# The Law of Genius and Home Runs Refuted

John DiNardo and Jason Winfree*

August 23, 2007

**Abstract**

In a lively, provocative paper, DeVany (2007) claims *inter alia* that the size distribution of home runs follows a continuous "power law" distribution which is nested in a larger class of "stable" statistical distributions characterized by an infinite variance.

He uses this putative fact about the size distribution of home runs to argue that concern about the use of steroids to enhance home run ability is necessarily misplaced. In this paper, we show that the initial claim is false and argue that the subsequent claim about the potential importance of steroid use does not follow from the first. We also show that the method used to establish that the size distribution of home runs is characterized by an infinite variance is unreliable and will find evidence "consistent" with infinite variance in all but the most trivial of data sets generated by processes with finite variance. Despite a large and growing literature that spans several fields and uses methods and arguments similar to DeVany's, we argue that mere inspection of the unconditional distribution of some human phenomenon is unlikely to yield much insight.

## 1 Introduction

> Empirical regularities in biology, as in other fields, can be extremely interesting. In particular, such regularities may suggest the operation of fundamental laws. Unfortunately, apparent regularities sometimes cannot stand up under close scrutiny(Solow, Costello and Ward 2003)

.

A lively, provocative paper by Arthur DeVany (2007) in *Economic Inquiry* argues that:

- "the statistical law of home run hitting is the same as the laws of human accomplishment developed by Lotka . . . , Pareto . . . , Price . . . , and Murray . . . ." ,
- "there is no evidence that steroid use has altered home run" hitting,
- "the greatest accomplishments in [science, art, and music] all follow the same universal law of genius", and
- "the stable Paretian model developed here will be of use to economists studying extreme accomplishments in other areas."

which apparently follows from his claim that the size distribution of annual home run production has a finite mean but infinite variance and follows a "power law distribution." DeVany's argument is not unique: it is part of a large and growing literature where claims of the ubiquity of power laws are legion.[1]

*DiNardo: University of Michigan and NBER, jdinardo@umich.edu. Winfree: University of Michigan, jwinfree@umich.edu. No steroids were consumed during the production of this paper.

[1] A recent survey by Newman (2005) cites evidence that a diverse number of things allegedly "follow power–law" distributions including "city populations, the sizes of earthquakes, moon craters, solar flares, computer files, wars, the frequency of use of words in any human language, the frequency of occurrence of personal names in most cultures, the numbers of papers scientists write, the number of citations received by papers, the number of hits on web pages, the sales of books, music recordings and almost every other branded commodity, the numbers of species in biological taxa, and peoples annual incomes."

DeVany takes the additional step of connecting this statistical analysis to an argument about the effect of steroids on home run hitting by major league ballplayers: "Steroid advocates have to argue that the new records are not consistent with the law of home runs *and*, that the law itself has changed as a result of steroid use"

Our purpose of this paper is to suggest that the above should be met with a fair amount of skepticism.

First, we try to provide some background on previous attempts to identify the existence of universal laws and provide the intellectual context for DeVany's claims.

Second, we show that DeVany's claims follow from a flawed statistical inference procedure. His procedure, with probability one, would find evidence consistent with "infinite variance" for virtually any non–trivial data set. To do so, we first analyze the size distribution of a quantity which could not follow a power law distribution, and show that using DeVany's inference procedure we would be led to the same (incorrect) claim. We also discuss the important distinction – elided by DeVany and others writing in related literatures – between an unconditional distribution and a conditional distribution.

Third, we observe that the size distribution of home runs *can not* follow a power law distribution and show that the posited class of distributions provide an inadequate approximation to the data, *at best*.

Fourth, while concurring with DeVany's implicit criticism that "steroid advocates" who rely on recent "trends" to substantiate their views have not made their case, we suggest that the problem is that the question is ill–posed. The level and distribution in any given year is minimally a function of hundreds of things: the quality of pitching, the weather, the introduction of new ball parks, the number of games played, the distribution of baseball talent across the teams, etc.. To claim that only one "cause" is responsible for a trend involves some (possibly unstated) assumption about the myriad of other factors. Indeed, what is sauce for the goose is sauce for the gander: those seeking to support *or* deny the claim that increased use of steroids have led to increased home run hitting will have to employ considerably more "shoe leather" than mere statistical analysis of the unconditional distribution of home runs per player or time trends in home run hitting.

Fifth, we conduct a brief review of the evidence on the question of whether the causal effect of "judicious" steroid use on home run production is positive. Our answer that better-posed question is "yes, maybe."

We conclude by observing that neither examination of time trends in annual home run production nor examination of the unconditional distribution of home runs will settle the dispute between "steroid advocates" and "steroid opponents" and that more convincing evidence will have to be sought elsewhere.

## 2 Does a Power Law Imply "Self Organizing Criticality", etc?

We are not the first to argue that claims about universal laws should be met with some skepticism. Indeed, our criticisms are depressingly familiar.[2]

> The stringency with which the goodness of a fitted model should be assessed depends to a degree on the claims that are being made about the model. The claim that a model is correct, as opposed merely to providing a useful approximation, should be subjected to particularly close scrutiny. Such claims have been made about the power law model for size-frequency data without adequate scrutiny. (Solow et al. 2003)

Claims about the ubiquity of statistical distributions have a long history. A classic example is from Feller (1940).

> "The logistic distribution function . . . may serve as a warning. An unbelievably huge literature tried to establish a transcendental "law of logistic growth"; measured in appropriate units, practically all growth processes were supposed to be represented by a function of [a particular distributional form] Lengthy tables, complete with chi-square tests, supported this thesis for human population, for bacterial colonies, development of railroads, etc. Both height *and* weight

[2] See Keller (2005) for a useful review of some of the history.

of plants and animals were found to follow the logistic law even though it is theoretically clear that these two variables cannot be subject to the same distribution. Laboratory experiments on bacteria showed that not even systematic disturbances can produce other results. Population theory relied on logistic extrapolations (even though they were demonstrably unreliable). The only trouble with the theory is that not only the logistic distribution but ... other distributions can be fitted to the *same material with the same or better goodness of fit.* In this competition the logistic distribution plays no distinguished role whatever; most contradictory theoretical models can be supported by the same observational material.

> Theories of this nature are short-lived because they open no new ways, and new confirmations of the same old thing soon grow boring. But the naive reasoning as such has not been superseded by common sense, and so it may be useful to have an explicit demonstration of how misleading a mere goodness of fit can be." Feller (1940) as cited in Brock (1999)

Brock (1999), cited by DeVany, cites Feller to warn economists and others against making precisely the types of claims DeVany makes:

> I will make the general argument here that, while useful, these "regularities" or "transcendental laws" must be handled with care because ... most of them are "unconditional objects" i.e. they only give properties of stationary distributions, e.g., "invariant measures," and, hence, can not say much about the dynamics of the stochastic process which generated them. To put it another way, they have little power to discriminate across broad classes of stochastic processes.

Even active researchers in the area have begun to observe "that research into power laws ... suffers from glaring deficiencies" (Mitzenmacher 2006) Nonetheless, a long history of researchers making extravagant claims about phenomenon that derive from the resemblance of their size distribution to some statistical distribution has not slowed down the making of the claims. Feller's (1940) rejection of "universal models of growth", Solow et. al.'s (2003) rejection of power laws in biology, Miller and Chomsky's (1963) rejection of the usefulness of Zipf's law of word length (Zipf 1932) are prior (apparently failed) attempts to raise the level of discourse and raise the quality of attempts to "validate" or subject such theorizing to "severe testing" (Mayo 1996).[3]

Our argument is complicated by at least two issues:

1. DeVany argues that "steroid advocates" are wrong. Unfortunately, he cites no one actually making the claims he attributes to such advocates.

2. DeVany makes claims about the size distribution of home runs and refers vaguely to notions of "self organized criticality" without spelling out the implications of such notions for hypotheses about the effect of steroids on home run hitting.[4]

An important concern, which we address in Section 2.1, revolves about (2). What is the law of genius? How would we know if some phenomenon was subject to such a law? Indeed, what does it mean to say, as DeVany does, that home runs are "more like the movies ... or ... earthquakes ... than dry cleaning?"

A good introduction to "complexity theory" can be found in Krugman (1996). And though we can not recapitulate the logic entirely, we sketch the notion of "self organized criticality" which we believe is key to understanding the implicit argument DeVany makes. Only then is it possible to understand why some might find it plausible to assert that "the law of home runs" might look something like "the law of earthquakes" and why such an assertion might lead some to suggest that "steroids don't matter."

[3]While it is routinely claimed that the putative fact that size distribution of word lengths follows Zipf's implies something important about language, for example, Li (1992) observes that "probably few people pay attention to a comment by Miller in his preface to Zipf's book [(Miller 1965)] ..., that randomly generated texts, which are perhaps the least interesting sequences and unrelated to any other scaling behaviors, also exhibit Zipf's law."

[4]We would hasten to add that this omission may be for no other reason than editorial constraints as DeVany cites some of the relevant literature.

## 2.1 Sandpiles, Self Organized Criticality, and Home Runs?

To place both our arguments and DeVany's in context, it would be most helpful to provide a comprehensive review of some of the arguments made by students of "self organizing" or "complex" systems which lie at the heart of some of DeVany's analysis. We can not obviously do that here.[5]

Instead, we think we can convey much of the implicit logic at the core with a short description of the canonical example of a system displaying "self–organizing criticality" – Per Bak's sandpile(Bak 1996, Bak, Tang and Wiesenfeld 1988, Winslow 1997, Nagel 1992, Bretz, Cunningham, Kurczynski and Nori 1992).

