The Best, The Hottest, & The Luckiest: A Statistical Tale of Extremes

Sid Redner, Boston University, physics.bu.edu/~redner

Extreme Events: Theory, Observations, Modeling and Simulations

Palma de Mallorca, November 10-14, 2008

The Best

with S. Maslov, P. Chen, H. Xie

J. Informetrics **1**, 8-15 (2007)

Observations about scientific citations

Google page rank analysis for “hidden gems”

Phys. Rev. Citation Data (as of July 03)

353,268 papers, 3,110,839 cites

$$\langle \# \text{ cites} \rangle = 8.81, \ \langle \text{cite age} \rangle = 6.20$$

N.B.: ***Internal*** citations only; undercount by factor of 3-5.

(for highly cited HEP papers; SPIRES)

Phys. Rev. Citation Data (as of July 03)

353,268 papers, 3,110,839 cites

$$\langle \# \text{ cites} \rangle = 8.81, \ \langle \text{cite age} \rangle = 6.20$$

N.B.: ***Internal*** citations only; undercount by factor of 3-5.

(for highly cited HEP papers; SPIRES)

11 papers with > 1000 citations
79 papers with > 500 citations
237 papers with > 300 citations
2340 papers with > 100 citations
8073 papers with > 50 citations

Phys. Rev. Citation Data (as of July 03)

353,268 papers, 3,110,839 cites

$$\langle \# \text{ cites} \rangle = 8.81, \ \langle \text{cite age} \rangle = 6.20$$

N.B.: ***Internal*** citations only; undercount by factor of 3-5.

(for highly cited HEP papers; SPIRES)

11	papers with >	1000	citations
79	papers with >	500	citations
237	papers with >	300	citations
2340	papers with >	100	citations
8073	papers with >	50	citations
245459	papers with <	10	citations
84144	papers with	1	citation
23421	papers with	0	citations

PR papers with >1000 cites

	article	cites	⟨age⟩	title	author(s)
1	PR 140 A1133 (1965)	3227	26.7	Self Consistent Equations..	W. Kohn & L. J. Sham

PR papers with >1000 cites

	article			cites	⟨age⟩	title	author(s)
1	PR	140	A1133 (1965)	4598	26.7	Self Consistent Equations..	W. Kohn & L. J. Sham

PR papers with >1000 cites

	article	cites	⟨age⟩	title	author(s)
1	PR 140 A1133 (1965)	4598	26.7	Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR 136 B864 (1965)	2460	28.7	Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB 23 5048 (1981)	2079	14.4	Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL 45 566 (1980)	1781	15.4	Ground State of the Electron..	D. M. Ceperley & B. J. Alder
5	PR 108 1175 (1957)	1364	20.2	Theory of Superconductivity	Bardeen, Cooper, Schrieffer
6	PRL 19 1264 (1967)	1306	15.5	A Model of Leptons	S. Weinberg
7	PRB 12 3060 (1975)	1259	18.4	Linear Methods in Band Theory	O. K. Andersen
8	PR 124 1866 (1961)	1178	28.0	Effects of Configuration..	U. Fano
9	RMP 57 287 (1985)	1055	9.2	Disordered Electronic Systems	P. A. Lee & T. V. Ramakrishnan
10	RMP 54 437 (1982)	1045	10.8	Electronic Properties of..	T. Ando, A. B. Fowler, & F. Stern
11	PRB 13 5188 (1976)	1023	20.8	Special Points for Brillouin..	H. J. Monkhorst & J. D. Pack

Google Page Rank for Citations Brin & Page (1999)

Google Page Rank for Citations Brin & Page (1999)

Idea: *launch many surfers that surf forever until a steady-state surfer number G_i is reached at each site i of the network.*

Google Page Rank for Citations Brin & Page (1999)

Idea: *launch many surfers that surf forever until a steady-state surfer number G_i is reached at each site i of the network.*

Equation for number of surfers:

$$G_i = (1-d) \sum_{j \text{ nn } i} \frac{G_j}{k_j^{\text{out}}} + \frac{d}{N}$$

Google Page Rank for Citations Brin & Page (1999)

Idea: *launch many surfers that surf forever until a steady-state surfer number G_i is reached at each site i of the network.*

Equation for number of surfers:

random walk propagation

manna from heaven

$$G_i = (1-d)\sum_{j\,\mathrm{nn}\,i} \frac{G_j}{k_j^{\mathrm{out}}} + \frac{d}{N}$$

Google Page Rank for Citations Brin & Page (1999)

Idea: *launch many surfers that surf forever until a steady-state surfer number G_i is reached at each site i of the network.*

Equation for number of surfers:

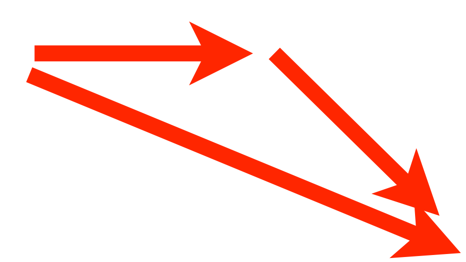
random walk propagation

manna from heaven

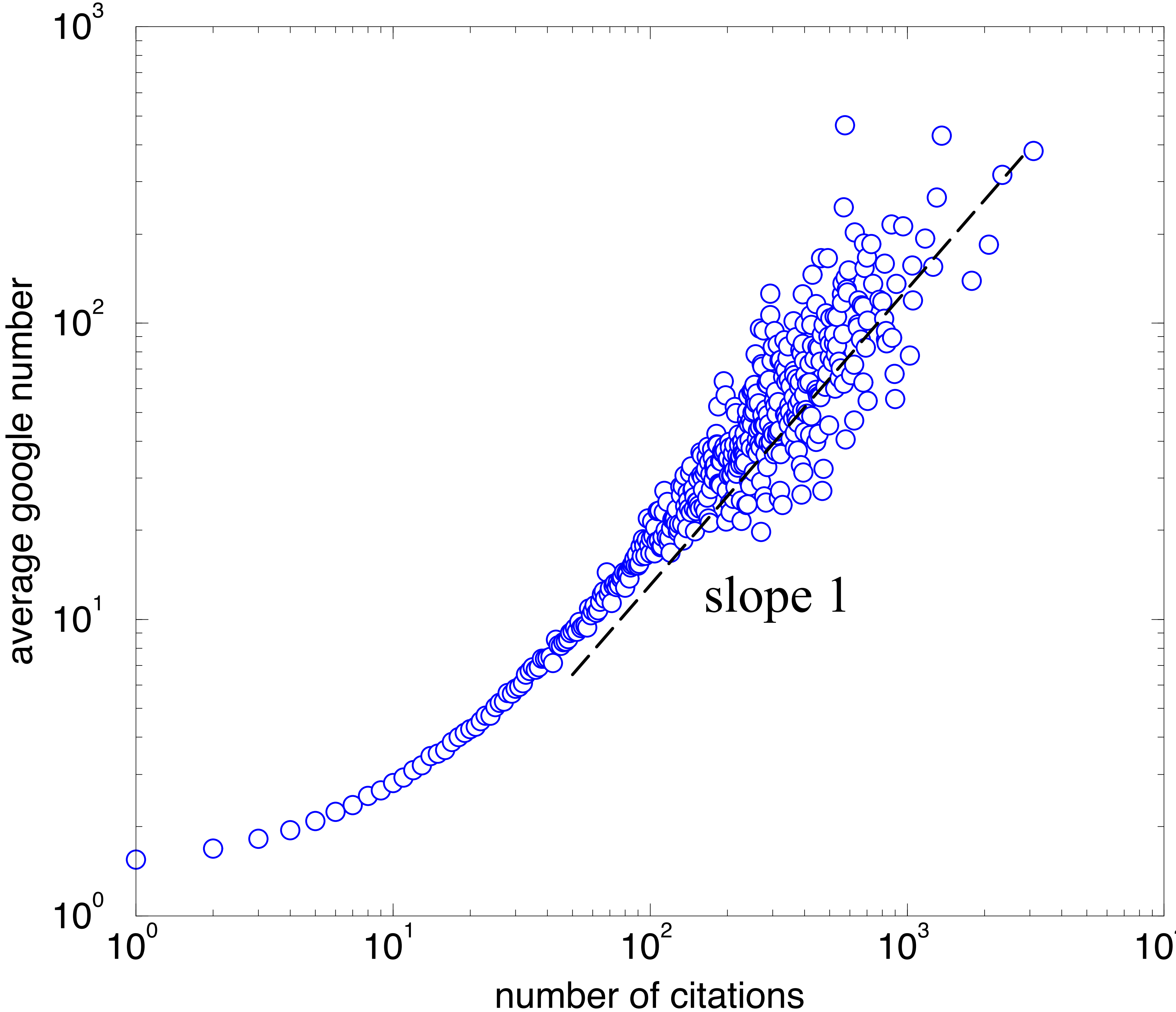
$$G_i = (1-d) \sum_{j \text{ nn } i} \frac{G_j}{k_j^{\text{out}}} + \frac{d}{N}$$

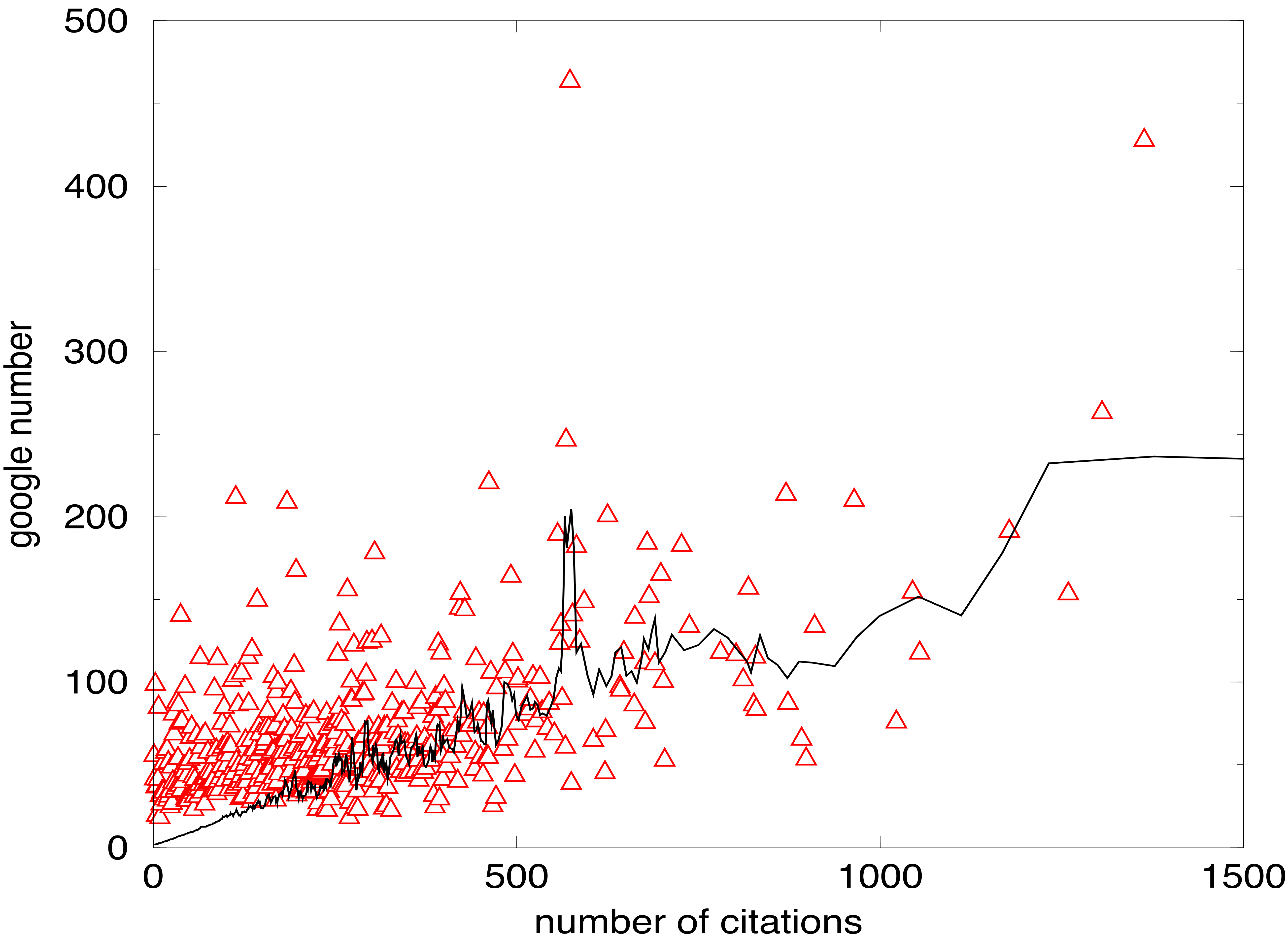
Brin/Page: d=0.15 *(bored after 6 clicks)*

We use: d=0.50 *(don't cite beyond 2 generations)*

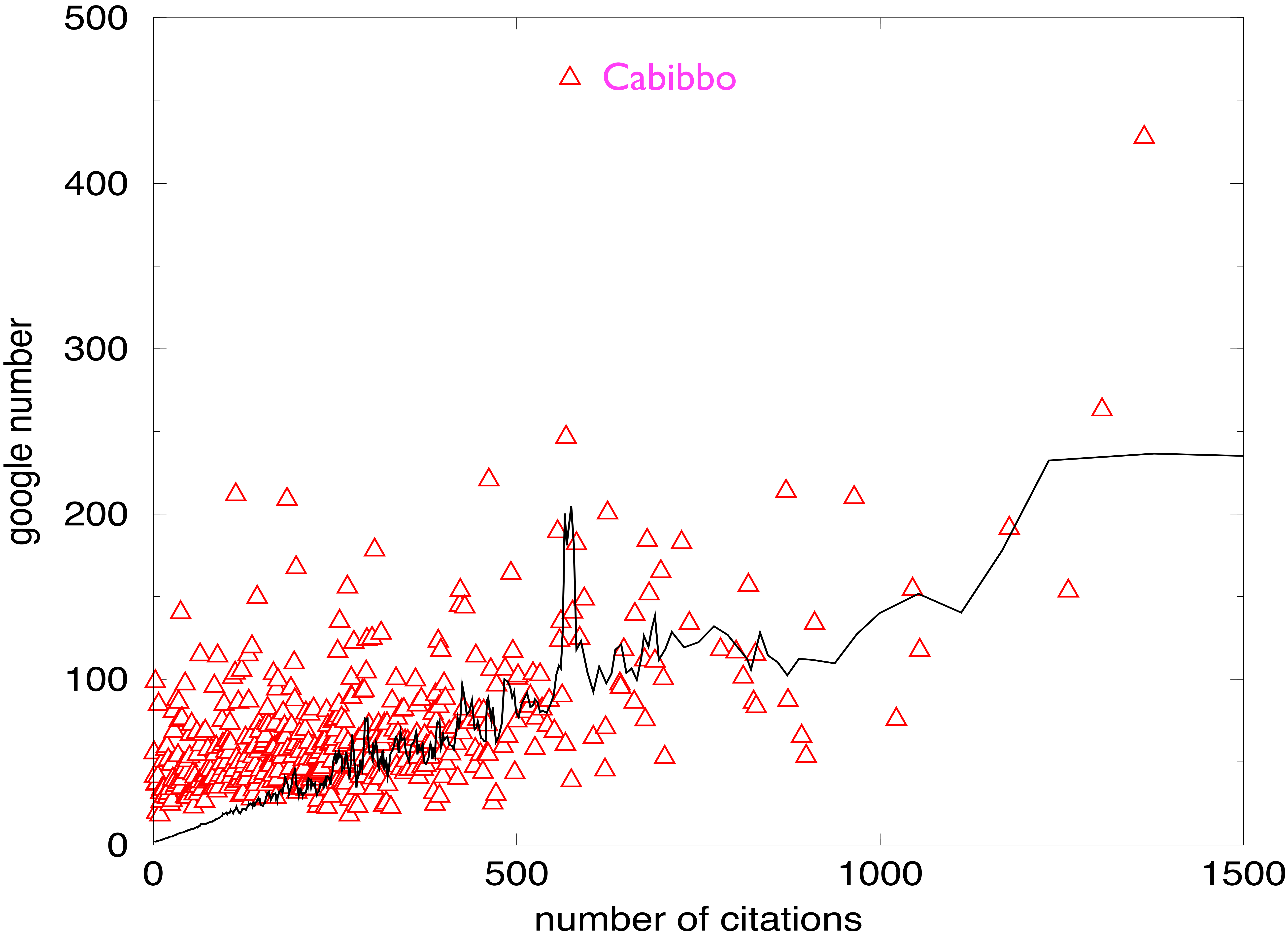


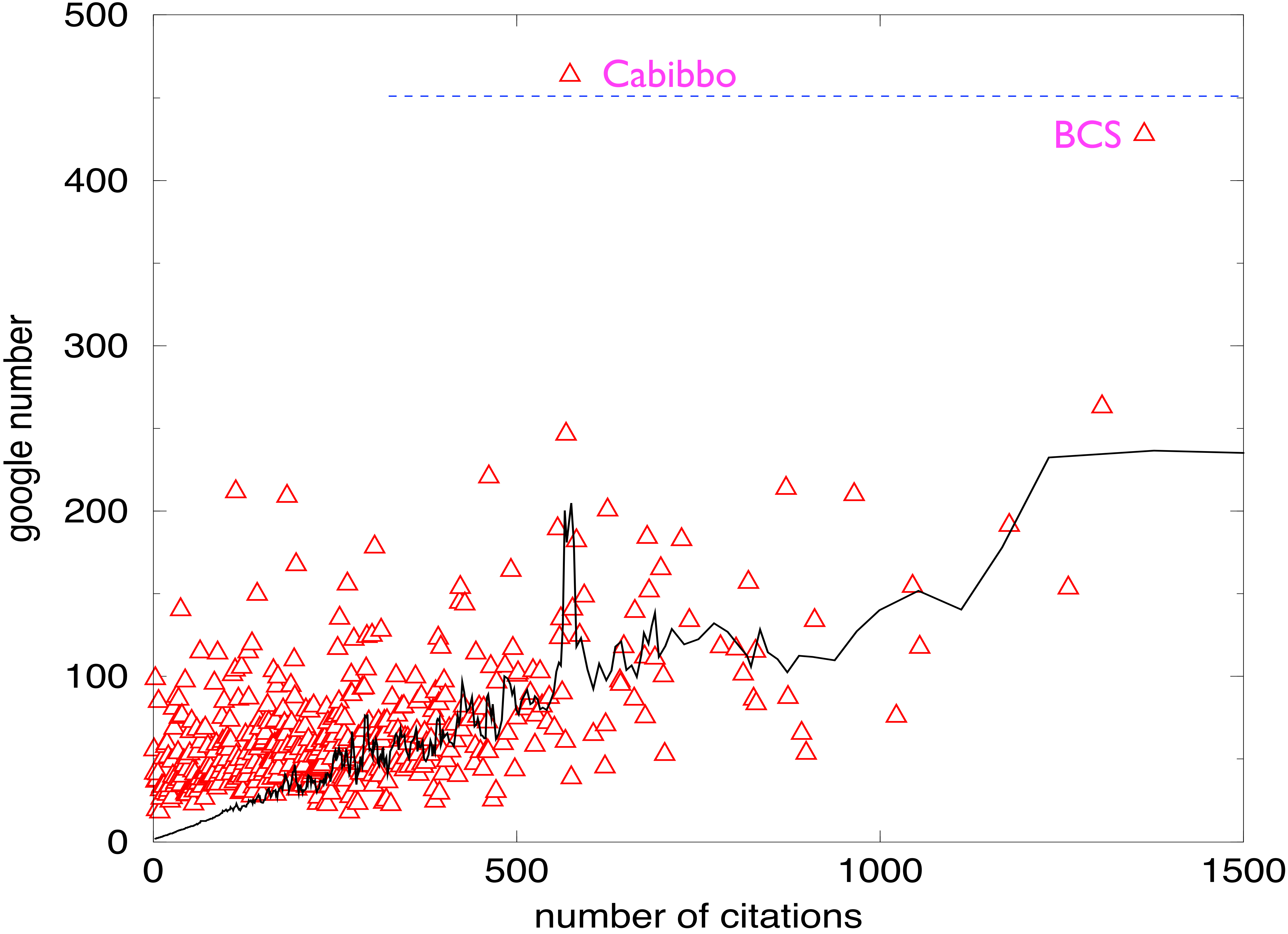
Correlation between Google & citation counts



