*Beyond the asterisk **
Adjusting for performance inflation in professional sports

SABERMETRICS, SCOUTING
AND THE SCIENCE OF BASEBALL

A weekend baseball seminar and benefit for the Jimmy Fund

[August 4-5, 2012]

Alexander M. Petersen
IMT Lucca
Lucca, Italy

Bridging the past and the present

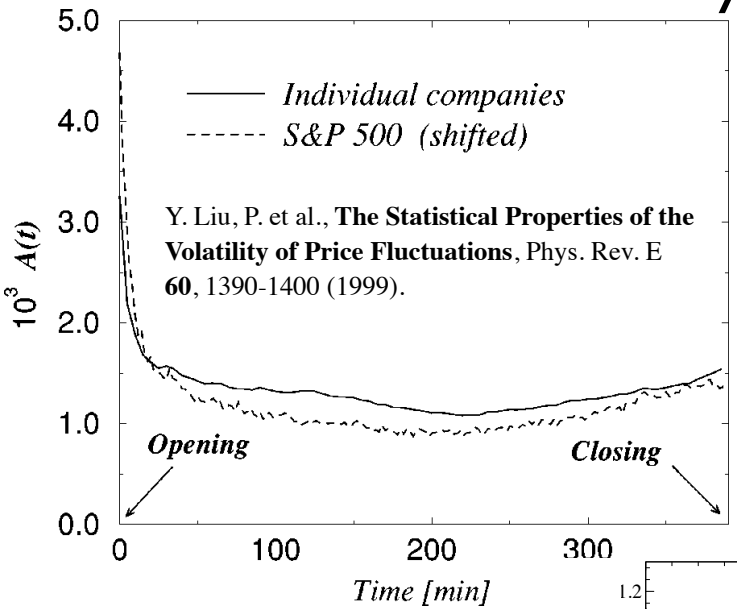
1. Method for "deflating" achievement metrics

3. Re-ranking The All-Time Greats

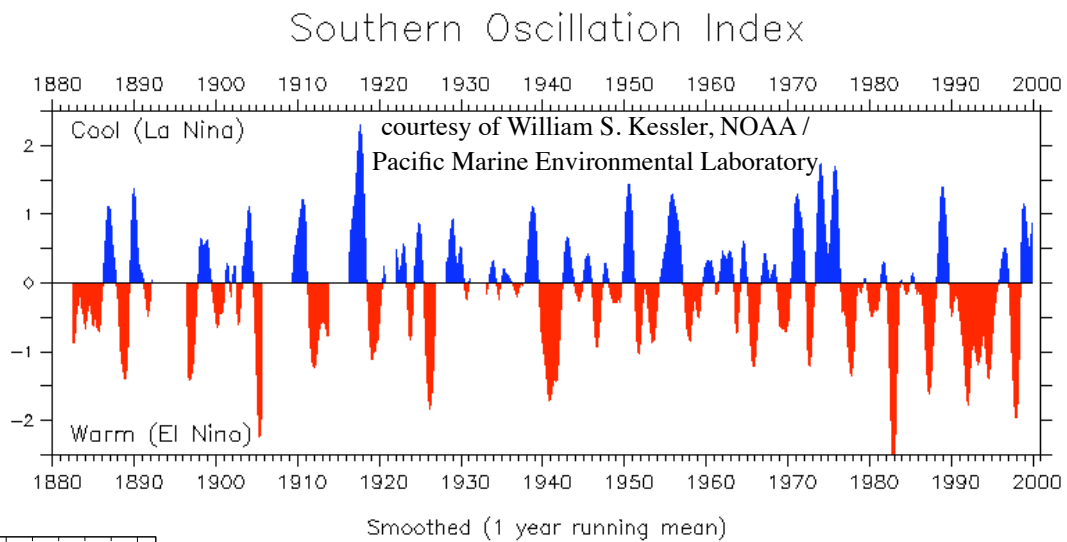
2. The Statistical Physics of Achievement

1. Establishing a baseline by removing trends

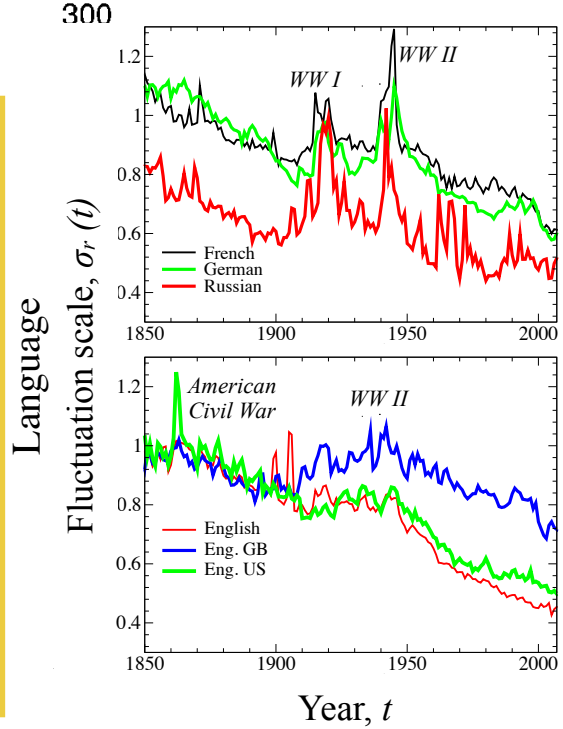
Financial Market Activity



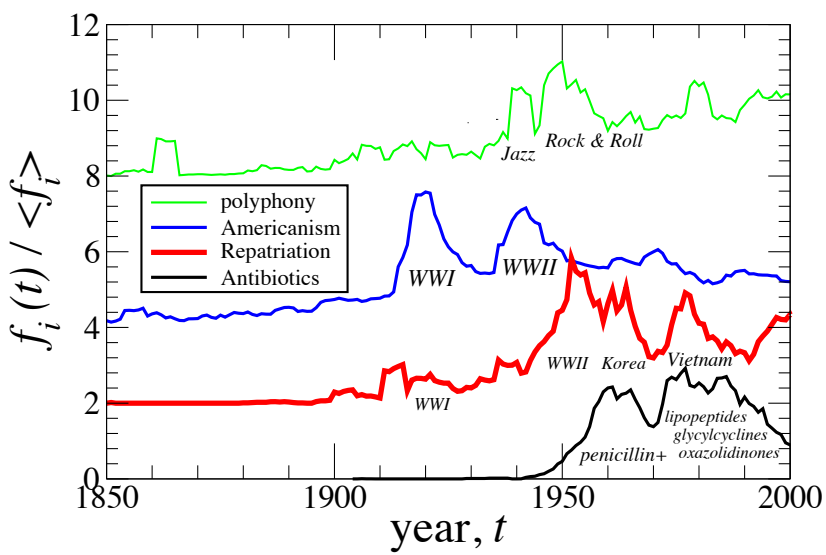
El Niño and La Nina



Time series in social and natural phenomena are typically non-stationary: there are underlying exogenous and endogenous factors that can significantly fluctuate!

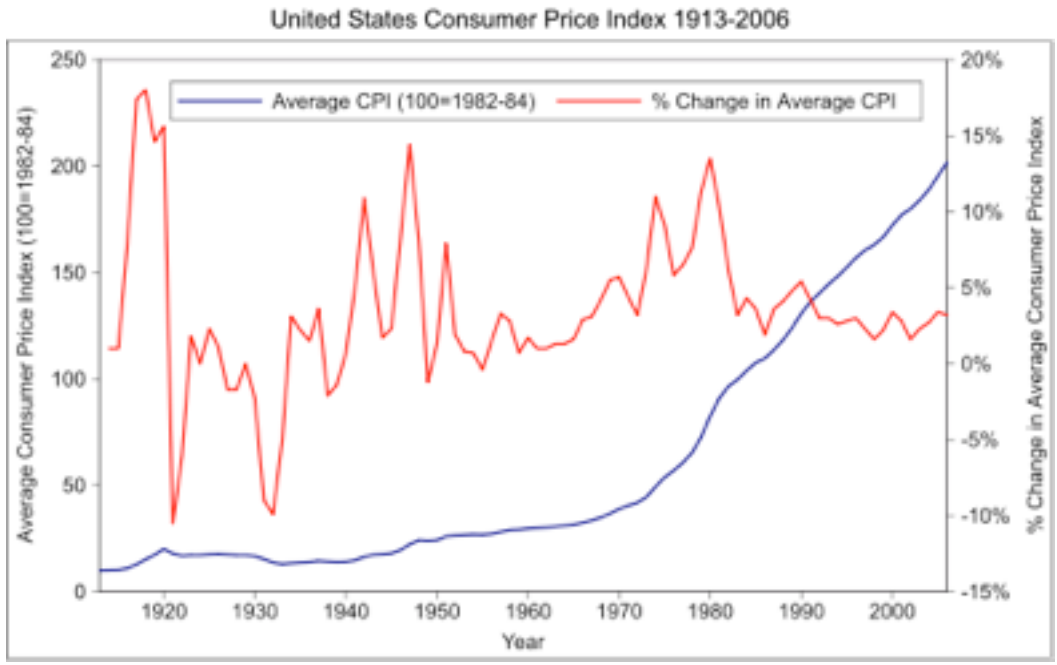


Culture



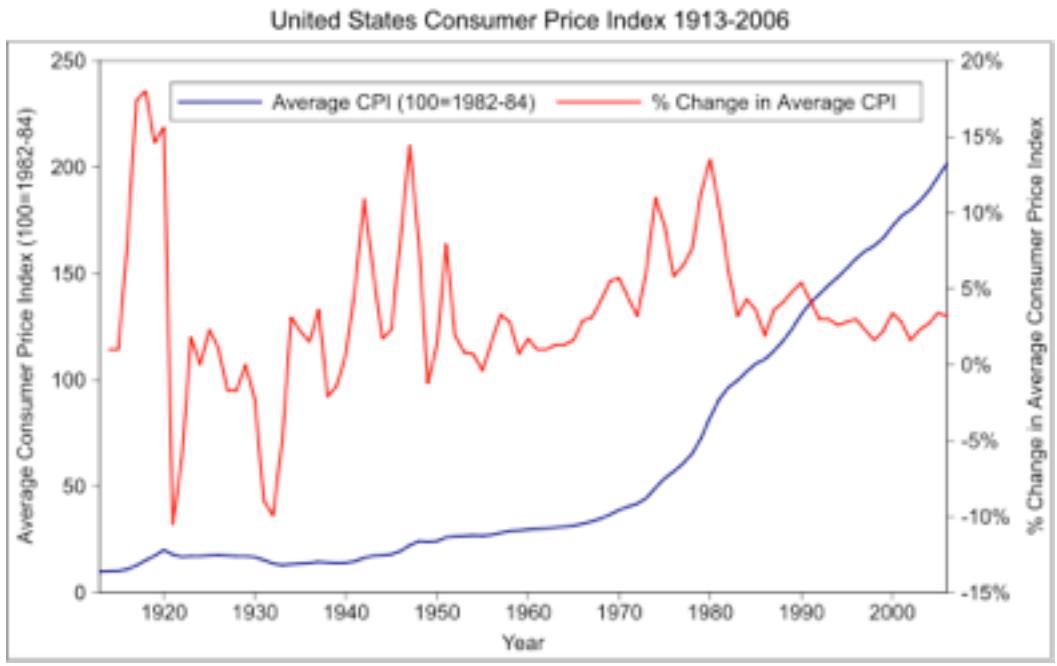
A. M. Petersen, J. Tenenbaum, S. Havlin, H. E. Stanley.
Statistical Laws Governing Fluctuations in Word Use from Word Birth to Word Death.
Scientific Reports 2, 313 (2012).

