Steroids, Home Runs and the Law of Genius

Arthur De Vany

Professor Emeritus
Department of Economics
Institute for Mathematical Behavioral Sciences
University of California, Irvine
www.arthurdevany.com

`asdevany@uci.edu`

ABSTRACT

The greatest home run hitters are as rare as great scientists, artists, or composers. The greatest accomplishments in these fields all follow the same universal law of genius, as I show in this paper. There is no evidence that steroid use has altered home run hitting and those who argue otherwise are profoundly ignorant of the statistics of home runs, the physics of baseball, and of the physiological effects of steroids. There is no standard for great accomplishments, in home runs, in the sciences, and in the arts. Genius has its own way and the great achievements of McGwire, Sosa, and Bonds (they did it in that order) are of a piece with genius in other fields — they are the Bach, Beethoven, and Mozart of home runs.

The law of home runs established here generalizes the laws of extreme accomplishment developed by Pareto, Lotka, Price and Charles Murray. Economics seems to have little place for genius, but the growing importance of intellectual property calls for a greater understanding of extreme accomplishment. The stable Paretian model developed here will be of use to economists studying extreme accomplishment in other areas and has led to a new understanding of the motion picture industry (De Vany, 2004).

1. Introduction

Six bills are before Congress to impose drug testing and penalties on major league baseball and other professional sports (Kiele 2005). A Senate hearing was held this year in which Jose Canseco, Mark McGwire, and Rafael Palmeiro testified about steroid use in MLB. Many appear to believe that steroid use has caused an increase in home runs. Arguing that Palmeiro's suspension for testing positive for steroids has pushed the issue into the policy arena, Senators Stevens, McCain, and Hall of Fame pitcher and now Senator Jim Bunning are cosponsoring drug-testing bills. The intent of Congress can be inferred from the title of the hearings: "Restoring Faith in America's Pastime".

Before we can reach any conclusions about the contribution of steroids to performance in professional baseball, we first must know something about home run hitting. What was home run hitting like before there were steroids? What is it like now that there is some evidence of steroid use? In a nutshell, the answer is that there are no differences; I will demonstrate that conclusion in this paper.

What then accounts for the rise in MLB home runs? There are more games now; both games and home runs have doubled. Home runs per player have not changed in over forty years; the statistics are the same and year to year changes are just part of the natural variation. What about the spate of new records? Intermittence is part of home run hitting. When Babe Ruth set records in the 1920s they came in clusters too. The same thing happened when Roger Bannister broke the 4 minute mile; others followed in a burst.

Hitting home runs is an extraordinary feat. Hitting many of them is like winning many PGA championships, several tennis Grand Slams, and multiple World Chess championships. It is of a piece with other high accomplishment as measured by a scientist's citations in the scientific literature, recognition and productivity in the arts, and box office revenues in the movies. I show that the statistical law of home run hitting is the same as the laws of human accomplishment developed by Lotka (Nicholls 1926), Pareto (Pareto 1897), Price (Price 1963), and Murray (Murray 2003). My generalization of these laws is a stable Paretian probability distribution with a finite mean and an infinite variance. This makes it a "wild" statistical distribution, far different from the normal (Gaussian) distribution that people are tempted to use in their reasoning about home runs and most other things. Things are not so orderly in home runs; they are rather more like the movies (De Vany 2003) or earthquakes (Samorodnitsky and Taqqu 1994) than dry cleaning. Too many sportswriters are reading slight variations in averages as significant evidence of something (other than their own ignorance of statistics) as though home run hitting could be measured like beer sales at a stadium. There is no standard for the performances of the greatest home run hitters.

2.   A Brief Survey of Home Run Hitting

In 1961 there were 2730 MLB home runs hit in 1430 games with 1.909 home runs per league game, 0.041 home runs per player game, and a maximum 61 home runs by a single player.[1] Forty years later, in 2001, 5458 home runs were hit in 2429 games with 2.247 home runs per game, 0.0413 home runs per player game, and a maximum of 73 home runs by a single player. Home runs per at bat are the same in both years, 0.075 for players with 200 or more at bats.[2] Home runs per hit are 0.110 in 1961 and 0.125 in 2001, both well within a standard deviation of the 40 year average.

---

[1]The data are from the Baseball1.com web site, www.baseball1.com, archives.

[2]At bats per game have risen by a bit more than half an at bat from 1961 to 2001; at bats were above the 40 year average in both years.

Babe Ruth's record was exceeded in both years.

The annual variation in home runs is driven by the great performances of a few players as in the Maris/Mantle/Gentile year of 1961 (61, 54, and 46 home runs respectively) or the McGwire/Sosa years of 1998 (70 and 66 with 56 from Griffey) and 1999 (65 and 63 with 45 from Vaughn). Or the Bonds/Sosa/Thome year of 2001 (73, 64, and 49). Bonds, McGwire and Sosa are truly exceptional. Their hitting is 10 standard deviations above the mean (but, caution, home run hitting does not follow a normal distribution so the standard deviation is a measure over the sample, not a property of the distribution itself). Relative to hitters with 200 or more at bats, their performances are about 7 standard deviations above the mean. These guys are profoundly different from the average player. But, this is true of Ruth, Foxx, Gehrig, Greenberg, Williams, Mantle, Maris, Mays, Kiner, Aaron, Schmidt, and all the well-known hitters.

Every big time home run hitter hits far above the average player, who hits just over 3 home runs per year. Even Barry Bonds record-setting 73 home runs is a smaller leap in his performance than was Roger Maris' 61 home runs. Mark McGwire's (then) record 70 home runs was only 12 home runs above his 58 of the year before (scattered over both leagues because of a mid-year trade) and just 1.5 standard deviations above his mean. (Remember, these are only sample statistics since the standard deviation of the distribution does not exist. I use them only because they are familiar to most readers.)

Among the premiere home run hitters, home runs per hit has hardly changed for more than 40 years; if anything there was a slight dip in the 1970s and early 1980s. The notable exceptions are Mark McGwire, Barry Bonds, and Sammy Sosa. McGwire's home runs per hit has always been larger than life, never less than 0.303 from 1987, rising into the 0.4 range from 1995 and on, with a peak of 0.52 in his shortened 2001 year. Bonds and Sosa also consistently were above 0.3 with Bonds reaching his career high of 0.467 in his record setting year of 2001. But, in earlier years, Killebrew, Maris, Mantle, Kingman, Schmidt, Aaron, Jackson, Stargell, Fielder, Buhner, and Williams exceeded 0.3 home runs per hit and hit from 40 to 61 home runs. The less extraordinary hitters in the 20th, 50th, 70th, and even the 90th percentiles in home runs per hit have not changed over 40 years of MLB hitting.

The league leading home run output has moved up and down a bit more, declining a bit through the late 1970s and early 1980s and only coming back up to the 1961 level in the late 1990s. There is evidence the hitters began "swinging for the fences" more, starting in 1993. Modern players strike out more, but they are not as efficient in hitting home runs as players of the past. The number of home runs hit per strike out is a bit less now than it was in 1959–1962, amazing since the strike zone was shrunken a few years ago.

On the whole, there is no evidence of increased home run hitting in MLB that cannot be accounted for by the expanded schedule–more teams and games–a slight rise in the number of at bats per game, a rise in strike outs in spite of a smaller strike zone, and the feats of three magical hitters during four years when they made history. Steroids do not come into the picture, nor is

there any need to invoke explanations that go beyond the natural variation of home run hitting, at bats, chance, and the laws of extreme human accomplishment. So, why all this attention on home runs?

3. Records and Variations

The controversy over steroid use seems to be about records. The records seem to be falling too rapidly for the critics of baseball. The Babe's record of 60 home runs stood for so many years that it seemed to be inviolable. It took 34 years before Roger Maris broke Babe Ruth's 1927 record of 60 home runs in 1961 with 61 home runs. It took The Babe 36 years to break the former record. If the statistical law of doubling time between records were true, then Maris' record should have held for 68 years. But, it was broken 37 years later in 1998 by two players, Mark McGwire (70 home runs) and Sammy Sosa (66 home runs). And twice again in 1999 by McGwire (66 home runs) and Sosa (63 home runs). And then by Barry Bonds in 2001 (73 home runs) and Sammy Sosa (64 home runs). The Babe's record held for 34 years, Roger Maris' record stood 37 years (longer than The Babe's by 3 years), but Mark McGwire's record stood just 3 years. The Babe's record was broken 7 times since 1927 by 4 players, but 6 of those record-breaking performances occurred in a brief and remarkable span of 4 years from 1998 through 2001. Sosa broke the Ruth/Maris record 3 times in 4 years, McGwire broke it twice, and Bonds broke it once while setting a new all-time record. It was a historic period in baseball of extraordinary accomplishment by three amazing players.

Yet, one can go back in time and find comparable record-setting. In 1919 Babe Ruth hit 29 home runs (140 games) to break Ned Williamson's 36 year-old record of 27 home runs in 1884 (112 game schedule). Ruth broke his own record the next year by hitting 54 (154 game schedule). He was aided in his 1920 record by an expanded game schedule (154 up from 140 the year before) and a change in the baseball. The coefficient of restitution of the baseball was increased in 1920, making it rebound off the bat with higher velocity. He broke the record again in 1921 with 59 home runs (154 game schedule) and then set his historic mark of 60 in 1927. Ned Williamson's record stood for 36 years and was broken three years in succession by Ruth, who then moved it to the magical 60 home runs just 6 years later. In a 9 year period the home run record was advanced 4 times, not so different from the latest run of records.

4. The Babe

It seems fitting to begin with The Babe as a standard for home runs; his career home runs are shown in Figure 2. He averaged 37.05 home runs per year in his 19 years as a hitter (he was a pitcher before that). This was in a era when the average number of home runs a year by players with 200 or more at bats was 6.11. In the 17 years that he had at least 200 at bats, The Babe hit an average of 41.12 home runs a year, had a 0.342 batting average, 126.5 runs batted in, 161.88

hits and struck out just 72.82 times a year. There was nothing like him and he still is the greatest there ever was.

The pace he sustained over that 17 year stretch has never been matched. But, even his great home run hitting was volatile: he hit as few as 11 home runs and as many as 60 home runs in years he had 200 or more at bats. The standard deviation of his home runs is 13.56, about one third of his average. His 60 home runs was an event about one and a half standard deviations above his average. The distribution of his home runs is unlike any one else's. Most hitters have the bulk of their home runs piled on the lower numbers, with a few years of high output. The Babe had his piled on the high numbers, with a lot of years of high output and only a few of low output. His frequency distribution seems as though it is pushed against the limits of human performance. He hit 46 home runs three times, 47 home runs one time, 49 home runs one time, 54 home runs twice, 59 home runs one time, and his famous 60 home runs just once. Only Sammy Sosa comes close and not very close at that (of which, more later).

Consider that Ruth played 151 games and had 540 at bats in his record year. He hit 0.39 home runs per game and 0.11 home runs per at bat. When Maris hit his 61 home runs, he played 161 games and had 590 at bats. He hit 0.38 home runs per game and 0.10 home runs per at bat. Neither played all the games in the season, 154 in 1927 and 162 in 1961. Had Ruth played the same number of games as Maris and maintained his home run rate (big assumptions), he might have hit 63 home runs.[3]

In his most productive year of 1998, Sammy Sosa hit 66 home runs in 159 games with 643 at bats. He hit 0.45 home runs per game and 0.10 home runs per at bat, exactly the home runs per at bat that Maris hit in 1961. If Ruth had had Sosa's at bats he could have hit 71 home runs. More famously, in 1998, Mark McGwire hit 70 home runs in 155 games and 509 at bats. He hit 0.45 home runs per game and 0.14 home runs per at bat. Had Ruth had McGwire's at bats he might have hit only 51 home runs. Had McGwire had Ruth's at bats, he might have hit 76 home runs. Had McGwire had Sosa's at bats, he might have hit 84 home runs.

And then there is Barry Bonds. He hit 73 home runs in 153 games and 467 at bats. He hit 0.48 home runs per game and 0.15 home runs per at bat. Had he had Ruth's at bats, he might have hit 81 home runs. Had he had Sosa's at bats, he might have hit 96 home runs.