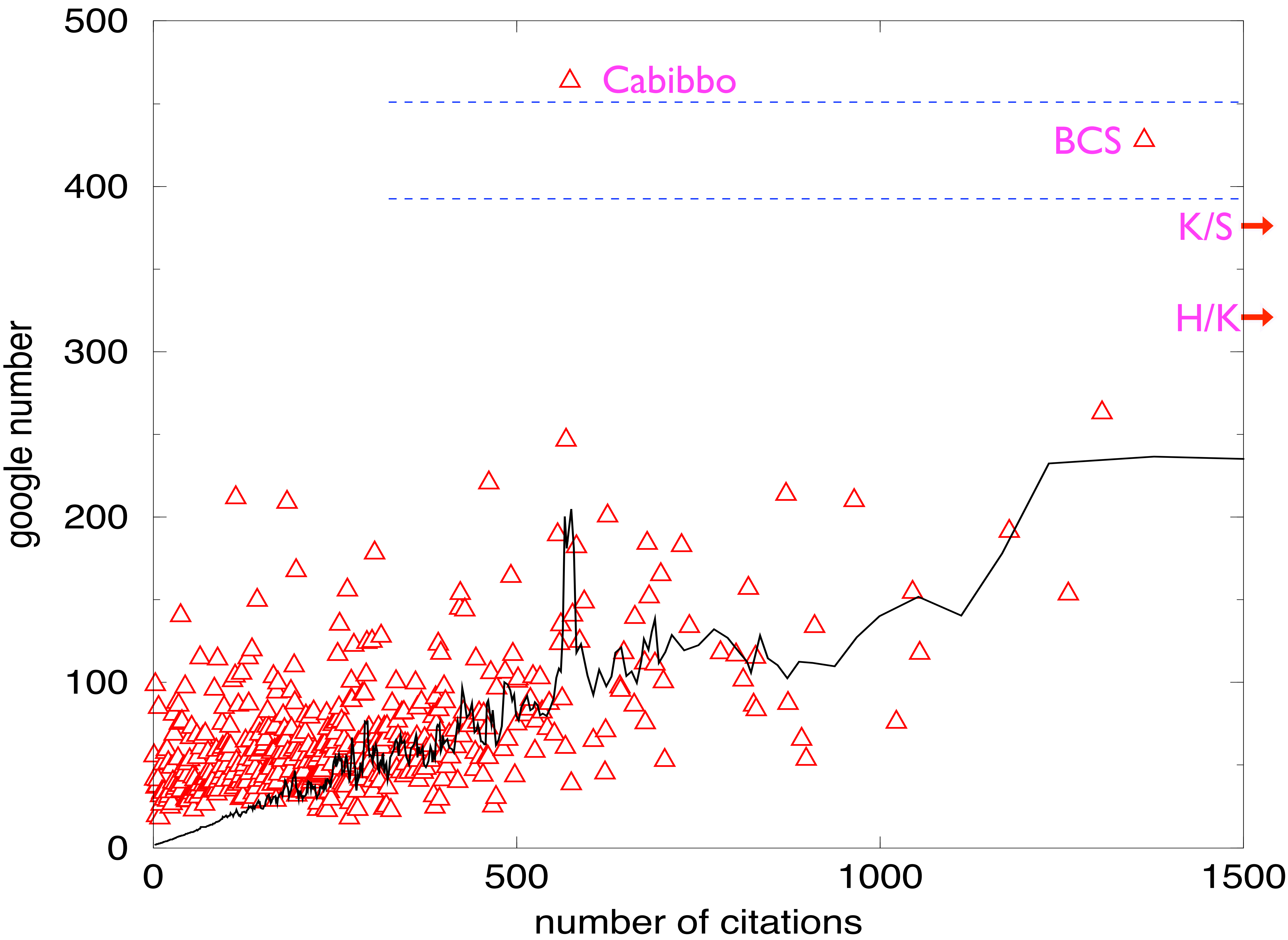


500
400
300
200
100
0
0
500
1000
1500
google number
number of citations

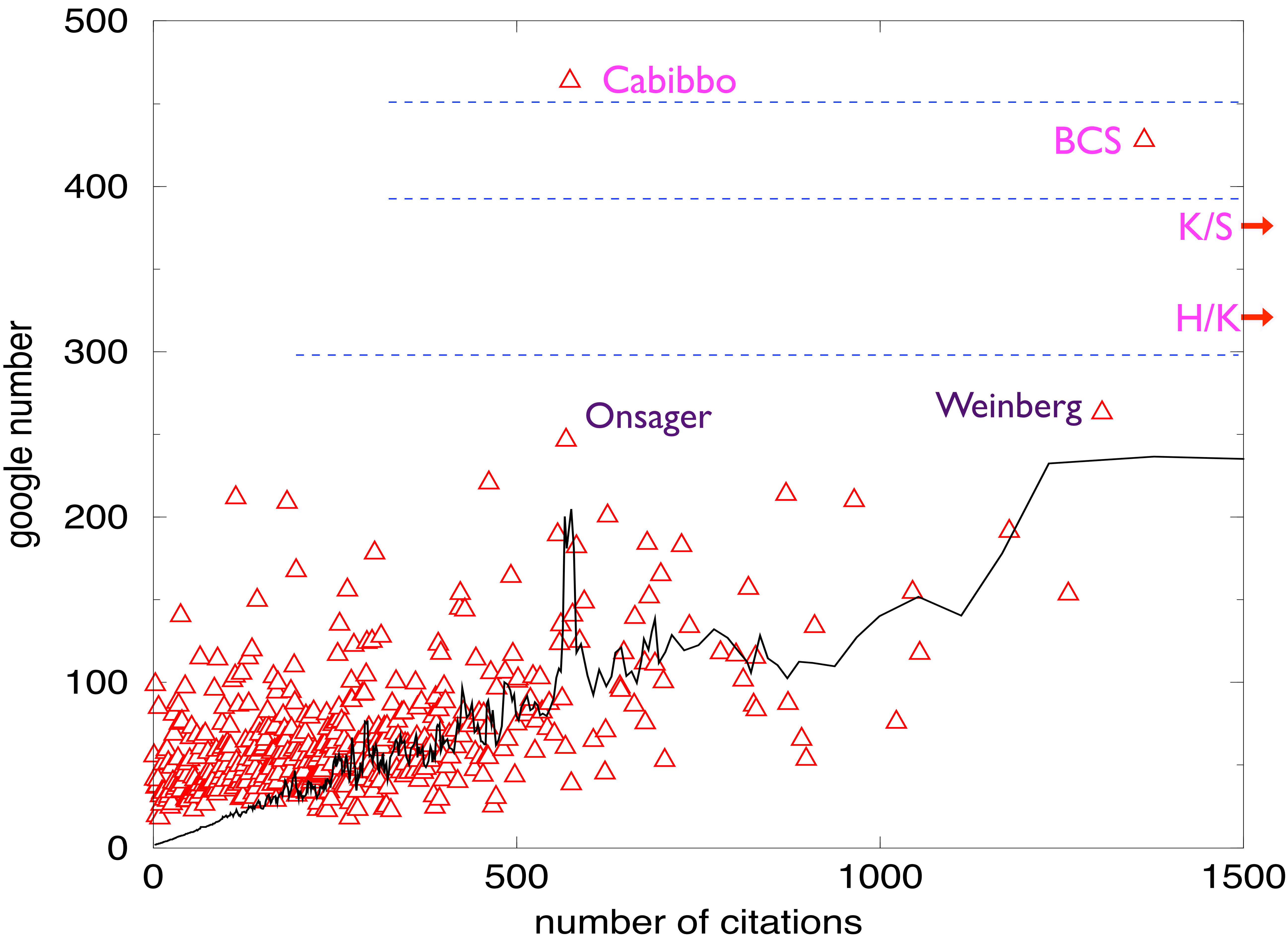

Cabibbo
google number
number of citations
0
100
200
300
400
500
0
500
1000
1500

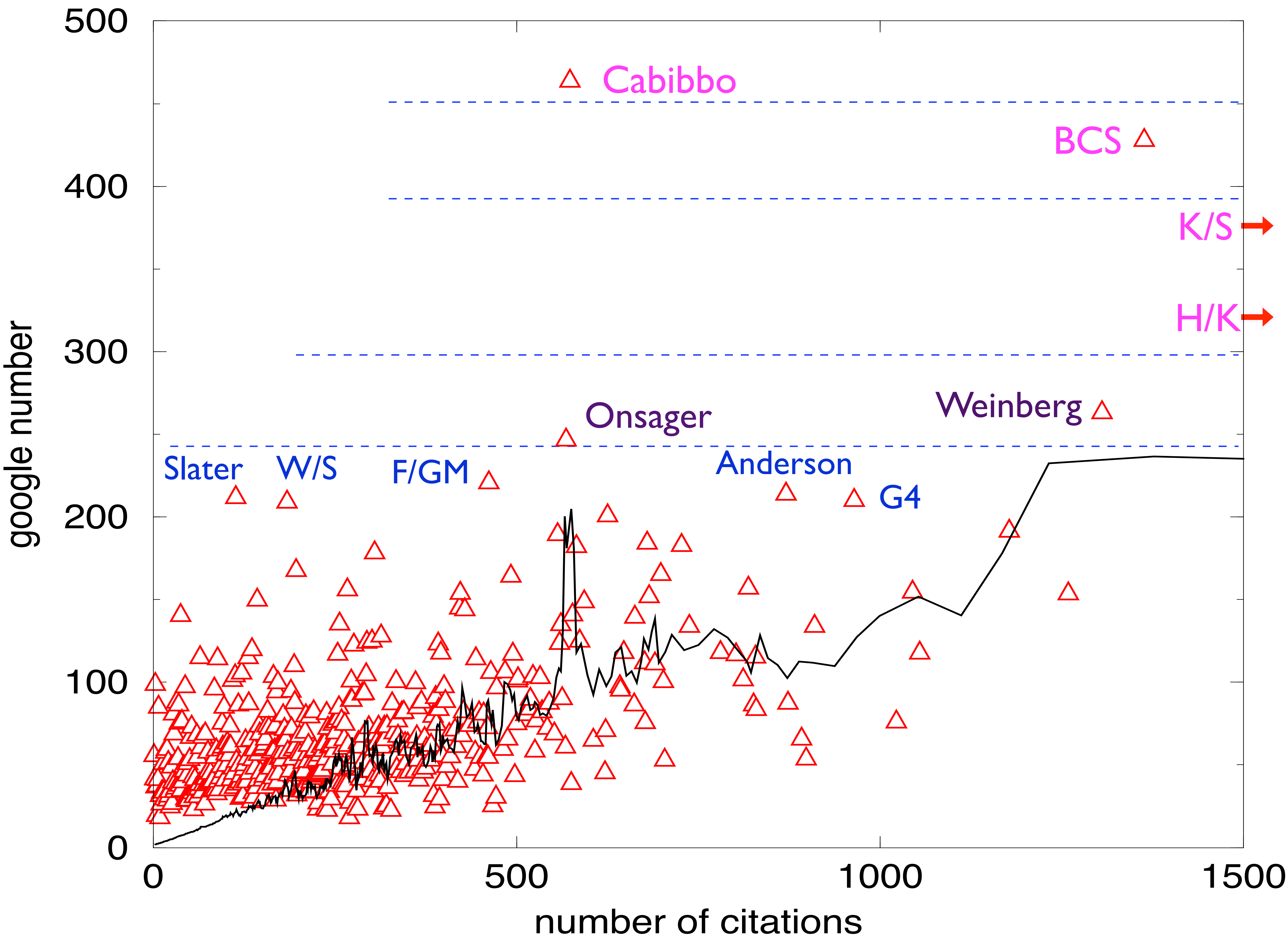


Cabibbo
BCS
google number
number of citations
0
100
200
300
400
500
0
500
1000
1500

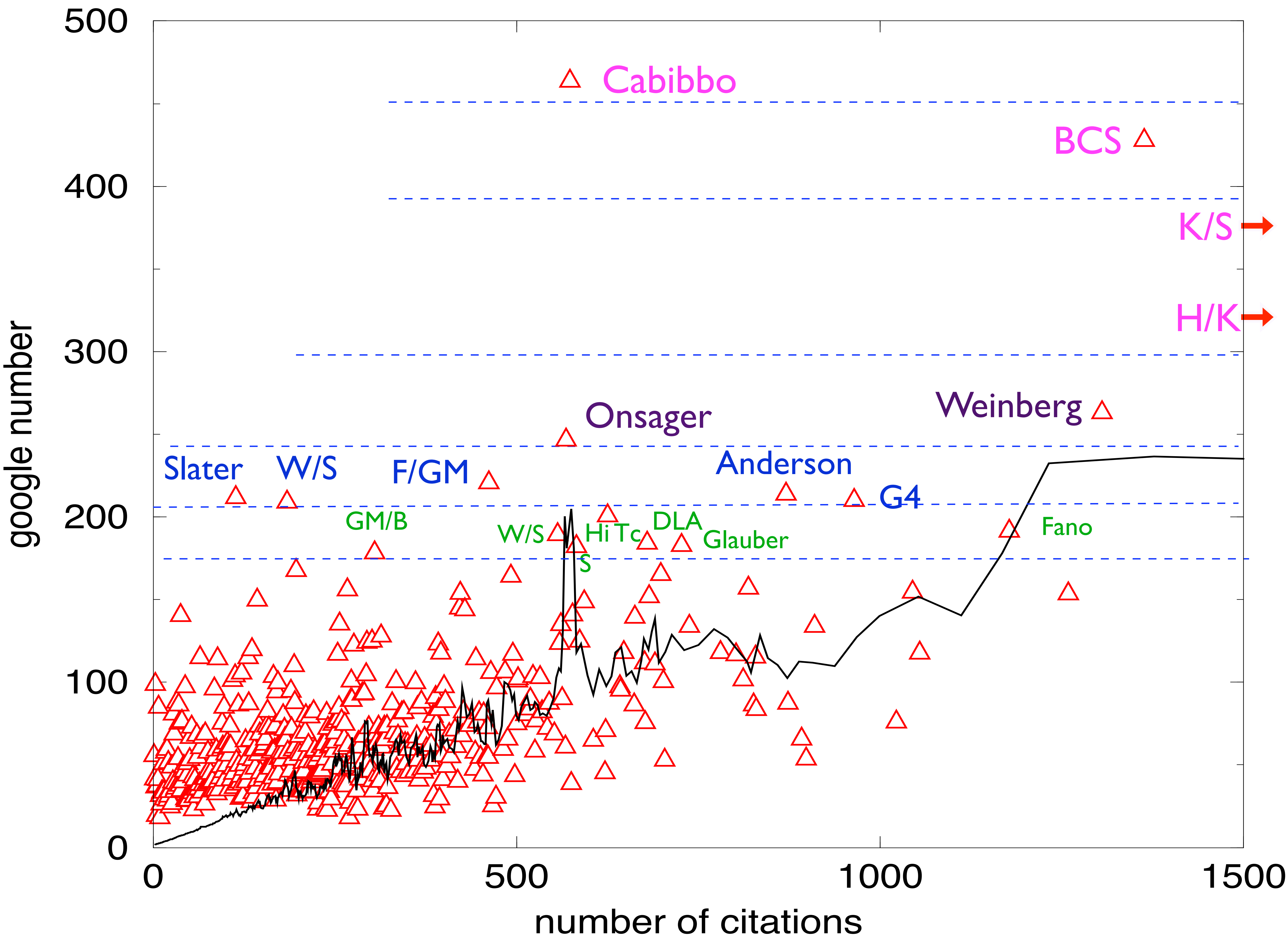


Cabibbo
BCS
K/S
H/K
google number
number of citations
0
100
200
300
400
500
0
500
1000
1500


Cabibbo
BCS
K/S
H/K
Onsager
Weinberg
google number
number of citations
0
100
200
300
400
500
0
500
1000
1500



500
400
300
200
100
0
0
500
1000
1500
google number
number of citations
Cabibbo
BCS
K/S
H/K
Onsager
Weinberg
Slater
W/S
F/GM
Anderson
G4



500
400
300
200
100
0
google number
0
500
1000
1500
number of citations
Cabibbo
BCS
K/S
H/K
Onsager
Weinberg
Slater
W/S
F/GM
Anderson
G4
GM/B
W/S
Hi Tc
S
DLA
Glauber
Fano

The Hottest

with M. Petersen

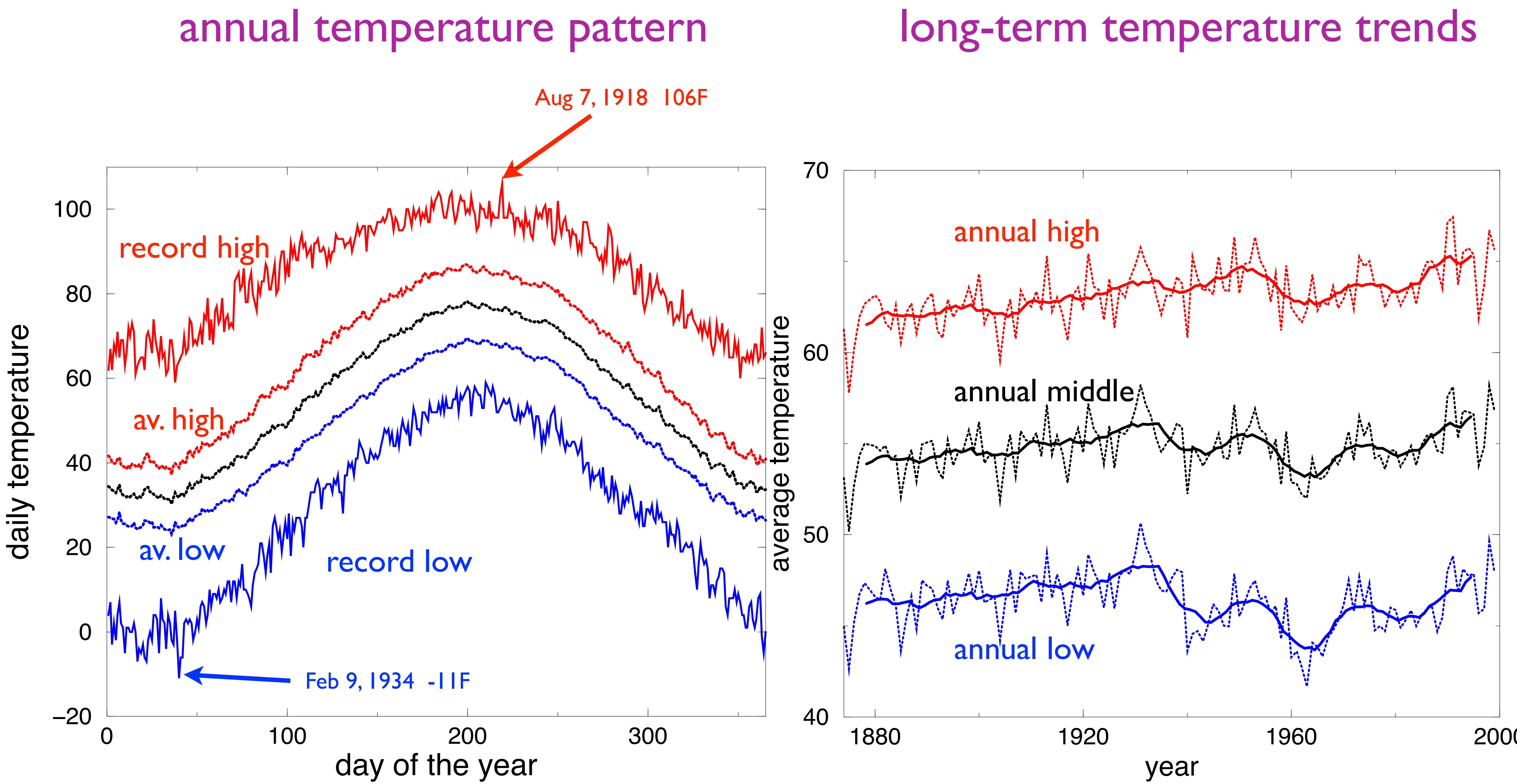
Phys. Rev. E **74**, 061114 (2006)