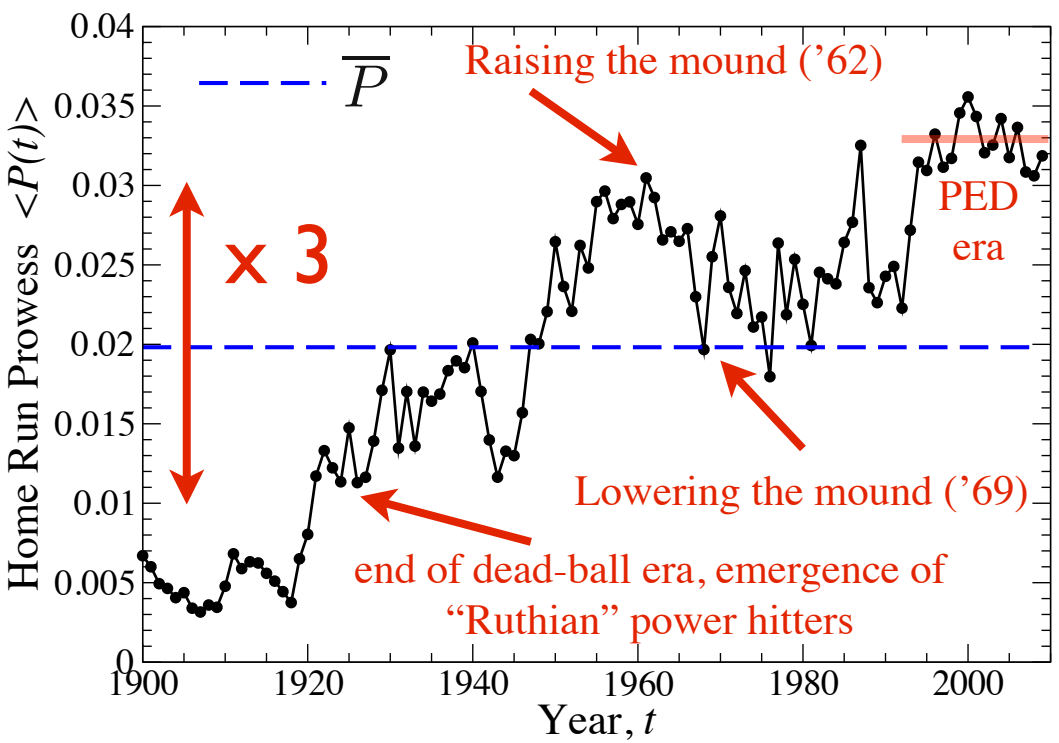
Accounting for Inflation



Just as the price of a candy bar has increased by a factor of ~ 20 over the last 100 years (roughly 3% inflation rate),

Accounting for Inflation



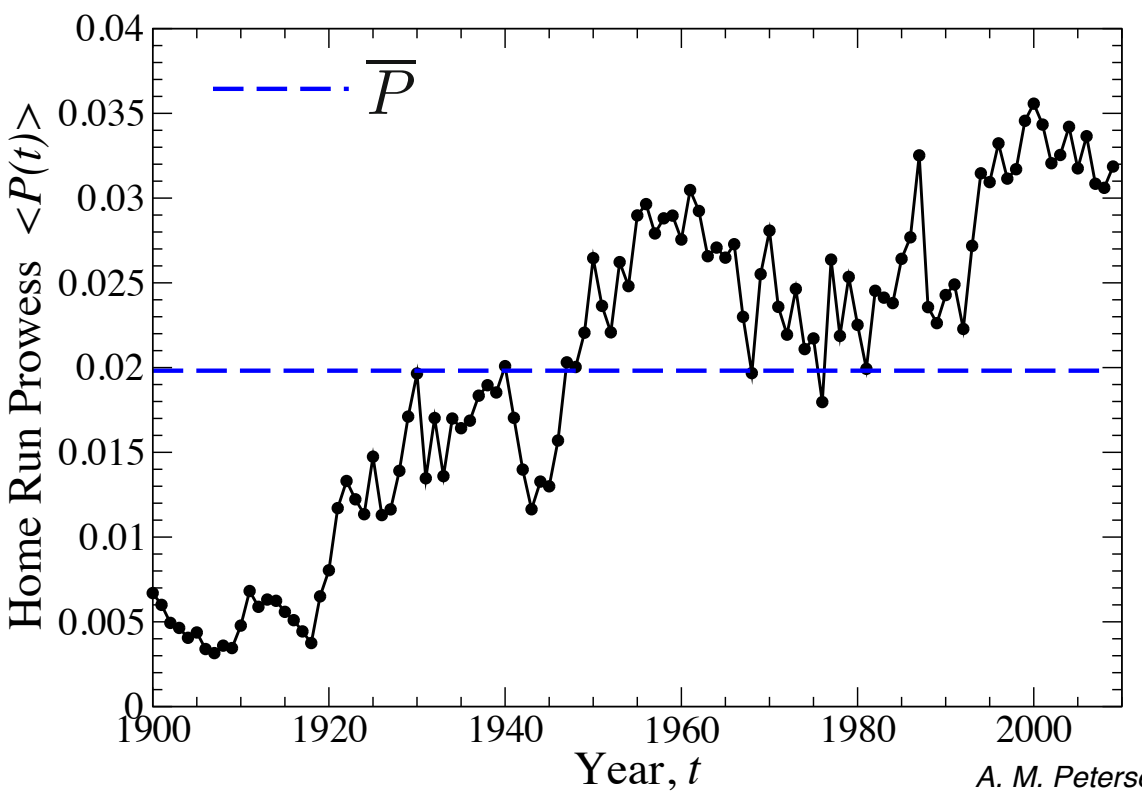


Just as the price of a candy bar has increased by a factor of ~ 20 over the last 100 years (roughly 3% inflation rate), the home run hitting ability of players has also increased by a significant factor over the same period

Detrending method

Time-dependent economic, technological, and social factors can artificially inflate or deflate quantitative measures for single season and career achievement.

$$x_i(t) = \#\ of\ successes$$
$$y_i(t) = \#\ of\ opportunities$$
$$P_i(t) = x/y = success\ rate$$



We first calculate the prowess $P_i(t)$ of an individual player i as

$$P_i(t) \equiv x_i(t)/y_i(t) \ ,$$

where $x_i(t)$ is an individual's total number of successes out of his/her total number of opportunities $y_i(t)$ in a given year t. To compute the league-wide average prowess, we then compute the weighted average for season t over all players

$$\langle P(t) \rangle \equiv \frac{\sum_i x_i(t)}{\sum_i y_i(t)} = \sum_i w_i(t) P_i(t) \ ,$$

where

$$w_i(t) = \frac{y_i(t)}{\sum_i y_i(t)} \ .$$

The index i runs over all players with at least y' opportunities during year t, and $\sum_i y_i$ is the total number of opportunities of all $N(t)$ players during year t. We use a cutoff $y' \equiv 100$ which eliminates statistical fluctuations that arise from players with very short seasons.

We now introduce the detrended metric for the accomplishment of player i in year t,

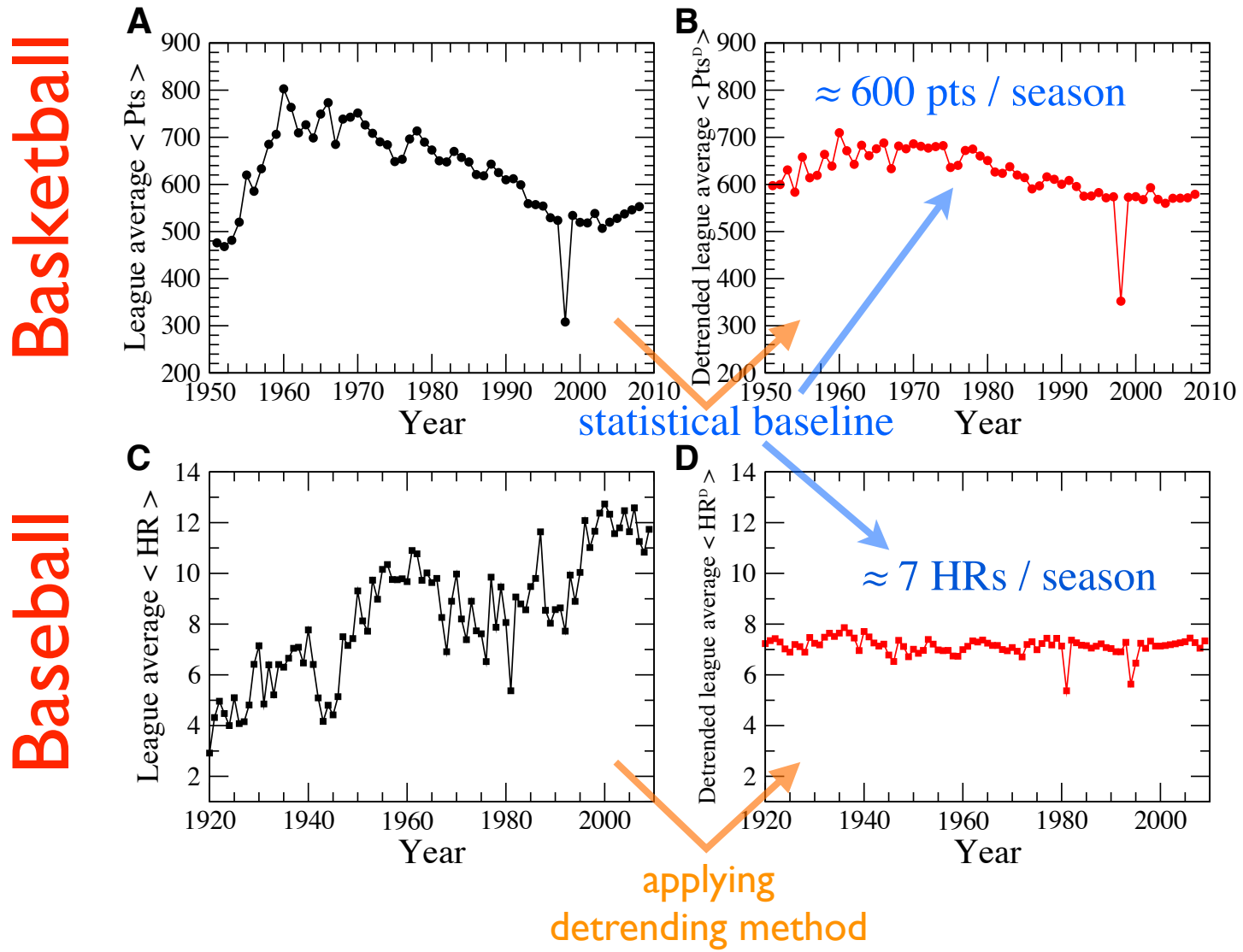
$$x_i^D(t) \equiv x_i(t) \ \frac{\overline{P}}{\langle P(t) \rangle}$$

where $\overline{P}$ is the average of $\langle P(t) \rangle$ over the entire period,

$$\overline{P} \equiv \frac{1}{110} \sum_{t=1900}^{2009} \langle P(t) \rangle \ .$$

A. M. Petersen, O. Penner, H. E. Stanley , "'Methods for detrending success metrics to account for inflationary and deflationary factors." Eur. Phys. J. B 79, 67-78 (2011).