These calculations give us an idea of the possibilities that exist within the physical constraints of the sport of modern baseball. The examples show that small variations in the variables can lead to large differences. When a positive variation in at bats coincides with a small rise in home runs per at bats, big differences in home runs can result and records will fall. Variation in these numbers is the norm, there is no constancy in hitting performance and, if there were, home run

_____

[3]That was the argument behind marking Roger Maris' record with an asterisk. Few players actually play all the games in a season, so this controversy seemed more a matter of preserving The Babe's history than recognizing records.

hitting would be almost boring. Records are all about variation and numbers of chances.

Home run hitting chances have doubled as the number of teams in MLB nearly doubled from 1959 to 2005. The number of players has more than doubled. Players are recruited from around the world now so that baseball draws from a wider and more diverse pool of talent now than ever before. If there was one Babe Ruth in 120 million people in the United States in 1927, there may be many of them in a world of over 6 billion people. Records should fall.

5. The Distribution of Home Runs

The first thing we must do is face up to the over arching fact about hitting home runs; few do it well, even among the best players in the world. It is a difficult task to hit home runs in the major leagues. To hit many home runs in a year is an extraordinarily difficult task. Over 60 percent of players hit fewer than 10 home runs a year. Less than 20 percent hit more than 20 home runs a year. The number who hit more than 30 is a select group, hitting 40 or more home runs is a task accomplished by the elite hitters. In many years, 31 to 40 home runs led MLB. In only 13 of the past 45 years did the MLB leading home run hitter hit 50 or more home runs. Only 5 players in the history of MLB hit 60 or more home runs in a single season, just 4 of them accomplished this feat 7 times in over 46,000 player seasons since 1959 through 2004.

This means that it is useless to try to analyze home run hitting as though it were a "normal" activity where we might use averages and standard deviations. These sorts of statistics are meaningless because home run hitting is far from a normally distributed activity. Few elite accomplishments of great difficulty are as has been established by scientists of human accomplishment such as Lotka, Pareto, Price, or Murray.

A picture makes this clear. In Figure 1 the distribution (histogram of empirical frequencies) of home runs is shown for two record home run years in MLB. You can't tell them apart because the distribution has not changed in the 40 years that separate these years (this is confirmed below more rigorously). Note the high peak to the left at from 0 to 2 home runs and the sharp fall off to the right at higher numbers. The distribution has a high, sharp, and narrow peak at the left and a lot of skew to the right. The mean is all the way out to the 75th percentile which shows how unreliable the mean is—you are talking about the top 25 percent of players when you are talking about the MLB average number of home runs hit in a year.

Even more troublesome is that long and thick tail to the right. This upper tail of the distribution is called "heavy" because it does not decline as fast as it would if home runs were normally distributed. I show later that the upper tail declines at $X^{-1.64}$, a slow power law decay relative to the rapid exponential decay $\exp \frac{-X^2}{2}$ of the normal distribution. This means that the average is strongly affected by extreme values and that the probability mass far out in the tails of the home run distribution far exceeds the mass of a normal distribution. Put another way, hitting 60 home runs is 8 standard deviations above the mean, an event that would not occur in millions of years

of baseball if home run hitting were normally distributed.

The distribution of home runs is skewed far to the right, not symmetrical like a normal distribution. Its shape is described by its kurtosis, a name for its peaked and skewed shape. A normal distribution has a skew of 0, but the distribution of home runs has a skew of 3. The normal distribution has a kurtosis of 3 but the distribution of home runs has a kurtosis of 13. So, nothing is normal here, as we should expect because home run hitting is an extreme form of human accomplishment.

This makes the task of ascertaining any changes in home run hitting difficult. The object that we must analyze is the distribution, not the usual means that people are familiar with. If there is a change in home run hitting over the years it is to be found in a shift of the distribution away from low numbers to higher ones. The bottom of the distribution must pull up and the far right tail must push out or become more heavy. We can examine this problem by looking at the features of the distribution shown in Tables 3 and 4. This table tabulates for each year from 1959 through 2004 the mean, standard deviation, skewness, 75th percentile, kurtosis, and maximum home runs.

The mean and the 75th percentile are close to one another and they track one another over the years. The standard deviation is large, around twice the mean, and increases a bit relative to the mean when the maximum home runs (max) rises. The mean is influenced by the max home runs, rising when max is high and falling when it is low. This is typical of a process that is driven by extreme events because the mean and 75th percentile are heavily influenced by the few extreme accomplishments in the upper tail, far above the mean but so large they exert influence. The distribution is skewed to the right, with a positive skew of approximately 3. The excess kurtosis, a measure of skew and a heavy tail is about 13 over the whole sample. Far from normal where the excess kurtosis would be 0.[4]

The 75th percentile typically varies between 3 and 4, with a couple of years at 2 when the max home runs is low. All the distributions for all the years have high kurtosis, positive skew, a high standard deviation relative to the mean, and a mean near the 75th percentile.

Clearly, these measures of the shape and limits of the distribution of home runs shows that nothing has changed. The distribution has the same shape year after year. Even the slight variation of mean home runs is consistent with the shape and strong influence of extreme events. In fact, the mean has its own statistical distribution because it is a linear transformation of stable distributed variables; thus, it too must have a stable distribution. I show that the distribution of home runs is a stable distribution.

---

[4]The excess kurtosis is kurtosis - 3, a way of stating kurtosis relative to the normal distribution, which has a kurtosis of 3. All kurtosis statistics mentioned in this paper are excess kurtosis.

6.   The Law of Home Runs

Lotka, Price, Pareto and, most recently, in his wonderful book, *Human Accomplishment*, Charles Murray have studied the extremes of human accomplishment in literature, science, and the arts. Murray also developed some of the statistics of sports accomplishment.

The models of these authors are power law statistical distributions of the form $Prob[x] \sim x^{-\alpha}$. For example, Lotka's law of the number of authors who published n works is $Prob[n] \sim C/n^{\alpha}$. Lotka found that $\alpha \simeq 2$ in his investigation of article publishing. Pareto's law is also a power law of the form $Prob(x \geqslant k) \sim Cx^{-\alpha}/k$. Pareto found a value of $\alpha \simeq 1.5$ in the distribution of income. Defining $w[n]$ as the function representing the work done by n people, Price found that $Prob[w(n) = 0.5] \sim n^{-.5}$ which is to say that half the work done by n scientists is done by a number equal to the square root of n. Price's law is a power law expressed in terms of the median of the distribution. Murray found that the number of PGA victories, number of victories in PGA majors, number of Boston and New York Marathon victories, batting championships, tennis Grand Slams, and points won in World Chess Championships are all power laws of the Lotka/Pareto form. Unfortunately, he does not report his various estimates of the α values for these accomplishments. The Lotka/Pareto/Murray laws are asymptotic approximations to the upper tail of a stable distribution, to which I now turn.

The number of home runs a player hits in a year or a career is the sum of independent random variables. We know that sums of independently distributed random variables converge to a stable distribution. The normal distribution is a member of the class of stable distributions, but we can easily see the home run distribution is not normal. We require a more general law. The general law of large numbers says that sums of iid random variables converge to a stable distribution. The generalization from normal to the broader class of stable distributions comes from the recognition that the random variables need not have a finite variance or even a finite mean.

A stronger statement is this: the distribution function F possesses a domain of attraction if and only if it is stable. The limiting distribution function of independent r.v.'s X_k belongs to the domain of attraction of F if there exist normalizing constants $a_n, b_n > 0$ such that the distribution of $(\sum_{i=1}^{n} X_i - a_n)/b_n \to F$. Thus, we might expect that extraordinary accomplishment might converge to a stable distribution because it is, in some sense, unlimited in the amount by which a singular accomplishment might exceed others. This is the intuitive meaning of the infinite variance and of the possibly infinite mean.

The Normal (Gaussian) distribution is stable (as it must be because it is the attractor of sums of random variables), but it is the only stable distribution with a finite variance. The Pareto, Lévy, and Cauchy distributions are the other (named) stable distributions. But, there are many other members of the stable class, whose functional forms cannot be given. Since they do not have specific functional forms, they must be expressed through the more general characteristic function.

Lévy (Levy 1954) characterized the class of stable distributions through their characteristic

functions. A stable distribution $X \sim \mathbf{S}(\alpha, \beta, \gamma, \delta)$ is a four-parameter distribution with characteristic function given by

$$C(t) = \begin{cases} \exp(-\gamma^\alpha |t|^\alpha [1 + i\beta(\tan \frac{\pi\alpha}{2})(\text{sign } t)((\gamma|t|)^{1-\alpha} - 1)] & \text{if } \alpha \neq 1 \\ \exp(-\gamma|t|[1 + i\beta\frac{2}{\pi}(\text{sign } t)(\ln|t| + \ln\gamma)] & \text{if } \alpha = 1 \end{cases}$$

Estimation of the stable distribution of home run hitting produced values of the parameters shown in Table 1.[5]

Table 1: Maximum Likelihood Estimates of the Parameters of the Law of Home Runs

	α	β	γ	δ	
	Index	Skewness	Scale	Location	Log-Likelihood
HOME RUN HITTING					
α-stable	1.6422	1.00	6.219	12.30	-39294.2

The exponent α is a measure of the probability weight in the upper and lower tails of the distribution; it has a range of $0 < \alpha \leq 2$ and the variance of the stable distribution is infinite when $\alpha < 2$. The basin of attraction is characterized by the tail weight of the distribution (α). This remarkable feature tells us that the weight assigned to extreme events is the key distinguishing property of a stable probability distribution. The skewness coefficient $-1 \leq \beta \leq 1$ is a measure of the asymmetry of the distribution. Stable distributions need not be symmetric; they may be skewed more in their upper tail than in their lower tail. The scale parameter γ must be positive. It expands or contracts the distribution in a non-linear way about the location parameter δ which is the center of the distribution.

The tails of a stable distribution are Paretian and moments of order ≥ 2 do not exist when $\alpha < 2$. This is typical of many extraordinary accomplishments, as seen in the works of Lotka, Pareto, and Murray. Its mean need not exist for values of $\alpha < 1$. When $\alpha = 2$, the stable distribution is the normal distribution with a finite variance. The parameter α is called the tail weight because it describes how rapidly the upper tail of the distribution decays with larger outcomes of the random variable; smaller α implies a less rapid decay of probability.

The distribution of home runs is skewed to the right with a value of $\beta = 1.00$. The probability density function is shown in Figure 4. Note the peak probability (mode) is well to the left at 7.8 home runs; this is the statistically most common number of home runs by a player in a year. The fit to the data is remarkable as shown by the cumulative theoretical and empirical distributions displayed in Figure 5.

---

[5]I used Dr. Robert Rimmer's *Mathematica* package to estimate the parameters of the stable distribution. It is described in the *Mathematica Journal* (Rimmer and Nolan 2005).

Its central mass is located at $\delta = 12.3$., just above the average. The important tail index is $\alpha = 1.642$. Since $1 < \alpha < 2$ the distribution has a finite mean but an infinite variance. This reflects the large probability mass that is located out in the right hand tail. It is a warning that we can never sample the whole distribution so our experience does not inform us of the events that are possible. And the infinite variance implies that predictions lack any precision; a prediction that player Z will hit K home runs next season, plus or minus infinity is no prediction at all.

Some other diagnostics are in order. First, consider the change in home run hitting by taking each player's home runs for each year and then taking the first difference. Figure 6 shows the change in home runs for each year of the sample. There is no change; the series is centered on zero and has no drift whatsoever. In fact, the largest positive deviation in home runs is by Roger Maris in his record year of 1961. Mark McGwire's record of 70 in 1998 is the next largest positive leap. Leaps and falls are pretty evenly divided, there are 5 players above a positive change of 20 home runs and 6 players below a negative change of 20 home runs. Barry Bonds leap in his record year is exceeded by Maris and McGwire and matched by 7 other players in other years. Nothing so unusual here.

