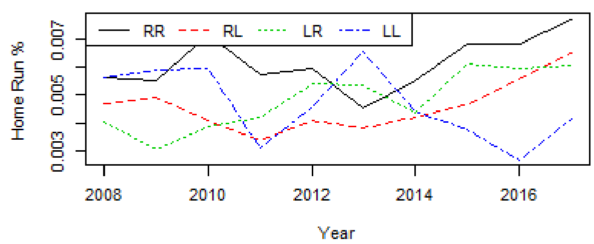

Sinker Horizontal Break
RR
RL
LR
LL
Horizontal Break
1.05
1.15
1.25
2008
2010
2012
2014
2016
Year


Slider Horizontal Break
RR
RL
LR
LL
Horizontal Break
0.40
0.50
2008
2010
2012
2014
2016
Year

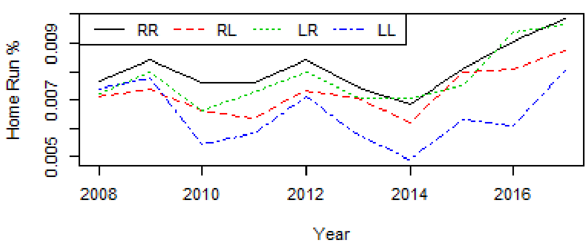
Change-Up Home Run %



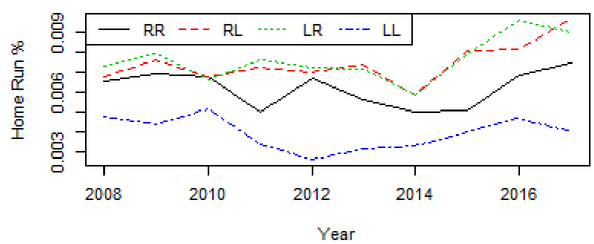
Curveball Home Run %



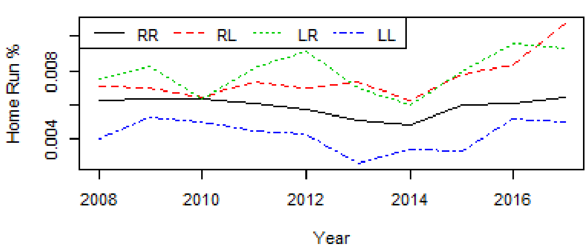
Four Seam Fastball Home Run %



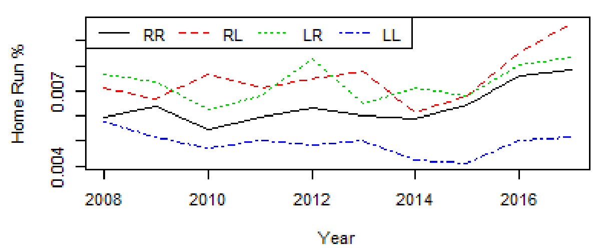
Two Seam Fastball Home Run %



Sinker Home Run %



Slider Home Run %



Appendix E: Pitch Sequencing

There is no doubt that some of the changes in annual mean pitch components are due to different proportions of pitches, since different pitch types vary in their pitch components (see Table 3). Differences in pitch proportions are shown in Figure 3. The primary ten-year observation is that sinkers (SI) have reduced while two-seam fastballs (FT) have increased[34].

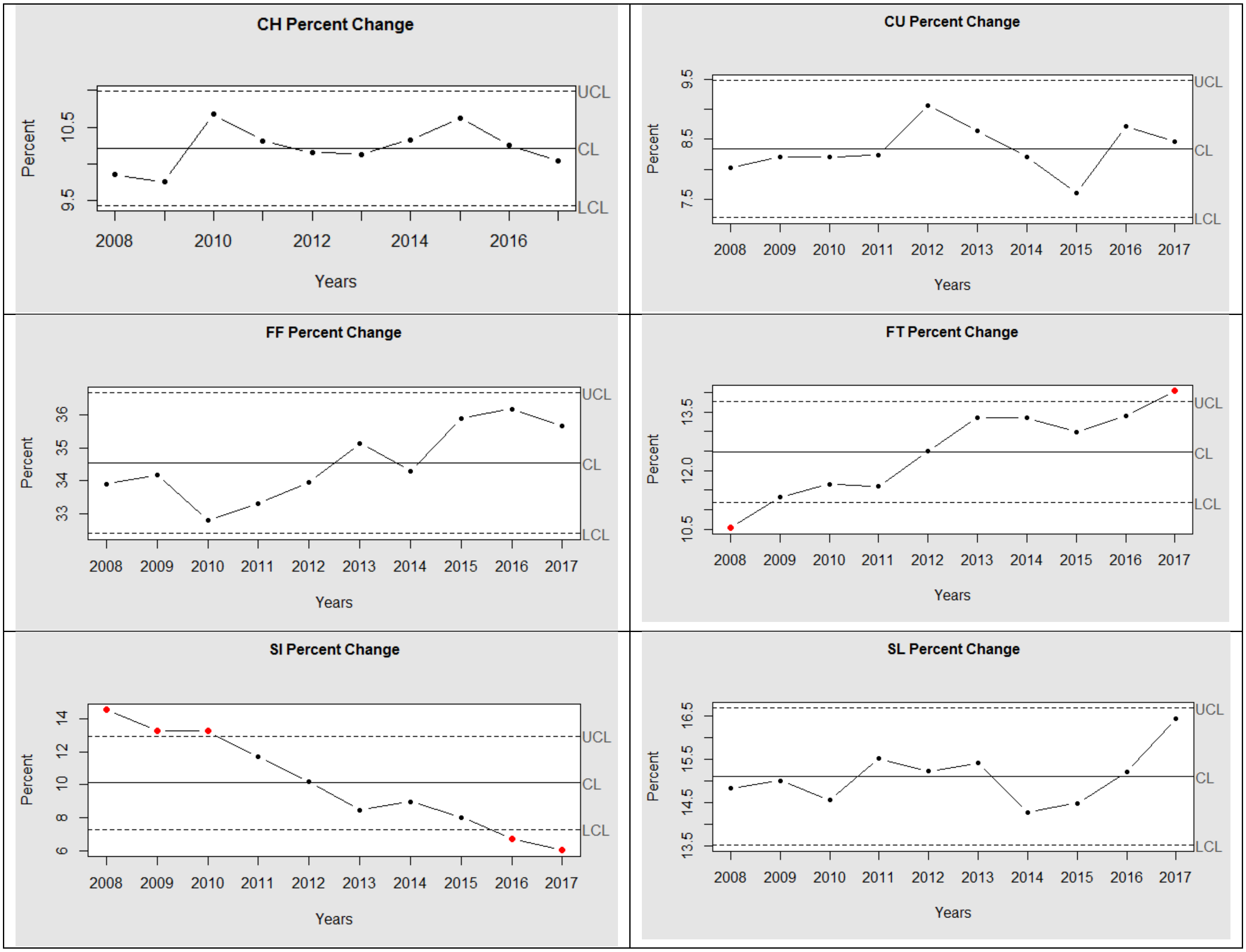


Figure 4. Variation in annual proportions of pitches by pitch type. CH = change-up; CU = curveball; FF = four-seam fastball; FT = two-seam fastball; SI = sinker; SL = slider.

[34] All pitch classifications of this paper use the classification provided by PITCHf/x data, with no modifications. We have strong evidence that there have been some adjustments to the classification algorithm over the years, and use it cautiously. For example, Zach Britton (2011-2017) shows hundreds of sinkers in 2014-15 and a dip in four-seam fastballs, but zero sinkers in other years. We do not believe Zach experimented with sinkers for a couple of years and gave them up.

For 2017, though, we see that the proportion of sinkers (SI) and curveballs (CU) decreased. This would contribute to a decrease in vertical break, on average, and an increase in horizontal break. Also, the sharp rise in sliders (SL) would contribute to an increase in horizontal break. The decrease in four seam fastballs (FF, -.5%) is balanced by an increase in two seam fastballs (FT, +.6%). These changes are accounted for in our final model of Section 4.