Accounting for socio-technological factors that underly achievement



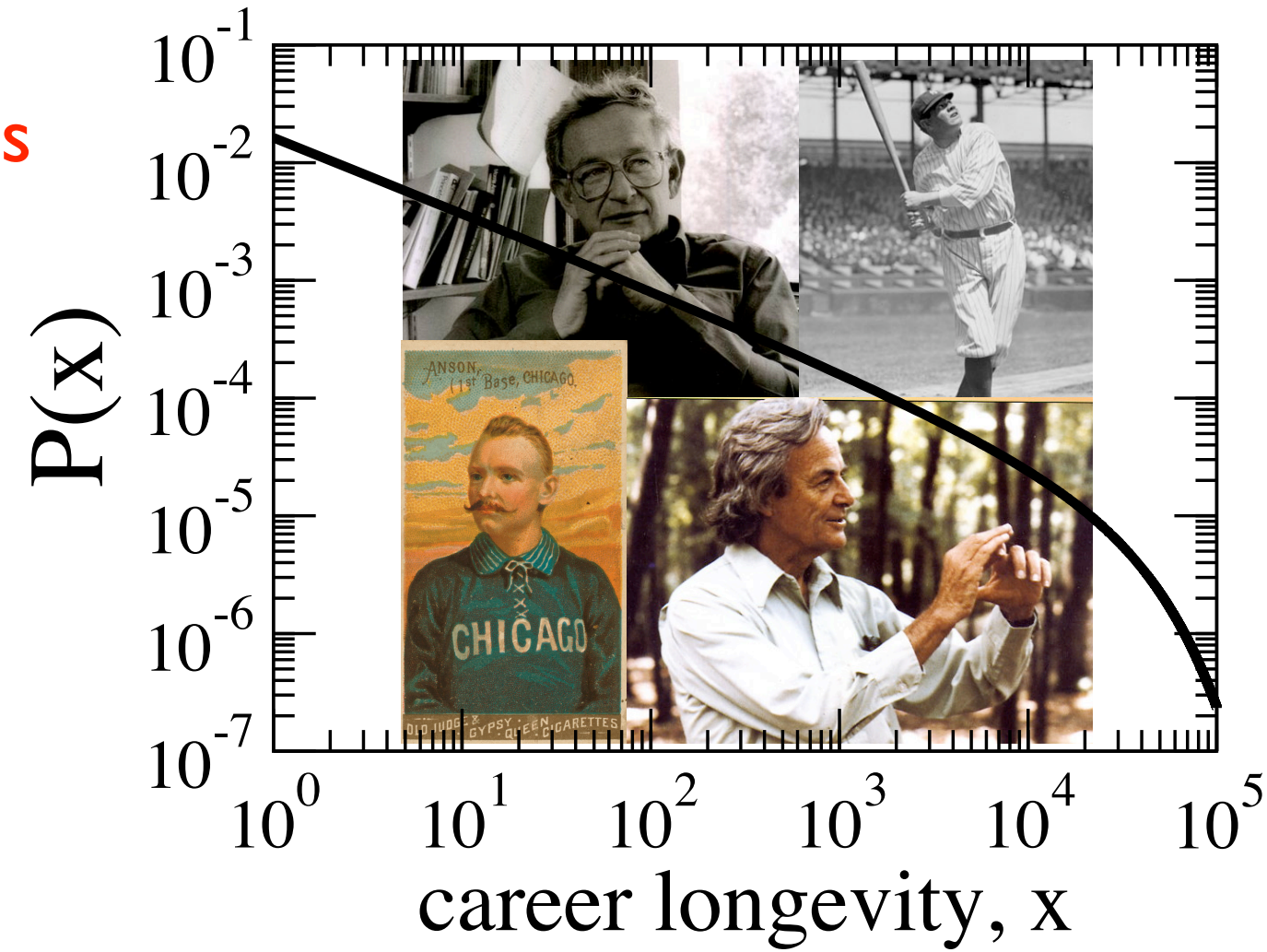
Quantitative measures for success are important for comparing both individual and group accomplishments, often achieved in different time periods.

However, the evolutionary nature of competition results in a non-stationary rate of success, that makes comparing accomplishments across time statistically biased.

While there is much speculation and controversy surrounding the causes for changes in player ability, we do not address these individually. In essence, we blindly account for not only the role of PED, but also changes in the physical construction of bats and balls, sizes of ballparks, talent dilution of players from expansion, etc.

2. the "Socio-physics" of Careers

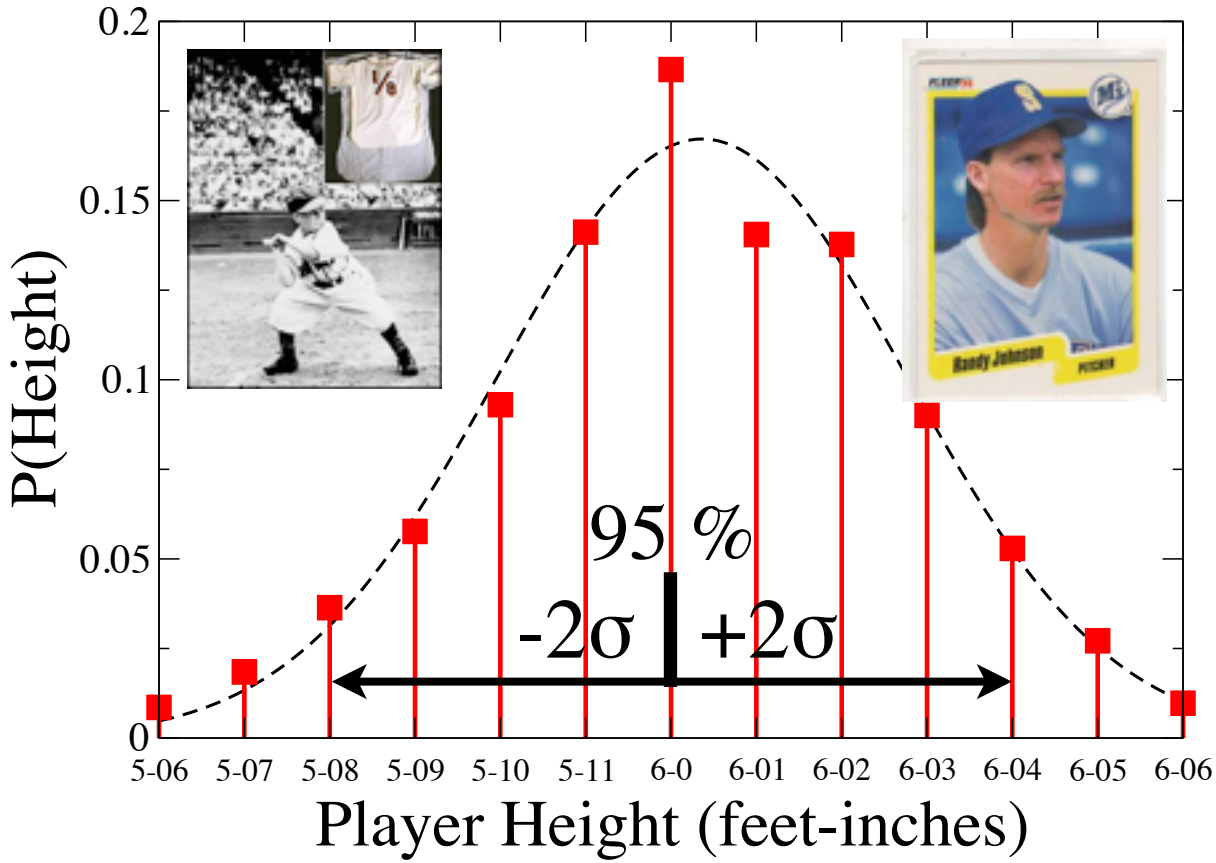
Statistical Physics approach to understanding longevity and success in competition driven systems



Not surprisingly, player height is governed by a standard bell-shaped distribution

The ratio of the tallest baseball player (Jon Rauch, 6 feet 11 inches) to the shortest baseball player (Eddie Gaedel, 3 feet 7 inches) is roughly 2.

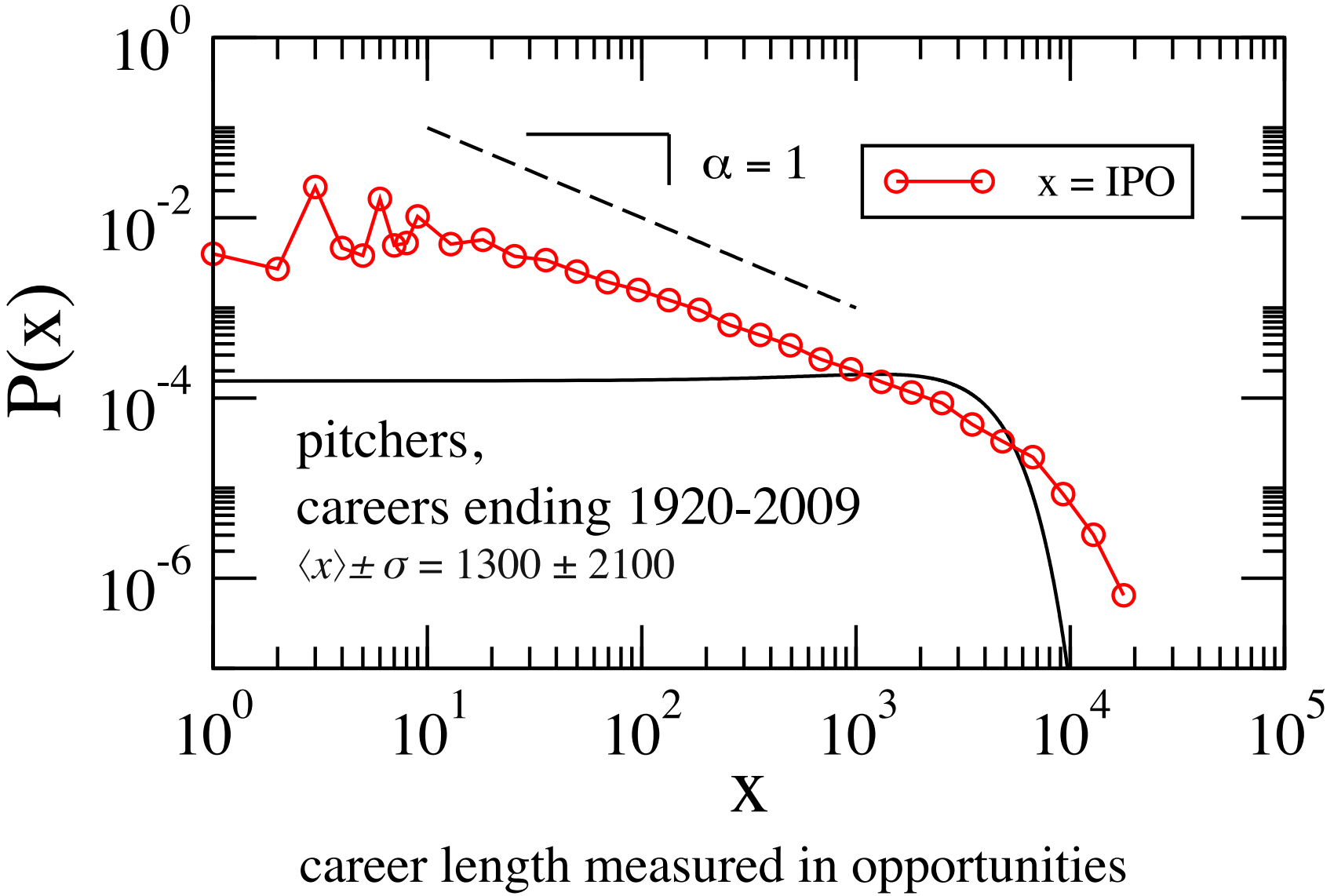
The relatively small value of this height ratio follows from the properties of the Gaussian distribution, which is well-suited for the description of height in a human population.