The changes are fairly extreme among players, but there is no trend at all. The kurtosis of the distribution of changes is large, 7.98, meaning the distribution is far from normal. It shows more extreme deviations than would be true if home run hitting were normally distributed. The mean change is zero, with a standard deviation of 10.03. The frequency distribution of changes in home runs hit by players is almost a classic picture of a stable distribution, with a narrow, high central peak and long tails; see Figure 7.

7. Home Run Records

What about home run records? As the perevious section demonstrated, home run hitting is a "wild" process. But, it has features that are regular too. Both of the distributions of the number of home runs by players and the change in home runs are far from normal. Hard things never are normal. Both processes follow a Paretian Stable distribution and have infinite variance. The changes are stationary over time but they have a large variation and the variation tends to cluster, like volatility in markets.

The most home runs hit in a season will have less variability than home runs. The maximum order statistic is known to follow a generalized Pareto distribution. A generalization of this distribution is the Stable Paretian distribution whose parameters are estimated in Table 2. It is a less skewed distribution than the distribution of home runs itself and has a smaller scale, but is located much farther to the right at 47.77 home runs. Its tail weight α is almost the same, 1.56 as the 1.64 for home runs.

The time series of MLB leading home runs shown in Figure 8 is rugged, with leaps and valleys, and there is no trend. There is a leap in Roger Maris' 1961 record year and then leaps in three

Table 2: Maximum Likelihood Parameter Estimates of the Distribution of League Leading Home Runs

	α Index	β Skewness	γ Scale	δ Location	Log-Likelihood
HOME RUN HITTING					
α-stable	1.5659	0.7049	4.1548	47.7713	-156.73

other years of 1998, 1999, and 2001. The low of 31 is two standard deviations below the mean of 48 and the high of 73 is just under three standard deviations above the mean. These are sample statistics, the standard deviation of the distribution does not exist because $\alpha < 2$.

8. Measuring Home Run Effectiveness

We now have ample warning that the distribution of home runs is a "wild" distribution that throws "normal" analysis out the window. But there are still some things we can try to discover in terms of hitting strategy and effectiveness of modern players over old timers. Maybe they are striking out more often and hitting more home runs because they are "swinging for the fences." Maybe they are more powerful in some measurable way.

Perhaps the purest measure of power hitting is home runs per hit. This takes base on balls, at bats, number of games played, hitting average and other factors out of the measure. With that said, I also develop the numbers for home runs per at bat since this may be both an attribute of power and an indicator of strategy. As a measure of home run efficiency, I also consider home runs per strike out. I consider home runs per strike out to be a strategic variable as well as a measure of skill and power. If a player is trying to hit home runs, he will strike out more often. If he is not skilled or powerful, the hard swings will not produce results and he will have few home runs per strike out. Some players change their strategy over time as they gain skill and mature. Strike outs, home runs per hit and per at bat, and home runs per strike out give us a sense of their strategy and effectiveness.

It turns out that all these measures have the same kind of skewed, heavy tailed distributions that we have already encountered. Seeing the same shape once again is a powerful confirmation of our finding that home run hitting is an stable Paretian distribution because this implies that any way you look at the process you should that the distribution has the same shape.

Table 5 summarizes the statistics of home runs, home runs per hit, home runs per at bat, at bats per player season, base on balls per year, strike outs and home runs per strike out. It is useful to look at these statistics before we look more closely at the players who produce them. None of these variables follow a normal distribution according to standard tests. In addition to these

distributional statistics, I have tabulated the annual figures by year of these variables for all hitters with 200 or more at bats from 1959 through 2004. The tabulations appear in Tables 6, 7, and 8. The data cover 45 years and 11992 player years.

Home runs per hit has a mean of .101, a large standard deviation (relative to the mean) of .07, positive skew and excess kurtosis. It reaches slight peaks in the 1960s, in 1987, and in the 2000s. The least value for a player is 0 and the most is an astonishing .52. Even a value as low as .29 is in the 99th percentile. The median is below the mean another indicator that the probability tail to the right is heavy.

Home runs per at bat (which does not count a walk as an at bat) have a mean of 0.027 and a large standard deviation of 0.019. The median is below the mean and the skew and kurtosis statistics verify that the distribution has a peak to the left and a long tail to the right, truncated at 1.

At bats per player is rather stable, partly because I have considered only players with 200 or more at bats. It is near its long term average of 416 at the beginning and end of the 45 years of the data. Base on balls is also rather stable.

Strike outs tell a different story. They begin to increase to more than 70 strike outs per game beginning in 1996 and remain above the long run average of 64 strike outs per game through 2004. Modern hitters strike out more than the old timers (even Ruth only averaged 68 strike outs), in spite of the smaller strike zone (see below). Their home run efficiency per strike out is a bit less than it was in the 1960s and right at or just above the long run average of 0.18 home runs per strike out. More strike outs and less home run efficiency per strike out suggests a strategic change in baseball in the mid nineties toward the long ball.

Home runs per player have drifted up a bit, but we now know that this average is dominated by the elite home run hitters. And, we know that the mean is volatile and should vary according to the law of home runs. Home runs per hit have hardly changed over the years; home runs per at bat are remarkably stable, and home runs per strike out are only slightly above the long run mean in the latter years, but still lower than in the 1960s. Nothing has changed.

9.   MLB Home Run Production

Now the question becomes what is the source of the slight change in total home runs hit in the major leagues over time? Has home run production increased at all levels of hitting in the big leagues? Or is it a few premiere hitters? Here are some things to think about.

I graphed home run hitting in percentiles over the time period of 1959 to 2004 in Figure 3. The figure shows home runs per hit in the 20th, the 50th, the 70th, the 90th percentiles of all hitters and the most home runs hit in a year by a single player. Only the maximum shows a peak during the McGwire, Sosa, and Bonds great years. Otherwise, there is no trend toward a higher number

of home runs per hit over this long time period. Even the 90th percentile hitters did not exceed slightly more than 0.20 home runs per hit, nowhere near the levels attained by a few, unique power hitters, and their production is unchanged over time. Even the 90th percentile hitter looks like an underachiever compared to these guys.

Many changes in the Major Leagues have taken place over the many years of the hitting data I have analyzed. The season has lengthened from 154 to 162 games per year. The number of teams has grown from 16 to 30. There have been 4 player strikes with shortened seasons. There have been changes in ballparks, modest changes in bats, and allegations of a change to the ball. The American League adopted the designated hitter in 1973. In addition, the players have better hitting mechanics and train for power. To quantify the effects of these changes, I have developed a simple empirical model of the total home runs hit in MLB each year.

The model of annual MLB home run production is:

$$HR = \beta_0 + \beta_1 teams + \beta_2 games + \beta_4 maxhr + \epsilon$$

The estimates of the model are in Table 9. I have used maximum likelihood estimates of the general linear model because the error term is not normally distributed. Nonetheless, these estimates differ little from the least squares estimates where the model accounts for about 83 percent of the variation in the data.

What do these results indicate? The coefficients are all estimated with small standard errors. The teams coefficient indicates that adding another team to MLB adds 115.56 home runs to average MLB home runs per year. An additional game (mlbg) increases average home run output by 1.08. The variable maxhr is the most home runs hit in the year by a player. This variable captures the effect of the extreme performances on the league total. A one unit increase in the number of home runs hit by the league-leading hitter raises total MLB home runs by 49.11. This is remarkable. It suggests some kind of home run competition among players of the sort that went on between Roger Maris and Mickey Mantle in 1961 or Sammy Sosa and Mark McGwire in 1998 and 1999. Or Sammy Sosa and Barry Bonds in 2001.

But it goes beyond that to suggest that Ken Griffey (56) and other players may have been caught up in the contest and hit more home runs too. Hitting home runs is more than a matter of power, it depends on strategy. On the third strike, for example, some players try to put the ball in play rather than go for a home run. Swinging harder and/or with a slight upper cut is another way to alter hitting strategy. A player may see others hitting home runs and begin to train and practice more effectively. The size of the coefficient, 49 more home runs in the league when the leading hitter hits one more, is some evidence that players follow the leading home run hitter or try to surpass him. It is said that when Honus Wagner saw Babe Ruth hitting home runs he altered his hitting and increased his home run output from less than 10 a year to 40 a year.

10. Self-similarity and Intermittence

The Paretian upper tail of the stable distribution law of home runs has two intriguing properties. The probability distribution is self-similar, meaning that parts of the distribution look just like the whole. If you look at the distribution through a narrow window or a broad one, you will see the same distribution. For example, estimating the stable distribution far into the upper tail beyond 50 home runs (the top 0.15 percent of home runs) also yields a stable distribution with a similar tail coefficient of $\alpha = 1.535$ and an appropriately rescaled location value of $\delta = 58.58$. In this region, you see the pure Paretian tail. The distribution has the same shape, the crucial feature of self-similarity.

The Richter law of earthquakes is a Pareto or power law distribution.[6] It says earthquakes of all sizes may occur and there is no typical earthquake. The law of home runs says the same thing; there is no typical home run year in baseball or typical home run hitter either. Earthquakes can come in bursts and there can be longer intervals of calm. The number of quakes per day, week, month, or year is random and they have the same distribution (similarity again). There is some longer term correlation between quakes. So, earthquakes can come in clusters and bursts of all sizes. They are not spread out like margarine on bread, all even and smooth.

That is the nature of the wild statistical distributions that describe earthquakes, storms and rainfall, disasters, stock markets (where there is evidence of volatility clustering), and home runs. They are all "bursty" processes and, according to the law, the size of fluctuations is greater among the most prolific home run hitters. This is apparent in Figure 3 where the maximum home runs per season is more volatile than the 90th and lower quantiles. This is not an anomaly. It is implied by the law of home runs; a power law tail implies that fluctuations increase with higher rank (Sornette 2000). So clusters of events occur, they have no typical size, and the fluctuations are larger among the best home run hitters. Thus records are made and fall in an unpredictable and intermittent manner. The burst of home run records in 1998 through 2001 is well within our expectation from the statistical law of home runs. There is no need for recourse to an external cause like steroids to explain it. This cluster of records looks very much like the cluster from 1921 to 1927.

11. A Closer Look at Palmeiro, Bonds, Sosa, and McGwire

Because of his suspension and status as a home run hitter I look more closely at Rafael Palmeiro's career (see the Appendix: Table 10. He began play in 1986 with 3 home runs in 73 at bats. In all his succeeding years he appeared at bat more than 200 times, commonly he appeared in from 400 to 600 at bats. The increase in his at bats reflects the expansion of the number of

---

[6]The upper tail of the distribution of home runs is a power law distribution that is known to be scale-free and subject to bursts or avalanches (Bak and Chen 1991). See William Brock (Brock 1993) and Schroeder (Schroeder 1990) on power laws, flicker distributions and intermittence.

games per season and his durability as a player.

His peak home runs per hit, per at bat, and per opportunity occurred in 1999 and 2002 when he hit 47 home runs in 565 and 546 at bats respectively. He averaged 531 at bats through his career. This high average has been an important part of his productivity as a home run hitter. His average of 29 home runs per year is well above the MLB average of 11.7 for players with at least 200 at bats. His strike outs are average, but they did peak during his prime home run years; by 1996 he was striking out at least 90 times when he had at 600 at bats or more. His home run efficiency was highest in 1995 (37 home runs) and in 1999 (47 home runs), exceeding .60 in each year. In his other big year in 2001 with 47 home runs, he managed to have 600 at bats with 90 strike outs, with a home run efficiency of 0.52.

There is no break or jump in Palmeiro's home runs per hit, the most reliable measure of power, anywhere in the series. His most powerful years were from 1995 through 2003, when he exceeded 0.21 home runs per hit. He hit more than 37 home runs ten times in the stretch from 1993 through 2003, with a drop to 23 in 1994, a year in which he had only 436 at bats. His peak years of 47 home runs are one standard deviation above his long run average. 2004 was one of his least productive years as a power hitter, just 23 home runs in 550 at bats and a home run efficiency of only 0.38 a level he hit in 1991 and that he had exceeded every year but one since 1993. If he used steroids during the year, there is scant evidence that it was effective.