Tesfatsion (2007) provides a nice intuitive explanation which covers most of the important points:

> When you first start building a sand pile on a tabletop of finite size, the system is weakly interactive. Sand grains drizzled from above onto the center of the sand pile have little effect on sand grains toward the edges. However, as you keep drizzling sand grains onto the center, a small number at a time, eventually the slope of the sand pile "self organizes" to a critical state where breakdowns of all different sizes are possible in response to further drizzlings of sand grains and the sand pile cannot grow any larger in a sustainable way. Bak refers to this critical state as a state of self-organized criticality (SOC), since the sand grains on the surface of the sand pile have self-organized to a point where they are just barely stable.
>
> What does it mean to say that "breakdowns of all different sizes" can happen at the SOC state?
>
> Starting in this SOC state, the addition of one more grain can result in an "avalanche" or "sand slide," i.e., a cascade of sand down the edges of the sand pile and (possibly) off the edges of the table. The size of this avalanche can range from one grain to catastrophic collapses involving large portions of the sand pile. The size distribution of these avalanches follows a power law over any specified period of time T. That is, the frequency of a given size of avalanche is inversely proportional to some power of its size, so that big avalanches are rare and small avalanches are frequent. For example, over 24 hours you might observe one avalanche involving 1000 sand grains, 10 avalanches involving 100 sand grains, and 100 avalanches involving 10 sand grains. . . . At the SOC state, then, the sand grains at the center must somehow be capable of transmitting disturbances to sand grains at the edges, implying that the system has become strongly interactive. The dynamics of the sand pile thus transit from being purely local to being global in nature as more and more grains of sand are added to the sand pile(Tesfatsion 2007)

.

Stipulating to this being an accurate description of avalanches in sandpiles,[6] and stipulating to the

[5] Krugman (1996) provides a sober yet optimistic discussion of this approach from an economists perspective. For an enthusiastic appraisal and simple introduction see Bak (1996) or Bak and Chen (1991). Krugman (1996) identifies three components of the complex system:

1. "Complicated feedback systems often have surprising properties."
2. "Emergence – [situations in which] large interacting ensembles of individuals [or neurons, magnetic dipoles, . . . "] exhibit collective behavior very different from [what one might have] expected by simply scaling up the behavior of the individual units."
3. "Self–organizing systems: systems that, even when they start from an almost homogeneous or almost random state, spontaneously form large scale patterns."

As Krugman observes, these components, especially the first two are not unique to complex systems. The standard general equilibrium model, for example, can be described as displaying complex feedback (everything depends on everything else). As to "emergence", it is possible to view the pareto optimality as "emergent" behavior generated by self–interested agents. Despite having some of the features associated with complex systems, neither of these would usually be viewed as examples of "complex systems."

(We don't mean to suggest that complex systems have not been developed or used by economists. An example of a classic model exhibiting all three components (and generally considered to be an example of a this approach) is Schelling's famous model of segregation, Schelling (1969) and (Schelling 1978).)

[6] In actual practice, while such a process is rather easy to generate in a computer simulation (Winslow 1997), even in the laboratory experiments with sandpiles require a fair amount of tweaking to behave in the idealized way described above.(Nagel 1992, Bretz et al. 1992).

ubiquity of such SOC in diverse fields and situations, some of the leaders in this field have drawn some rather wide-ranging implications for science or social science:

> If this picture is correct for the real world, then we must accept instability and catastrophes as inevitable in biology, history, and economics. Because the outcome is contingent upon specific minor events in the past, we must also abandon any idea of detailed long–term determinism or predictability. *Large catastrophic events occur as a consequence of the same dynamics that produces ordinary events. This observation runs counter to the usual way of thinking about large events, which . . . looks for specific reasons (for instance, a falling meteorite causing the extinction of dinosaurs) to explain large, catastrophic events.* Bak (1996) (page 32, emphasis added).

To put it yet a different way, the sandpile forms, experiences avalanches, etc. as a consequence of a *single* causal process. Great catastrophes arise from the *identical* mechanism as the periods of non–catastrophes.

We think DeVany means to make a similar argument regarding the production of home runs: home runs are the "catastrophe" in a SOC process. Applying Bak's and DeVany's logic to home run production, we might be led to conclude that the process that produces a year with few home runs for an individual batter can be identical to the process that produces a year with an extremely large number of home runs. Moreover, a further hunt for causes for extreme events might be unwarranted.

As we discuss in detail below, this seems an unwise inferential leap. Even in the case of sandpiles, the fact that avalanches *can* arise from the same causes that generate periods of low avalanche activity does not necessarily imply that other causes aren't or can't be at work. We conjecture, for example, that the introduction of a typical three–year–old with a plastic shovel into a sandpile laboratory might predictably lead to avalanches even in a system that until that time exhibited SOC. At a minimum, we doubt that many parents would accept without question a three–year–old's denial of involvement with the sandpile avalanche on the grounds that he or she could not have caused the avalanche since the sandpile exhibited SOC, – especially if the three–year–old is observed in the vicinity of the avalanche with sand all over their clothes.

# 3 A Powerless Power Law Test

The bulk of the statistical analysis in DeVany is in Section 5 "The Distribution of Home Runs" and Section 6 "The Law of Home Runs." The core of the statistical argument and upon which the subsequent statistical analysis rests is that the *unconditional* distribution of home runs hit in a year follows a so-called "stable distribution." [7]In particular, the claim is made that the distribution of home runs is characterized by a subset of this class of "stable distributions" in which the variance of home runs is infinite. Consequently DeVany infers that "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 . . . or earthquakes . . . than dry cleaning."

How does DeVany establish that the size distribution of home runs follows a power law? The method is simple. Fit the data to a "stable" distribution and check whether the estimated parameters are consistent with a stable distribution with infinite variance. If so, conclude that the data is generated from a 'wild' statistical distribution and follows the "universal law of genius."

> The exponent 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 ($\alpha$. ). This remarkable feature tells us that the weight assigned to extreme events is the key distinguishing

[7] "Stable distributions are a rich class of probability distributions that allow skewness and heavy tails and have many intriguing mathematical properties." (Nolan 2007). One difficult aspect of these distributions is that, except in a few special cases, there exists no closed–form expression for the probability density and distribution functions.

Table 1: $X$ Versus the Home Run Data: Fitted to the "Stable" Distribution

| | Index–$\alpha$ | $\beta$ | Scale | Location |
|---|---|---|---|---|
| DeVany's Data | 1.6422 | 1.00 | 6.219 | 12.30 |
| $X$ | 1.0006 | 0.912 | 7.90609 | 2.297 |

MLE estimates of the four–parameter Stable distribution.The estimates in the first row replicate DeVany (2007) using home run data from 1950–2004. The estimates in the second row use data on variable "$X$"; see text for details.

> property of a stable probability distribution. ...[8] The tails of a stable distribution are Paretian and moments of order $\geq 2$ 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 $\alpha$ 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 implies a less rapid decay of probability.

Put more simply, DeVany's procedure is: *Estimate the four parameter Stable distribution. If the estimated value of $\alpha < 2$ conclude that the distribution of home runs has infinite variance.*

## 3.1 Does $X$ follow the law of Genius?

To make clear why this analysis is problematic, we perform a similar analysis on a different random variable, which we call $X$ for the moment. Following DeVany, we display a smoothed histogram of $X$ (using the conventional "Silverman rule–of–thumb bandwidth") and compare it the normal distribution implied by the empirical mean and variance of "X" in Figure 1.

As is true with size distribution of home runs, the size distribution of $X$ is decidedly non–normal. As with the home run data, the upper tail is poorly–fit by the normal distribution. Table 1 repeats the more formal analysis in DeVany (2007). The table displays our estimates of the four parameters of the "Stable" distribution by maximum likelihood using the same program as DeVany (2007) but using data $X$.[9] We display our results for $X$ along side DeVany's results using the same data on individual home run hitting. (See DeVany (2007), Table 1 ).[10] While the distribution of $X$ and the distribution of and home runs aren't identical, they both have the properties which are "consistent" with the random variable $X$ possessing an infinite variance, namely that the estimated value of $\alpha$ (one of the four parameters of the Stable distribution) is less than 2.

Does $X$ follow a power law? No. We defined $X$ as the the number of mentions (times five) of the word "normal" or "normality" on a page of the web-draft of DeVany (2007).[11] Surely $X$ does not possess infinite variance: presumably the number of words that *Economic Inquiry* will allow to be printed on a page is finite;

[8]The stable distribution has a total of four parameters. For the other 3 parameters "...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 $\gamma$ must be positive. It expands or contracts the distribution in a non-linear way about the location parameter $\delta$ which is the center of the distribution."(DeVany 2007). Following DeVany we limit our discussion to just the one parameter, $\alpha$.

[9]See Rimmer and Nolan (2005) for details.