A demonstration of a probability density function that has a characteristic scale. The pdf of Major League Baseball player height. The data are fit well by a Gaussian "bell-curve" pdf (dashed line) with an average height of 6.0 feet ±2 inches. Data courtesy of baseball-almanac.com, accessed at:
http://www.baseball-almanac.com/charts/heights/heights.shtml

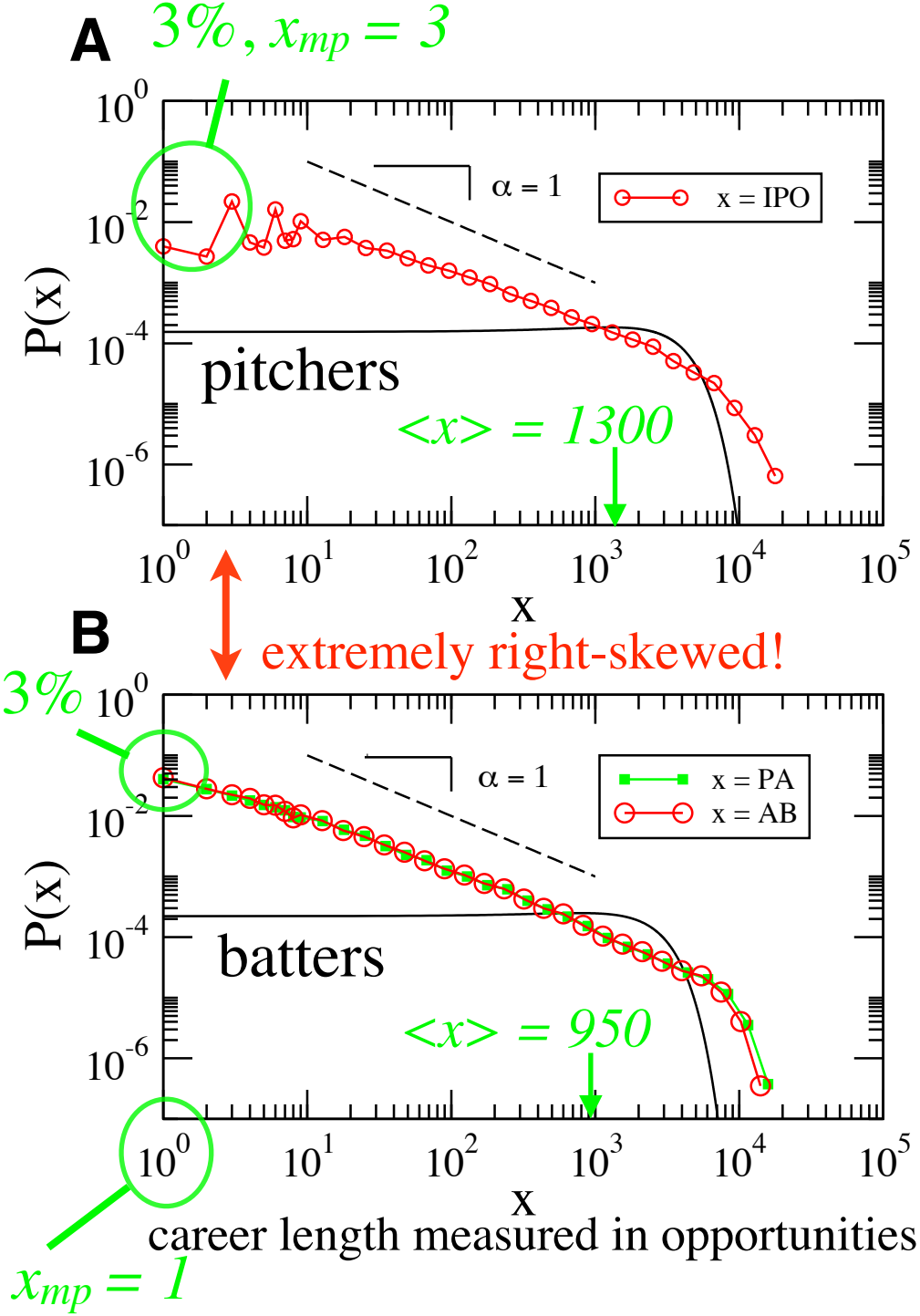
.... but how about the longest and shortest MLB career?

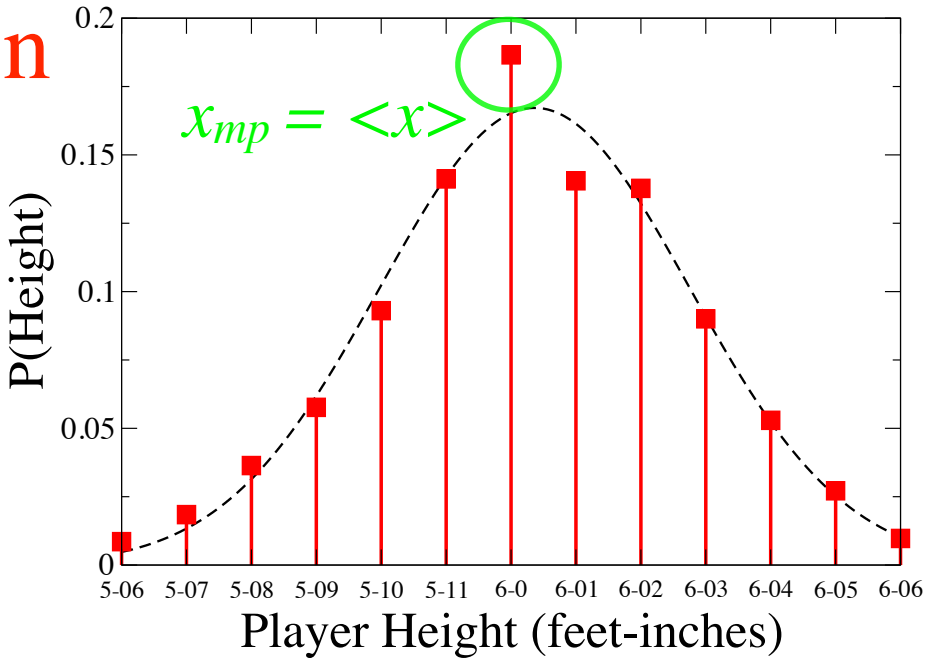
Career longevity distribution



career length measured in opportunities

Career longevity distribution



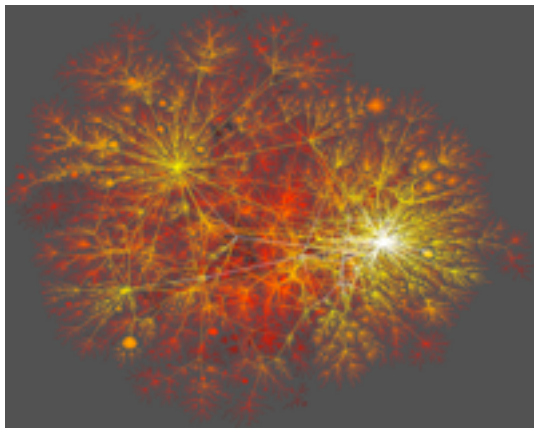


In order to emphasize the disparity between the long and short careers, consider the ratio of the longest career (Pete Rose, 14,053 at-bats) to the shortest career (many individuals with one at-bat), which is roughly 1x 10^4. For comparison, the ratio of the tallest baseball player (Jon Rauch, 6 feet 11 inches) to the shortest baseball player (Eddie Gaedel, 3 feet 7 inches) is roughly 2.

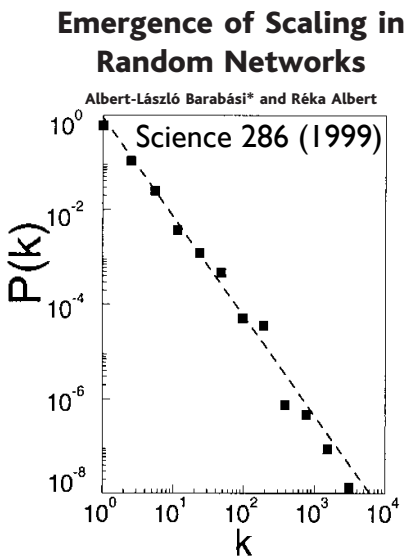
the probability density function (pdf) $P(x)$ is defined so that the probability of observing an event in the interval $(x, x + \delta x)$ is $P(x)\delta x$.

Heavy-tailed distributions in social and physical phenomena

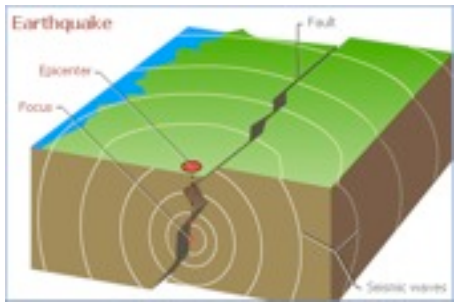
Complex networks

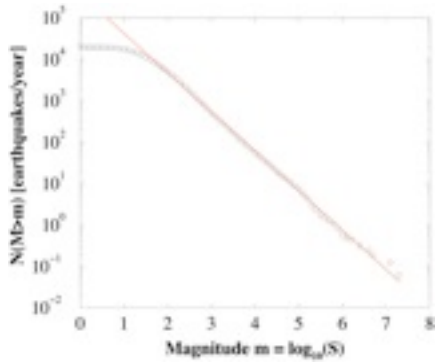


Snapshot of Internet network
courtesy k.c. claffy

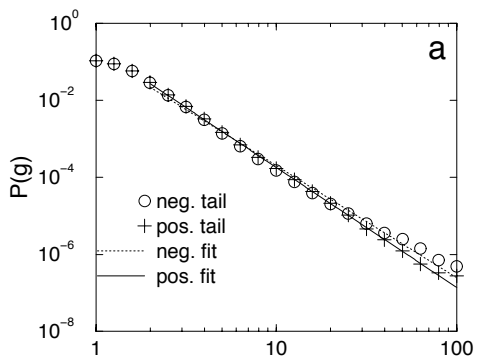


Geophysical and Financial Shocks





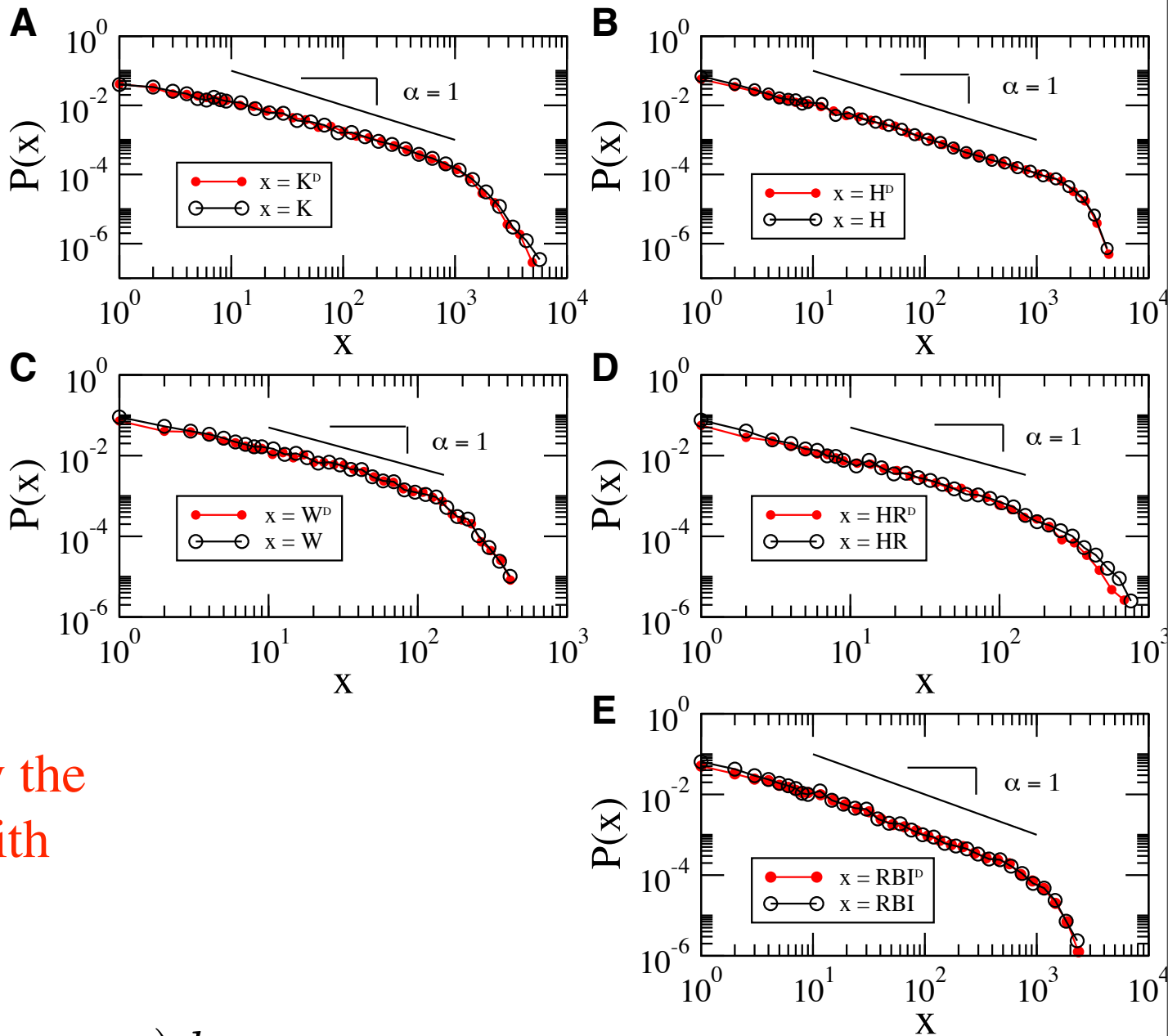
Unified scaling law for earthquakes, K. Christensen et al., PNAS 99 (2002)