Some facts about urban temperatures

Evolution of record temperature events

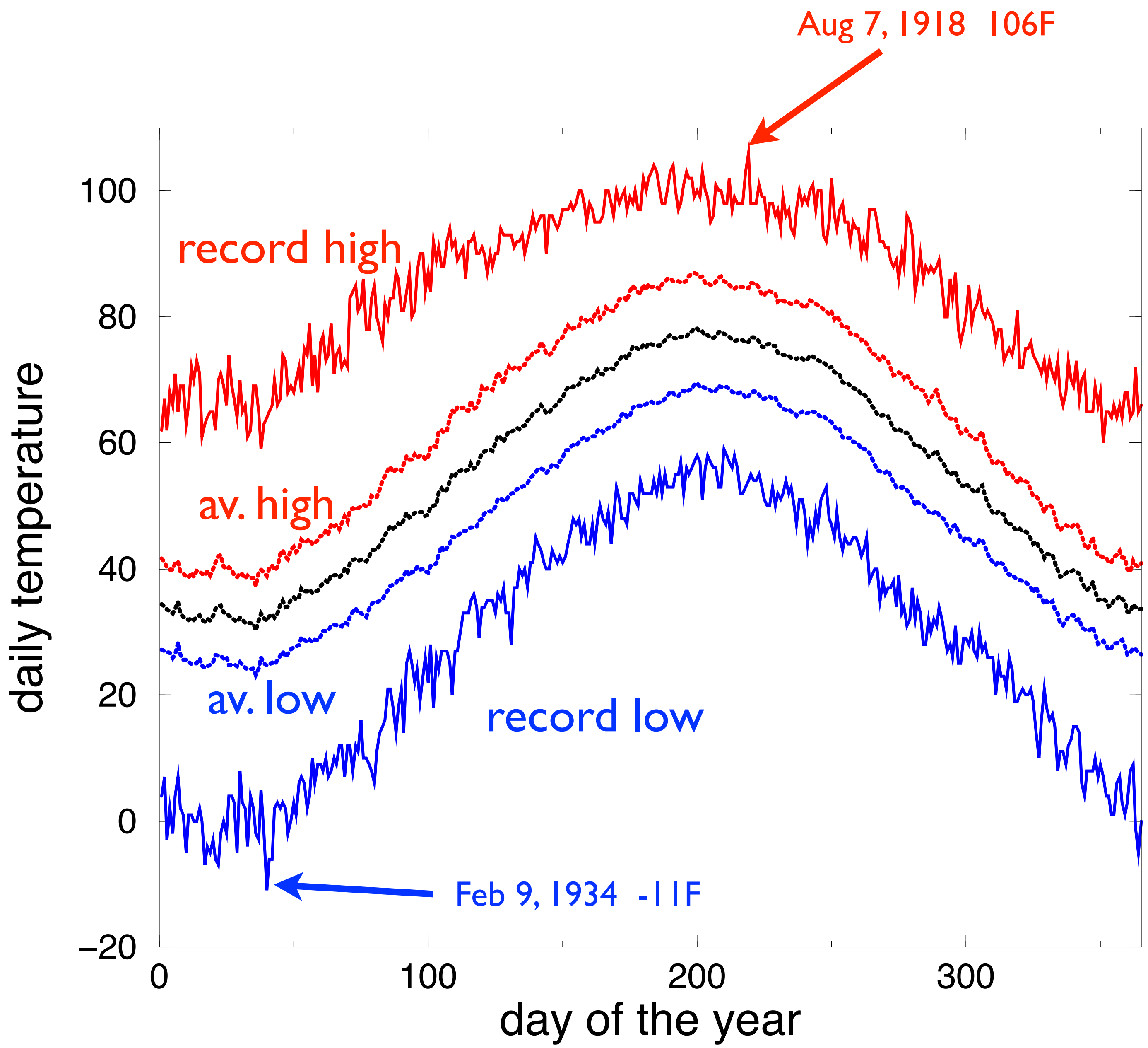
Role of global warming on records

Philadelphia Temperatures



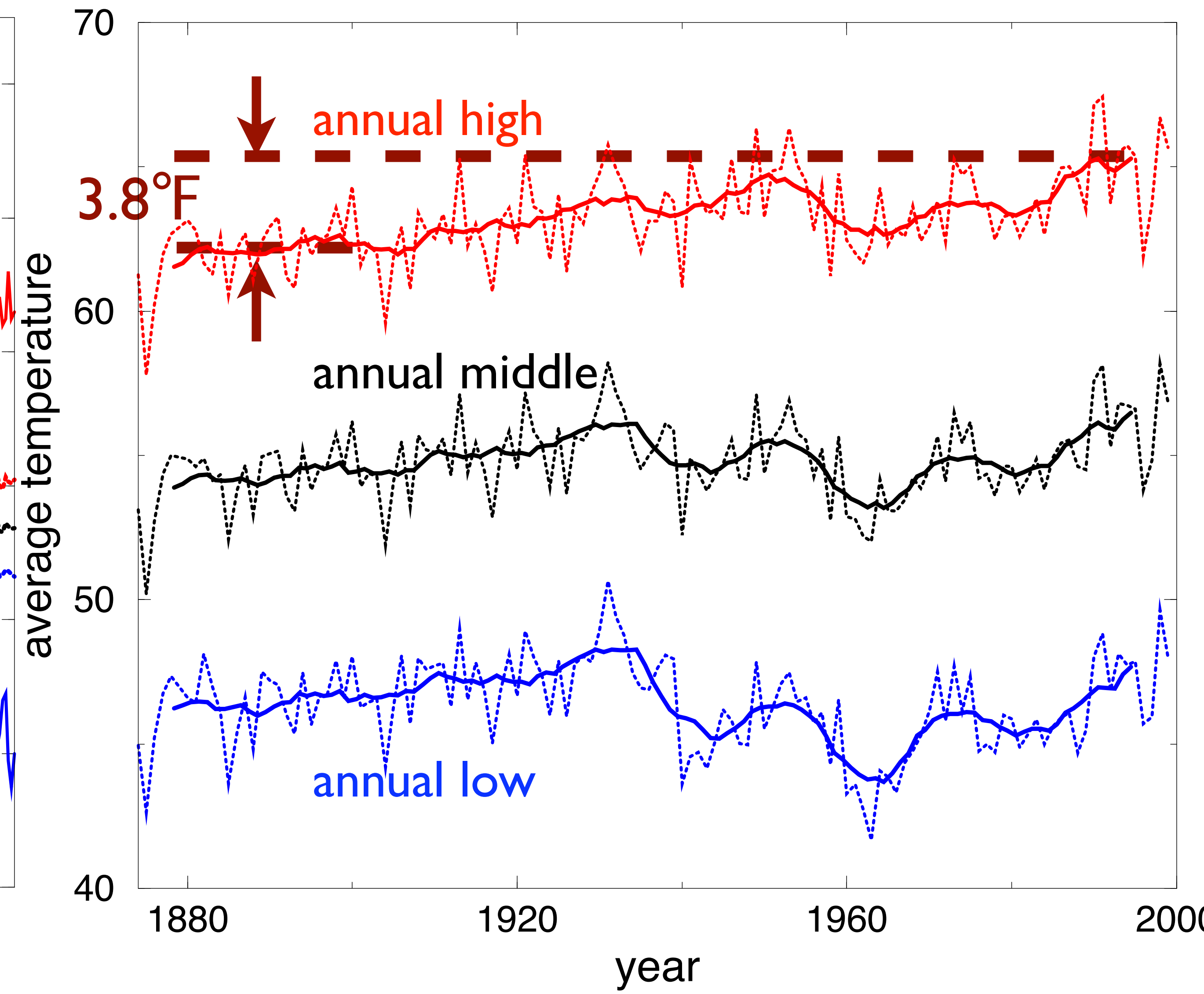
Philadelphia Temperatures

annual temperature pattern



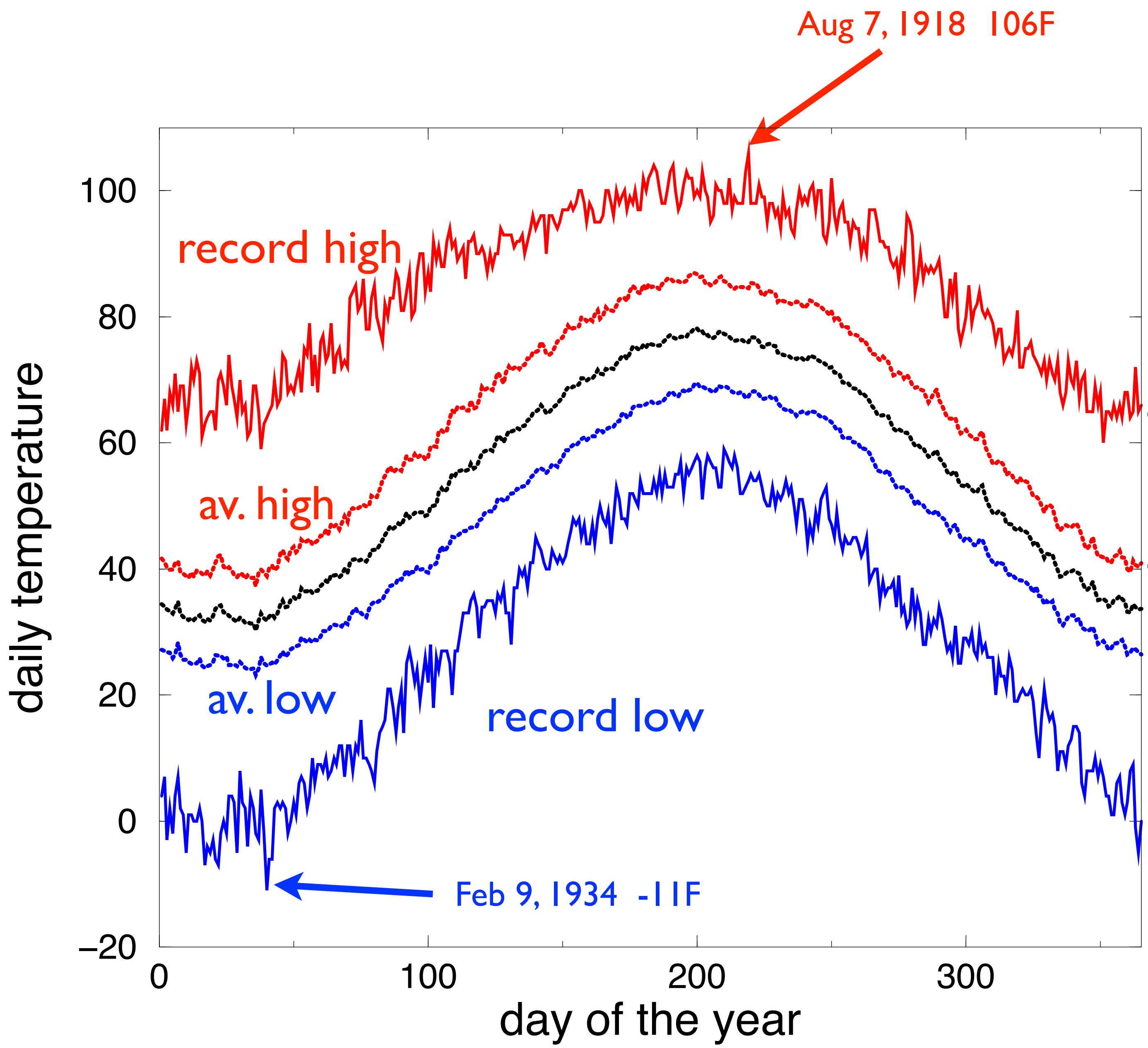
long-term temperature trends

IPCC 2001: global warming 1.1°F over past century



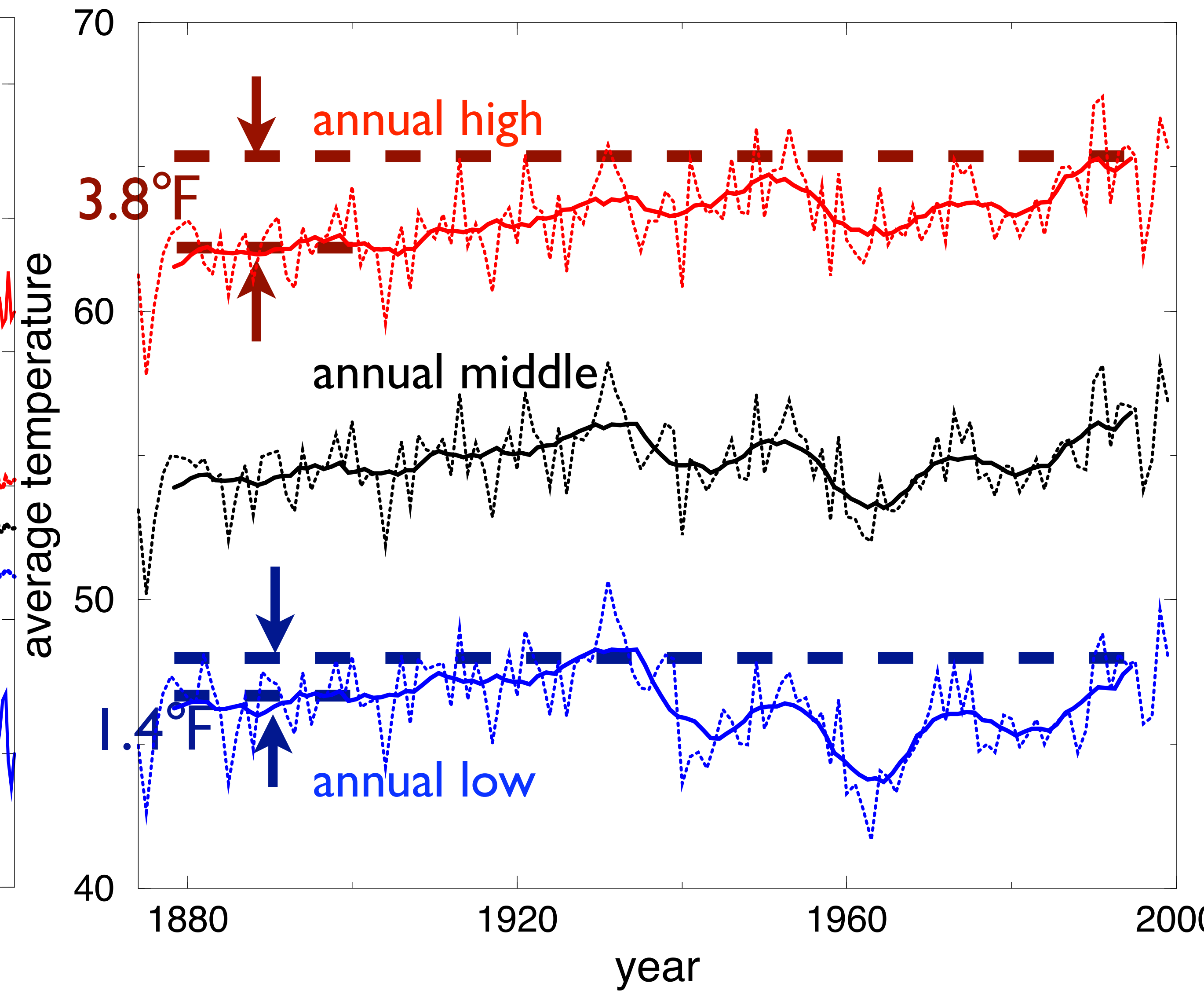
Philadelphia Temperatures

annual temperature pattern



long-term temperature trends

IPCC 2001: global warming 1.1°F over past century



Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*

Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*

Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*

Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*

Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*

Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*

Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*

Evolution of Temperature Records

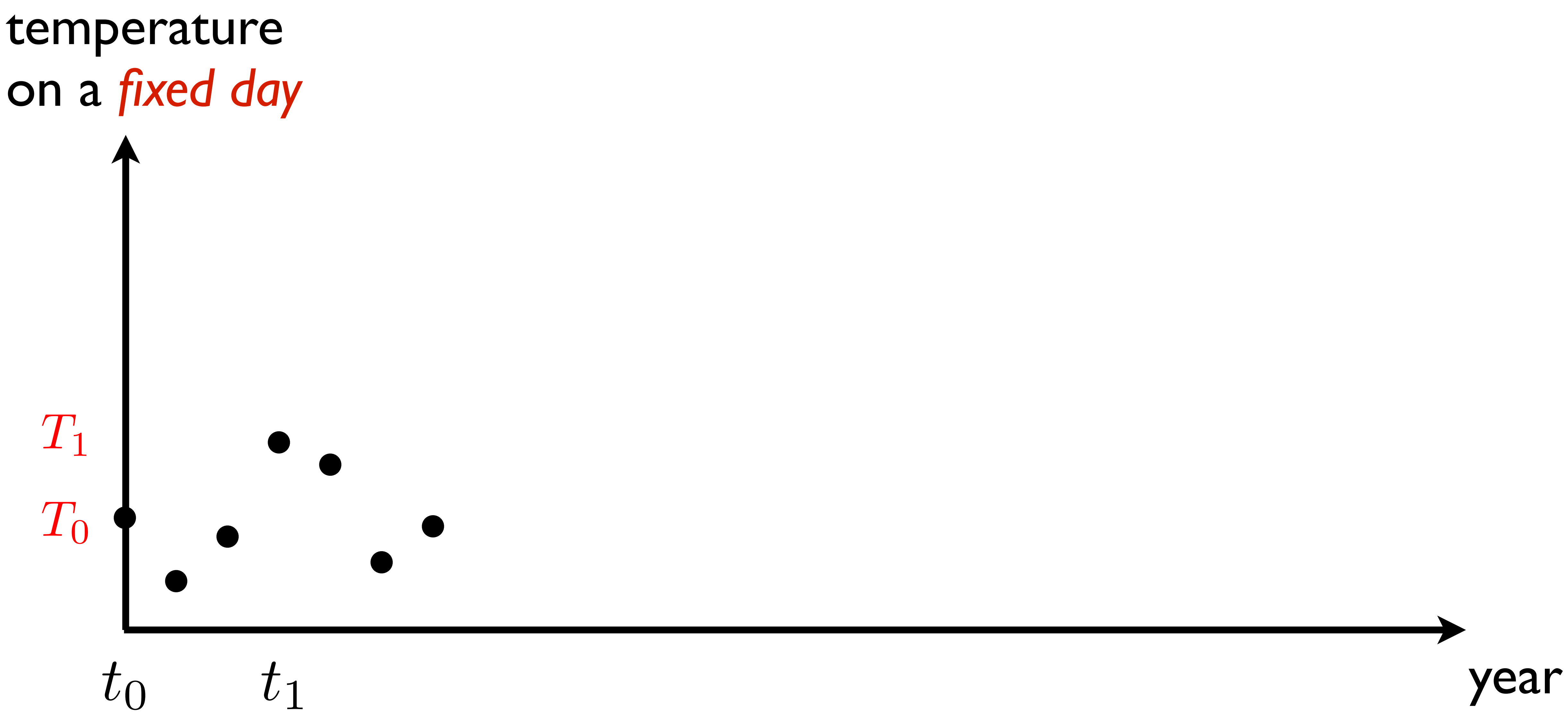
(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*

Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

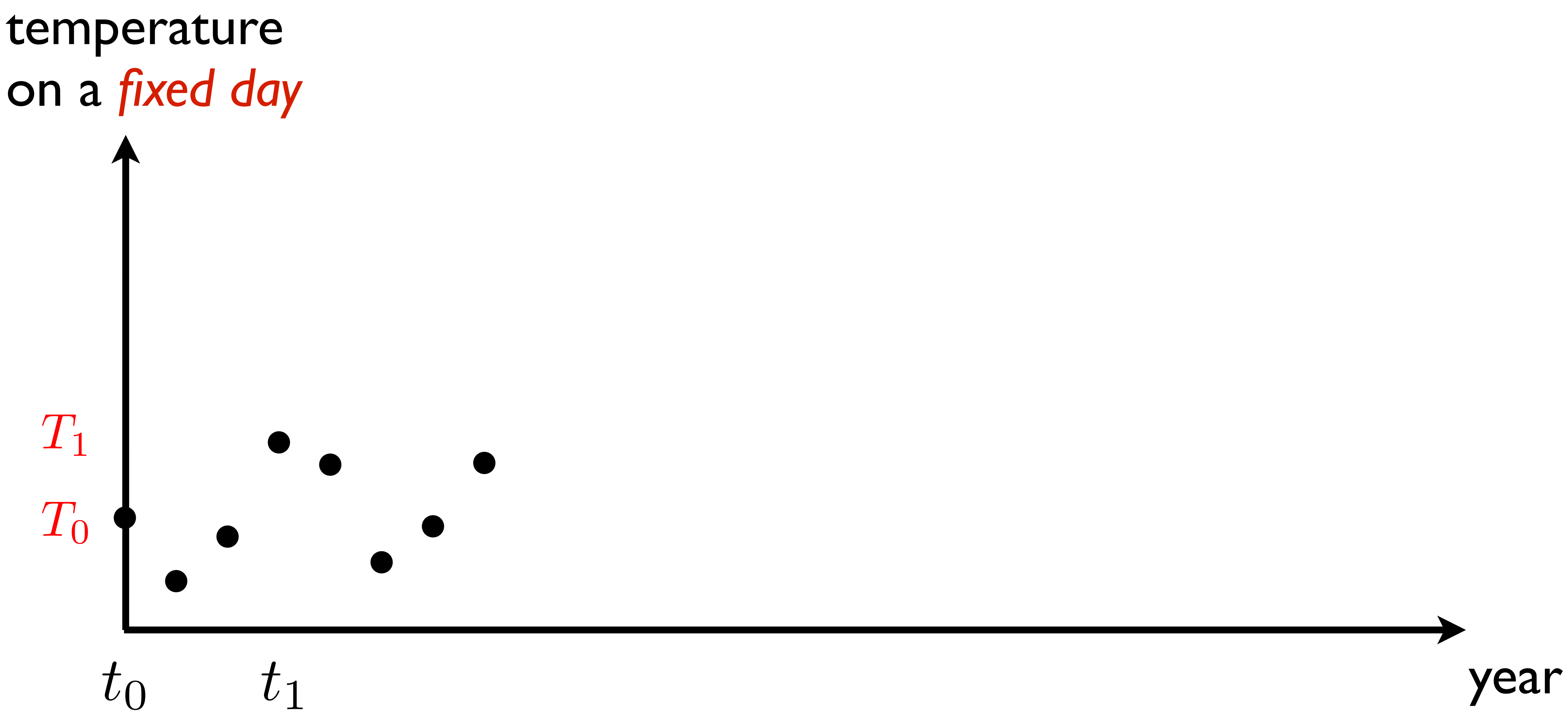
assumption: *daily temperatures are iid continuous variables*



Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

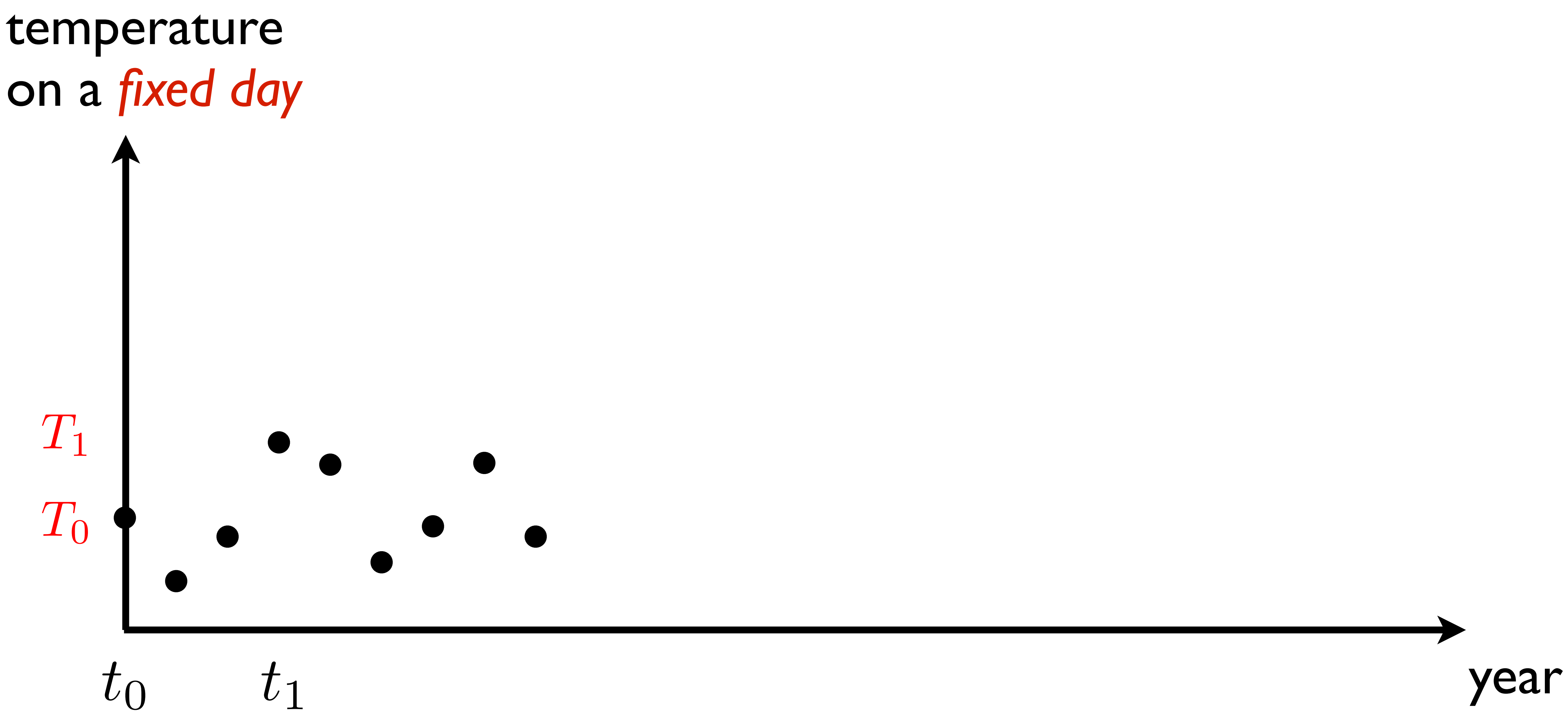
assumption: *daily temperatures are iid continuous variables*



Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

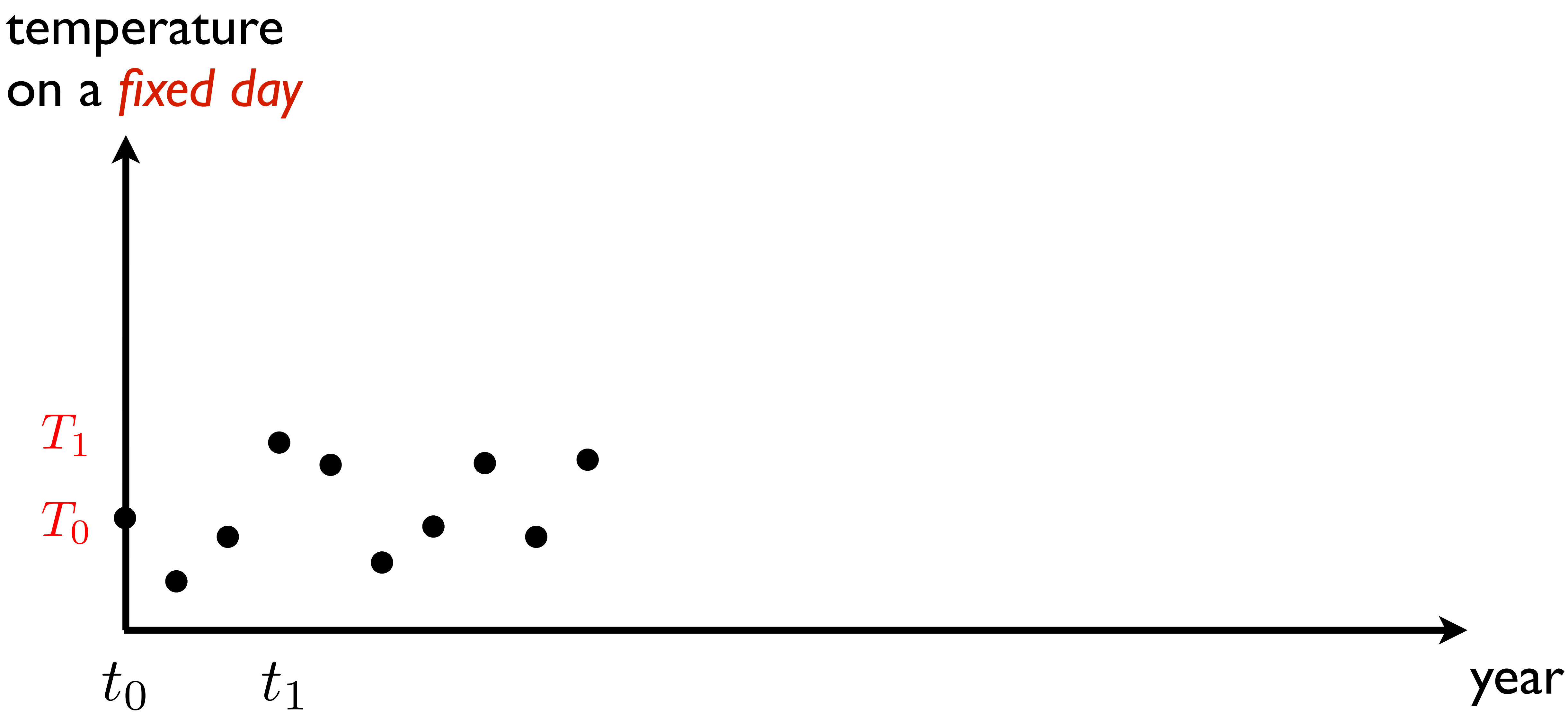
assumption: *daily temperatures are iid continuous variables*



Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*



Evolution of Temperature Records

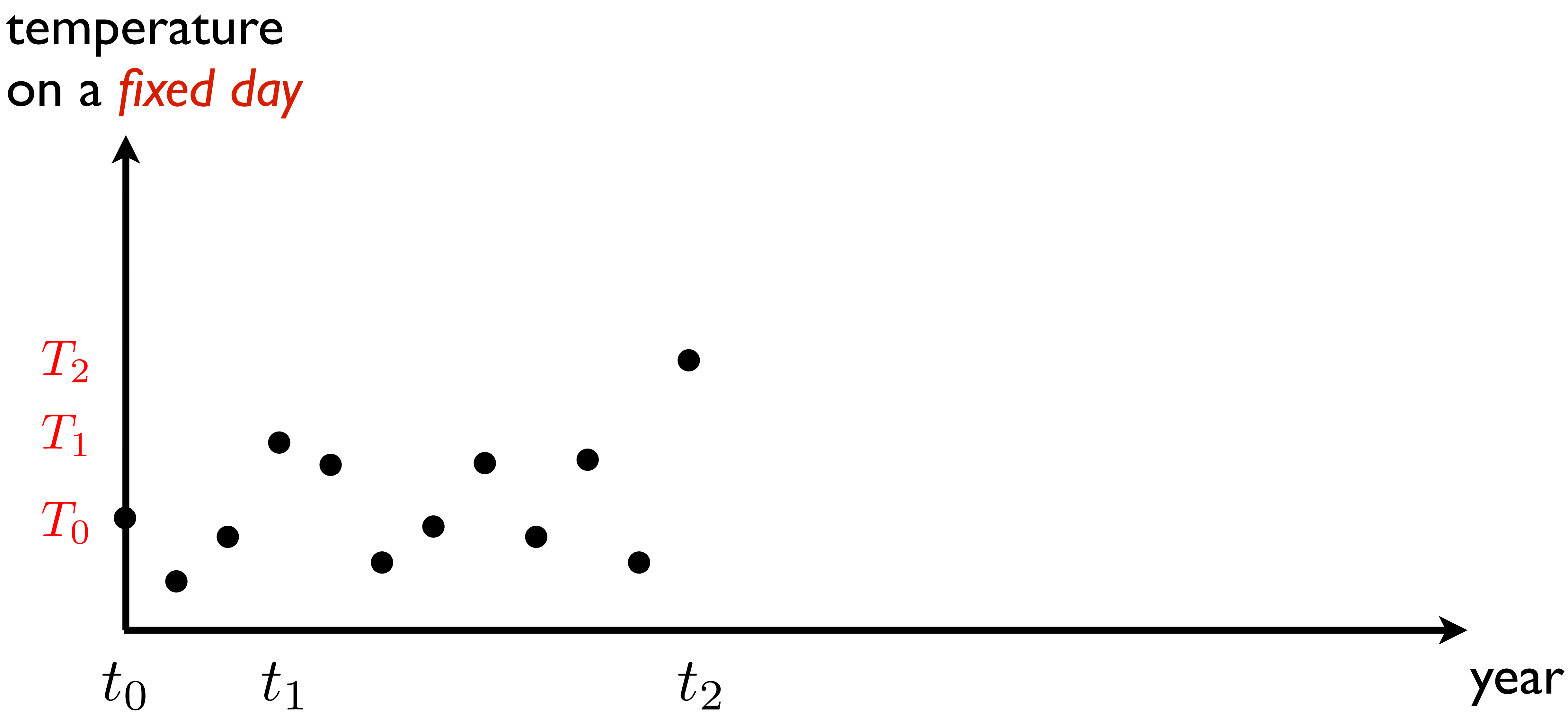
(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*

Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

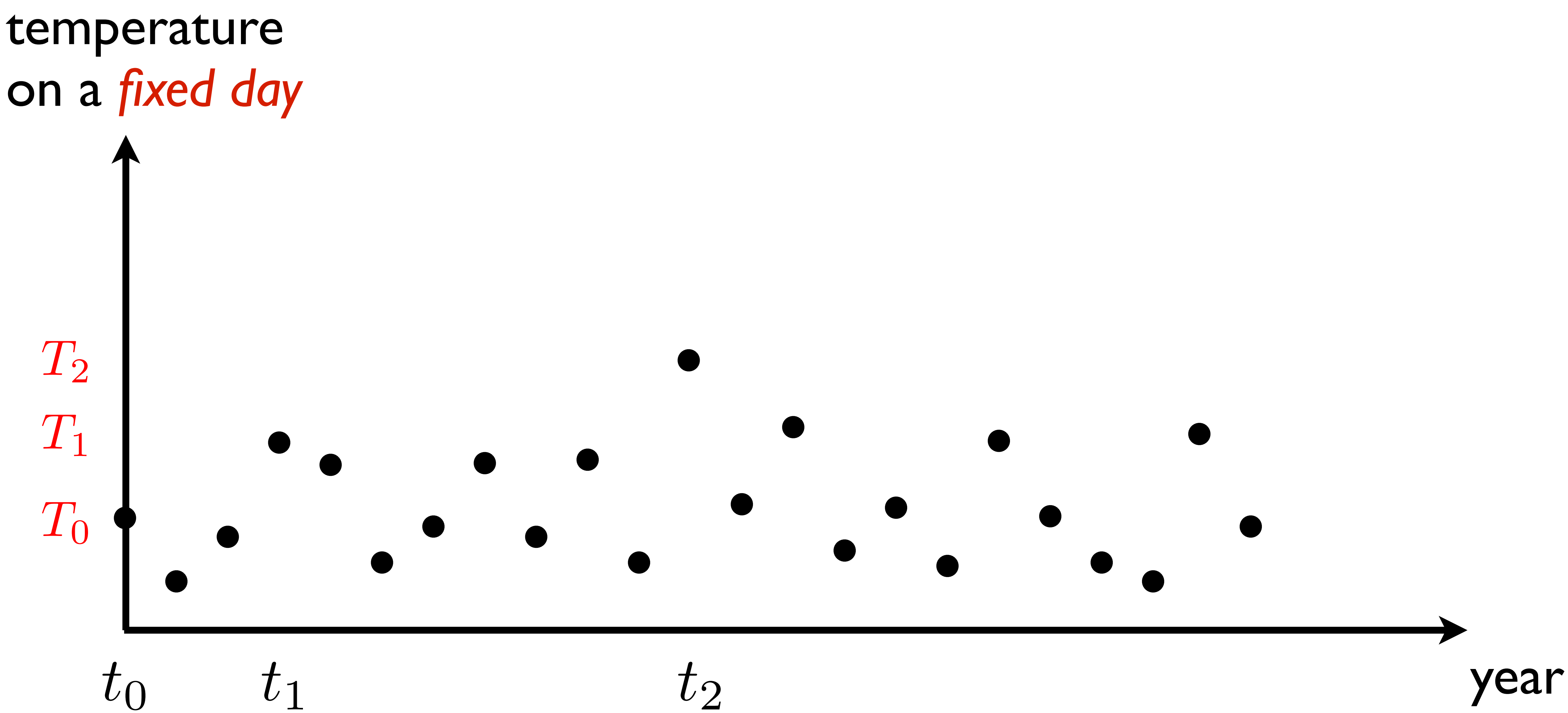
assumption: *daily temperatures are iid continuous variables*



Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

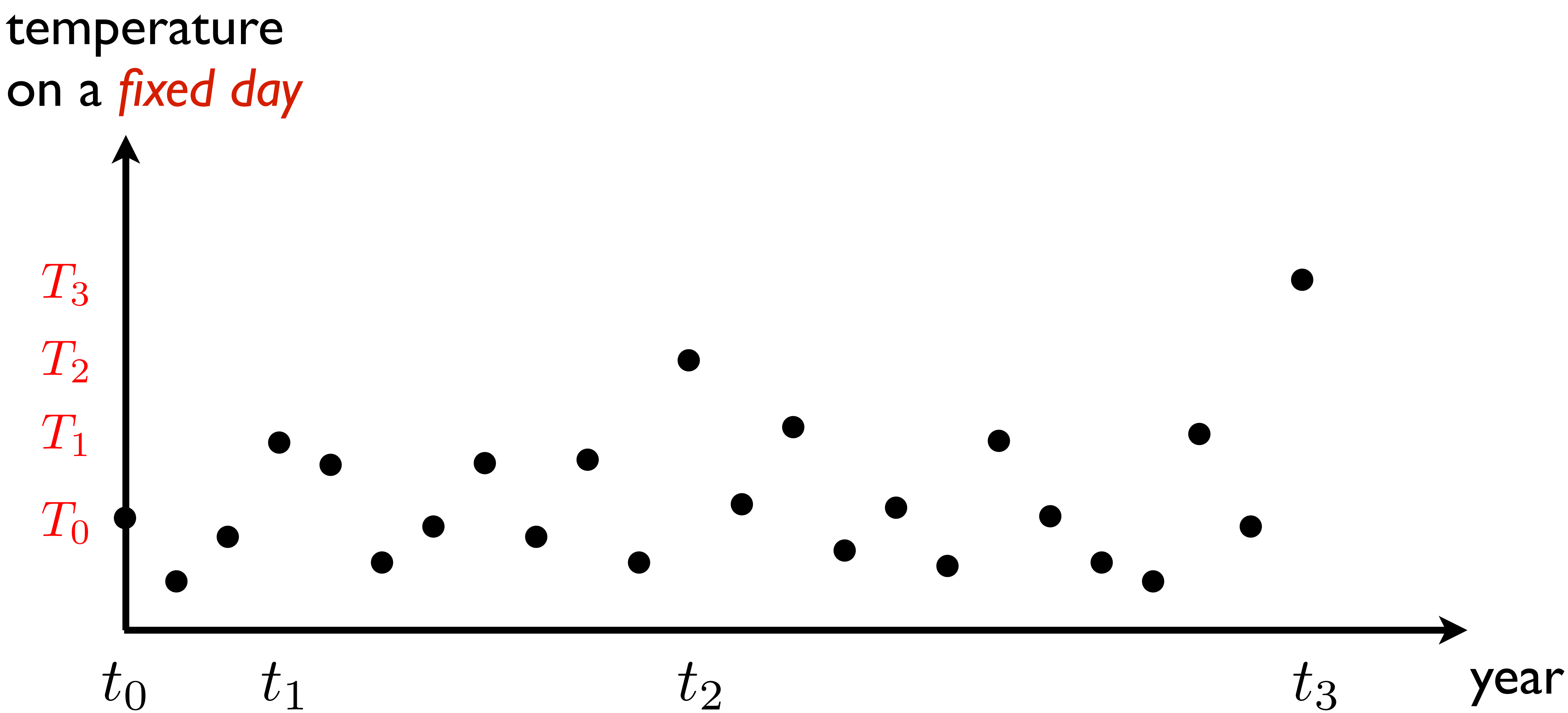
assumption: *daily temperatures are iid continuous variables*



Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

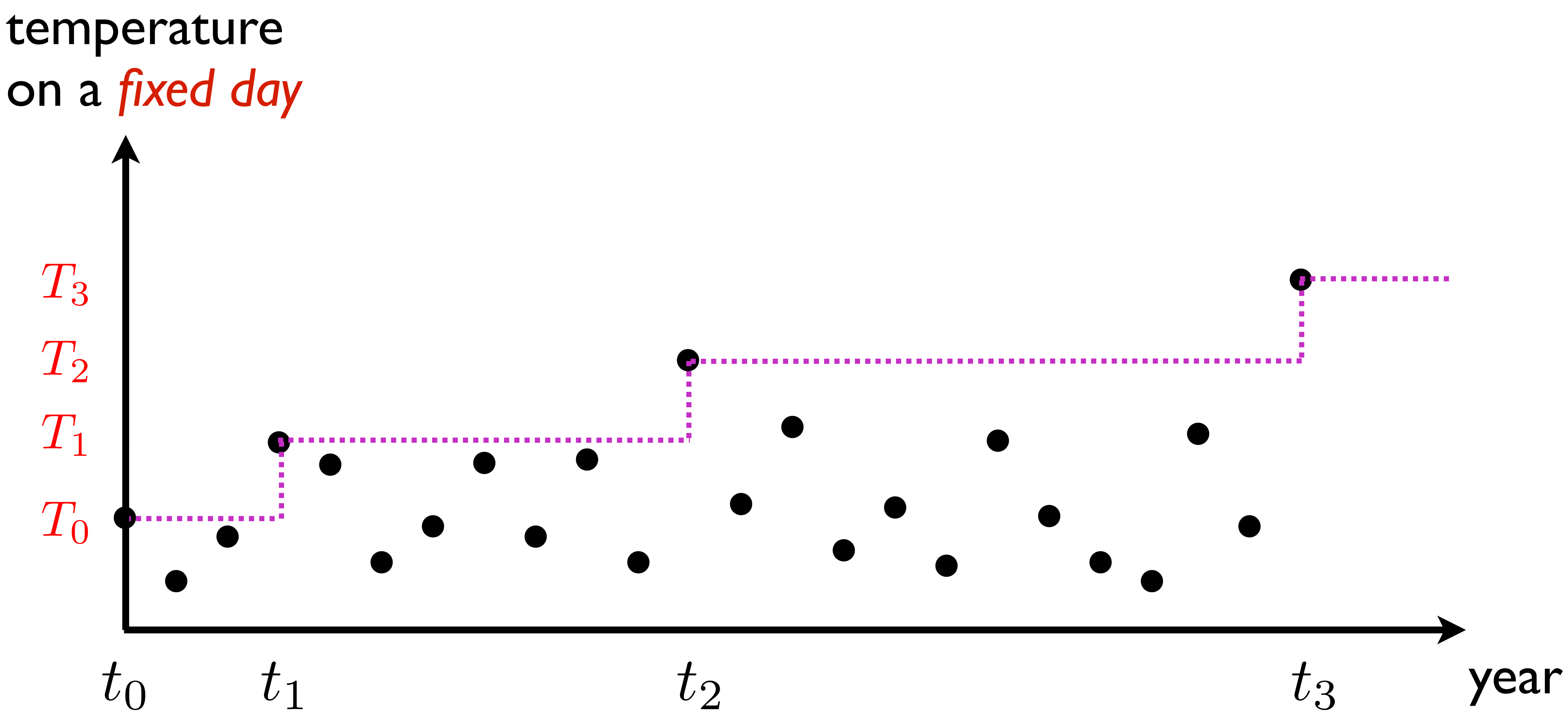
assumption: *daily temperatures are iid continuous variables*



Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

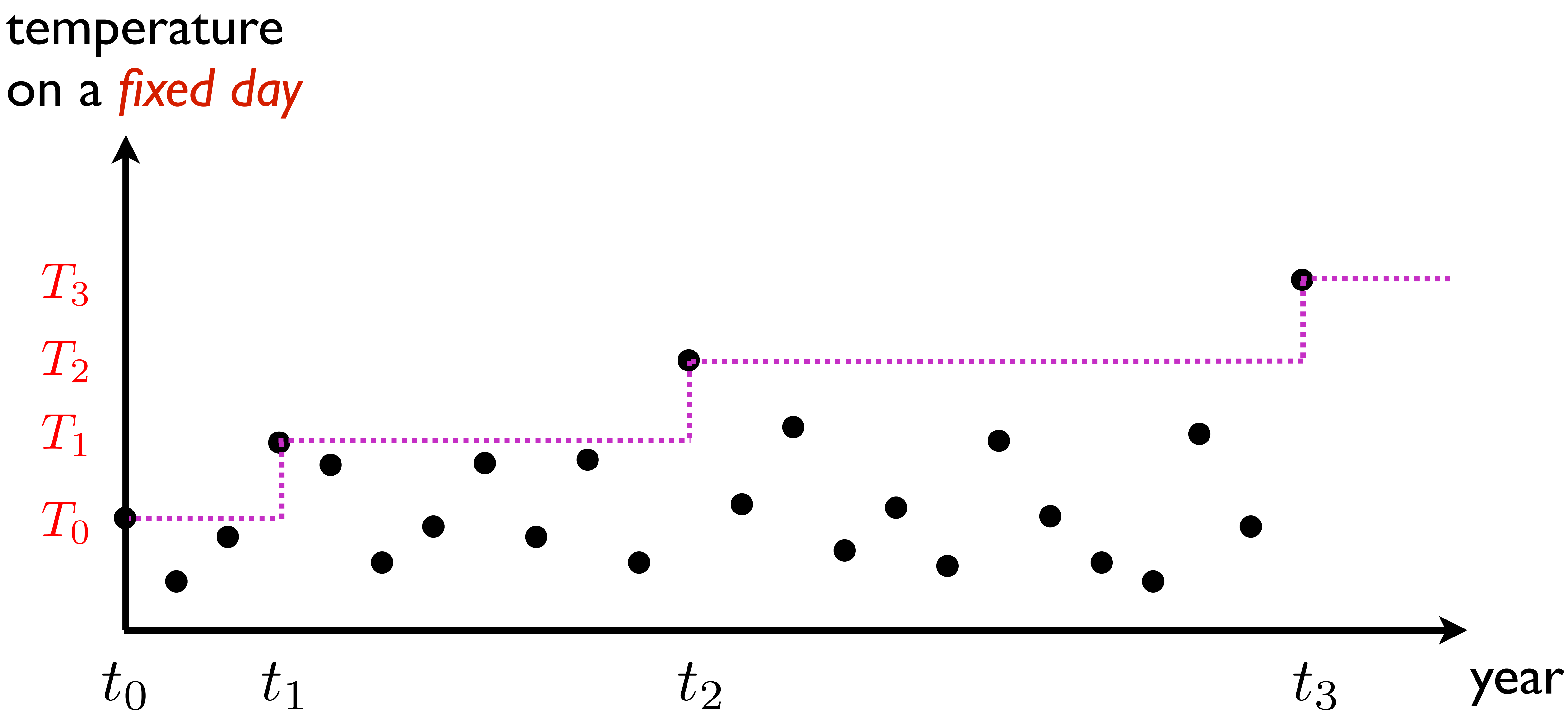
assumption: *daily temperatures are iid continuous variables*



Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*

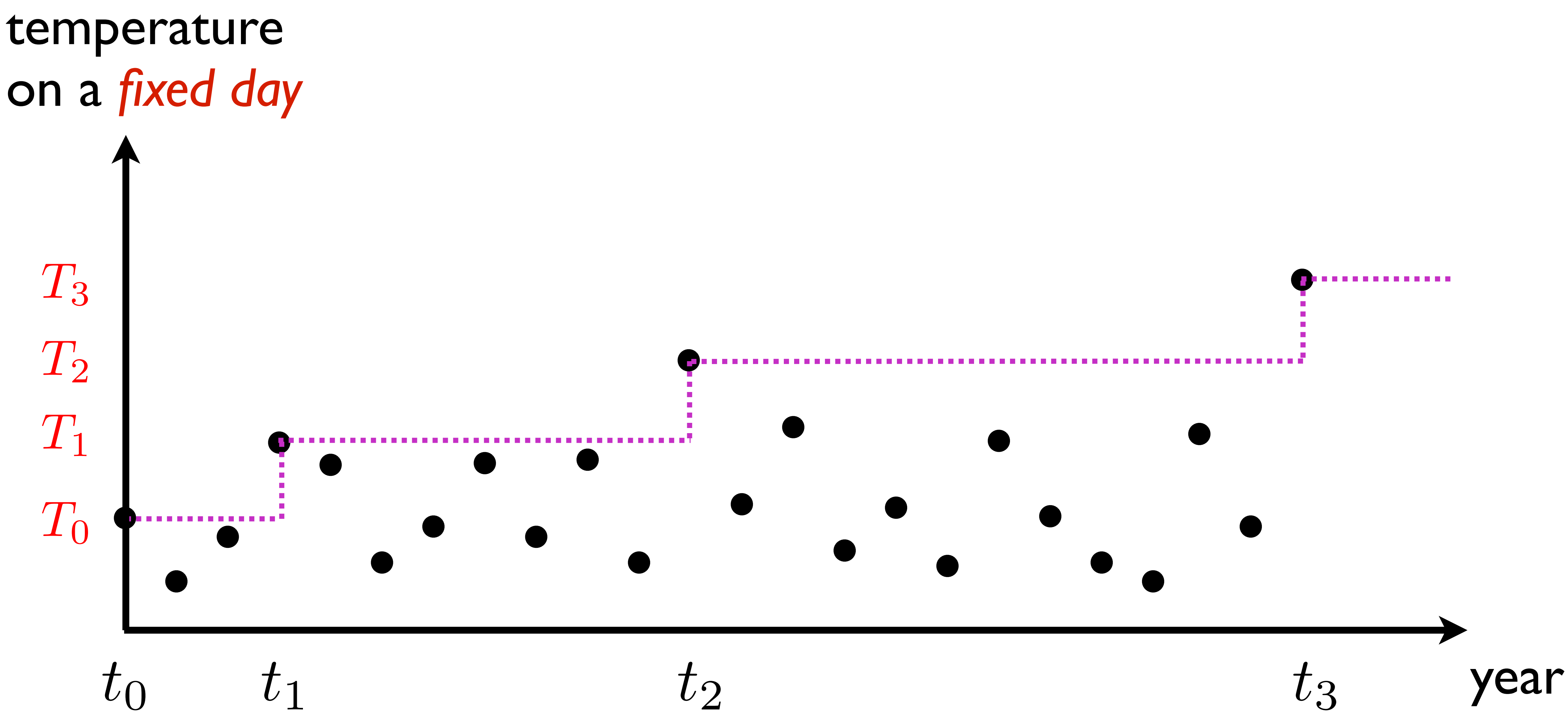


Goal: compute $T_k, \mathcal{P}_k(T)$ $t_k, p_k(t)$

Evolution of Temperature Records

(Arnold et al., *Records*, 1998)

assumption: *daily temperatures are iid continuous variables*



Goal: compute $T_k, \mathcal{P}_k(T)$ $t_k, p_k(t)$

distribution independent!

Record Time Statistics

(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)

Record Time Statistics

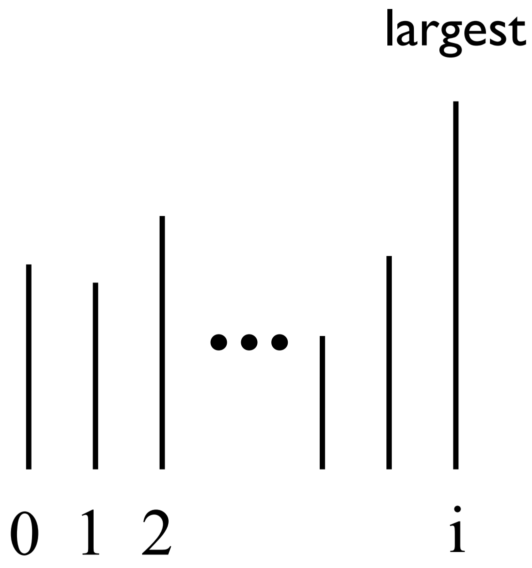
(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)

define $$\sigma_i = \begin{cases} 1 & \text{if record in } i^{\text{th}} \text{ year} \\ 0 & \text{otherwise} \end{cases}$$

Record Time Statistics

(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)

define $$\sigma_i = \begin{cases} 1 & \text{if record in } i^{\text{th}} \text{ year} \\ 0 & \text{otherwise} \end{cases}$$



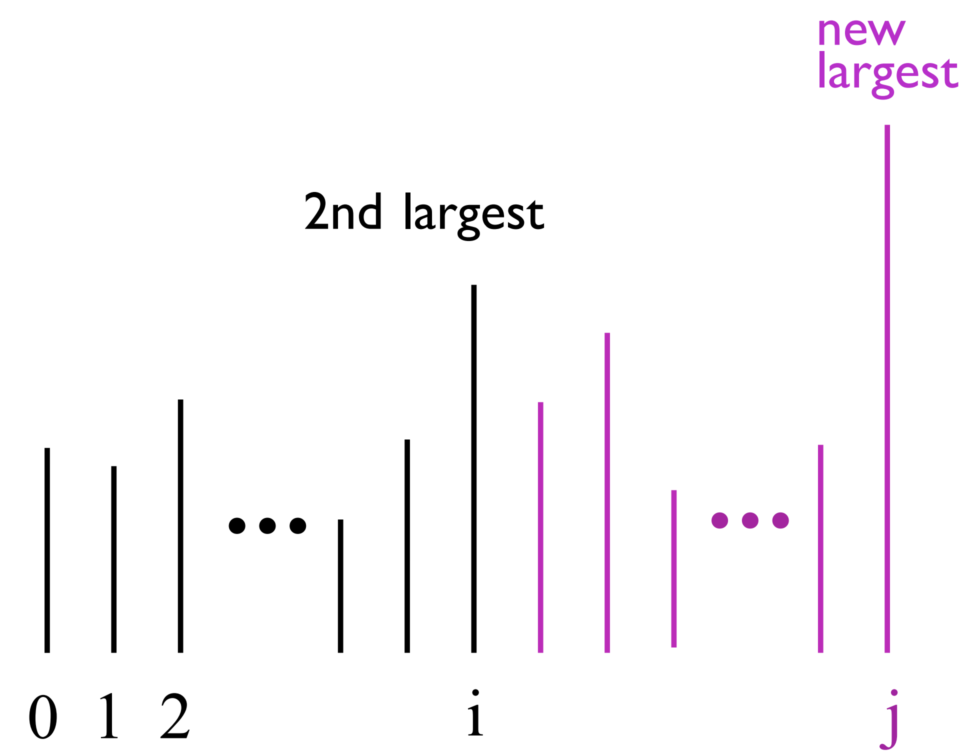
then $$\langle \sigma_i \rangle = \frac{1}{i+1}$$

Record Time Statistics

(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)

define $\sigma_i = \begin{cases} 1 & \text{if record in } i^{\text{th}} \text{ year} \\ 0 & \text{otherwise} \end{cases}$

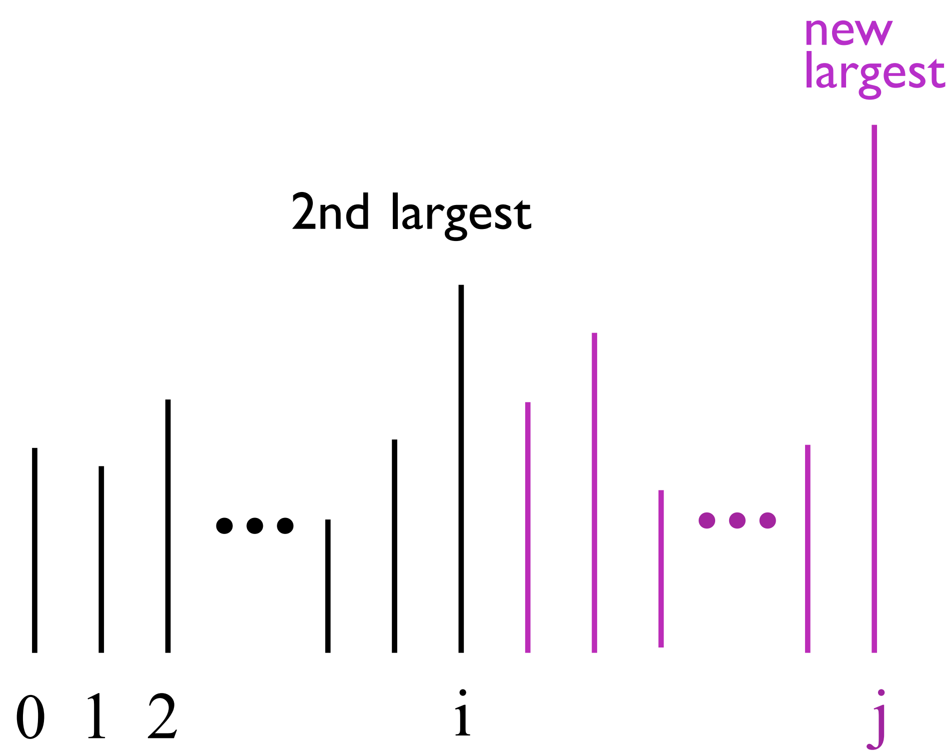
then $\langle \sigma_i \rangle = \dfrac{1}{i+1}$ $\quad \langle \sigma_i \sigma_j \rangle = \langle \sigma_i \rangle \langle \sigma_j \rangle$



Record Time Statistics

(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)

define $\sigma_i = \begin{cases} 1 & \text{if record in } i^{\text{th}} \text{ year} \\ 0 & \text{otherwise} \end{cases}$



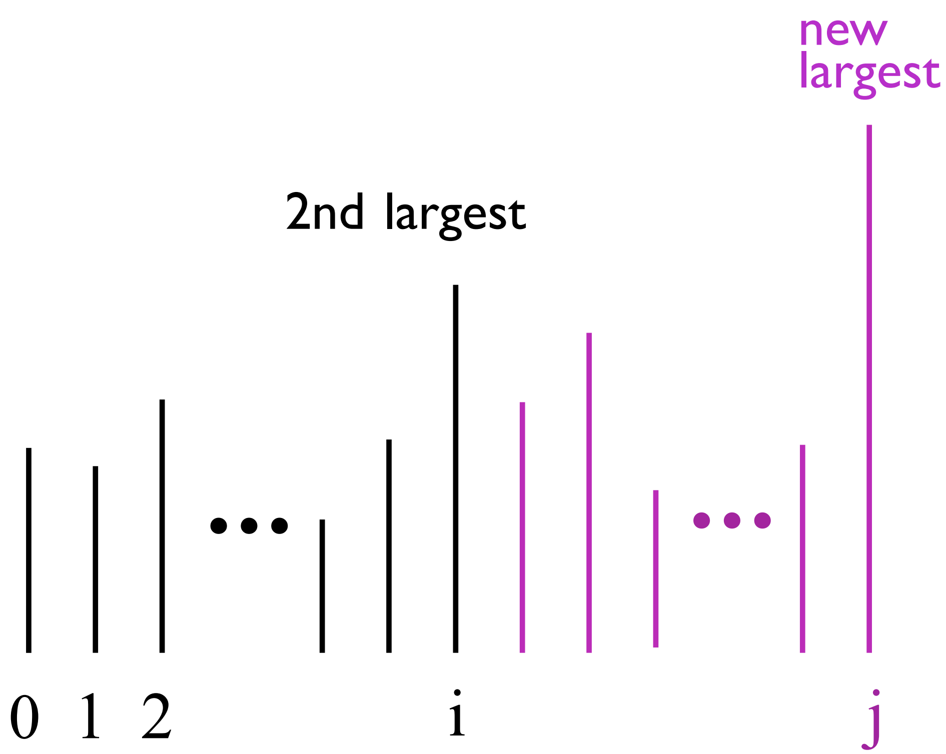
then $\langle \sigma_i \rangle = \dfrac{1}{i+1}$ $\quad \langle \sigma_i \sigma_j \rangle = \langle \sigma_i \rangle \langle \sigma_j \rangle$

therefore $\langle n(t) \rangle = \sum_{i=1}^{t} \langle \sigma_i \rangle \sim \ln t$ $\quad P(n) = \dfrac{(\ln t)^n}{n!} e^{-\ln t}$

Record Time Statistics

(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)

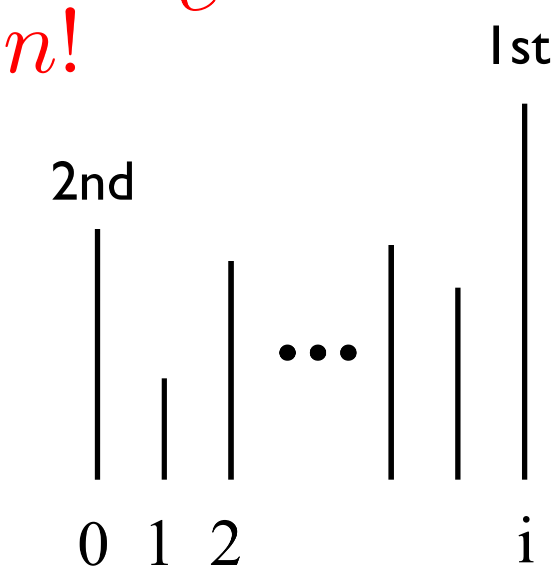
define $\sigma_i = \begin{cases} 1 & \text{if record in } i^{\text{th}} \text{ year} \\ 0 & \text{otherwise} \end{cases}$



then $\langle \sigma_i \rangle = \dfrac{1}{i+1}$ $\langle \sigma_i \sigma_j \rangle = \langle \sigma_i \rangle \langle \sigma_j \rangle$

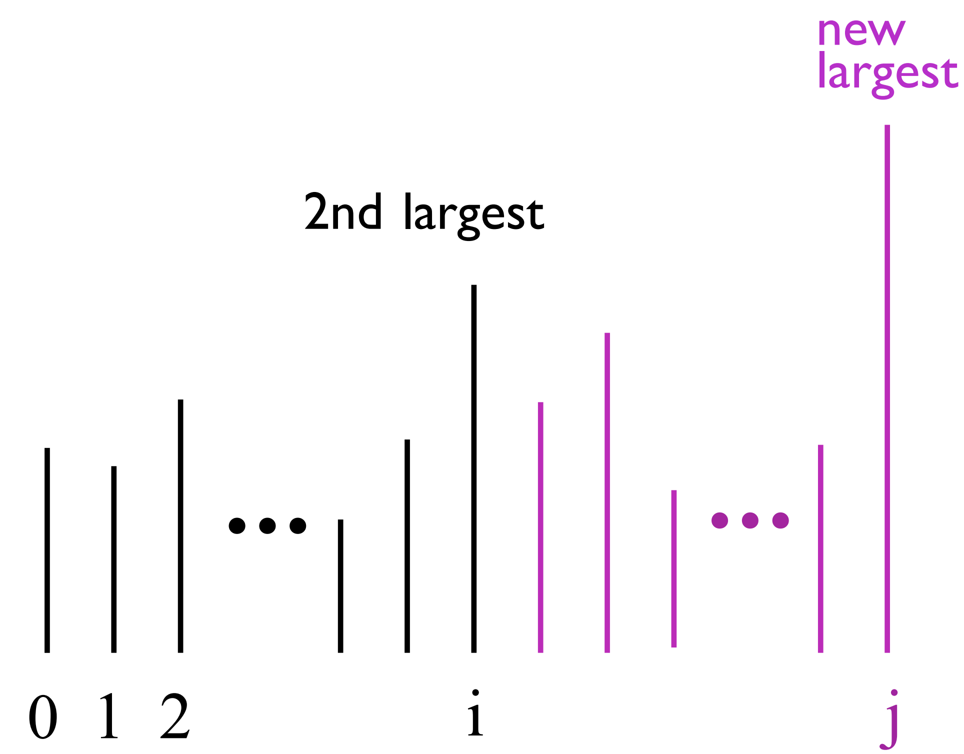
therefore $\langle n(t) \rangle = \sum_{i=1}^{t} \langle \sigma_i \rangle \sim \ln t$ $P(n) = \dfrac{(\ln t)^n}{n!} e^{-\ln t}$

probability that current record is broken in ith year:



Record Time Statistics

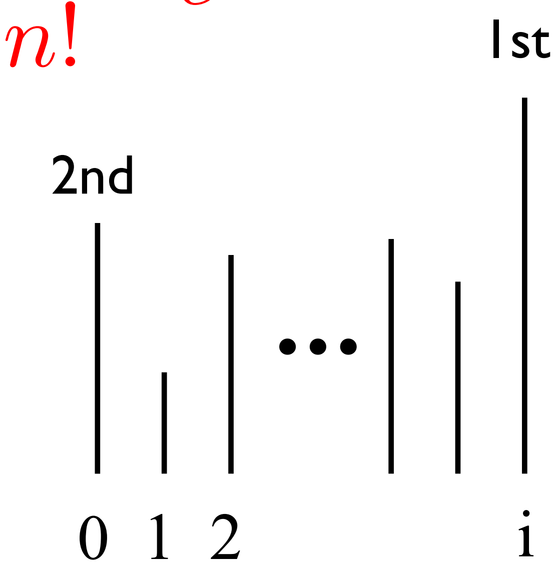
(Glick 1978, Sibani et al 1997, Krug & Jain 05, Majumdar)



define $\sigma_i = \begin{cases} 1 & \text{if record in } i^{\text{th}} \text{ year} \\ 0 & \text{otherwise} \end{cases}$

then $\langle \sigma_i \rangle = \dfrac{1}{i+1}$ $\quad \langle \sigma_i \sigma_j \rangle = \langle \sigma_i \rangle \langle \sigma_j \rangle$

therefore $\langle n(t) \rangle = \displaystyle\sum_{i=1}^{t} \langle \sigma_i \rangle \sim \ln t$ $\quad P(n) = \dfrac{(\ln t)^n}{n!} e^{-\ln t}$



probability that current record is broken in ith year:

$$= \frac{1}{i(i+1)} \rightarrow \text{time until next record} = \infty$$

Records If Global Warming Is Occurring

assume temperature distribution *as a soluble example only*

$$p(T;t) = \begin{cases} e^{-(T-vt)} & T > vt \\ 0 & T < vt \end{cases}$$

exceedance probability

$$\begin{aligned} p_>(T_k; t_k + j) &= \int_{T_k}^{\infty} e^{-[T-v(t_k+j)]}\, dT \\ &= e^{-(T_k - vt_k)}\, e^{jv} \equiv X\, e^{jv} \end{aligned}$$

prob of record at year n

$$\begin{aligned} q_n(T_k) &\equiv p_>(T_k)\, p_<(T_k)^{n-1} \\ &\to e^{nv}\, X \prod_{j=1}^{n-1} \underbrace{(1 - e^{jv}\, X)}_{\text{when 0 record must occur}} \end{aligned}$$