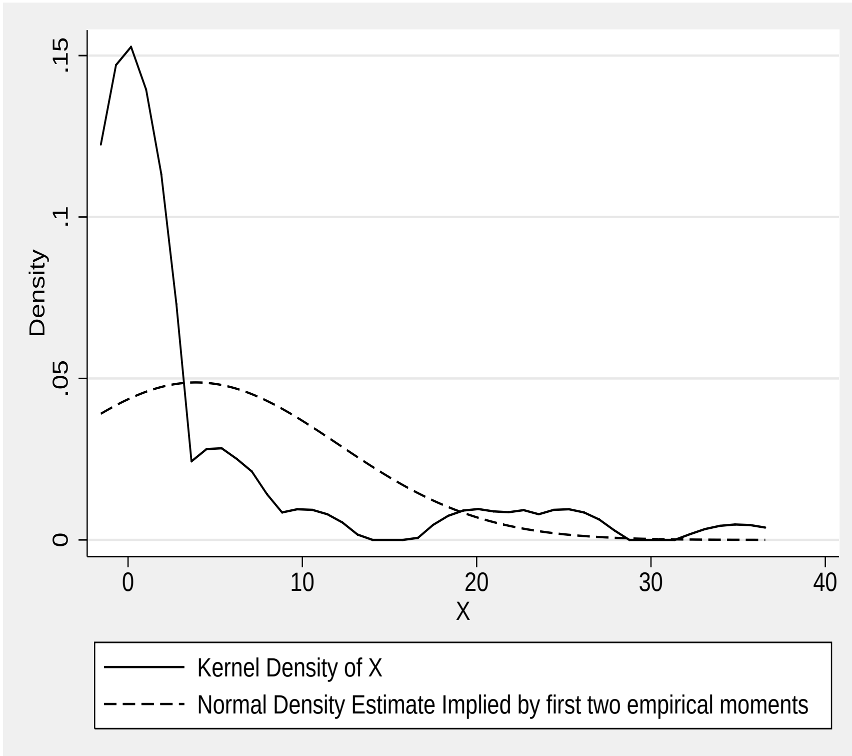
[10]Our estimates of the four parameters are identical to those estimated by DeVany (2007), although our calculated value of the maximized log–likelihood function is somewhat larger than reported in the paper.

[11]The data we used was:

| Page | 2 | 3 | 6 | 7 | 8 | 9 | 10 | 11 | 13 | 16 | 20 | 22 | 41 | The other 32 pages |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of Mentions | 1 | 1 | 5 | 7 | 5 | 1 | 4 | 2 | 1 | 1 | 2 | 4 | 1 | 0 |

We multiplied the number of mentions by five. The data was collected using the (undated) web draft which was created on 6–14–2006. For the kernel density estimate used an Epanechnikov kernel and a bandwidth of 1.558. The normal density estimate used the sample mean of $X$ which was 3.89 and had a sample standard deviation of 8.18.

Figure 1: Does $X$ Follow A Power Law




an author that proposed to submit an article including nothing but the words "normal" and "normality" would stand a low chance of having the article included in the journal.

Why is DeVany's procedure flawed? Most simply, observing that the estimated value of $\alpha$ is less than 2 can only be construed as evidence for an infinite variance conditional on the data actually following the "stable" distribution. Among *stable* distributions (see Nolan (2007)), those consistent with finite variances only occur on the boundary of the parameter space − when $\alpha = 2$. Since $\alpha \leq 2$ − by definition − DeVany's procedure will always provide evidence for an infinite variance unless it reaches the boundary. Putting aside the considerable difficulties in maximum likelihood estimation when the true value of the parameter lies on the boundary of the parameter space, even a variable "just shy of normality" − i.e. $\alpha = 1.\overline{9}$ − is "consistent" with an infinite variance. More importantly, if the the data is not from the stable distribution – say, uniformly distributed, exponentially distributed, etc. − such a procedure will almost surely result in a estimated value of $\alpha < 2$.

# 4 Other Problems with the Analysis

There are other significant problems with the analysis in DeVany (and in much of literature which purports to have found evidence for the workings of "power laws"):

1. There is a failure to distinguish between conditional and unconditional distributions. If the number of at bats, for example, were allowed to follow a power law, the relationship between home run hitting and at bats could be non-stochastic, deterministic, and purely mechanical and the unconditional distribution would follow a power law. Home run hitting in such a situation would be more like dry–cleaning than "genius" despite the fact that the size distribution of home runs followed a power law.

2. Like much of the literature, DeVany does not contemplate the possibility that the observed size distribution of home runs is a mixture of many different – individual – (non–power law) statistical distributions. Hence, estimating the parameters of a single (falsely imposed) statistical distribution can not, in general, be adequate for reliable inference about the potential existence of a "fundamental" law.[12]

One important problem with the analysis is that it fails to distinguish between a *conditional* probability density function and an *unconditional* probability density function. To make this point as transparently as possible, consider the unrealistic case where the "law of home runs" is strictly *deterministic* –as "non–wild" as it can be – and steroids matter in a purely *mechanical* way *conditional* on the number of at bats.

For instance let the total number of home runs hit by player $i$ in a given year ($h_i$) be given by *deterministic* function, more akin to "dry cleaning" than "genius":

$$h_i = (a + s_i)x_i \tag{1}$$

where $a$ is a positive constant, $s_i$ is zero or a positive constant such that ($a + s_i \leq 1$) as the ball player does not or does use steroids. $x_i$ is the (identical) number of at bats a player gets. *Conditional* on the number of at bats, the distribution of home runs would have *zero variance.*

If we allowed for the impossible – (1) that there was no bound to the number of at bats player $i$ could face, and (2) that the distribution of at bats had infinite variance *then* an econometrician analyzing data arising from this data generation process would learn the following:

1. if all players always used steroids, the unconditional variance of home runs would be infinite

2. if some players did not use steroids, *conditional* on $x_i$, the variance of home runs would be no more than the necessarily finite variance of $s_i$. "Genius" would be irrelevant.

The moral of this story is that mere inspection of the variance of the unconditional distribution of total home runs could tell us *nothing* about whether steroids matter.

We note in passing, moreover, there is no reason to believe that a single law applies to all baseball players in all times and all situations. Thus far, following DeVany, we have focused our attention on the unrealistic case where each player is an i.i.d. draw from the same distribution. Even if we stipulate that the same class of statistical distributions characterize every baseball player, the evidence is inconsistent with each player drawing from a distribution with the same parameters. In such a world, we would also expect that the individual with, say, the maximum home runs in a season would be essentially chosen at random from all players. Today's home run leader might be next season's zero home run hitter. Rather, we might even want to consider a finite mixture distribution of the class:

$$\sum_{i=1}^{N} p_i f(x; \theta_i)$$

where $p_i$ would represent the fraction of the observations contributed by player $i$ and $f(x; \theta_i)$ is the probability mass function with parameters $\theta$ that might vary from player to player. The joint assumption that all players follow the same distribution $f(\cdot)$ and that $\theta_i = \theta \ \forall i$ is clearly violated. Some players are never going to hit a lot of home runs. Moreover, in general, it seems too much to expect to recover parameters useful for demonstrating the existence of a fundamental law by falsely imposing a single distribution on data which is generated by a mixture of several distributions.

## 4.1 Power Law as "Law" and "Approximation"

In section 3 we documented the difficulties with DeVany's inference procedure as well as the more general problem of reasoning about the existence of a "law" from the unconditional distribution from a quantity.

[12] It is possible, however, that aggregation of objects following their own power law could itself produce another power law. See Gabaix (1999) for example.

Thus far we have argued that the inference *procedure* was faulty. Nonetheless it remains possible that the *inference* drawn from such a procedure might be correct: a broken clock is still correct twice a day, as the adage goes.

Unfortunately, such is not the case here. The problem is so grave that it is considerable work to even contemplate a situation in which it might be reasonable to characterize the distribution of home runs as following a "power law" and hence having a tail that is "subject to bursts or avalanches". That is, the distribution of home runs *can't* follow the distibution that DeVany posits and even if it could, he is not licensed to draw the inferences that he does about the nature of home run production.

We agree with criticism of related work on SOC that a serious problem with this literature is the unwillingness to put the proposition that an outcome follows a power law to even a minimally severe test. In a nice discussion, Solow et al. (2003) suggest that the problem with much of the "power law" literature in biology is the failure to evaluate the power of the power law against an explicit *alternative*. [13] Indeed, DeVany, following a tradition in the "complex studies" literature, considers no alternative to a "stable" distribution characterized by infinite variance but the normal.

While we wholeheartedly concur with this critical judgement (and compare a power law to other distributions) we wish to emphasize that in the *present case* such an analysis is superflous: there are other even more insurmountable obstacles.

## 4.2 Why Home Runs are Immediately Inconsistent with a "Stable Distribution"

The most immediate problem is that the size distribution home runs is *immediately* inconsistent with the posited distribution in DeVany (2007) even before approaching a systematic analysis of the data:

1. The number of home runs is bounded below by 0. Indeed, Figure 4 of DeVany displays only *part* of the estimated probability density function, that part where the number of home runs is greater than or equal to zero. The estimated power law distribution, if it were to be taken literally, predicts that 11 percent of baseball players would have a negative quantity of home runs. We think it safe to assume that negative home runs don't exist.