Inverse cubic law for the distribution of stock price variations, P. Gopikrishnan et al., EPJB 3 (1998)

- An key feature of extremely skewed $P(x)$ distributions (i.e., scale-free power law $P(x) \sim 1 / x^{\alpha}$), is the large disparity between the most probable value and the mean/ median value of the distribution: $\Rightarrow$ the most probable value $x_{mp} = Min(x)$, and the mean value $\langle x \rangle >> x_{mp}$.
- This is in stark contrast to the Gaussian (Normal) distribution pdf for which the mean value and the most probable value coincide, $x_{mp} = \langle x \rangle$.
- For Baseball, the approximate power-law behavior can be roughly phrased as such: For every Mickey Mantle (~8000 career at-bats), there are roughly 10 players with careers similar to Doc "the Punk" Gautreaus (~800 career at-bats); and for every Doc "the Punk" Gautreau there are roughly 10 players with careers similar to Frank "the Jelly" Jelincich (8 career at-bats with one hit!). This statistical property arises from the ratio of frequencies $P(x)/P(y) \sim (y / x)^{\alpha} = (y / x)$ for exponent $\alpha \approx 1$

Career success distributions

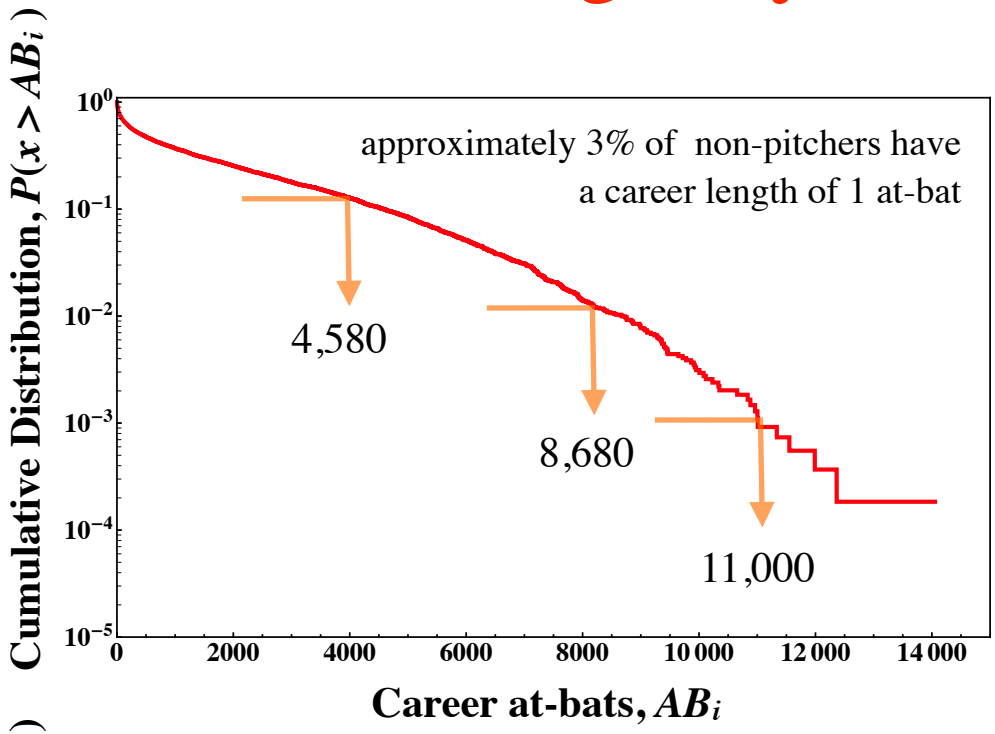


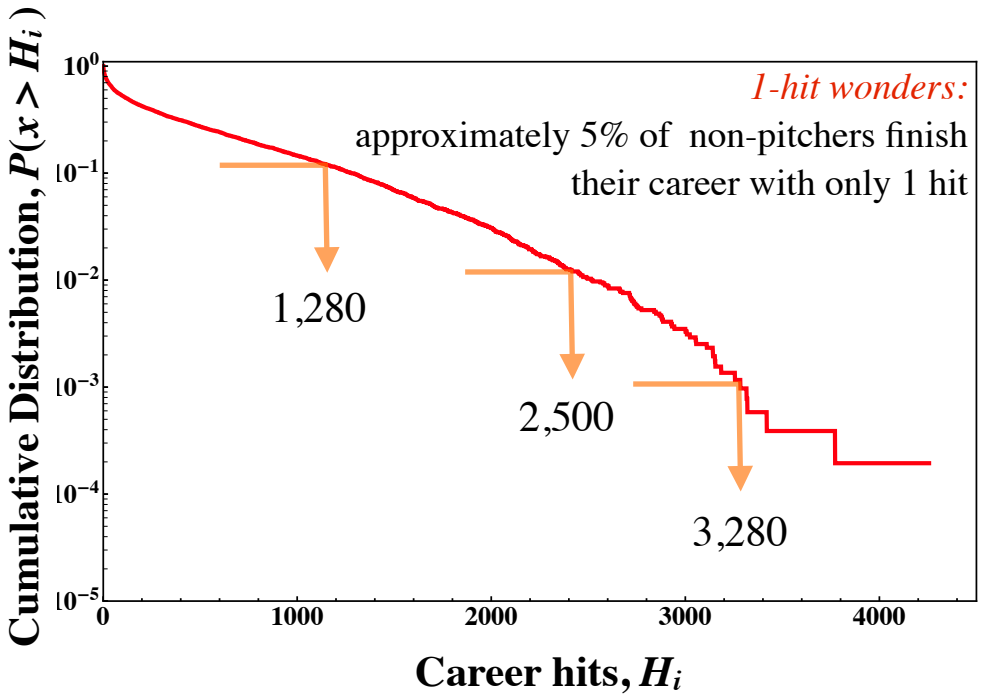
can be modeled well by the Gamma distribution with scaling exponent $\alpha < 1$

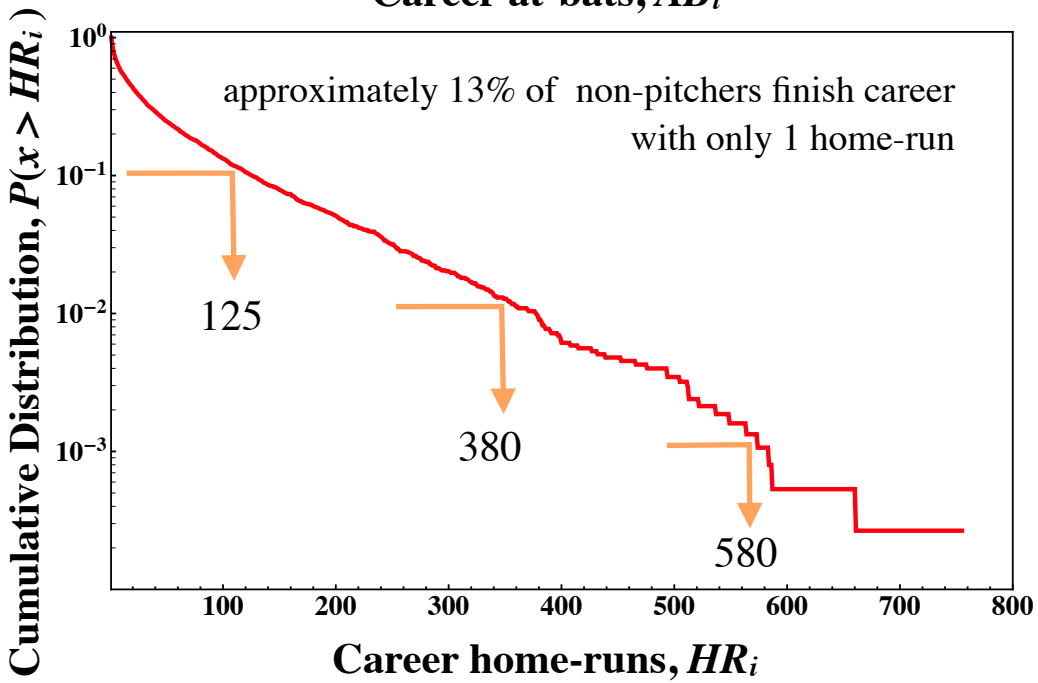
$$P(x)dx \approx \text{Gamma}(x;\alpha,x_c)dx = \frac{(x/x_c)^{-\alpha}e^{-x/x_c}}{\Gamma(1-\alpha)}\frac{dx}{x_c} \propto x^{-\alpha}\ e^{-x/x_c}\ ,$$

and with a scale factor x_c representing a "tipping point" in the career, which distinguishes the veterans from the newcomers