Appendix F: Kolmogorov-Smirnov Test Results

Each row below contains the results of the three Kolmogorov-Smirnov tests:

1. KS1617: Test of the difference in the distribution from 2016 to 2017
2. KS1516: Test of the difference in the distribution from 2015 to 2016
3. KS1517: Test of the difference in the distribution from 2015 to 2017

Each row is for one pitcher-pitch type combination and one pitch component. As a result there is one row for each graph of Appendix B: 49 pitcher-pitch type combinations * 6 = 294 rows.

For example, in row three we have the Tot.brk for Christopher Archer's CH (Change-Up). The p-values for the three tests are 0.5131, 0.0389, and 0.4192, respectively. This means that there is a statistically significant vertical break between 2015 and 2016, at the 5% level of significance. Comparing this with the 3rd graph in Appendix B, we see this is identifying the black and blue distributions as statistically significantly different. We would probably not observe this by eye. We set the Flag on this row to FALSE, because we only want to flag the rows where 2017 is different from both 2015 and 2016, but 2015 and 2016 are the same. This is a consistent behavior for the pitcher across two years which changed in 2017. The first Flag=TRUE occurs in row 17.

	Pitcher	Component	Type	KS1617	KS1516	KS1517	Flag
1	Christopher Archer	Rise	CH	1.0000	1.0000	1.0000	FALSE
2	Christopher Archer	BreakPoint	CH	0.0000	0.0003	0.0002	FALSE
3	Christopher Archer	Tot.brk	CH	0.5131	0.0389	0.4192	FALSE
4	Christopher Archer	Location	CH	0.4838	0.0402	0.1049	FALSE
5	Christopher Archer	H.brk2	CH	0.0000	0.0008	0.0000	FALSE
6	Christopher Archer	Speed	CH	0.0000	0.0000	0.5354	FALSE
7	Christopher Archer	Rise	FF	1.0000	1.0000	1.0000	FALSE
8	Christopher Archer	BreakPoint	FF	0.0000	0.0000	0.0052	FALSE
9	Christopher Archer	Tot.brk	FF	0.0000	0.0000	0.3567	FALSE
10	Christopher Archer	Location	FF	0.2895	0.0314	0.7708	FALSE
11	Christopher Archer	H.brk2	FF	0.0000	0.0000	0.0000	FALSE
12	Christopher Archer	Speed	FF	0.0000	0.0000	0.0000	FALSE
13	Christopher Archer	Rise	SL	1.0000	0.9633	0.9903	FALSE
14	Christopher Archer	BreakPoint	SL	0.0008	0.0000	0.0008	FALSE
15	Christopher Archer	Tot.brk	SL	0.5091	0.1241	0.0155	FALSE
16	Christopher Archer	Location	SL	0.6456	0.5111	0.5219	FALSE
17	Christopher Archer	H.brk2	SL	0.0000	0.1827	0.0000	TRUE
18	Christopher Archer	Speed	SL	0.0000	0.0000	0.0000	FALSE
19	Ervin Santana	Rise	CH	1.0000	1.0000	1.0000	FALSE
20	Ervin Santana	BreakPoint	CH	0.3245	0.6111	0.6454	FALSE
21	Ervin Santana	Tot.brk	CH	0.0000	0.0000	0.0784	FALSE
22	Ervin Santana	Location	CH	0.0448	0.0364	0.7201	FALSE
23	Ervin Santana	H.brk2	CH	0.0000	0.4088	0.0000	TRUE
24	Ervin Santana	Speed	CH	0.0000	0.8764	0.0000	TRUE
25	Ervin Santana	Rise	FF	1.0000	1.0000	1.0000	FALSE
26	Ervin Santana	BreakPoint	FF	0.0000	0.0000	0.0000	FALSE
27	Ervin Santana	Tot.brk	FF	0.0000	0.0000	0.0028	FALSE
28	Ervin Santana	Location	FF	0.2366	0.8163	0.6133	FALSE
29	Ervin Santana	H.brk2	FF	0.0000	0.1444	0.0000	TRUE
30	Ervin Santana	Speed	FF	0.0000	0.0000	0.0000	FALSE
31	Ervin Santana	Rise	SL	1.0000	0.9771	0.7308	FALSE
32	Ervin Santana	BreakPoint	SL	0.0000	0.0000	0.0000	FALSE
33	Ervin Santana	Tot.brk	SL	0.0000	0.0000	0.1985	FALSE
34	Ervin Santana	Location	SL	0.0019	0.0071	0.1702	FALSE
35	Ervin Santana	H.brk2	SL	0.0000	0.0000	0.0000	FALSE
36	Ervin Santana	Speed	SL	0.0000	0.0000	0.0000	FALSE
37	Justin Verlander	Rise	CH	1.0000	1.0000	1.0000	FALSE
38	Justin Verlander	BreakPoint	CH	0.0000	0.0000	0.0000	FALSE
39	Justin Verlander	Tot.brk	CH	0.0008	0.0005	0.4057	FALSE
40	Justin Verlander	Location	CH	0.6897	0.1194	0.1395	FALSE
41	Justin Verlander	H.brk2	CH	0.1293	0.0000	0.0000	FALSE
42	Justin Verlander	Speed	CH	0.0000	0.0000	0.0000	FALSE
43	Justin Verlander	Rise	CU	0.0038	0.0000	0.0007	FALSE
44	Justin Verlander	BreakPoint	CU	0.0000	0.0000	0.0008	FALSE
45	Justin Verlander	Tot.brk	CU	0.0141	0.4093	0.0017	TRUE
46	Justin Verlander	Location	CU	0.4966	0.5326	0.7905	FALSE
47	Justin Verlander	H.brk2	CU	0.0000	0.0000	0.0433	FALSE
48	Justin Verlander	Speed	CU	0.0000	0.0001	0.0000	FALSE
49	Justin Verlander	Rise	FF	1.0000	1.0000	1.0000	FALSE
50	Justin Verlander	BreakPoint	FF	0.0000	0.0000	0.0000	FALSE
51	Justin Verlander	Tot.brk	FF	0.7002	0.0000	0.0000	FALSE
52	Justin Verlander	Location	FF	0.3950	0.5896	0.2476	FALSE
53	Justin Verlander	H.brk2	FF	0.0000	0.0000	0.0000	FALSE
54	Justin Verlander	Speed	FF	0.0000	0.0000	0.0000	FALSE
55	Justin Verlander	Rise	SL	1.0000	1.0000	1.0000	FALSE
56	Justin Verlander	BreakPoint	SL	0.0000	0.0000	0.0000	FALSE
57	Justin Verlander	Tot.brk	SL	0.0000	0.0005	0.0000	FALSE
58	Justin Verlander	Location	SL	0.7105	0.3531	0.4361	FALSE
59	Justin Verlander	H.brk2	SL	0.0775	0.1873	0.0303	FALSE
60	Justin Verlander	Speed	SL	0.0000	0.0000	0.0000	FALSE
61	Kyle Gibson	Rise	CH	1.0000	1.0000	1.0000	FALSE
62	Kyle Gibson	BreakPoint	CH	0.0000	0.0000	0.2271	FALSE
63	Kyle Gibson	Tot.brk	CH	0.0000	0.8509	0.0000	TRUE