2. The number of home runs by a given player is discrete, not continuous as posited by the class of stable functions DeVany choose to estimate. No one will ever hit 1.2 home runs in a season. Somewhat surprisingly, DeVany makes a related observation regarding team production of home runs when he dismisses the "home runs per game statistic"[14]

3. If we are willing to assume that the number of games, at bats, etc. in a given year is bounded from below by 0 and above by some arbitrarily large value $\overline{M}$ then it immediately follows that for any discrete distribution the variance is bounded by $\frac{\overline{M}^2}{4}$ which is the variance of the Bernoulli distribution with equal sized mass points at 0 and $\overline{M}$. Indeed, there is a long literature on establishing bounds for the variance of distributions with finite domain – see for example, Muilwijk (1966),Gray and Odell (1967), Jacobson (1969) – where the bounds can be tightened under various conditions (assumptions about symmetry, unimodality, etc.).[15]

[13]Specifically, they considered data from Yule (1925), an early proponent of a power law hypothesis. E. Udny Yule is better known perhaps for his work in Economics where he documented a positive correlation between the degree of pauperism in a district and the generosity of provision of food for the poor; this was used to argue that there was a *causal* relationship between the generosity of such relief and the degree of pauperism in Yule (1899). See Freedman (1999) for a discussion. Yule used data representing the frequencies of genera of different sizes for snakes, lizards, and two Coleopterans (Chrysomelidae and Cerambycinae). When Solow et al. (2003) examined 4 of Yule's cases they were able to reject the discrete "power law" distribution proposed by Yule (1925) versus a discrete non–parametric alternative in three of the four cases. This is also the same non–parametric alternative we estimate in the column of Table 2 labeled "Pool–Adjacent Violators"

[14]From page 22 of DeVany (2007): "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."

[15]**N. B.** The existence of bounds somewhere in the data generation process is not necessarily inconsistent with some version

Table 2: Maximum Likelihood Estimates of the Size Distribution of Home Runs per Player in Major League Baseball – 1950–2004

| | Estimations including zero HR Observations | | | | Estimations not including zero HR Observations | |
|---|---|---|---|---|---|---|
| Distribution | Stable[2] | Stable[3] | Stable[4] | Negative Binomial | Discrete Power Law | Pool-adjacent violators[5] |
| Index ($\alpha$) | 1.6422 | 1.64221 | 1.64221 | | 1.378 | |
| $\beta$ | 1.00 | 1 | 1 | | | |
| Scale | 6.219 | 6.21928 | 6.21928 | | | |
| Location | 12.30 | 12.3041 | 12.3041 | | | |
| r | | | | 1.506172 | | |
| p | | | | 0.1141677 | | |
| Log Likelihood | -39294.2 | -43812.4 | -43217.8 | -41780.7 | -47552.8 | -39528.0 |
| Number of obs | 11992 | 11992 | 11992 | 11992 | 11552 | 11552 |

[1]Version 5.3 of the data was obtained at `http://baseball1.com/content/view/57/82/` Following DeVany (2007) we drop observations in the year 2005 or persons with less than 200 at bats. Therefore all player-years with at least 200 at-bats from 1959 to 2004 were in the sample. We note that the data also includes multiple observations from some players in the same year if they played for multiple teams, or had multiple "stints". This also implies that a player's home run total is only for a specific team for that year and not necessarily the entire season.

[2]Estimates reported in De Vany (2007).

[3]Estimates from using the `Sloglikelihood` command to calculate the maximum likelihood value in *Mathematica* (Rimmer and Nolan 2005)

[4]. The maximized value of the log likelihood function is calculated by adding the log of the probability distribution function at each home run value observed in the data.

[5]As described in Solow et al. (2003). See also text.

4. There is no single description of a "power law." Indeed, in the case of discrete variables it is common to define a power law as a probability mass function (see Newman (2005)) such that:

$$f(x) \propto x^{-\alpha} \tag{2}$$

This is of course problematic if $x$ can take the value of 0. One may choose the expedient of focusing on observations above which exceed some threshold (and above zero) in the discrete case, and describing the results as consistent with "the upper tail following a power law" (even if the distribution above some threshold follows some other non power law distribution [16]) but an estimation procedure that allows one an extra degree of freedom to choose this threshold after looking at the data is obviously not going to be very powerful.[17]

## 4.3 Fitting Unconditional Distributions

Despite the substantial caveats we have enumerated, we present several different attempts at fitting single distributions to home run data in Table 2.

of a power law. For example, a random walk model of *growth* with a (lower) barrier could produce a size distribution consistent with Zipf's laws. See, for example, Gabaix (1999). More descriptively accurate models would have to allow for the "birth" and "death" of new ballplayers.

[16]See Nolan (2007)

[17]See, in particular, our analysis of cumulative density function below.

In the first panel, we consider the data *including* zeros. In the second panel, we conduct an analysis excluding data on individuals who hit no home runs ("Excluding Zeros"). In the first column of the table we present DeVany's estimates of the stable distribution. In the next two columns we reproduce our estimates using two variants of the same *Mathematica* program used by DeVany to generate his results. In the next column we report the maximum likelihood estimates of the two parameter negative binomial distribution.

Next we repeat the exercise with a sample that excludes all individuals with zero home runs and present the results of fitting an appropriate version of a *discrete* power law. Following Solow et al. (2003), we estimate the parameters of an alternative class of distributions that that has declining tails. Specifically, we fit the size distribution of home runs subject only to the constraint that frequency with which individuals hit a specific number of home runs is non–increasing in the number of home runs. That is, if $p_k$ is the probability of a player hitting $k$ home runs, and $n$ is the highest number of home runs that can be it in a season,

$$p_n < p_{n-1} - \ldots < p_k < p_{k-1} < p_{k-2} < \ldots < p_0 \quad (3)$$

This is a nice class of distributions to consider since the condition in equation (3) is necessary for the size distribution to be from a power law class, but not *sufficient*.

We draw several conclusions from this statistical analysis:

1. With the exception of the maximized value of the likelihood function (which we suspect is a mere typographical error in the original DeVany manuscript), our estimates of the parameters are essentially identical.[18]

2. Despite having four parameters, the "stable" distribution does a poor job of "fitting" the data. The negative binomial distribution, with only two parameters, for example results in a higher value of the maximized log likelihood. If you were to believe that the "stable distribution" or the negative binomial distribution were the only two hypothesis to be considered, considered them equally likely, (and were willing to overlook the negative and fractional home run predictions of the stable distribution!) the "weight of the evidence" (Good 1981, Peirce 1878) would *still* be against the power law distribution.[19] Of course, if you were to allow other possibilities you would certainly reject the stable distribution and quite possibly the negative binomial distribution. We also illustrate this point in Figure 2 with a graph of the estimated stable distribution, negative binomial distribution, and the histogram of the data.

   Clearly, the negative binomial distribution estimates the actual distribution better than the stable distribution. It does not mistakenly predict negative home runs.

3. The situation looks no better when we focus just on the positive observations. As before, the "weight of the evidence" is against the power law distribution.

   We would like to stress that the problem is not unique to DeVany.

   > While the arguments found in the statistics literature concerning the use of scaling distributions for modeling high variability/infinite variance phenomena have hardly changed since Mandelbrots attempts in the 1960s to bring scaling distributions into mainstream statistics, discovering and explaining strict power law relationships has become a minor industry in the complex science literature. Unfortunately, a closer look at the fascination within the complex science community with power law relationships reveals a very cavalier attitude toward inferring power law relationships or strict power law distributions from measurements. (Willinger, Alderson, Doyle and Li 2004)

[18] We corresponded briefly with Professor DeVany on the subject. We have not been able to determine the source of the discrepancy in the estimate of the maximized value of the log likelihood function but we suspect it is typographical error in the DeVany manuscript given the almost exact correspondence between his and our estimates of the parameters of the distribution and the fact that we appear to be using the exact same data set (judging by the number of observations and sample means DeVany reports.)

[19] When comparing only two statistical hypothesis, the difference in the value of the log likelihood function can be interpreted as the (Bayesian) posterior log odds ratio if the the initial probabilities attached to the two possibilities were 0.5.