Longevity underlies career totals





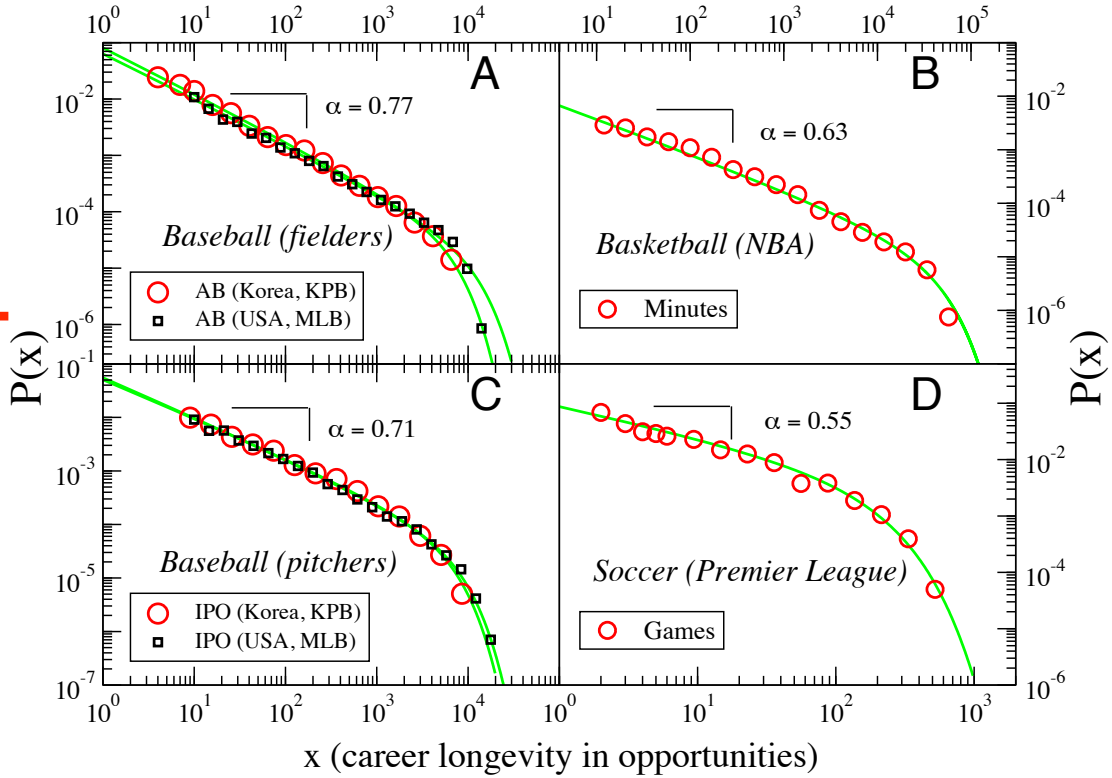


Cap Anson
10,277 at-bats
27 seasons
3,418 hits

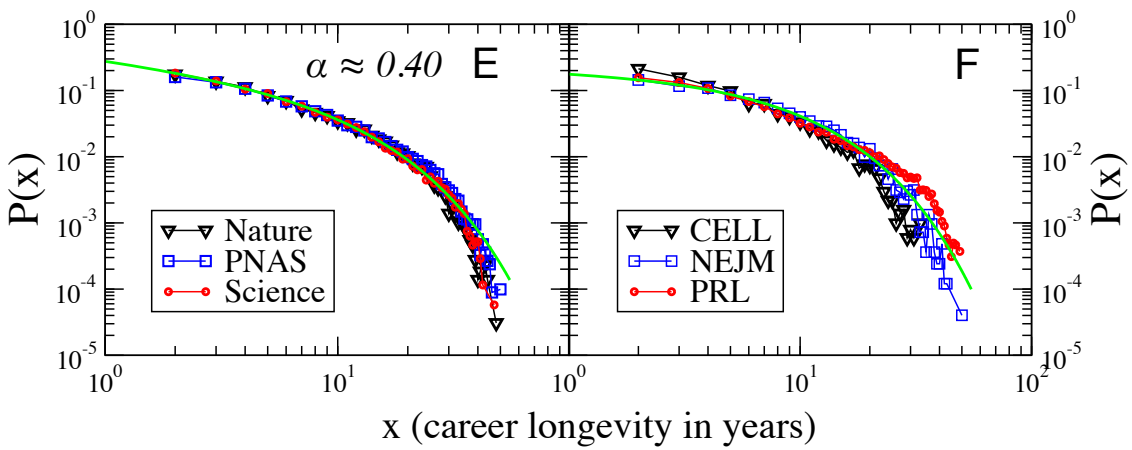
x = career success total
typically proportional to the career length

Statistical law for career longevity

Pro Sports



Academia



opportunities ~ time duration

A. M. Petersen, W.-S. Jung, J.-S. Yang, H. E. Stanley, "Quantitative and empirical demonstration of the Matthew effect in a study of career longevity." Proc. Natl. Acad. Sci. USA 108, 18-23 (2011).

Major League Baseball

- 130+ years of player statistics, ~ 15,000 careers

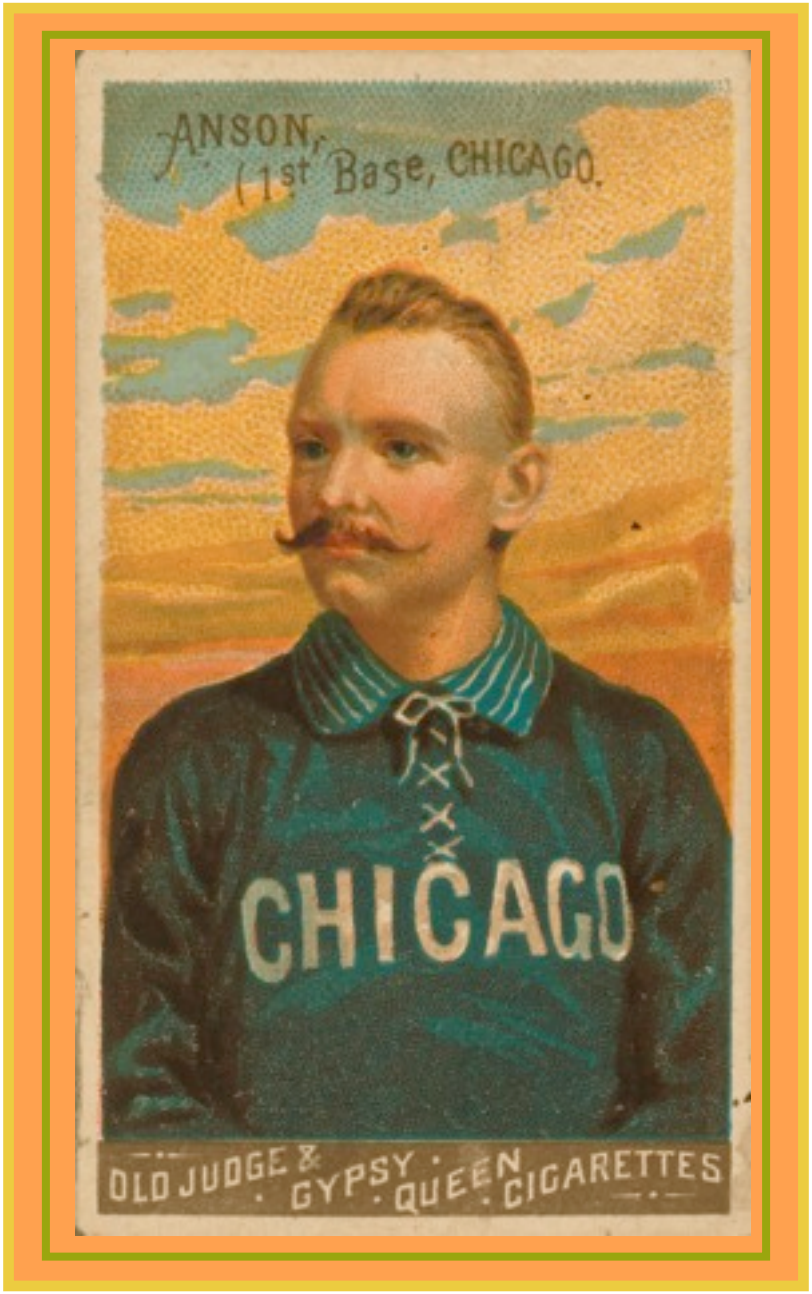
``One-hit wonders"

- 3% of all fielders finish their career with ONE at-bat!
- 3% of all pitchers finish their career with less than one inning pitched!

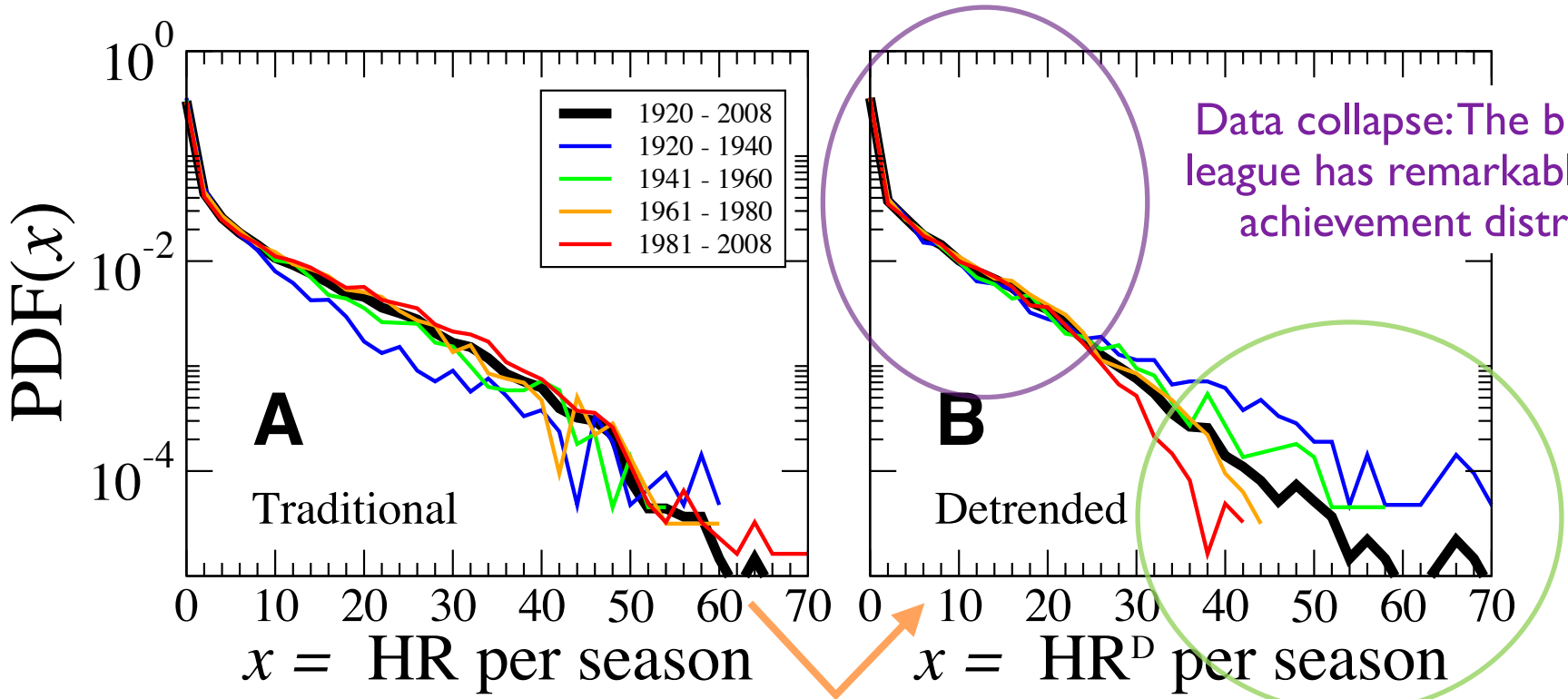
``Iron horses"

- Lou Gehrig (the Iron Horse): NY Yankees (1923-1939)
- Played in 2,130 consecutive games in 15 seasons! 8001 career at-bats!
- Career & life stunted by the fatal neuromuscular disease, amyotrophic lateral sclerosis (ALS), aka Lou Gehrig's Disease

3. Re-ranking the all-time greats



Single season success distributions



Data collapse: The bulk of the league has remarkably uniform achievement distribution

applying detrending method

But in comparison to the bulk, the extremely rare achievements gain new perspective lighting

Detrending amplifies relatively significant achievements using the local ability average as a baseline

Career Wins: not affected by detrending

Table S5. Ranking of career wins (1890–2009).