	Pitcher	Component	Type	KS1617	KS1516	KS1517	Flag
64	Kyle Gibson	Location	CH	0.2254	0.0158	0.0008	FALSE
65	Kyle Gibson	H.brk2	CH	0.0000	0.0000	0.0003	FALSE
66	Kyle Gibson	Speed	CH	0.0000	0.0000	0.0003	FALSE
67	Kyle Gibson	Rise	CU	0.0132	0.9544	0.5076	FALSE
68	Kyle Gibson	BreakPoint	CU	0.0000	0.0261	0.0000	FALSE
69	Kyle Gibson	Tot.brk	CU	0.1521	0.5084	0.3983	FALSE
70	Kyle Gibson	Location	CU	0.7562	0.5638	0.8239	FALSE
71	Kyle Gibson	H.brk2	CU	0.0000	0.0112	0.0025	FALSE
72	Kyle Gibson	Speed	CU	0.0000	0.0468	0.0000	FALSE
73	Kyle Gibson	Rise	FF	1.0000	1.0000	1.0000	FALSE
74	Kyle Gibson	BreakPoint	FF	0.0000	0.0000	0.1737	FALSE
75	Kyle Gibson	Tot.brk	FF	0.0000	0.0000	0.0000	FALSE
76	Kyle Gibson	Location	FF	0.2759	0.8026	0.6765	FALSE
77	Kyle Gibson	H.brk2	FF	0.0000	0.0125	0.0000	FALSE
78	Kyle Gibson	Speed	FF	0.0000	0.0000	0.0000	FALSE
79	Kyle Gibson	Rise	FT	1.0000	1.0000	1.0000	FALSE
80	Kyle Gibson	BreakPoint	FT	0.0000	0.0000	0.0000	FALSE
81	Kyle Gibson	Tot.brk	FT	0.0000	0.0000	0.0000	FALSE
82	Kyle Gibson	Location	FT	0.4079	0.9450	0.3132	FALSE
83	Kyle Gibson	H.brk2	FT	0.0000	0.0000	0.0000	FALSE
84	Kyle Gibson	Speed	FT	0.0000	0.0000	0.0000	FALSE
85	Kyle Gibson	Rise	SL	1.0000	1.0000	1.0000	FALSE
86	Kyle Gibson	BreakPoint	SL	0.0000	0.0000	0.0000	FALSE
87	Kyle Gibson	Tot.brk	SL	0.0000	0.0000	0.1330	FALSE
88	Kyle Gibson	Location	SL	0.4241	0.5544	0.7214	FALSE
89	Kyle Gibson	H.brk2	SL	0.0000	0.0000	0.0196	FALSE
90	Kyle Gibson	Speed	SL	0.0182	0.0000	0.0000	FALSE
91	Marco Estrada	Rise	CH	1.0000	1.0000	1.0000	FALSE
92	Marco Estrada	BreakPoint	CH	0.0873	0.0000	0.0000	FALSE
93	Marco Estrada	Tot.brk	CH	0.0000	0.0000	0.0325	FALSE
94	Marco Estrada	Location	CH	0.0661	0.0571	0.9129	FALSE
95	Marco Estrada	H.brk2	CH	0.0000	0.0000	0.0000	FALSE
96	Marco Estrada	Speed	CH	0.0000	0.0000	0.0000	FALSE
97	Marco Estrada	Rise	CU	0.5683	0.0000	0.0021	FALSE
98	Marco Estrada	BreakPoint	CU	0.0929	0.0000	0.0003	FALSE
99	Marco Estrada	Tot.brk	CU	0.0042	0.0000	0.0055	FALSE
100	Marco Estrada	Location	CU	0.7690	0.9999	0.6862	FALSE
101	Marco Estrada	H.brk2	CU	0.0000	0.0000	0.0000	FALSE
102	Marco Estrada	Speed	CU	0.0000	0.0000	0.3700	FALSE
103	Marco Estrada	Rise	FC	0.9986	1.0000	1.0000	FALSE
104	Marco Estrada	BreakPoint	FC	0.0000	0.0000	0.0000	FALSE
105	Marco Estrada	Tot.brk	FC	0.0674	0.0660	0.0025	FALSE
106	Marco Estrada	Location	FC	0.0983	0.2890	0.0322	FALSE
107	Marco Estrada	H.brk2	FC	0.0005	0.6706	0.1253	FALSE
108	Marco Estrada	Speed	FC	0.0000	0.0000	0.0000	FALSE
109	Marco Estrada	Rise	FF	1.0000	1.0000	1.0000	FALSE
110	Marco Estrada	BreakPoint	FF	0.0000	0.0000	0.0000	FALSE
111	Marco Estrada	Tot.brk	FF	0.0004	0.0000	0.0000	FALSE
112	Marco Estrada	Location	FF	0.2844	0.2105	0.8566	FALSE
113	Marco Estrada	H.brk2	FF	0.0000	0.0000	0.0000	FALSE
114	Marco Estrada	Speed	FF	0.0000	0.0000	0.0000	FALSE
115	Max Scherzer	Rise	CH	0.0693	0.9520	0.0694	FALSE
116	Max Scherzer	BreakPoint	CH	0.0000	0.0128	0.0091	FALSE
117	Max Scherzer	Tot.brk	CH	0.0000	0.4950	0.0008	TRUE
118	Max Scherzer	Location	CH	0.7316	0.7792	0.1468	FALSE
119	Max Scherzer	H.brk2	CH	0.0000	0.0000	0.0000	FALSE
120	Max Scherzer	Speed	CH	0.0000	0.0001	0.2939	FALSE
121	Max Scherzer	Rise	CU	0.0000	0.0922	0.0000	TRUE
122	Max Scherzer	BreakPoint	CU	0.0000	0.0051	0.0000	FALSE
123	Max Scherzer	Tot.brk	CU	0.0000	0.3299	0.0001	TRUE
124	Max Scherzer	Location	CU	0.0348	0.6640	0.3374	FALSE
125	Max Scherzer	H.brk2	CU	0.0000	0.0000	0.0000	FALSE
126	Max Scherzer	Speed	CU	0.0001	0.0000	0.0000	FALSE

	Pitcher	Component	Type	KS1617	KS1516	KS1517	Flag
127	Max Scherzer	Rise	FF	0.3901	0.1038	0.9975	FALSE
128	Max Scherzer	BreakPoint	FF	0.0000	0.0000	0.0000	FALSE
129	Max Scherzer	Tot.brk	FF	0.0000	0.0000	0.0000	FALSE
130	Max Scherzer	Location	FF	0.2593	0.6411	0.0487	FALSE
131	Max Scherzer	H.brk2	FF	0.0000	0.0102	0.0000	FALSE
132	Max Scherzer	Speed	FF	0.0005	0.0000	0.0001	FALSE
133	Max Scherzer	Rise	SL	0.0000	0.0249	0.0004	FALSE
134	Max Scherzer	BreakPoint	SL	0.0000	0.0064	0.0003	FALSE
135	Max Scherzer	Tot.brk	SL	0.0000	0.5729	0.0000	TRUE
136	Max Scherzer	Location	SL	0.0608	0.9019	0.0135	FALSE
137	Max Scherzer	H.brk2	SL	0.0000	0.3255	0.0000	TRUE
138	Max Scherzer	Speed	SL	0.0033	0.0559	0.1211	FALSE
139	Michael Wacha	Rise	CH	1.0000	1.0000	1.0000	FALSE
140	Michael Wacha	BreakPoint	CH	0.0000	0.0000	0.0060	FALSE
141	Michael Wacha	Tot.brk	CH	0.0052	0.1012	0.0094	TRUE
142	Michael Wacha	Location	CH	0.4979	0.7171	0.8737	FALSE