Figure 2: Empirical and fitted MLE estimates of Probability Density Function

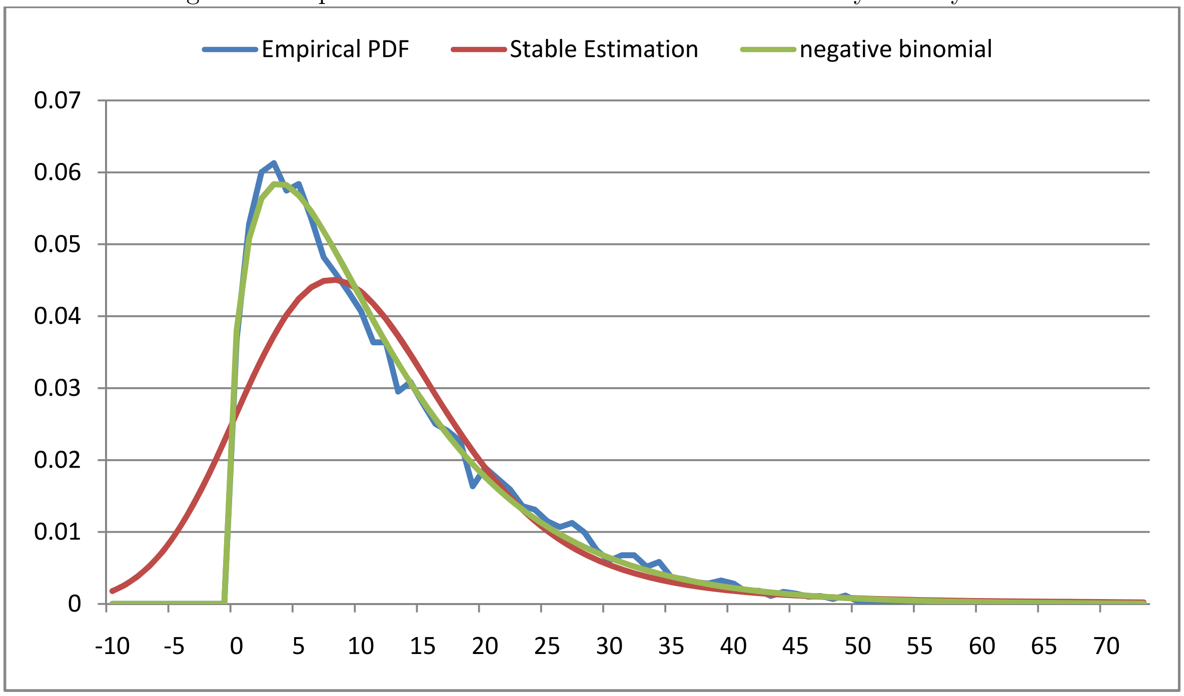



Figure 3: Plot of Fitted Stable Distribution Estimates and Empirical CDF from Devany (2007)

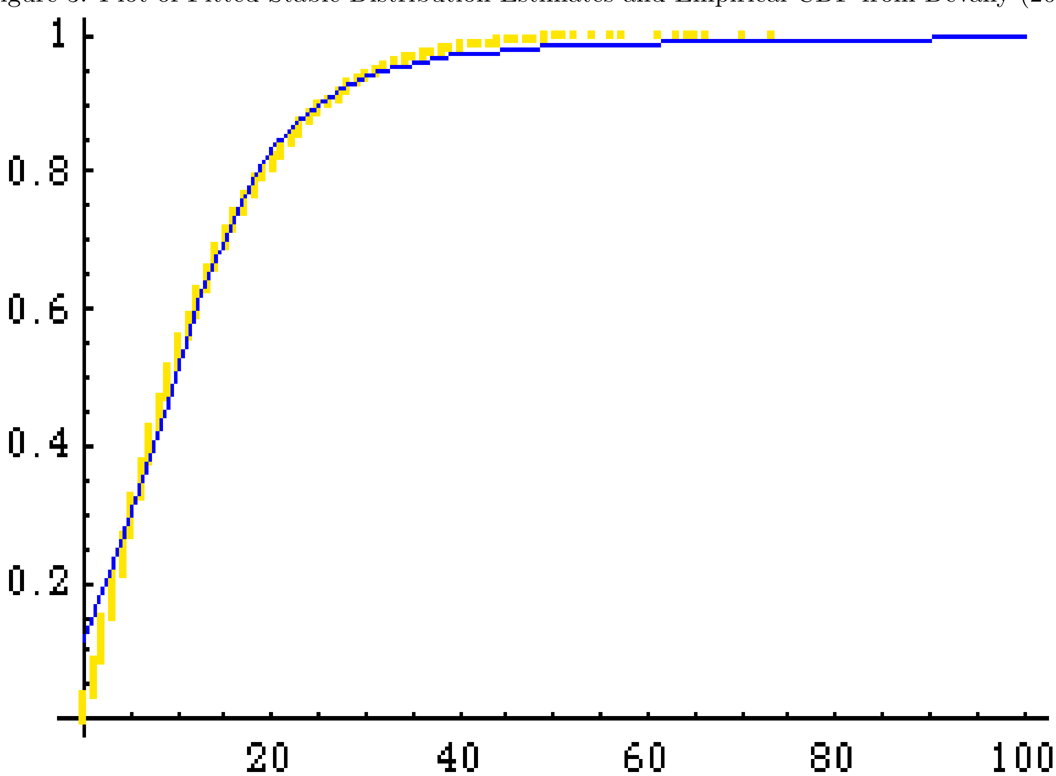


Figure 3, taken directly from of DeVany (2007)[20], is a case in point. As he describes it, this figure displays a "remarkable" fit of the "cumulative theoretical and empirical distributions." One need not cavil about the definition of "remarkable" to demonstrate that with a more appropriate metric of "fit", the home run data is not well approximated by a power law.

The problem with DeVany's figure is, as Willinger et al. (2004) demonstrates, that such a display is quite powerless; with such a plot it is difficult to distinguish power law from non power law data or discriminate among power laws (i.e. different values of $\alpha$.) Even if we stipulate to "ignoring the zeroes," it is easy to generate a more powerful visual test of the proposition.

One aspect of "self–similarity" – as this property is referred to in the complex systems literature[21] – is that the definition in equation 2 implies the "complementary cumulative density function" (CCDF) is *linearly* related to size:

$$\log(1 - P(x <= x_0) \approx a - \alpha \log(x_0) \tag{4}$$

where $(1 - P(x <= x_0) \equiv P(x > x_0)$ is the CCDF, or one minus the cumulative probability of hitting at least $x_0$ home runs. The approximatioin becomes exact as $x_0 \longrightarrow \infty$. This property suggests a useful visual display to assess the fit of the data to a power law: one merely plots the natural logarithm of the CCDF against the log of size. This particular display highlights the fit (or lack of fit) in the *tails* of the distribution and makes it relatively easy to distinguish the fit of the tail to different choices of $\alpha$.

As Figure 4 demonstrates, the power law provides a poor approximation globally and in the tails of the distribution. The most appropriate power law – the simple discrete version of the power law – gives the *worst* fit to the data, globally and in the all–important tail. The "inappropriate" power law (the continuous stable version) gives a slightly better fit but fits quite poorly in the tail. The negative binomial distribution – which is as well behaved as it is possible for a distribution to be – seems more deserving of the moniker "remarkable" than the power law distributions in terms of quality of fit.

Figure 4 also helps explain why it is much easier to find a power law if one is allowed to characterize *part* of the distribution *that one chooses after the fact* as being a power law: it is easy to convince oneself that even a very convex shape is linear if one can systematically ignore part or most of the curve.[22]

There is another, informal, yet instructive way to evaluate how well the continuous stable distribution works as "the law of genius." Under the hypothesis that the fitted continuous stable law distribution is correct, we can use estimates from the CDF to generate predictions for the number of "genius" home–run hitters we should have expected to see over the period from 1959 to 2004. We can also do the same with our "dry cleaning" distribution, the negative binomial distribution.

For example, according to the negative binomial distribution, the expected number of players that would have hit 100 or more home runs is 0.23; the expected number who would hit more than 1000 according to the same estimates is essentially zero.

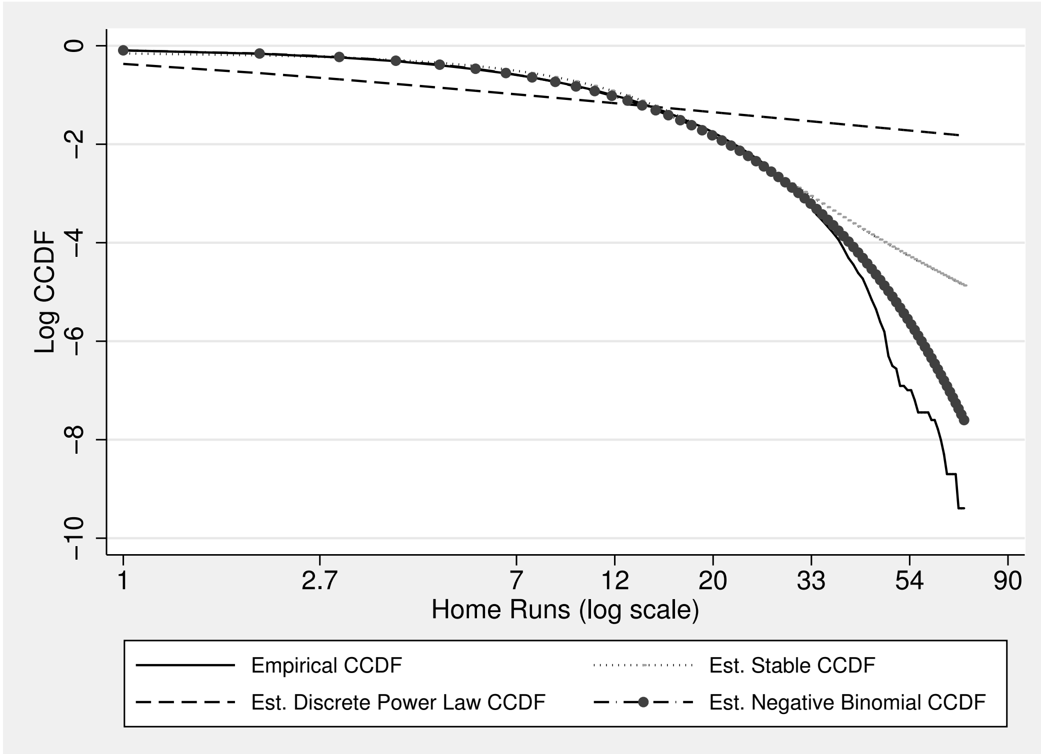
This is arguably a sign of bad fit for the negative binomial distribution. However bad the fit, the continuous stable law fits remarkably worse; our estimates from that distribution suggest that there should have been over 48 players to hit 100 home runs or more. Again, according to that same distribution, we would expect .88 players, to hit 1000 home runs or more! Worse yet is the estimated discrete power law distribution: By that distribution we would have expected to see 1709 players hit 100 or more home runs and 716 players would have to hit 1000 or more home runs! This would be unthinkable to most baseball fans, especially since record for *at-bats* is 705. (An important distinction between the continuous and discrete

[20] It is labeled Figure 5 in his paper.