(over)estimate of record time

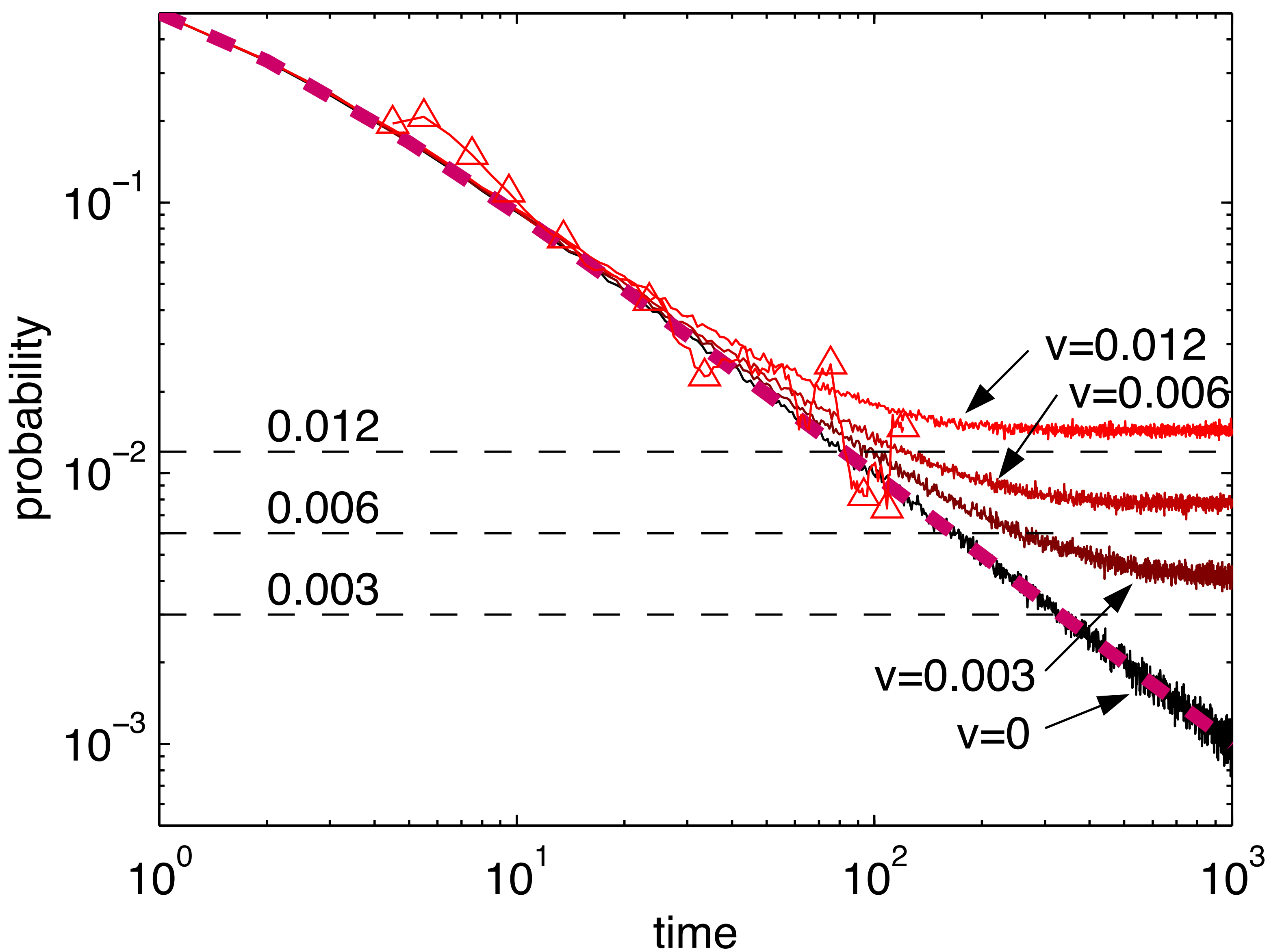
$$e^{jv} X = 1 \quad \to \quad (t_{k+1} - t_k)v = T_k - vt_k$$

$$\to t_k \sim \frac{k}{v}$$

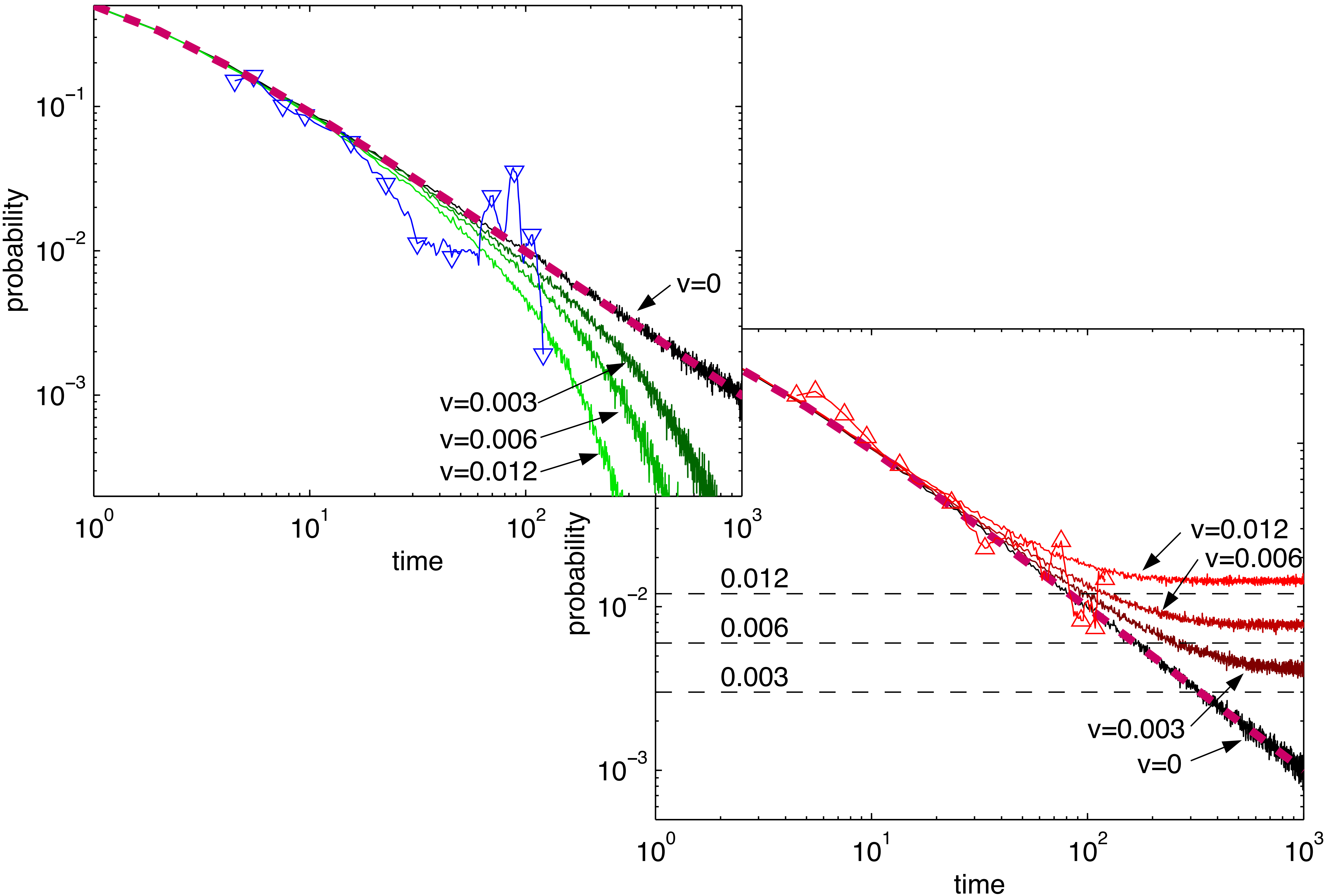
records ultimately occur at constant rate

(Ballerini & Resnick, 85; Borokov, 99)

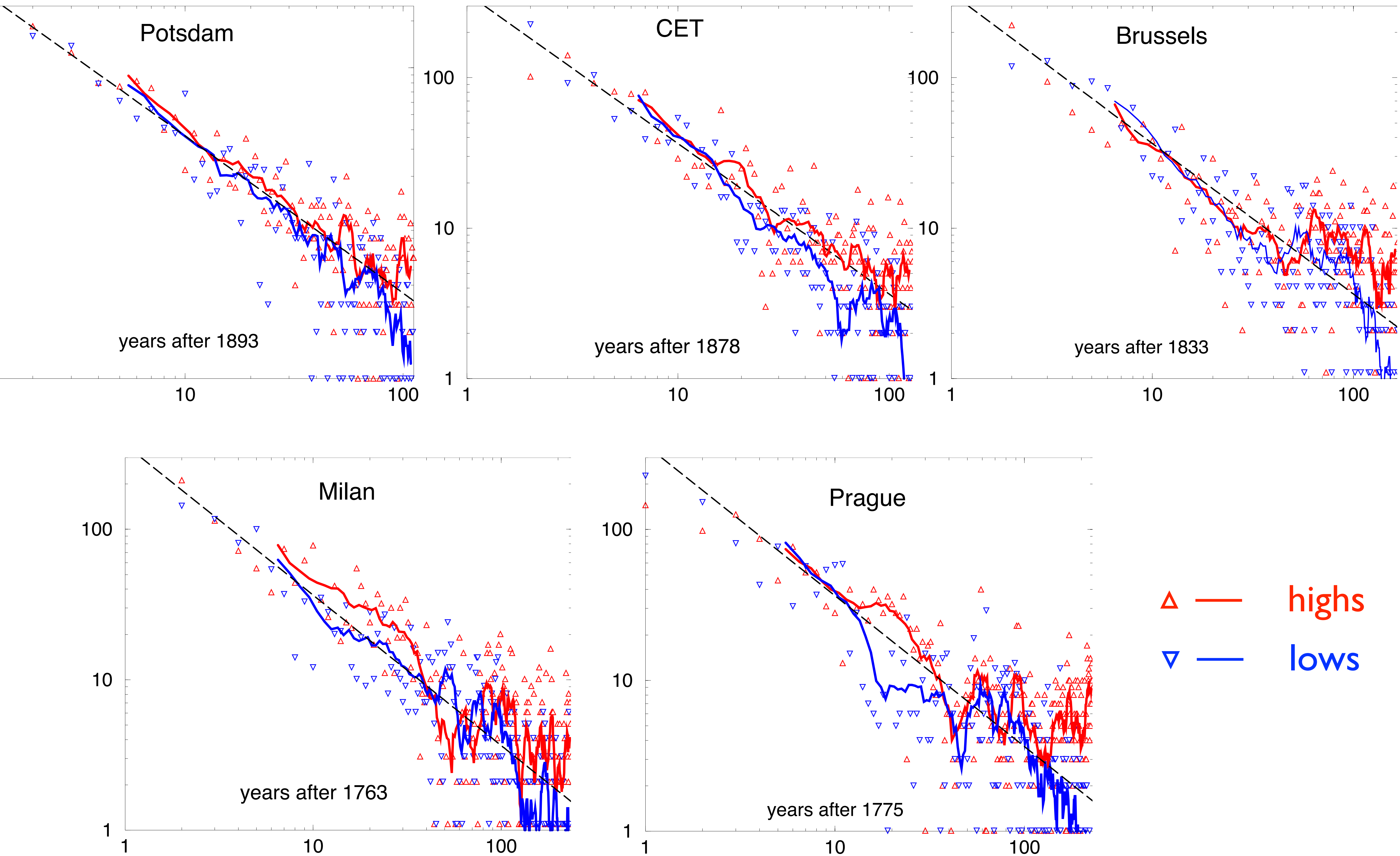
Frequency of Record High Temperatures



Frequency of Record Low Temperatures



Record Frequencies: Multiple Stations



The Luckiest

with C. Sire

Eur. Phys. J B to appear very soon

Statistics of team win/loss records

Consecutive-game winning/losing streaks

Fun Facts About Team Win/Loss Records

Best all-time record:

1906 Chicago Cubs: 116/36 .763

Worst all-time record:

1916 Philadelphia Athletics: 36/117 .235

Historically most successful 1901-1960:

NY Highlanders/Yankees 1903-1960: 5105/8809 .5795

Historically least successful 1901-1960:

Philadelphia Phillies 1901-1960: 3828/8786 .4357

Boston Braves 1905-1952: 2627/5923 .4435

Bradley-Terry Competition Model

Zermelo (1929)
Bradley & Terry (1952)

Bradley-Terry Competition Model

Zermelo (1929)
Bradley & Terry (1952)

- each team i has strength x_i, $i=1,2,....,n$
  (**fixed** in each season, can change between seasons)

Bradley-Terry Competition Model

Zermelo (1929)
Bradley & Terry (1952)

- each team i has strength x_i, i=1,2,....,n
  (**fixed** in each season, can change between seasons)

- probability that team i beats j:

$$p_{ij} = \frac{x_i}{x_i + x_j}$$

actually, assume only:

$$p_{ij} = \frac{f(x_i)}{f(x_i) + f(x_j)}$$

numerical studies: Anthology of Statistics in Sports, Albert et al. (2005)

Bradley-Terry Competition Model

Zermelo (1929)
Bradley & Terry (1952)

- each team i has strength x_i, i=1,2,....,n
  (**fixed** in each season, can change between seasons)

- probability that team i beats j:

$$p_{ij} = \frac{x_i}{x_i + x_j}$$

actually, assume only:

$$p_{ij} = \frac{f(x_i)}{f(x_i) + f(x_j)}$$

numerical studies: Anthology of Statistics in Sports, Albert et al. (2005)

- **uniform** strength distribution:

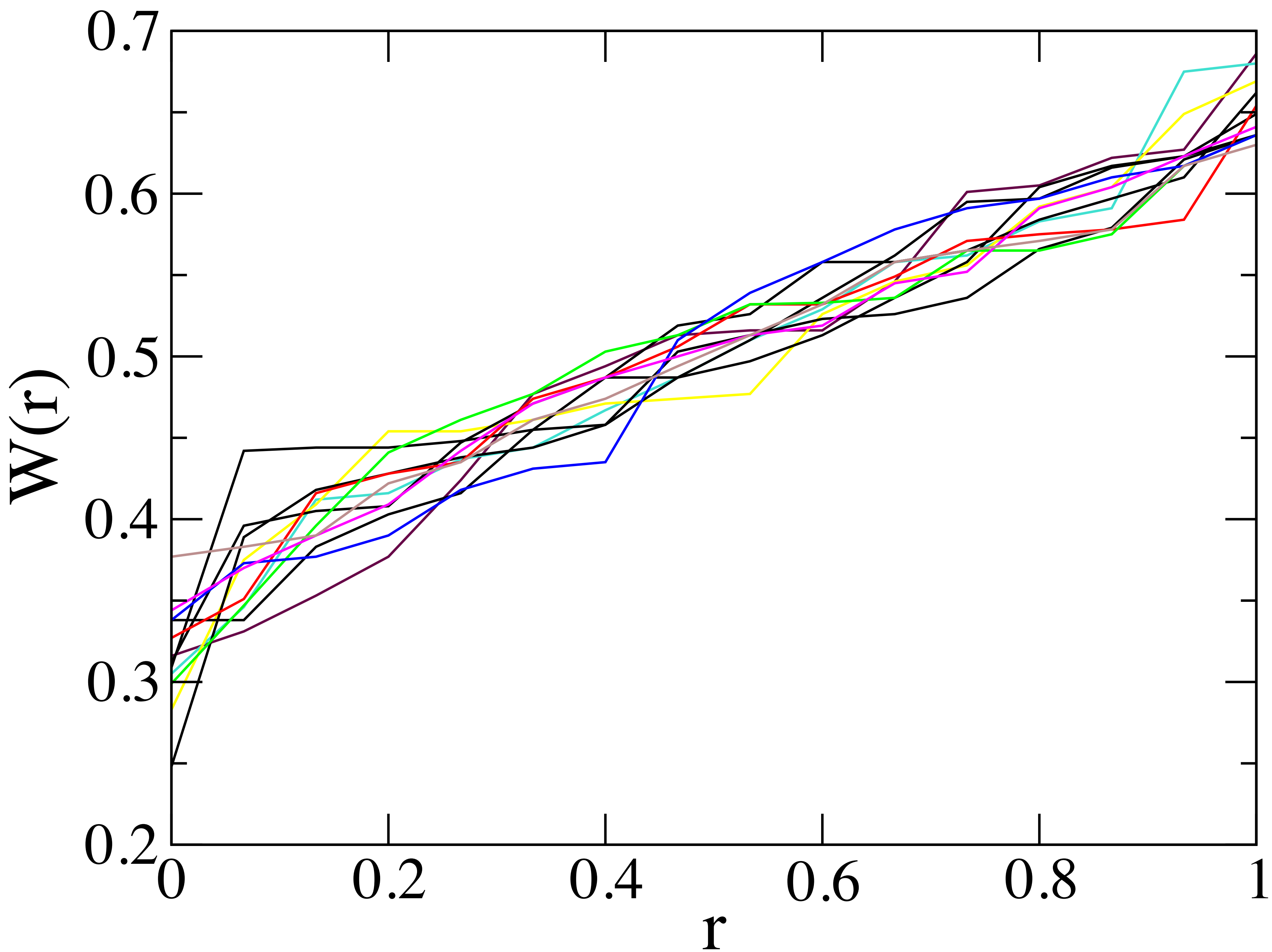
$$x \in [\epsilon, 1] \text{ with } 0 \leq \epsilon \leq 1$$