	Pitcher	Component	Type	KS1617	KS1516	KS1517	Flag
143	Michael Wacha	H.brk2	CH	0.0011	0.0000	0.0000	FALSE
144	Michael Wacha	Speed	CH	0.0000	0.0000	0.0000	FALSE
145	Michael Wacha	Rise	CU	0.5459	0.0025	0.0029	FALSE
146	Michael Wacha	BreakPoint	CU	0.3257	0.0034	0.0035	FALSE
147	Michael Wacha	Tot.brk	CU	0.8830	0.0974	0.1011	FALSE
148	Michael Wacha	Location	CU	0.7148	0.9776	0.7552	FALSE
149	Michael Wacha	H.brk2	CU	0.5121	0.7525	0.9730	FALSE
150	Michael Wacha	Speed	CU	0.0000	0.0000	0.0000	FALSE
151	Michael Wacha	Rise	FC	1.0000	0.9979	1.0000	FALSE
152	Michael Wacha	BreakPoint	FC	0.0008	0.0000	0.0000	FALSE
153	Michael Wacha	Tot.brk	FC	0.0000	0.0070	0.1887	FALSE
154	Michael Wacha	Location	FC	0.5750	0.8372	0.2394	FALSE
155	Michael Wacha	H.brk2	FC	0.1721	0.8238	0.0858	FALSE
156	Michael Wacha	Speed	FC	0.0034	0.0000	0.0000	FALSE
157	Michael Wacha	Rise	FF	1.0000	1.0000	1.0000	FALSE
158	Michael Wacha	BreakPoint	FF	0.0000	0.0000	0.6954	FALSE
159	Michael Wacha	Tot.brk	FF	0.0005	0.0000	0.0000	FALSE
160	Michael Wacha	Location	FF	0.5015	0.0609	0.1461	FALSE
161	Michael Wacha	H.brk2	FF	0.0000	0.0000	0.0000	FALSE
162	Michael Wacha	Speed	FF	0.0000	0.0000	0.0000	FALSE
163	R.A. Dickey	Rise	FF	0.3660	0.1255	0.0014	FALSE
164	R.A. Dickey	BreakPoint	FF	0.0000	0.0000	0.0000	FALSE
165	R.A. Dickey	Tot.brk	FF	0.2425	0.0112	0.1017	FALSE
166	R.A. Dickey	Location	FF	0.6851	0.1430	0.6418	FALSE
167	R.A. Dickey	H.brk2	FF	0.0000	0.0287	0.0000	FALSE
168	R.A. Dickey	Speed	FF	0.0000	0.0000	0.0000	FALSE
169	R.A. Dickey	Rise	KN	0.0098	0.0065	0.0025	FALSE
170	R.A. Dickey	BreakPoint	KN	0.0133	0.0076	0.0133	FALSE
171	R.A. Dickey	Tot.brk	KN	0.0000	0.0000	0.0504	FALSE
172	R.A. Dickey	Location	KN	0.1629	0.1344	0.1525	FALSE
173	R.A. Dickey	H.brk2	KN	0.1662	0.0307	0.5392	FALSE
174	R.A. Dickey	Speed	KN	0.0000	0.2393	0.0000	TRUE
175	James Paxton	Rise	CH	1.0000	1.0000	1.0000	FALSE
176	James Paxton	BreakPoint	CH	0.0016	0.0000	0.0065	FALSE
177	James Paxton	Tot.brk	CH	0.4920	0.6158	0.8991	FALSE
178	James Paxton	Location	CH	0.8850	0.1182	0.1565	FALSE
179	James Paxton	H.brk2	CH	0.1186	0.0000	0.0000	FALSE
180	James Paxton	Speed	CH	0.0000	0.0000	0.0011	FALSE
181	James Paxton	Rise	FF	1.0000	1.0000	1.0000	FALSE
182	James Paxton	BreakPoint	FF	0.0000	0.0000	0.0000	FALSE
183	James Paxton	Tot.brk	FF	0.0308	0.0000	0.0000	FALSE
184	James Paxton	Location	FF	0.2404	0.0202	0.0278	FALSE
185	James Paxton	H.brk2	FF	0.0000	0.0000	0.0000	FALSE
186	James Paxton	Speed	FF	0.0000	0.0000	0.0000	FALSE
187	James Paxton	Rise	KC	0.0650	0.0000	0.0104	FALSE

	Pitcher	Component	Type	KS1617	KS1516	KS1517	Flag
188	James Paxton	BreakPoint	KC	0.0000	0.0000	0.0000	FALSE
189	James Paxton	Tot.brk	KC	0.9751	0.0000	0.0000	FALSE
190	James Paxton	Location	KC	0.2776	0.0002	0.0004	FALSE
191	James Paxton	H.brk2	KC	0.0000	0.0000	0.0000	FALSE
192	James Paxton	Speed	KC	0.0000	0.0000	0.0000	FALSE
193	Johnny Cueto	Rise	CH	0.8102	0.7232	0.2128	FALSE
194	Johnny Cueto	BreakPoint	CH	0.0000	0.0000	0.0000	FALSE
195	Johnny Cueto	Tot.brk	CH	0.1523	0.0000	0.0026	FALSE
196	Johnny Cueto	Location	CH	0.9451	0.0844	0.1786	FALSE
197	Johnny Cueto	H.brk2	CH	0.0011	0.2746	0.0588	FALSE
198	Johnny Cueto	Speed	CH	0.0000	0.0000	0.0000	FALSE
199	Johnny Cueto	Rise	FC	1.0000	0.9820	0.9502	FALSE
200	Johnny Cueto	BreakPoint	FC	0.0000	0.0000	0.0000	FALSE
201	Johnny Cueto	Tot.brk	FC	0.2506	0.0000	0.0000	FALSE
202	Johnny Cueto	Location	FC	0.7978	0.7125	0.6047	FALSE
203	Johnny Cueto	H.brk2	FC	0.0000	0.0002	0.0418	FALSE
204	Johnny Cueto	Speed	FC	0.0000	0.0000	0.0004	FALSE
205	Johnny Cueto	Rise	FF	1.0000	1.0000	1.0000	FALSE
206	Johnny Cueto	BreakPoint	FF	0.0000	0.0000	0.0000	FALSE
207	Johnny Cueto	Tot.brk	FF	0.0000	0.0000	0.0727	FALSE
208	Johnny Cueto	Location	FF	0.3101	0.6770	0.7345	FALSE
209	Johnny Cueto	H.brk2	FF	0.0000	0.0001	0.0000	FALSE
210	Johnny Cueto	Speed	FF	0.0000	0.0000	0.0000	FALSE
211	Johnny Cueto	Rise	FT	1.0000	1.0000	0.9979	FALSE
212	Johnny Cueto	BreakPoint	FT	0.0000	0.0000	0.0000	FALSE
213	Johnny Cueto	Tot.brk	FT	0.0219	0.0277	0.6468	FALSE
214	Johnny Cueto	Location	FT	0.3770	0.3062	0.7703	FALSE
215	Johnny Cueto	H.brk2	FT	0.0000	0.0000	0.0000	FALSE
216	Johnny Cueto	Speed	FT	0.0000	0.0000	0.0000	FALSE
217	Johnny Cueto	Rise	SL	0.0390	0.0106	0.0000	FALSE
218	Johnny Cueto	BreakPoint	SL	0.0000	0.0108	0.0000	FALSE
219	Johnny Cueto	Tot.brk	SL	0.2681	0.5670	0.1124	FALSE
220	Johnny Cueto	Location	SL	0.9501	0.9427	0.9116	FALSE
221	Johnny Cueto	H.brk2	SL	0.0001	0.6041	0.0378	TRUE
222	Johnny Cueto	Speed	SL	0.0015	0.0001	0.0000	FALSE
223	Jon Lester	Rise	CH	1.0000	1.0000	0.9883	FALSE
224	Jon Lester	BreakPoint	CH	0.0017	0.4666	0.0492	TRUE
225	Jon Lester	Tot.brk	CH	0.8118	0.1951	0.0997	FALSE
226	Jon Lester	Location	CH	0.7166	0.0150	0.0030	FALSE
227	Jon Lester	H.brk2	CH	0.0067	0.2630	0.0031	TRUE
228	Jon Lester	Speed	CH	0.8717	0.0775	0.1649	FALSE
229	Jon Lester	Rise	CU	0.6357	0.0572	0.0123	FALSE
230	Jon Lester	BreakPoint	CU	0.1126	0.0703	0.0084	FALSE
231	Jon Lester	Tot.brk	CU	0.0004	0.0012	0.4086	FALSE
232	Jon Lester	Location	CU	0.0223	0.1993	0.6120	FALSE
233	Jon Lester	H.brk2	CU	0.0000	0.0001	0.0000	FALSE
234	Jon Lester	Speed	CU	0.0000	0.0048	0.0002	FALSE
235	Jon Lester	Rise	FC	0.9294	1.0000	0.8730	FALSE
236	Jon Lester	BreakPoint	FC	0.0000	0.0000	0.0000	FALSE
237	Jon Lester	Tot.brk	FC	0.0107	0.0180	0.0000	FALSE
238	Jon Lester	Location	FC	0.0221	0.9103	0.0109	TRUE
239	Jon Lester	H.brk2	FC	0.0000	0.0000	0.0011	FALSE
240	Jon Lester	Speed	FC	0.0000	0.0000	0.0194	FALSE
241	Jon Lester	Rise	FF	0.9992	1.0000	0.9989	FALSE
242	Jon Lester	BreakPoint	FF	0.0000	0.2013	0.0000	TRUE
243	Jon Lester	Tot.brk	FF	0.0000	0.3670	0.0000	TRUE
244	Jon Lester	Location	FF	0.6694	0.5467	0.3837	FALSE
245	Jon Lester	H.brk2	FF	0.2547	0.0000	0.0000	FALSE
246	Jon Lester	Speed	FF	0.0000	0.1206	0.0000	TRUE
247	Jon Lester	Rise	SI	1.0000	1.0000	1.0000	FALSE
248	Jon Lester	BreakPoint	SI	0.0000	0.1213	0.0000	TRUE
249	Jon Lester	Tot.brk	SI	0.0656	0.3420	0.0914	FALSE
250	Jon Lester	Location	SI	0.1034	0.6624	0.2367	FALSE