[21] DeVany discusses this briefly in section 10 and on page 11: "[if the distribution is from the stable distribution] this implies that any way you look at the process you should that the distribution has the same shape."

[22] The fit of the continuous stable distribution fit to the non–zero observations is no better than that produced by using all the observations as DeVany does and to avoid clutter, it is dropped from Figure 4.

Figure 4: Log Log Plot of Complementary Cumulative Distribution Function of Home Runs




versions of the power law distributions being discussed is that the former has more parameters. It is not surprising it fits better, even for discrete data.)

We hasten to add that although the news is unremittingly bad for the power law distribution that we and DeVany have estimated, we do not mean suggest that we believe that any of our alternative distributional choices are realistic or even particularly useful. Indeed, the whole idea of fitting a parametric model of the size distribution of home runs seems like a really bad idea (except perhaps as a "quick and dirty" way to communicate some features of the data.) Like any human endeavor (and much else) home run hitting is a process so ill–understood that it would be a miracle if any simple parametric model (such as the stable distribution) were able to characterize it.[23]

Apropos of why one would expect some outcome to be distributed as a power law or some other class of distributions, it is also important to remember that the ubiquity of the normal distribution in statistical analysis does *not* arise because the characteristics of the *objects of study* are distributed normally – rather they often follow because we are studying systems that can be well approximated by "chance set ups" (Hacking 1965) – the randomized controlled trial is the canonical example of such a set up – and the *sample means* of such a process can be shown, by some variant of the Central Limit Theorem[24], to be approximately normal even when the outcomes under consideration are *not* distributed normally, as long as the outcomes

[23] Indeed, one of the serious problems with the "power law" hypothesis is that it would be difficult to learn about without enormous amounts of data. The "wild" distributions discussed by DeVany take their character from the extreme tails of the distribution. Such phenomenon, are consequently "rare" and therefore quite difficult to learn about. Heyde and Kou (2004), for example, observe that there are good reasons to doubt simple comparisons of likelihoods in this context. In part, this is a problem because of the importance of correctly characterizing the tails of the distribution. A sharp ability to discriminate between a tail following a power law distribution and a tail following an exponential distribution generally requires enormous amounts of data, at a minimum.(Heyde and Kou 2004).

[24] In fact, the notion of Lévy–stable or stable distribution is so named since it is a CLT of sorts for variables with infinite variance. (Gnedenko 1943, Gnedenko and Kolmogorov 1954)

have finite variance.[25] As we discussed in Section 4, assessing whether an outcome has a finite variance can often be established merely by demonstrating that outcome is bounded.

# 5 "Well, Who Are You Gonna Believe? Me Or Your Own Eyes?"[26]

As we have argued thus far, mere inspection of the size distribution of a random variable is insufficient to draw any conclusions about the *process* generating the data. While some distributions might allow for a rough approximation of the data, and this may be sufficient for some purposes, an *approximation* is *not* adequate for the purpose of drawing some sort of "causal" inference. That is, it may be fair to say that Zipf's power law – $P(Size > S) \propto \frac{1}{S}$ – provides a rough approximation to the size distribution of cities (Gabaix 1999), but quite another (inappropriate) matter to infer anything about the mechanism of city growth directly from that fact. To take one example from economics, Gabaix (1999) demonstrates that the mechanisms that could induce a Zipf's law for cities could be very different and result in very different inferences: "[although] the models [might be] mathematically similar, they [may be] economically completely different."[27]

Of course, it is difficult to establish the effect steroids had on the distribution of home runs. As Fair (2007) observes performing an analysis not unlike what we do here, "Since there is no direct information about drug use in the data used in this paper, . . . [such evidence] can only be interpreted [at best] as showing patterns for some players that are consistent with such use, not confirming such use."(words in brackets have been added to the original.) Moreover, as we hope we may have clarified, it makes more sense to ask whether the "treatment effect" of judicious steroid use on hitters is positive for home run production than to ask questions about the unconditional distribution of home runs.

DeVany correctly states that in order for steroids to increase the number of home runs, players must have used steroids and steroids must have helped those players. The fact that some Major League Baseball (MLB) players have used steroids is not subject to debate. Some have, in fact, admitted to using steroids. The fraction of ballplayers using steroids is the subject of some controversy, but the fact of their use is widely held among the ballplayers themselves. A USA today survey in 2005 of 568 of 700 players showed that 91.9 percent of players believed that some players had used steroids, 4.4 percent did not answer(Jenkins 2005). 79.2 percent of players felt as though steroids played some role in record performances. As Table 3 illustrates, many of baseball's top home run hitters have either tested positive for steroids, admitted to steroid use, or there has been other evidence of steroid use.

It would seem much more difficult to prove that steroid use has helped these players. Mere reference to the time series behavior of home runs per year, for example, is not enough: during the "steroid era" [28] there have been new stadiums built, team expansion, a possible change in the baseball, changes in pitching, etc.. Any of these certainly play a role in the distribution of home runs.

On the other hand, there is some evidence that steroids increase muscle mass (Hartgens and Kuipers 2004). It is certainly possible that increased muscle mass (in combination with other activities) might make it easier for a professional baseball player to hit more home runs.

The closest thing to an experiment is Major Leauge Baseball's more strenuous anti–steroid policy implemented in 2005; the number of home runs did in fact decrease during the first year this policy. (Tainsky and Winfree 2007). The limitations of such a "before and after" research design are clear: just to name one limitation, there is no strong evidence of a "first stage" – reliable data in which ballplayers are using steroids

[25] Alternatively, if the log likelihood of the data generation process is approximately quadratic with a constant Hessian, it can be shown that the maximum likelihood estimator of a quantity is approximately normal.(LeCam 1986, LeCam and Yang 2000, Geyer May 20, 2005).

[26] From the Marx Brothers movie "Duck Soup" (1933). In the movie the words are spoken by the character *Chicolini* played by Chico Marx.

[27] Indeed, Gabaix (1999) states simply that "economic models [for describing the size distribution of cites] have been inadequate. See also Krugman (1996).

[28] It is difficult to define the period of the "steroid era". In 1991, MLB's commissioner's office acknowledged the harmful effects of steroids. However, it is widely believed that some steroid use was prevalent before this. We consider the "steroid era" to have began in the late 1980s or early 1990s, but the lack of a specific date has no implications for our analysis

Table 3: Major League Baseball Players under suspicion of steroids

| Player | Career HR | Led league in HRs | MVP awards | Steroid Use? |
|---|---|---|---|---|
| Barry Bonds* | 734 | 2 | 7 | According to the *San Francisco Chronicle* admitted to previously unknowingly using steroids[1] |
| Sammy Sosa* | 588 | 2 | 1 | Suddenly could not speak English[2] at Congressional Hearings |
| Mark McGwire | 583 | 4 | 0 | "Did not want to talk about the past" at Congressional Hearings |
| Rafael Palmeiro | 569 | 0 | 0 | Tested positive during the 2005 season |
| Jose Canseco | 462 | 2 | 1 | Admitted to steroid use |
| Gary Sheffield* | 455 | 0 | 0 | Admitted to previously unknowingly using steroids |
| Jason Giambi* | 350 | 0 | 1 | Linked to the BALCO investigation; apologized but did not specify for what |
| Ken Caminiti | 239 | 0 | 1 | Admitted to steroid use |

* Denotes current player.
Data obtained from `www.baseballreference.com.`
[1](Williams, Fainaru-Wada and Chronicle Staff Writers 2004)
[2] Press accounts about Sosa's testimony to the U.S. Congress on steroid use were near unanimous in describing his testimony as "non–responsive" also observing that he relied on a lawyer to speak on his behalf citing difficulties with English. See Mariotti (2005) for example.

are obviously not available. Nonetheless, the importance of steroid use can not be ruled out.

One argument we have *not* tried to make is that steroids have been the *sole* cause of the increase in home runs. Home runs have at least as many "causes" as players and (potentially) combinations of ballplayers. Identification of endogenous "social effects" would be quite difficult, at best (Manski 1993). It is possible that the mere *possibility* of engaging in steroid use by some hitters would lead to changes in strategy, and these strategies might have consequences for the size distribution of home runs. On the other hand, we think the evidence is at least "suggestive" that *some* players may have used steroids, and those that may have, it *may* have increased the number of home runs they hit relative to a "no steroid counterfactual" (say, a rigorous drug testing program.)

Consider the following (merely) suggestive evidence: for example, those players who have been identified as potential steroid users have had rates of home run production that, in the main, exceed the best home run hitters of the pre–steroid era. Table 4 shows the top ten seasons for at-bats per home run in both the "pre-steroid era" and "steroid era".

As table 4 documents, the top five seasons with regards to at-bats per home run were during the steroid era; the range of years spanned is only nine years. By comparison, the number of seasons spanned by the pre–steroid error is 42 years.