fundamental parameter: ε

common assumption: log-normal distribution

Win Fraction Versus Rank

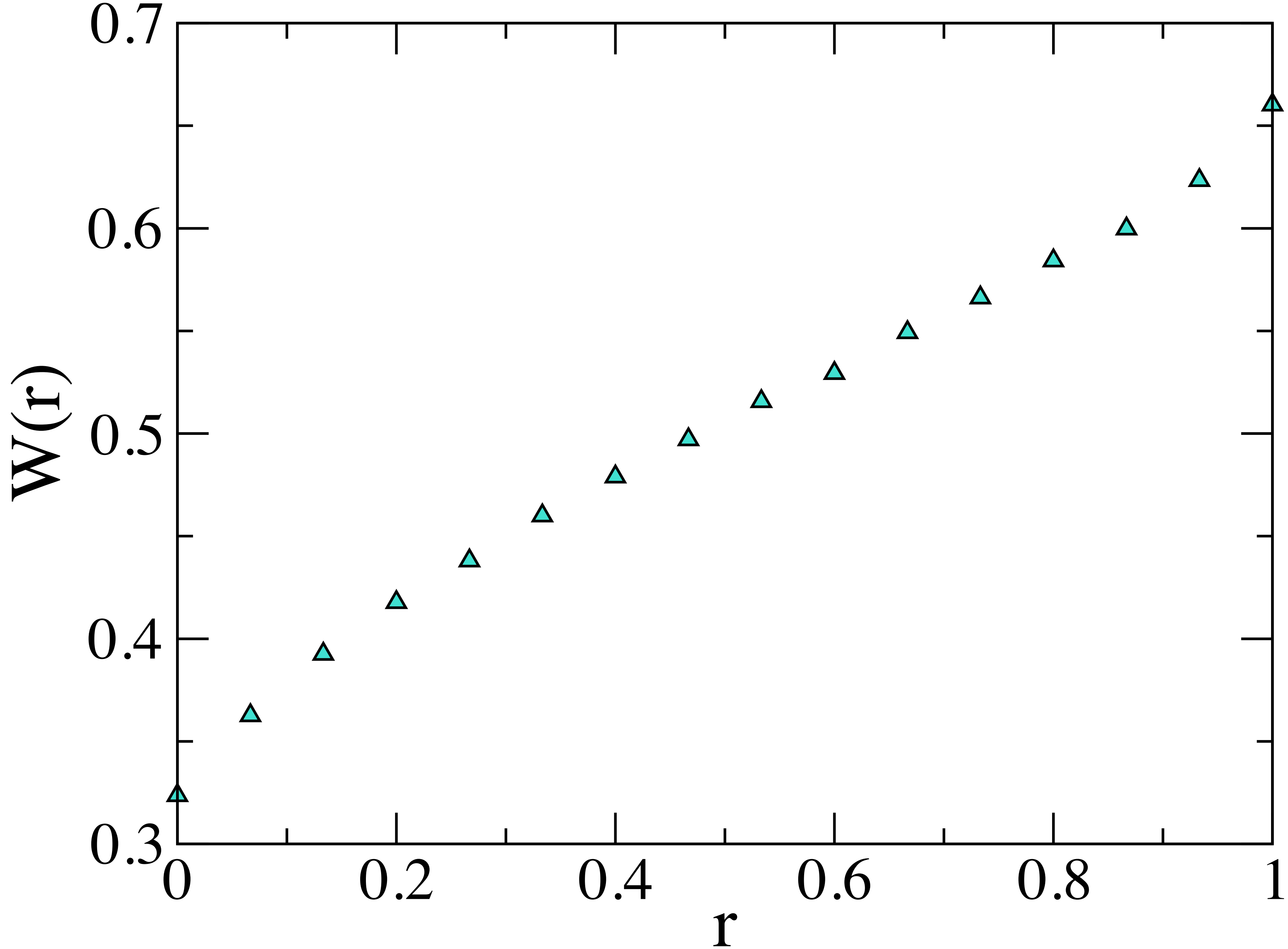
MLB data for 1905, 10, 15,..., 1960



data from www.shrpsports.com

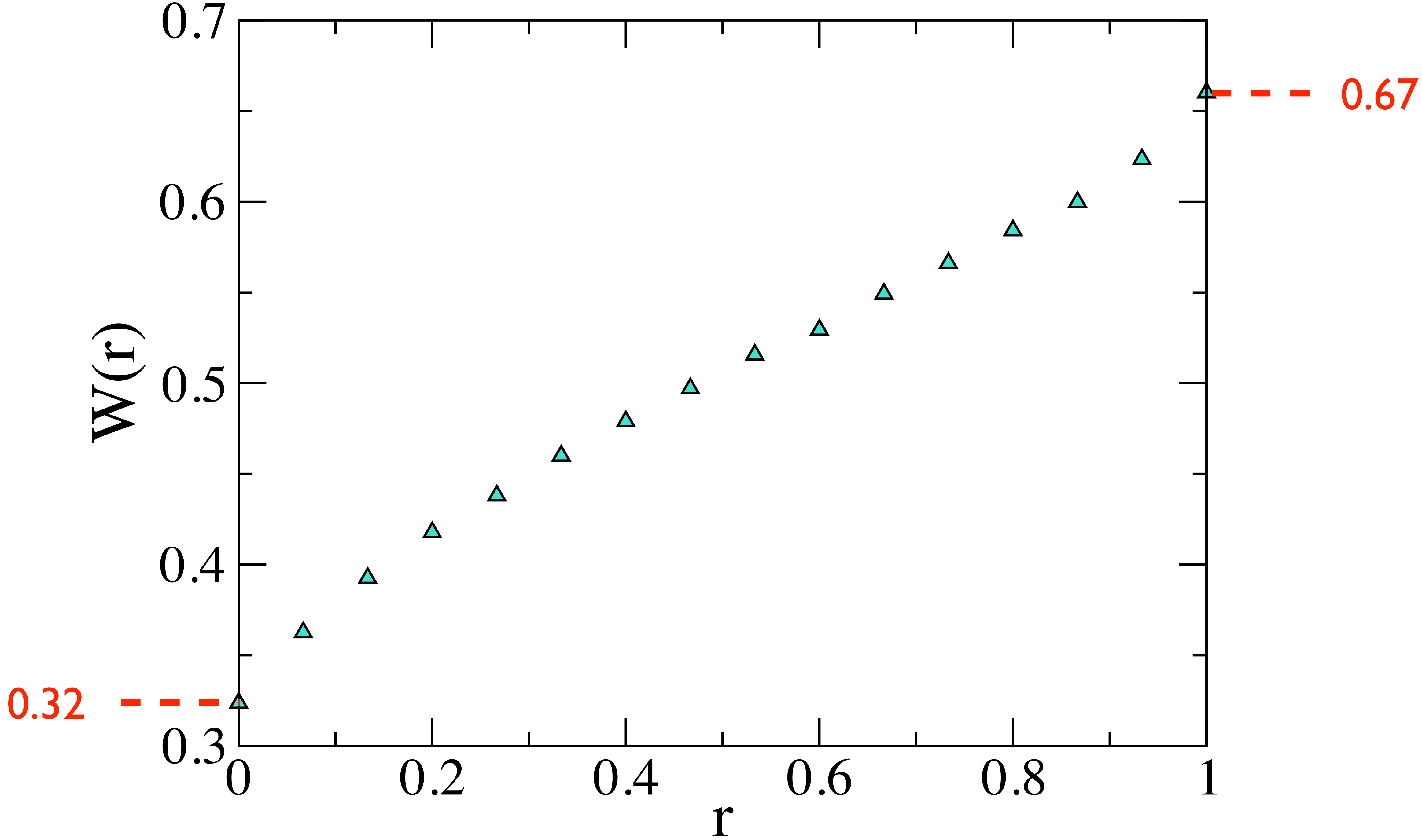
Average Win Fraction Versus Rank

MLB 1901-60



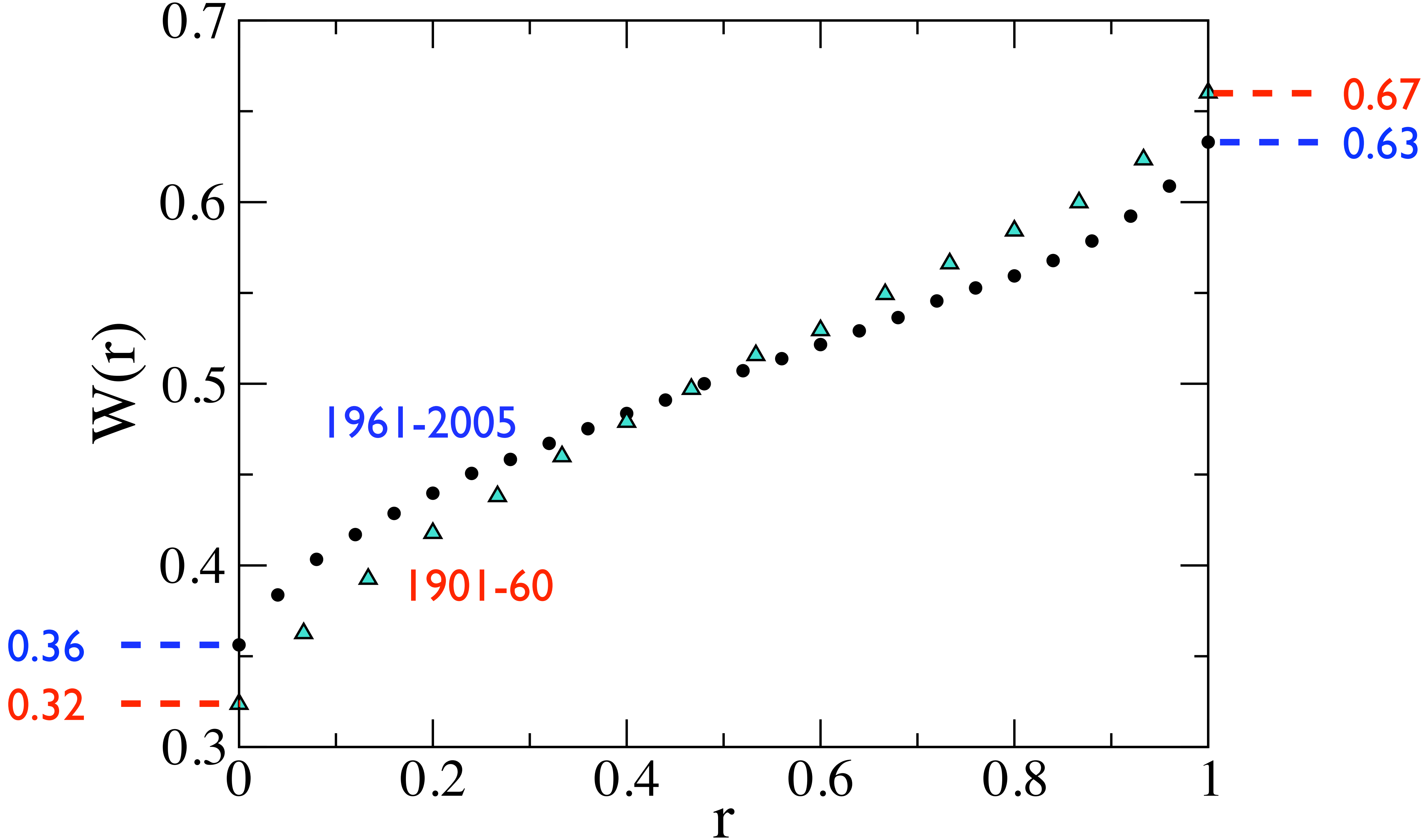
Average Win Fraction Versus Rank

MLB 1901-60



Average Win Fraction Versus Rank

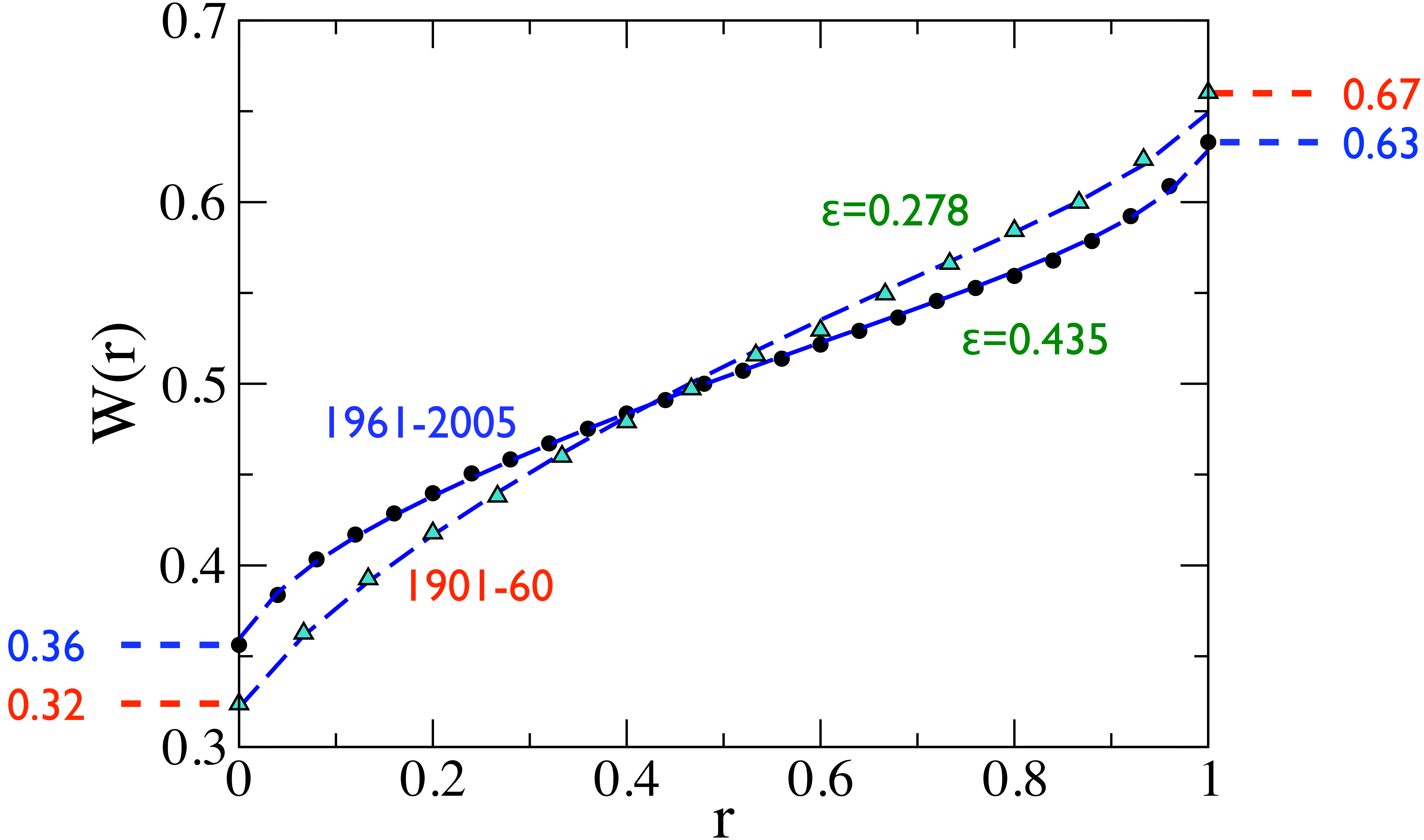
MLB 1901-60, 1961-2005



Average Win Fraction Versus Rank

Simulations of BT model for 10^6 seasons

balanced schedule



Theory for the Average Win Fraction W

Theory for the Average Win Fraction W

for a team of strength x:

$$W(x) \quad = \quad \frac{1}{N}\sum_{j=1}^{N}\frac{x}{x+x_j}$$

Theory for the Average Win Fraction W

for a team of strength x:

$$W(x) = \frac{1}{N}\sum_{j=1}^{N}\frac{x}{x+x_j}$$

$$\to \frac{1}{1-\epsilon}\int_{\epsilon}^{1}\frac{x}{x+y}\,dy$$

- infinite season length
- infinite number of teams
- balanced schedule
- uniform strengths on [ε,1]

Theory for the Average Win Fraction W

for a team of strength x:

$$W(x) = \frac{1}{N}\sum_{j=1}^{N}\frac{x}{x+x_j}$$

$$\to \frac{1}{1-\epsilon}\int_{\epsilon}^{1}\frac{x}{x+y}\,dy$$

$$= \frac{x}{1-\epsilon}\ln\left[\frac{x+1}{x+\epsilon}\right]$$

- infinite season length
- infinite number of teams
- balanced schedule
- uniform strengths on [ε,1]

Theory for the Average Win Fraction W

for a team of strength x:

$$W(x) = \frac{1}{N}\sum_{j=1}^{N}\frac{x}{x+x_j}$$

$$\rightarrow \frac{1}{1-\epsilon}\int_{\epsilon}^{1}\frac{x}{x+y}\,dy$$

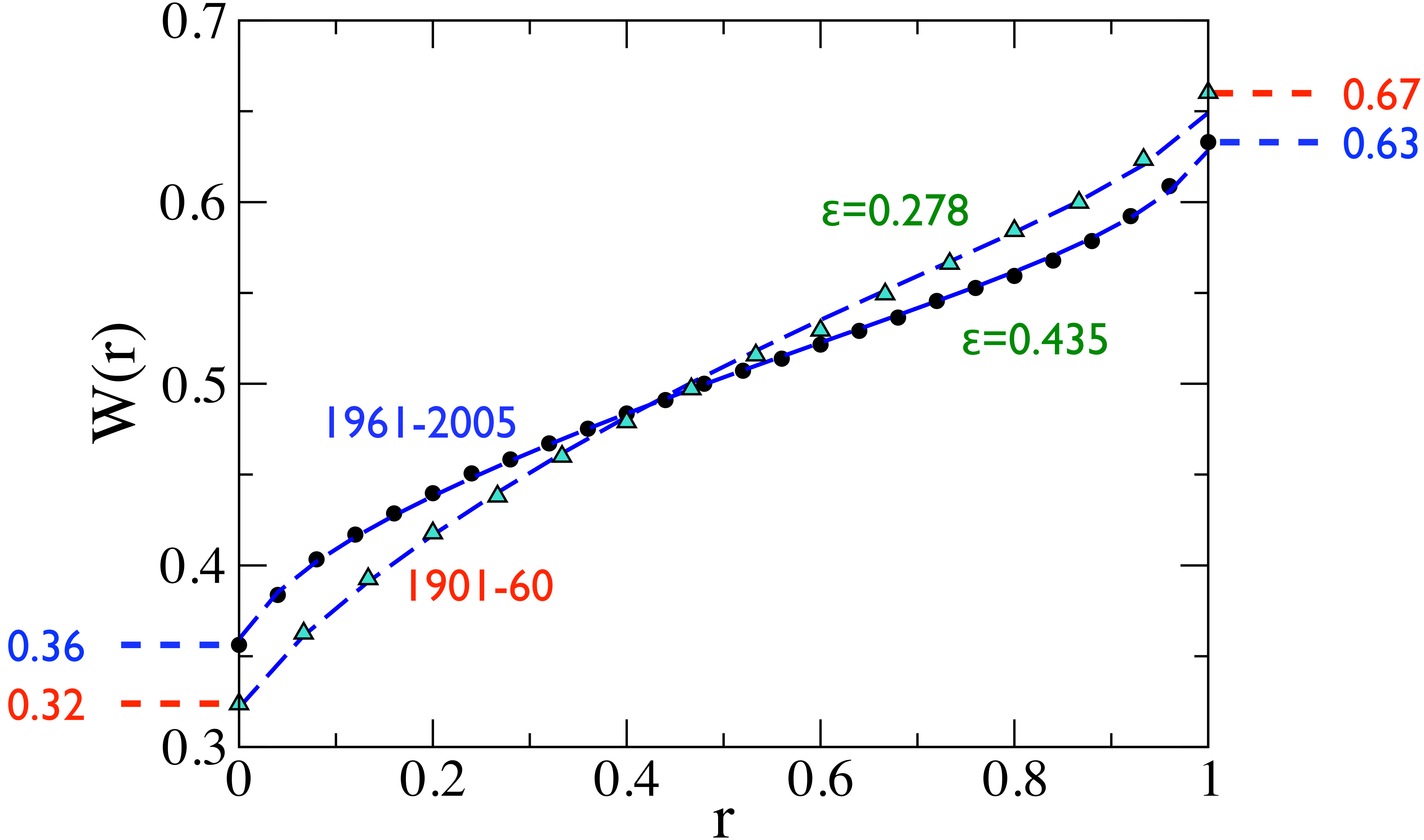
$$= \frac{x}{1-\epsilon}\ln\left[\frac{x+1}{x+\epsilon}\right]$$

- infinite season length
- infinite number of teams
- balanced schedule
- uniform strengths on [ε,1]

strength x → rank r: x=ε+(1-ε)r →

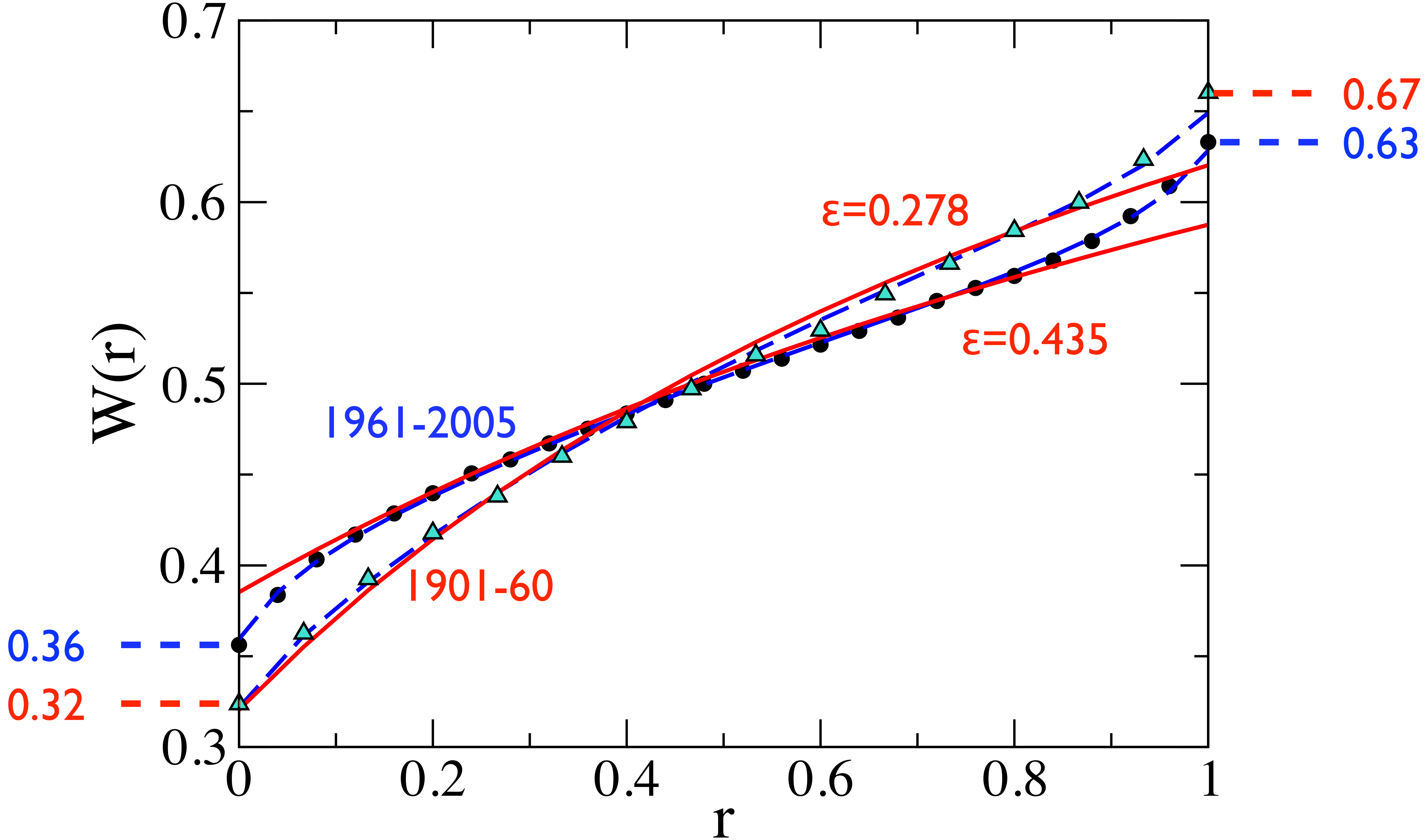
$$W(r) = [\epsilon + (1-\epsilon)r]\,\ln\left[\frac{1+\epsilon+(1-\epsilon)r}{2\epsilon+(1-\epsilon)r}\right]$$