In the Appendix in Tables 11, 12, 13 I also do the same tabulation for Bonds, Sosa, and McGwire. Bonds has the highest home run average of 37 and, by far, the largest number of base on balls of 121 per year. His average 478 at bats per season trails Palmeiro by about 50 but is just ahead of Sosa at 471 and far ahead of McGwire who only averaged 363 at bats per season. McGwire and Sosa are tied in home runs per year at 34, ahead of Palmeiro, but behind Bonds. Sosa and McGwire had to hit more home runs per at bat and per hit to come so close to Bonds' hitting because they had fewer at bats than he during these years.[7]

In his record year of 2001, Bonds had the most strike outs of his career since 1989. His home run efficiency was a high 0.78. He was swinging for home runs and did it well. But, in 1994 he had a higher home run efficiency equal to 0.86, but so few at bats that he hit only 37 home runs. By 1993 his home runs per at bat were beginning to reach Ruthian levels. He hit 0.09 home runs per at bat that year and kept it near or above that level since. By 1999 he broke into the magic 0.30 or more home runs per hit to stay (he had previously done it in 1994). From 2001 through 2004, he exceeded even Ruth in home runs per hit. His at bats and strike outs decline dramatically from 2001 onward. But, he didn't have as many big years as The Babe, just one huge year and then some good ones.

Sosa has always been a free swinger, striking out an average 124.12 times a year and more than 170 times three times. That is one reason why he hit more home runs than anybody in MLB from

---

[7]Now Bonds' at bats are among the lowest because his is walked so often.

1996 through 2004 when he hit an average of 49.22 home runs a year. His home run efficiency is lower than Palmeiro's or Bonds', but for him the strike out home run trade-off was very favorable. They pay you for home runs not for avoiding strike outs and he hit more than anybody during his best years.

McGwire had a stunning streak from 1995 through 2001 in which he hit 0.4 or more home runs for every hit. But, in only two years did he hit more home runs than he hit in 1987 (52 is essentially a tie with 49 from a statistical point of view). His record year in 1998 of 70 home runs is two standard deviations above his average. 1999 is about one and three quarters above his average. Yet, in neither of those years does he match his home runs per hit of 2001. His record years of 1998 and 1999 coincide with over 500 at bats, a figure he previously attained only three times early in his career. In contrast, Bonds, Sosa, and Palmeiro appeared at bat more than 500 times many times in their careers, though Bonds' at bats in his last two years were hampered by many walks and fell below 400. McGwire was clearly swinging for home runs in his big years of 1998 and 1999; his strike outs jumped far above his average to 155 and 141 respectively.

Finally, let us compare these players to home runs hit in the past. In Figure 9 we see the most home runs hit each season in MLB and the number of home runs hit by Sosa, McGwire, Bonds. The league-leading home run series is centered at the mean of 48, with peaks and valleys through the years. Maris' record is at the 90th percentile, McGwire's is at the 95th and Bonds' record is at the 99th percentile. The distribution is not normal and it has a narrow central peak with a long, heavy tail to the right. These are the classic signs of a Paretian or more general stable distribution.

In their best home run years, all these players were clearly going for it and paying the small price of more strike outs. They had high home run efficiency. McGwire's home run efficiency peaked in 1995 at 0.51, but it was already at 0.47 in 1993. Bonds peaked at 1.10. Palmeiro peaked at 0.68 and Sosa at 0.42. These guys are far above other players who have an average efficiency 0.176.

12. The Baseball, Ball Parks, The Strike Zone, Bats, Players, and Steroids

We have to go over a few of the many explanations that have been offered for an increase in MLB home runs. As we should expect, they do not stand up to scrutiny because there have been no fundamental changes.

The Baseball The coefficient of restitution of the baseball was increased in 1920, making it rebound off the bat with higher velocity. The number of games increased from 140 to 154. That year, Babe Ruth hit 54 home runs, breaking his record of 29 from the year before. So, he benefited from a hotter ball and more games in breaking his record.

In the modern era, claims persist that the ball has been made hotter off the bat. If it were

lighter or had more rebound, the effect would be to increase home runs. Yet, there is no evidence to support this claim. Each time it is tested it is rejected.

In response to a controversy about the "hot" 1977 baseball, it was tested against the 1976 ball. The 1977 ball rebounded 1.3 percent higher in a small sample of balls tested. This is well within the random variation among baseballs; the MLB standard permits a variation of 3.2 percent. Another controversy concerned the 1999 and 2000 baseballs. Tests again showed there was no difference. Dr. Jim Sherwood's measurements confirm that the 1999 and 2000 baseballs have the same coefficient of restitution (COR). A test of balls a decade apart, of 1988 and 1998 baseballs, by R. C. Larsen showed no difference (Adair 2002).

Current MLB standards set a rebound velocity of 54.6 percent of the speed of the ball at impact, plus or minus 3.2 percent. Weight and COR variations among balls could cause two conforming balls, one light and hard, and the other heavy and soft, to differ by as much as 49 feet distance with the same impact. It also is true that balls are no longer used as many times in modern MLB. Impacts and scuffs can alter the ball flight and lower the COR.

The baseball is smoother say umpires such as Mike Reilly (Kaat 2000). Testing by Jim Kaat and Popular Mechanics showed that 2000 MLB baseballs were slicker than 1999 baseballs; they had a slightly lower coefficient of friction. There were no differences, however, in batted-ball speeds. A smoother ball will travel a shorter distance because it does not break the airflow into a turbulent flow where resistance drops (Adair 2002). On the other hand, the test showed that the seams were a bit higher on the 2000 baseball, but this may have been a matter of the sampling and measurement. Higher seams would break the airflow and make the ball travel farther as its airflow becomes turbulent. A smoother ball may also move differently when thrown by the pitcher; the late-breaking cut fast balls and back-up sliders that are thrown today may not have been possible years ago. A slicker, harder-thrown ball will break later and a ball with higher seams will break more. On the whole a smoother ball with slightly higher seams would probably be of more benefit to the pitcher than the hitter.

In short, aside from the change to the baseball in 1920 there are no confirmed tests that show any other changes of significance in the baseball. But, with each apparent surge in home runs, and I stress apparent, people seem to look for an explanation. Tests from 1976 to 2004 indicate that the baseball has not changed in a way that would permit more home runs to be hit.

Ball Parks The House that Ruth built, Yankee Stadium, opened in 1923 and was suited to his home run stroke. Only 295 feet down the right field line and 350 to short right center to a pull hitter like the Babe must have been a very friendly hitting environment. Maris, who broke the Babe's record in 1961, was also a left-handed hitter and he set his record playing home games in Yankee Stadium. Hank Greenberg came close to the record by hitting 58 in 1938. He was a right-handed hitter playing in a not-so-home-run-friendly Briggs Stadium which was 340 feet down the left field line and 365 to left center.

The only rigorous research on ball parks shows that the average of the required ball velocities to hit a home run the park dimensions (left field, left center, center, etc.) of MLB parks varies little, from a ball velocity off the bat of 107 to 110 feet per second. There is no comparable historical analysis that would permit a conclusion as to the effects of the various dimensional changes that have taken place in MLB ball parks over the years.

One change that in ballparks that might boost home runs is the addition of Denver to major league baseball. The high altitude and low air density allow a struck ball to carry farther than a sea level. Atlanta is located at a higher altitude too. But, the addition of other parks at low altitude, such as Miami and Houston, seem to have offset Denver and Atlanta, leaving a only a small net change in average park altitude.

The Strike zone   The strike zone has gotten smaller. Whereas the height of the strike zone was the top of the letters or the arm pits, it has moved down to half the distance between the shoulders and the belt.

Both Barry Bonds and Babe Ruth are 6-ft.-2in. in height. With a 17-in.-wide plate, Babe Ruth had a strike zone of roughly 545 square inches. In 2001, Barry Bonds, had a strike zone of about 410 square inches. The high fast ball at the letters that used to be a strike out pitch is not available to pitchers in recent years. In spite of the shrinking strike zone, strike outs per player has drifted upward over the years (see Table 5) from 56 per player in 1961 to 77 in 2001. Swinging for the fences does seem to produce more home runs and strike outs. Remarkably, home runs per strike out are down from 0.23 home runs per strike out in 1961 to 0.19 in 2001. Modern hitters are a bit less efficient than players of the 1960s in hitting home runs. This suggests they really are swinging for the fences because the smaller strike zone should have increased home runs hit per strike out.

The smaller strike zone may be an explanation for the very slight upward drift in the number of at bats players get per game; it has increased by 0.6 at bats per game.

Bats   MLB rules governing baseball bats have not changed. Bats must be made of wood and the barrel of the bat must not exceed 2.75 inches in diameter. Earlier bats tended to be made from hickory. Ash became more common later and is the main wood used in bats today. Maple is used by a few players. Maple is the denser of these woods, followed by hickory and then ash.

Babe Ruth used a hickory bat weighing 47 ounces to hit 60 home runs. A bat of that size had a full 2.75 inch barrel. Roger Maris used an ash bat weighing 33 ounces to hit 61 home runs. Barry Bonds used a maple bat weighing 33 ounces to hit 73 home runs. To make a bat as light as 33 ounces, it has to have a smaller barrel diameter. Consequently, the ash bat Maris used and the maple bat Barry Bonds used had a barrel diameter of just 2.50 inches, a loss of hitting area of about 4 percent (Adair 2002). Henry Aaron hit more home runs than anyone with a bat weighing from 31 to 32 ounces.

A heavier bat will hit the ball farther if it is swung at the same speed as a lighter one. Since the kinetic energy is mass times velocity, there is a trade off between bat speed, weight and power. Modern players, facing a large variety of pitch speeds from 70 mph change ups to 98 mph fastballs, use lighter bats than players of the past. A light bat will actually drive a slower pitch farther than a heavy bat. A light bat gives the hitter a few milliseconds longer to judge a pitch, but its main contribution may be in the higher precision of the swing path. Home runs are hit by striking the ball about half an inch below center, generating lift and back spin. Mammoth home runs must be launched at around 35 degrees. So a slight upswing and a strike just below center are optimal for hitting home runs (see (Williams 1986) and (Adair 2002)).

Players   Babe Ruth was 6' 2" inches tall and weighed 251 pounds when he hit 60 home runs in 1927. He was quite a bit more massive than the typical MLB player of the time. Mark McGwire was 6' 5" and weighed 225 when he hit 70 home runs in 1998. Barry Bonds was 6'2", like Ruth, and weighed 220 pounds. Sammy Sosa is 6'0" and also weighs 220 pounds. Home run hitters are more massive than the typical player.

With the same bat speed, a heavier player will generate more energy to drive the ball. In his *Physics of Baseball* Dr. Adair (Adair 2002) shows that total energy is approximately proportional to player weight. A 140-pound player swinging a 32 ounce bat would drive a 90-mph fastball 338 feet, a long out or off the fence if pulled down the line in most parks. A 225 pound player swinging the same bat at the same speed would drive the ball 382 feet, a home run to all but deep center in most parks. A more massive player can also swing a heavier bat, which adds to distance. But, the strong qualifier is that the bat must be swung at the same speed and a heavier player may not have the same quickness to accelerate the bat as a lighter player.

If there were more home runs in MLB, an explanation could be found in the mass of the modern MLB player. Modern players are larger than players of even as few as 30 years ago, just as the current generation of Americans is taller and heavier than previous generations. On the other hand, it is more difficult for a larger player to put the fine kinetics of a beautiful and powerful swing together or to react and accelerate the bat quickly. If size were all that mattered, baseball players would look like NFL football players. Longer limbs make for longer levers that are more difficult to move accurately and quickly.[8]

Steroids   At this point, there is nothing left to explain and no need to invoke any external source such as steroids for MLB home run hitting. There have been no changes in home run hitting. That should be the end of it. But, it isn't. Records have fallen and for the critics of baseball there must be an explanation, a cause. And they seem to have the explanation they seek in steroids. Let's

---

[8]A 6'4" in height, Ted Williams (Williams 1986) kept his hands very close to his body in his hitting stance, stressing the quickness that gave his swing.

look at the argument and some facts.