We note that until the late 1990's when suspicions of steroid use became more common, Babe Ruth remained a complete anomaly. He is less anomalous in the "steroid era." Not only do the top home run hitters need fewer at-bats per home run, but these players achieve this at an older age. Aside from Babe Ruth, there are no top home run hitters over 30 in the "pre-steroid era"; all such players are over 30 during the "steroid era".

Finally, we show pictures of the top four career home run leaders that played in the steroid era, Ken Griffey Jr., Barry Bonds, Mark McGwire, and Sammy Sosa, in Figure 6 to Figure 12. We show pictures of all of them early and late in their career. Of these four, we think most observers would argue that Ken Griffey Jr. is the only one that has not drastically increased their size. Perhaps coincidentally, Griffey is the only one of the four players we consider whose home run production has decreased with age and whose

Table 4: Top Ten Seasons with Fewest At-bats per Home Run, Pre and Post Steroid Era

| Pre Steroid Era | | | | | Steroid Era | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Overall Rank | Player | Age | At-Bat per HR | Year | Overall Rank | Player | Age | At-Bat per HR | Year |
| 6 | Babe Ruth | 25 | 8.46 | 1920 | 1 | Barry Bonds | 36 | 6.52 | 2001 |
| 9 | Babe Ruth | 32 | 9 | 1927 | 2 | Mark McGwire | 34 | 7.27 | 1998 |
| 11 | Babe Ruth | 26 | 9.15 | 1921 | 3 | Mark McGwire | 35 | 8.02 | 1999 |
| 14 | Mickey Mantle | 29 | 9.52 | 1961 | 4 | Mark McGwire | 32 | 8.13 | 1996 |
| 15 | Hank Greenberg | 27 | 9.59 | 1938 | 5 | Barry Bonds | 39 | 8.29 | 2004 |
| 16 | Roger Maris | 26 | 9.67 | 1961 | 7 | Barry Bonds | 38 | 8.67 | 2003 |
| 20 | Babe Ruth | 33 | 9.93 | 1928 | 8 | Barry Bonds | 37 | 8.76 | 2002 |
| 22 | Jimmie Foxx | 24 | 10.09 | 1932 | 10 | Sammy Sosa | 32 | 9.02 | 2001 |
| 23 | Ralph Kiner | 26 | 10.17 | 1949 | 12 | Jim Thome | 31 | 9.23 | 2002 |
| 24 | Mickey Mantle | 24 | 10.25 | 1956 | 13 | Mark McGwire | 33 | 9.31 | 1997 |
| Averages | | | | | | | | | |
| 16 | | 27.2 | 9.58 | 1939.3 | 6.5 | | 34.7 | 8.32 | 2000.3 |

"at bats per home run" statistic failed to decrease as he aged, notwithstanding the fact that Griffey was by far the most prolific hitter before the age of thirty. Up to and including the season he turned thirty, Griffey had 398 home runs compared to 259 for Bonds, 229 for McGwire, and 273 for Sosa. However, in the seasons after, Griffey was the least prolific with 165 home runs compared to 475 for Bonds, 354 for McGwire, and 315 for Sosa (as of the end of the 2006 season). Although injuries may have played a role in this, the same is true if we look at at-bats per home run. Again if we look at the same year comparisons, Griffey went from 14.65 at-bats per home run before he reached thirty years old to 14.95 at-bats per home run in seasons after. This is in comparison with 17.43 to 10.51 for Bonds, 14.00 to 8.42 for McGwire, and 17.08 to 11.86 for Sosa. Clearly many of the players that are under suspicion of steroid use have seen unprecedented offensive statistics at an age when most players are declining in their ability.

# 6 Conclusion

We do not believe that we have made the case that some current or former MLB players benefited from steroid use, although we think it is possible, in principle. Proof of such a claim would take a great deal more work than casual inspection of statistics on home run hitting than we have engaged in here.[29] A higher standard of evidence is needed to establish or refute such a claim.

As we have demonstrated, *none* of the statistical analysis provided in DeVany (2007) speaks to the claim that the causal impact of the judicious use of steroids on home run hitting is zero. Inferring the existence of fundamental causal laws – i.e. the law of genius – from the statistical distribution of some outcome is difficult, at best.

The view that aspects of the human condition or human behavior could be summarized by autonomous statistical laws has a long and not entirely distinguished history. It is ironic, given the aspersions cast on the normal distribution in DeVany (2007), that Galton's explorations into the normal distribution were in part motivated by a quest similar to DeVany's – to explain the "exceptional" and "human genius".[30] Galton worried about breeding mediocrity. Others took the existence of apparently stable (i.e. non–changing)

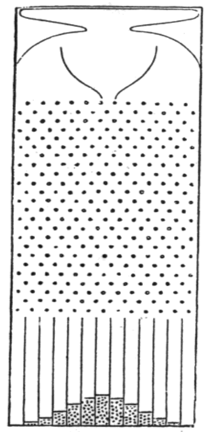
[29] N.B. It is certainly possible that even if some players could benefit from judicious use of steroids, it could also be true that some people would not benefit. In the terminology of the "treatment effect" literature, their might well exist "treatment effect heterogeneity." The effect of steroids on a typical MLB ballplayer might be positive while the effect of steroids on the home run production of the authors of this paper might well be zero or negative.

[30] See the discussion, especially Chapter 21, in Hacking (1990). It is no accident, for example, that one of Galton's most significant efforts was entitled "Hereditary Genius: An Inquiry into its Laws and Consequences" (Galton 1882).

distributions as vitiating free will.[31] Indeed, using different language, Galton (1882) was among the first to use simulation to display an "emergent" system. Galton's famous quincunx – a vertical board with equally spaced pegs and a hole at the top in which marbles could be placed. The marbles entered the top of the device and were allowed to fall randomly[32] to reach the bins at the bottom. A figure from his book (Galton 1894) is displayed in figure 5. The normal distribution that resulted was described as "order out of chaos." [33]

Today, we think of it as a useful mechanical model of the normal distribution as the limiting distribution of the binomial and few would attribute any "deeper" rational for this behavior.

Figure 5: Galton's Quincunx from *Natural Inheritance*



We believe it is fair to say that there have been no convincing evidence of the existence of any causal laws regarding any aspect of the human condition regulated by the normal distribution (or any other distribution) since such ideas were proposed in the nineteenth century. The class of "stable" distributions investigated by DeVany (2007) may prove to be an exception, although we think it quite unlikely. If nonetheless, economists are to take up DeVany's suggestion that the "stable Paretian model developed here will be of use to economists studying extreme accomplishments in other areas" we can only hope such claims will be subject to far more rigorous scrutiny than they have up to this point. Until then, we think it is wise to treat such claims with great skepticism.

[31] See Hacking (1990) and DiNardo (2007) for discussion and citations. Hacking nicely summarizes one example of this view which arose during the nineteenth century: "A problem of statistical fatalism arose. If it were a law that each year so many people must kill themselves in a given region, then apparently the population is not free to refrain from suicide."

[32] A sufficient condition for the marbles to be distributed normally at the bottom of the quincunx is that when a marble arrives at any peg, each marble has the *same* probability of heading left *or* right.

[33] Galton's description of the normal distribution ("Law of Frequency of Error") echoes language used to describe self-organized criticality. From Galton (1894): "*Order in Apparent Chaos* – I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the 'Law of Frequency of Error.' The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever n large sample of chaotic elements are taken in hand and marshaled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along." (Page 66).

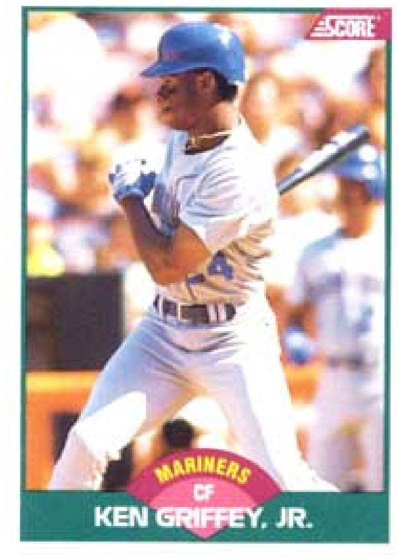



Figure 6: Ken Griffey – Early Career

Figure 7: Ken Griffey – Late career

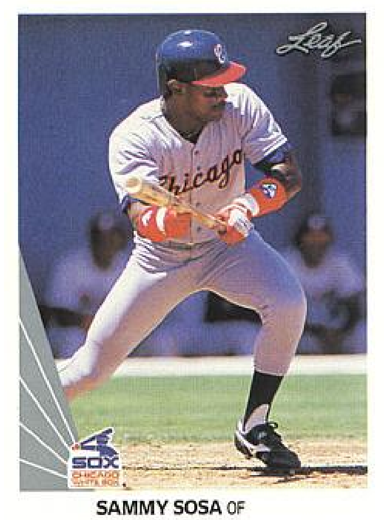



Figure 8: Sammy Sosa– Early Career


Figure 9: Sammy Sosa– Late career

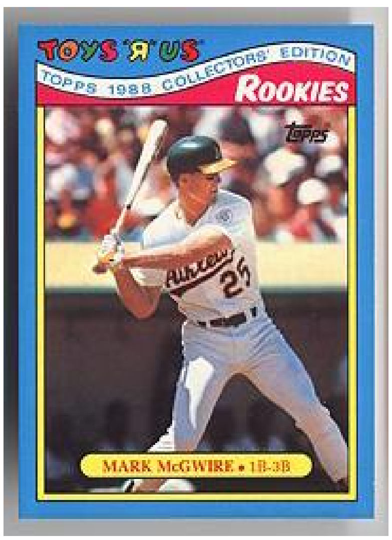



Figure 10: McGwire – Early Career


Figure 11: McGwire – Late career


Figure 12: Barry Bonds – Early and Late career

# References


**Bak, Per**, *How Nature Works: The science of self-organized criticality*, New York: Springer–Verlag, 1996.

——— **and Kan Chen**, "Self–organized Criticality," *Scientific American*, 1991, *264* (1), 26–33.