Average Win Fraction Versus Rank

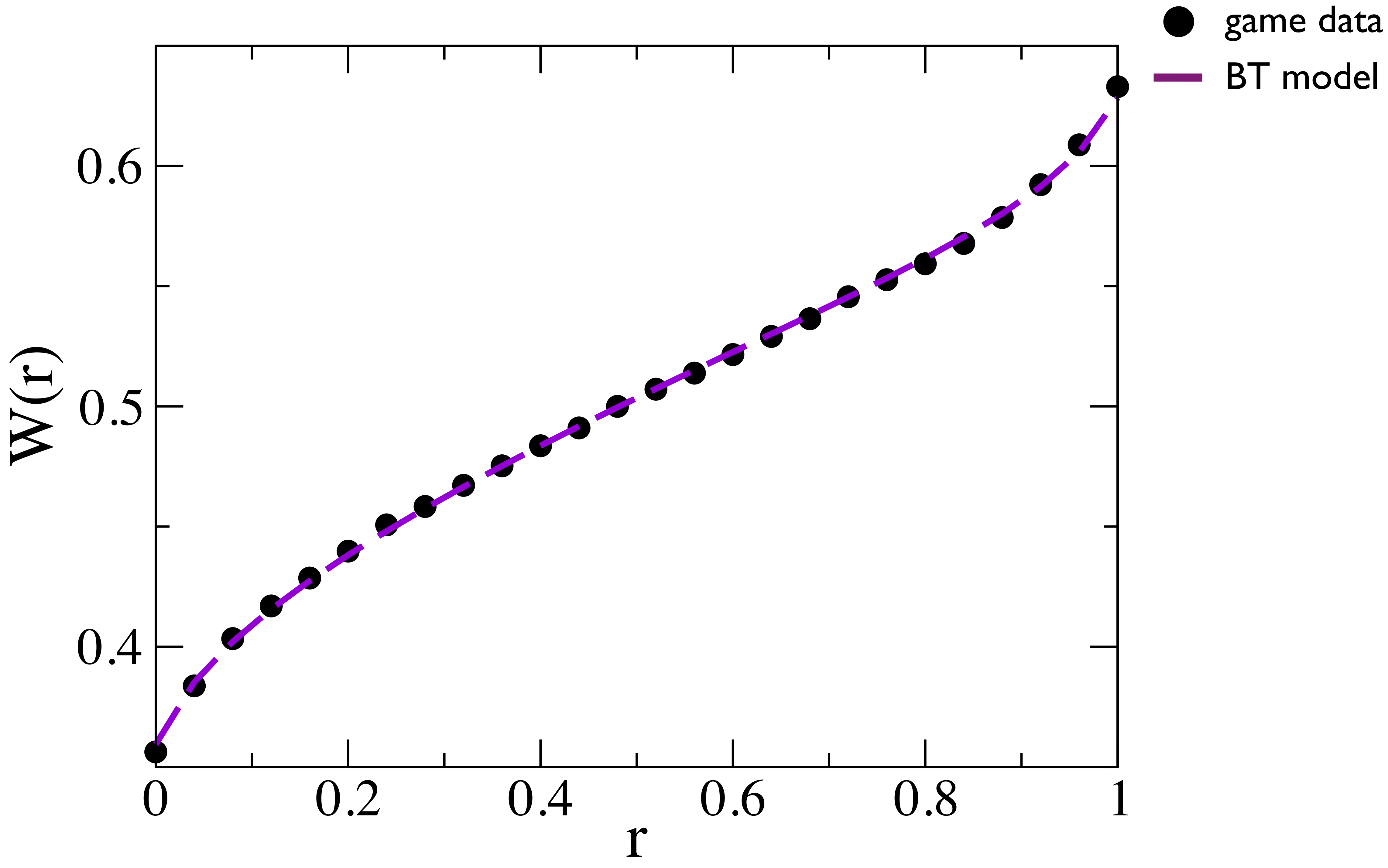


Average Win Fraction Versus Rank

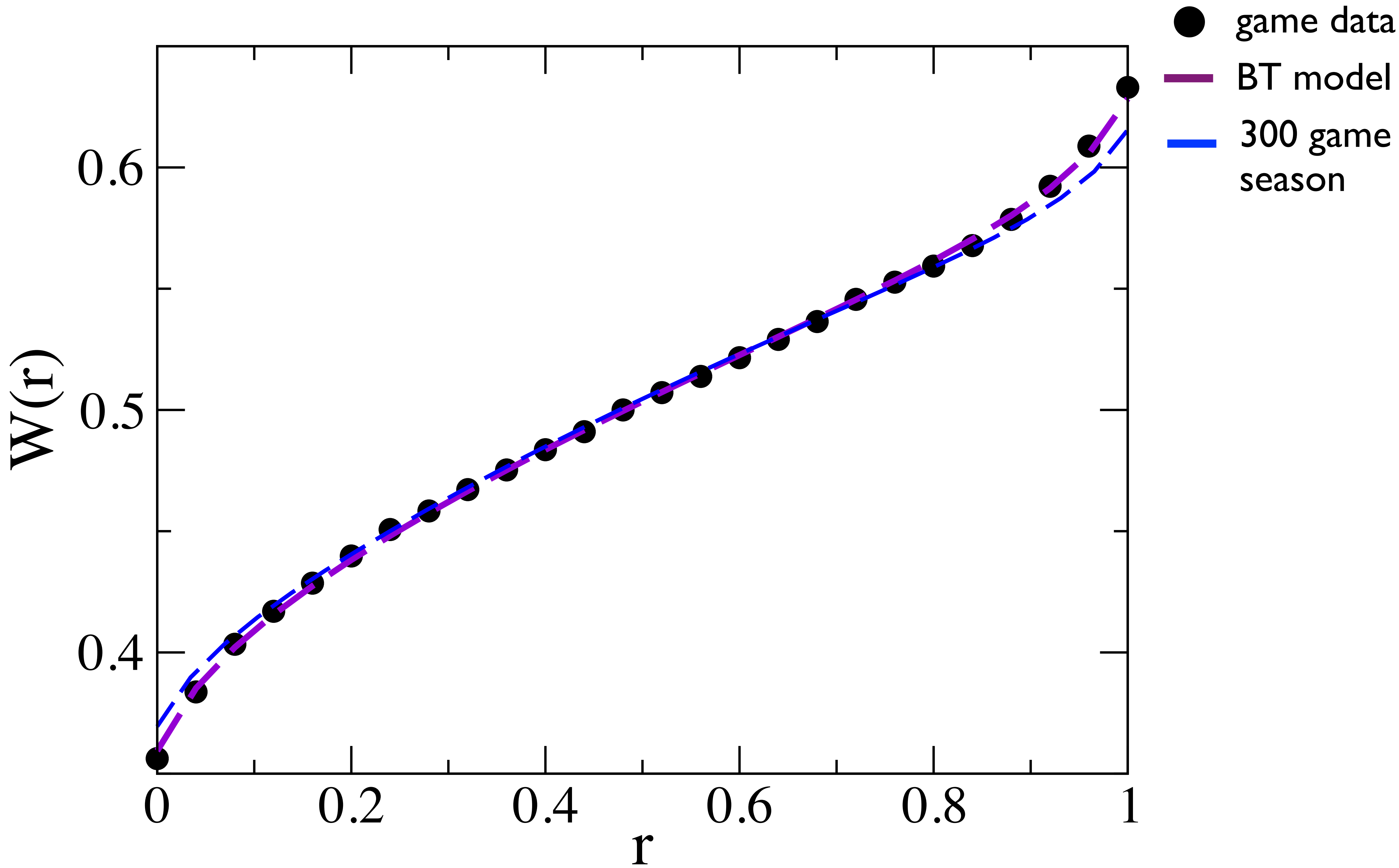
BT model analytics



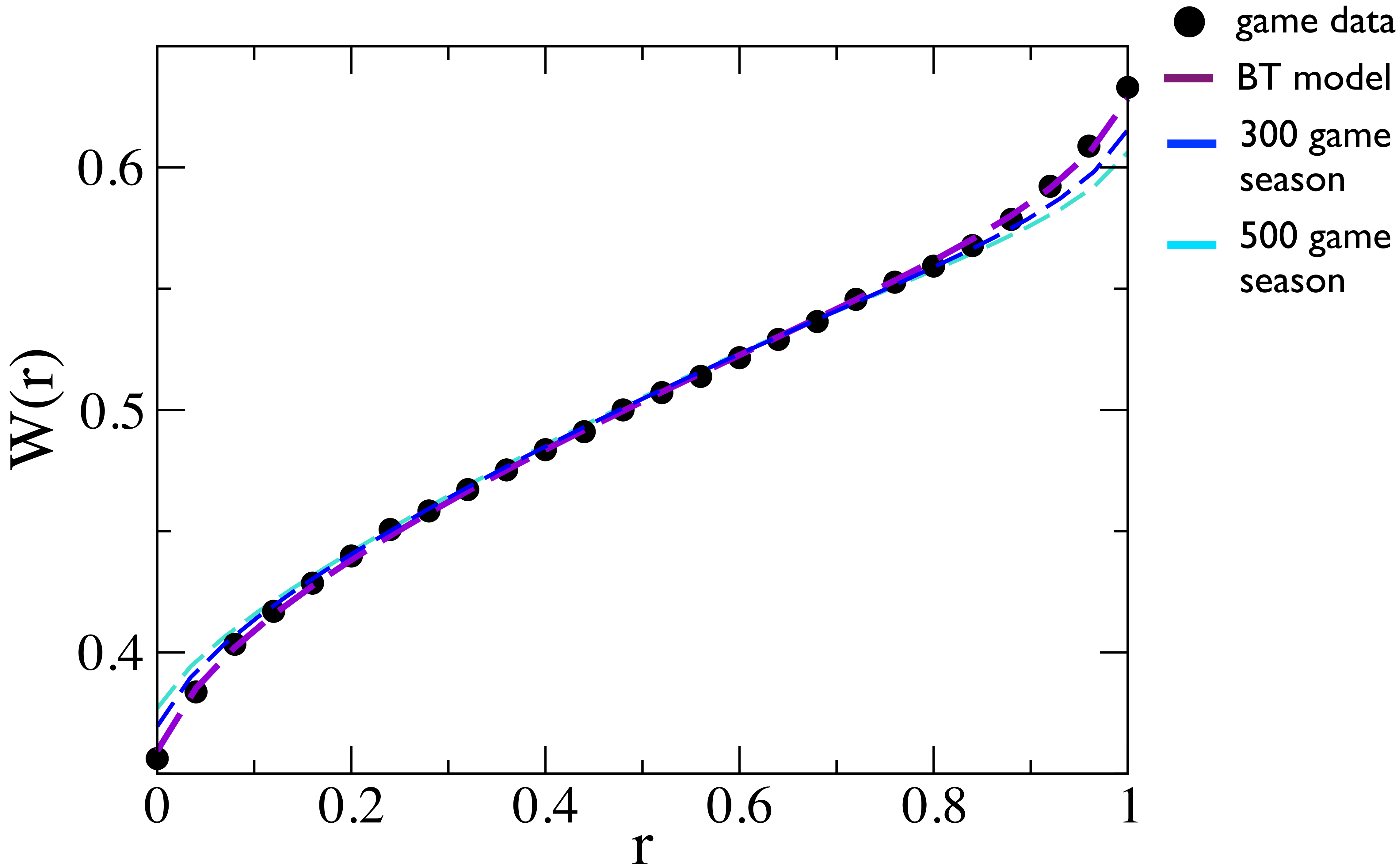
Role of Season Length on Win Fraction



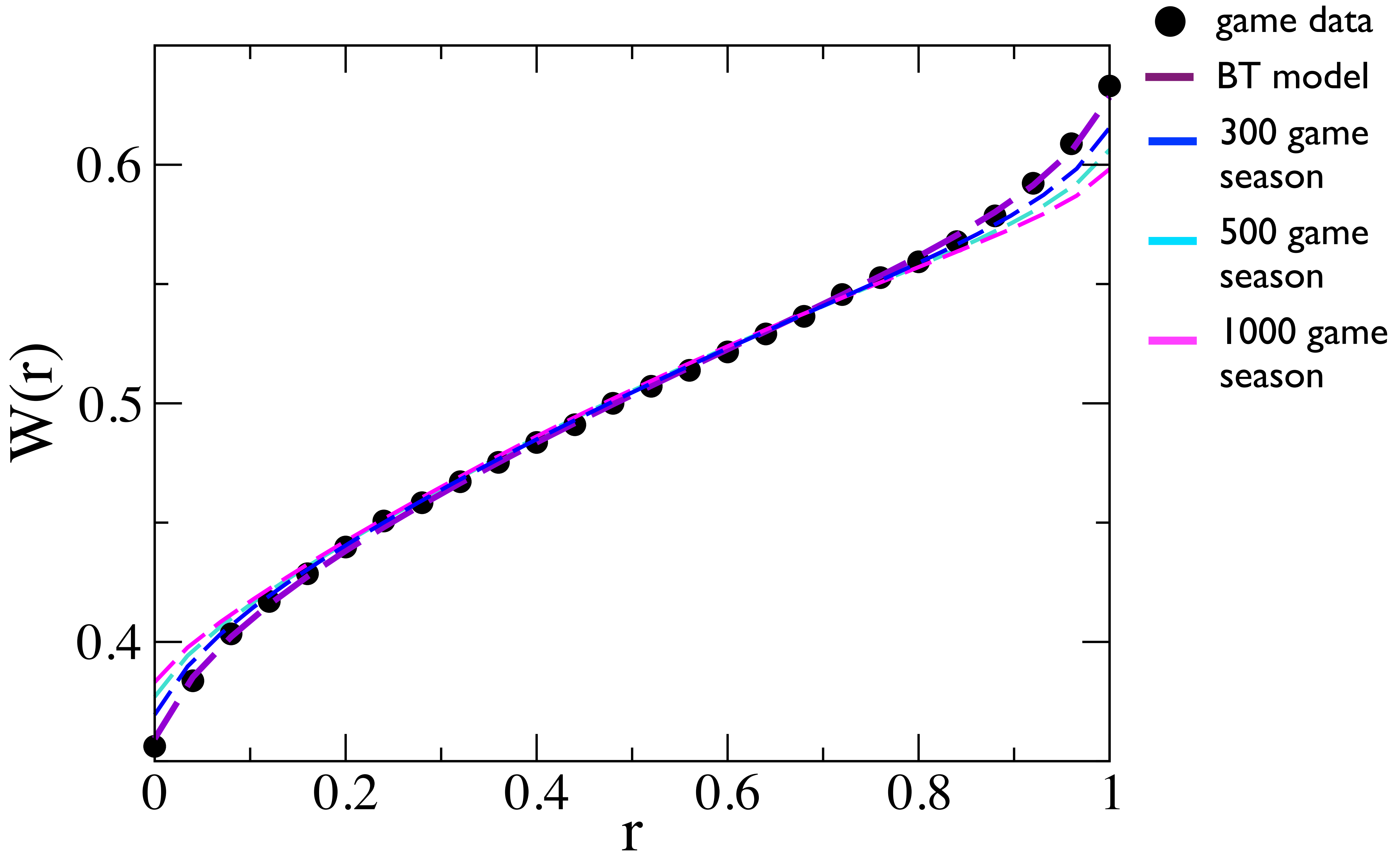
Role of Season Length on Win Fraction



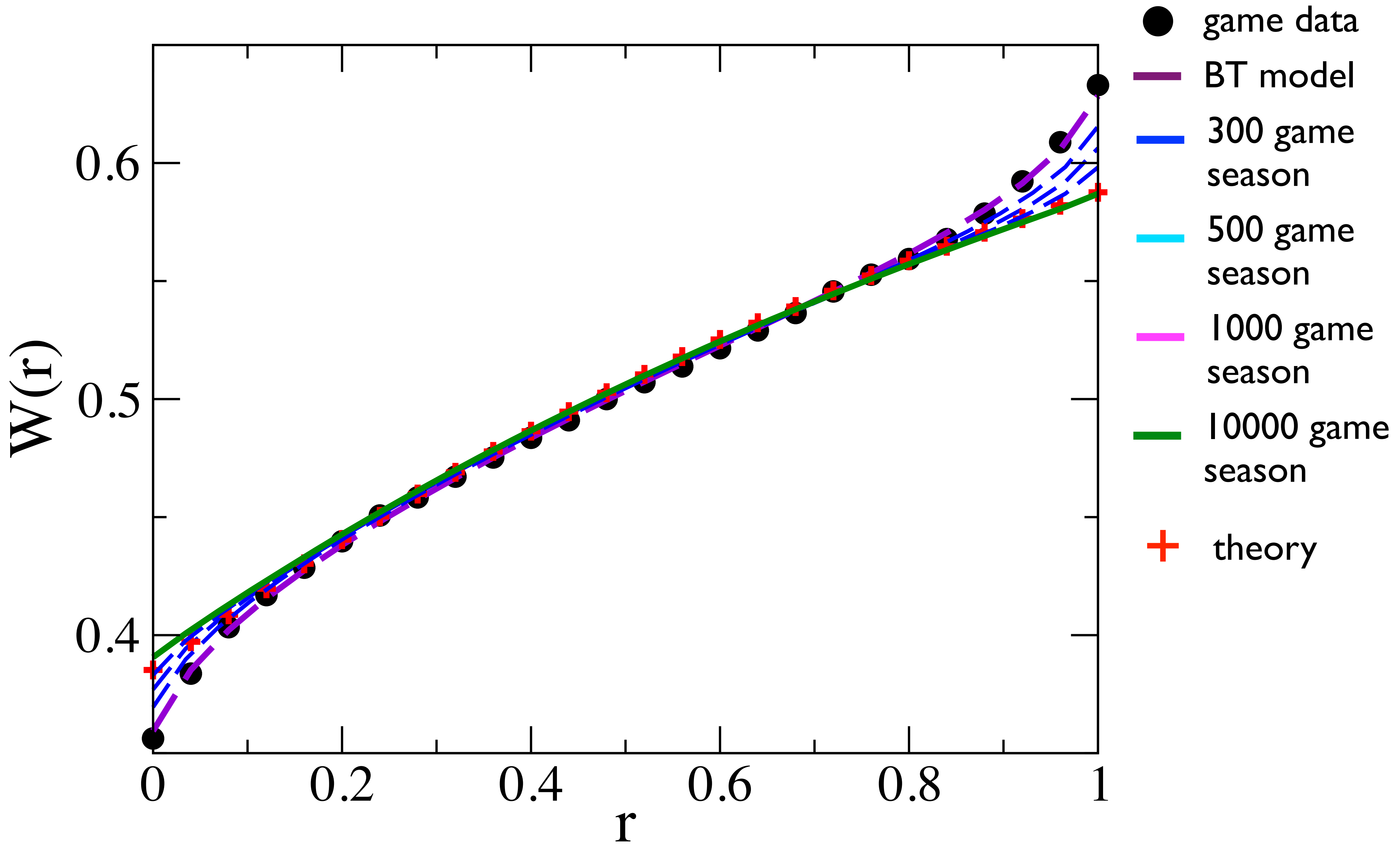
Role of Season Length on Win Fraction



Role of Season Length on Win Fraction



Role of Season Length on Win Fraction



Fun Facts About Winning and Losing Streaks

Fun Facts About Winning and Losing Streaks

longest win streaks

26	1916 New York Giants	(1 tie)
21	1935 Chicago Cubs	
20	2002 Oakland Athletics	
19	1906 Chicago White Sox	(1 tie)
19	1947 New York Yankees	
18	1904 New York Giants	
18	1953 New York Yankees	
17	1907 New York Giants	
17	1912 Washington Senators	
17	1916 New York Giants	
17	1931 Philadelphia Athletics	
16	1909 Pittsburgh Pirates	
16	1912 New York Giants	
16	1926 New York Yankees	
16	1951 New York Giants	
16	1977 Kansas City Royals	
15	1903 Pittsburgh Pirates	
15	1906 New York Highlanders	
15	1913 Philadelphia Athletics	
15	1924 Brooklyn Dodgers	
15	1936 Chicago Cubs	
15	1936 New York Giants	
15	1946 Boston Red Sox	
15	1960 New York Yankees	
15	1991 Minnesota Twins	
15	2000 Atlanta Braves	
15	2001 Seattle Mariners	

Fun Facts About Winning and Losing Streaks

longest win streaks

26	1916 New York Giants	(1 tie)
21	1935 Chicago Cubs	
20	2002 Oakland Athletics	
19	1906 Chicago White Sox	(1 tie)
19	1947 New York Yankees	
18	1904 New York Giants	
18	1953 New York Yankees	
17	1907 New York Giants	
17	1912 Washington Senators	
17	1916 New York Giants	
17	1931 Philadelphia Athletics	
16	1909 Pittsburgh Pirates	
16	1912 New York Giants	
16	1926 New York Yankees	
16	1951 New York Giants	
16	1977 Kansas City Royals	
15	1903 Pittsburgh Pirates	
15	1906 New York Highlanders	
15	1913 Philadelphia Athletics	
15	1924 Brooklyn Dodgers	
15	1936 Chicago Cubs	
15	1936 New York Giants	
15	1946 Boston Red Sox	
15	1960 New York Yankees	
15	1991 Minnesota Twins	
15	2000 Atlanta Braves	
15	2001 Seattle Mariners	

longest losing streaks

23	1961 Philadelphia Phillies	
21	1988 Baltimore Orioles	
20	1906 Boston Americans	
20	1906 Philadelphia As	
20	1916 Philadelphia As	
20	1969 Montreal Expos	(first year)
19	1906 Boston Beaneaters	
19	1914 Cincinnati Reds	
19	1975 Detroit Tigers	
19	2005 Kansas City Royals	
18	1920 Philadelphia As	
18	1948 Washington Senators	
18	1959 Washington Senators	
17	1926 Boston Red Sox	
17	1962 NY Mets	(first year)
17	1977 Atlanta Braves	
16	1911 Boston Braves	
16	1907 Boston Doves	
16	1907 Boston Americans	(2 ties)
16	1944 Brooklyn Dodgers	(1 made-up game)
15	1909 St. Louis Browns	
15	1911 Boston Rustlers	
15	1927 Boston Braves	
15	1927 Boston Red Sox	
15	1935 Boston Braves	
15	1937 Philadelphia As	
15	2002 Tampa Bay	
15	1972 Texas Rangers	(first year)

Fun Facts About Winning and Losing Streaks

longest win streaks

26	1916 New York Giants	(1 tie)
21	1935 Chicago Cubs	
20	2002 Oakland Athletics	
19	1906 Chicago White Sox	(1 tie)
19	1947 New York Yankees	
18	1904 New York Giants	
18	1953 New York Yankees	
17	1907 New York Giants	
17	1912 Washington Senators	
17	1916 New York Giants	
17	1931 Philadelphia Athletics	
16	1909 Pittsburgh Pirates	
16	1912 New York Giants	
16	1926 New York Yankees	
16	1951 New York Giants	
16	1977 Kansas City Royals	
15	1903 Pittsburgh Pirates	
15	1906 New York Highlanders	
15	1913 Philadelphia Athletics	
15	1924 Brooklyn Dodgers	
15	1936 Chicago Cubs	
15	1936 New York Giants	
15	1946 Boston Red Sox	
15	1960 New York Yankees	
15	1991 Minnesota Twins	
15	2000 Atlanta Braves	
15	2001 Seattle Mariners	

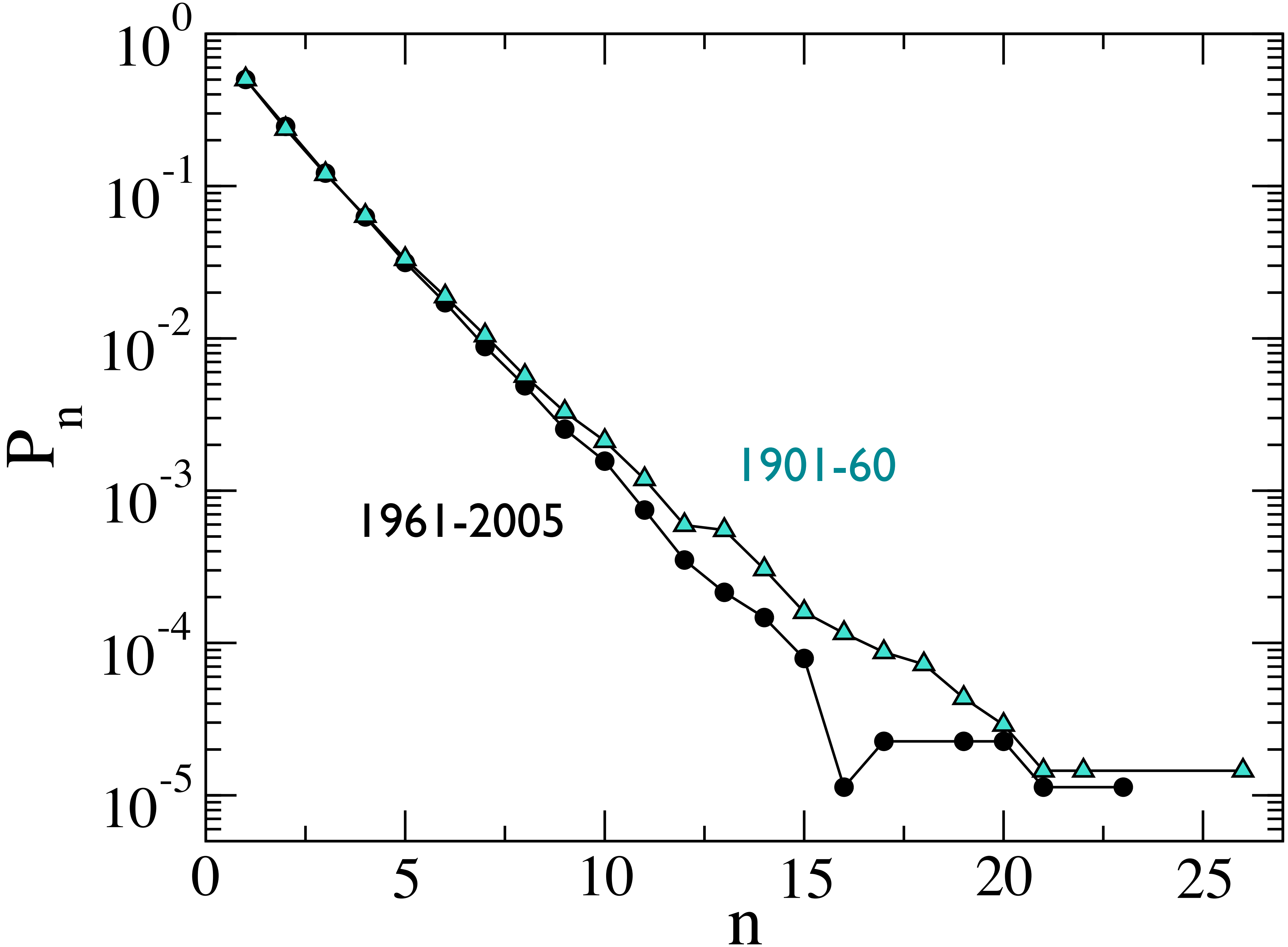
longest losing streaks

23	1961 Philadelphia Phillies	
21	1988 Baltimore Orioles	
20	1906 Boston Americans	
20	1906 Philadelphia As	
20	1916 Philadelphia As	
20	1969 Montreal Expos	(first year)
19	1906 Boston Beaneaters	
19	1914 Cincinnati Reds	
19	1975 Detroit Tigers	
19	2005 Kansas City Royals	
18	1920 Philadelphia As	
18	1948 Washington Senators	
18	1959 Washington Senators	
17	1926 Boston Red Sox	
17	1962 NY Mets	(first year)
17	1977 Atlanta Braves	
16	1911 Boston Braves	
16	1907 Boston Doves	
16	1907 Boston Americans	(2 ties)
16	1944 Brooklyn Dodgers	(1 made-up game)
15	1909 St. Louis Browns	
15	1911 Boston Rustlers	
15	1927 Boston Braves	
15	1927 Boston Red Sox	
15	1935 Boston Braves	
15	1937 Philadelphia As	
15	2002 Tampa Bay	
15	1972 Texas Rangers	(first year)

27 ≥15-game win/loss streaks between 1901-30
13 between 1931-60
15 between 1961-present

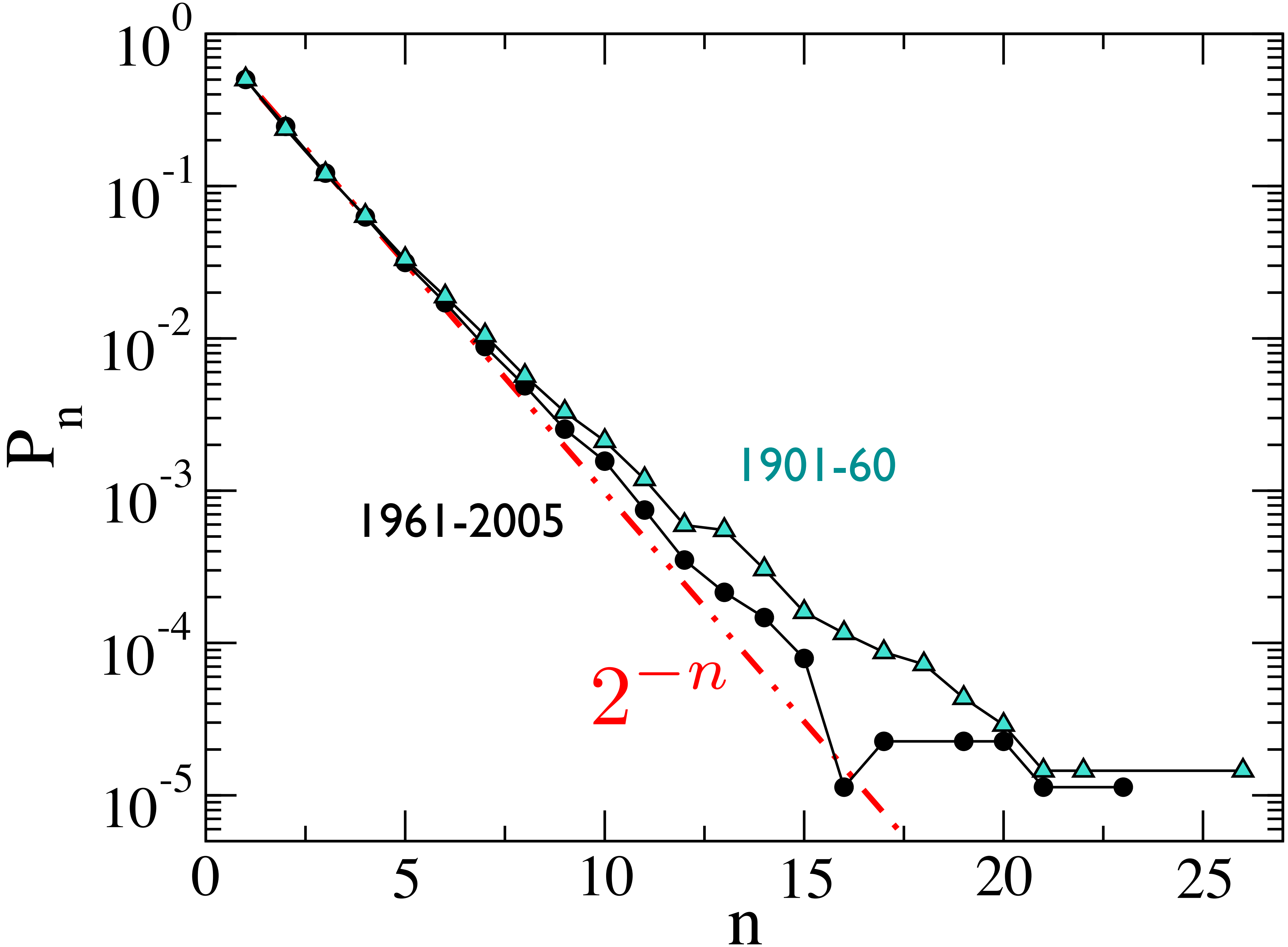
Winning and Losing Streak Distributions

MLB data



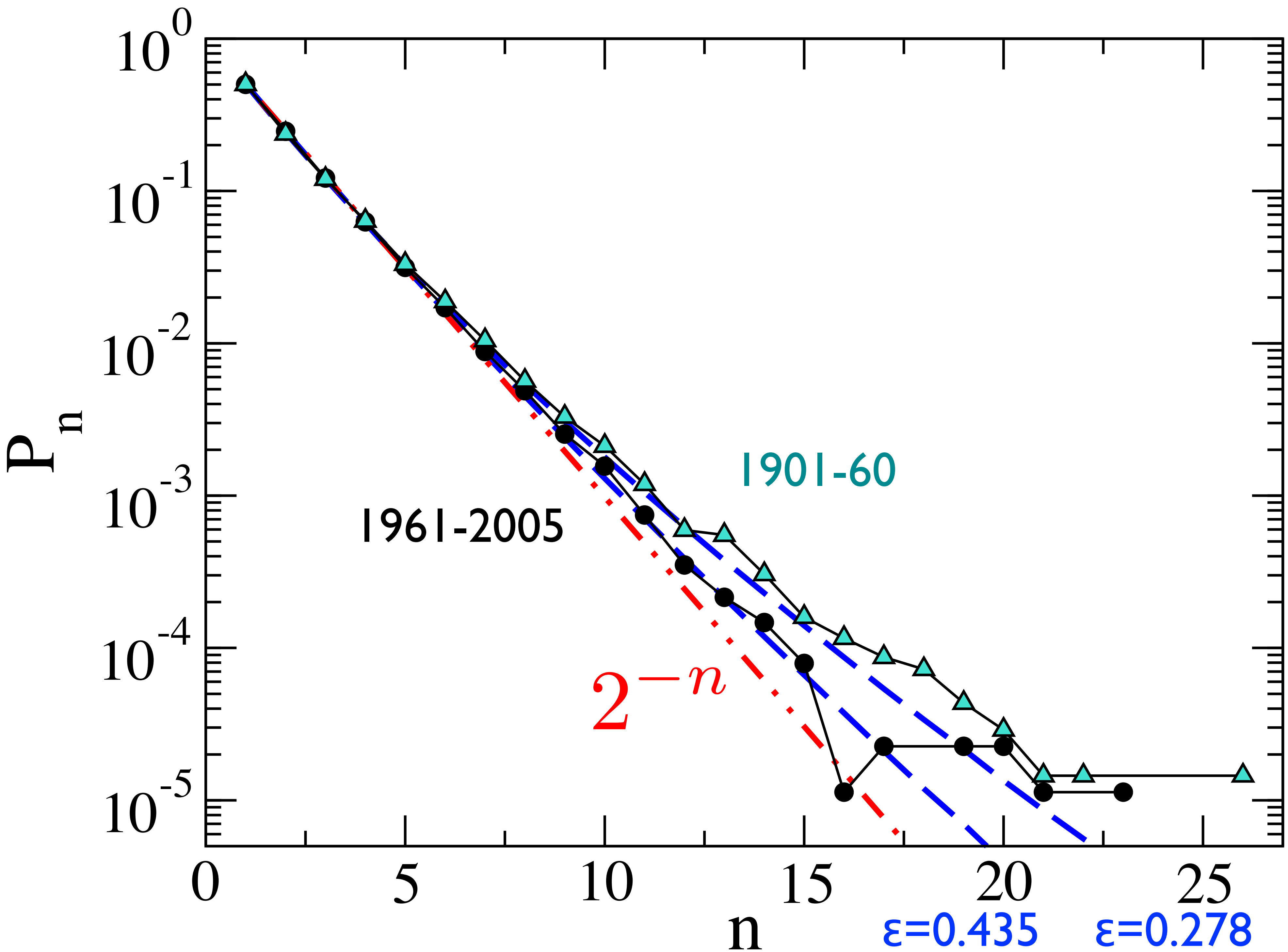
Winning and Losing Streak Distributions

MLB data



Winning and Losing Streak Distributions

BT model simulations



Streak Length Distribution in BT Model

n wins in a row — initial loss — final loss

$$P_n(x) = \prod_{j=1}^{n} \frac{x}{x + x_j} \; \frac{x_0}{x + x_0} \frac{x_{n+1}}{x + x_{n+1}}$$

Streak Length Distribution in BT Model

n wins in a row — initial loss — final loss

$$P_n(x) = \prod_{j=1}^{n} \frac{x}{x+x_j} \quad \frac{x_0}{x+x_0} \frac{x_{n+1}}{x+x_{n+1}}$$

$$\langle P_n(x) \rangle_{\{x_j\}} = x^n \left\langle \frac{1}{x+y} \right\rangle^n \left\langle \frac{y}{x+y} \right\rangle^2$$

$$\left\langle \frac{1}{x+y} \right\rangle = \frac{1}{1-\epsilon} \int_\epsilon^1 \frac{dy}{x+y} = \frac{1}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)$$

$$\left\langle \frac{y}{x+y} \right\rangle = 1 - \frac{x}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)$$