Traditional Rank						Detrended Rank		
Rank	Name	Final Season (L)	Career Metric	Rank*(Rank)	% Change	Name	Final Season (L)	Career Metric
1	Cy Young	1911 (22)	511	1(1)	0	Cy Young	1911 (22)	510
2	Walter Johnson	1927 (21)	417	2(2)	0	Walter Johnson	1927 (21)	420
3	Christy Mathewson	1916 (17)	373	3(3)	0	Christy Mathewson	1916 (17)	376
3	Pete Alexander	1930 (20)	373	4(3)	-33	Pete Alexander	1930 (20)	375
5	Pud Galvin	1892 (15)	364	5(5)	0	Pud Galvin	1892 (15)	365
6	Warren Spahn	1965 (21)	363	6(6)	0	Warren Spahn	1965 (21)	362
7	Kid Nichols	1906 (15)	361	7(7)	0	Kid Nichols	1906 (15)	359
8	Greg Maddux	2008 (23)	355	8(8)	0	Greg Maddux	2008 (23)	351
9	Roger Clemens	2007 (24)	354	9(9)	0	Roger Clemens	2007 (24)	350
10	Tim Keefe	1893 (14)	342	10(10)	0	Tim Keefe	1893 (14)	342
11	Steve Carlton	1988 (24)	329	11(11)	0	Steve Carlton	1988 (24)	329
12	John Clarkson	1894 (12)	328	12(13)	7	Eddie Plank	1917 (17)	328
13	Eddie Plank	1917 (17)	326	13(12)	-8	John Clarkson	1894 (12)	327
14	Don Sutton	1988 (23)	324	14(14)	0	Don Sutton	1988 (23)	324
14	Nolan Ryan	1993 (27)	324	14(14)	0	Nolan Ryan	1993 (27)	324
16	Phil Niekro	1987 (24)	318	16(16)	0	Phil Niekro	1987 (24)	318
17	Gaylord Perry	1983 (22)	314	17(17)	0	Gaylord Perry	1983 (22)	314
18	Tom Seaver	1986 (20)	311	18(18)	0	Tom Seaver	1986 (20)	311
19	Charley Radbourn	1891 (11)	309	19(19)	0	Charley Radbourn	1891 (11)	308
20	Mickey Welch	1892 (13)	307	20(20)	0	Mickey Welch	1892 (13)	307
21	Tom Glavine	2008 (22)	305	21(21)	0	Tom Glavine	2008 (22)	302
22	Randy Johnson	2009 (22)	303	22(25)	12	Bobby Mathews	1887 (15)	300
23	Early Wynn	1963 (23)	300	22(23)	4	Early Wynn	1963 (23)	300
23	Lefty Grove	1941 (17)	300	24(23)	-4	Lefty Grove	1941 (17)	299
25	Bobby Mathews	1887 (15)	297	24(22)	-9	Randy Johnson	2009 (22)	299
26	Tommy John	1989 (26)	288	26(26)	0	Tommy John	1989 (26)	288
27	Bert Blyleven	1992 (22)	287	27(27)	0	Bert Blyleven	1992 (22)	287
28	Robin Roberts	1966 (19)	286	28(28)	0	Robin Roberts	1966 (19)	285
29	Tony Mullane	1894 (13)	284	29(29)	0	Fergie Jenkins	1983 (19)	284
29	Fergie Jenkins	1983 (19)	284	30(31)	3	Jim Kaat	1983 (25)	283

not surprising, since pitcher wins is largely dependent on team factors

Career Hits: not affected by detrending

Table S2. Ranking of career hits (1871–2009).

Traditional Rank						Detrended Rank		
Rank	Name	Final Season (L)	Career Metric	Rank* (Rank)	% Change	Name	Final Season (L)	Career Metric
1	Pete Rose	1986 (24)	4256	1(1)	0	Pete Rose	1986 (24)	4409
2	Ty Cobb	1928 (24)	4189	2(2)	0	Ty Cobb	1928 (24)	4166
3	Hank Aaron	1976 (23)	3771	3(3)	0	Hank Aaron	1976 (23)	3890
4	Stan Musial	1963 (22)	3630	4(4)	0	Stan Musial	1963 (22)	3661
5	Tris Speaker	1928 (22)	3514	5(6)	16	Carl Yastrzemski	1983 (23)	3537
6	Carl Yastrzemski	1983 (23)	3419	6(8)	25	Honus Wagner	1917 (21)	3484
7	Cap Anson	1897 (27)	3418	7(7)	0	Cap Anson	1897 (27)	3464
8	Honus Wagner	1917 (21)	3415	8(5)	-60	Tris Speaker	1928 (22)	3449
9	Paul Molitor	1998 (21)	3319	9(11)	18	Willie Mays	1973 (22)	3375
10	Eddie Collins	1930 (25)	3315	10(9)	-11	Paul Molitor	1998 (21)	3361
11	Willie Mays	1973 (22)	3283	11(12)	8	Eddie Murray	1997 (21)	3303
12	Eddie Murray	1997 (21)	3255	12(13)	7	Nap Lajoie	1916 (21)	3291
13	Nap Lajoie	1916 (21)	3242	13(10)	-30	Eddie Collins	1930 (25)	3266
14	Cal Ripken	2001 (21)	3184	14(15)	6	George Brett	1993 (21)	3222
15	George Brett	1993 (21)	3154	15(14)	-7	Cal Ripken	2001 (21)	3219
16	Paul Waner	1945 (20)	3152	16(17)	5	Robin Yount	1993 (20)	3209
17	Robin Yount	1993 (20)	3142	17(18)	5	Tony Gwynn	2001 (20)	3175
18	Tony Gwynn	2001 (20)	3141	18(19)	5	Dave Winfield	1995 (22)	3171
19	Dave Winfield	1995 (22)	3110	19(23)	17	Lou Brock	1979 (19)	3150
20	Craig Biggio	2007 (20)	3060	20(22)	9	Rod Carew	1985 (19)	3149
21	Rickey Henderson	2003 (25)	3055	21(27)	22	Roberto Clemente	1972 (18)	3107
22	Rod Carew	1985 (19)	3053	22(26)	15	Al Kaline	1974 (22)	3094
23	Lou Brock	1979 (19)	3023	23(21)	-9	Rickey Henderson	2003 (25)	3089
24	Rafael Palmeiro	2005 (20)	3020	24(20)	-20	Craig Biggio	2007 (20)	3060
25	Wade Boggs	1999 (18)	3010	25(25)	0	Wade Boggs	1999 (18)	3053
26	Al Kaline	1974 (22)	3007	26(29)	10	Sam Crawford	1917 (19)	3046
27	Roberto Clemente	1972 (18)	3000	27(30)	10	Frank Robinson	1976 (21)	3040
28	Sam Rice	1934 (20)	2987	28(24)	-16	Rafael Palmeiro	2005 (20)	3034
29	Sam Crawford	1917 (19)	2961	29(16)	-81	Paul Waner	1945 (20)	2968
30	Frank Robinson	1976 (21)	2943	30(42)	28	Brooks Robinson	1977 (23)	2955

not so surprising since career hits is closely related to career length, which hasn't changed significantly

Career Strikeouts: affected by distinct pitcher eras

Table S10. Ranking of season strikeouts for the Modern Era (1920–2009).

Traditional Rank				Detrended Rank				
Rank	Name	Season (Y#)	Season Metric	Rank* (Rank)	% Change	Name	Season (Y#)	Season Metric
1	Nolan Ryan	1973 (7)	383	1(72)	98	Dazzy Vance	1924 (5)	443
2	Sandy Koufax	1965 (11)	382	2(6)	66	Bob Feller	1946 (8)	407
3	Randy Johnson	2001 (14)	372	3(197)	98	Dazzy Vance	1925 (6)	368
4	Nolan Ryan	1974 (8)	367	4(4)	0	Nolan Ryan	1974 (8)	335
5	Randy Johnson	1999 (12)	364	5(79)	93	Bob Feller	1941 (6)	334
6	Bob Feller	1946 (8)	348	6(1)	-500	Nolan Ryan	1973 (7)	333
7	Randy Johnson	2000 (13)	347	7(75)	90	Bob Feller	1940 (5)	325
8	Nolan Ryan	1977 (11)	341	8(133)	93	Van Mungo	1936 (6)	323
9	Randy Johnson	2002 (15)	334	9(47)	80	Hal Newhouser	1946 (8)	322
10	Nolan Ryan	1972 (6)	329	10(102)	90	Bob Feller	1939 (4)	321
10	Randy Johnson	1998 (11)	329	11(435)	97	Lefty Grove	1926 (2)	317
12	Nolan Ryan	1976 (10)	327	12(124)	90	Bob Feller	1938 (3)	316
13	Sam McDowell	1965 (5)	325	12(400)	97	Dazzy Vance	1923 (4)	316
14	Curt Schilling	1997 (10)	319	12(367)	96	Dazzy Vance	1928 (9)	316
15	Sandy Koufax	1966 (12)	317	15(12)	-25	Nolan Ryan	1976 (10)	310
16	Curt Schilling	2002 (15)	316	16(8)	-100	Nolan Ryan	1977 (11)	301
17	J.R. Richard	1979 (9)	313	17(578)	97	Dazzy Vance	1927 (8)	299
17	Pedro Martinez	1999 (8)	313	18(175)	89	Bobo Newsom	1938 (8)	298
19	Steve Carlton	1972 (8)	310	18(382)	95	Dizzy Dean	1933 (3)	298
20	Mickey Lolich	1971 (9)	308	18(17)	-5	J.R. Richard	1979 (9)	298
20	Randy Johnson	1993 (6)	308	21(2)	-950	Sandy Koufax	1965 (11)	294
22	Mike Scott	1986 (8)	306	21(251)	91	Hal Newhouser	1945 (7)	294
22	Sandy Koufax	1963 (9)	306	23(269)	91	Lefty Grove	1930 (6)	293
24	Pedro Martinez	1997 (6)	305	24(26)	7	J.R. Richard	1978 (8)	289
25	Sam McDowell	1970 (10)	304	24(600)	96	Lefty Grove	1928 (4)	289
26	J.R. Richard	1978 (8)	303	26(767)	96	Lefty Grove	1927 (3)	282
27	Nolan Ryan	1989 (23)	301	27(484)	94	Dizzy Dean	1932 (2)	274
27	Vida Blue	1971 (3)	301	28(499)	94	Red Ruffing	1932 (9)	273
29	Curt Schilling	1998 (11)	300	29(37)	21	Steve Carlton	1980 (16)	272
30	Randy Johnson	1995 (8)	294	30(449)	93	George Earnshaw	1930 (3)	271
31	Curt Schilling	2001 (14)	293	31(10)	-210	Nolan Ryan	1972 (6)	270

The competitive (dis)advantage associated with particular eras (raised mound 1962-69, deadball era 1900-20) is evident in this re-ranking

Season Home Runs: case of extreme inflation

Table S6. Ranking of season home runs for the Modern Era (1920–2009).