	Pitcher	Component	Type	KS1617	KS1516	KS1517	Flag
251	Jon Lester	H.brk2	SI	0.1291	0.0076	0.0001	FALSE
252	Jon Lester	Speed	SI	0.0000	0.0496	0.0000	FALSE
253	Julio Teheran	Rise	CH	0.9978	0.0186	0.0617	FALSE
254	Julio Teheran	BreakPoint	CH	0.0346	0.0186	0.1017	FALSE
255	Julio Teheran	Tot.brk	CH	0.9434	0.5055	0.7699	FALSE
256	Julio Teheran	Location	CH	0.4625	0.0791	0.5490	FALSE
257	Julio Teheran	H.brk2	CH	0.0000	0.1039	0.0000	TRUE
258	Julio Teheran	Speed	CH	0.0288	0.0907	0.0157	TRUE
259	Julio Teheran	Rise	CU	0.6091	0.0001	0.0000	FALSE
260	Julio Teheran	BreakPoint	CU	0.7171	0.0001	0.0021	FALSE
261	Julio Teheran	Tot.brk	CU	0.0004	0.0048	0.4396	FALSE
262	Julio Teheran	Location	CU	0.9549	0.7036	0.7917	FALSE
263	Julio Teheran	H.brk2	CU	0.7006	0.0803	0.0045	FALSE
264	Julio Teheran	Speed	CU	0.0000	0.0520	0.0000	TRUE
265	Julio Teheran	Rise	FF	0.9998	0.4774	0.3128	FALSE
266	Julio Teheran	BreakPoint	FF	0.0000	0.0000	0.0000	FALSE
267	Julio Teheran	Tot.brk	FF	0.0000	0.1404	0.0003	TRUE
268	Julio Teheran	Location	FF	0.1088	0.1153	0.6601	FALSE
269	Julio Teheran	H.brk2	FF	0.0000	0.0101	0.0000	FALSE
270	Julio Teheran	Speed	FF	0.0000	0.0000	0.0132	FALSE
271	Julio Teheran	Rise	FT	0.2851	0.8684	0.5510	FALSE
272	Julio Teheran	BreakPoint	FT	0.0000	0.0000	0.4144	FALSE
273	Julio Teheran	Tot.brk	FT	0.9834	0.8489	0.6939	FALSE
274	Julio Teheran	Location	FT	0.0539	0.5095	0.4070	FALSE
275	Julio Teheran	H.brk2	FT	0.0000	0.0000	0.0000	FALSE
276	Julio Teheran	Speed	FT	0.0000	0.0001	0.0000	FALSE
277	Julio Teheran	Rise	SL	0.0253	0.0000	0.0000	FALSE
278	Julio Teheran	BreakPoint	SL	0.0003	0.0000	0.0000	FALSE
279	Julio Teheran	Tot.brk	SL	0.7805	0.0560	0.0264	FALSE
280	Julio Teheran	Location	SL	0.0392	0.8700	0.0874	FALSE
281	Julio Teheran	H.brk2	SL	0.0000	0.0000	0.0000	FALSE
282	Julio Teheran	Speed	SL	0.0000	0.0000	0.0000	FALSE
283	Justin Grimm	Rise	CU	0.9973	0.4366	0.6434	FALSE
284	Justin Grimm	BreakPoint	CU	0.0000	0.0063	0.0000	FALSE
285	Justin Grimm	Tot.brk	CU	0.1410	0.8400	0.1303	FALSE
286	Justin Grimm	Location	CU	0.1368	0.5227	0.2401	FALSE
287	Justin Grimm	H.brk2	CU	0.0000	0.0019	0.0000	FALSE
288	Justin Grimm	Speed	CU	0.0001	0.0000	0.0000	FALSE
289	Justin Grimm	Rise	FF	1.0000	1.0000	1.0000	FALSE
290	Justin Grimm	BreakPoint	FF	0.0343	0.0000	0.0000	FALSE
291	Justin Grimm	Tot.brk	FF	0.5331	0.0265	0.0433	FALSE
292	Justin Grimm	Location	FF	0.3057	0.7943	0.8093	FALSE
293	Justin Grimm	H.brk2	FF	0.2577	0.0871	0.8407	FALSE
294	Justin Grimm	Speed	FF	0.0000	0.0000	0.0000	FALSE

Appendix G: Logistic Regression Model Validation

The purpose of this Appendix is to explain and show the results of the logistic regression model validation study. For an analysis of the model coefficients, see Appendix H.

Below are the logistic regression models. Two are shown for each pitch type. The first is the model based on the entire dataset and the second model is the first sample set whose predictions were validated (fell within the 95% confidence interval). The first predicts the total number of homeruns exactly; the second predicts the home runs of the validation sample and are shown at bottom. The reason for showing the first model is the coefficients exactly describe the relationship between the variables which explain the home runs exactly. The reason for showing the second model is to gain a feel for the behavior of the sample.

Remarks:

1. We did not detect any important interactions in model development. We did not look for quadratic or higher order terms.
2. The reason for the NA in the LL row is because the four pitcher-batter match-ups are categorical and an arbitrary one is set to the baseline, or zero. All statistical outputs report this as, "`Coefficients: (1 not defined because of singularities)`"
3. It is of interest to see where the sample models coefficients vary from the full data model. For this sample set, the coefficients with a different sign than the full data coefficients are marked in red: curveball (intercept + breakpt), two-seam fastball (rise*), and slider (location). Of these variations, only rise was statistically significant. What this implies is that the statistically significant components are being detected by the sample.
4. Here is a comparison table for all of the coefficients for the full data model.