Streak Length Distribution in BT Model

n wins in a row — initial loss — final loss

$$P_n(x) = \prod_{j=1}^{n} \frac{x}{x+x_j} \quad \frac{x_0}{x+x_0} \frac{x_{n+1}}{x+x_{n+1}}$$

$$\langle P_n(x) \rangle_{\{x_j\}} = x^n \left\langle \frac{1}{x+y} \right\rangle^n \left\langle \frac{y}{x+y} \right\rangle^2$$

$$\left\langle \frac{1}{x+y} \right\rangle = \frac{1}{1-\epsilon} \int_\epsilon^1 \frac{dy}{x+y} = \frac{1}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)$$

$$\left\langle \frac{y}{x+y} \right\rangle = 1 - \frac{x}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)$$

$$\langle P_n \rangle = \frac{1}{1-\epsilon} \int_\epsilon^1 \left[1 - \frac{x}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)\right]^2 e^{n g(x)}\, dx$$

$$g(x) = \ln x + \ln\left[\frac{1}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)\right]$$

monotonic

Streak Length Distribution in BT Model

n wins in a row — initial loss — final loss

$$P_n(x) = \prod_{j=1}^{n} \frac{x}{x+x_j} \quad \frac{x_0}{x+x_0} \frac{x_{n+1}}{x+x_{n+1}}$$

$$\langle P_n(x)\rangle_{\{x_j\}} = x^n \left\langle \frac{1}{x+y} \right\rangle^n \left\langle \frac{y}{x+y} \right\rangle^2$$

$$\left\langle \frac{1}{x+y} \right\rangle = \frac{1}{1-\epsilon} \int_\epsilon^1 \frac{dy}{x+y} = \frac{1}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)$$

$$\left\langle \frac{y}{x+y} \right\rangle = 1 - \frac{x}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)$$

$$\langle P_n \rangle = \frac{1}{1-\epsilon} \int_\epsilon^1 \left[1 - \frac{x}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)\right]^2 e^{ng(x)}\, dx$$

$$g(x) = \ln x + \ln\left[\frac{1}{1-\epsilon} \ln\left(\frac{x+1}{x+\epsilon}\right)\right]$$

$$\sim \frac{1}{ng'(1)}\, e^{ng(1)}$$

monotonic

Asymptotic Behavior

$$\langle P_n \rangle \sim \frac{1}{n g'(1)} \; e^{n g(1)}$$

$$g(x) = \ln x + \ln \left[\frac{1}{1-\epsilon} \ln \left(\frac{x+1}{x+\epsilon} \right) \right]$$

$$g(1) = -\ln(1-\epsilon) + \ln \ln \left(\frac{2}{1+\epsilon} \right)$$

Asymptotic Behavior

$$\langle P_n \rangle \sim \frac{1}{n g'(1)} \; e^{n g(1)}$$

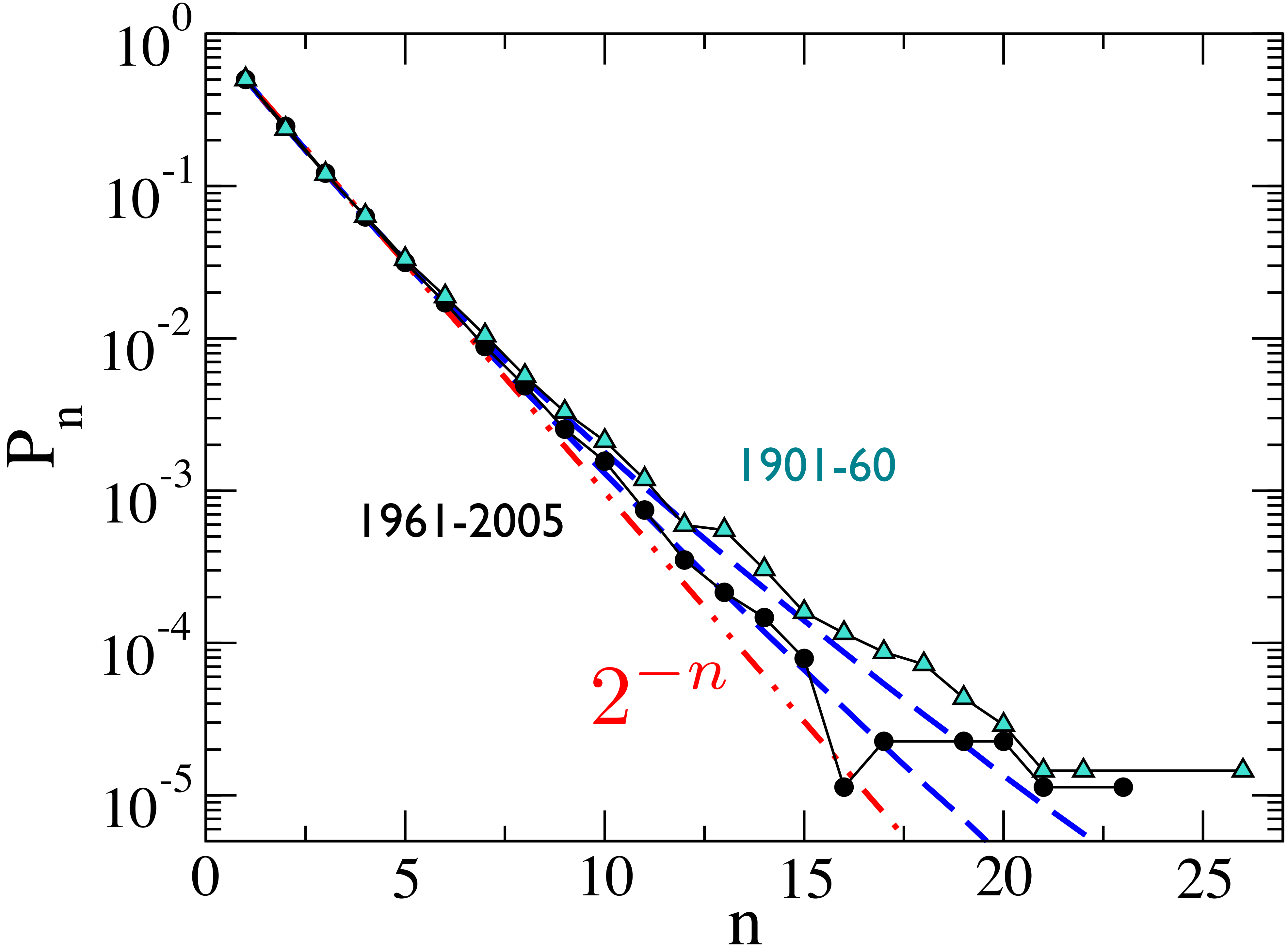
$$g(x) = \ln x + \ln \left[\frac{1}{1-\epsilon} \ln \left(\frac{x+1}{x+\epsilon} \right) \right]$$

$$g(1) = -\ln(1-\epsilon) + \ln \ln \left(\frac{2}{1+\epsilon} \right)$$

fastest decay when ε=1: $\langle P_n \rangle \sim 2^{-n}$

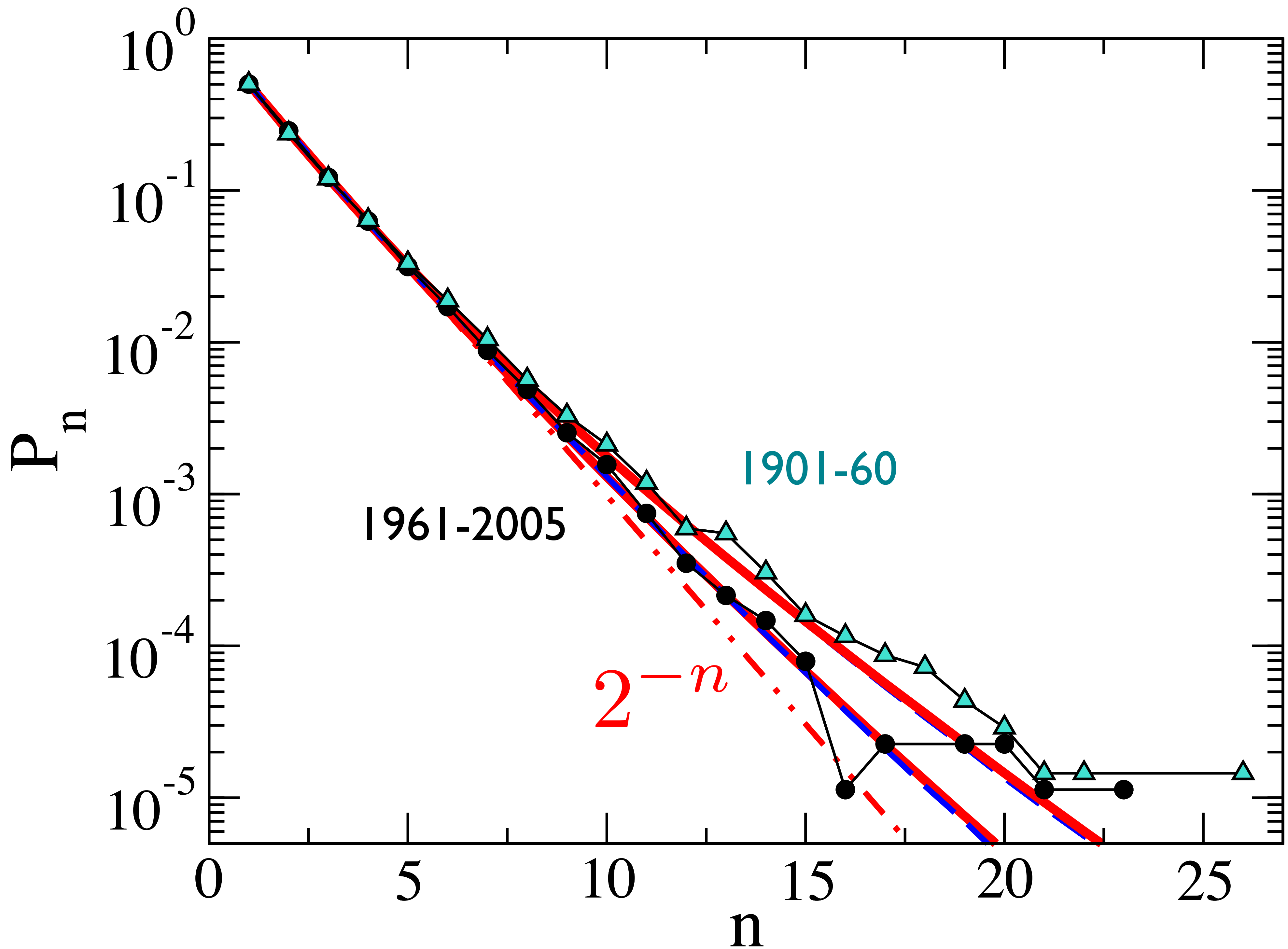
slowest decay when ε=0: $\langle P_n \rangle \sim (\ln 2)^n \approx (0.693)^n$

Winning and Losing Streak Distributions



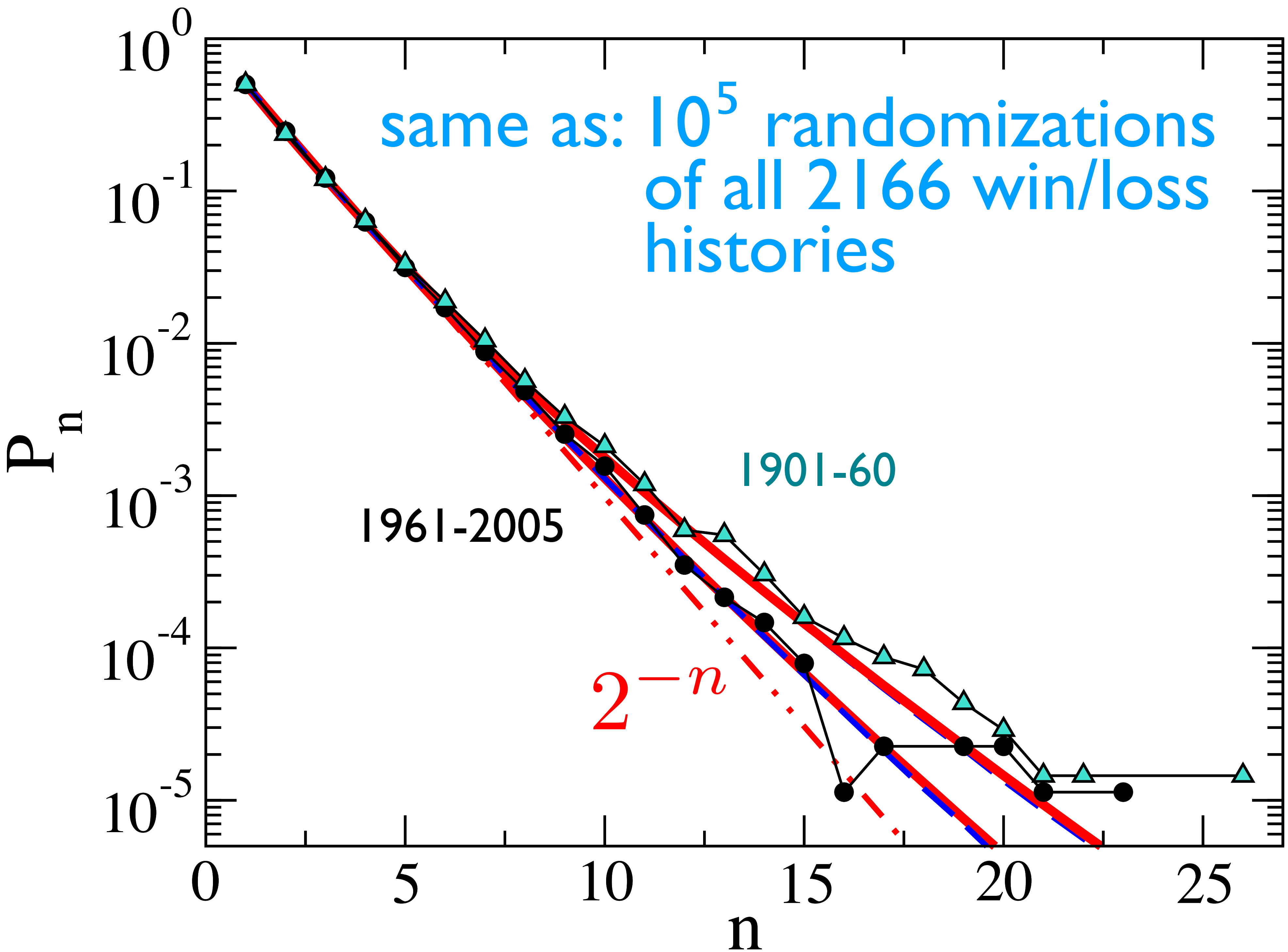
Winning and Losing Streak Distributions

numerically integrate $\langle P_n \rangle = \frac{1}{1-\epsilon}\int_{\epsilon}^{1}\left[1-\frac{x}{1-\epsilon}\ln\left(\frac{x+1}{x+\epsilon}\right)\right]^2 e^{ng(x)}\,dx$



Winning and Losing Streak Distributions

numerically integrate $\langle P_n \rangle = \frac{1}{1-\epsilon}\int_\epsilon^1 \left[1 - \frac{x}{1-\epsilon}\ln\left(\frac{x+1}{x+\epsilon}\right)\right]^2 e^{ng(x)}\,dx$



Summary/Outlook

Summary/Outlook

Citations:

Page rank analysis: *helps uncover hidden "gems"*

Future: *contextual analysis, specialization*

Summary/Outlook

Citations:

Page rank analysis: *helps uncover hidden "gems"*

Future: *contextual analysis, specialization*

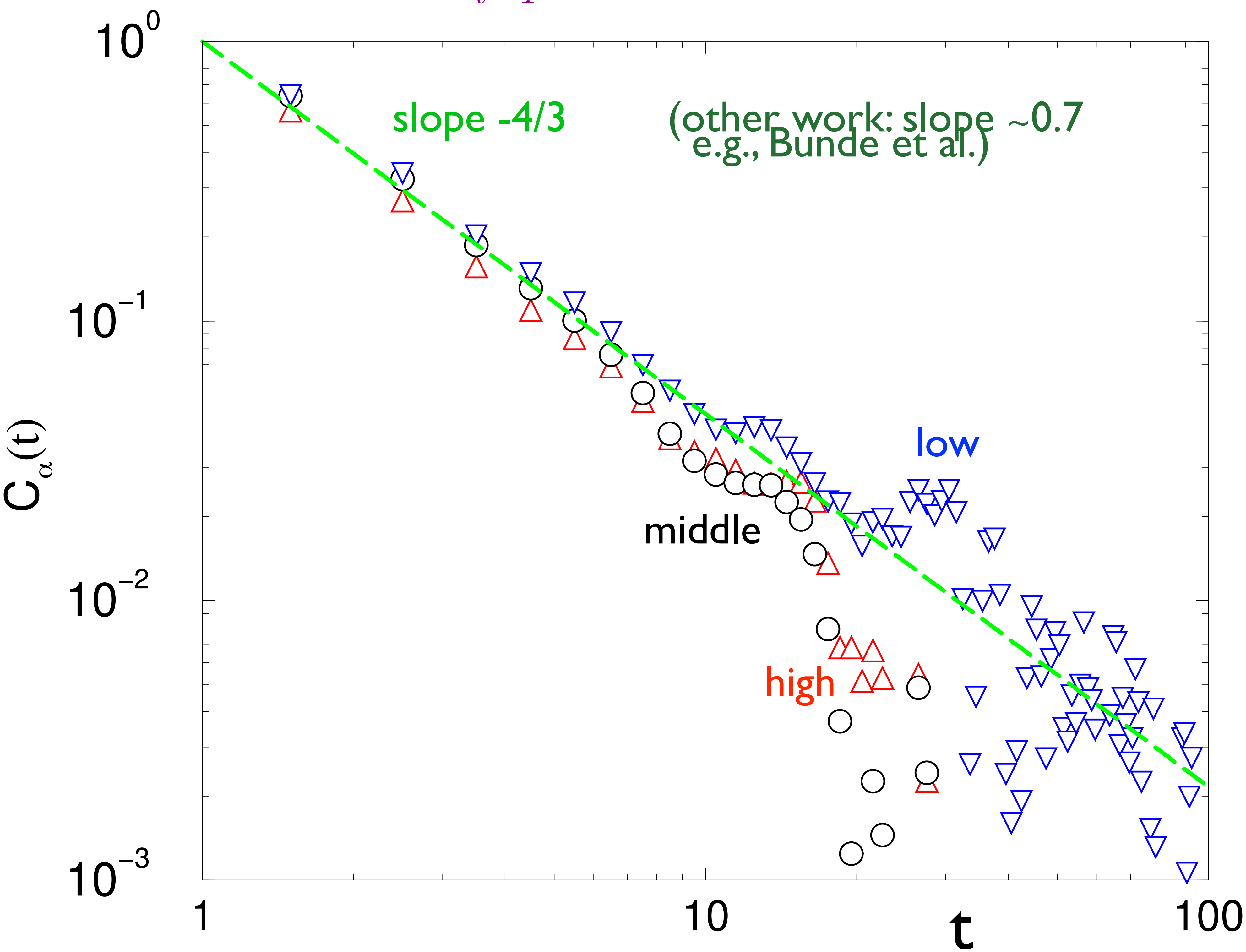
Climate:

Global warming seems to affect temperature records

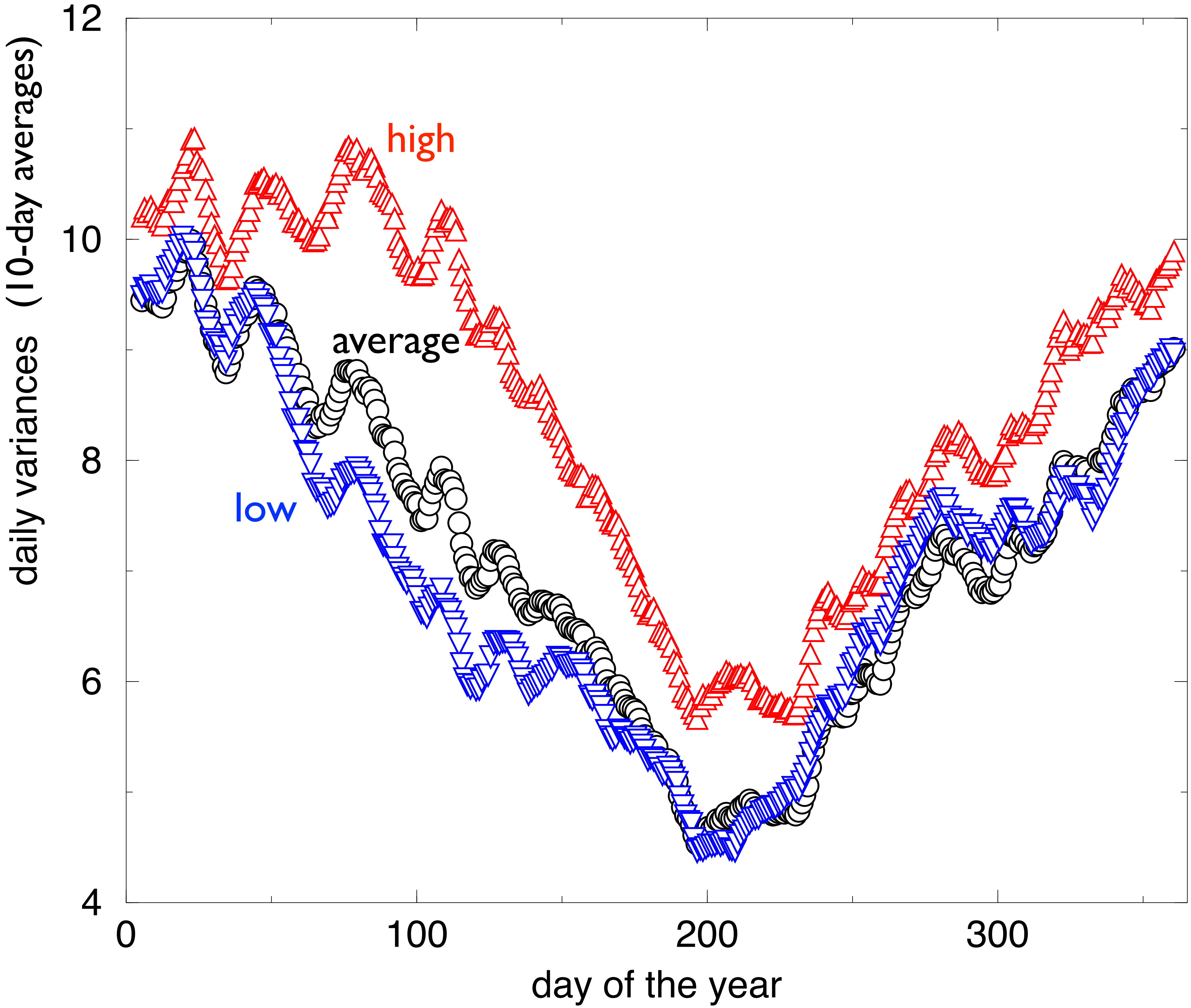
Open: *seasonality, high/low asymmetry, changing variability*

Inter-day Temperature Correlations

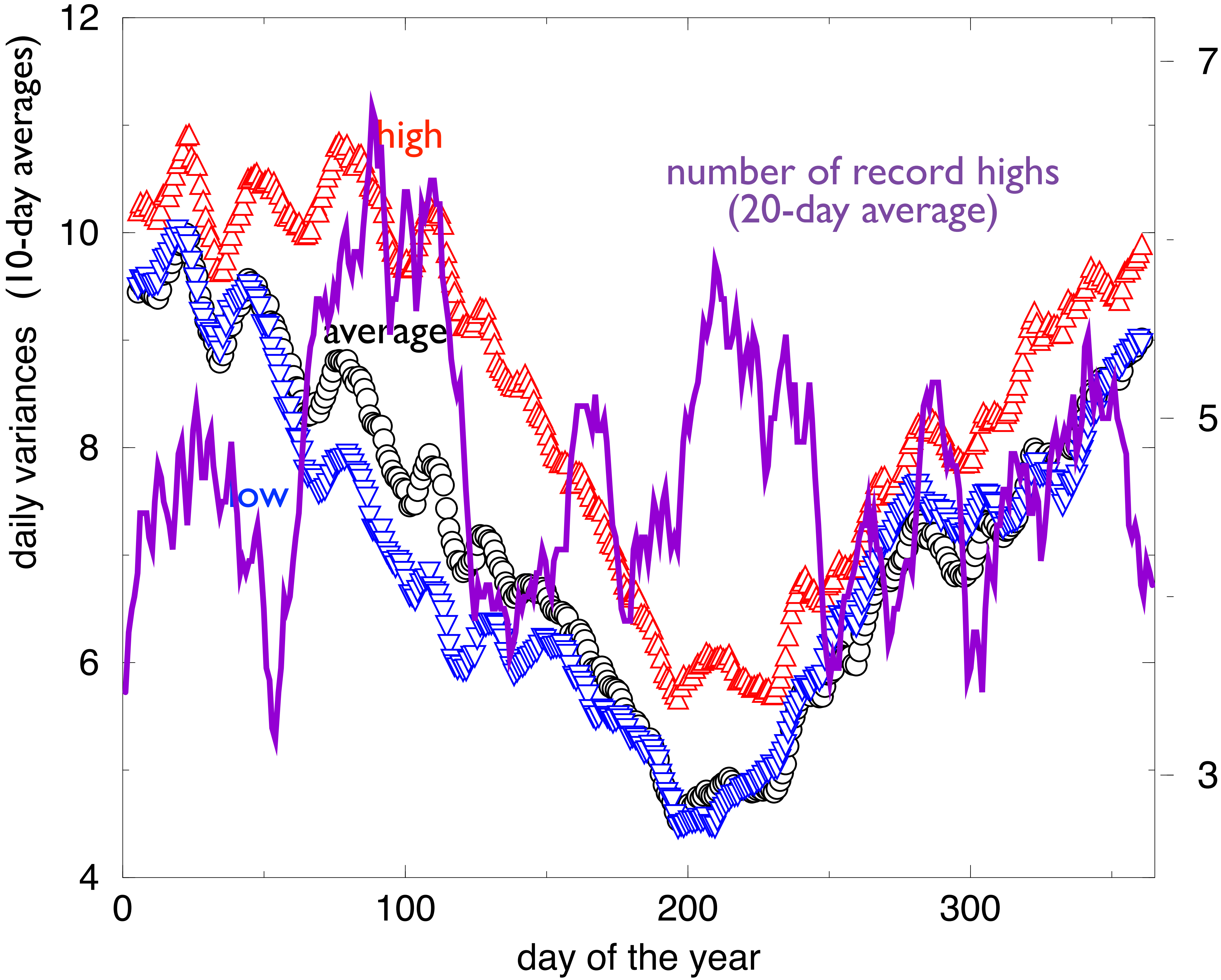
$$C_\alpha(t) = \frac{1}{365}\sum_{i=1}^{365} \frac{\langle T_i T_{i+t}\rangle - \langle T_i\rangle\langle T_{i+t}\rangle}{\langle T_i^2\rangle - \langle T_i\rangle^2} \quad \alpha = \text{low, middle, high}$$



Seasonal Variance



Seasonal Variance & Record Numbers



Summary/Outlook

Citations:

Page rank analysis: *helps uncover hidden "gems"*

Future: *contextual analysis, specialization*

Climate:

Global warming seems to affect temperature records

Open: *seasonality, high/low asymmetry, changing variability*

Summary/Outlook

Citations:

Page rank analysis: *helps uncover hidden "gems"*

Future: *contextual analysis, specialization*

Climate:

Global warming seems to affect temperature records

Open: *seasonality, high/low asymmetry, changing variability*

Competition:

Statistical model → *win/loss records, streak distributions*

keener competition; harder to maintain advantage