Traditional Rank				Detrended Rank				
Rank	Name	Season (Y#)	Season Metric	Rank* (Rank)	% Change	Name	Season (Y#)	Season Metric
1	Barry Bonds	2001 (16)	73	1(19)	94	Babe Ruth	1920 (7)	133
2	Mark McGwire	1998 (13)	70	2(8)	75	Babe Ruth	1927 (14)	102
3	Sammy Sosa	1998 (10)	66	3(9)	66	Babe Ruth	1921 (8)	100
4	Mark McGwire	1999 (14)	65	4(72)	94	Babe Ruth	1926 (13)	82
5	Sammy Sosa	2001 (13)	64	5(94)	94	Babe Ruth	1924 (11)	80
6	Sammy Sosa	1999 (11)	63	5(72)	93	Lou Gehrig	1927 (5)	80
7	Roger Maris	1961 (5)	61	7(19)	63	Babe Ruth	1928 (15)	77
8	Babe Ruth	1927 (14)	60	8(61)	86	Jimmie Foxx	1933 (9)	70
9	Babe Ruth	1921 (8)	59	9(94)	90	Babe Ruth	1931 (18)	68
10	Mark McGwire	1997 (12)	58	9(94)	90	Lou Gehrig	1931 (9)	68
10	Ryan Howard	2006 (3)	58	11(10)	-10	Jimmie Foxx	1932 (8)	67
10	Hank Greenberg	1938 (7)	58	12(215)	94	Cy Williams	1923 (12)	66
10	Jimmie Foxx	1932 (8)	58	12(215)	94	Babe Ruth	1923 (10)	66
14	Alex Rodriguez	2002 (9)	57	14(181)	92	Rogers Hornsby	1922 (8)	62
14	Luis Gonzalez	2001 (12)	57	15(10)	-50	Hank Greenberg	1938 (7)	60
16	Hack Wilson	1930 (8)	56	16(301)	94	Ken Williams	1922 (7)	58
16	Ken Griffey	1998 (10)	56	16(592)	97	Rudy York	1943 (8)	58
16	Ken Griffey	1997 (9)	56	18(42)	57	Lou Gehrig	1936 (14)	57
19	Babe Ruth	1928 (15)	54	18(42)	57	Lou Gehrig	1934 (12)	57
19	Babe Ruth	1920 (7)	54	20(16)	-25	Hack Wilson	1930 (8)	56
19	Alex Rodriguez	2007 (14)	54	21(135)	84	Hank Greenberg	1946 (12)	55
19	David Ortiz	2006 (10)	54	21(401)	94	Tilly Walker	1922 (12)	55
19	Mickey Mantle	1961 (11)	54	23(94)	75	Babe Ruth	1929 (16)	53
19	Ralph Kiner	1949 (4)	54	23(899)	97	Charlie Keller	1943 (5)	53
25	Jim Thome	2002 (12)	52	25(301)	91	Rogers Hornsby	1925 (11)	52
25	Alex Rodriguez	2001 (8)	52	25(36)	30	Jimmie Foxx	1938 (14)	52
25	Mark McGwire	1996 (11)	52	25(519)	95	Babe Ruth	1922 (9)	52
25	Willie Mays	1965 (14)	52	28(135)	79	Jimmie Foxx	1934 (10)	51
25	Mickey Mantle	1956 (6)	52	28(1023)	97	Hack Wilson	1927 (5)	51
25	George Foster	1977 (9)	52	28(1023)	97	Cy Williams	1927 (16)	51
31	Johnny Mize	1947 (9)	51	28(457)	93	Ted Williams	1942 (4)	51
31	Willie Mays	1955 (4)	51	32(161)	80	Chuck Klein	1929 (2)	50
31	Ralph Kiner	1947 (2)	51	32(31)	-3	Johnny Mize	1947 (9)	50
31	Andruw Jones	2005 (10)	51	32(31)	-3	Ralph Kiner	1947 (2)	50

.....Steroids era players show a relative decrease in their achievement significance;
Nevertheless, their achievements are still monumental in magnitude!

the big debate...Career Home Runs....

Traditional Rank				Detrended Rank			
Rank	Name	Final Season (L)	Career Metric	Rank*(Rank)	Name	Final Season (L)	Career Metric
1	Barry Bonds	2007 (22)	762	1(3)	Babe Ruth	1935 (22)	1215
2	Hank Aaron	1976 (23)	755	2(23)	Mel Ott	1947 (22)	637
3	Babe Ruth	1935 (22)	714	3(26)	Lou Gehrig	1939 (17)	635
4	Willie Mays	1973 (22)	660	3(17)	Jimmie Foxx	1945 (20)	635
5	Ken Griffey Jr.	2009 (21)	630	5(2)	Hank Aaron	1976 (23)	582
6	Sammy Sosa	2007 (18)	609	6(124)	Rogers Hornsby	1937 (23)	528
7	Frank Robinson	1976 (21)	586	7(192)	Cy Williams	1930 (19)	527
8	Alex Rodriguez	2009 (16)	583	8(1)	Barry Bonds	2007 (22)	502
8	Mark McGwire	2001 (16)	583	9(4)	Willie Mays	1973 (22)	490
10	Harmon Killebrew	1975 (22)	573	10(18)	Ted Williams	1960 (19)	482
11	Rafael Palmeiro	2005 (20)	569	11(13)	Reggie Jackson	1987 (21)	478
12	Jim Thome	2009 (19)	564	12(14)	Mike Schmidt	1989 (18)	463
13	Reggie Jackson	1987 (21)	563	13(7)	Frank Robinson	1976 (21)	444
14	Mike Schmidt	1989 (18)	548	14(10)	Harmon Killebrew	1975 (22)	437
15	Manny Ramirez	2009 (17)	546	15(577)	Gavvy Cravath	1920 (11)	433
16	Mickey Mantle	1968 (18)	536	16(718)	Honus Wagner	1917 (21)	420
17	Jimmie Foxx	1945 (20)	534	17(18)	Willie McCovey	1980 (22)	417
18	Ted Williams	1960 (19)	521	18(557)	Harry Stovey	1893 (14)	413
18	Frank Thomas	2008 (19)	521	19(5)	Ken Griffey Jr.	2009 (21)	411
18	Willie McCovey	1980 (22)	521	20(28)	Stan Musial	1963 (22)	410

...for extensive top-50 tables for Hits, HR, RBI, K, W calculated for single seasons and also over entire the career consult the papers downloadable at:

http://physics.bu.edu/~amp17/webpage_files/publications.html

A. M. Petersen, O. Penner, H. E. Stanley.
Methods for detrending success metrics to account for inflationary and deflationary factors
Eur. Phys. J. B 79, 67-78 (2011). DOI: 10.1140/epjb/e2010-10647-1

and an analogous statistical analysis of basketball career statistics:

A. M. Petersen, O. Penner.
A method for the unbiased comparison of MLB and NBA career statistics across era
Presented at the MIT Sloan Sports Analytics Conference 2012 (2012).

... aside from being fun... Baseball is a historical treasure

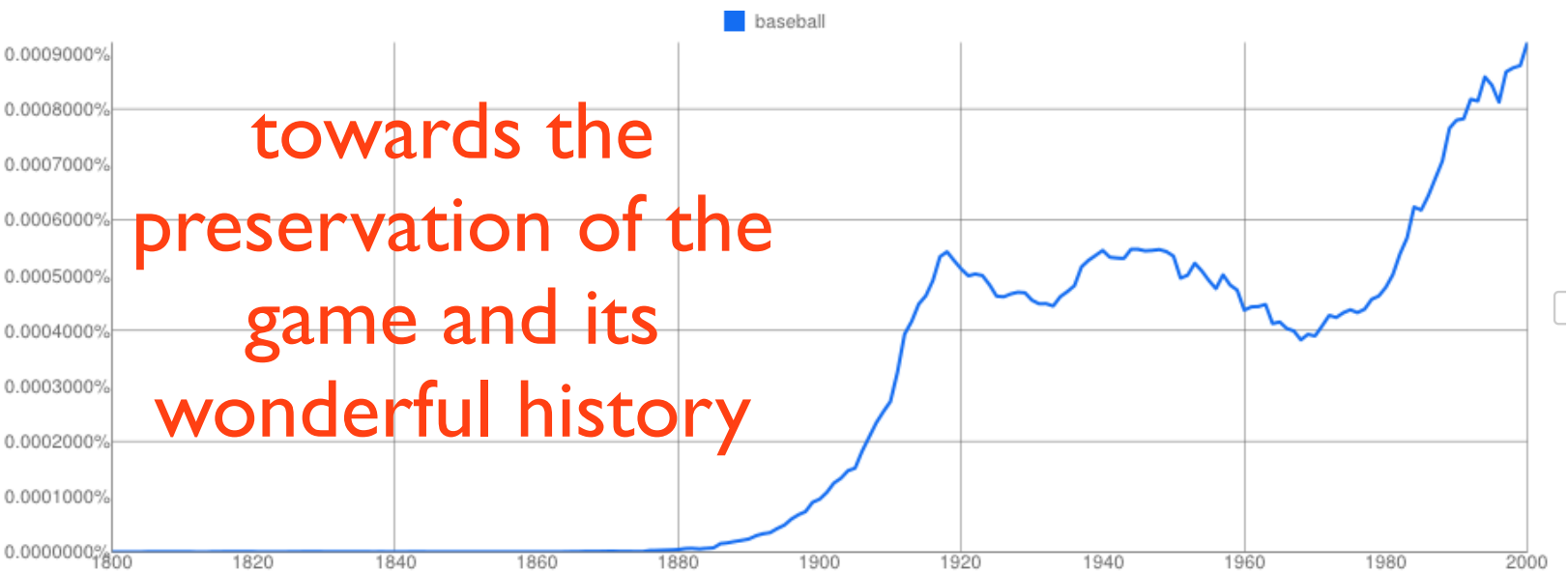
Google books Ngram Viewer

http://books.google.com/ngrams/

Graph these **case-sensitive** comma-separated phrases: baseball

between 1800 and 2000 from the corpus English One Million with smoothing of 3.

Search lots of books



towards the preservation of the game and its wonderful history

Closing remarks....

Relevant cultural questions:

(i) How to quantitatively account for economic, technological, and social factors that influence the rate of success in competitive professions.

(ii) How to use career statistics in an unbiased fashion to help in the both the standard, as well as, retroactive induction of athletes into a Hall of Fame. This is particularly important given the "inflation" observed for home runs in Major League Baseball, a phenomena that is believed to be related to the widespread use of Performance Enhancing Drugs (PED).

Relevant bar-stool debates:

Who was The Greatest Slugger of All-Time ?????????

Thank You!

and also thanks to my collaborators:

Woo-Sung Jung, Orion Penner, Gene Stanley, Sauro Succi, Fengzhong Wang, Jae-Sook Yang, Massimo Riccaboni and Fabio Pammolli

http://physics.bu.edu/~amp17/

I) A. M. Petersen, W.-S. Jung, J.-S. Yang, H. E. Stanley, "Quantitative and empirical demonstration of the Matthew effect in a study of career longevity." Proc. Natl. Acad. Sci. USA 108, 18-23 (2011).

II) A. M. Petersen, W.-S. Jung, H. E. Stanley, "On the distribution of career longevity and the evolution of home run prowess in professional baseball." Europhysics Letters 83, 50010 (2008).

III) A. M. Petersen, O. Penner, H. E. Stanley , "'Methods for detrending success metrics to account for inflationary and deflationary factors." Eur. Phys. J. B 79, 67-78 (2011).

IV) A. M. Petersen, O. Penner. "A method for the unbiased comparison of MLB and NBA career statistics across era." Presented at the MIT Sloan Sports Analytics Conference, 2012.

V) A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli "'Persistence and Uncertainty in the Academic Career." Proc. Natl. Acad. Sci. USA 109, 5213 - 5218 (2012).

“Beyond the asterisk* : Adjusting for performance inflation in professional sports”

The evaluation of success depends on many factors, some time dependent, others time independent. In order to compare human achievements from different time periods, success metrics should be normalized to a common index so that the time dependent factors do not bias the comparison of the statistical measures. This consideration is particularly relevant to career achievement records in Major League Baseball (MLB), which are of significant cultural importance. I will present a novel approach which removes the time-dependent factors by normalizing a player’s annual achievement by the local ability average. Using empirical career data for more than 15,000 MLB player careers, our method yields “detrended” success measures that are more appropriate for comparing and evaluating the relative merits of players from different historical eras. In particular, this study addresses two relevant cultural questions: (i) How to quantitatively account for economic, technological, and social factors that influence the rate of success in competitive professions, and (ii) How to use career statistics in an unbiased fashion to help in the both the standard, as well as, retroactive induction of athletes into a Hall of Fame; This is particularly important given the “steroids-era” inflation observed for home runs in Major League Baseball, a phenomena that is believed to be related to the widespread use of Performance Enhancing Drugs (PED).