First, there is no causal explanation required for home run records. Under a random process, the realizations of the events just happen. Explaining why you drew a Spade instead of a Diamond when the draw is random is silly. No explanation is required; it is a realization of a random process.

Home run hitting is one of the most random and uncertain of human activities. The law of home runs has an infinite variance, which is about as uncertain as anything can be. That means there is no basis for anyone to say how long records ought to last or how often they should be broken. We have only probabilities. And just as you can get three or five heads in a row in a series of coin tosses, you can get clusters of home run records. It is nothing more than a random series of events. No more explanation is required for home run records than for three heads in a row. Under the law of home runs, new records have non-vanishing probability, far more than one would expect if they were thinking in terms of averages and normal distributions. The probability of new records is far higher than a normal distribution would make possible.

Steroid advocates have to argue that the new records are not consistent with the law of home runs, that the law itself has changed as a result of steroid use. They have to show that the Sosa, Bonds, and McGwire performances are impossible under the statistics of home run hitting. But, this is not true. The cluster of recent records is perfectly consistent with, and even implied by, the statistics of home run hitting. It also is consistent with bursts of home run records in the past.

There are two other hurdles the steroid hypothesis must clear. One must show positive proof of steroid use by a record setting home run hitter and, further, that steroids are the cause. There is slim evidence on the first score. We should note that only 9 players have been suspended for testing positive for steroids in thousands of tests; approximately 3990 players were in the big leagues over the past three years. That is only a positive rate of 0.002. The test surely has a false positive rate higher than that. And there are contaminants in most supplements that nearly all players take that would cause them to test positive even if they are not taking steroids.

As to steroids themselves, they are highly overrated as aids to home run hitting. McGwire, Ruth, Maris, Sosa and Bonds rarely hit balls to the warning track. When they connected the balls were gone, far over the wall. This was true throughout their careers. Bonds and Maris were line drive hitters earlier in their careers, but once they developed as home run hitters (a home run hitter swings slightly upward to launch the ball at a 35 degree angle and hits the ball just below center as opposed to a line drive hitter who has a level swing and hits the ball in the center) their home runs landed in the seats, not on the warning track. Were steroids effective at increasing bat speed (and they have not been shown to), they could increase home runs by lengthening warning track hits by 10 to 20 feet. Dr. Adair estimates that about 1 hit lands on the track for every 10 home runs. For Ruth, Maris, et al far fewer balls land on the track for every home run they hit. Suppose it is 1 out of 25. This means that, at most, higher bat speed would increase home runs by 1 to 3 a year in these great hitters. So, the test of steroids would be to calculate the number of warning track hits relative to home runs for a known steroid user and show that what were previously warning

track shots become home runs. No one has produced this evidence.

But, even that may not be good evidence, because there are no strong reasons why steroids should increase bat speed. Steroids encourage hypertrophy in body builders and strength athletes, who use massive amounts, train long and hard, and do very high volume work. No baseball player can afford to train that long and the training would be highly counterproductive to his baseball playing. Body builders and strength athletes (power lifters) are not particularly quick. While there is evidence that steroids increase muscle mass of body builders and strength athletes, this muscle hypertrophy will not increase the important fast twitch muscle fibers that home run hitters rely on. Body builder and strength athlete exercises produce a higher volume of slow twitch muscle fiber, the antithesis of rapid power production.[9]

Hitting an 85-mph pitch 400 feet requires a bat velocity of 70 mph. The bat must be accelerated to this velocity in 200 milliseconds. Such a feat requires rapid force generation that can only be supplied by fast twitch muscle fibers. Body building, and the slow twitch fiber composition that it produces, could not produce the power and speed that Mark McGwire, Barry Bonds, and Sammy Sosa exhibited in their prime years. Speed strength is developed using rapid movements that do not result in hypertrophy. And there is no evidence that steroid use directly promotes speed strength. Even these modern players are not as muscular as Ted Kluszewski or Steve Bilko were in their prime. In fact, they look like slightly taller versions of Mickey Mantle, a densely muscled player of the past.

Again, The Babe has the final word on steroids. The last home run he hit, number 714, in his last game in 1935 with the Boston Braves in a season of just 72 at bats, was one of the longest of his career. Using a 36 ounce bat, he hit the first home run ever hit over the right field roof of Forbes Field in Pittsburgh.[10] Films show that he was a shadow of his former self by this time and he likely already had the cancer that would kill him a 13 years later.

Other long home runs from earlier years show that the longest home runs have not changed. Hank Greenberg in 1938 and Vic Wertz in 1948 hit balls that were caught in center field at more than 450 feet. Joe Adcock, Lou Brock, Hank Aaron, and Luke Easter all hit home runs over the nine-foot screen in front of the center-field bleachers 480 feet from home plate in the Polo Grounds. Mickey Mantle hit a ball out of Griffith Park in 1953 that is estimated reliably to have gone 506 feet, aided by a 20-mph wind (Adair 2002). Dick Stuart hit a home run over the 457 foot left-center field corner of Forbes Field in 1957.[11] These home runs are comparable to any hit today.

----

[9]The literature on highly hypertrophied body builders demonstrates slower contraction/relaxation speeds and a muscle fiber shift away from the fastest type IIx/IIb muscle fibers toward slower fast twitch type IIa and even the slow twitch type I muscle fibers. See the review by Angus Ross and Micheal Leveritt (Ross and Leveritt 2001).

[10]Right field was 300 feet down the line and 375 to right-center. The roof was 86 feet high. Only 10 players have duplicated the feat since, Mickey Mantle, Eddie Mathews, Willie McCovey, and Willie Stargell among them. Thanks to Michael Stima for this information.

[11]The ball was estimated to land 500 feet from home plate. Michael Stima, personal communication.

Home Runs per Game   The measure that many proponents of the steroid explanation seem to pin their theory to is home runs per game. This reasoning is flawed. First, home runs per game tell us nothing about records. Records are about extreme performance, not some kind of average. Records are made by extraordinary performances that have no relation to average home runs per game.

Second, home runs per game is a strange statistic. Home runs per game is simply total MLB home runs divided by total MLB games. Why should this almost meaningless calculation be a well-behaved measure of performance? It turns out to have a strange statistical distribution that has four peaks (two to the left of the mean and two to the right) and a deep valley just above the mean. The density in shown in Figure 10 where I have overlaid the normal distribution for comparison. It is far from normal, it is multi-modal, left skewed (skewness of 0.19) and has more kurtosis (1.91) than a normal distribution. It is a monster that cannot be used to set a standard.

If you think for a moment about the constraints of a ball game, it becomes obvious that home runs per game cannot be a well-behaved statistic that can be used to make sharp comparisons. The number of home runs in a game is an integer, not a continuous variable. The number of league games is an integer too. Dividing these numbers will give rational numbers, but they will not be distributed normally and will have strong modes at a few typical values. The progression of a game sets constraints on home runs. If one home run is hit, the odds that two will be hit go up. And if two are hit, the odds that three will be hit go up. And so on. But, as more home runs are hit in a game, the odds that greater numbers will be hit decline because innings and at bats run out.

No valid comparisons can be made using MLB home runs per game.

13.   Conclusions

Why do all the attempts at explaining an increase in MLB home runs seem to fail? I think it is because they are attempting to explain something that has not happened. There are no more home runs in baseball than before when the problem is properly analyzed. Once the number of games and other variables are factored into home run productivity and the random nature of home run hitting is taken into account, there is no change. That may be why all the attempts at explanation, such as steroids, hotter balls, altered strike zones, and home-run-friendly ball parks fall apart when they are tested.

Many sportswriters and fans make two mistakes. One, they think home runs have increased when they have not. I show here that the probability distribution has not changed. And, two, they think a random event, such as a new record, calls for some kind of external explanation. The new records are completely within the variation of outcomes under the stable probability law of home runs. You do not need an explanation for a random event when you have evidence that the law of home runs has not changed in 40 years. If you roll box cars three times in a row do you require an explanation? Only people who fail to understand chance do.

The same law of home runs holds now that held 40 years ago. Year to year differences in home runs require no explanation; they are all within the variation of the outcomes under the stable probability distribution of home runs. The burst of new records does not require an external explanation; they are part of the pattern that comes from the nature of the law of home runs. People often look for an "explanation" of purely random events. Looking for a "hotter ball" or steroids as a cause of a random event is just another example of the human failing to understand chance and randomness.

The pace of new records in recent years is due to the extraordinary accomplishments of three prodigious hitters. We have lucky enough to see three Babe Ruth's in this generation. Hitters such as these may never appear again. You cannot take an ordinary player and turn them into home run hitters of the accomplishment of Bonds, McGwire, and Sosa by dosing them with steroids. It may even be harmful. Home run hitting of that magnitude is as incomparably rare as the greatest works of science or art.[12]

It took more than 80 years and more than 100,000 player-years of baseball for 5 home run hitters of the caliber of Ruth, Maris, McGwire, Sosa, and Bonds to appear. We have been lucky enough to witness three of them in the past few years. To diminish their accomplishments on the basis of speculations and rumors about steroids, as members of Congress and the media have done, has the facts and the science wrong and is profoundly ignorant of the statistical laws of human accomplishment. McGwire, Sosa, and Bonds are home run geniuses, not ordinary hitters on steroids; they are the Bach, Mozart, and Beethoven of home runs.

Will another player of the prowess of Babe Ruth, Roger Maris, Mark McGwire, Sammy Sosa or Barry Bonds ever appear again? Even greater performances are possible because the law of home runs has a long Paretian upper tail with positive and non-vanishing probability mass. But, the chances are small. The law of home runs says that the probability that Babe Ruth's record of 60 home runs would be broken is 0.0109. Given more than 100,000 tries in nearly 80 years, it was almost sure to fall. Barry Bonds' record of 73 will be harder to break. According to the law of home runs, the probability that his record will be broken is 0.007206.

---

[12]Even Shakespeare had his doubts; there was a long controversy over his extraordinary works that led some scholars to contend that no one person could have accomplished what he did. The claim circulated that Francis Bacon must have written some Shakespearan plays. Maybe Shakespeare was on steroids.

REFERENCES

Adair, R. (2002). *The Physics of Baseball*. HarperCollins, New York.

Bak, P. and Chen, K. (1991). Self-organized criticality. *Scientific American*, 264(1):26–33.

Brock, W. A. (1993). Scaling laws for economists. Technical report, University of Wisconsin, Madison.

De Vany, A. (2003). *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*. Routledge, London.

Kaat, J. (2000). Baseball's new baseball. *Popular Mechanics*.

Kiele, K. (2005). Drug testing on Senator Mccains mind. *USA Today*, 2:1.

Levy, P. (1954). *Theorie de l'addition des variables aleatoires*. Gauthier-Villars, second edition.

Murray, C. (2003). *Human Accomplishment*. HarperCollins, New York.

Nicholls, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, pages 16:317–323.

Pareto, V. (1897). *Cours d'Economique Politique*. MacMillan, Paris.

Price, D. (1963). *Little Science, Big Science*. Columbia University Press, New York.

Rimmer, R. and Nolan, J. (2005). Stable distributions in Mathematica. *Mathematica Journal*, 9(4).

Ross, A. and Leveritt, M. (2001). Long-term metabolic and skeletal muscle adaptations to short-sprint training. *Sports Medicine*, 31:1062–1082.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York.

Schroeder, M. (1990). *Fractals, Chaos, Power Laws*. W. H. Freeman and Co., New York.

Sornette, D. (2000). *Critical Phenomena in Natural Sciences*. Springer, Berlin.

Williams, T. (1986). *The Science of Hitting*. Simon and Shuster, New York.