	CH	CU	FF	FT	SI	SL
Intercept	-0.256	2.857	-1.455	-6.184 ***	-4.491 .	1.120
rise	3.812 **	-4.750 ***	7.031 *	14.905 *	0.647	2.539
breakpt	-0.237 ***	-0.050	-0.223 ***	-0.290 ***	-0.093 *	-0.212 ***
tot.brk	-0.806 ***	-0.981 ***	-0.089 ***	-0.321 ***	-0.238 **	-0.738 ***
h.brk2	-0.369 ***	-0.576 ***	0.131 *	-1.334	-0.398 *	-0.269 **
loc	-0.026	0.033	-0.006	0.007	-0.010	0.014
start_speed	-0.054 ***	-0.124 ***	-0.071 ***	-0.056 ***	-0.071 ***	-0.074 ***
Height	0.554 *	0.974 ***	0.571 ***	1.172 ***	1.139 ***	0.487 **
RR	0.558 *	0.770 ***	0.247 **	0.652 **	0.305	0.446 **
RL	0.231	0.631 **	0.158 .	0.972 ***	0.851 ***	0.645 ***
LR	0.158	0.463	0.183 *	0.818 ***	0.634 **	0.530 ***
LL	NA	NA	NA	NA	NA	NA

Table: Full logistic regression model coefficients, summarized for comparison.

Change-up (CH)

Full Prediction Model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4058  -0.1586  -0.1253  -0.0969   3.4573

            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.25573    1.68862  -0.151 0.879626
rise         3.81226    1.30329   2.925 0.003443 **
breakpt     -0.23724    0.03688  -6.433 1.26e-10 ***
tot.brk     -0.80638    0.05684 -14.188  < 2e-16 ***
h.brk2      -0.36878    0.10785  -3.419 0.000628 ***
loc         -0.02637    0.02881  -0.915 0.360038
start_speed -0.05392    0.01164  -4.631 3.65e-06 ***
Height       0.55397    0.21722   2.550 0.010763 *
RR           0.55776    0.25483   2.189 0.028613 *
RL           0.23081    0.25232   0.915 0.360332
LR           0.15834    0.25551   0.620 0.535457
LL                NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Null deviance: 7525.5  on 69487  degrees of freedom
Residual deviance: 7234.4  on 69477  degrees of freedom
AIC: 7256.4

Number of Fisher Scoring iterations: 8
```

Sample Prediction Model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3929  -0.1577  -0.1256  -0.0975   3.3858

            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.76211    2.44010  -1.542  0.12313
rise         3.83592    1.66423   2.305  0.02117 *
breakpt     -0.21573    0.04952  -4.356 1.32e-05 ***
tot.brk     -0.76892    0.08118  -9.471  < 2e-16 ***
h.brk2      -0.26750    0.15510  -1.725  0.08457 .
loc         -0.04035    0.04095  -0.985  0.32440
start_speed -0.05399    0.01650  -3.273  0.00106 **
Height       1.00650    0.31170   3.229  0.00124 **
RR           1.02147    0.46130   2.214  0.02680 *
RL           0.75866    0.45798   1.657  0.09761 .
LR           0.71071    0.46099   1.542  0.12315
LL                NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Null deviance: 3730.3  on 34743  degrees of freedom
Residual deviance: 3590.9  on 34733  degrees of freedom
AIC: 3612.9

Number of Fisher Scoring iterations: 8
```

Curveball (CU)

Full Prediction Model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4768  -0.1363  -0.1101  -0.0880   3.5492

            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.85658    2.33286   1.224 0.220765
rise        -4.74697    1.34115  -3.539 0.000401 ***
breakpt     -0.05040    0.03550  -1.420 0.155668
tot.brk     -0.98060    0.10485  -9.353  < 2e-16 ***
h.brk2      -0.57587    0.12818  -4.493 7.04e-06 ***
loc          0.03329    0.03458   0.963 0.335723
start_speed -0.12453    0.01651  -7.545 4.53e-14 ***
Height       0.97421    0.27925   3.489 0.000485 ***
RR           0.77048    0.23291   3.308 0.000939 ***
RL           0.63109    0.23528   2.682 0.007313 **
LR           0.46251    0.25403   1.821 0.068654 .
LL                NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Null deviance: 4873.9  on 56362  degrees of freedom
Residual deviance: 4705.5  on 56352  degrees of freedom
AIC: 4727.5

Number of Fisher Scoring iterations: 8
```

Sample Prediction Model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3334  -0.1344  -0.1098  -0.0885   3.5338

            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.11627    3.42955  -0.325 0.744813
rise        -7.59211    2.30230  -3.298 0.000975 ***
breakpt      0.02863    0.05563   0.515 0.606832
tot.brk     -0.77642    0.15210  -5.105 3.31e-07 ***
h.brk2      -0.43729    0.18270  -2.394 0.016685 *
loc          0.01921    0.04988   0.385 0.700110
start_speed -0.07522    0.02478  -3.036 0.002399 **
Height       0.82820    0.40171   2.062 0.039239 *
RR           0.76666    0.33721   2.274 0.022993 *
RL           0.51666    0.34205   1.510 0.130926
LR           0.46767    0.36807   1.271 0.203877
LL                NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Null deviance: 2358.0  on 28181  degrees of freedom
Residual deviance: 2287.8  on 28171  degrees of freedom
AIC: 2309.8

Number of Fisher Scoring iterations: 8
```

Four seam fastball (FF)

Full Prediction Model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4161  -0.1488  -0.1383  -0.1282   3.6185

             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.455463    0.97350  -1.495 0.134892
rise         7.031198    2.78916   2.521 0.011705 *
breakpt     -0.223464    0.05499  -4.064 4.83e-05 ***
tot.brk     -0.088592    0.02518  -3.518 0.000435 ***
h.brk2       0.130535    0.05858   2.228 0.025869 *
loc         -0.005720    0.01552  -0.368 0.712546
start_speed -0.071052    0.00727  -9.775  < 2e-16 ***
Height       0.571065    0.11329   5.041 4.64e-07 ***
RR           0.247179    0.08668   2.852 0.004348 **
RL           0.157528    0.08850   1.780 0.075080 .
LR           0.182606    0.09293   1.965 0.049415 *
LL                 NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null deviance: 27160  on 248063  degrees of freedom
Residual deviance: 27007  on 248053  degrees of freedom
AIC: 27029

Number of Fisher Scoring iterations: 8
```

Sample Prediction Model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4398  -0.1486  -0.1373  -0.1265   3.8016

              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.8454627  1.3928635  -1.325  0.18519
rise         9.1940882  4.5531604   2.019  0.04346 *
breakpt     -0.2887096  0.0920175  -3.138  0.00170 **
tot.brk     -0.0966293  0.0357768  -2.701  0.00692 **
h.brk2       0.0873742  0.0836052   1.045  0.29599
loc         -0.0006439  0.0219955  -0.029  0.97665
start_speed -0.0747184  0.0104071  -7.180 6.99e-13 ***
Height       0.7074046  0.1614638   4.381 1.18e-05 ***
RR           0.2156443  0.1206126   1.788  0.07379 .
RL           0.1048190  0.1234823   0.849  0.39596
LR           0.0802937  0.1305786   0.615  0.53862
LL                  NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null deviance: 13455  on 124031  degrees of freedom
Residual deviance: 13366  on 124021  degrees of freedom
AIC: 13388

Number of Fisher Scoring iterations: 8
```

Two seam fastball (FT)

Full Prediction Model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.2919  -0.1450  -0.1266  -0.1097   3.4453

             Estimate Std. Error  z value Pr(>|z|)
(Intercept) -6.183748    1.70426  -3.628 0.000285 ***
rise        14.905262    7.21491   2.066 0.038838 *
breakpt     -0.290038    0.08114  -3.574 0.000351 ***
tot.brk     -0.320783    0.04356  -7.364 1.79e-13 ***
h.brk2      -0.133720    0.11449  -1.168 0.242808
loc          0.007211    0.02547   0.283 0.777053
start_speed -0.056419    0.01334  -4.229 2.35e-05 ***
Height       1.171988    0.19146   6.121 9.29e-10 ***
RR           0.652050    0.20622   3.162 0.001567 **
RL           0.971567    0.20497   4.740 2.14e-06 ***
LR           0.818335    0.21086   3.881 0.000104 ***
LL                 NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null deviance: 9805.1  on 99316  degrees of freedom
Residual deviance: 9660.2  on 99306  degrees of freedom
AIC: 9682.2