——— **, Chao Tang, and Kurt Wiesenfeld**, "Self–organized criticality," *Physical Review A*, July 1988, *38* (1), 364–374.

**Bretz, Michael, J.B. Cunningham, P.L. Kurczynski, and F. Nori**, "Imaging of Avalanches in Granular Materials," *Physical Review Letters*, October 1992, *69*, 2431–2434.

**Brock, William A.**, "Scaling in Economics: A Reader's Guide," Manuscript, Department of Economics, University of Wisconsin, Madison March 1999.

**DeVany, Arthur**, "Steroids, Home Runs and the Law of Genius," *Economic Inquiry*, 2007. *Forthcoming.*

**DiNardo, John**, "Interesting Questions in Freakonomics," *Journal of Economic Literature*, 2007, *Forthcoming.*

**Fair, Ray**, "Estimated Age Effects in Baseball," Unpublished Manuscript, Department of Economics, Yale University, `http://fairmodel.econ.yale.edu/rayfair/pdf/2005D.PDF` 2007.

**Feller, William**, "On the Logistic Law of Growth and Its Empirical Verifications in Biology," *Acta Biotheoretca*, 1940, *5*, 51–66.

**Freedman, David A.**, "From Association to Causation: Some Remarks on the History of Statistics," *Statistical Science*, 1999, *14*, 243–258.

**Gabaix, Xavier**, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, 1999, *114*, 739–767.

**Galton, Francis**, *Hereditary Genius: an Inquiry into its Laws and Consequences*, second ed., Macmillan, 1882. Third corrected proof of the first electronic edition, 2000. Based on the text of the second edition. Retrieved electronically from `http://galton.org/books/hereditary-genius/` 28 May 2007. Web site maintained by Gavan Tredoux.

——— , *Natural Inheritance*, Macmillan and Company, 1894.

**Geyer, Charles J.**, "Le Cam Made Simple: Asymptotics of Maximum Likelihood without the LLN or CLT or Sample Size Going to Infinity," Technical Report 643, School of Statistics, University of Minnesota, Minneapolis, MN 20 May 2005 May 20, 2005. Revised.

**Gnedenko, Boris Vladimirovich**, "Sur la distribution limite du terme maximum d'une serie aleatoire," *Annals of Mathematics (Second Series)*, July 1943, *44* (3), 423–453.

——— **and Andrey Nikolaevich Kolmogorov**, *Limit distributions for sums of independent random variables*, New York: Addison-Wesley, 1954.

**Good, Irving John**, "An Error By Peirce Concerning Weight of Evidence," *Journal of Statistical Computation and Simulation*, 1981, *13*, 155–157.

**Gray, H. L. and P. L. Odell**, "On Least Favorable Density Functions," *SIAM Review*, October 1967, *9* (4), 715–720.

**Hacking, Ian**, *The Logic of Statistical Inference*, Cambridge: Cambridge University Press, 1965.

——— , *The Taming of Chance* number 17. In 'Ideas in Context.', Cambridge, England: Cambridge University Press, August 1990.

**Hartgens, F. and H. Kuipers**, "Effects of androgenic–anabolic steroids in athletes," *Sports Medicine*, 2004, *34*, 513–554.

**Heyde, C. C. and S. G. Kou**, "On the controversy over tailweight of distributions," *Operations Research Letters*, 2004, *32*, 399–408.

**Jacobson, Harold I.**, "The Maximum Variance of Restricted Unimodal Distributions," *The Annals of Mathematical Statistics*, October 1969, *40* (5), 1746–1752.

**Jenkins, C.**, "Players Admit Steroids Changed Baseball," *USA Today*, March 15 2005. Retrieved from `http://www.usatoday.com/sports/baseball/2005-03-15-steroids-mlb-cover_x.htm` on 28 May 2007. Article updated 17 March 2005.

**Keller, Evelyn Fox**, "Revisiting "scale-free" networks.," *Bioessays*, Oct 2005, *27* (10), 1060–1068.

**Krugman, Paul**, *The Self–Organizing Economy*, Cambridge, MA: Blackwell, 1996.

**LeCam, Lucien**, *Asymptotic Methods in Statistical Decision Theory*, New York: Springer–Verlab, 1986.

___ **and Grace L. Yang**, *Asymptotics in Statistics: Some Basic Concepts*, second ed., New York: Springer–Verlag, 2000.

**Li, Wentian**, "Random Texts Exhibit Zipf's–Law–Like Word Frequency Distribution," *IEEE Transactions on Information Theory*, 1992, *38*, 1842–1845.

**Manski, Charles F.**, "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, July 1993, *60* (3), 531–542.

**Mariotti, Jay**, "Injury or not, Sosa Can't Hide From the Truth," *Chicago Sun–Times*, May 11 2005.

**Mayo, Deborah G.**, *Error and the Growth of Experimental Knowledge* Science and Its Conceptual Foundations, Chicago: University of Chicago Press, 1996.

**Miller, George A.**, "Introduction," in "The Psycho–biology of Language: An Introduction to Dynamic Philology," Cambridge, MA: MIT Press, 1965.

___ **and Noam Chomsky**, "Finitary Models of Language Users," in R.D. Luce, Robert R. Bush, and Eugene Galanter, eds., *Handbook of Mathematical Psychology*, Vol. 2, New York and London: Wiley and Sons, 1963, pp. 419–491.

**Mitzenmacher, Michael**, "Editorial: The Future of Power Law Research," *Internet Mathematics*, 2006, *2* (4), 525–534.

**Muilwijk, J.**, "Note on a theorem of M. N. Murthy and V. K. Sethi," *Sankhya, Series B*, 1966, *28*, 183. Part 1, 2.

**Nagel, Sidney R.**, "Instabilities in a sandpile," *Reviews of Modern Physics*, Jan 1992, *64* (1), 321–325.

**Newman, M. E. J.**, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, September 2005, *46* (5), 321–353.

**Nolan, J. P.**, *Stable Distributions - Models for Heavy Tailed Data*, Boston: Birkhäuser, 2007. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

**Peirce, Charles Sanders**, "The Probability of Induction," *Popular Science Monthly*, 1878, *12*, 705–718.

**Rimmer, R. H. and J. P. Nolan**, "Stable Distributions in Mathematica," *Mathematica Journal*, 2005, *9*, 776–789.

**Schelling, Thomas C.**, "Models of Segregation," *American Economic Review*, May 1969, *59* (2), 488–493.

———, *Micromotives and Macrobehavior*, New York: W. W. Norton & Company, 1978.

**Solow, Andrew R., Christopher J. Costello, and Michael Ward**, "Testing the Power Law Model for Discrete Size Data," *American Naturalist*, November 2003, *162* (5), 685–689.

**Tainsky, Scott and Jason A. Winfree**, "Financial Incentives and League Policy: The Example of Major League Baseball's Steroid Policy," Unpublished Manuscript, University of Michigan, School of Kinesiology, Ann Arbor, MI 2007.

**Tesfatsion, Leight**, "Introductory Notes on Complex Adaptive Systems and Agent-Based Computational Economics," Technical Report, Department of Economics, Iowa State University January 2007. `http://www.econ.iastate.edu/classes/econ308/tesfatsion/bat1a.htm`.

**Williams, Lance, Mark Fainaru-Wada, and Chronicle Staff Writers**, "What Bonds told BALCO grand jury.," *San Francisco Chronicle*, December 3 2004. Retrieved from `http://www.sfgate.com/cgi-bin/article.cgi?file=/c/a/2004/12/03/MNGGFA0UDU65.DTL` on 28 May 2007.

**Willinger, W., D. Alderson, J.C. Doyle, and L. Li**, "More "normal" than normal: scaling distributions and complex systems," in "Simulation Conference, 2004. Proceedings of the 2004 Winter," Vol. 1 5-8 Dec. 2004.

**Winslow, Nathan**, "Introduction to Self-Organized Criticality and Earthquakes," Discussion Paper, Department of Geological Sciences, University of Michigan 1997.

**Yule, G. Udny**, "A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis," *Philosophical Transactions of the Royal Society of London, Series B, Containing Papers of a Biological Character*, 1925, *213*, 21–87.

**Yule, G. Undy**, "An Investigation into the Causes of Changes in Pauperism in England, Chiefly During the Last Two Intercensal Decades (Part I.)," *Journal of the Royal Statistical Society*, June 1899, *62* (2), 249–295.

**Zipf, George Kingsley**, *Selective Studies and the Principle of Relative Frequency in Language*, Cambridge, MA: MIT Press, 1932.