Record theory and applications:
Global warming and market fluctuations

Joachim Krug
Institute of Theoretical Physics, University of Cologne

- What are records, and why do we care?
- Records in growing and improving populations
- Record-breaking temperatures and global warming
  with Gregor Wergen
- Records of random walks and financial data
  with Miro Bogner

Records in popular culture

Records in popular culture

1.3.2008: 60 New Delhi chefs create the world's largest biryani (13 tons)

http://www.guinnessworldrecords.com/

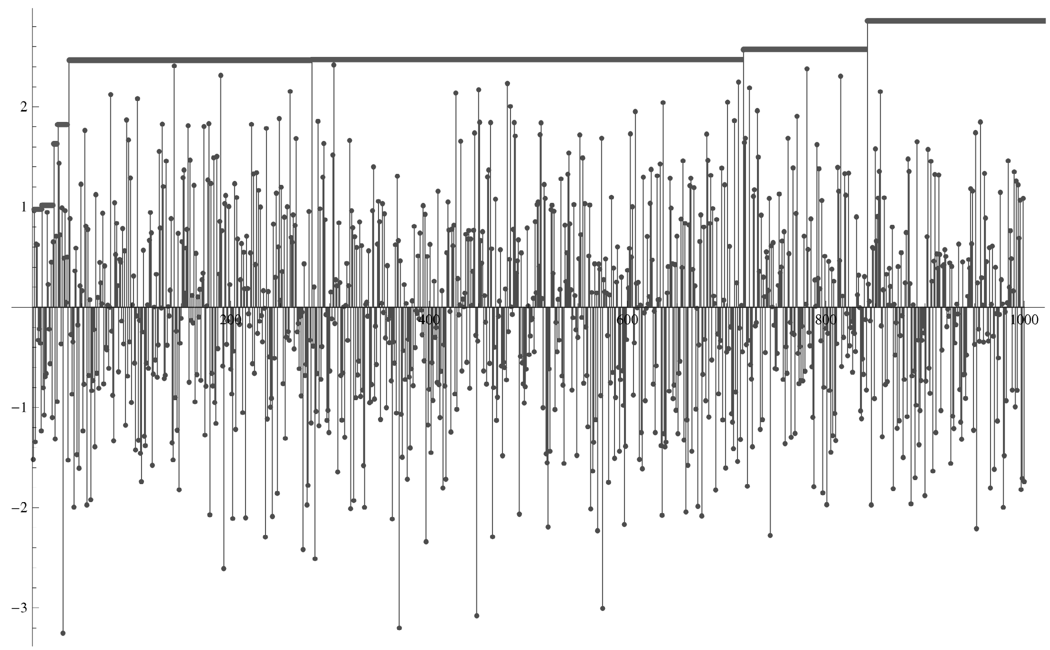
Records in popular culture

2007: Shailendra Singh Yadav (Kanpur) published most letters (214) in a single national newspaper (AJ Independent Hindi Daily) in a year.

http://www.guinnessworldrecords.com/

Basic facts about records I

- A record is an entry in a sequence of random variables (RV's) X_n which is larger (upper record) or smaller (lower records) than all previous entries



- If the RV's are independent and identically distributed (i.i.d.), the probability for a record at time n is $P_n = 1/n$ by symmetry

- This result is universal, i.e. independent of the underlying distribution (provided it is continuous)

Basic facts about records II: i.i.d. RV's

N. Glick, Am. Math. Mon. **85**, 2 (1978)

- The expected number of records up to time n is

$$\langle R_n \rangle = \sum_{k=1}^{n} \frac{1}{k} = \ln(n) + \gamma + \mathcal{O}(1/n)$$

where $\gamma \approx 0.5772156649....$ is the Euler-Mascheroni constant

- Record events are independent: The sequence of records is a Bernoulli process with success probability P_n, which converges to a Poisson process in logarithmic time for large n

- If n_k is the time of the k'th record, then $n_k/n_{k+1} \in [0,1]$ becomes a uniform RV for large k. As a consequence

$$\langle n_k \rangle|_{n_{k+1}} \approx \frac{1}{2} n_{k+1}, \quad \langle n_k \rangle|_{n_{k-1}} = \infty$$

$\Rightarrow$ records can only be "predicted" backwards in time

Beyond the i.i.d. model

Records in growing populations

M.C.K. Yang, J. Appl. Prob. **12**, 148 (1975)

- Motivation: Olympic records occur at an essentially constant (non-decreasing) rate