Number of Fisher Scoring iterations: 8
```

Sample Prediction Model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3288  -0.1493  -0.1290  -0.1105   3.4618

            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.33162    2.37261  -2.247 0.024630 *
rise        -0.32678   26.51543  -0.012 0.990167
breakpt     -0.24275    0.15649  -1.551 0.120846
tot.brk     -0.33501    0.05969  -5.612 2.00e-08 ***
h.brk2      -0.32295    0.15749  -2.051 0.040297 *
loc          0.03482    0.03474   1.003 0.316084
start_speed -0.06071    0.01860  -3.263 0.001101 **
Height       1.11484    0.26413   4.221 2.43e-05 ***
RR           0.77626    0.30212   2.569 0.010189 *
RL           1.14928    0.29990   3.832 0.000127 ***
LR           0.94837    0.30765   3.083 0.002052 **
LL                NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null deviance: 5110.4  on 49658  degrees of freedom
Residual deviance: 5020.1  on 49648  degrees of freedom
AIC: 5042.1

Number of Fisher Scoring iterations: 9
```

Sinker (SI)

Full Prediction Model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.2853  -0.1440  -0.1232  -0.1048   3.5445

            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.49071    2.57880  -1.741 0.081614 .
rise         0.64687    1.47016   0.440 0.659937
breakpt     -0.09337    0.03881  -2.406 0.016132 *
tot.brk     -0.23773    0.07403  -3.211 0.001322 **
h.brk2      -0.39803    0.18082  -2.201 0.027719 *
loc         -0.01046    0.03966  -0.264 0.791969
start_speed -0.07119    0.02016  -3.532 0.000413 ***
Height       1.13940    0.29536   3.858 0.000114 ***
RR           0.30544    0.23467   1.302 0.193073
RL           0.85135    0.23151   3.677 0.000236 ***
LR           0.63446    0.24325   2.608 0.009101 **
LL                NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null deviance: 4138.3  on 42804  degrees of freedom
Residual deviance: 4062.9  on 42794  degrees of freedom
AIC: 4084.9

Number of Fisher Scoring iterations: 8
```

Sample Prediction Model

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.970033   3.641283  -1.640 0.101101
rise         0.611441   2.062094   0.297 0.766837
breakpt     -0.093762   0.056174  -1.669 0.095091 .
tot.brk     -0.217046   0.104176  -2.083 0.037210 *
h.brk2      -0.551934   0.254639  -2.168 0.030196 *
loc         -0.007428   0.055927  -0.133 0.894336
start_speed -0.071615   0.028136  -2.545 0.010919 *
Height       1.397942   0.422554   3.308 0.000939 ***
RR           0.441417   0.329397   1.340 0.180221
RL           0.851783   0.328577   2.592 0.009533 **
LR           0.601841   0.346136   1.739 0.082080 .
LL                 NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null deviance: 2078.7  on 21402  degrees of freedom
Residual deviance: 2038.7  on 21392  degrees of freedom
AIC: 2060.7

Number of Fisher Scoring iterations: 8
```

Slider (SL)

Full Prediction Model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3930  -0.1501  -0.1229  -0.0995   3.5098

            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.11977    1.49743   0.748 0.454583
rise         2.53903    1.91372   1.327 0.184592
breakpt     -0.21224    0.03087  -6.876 6.17e-12 ***
tot.brk     -0.73846    0.05078 -14.543  < 2e-16 ***
h.brk2      -0.26852    0.10089  -2.661 0.007782 **
loc          0.01440    0.02326   0.619 0.535888
start_speed -0.07433    0.01139  -6.527 6.69e-11 ***
Height       0.48671    0.17982   2.707 0.006796 **
RR           0.44570    0.14035   3.176 0.001495 **
RL           0.64494    0.14632   4.408 1.05e-05 ***
LR           0.53035    0.15728   3.372 0.000746 ***
LL                NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null deviance: 11176  on 111642  degrees of freedom
Residual deviance: 10873  on 111632  degrees of freedom
AIC: 10895

Number of Fisher Scoring iterations: 8
```

Sample Prediction Model

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.70864    2.11556  -0.808 0.419289
rise         3.99531    2.41495   1.654 0.098045 .
breakpt     -0.21740    0.04212  -5.161 2.45e-07 ***
tot.brk     -0.67000    0.07150  -9.371  < 2e-16 ***
h.brk2      -0.28112    0.14260  -1.971 0.048675 *
loc         -0.01482    0.03286  -0.451 0.651978
start_speed -0.07177    0.01606  -4.469 7.85e-06 ***
Height       0.87830    0.25362   3.463 0.000534 ***
RR           0.46348    0.19851   2.335 0.019558 *
RL           0.70706    0.20644   3.425 0.000615 ***
LR           0.52336    0.22288   2.348 0.018869 *
LL                NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null deviance: 5616.5  on 55821  degrees of freedom
Residual deviance: 5470.2  on 55811  degrees of freedom
AIC: 5492.2

Number of Fisher Scoring iterations: 8
```

Final Results

Below are the observed HR, the HR explained by the model, and the differences. This particular sample model was quite accurate.

```
     Obs  Explained      Diff
CH   337  330.2965   6.703521
CU   214  197.5163  16.483701
FF  1220 1194.9813  25.018711
FT   404  447.5689 -43.568853
SI   178  179.7620  -1.762037
SL   484  490.6307  -6.630706
SUM 2837 2840.8     -3.8
```

Appendix H: Analysis of Logistic Regression Models

In Appendix G, we explained the logistic regression model validation. In this Appendix, we focus on the best model and seek to learn from it. To start, we look at the no-intercept model coefficients, below.

	CH	CU	FF	FT	SI	SL
rise	3.812 **	-4.750 ***	7.031 *	14.905 *	0.647	2.539
breakpt	-0.237 ***	-0.050	-0.223 ***	-0.290 ***	-0.093 *	-0.212 ***
tot.brk	-0.806 ***	-0.981 ***	-0.089 ***	-0.321 ***	-0.238 **	-0.738 ***
h.brk2	-0.369 ***	-0.576 ***	0.131 *	-1.334	-0.398 *	-0.269 **
loc	-0.026	0.033	-0.006	0.007	-0.010	0.014
start_speed	-0.054 ***	-0.124 ***	-0.071 ***	-0.056 ***	-0.071 ***	-0.074 ***
Height	0.554 *	0.974 ***	0.571 ***	1.172 ***	1.139 ***	0.487 **
RR	0.302	3.627	-1.208	-5.532 **	-4.185	1.565
RL	-0.025	3.488	-1.298	-5.212 **	-3.639	1.765
LR	-0.097	3.319	-1.273	-5.365 **	-3.856	1.650
LL	-0.256	2.857	-1.455	-6.184 ***	-4.491	1.120

Table: No-intercept logistic regression model coefficients, summarized for comparison.