Table 3: Historical Features of the Distribution of MLB Home Runs

year	mean	sd	skewness	p75	kurtosis	max
1959	3.560127	7.186415	2.935209	3	12.66449	46
1960	3.340659	6.757604	2.969537	3	12.93642	41
1961	3.911175	8.044041	3.150735	4	14.92974	61
1962	3.948684	7.611533	2.675909	4	10.99461	49
1963	3.595745	6.977719	2.780676	4	11.8588	45
1964	3.66313	7.34417	2.688984	3	10.74072	49
1965	3.579228	7.034111	2.664924	3	10.84267	52
1966	3.543928	7.2754	2.837837	4	11.72799	49
1967	2.924936	6.233929	3.11355	2	14.26803	44
1968	2.79021	5.85355	3.249436	3	15.28404	44
1969	3.346567	7.185666	3.201337	3	14.69004	49
1970	3.73123	7.579399	2.741924	3	10.8387	45
1971	3.242356	6.537536	2.984032	3	13.34342	48
1972	2.853604	5.830272	3.064857	3	13.88811	40
1973	3.477578	6.851032	2.631955	3	10.55625	44
1974	2.898249	5.753187	2.560656	3	9.743141	36
1975	2.974642	5.95591	2.798702	3	11.81099	38
1976	2.522573	5.256048	3.070518	2	13.962	38
1977	3.703252	7.2927	2.651926	3.5	10.53618	52
1978	3.079167	6.329498	2.862956	3	11.99323	46
1979	3.57232	7.203676	2.717527	3	10.98704	48
1980	3.249474	6.357614	2.835592	3	12.69256	48
1981	1.886653	3.902153	3.1204	2	14.5556	31
1982	3.40625	6.868028	2.711784	3	10.41664	39
1983	3.281312	6.436787	2.644809	3	10.48907	40
1984	3.310976	6.546558	2.667375	3	10.43311	43
1985	3.609218	7.025644	2.475744	4	8.941914	40
1986	3.749263	6.994448	2.34848	4	8.281518	40
1987	4.253817	8.233141	2.421392	4	8.777658	49
1988	3.072464	6.183647	2.658388	3	10.42949	42
1989	2.873253	5.915982	3.046326	3	13.88731	47

Source: /Users/asdevany/BaseballSoftball/HomeRunPaper/Batting195 9-2004 copy.dta

Table 4: Historical Features of the Distribution of MLB Home Runs

year	mean	sd	skewness	p75	kurtosis	max
1990	2.974888	6.295484	3.081658	3	14.17139	51
1991	3.115101	6.626898	2.8086	3	11.28122	44
1992	2.849906	5.979934	3.039026	3	13.60428	43
1993	3.415254	6.975209	2.881807	3	12.32715	46
1994	3.209709	6.232981	2.747431	4	11.88949	43
1995	3.256983	6.870624	2.956998	3	12.60222	50
1996	3.960096	8.393098	2.905057	3	11.82035	52
1997	3.754045	7.722999	2.856834	4	11.91095	56
1998	3.83056	8.217706	3.265603	3	16.2117	70
1999	4.255581	8.748636	2.804406	4	11.71542	65
2000	4.119392	8.373664	2.695368	4	10.51838	50
2001	4.076176	8.537626	3.146246	4	15.08906	73
2002	3.835481	7.797285	2.820342	4	11.95775	57
2003	3.865627	7.798382	2.717163	4	10.90832	47
2004	4.049777	8.059798	2.573488	4	9.748294	48
Total	3.458776	7.101802	2.933478	3	12.93026	73

Source: /Users/asdevany/BaseballSoftball/HomeRunPaper/Batting195 9-2004 copy.dta

Table 5: Summary Statistics 1959-2004 for All MLB Players with 200 or more at bats

Variable	Mean	Std. Dev.	N
HR	11.686	9.805	11992
hrhit	0.101	0.07	11992
hrab	0.027	0.019	11992
AB	416.552	129.337	11992
SO	64.031	29.38	11992
hrso	0.176	0.123	11992

Table 6: Hitting Statistics by Year

year	mlbhr	mlbhrg	hrg	abg	teams	mlbg	tmax
1959	2250	1.817448	.0372908	1.822867	16	1238	31.41297
1960	2128	1.714746	.0356123	1.809906	18	1241	29.22606
1961	2730	1.909091	.0410229	1.899026	20	1430	31.8553
1962	3001	1.851326	.038726	1.83662	20	1621	31.13816
1963	2704	1.670167	.0355125	1.878385	20	1619	26.59574
1964	2762	1.698647	.0342992	1.859165	20	1626	29.43236
1965	2688	1.656192	.0350058	1.799557	20	1623	26.25965
1966	2743	1.698452	.0330716	1.820053	20	1615	27.53618
1967	2299	1.419136	.0281007	1.824632	20	1620	27.16412
1968	1995	1.227692	.0277447	1.902467	20	1625	24.90629
1969	3119	1.602775	.0322289	1.821728	24	1946	30.53755
1970	3429	1.763889	.0357301	1.804204	24	1944	31.94995
1971	2863	1.477296	.0338321	1.94158	24	1938	27.80068
1972	2534	1.363098	.0299677	1.895238	24	1859	25.36149
1973	3102	1.5965	.0339817	1.747667	24	1943	26.57848
1974	2649	1.361954	.0279765	1.708371	24	1945	23.67615
1975	2698	1.395036	.0286942	1.729718	24	1934	24.68357
1976	2235	1.152656	.0253954	1.813754	24	1939	21.62077
1977	3644	1.732763	.0356719	1.695905	26	2103	29.22764
1978	2956	1.40628	.0290252	1.703205	26	2102	26.23958
1979	3433	1.635541	.0326174	1.683551	26	2099	28.44537

Source: /Users/asdevany/BaseballSoftball/HomeRunPaper/Batting1959-2 004 copy.dta

Table 7: Hitting Statistics by Year

year	mlbhr	mlbhrg	hrg	abg	teams	mlbg	tmax
1980	3087	1.466508	.0315221	1.747295	26	2105	25.29579
1981	1781	1.277618	.0282501	1.669916	26	1394	16.82097
1982	3379	1.603702	.0323181	1.699242	26	2107	28.33165
1983	3301	1.565197	.0316736	1.690239	26	2109	26.44632
1984	3258	1.547743	.0322248	1.687198	26	2105	26.10569
1985	3602	1.712791	.0342113	1.626393	26	2103	27.71343
1986	3813	1.813124	.0380635	1.671918	26	2103	27.58112
1987	4458	2.117815	.040573	1.643536	26	2105	33.62977
1988	3180	1.514286	.0313752	1.679026	26	2100	25.90242
1989	3083	1.463913	.0297005	1.644554	26	2106	27.91519
1990	3317	1.575772	.0310493	1.56201	26	2105	27.33094
1991	3383	1.60789	.0292512	1.569098	26	2104	28.94659
1992	3038	1.442545	.028964	1.641287	26	2106	25.36961
1993	4030	1.776113	.0358744	1.610495	28	2269	30.2678
1994	3306	2.067542	.0416931	1.684764	28	1599	26.39806

Source: /Users/asdevany/BaseballSoftball/HomeRunPaper/Batting1959-2 004 copy.dta

Table 8: Hitting Statistics by Year

year	mlbhr	mlbhrg	hrg	abg	teams	mlbg	tmax
1995	4081	2.024306	.0372218	1.529073	28	2016	29.7103
1996	4962	2.185903	.0414978	1.583147	28	2270	37.04629
1997	4640	2.048565	.0403417	1.691973	28	2265	33.18528
1998	5064	2.082237	.0396815	1.728729	28	2432	37.30408
1999	5528	2.274897	.0430927	1.686842	30	2430	37.93226
2000	5693	2.343763	.043494	1.678945	30	2429	36.92113
2001	5458	2.247015	.0413485	1.694899	30	2429	38.62435
2002	5059	2.085326	.0383969	1.674799	30	2426	34.57013
2003	5207	2.142798	.0397508	1.673171	30	2430	33.17892
2004	5451	2.24321	.0416819	1.685324	30	2430	33.41605
Average	3646.962	1.764264	.0351253	1.713224	25.6713	2038.301	29.58311

Source: /Users/asdevany/BaseballSoftball/HomeRunPaper/Batting1959-2 004 copy.dta

Table 9: Estimation results : Model of Annual MLB Home Runs

Variable	Coefficient	(Std. Err.)
teams	115.563	(40.257)
mlbg	1.081	(0.457)
maxhr	49.113	(8.881)
Intercept	-3896.503	(554.264)

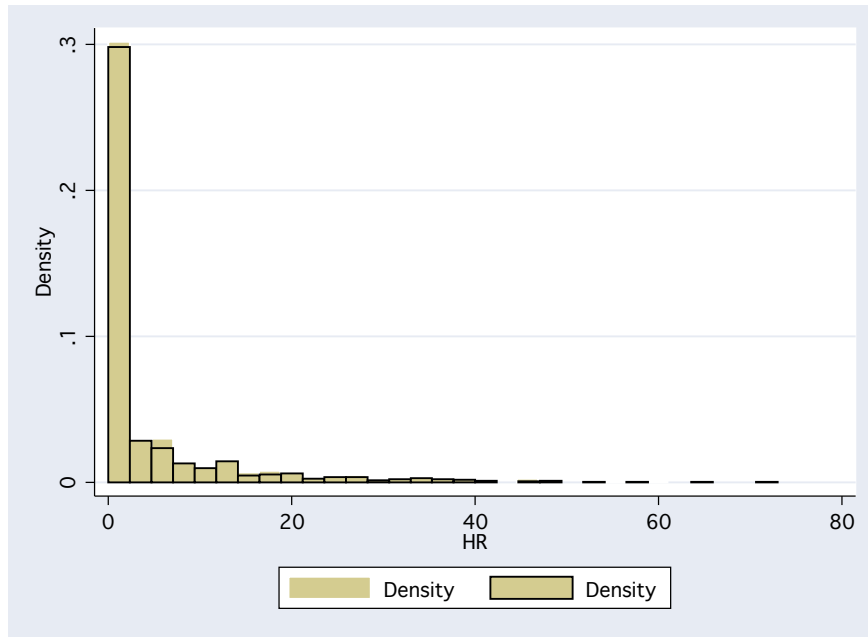Fig. 1.— The Distribution of Home Runs 1959–2004 for Players with 200 or More At Bats
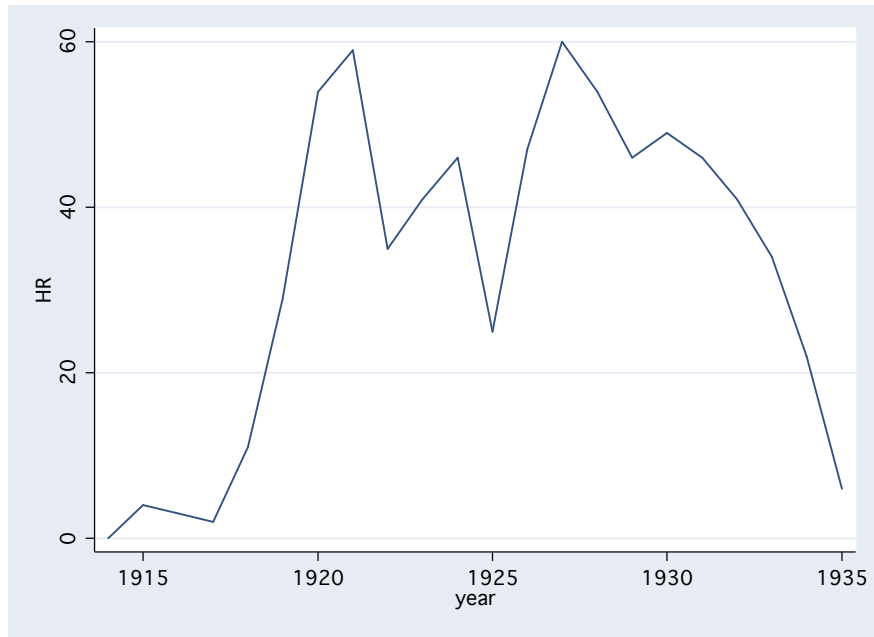
Fig. 2.— Babe Ruth's Home Runs

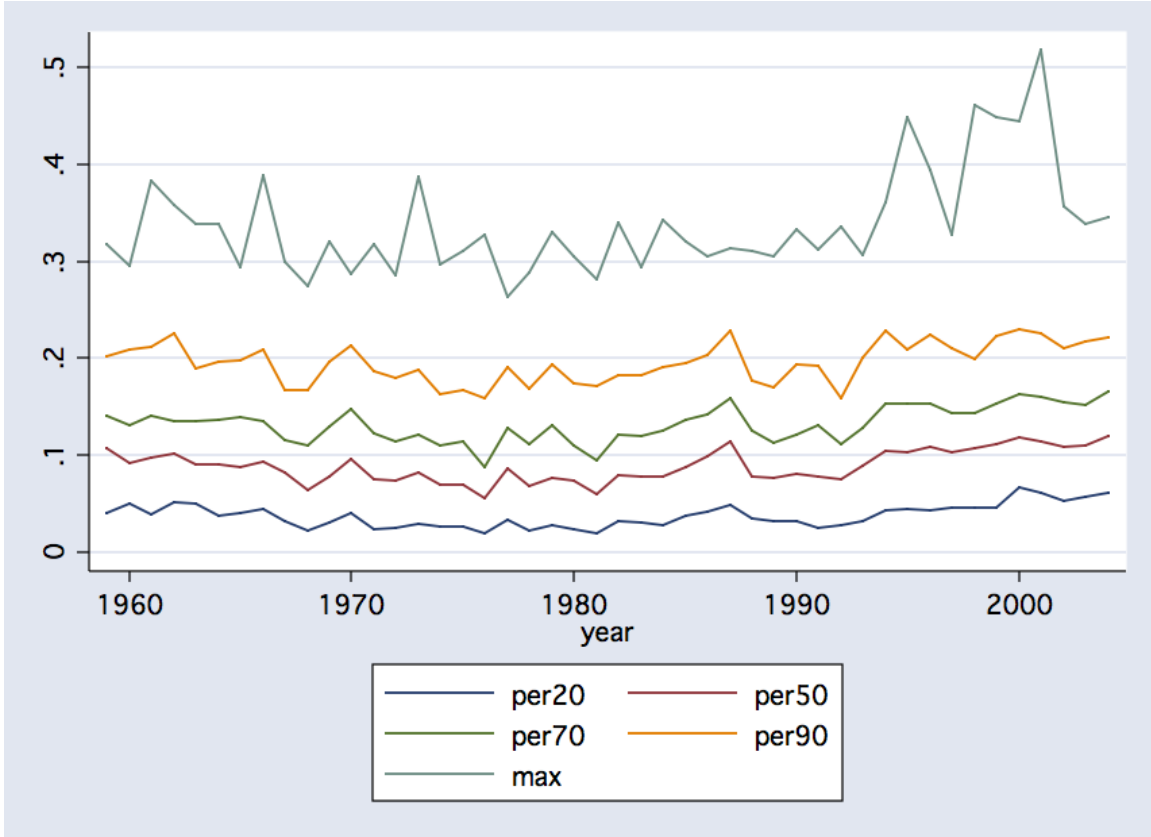Fig. 3.— Home Runs Per Hit over Time, Percentiles and Maximum 1959-2004

Fig. 4.— The Law of Home Runs: Stable Probability Density Function of Home Runs
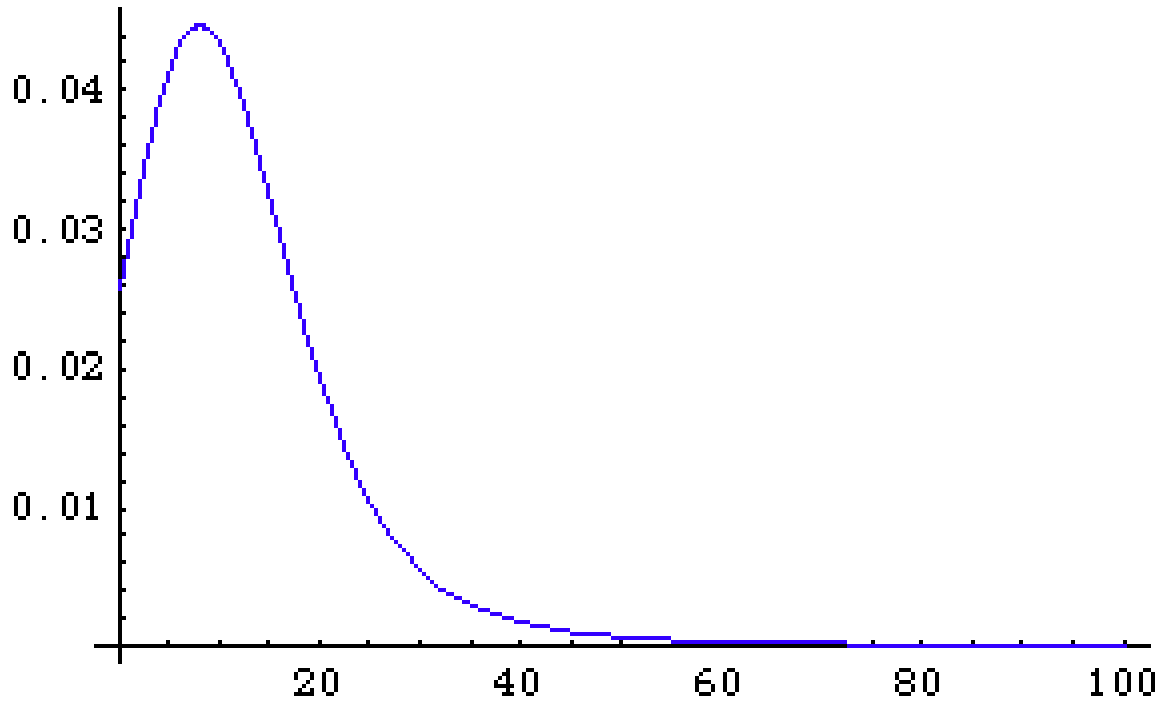
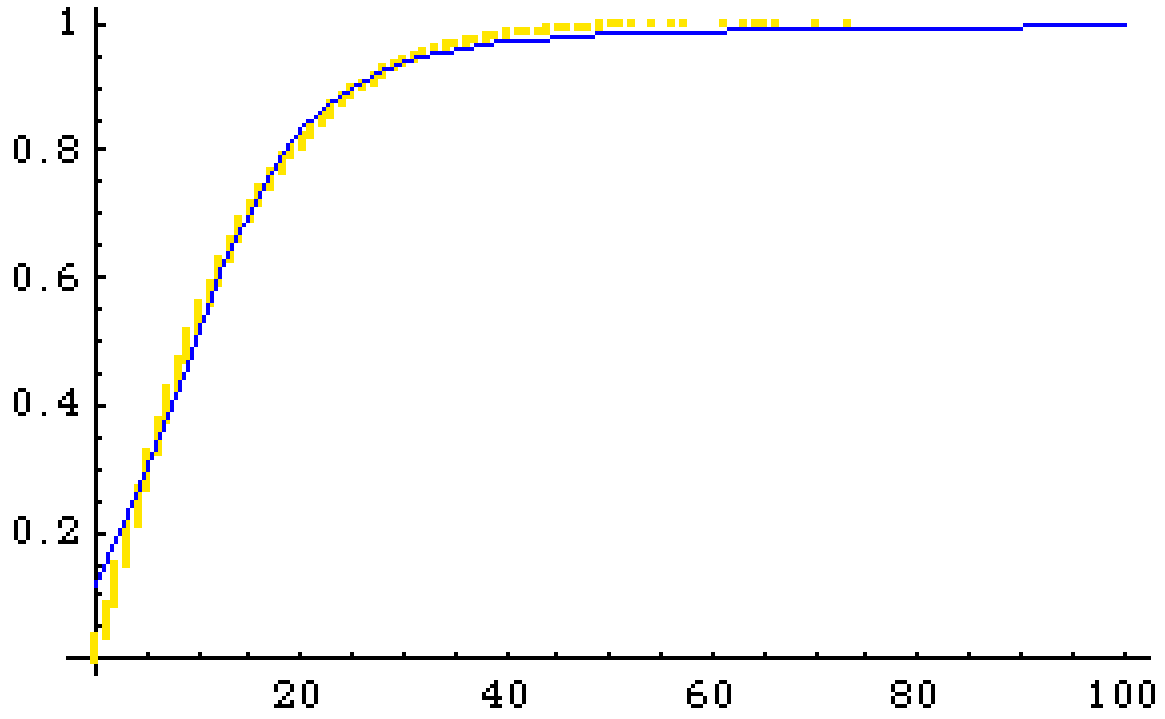Fig. 5.— Cumulative Fitted and Empirical Density of the Law of Home Runs

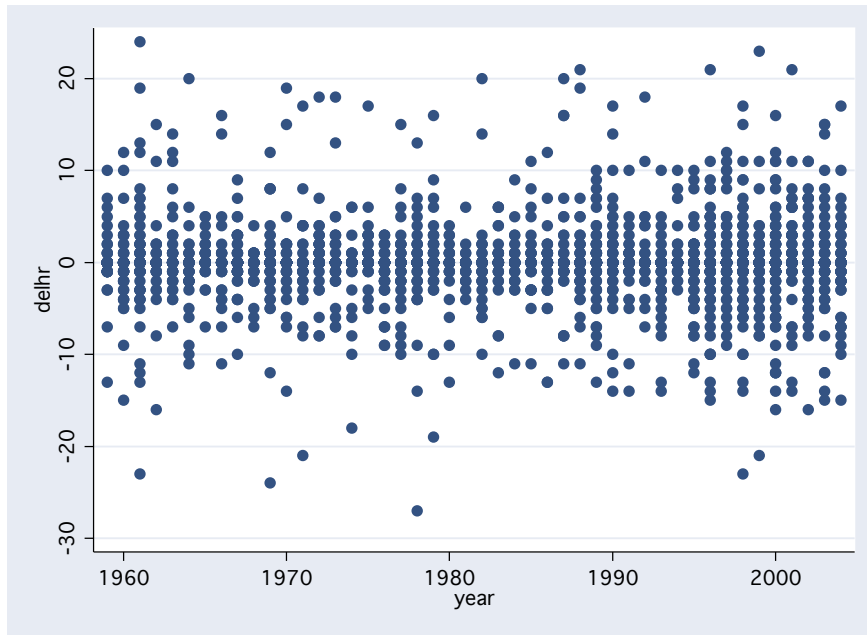Fig. 6.— Changes in Player Home Runs by Year