Remarks:

1. **Various Models Attempted**. To attempt to understand the relationship between HR%, the 6 pitch components, pitch type, and pitcher-batter match-up, we constructed and compared different models. More work could be done, here, but this is a record of what we did by the time of writing.
   a. Summary: We started with an intercept model, with pitcher-batter match-up, for each pitch type, in the hopes of getting any explanation at all. This model explained HR% effectively, so we made some others, as follows. The full model and no-intercept model have the same coefficients for the pitch components, so we chose to use the no-intercept model in the body of the paper.
   b. The zero-intercept model turns out to be equivalent to the intercept model when handedness is incorporated. The coefficients, and statistical significance, of the pitch components and height remain identical. The connection is seen by taking the intercept from the intercept model, move it to the LL and add it to the RR, RL, and LR coefficients. Then the statistical significance for the handedness match-ups changes, making it preferable for interpreting the importance of the handedness match-ups – which is significant only for the four seam fastball.
   c. Interactions, including tot.brk with h.brk2 and loc - nothing interesting.
   d. No intercept model (no interactions) removing pitcher batter match-ups – coefficients very similar to the full no-intercept model. This is important because the interpretations will be the same whether the match-ups are included or not. The predicted HRs are within 0.5.

e. Pitch components only (no Height or match-ups) – coefficients shift, but are still near those of the no-intercept model, always with the same sign. The predicted HRs are within 5.
   - ➔ **Conclusion**: The explanatory power of the model is good. The model coefficients shown are a good representation of the sign and magnitude. There were no immediately detected interaction terms for the pitch components.

2. **Components**. The only components that are statistically significant across all pitch types are: tot.brk, start_speed, and Height. Breakpt and h.brk2 are significant for five of the six; rise for four of six.
   a. **Rise:** Positive for all models except curveball, where it is negative.
   b. **Breakpt:** Negative for all models, confirming the intuition that the further until the break, the less likely it is to be hit for a home run.
   c. **Tot.brk:** Negative and the most statistically significant of four of the six models (not FF and SI). This confirms our claim of the greatest importance of vertical break, and the claim that reducing the vertical break increases home run probability.
   d. **H.brk2:** The components are negative and statistically significant for all off-speed pitches. It is positive for FF and not statistically significant for FT. This helps explain the puzzling phenomenon of increased horizontal break and more home runs – more *h.brk2* is helping the HRs off of FF, which is the highest proportion pitch type.
   e. **Loc**: Not statistically significant for any model.
   f. **Height**. Positive and statistically significant for all models. This means the taller batters are hitting relatively more home runs. This is probably because taller batters are bigger and stronger, on average, but part of it could be that they experience less vertical break, on average.
   g. **RR, RL, LR, LL:** There is no statistically significant effect of handedness match-up in the models except for the two seam fastball. What stands out across pitch types is that the LL match-up gives lower HR% than the other match-ups.
      - ➔ **Conclusion:** For the most part, the HR% model mirrors our QOP model, except *loc* was not significant. The exceptions are *rise* for CU and *h.brk2* for FF. *tot.brk* was the most statistically significant. *Height* is significant, and improves the models, but handedness match-up is not.

3. **Interpretation by Pitch Type.** In order to identify the features of the models for these pitch types, it is helpful to look at their variation from the essence of our QOP model, focusing on the signs of the coefficients:

   QOP = –Rise + Breakpt + Tot.brk + H.brk2 – Loc + Speed

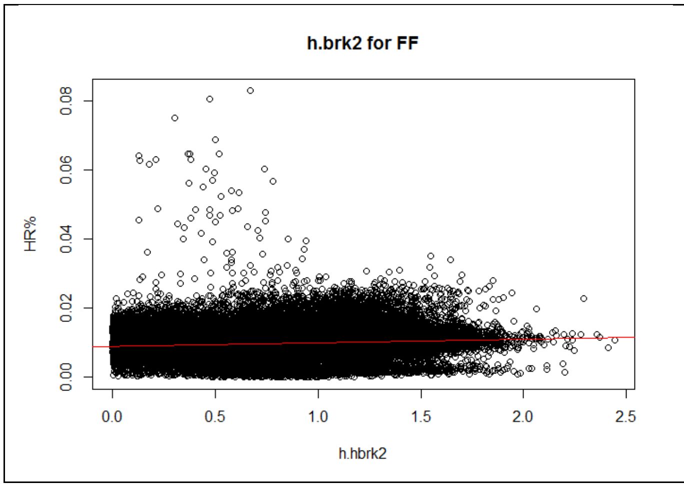
   Since for QOP bigger is better and for HR% bigger is worse, the coefficients switch, so that this is the expected sign of the coefficients:

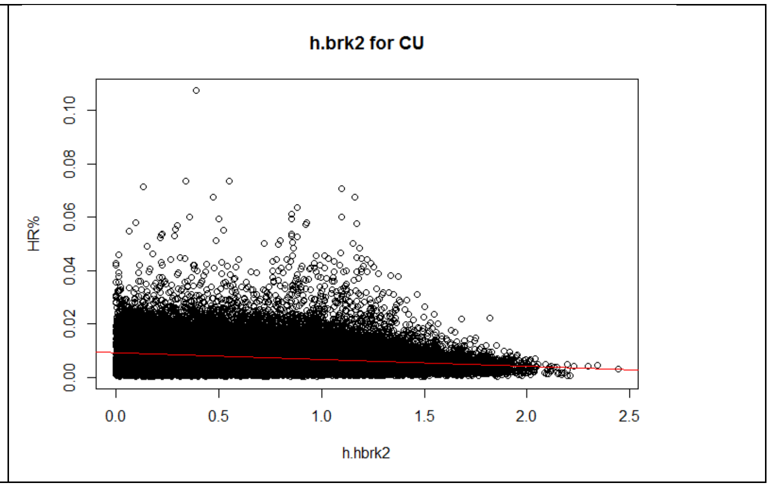
   HR% = Rise – Breakpt – Tot.brk – H.brk2 + Loc – Speed

   We will use this HR% model as a baseline, and added Height and {RR, RL, LR, LL}. Note that location is not statistically significant in any model and therefore will not be considered in this analysis.
   a. **CH:** Matches the pattern

b. **CU:** Matches the pattern except for *rise*, where more rise decreases the probability of HRs. This is unexpected, but it is the most statistically significant *rise* coefficient among the pitch types (p-value = 0.0004), so appears important for the underlying multivariate relationship. At the same time, the curveball has the highest *tot.brk* coefficient (-0.981) by a minimum of 22%. This seems to be linked to the nature of the curveball which has the smallest HR% (Appendix D).

c. **FF:** Matches the pattern except for *h.brk2*, where the increase in *h.brk2* increases *HR%.* This is one of the most striking surprises of these models. Here is a comparison of the *h.brk2* component for *FF* pitches and their model *HR%* versus the *FT* model:





There is an positive correlation for *FF*, and the contrast can be seen with the negative correlation with *CU.* While the scatterplots do show a relationship, it is weak. It should be remembered that this is only two dimensions of a multi-dimensional model.

d. **FT:** Matches the pattern

e. **SI:** Matches the pattern

f. **SL:** Matches the pattern

➔ **Conclusion:** All of the pitch types match the expected pattern, except *rise* for CU and *h.brk2* for FF.

Appendix I: Effect of New Pitchers

Since the thesis of this paper is that changes in 2017 pitching are resulting in lower pitch quality, as measured by QOP™, and better batter results, as measured by HR, another possible source of differences is the effect of new pitchers in 2017.

The PITCHf/x data show 539 returning pitchers with 635,386 pitches and 216 new pitchers with 94,010 pitches for a total of 755 pitchers pitching 729,396 pitches during regular and post season games[35]. Their statistics are as follows.

Component	Rise	Breakpt	Tot.brk	Loc	H.brk2	Speed	QOPA
Returning	0.01	1.24	3.60	1.93	0.75	87.87	4.57
New	0.01	1.22	3.53	1.91	0.69	88.14	4.51

Table. Mean pitch components of new versus returning pitchers in 2017.

The vertical break is lower for the new pitchers, which would contribute to the MLB-wide drop in Tot.brk in 2017. However, the horizontal break is also lower, which does not contribute to the MLB-wide drop. It should be noted that the QOPA is lower for the new pitchers, which brings down the QOPA average.

[35] For all analyses in this paper, we use regular and post-season games, including the all-star game. Pre-season games are not included. The reason is for consistency with our historic QOP analyses, and to maximize the number of pitches in the dataset. The one exception is except the home run Table 1 and Figure 1, where the commonly known regular season only numbers are used.