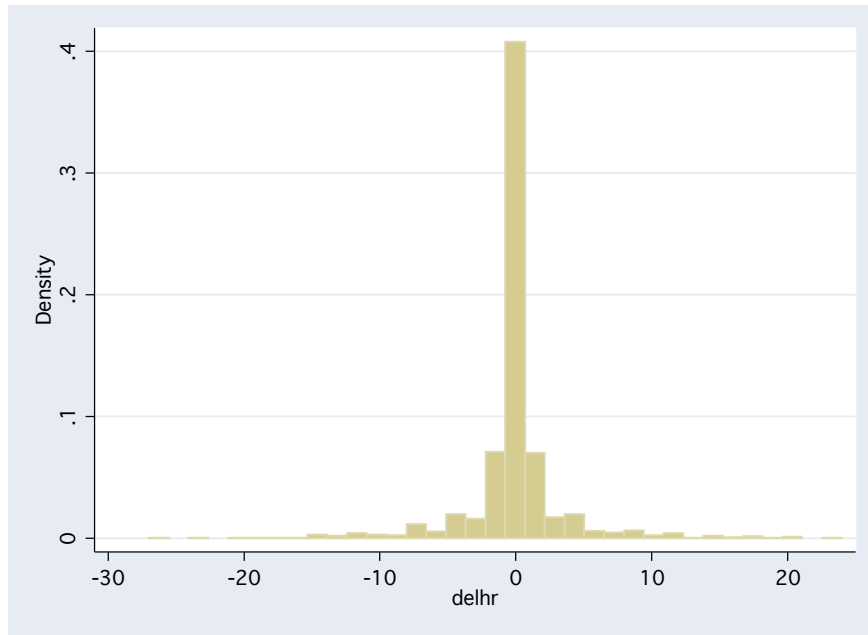Fig. 7.— The Law of Home Runs: Probability Density of Changes in Player Home Runs
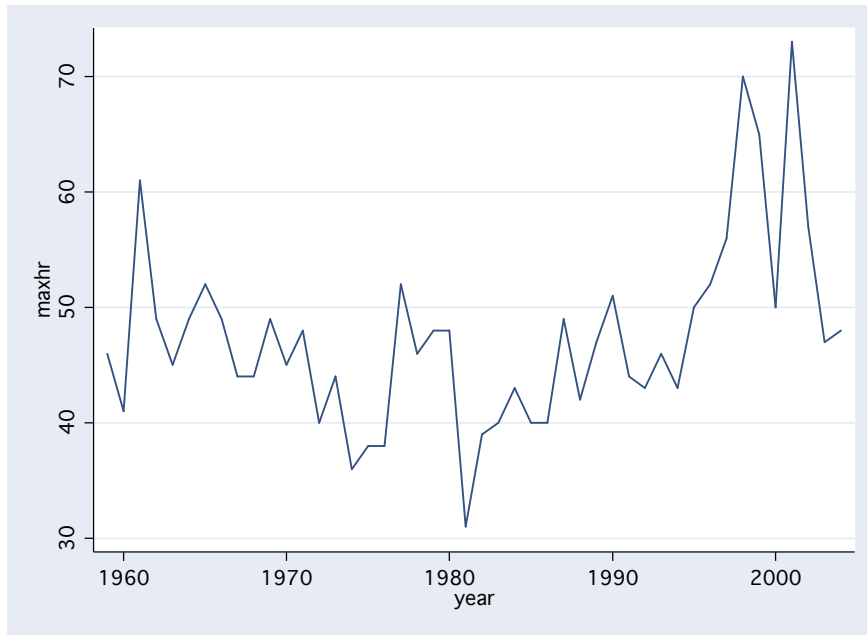
Fig. 8.— Maximum Home Runs by Year
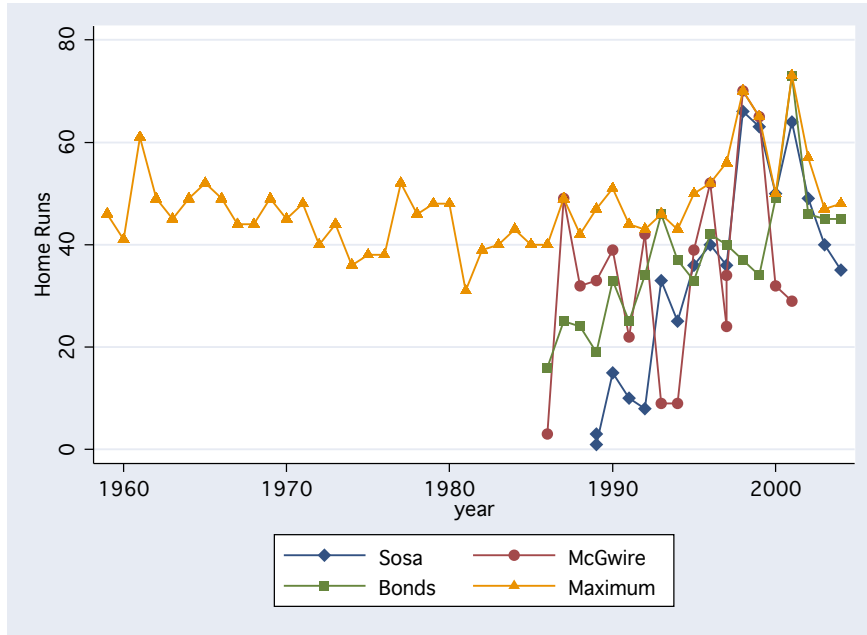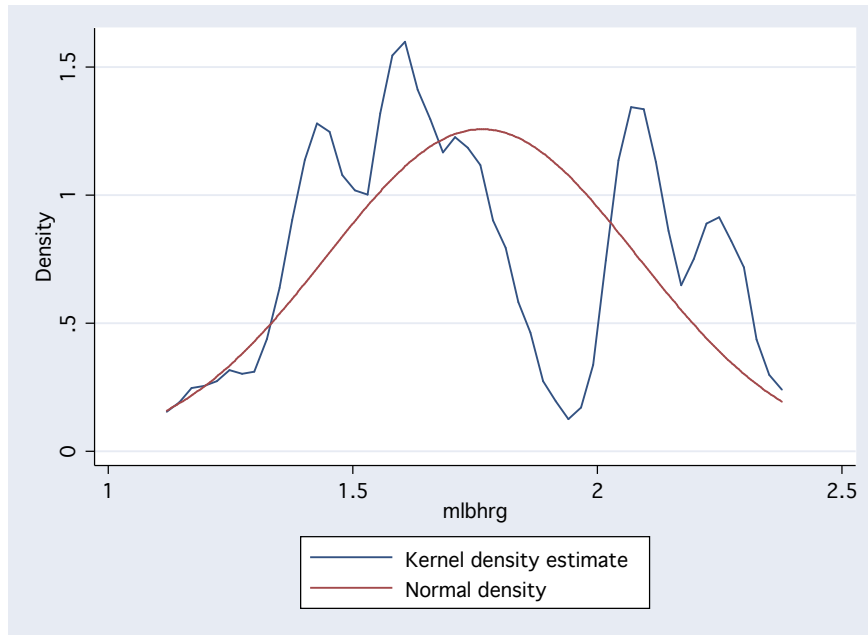
Fig. 9.— Bonds, McGwire, Sosa and MLB Leading Home Runs

Fig. 10.— Density of Home Runs per Game

14. Appendix

Table 10: Palmeiro's Hitting

year	HR	hrhit	hrab	AB	BB	SO	hrso
1986	3.00	0.17	0.04	73.00	4.00	6.00	0.50
1987	14.00	0.23	0.06	221.00	20.00	26.00	0.54
1988	8.00	0.04	0.01	580.00	38.00	34.00	0.24
1989	8.00	0.05	0.01	559.00	63.00	48.00	0.17
1990	14.00	0.07	0.02	598.00	40.00	59.00	0.24
1991	26.00	0.13	0.04	631.00	68.00	72.00	0.36
1992	22.00	0.13	0.04	608.00	72.00	83.00	0.27
1993	37.00	0.21	0.06	597.00	73.00	85.00	0.44
1994	23.00	0.17	0.05	436.00	54.00	63.00	0.37
1995	39.00	0.23	0.07	554.00	62.00	65.00	0.60
1996	39.00	0.22	0.06	626.00	95.00	96.00	0.41
1997	38.00	0.24	0.06	614.00	67.00	109.00	0.35
1998	43.00	0.23	0.07	619.00	79.00	91.00	0.47
1999	47.00	0.26	0.08	565.00	97.00	69.00	0.68
2000	39.00	0.24	0.07	565.00	103.00	77.00	0.51
2001	47.00	0.29	0.08	600.00	101.00	90.00	0.52
2002	43.00	0.29	0.08	546.00	104.00	94.00	0.46
2003	38.00	0.26	0.07	561.00	84.00	77.00	0.49
2004	23.00	0.16	0.04	550.00	86.00	61.00	0.38
Average	29.00	0.19	0.05	531.74	68.95	68.68	0.42

Source: /Users/asdevany/BaseballSoftball/HomeRunPaper/Batting1959-2004 copy.dta

Table 11: Bonds' Hitting

year	HR	hrhit	hrab	AB	BB	SO	hrso
1986	16.00	0.17	0.04	413.00	65.00	102.00	0.16
1987	25.00	0.17	0.05	551.00	54.00	88.00	0.28
1988	24.00	0.16	0.04	538.00	72.00	82.00	0.29
1989	19.00	0.13	0.03	580.00	93.00	93.00	0.20
1990	33.00	0.21	0.06	519.00	93.00	83.00	0.40
1991	25.00	0.17	0.05	510.00	107.00	73.00	0.34
1992	34.00	0.23	0.07	473.00	127.00	69.00	0.49
1993	46.00	0.25	0.09	539.00	126.00	79.00	0.58
1994	37.00	0.30	0.09	391.00	74.00	43.00	0.86
1995	33.00	0.22	0.07	506.00	120.00	83.00	0.40
1996	42.00	0.26	0.08	517.00	151.00	76.00	0.55
1997	40.00	0.26	0.08	532.00	145.00	87.00	0.46
1998	37.00	0.22	0.07	552.00	130.00	92.00	0.40
1999	34.00	0.37	0.10	355.00	73.00	62.00	0.55
2000	49.00	0.33	0.10	480.00	117.00	77.00	0.64
2001	73.00	0.47	0.15	476.00	177.00	93.00	0.78
2002	46.00	0.31	0.11	403.00	198.00	47.00	0.98
2003	45.00	0.34	0.12	390.00	148.00	58.00	0.78
2004	45.00	0.33	0.12	373.00	232.00	41.00	1.10
Average	37.00	0.26	0.08	478.84	121.16	75.16	0.54

Source: /Users/asdevany/BaseballSoftball/HomeRunPaper/Batting1959-2004 copy.dta

Table 12: Sosa's Hitting

year	HR	hrhit	hrab	AB	BB	SO	hrso
1989	2.00	0.08	0.02	91.50	5.50	23.50	0.08
1990	15.00	0.12	0.03	532.00	33.00	150.00	0.10
1991	10.00	0.16	0.03	316.00	14.00	98.00	0.10
1992	8.00	0.12	0.03	262.00	19.00	63.00	0.13
1993	33.00	0.21	0.06	598.00	38.00	135.00	0.24
1994	25.00	0.20	0.06	426.00	25.00	92.00	0.27
1995	36.00	0.24	0.06	564.00	58.00	134.00	0.27
1996	40.00	0.29	0.08	498.00	34.00	134.00	0.30
1997	36.00	0.22	0.06	642.00	45.00	174.00	0.21
1998	66.00	0.33	0.10	643.00	73.00	171.00	0.39
1999	63.00	0.35	0.10	625.00	78.00	171.00	0.37
2000	50.00	0.26	0.08	604.00	91.00	168.00	0.30
2001	64.00	0.34	0.11	577.00	116.00	153.00	0.42
2002	49.00	0.31	0.09	556.00	103.00	144.00	0.34
2003	40.00	0.28	0.08	517.00	62.00	143.00	0.28
2004	35.00	0.29	0.07	478.00	56.00	133.00	0.26
Average	33.76	0.23	0.06	471.82	50.35	124.12	0.24

Source: /Users/asdevany/BaseballSoftball/HomeRunPaper/Batting1959-2004 copy.dta

Table 13: McGwire's Hitting

year	HR	hrhit	hrab	AB	BB	SO	hrso
1986	3.00	0.30	0.06	53.00	4.00	18.00	0.17
1987	49.00	0.30	0.09	557.00	71.00	131.00	0.37
1988	32.00	0.22	0.06	550.00	76.00	117.00	0.27
1989	33.00	0.29	0.07	490.00	83.00	94.00	0.35
1990	39.00	0.32	0.07	523.00	110.00	116.00	0.34
1991	22.00	0.23	0.05	483.00	93.00	116.00	0.19
1992	42.00	0.34	0.09	467.00	90.00	105.00	0.40
1993	9.00	0.32	0.11	84.00	21.00	19.00	0.47
1994	9.00	0.26	0.07	135.00	37.00	40.00	0.22
1995	39.00	0.45	0.12	317.00	88.00	77.00	0.51
1996	52.00	0.39	0.12	423.00	116.00	112.00	0.46
1997	58.00	0.39	0.11	101.00	159.00	79.50	0.37
1998	70.00	0.46	0.14	509.00	162.00	155.00	0.45
1999	65.00	0.45	0.12	521.00	133.00	141.00	0.46
2000	32.00	0.44	0.14	236.00	76.00	78.00	0.41
2001	29.00	0.52	0.10	299.00	56.00	118.00	0.25
Average	34.29	0.36	0.10	363.94	77.47	93.88	0.36

Source: /Users/asdevany/BaseballSoftball/HomeRunPaper/Batting1959-2004 copy.dta