- Model: At each time n a new "generation" of N_n i.i.d. RV's becomes available simultaneously. By symmetry, the probability of a new record at time n is then

$$P_n = \frac{N_n}{\sum_{k=1}^{n} N_k}$$

- For an exponentially growing population, $N_n = a^n$, this yields

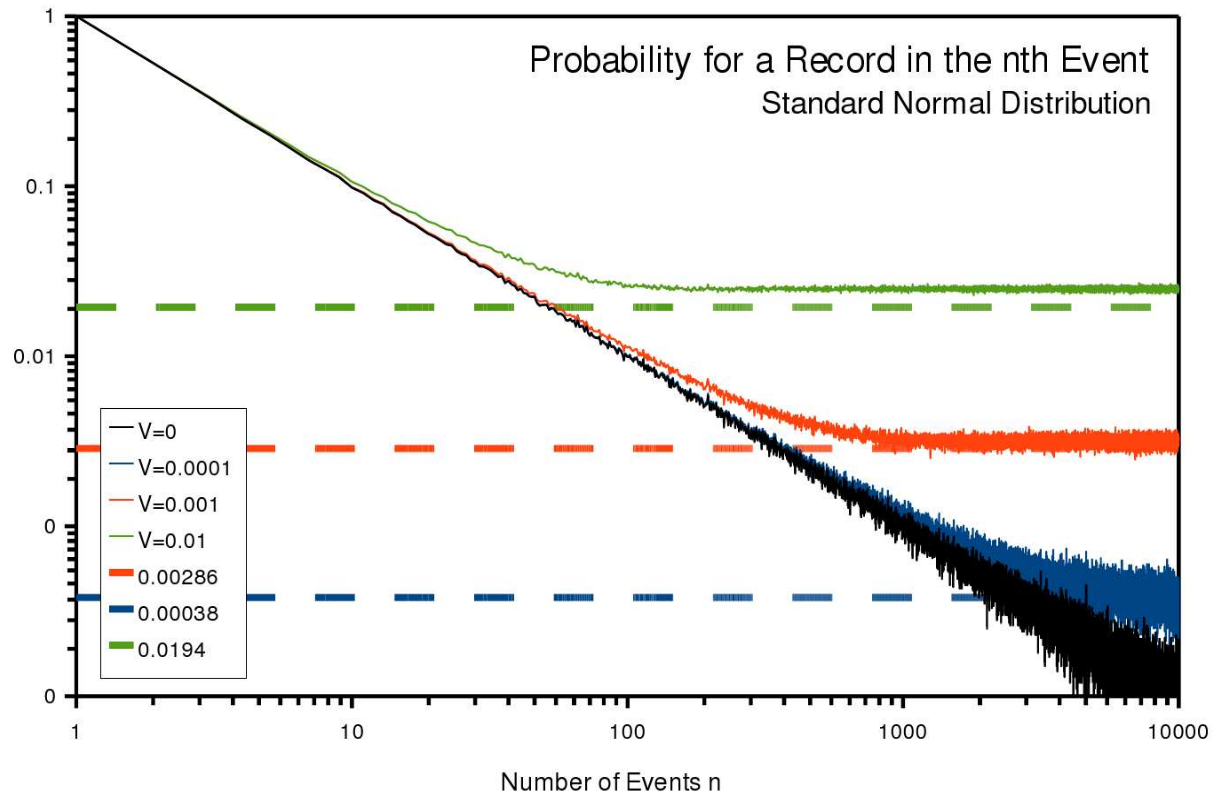
$$P_n = \frac{a^n(a-1)}{a(a^n-1)} \to \frac{a-1}{a} \text{ for } n \to \infty.$$

- The growth of the world population is insufficient to explain the occurrence rate of Olympic records under this model.

Records in improving populations

R. Ballerini & S. Resnick, J. Appl. Prob. **22**, 487 (1985)

- Let $X_n = Y_n + vn$ with i.i.d. RV's Y_n and a drift speed $v > 0$
- For large n the record probability approaches a finite limit $\lim_{n\to\infty} P_n(v) > 0$ which is however difficult to compute in general



An exactly solvable case

thanks to Jasper Franke

- Let Y_n have probability density $p(y)$ and probability distribution function $q(x) = \int^x dy\, p(y)$. Then

$$P_n(v) = \int dx_n\, p(x_n - vn) \prod_{k=1}^{n-1} q(x_n - vk) = \int dx\, p(x) \prod_{k=1}^{n-1} q(x + vk)$$

- For the Gumbel distribution $q(x) = \exp[-e^{-x/b}]$

$$\prod_{k=1}^{n-1} q(x - vk) = \exp[-e^{-x/b} \sum_{k=1}^{n-1} e^{-vk/b}] = q(x)^{\alpha_n} \quad \text{with} \quad \alpha_n = \sum_{k=1}^{n-1} (e^{-v/b})^k$$

$$\Rightarrow \; P_n(v) = \int_0^1 dq q^{\alpha_n} = \frac{1}{\alpha_n + 1} = \frac{1 - e^{-v/b}}{1 - e^{-nv/b}}$$

- Key parameter is the ratio v/b

Records from broadening distributions

JK, J. Stat. Mech. P07001 (2007)

- Let X_n be drawn from $p_n(x) = n^{-\alpha} f(x/n^{\alpha})$ with $\alpha > 0$

- Asymptotic growth of the number of records depends on the universality class of f in the sense of extreme value statistics.

  Fréchet class: $f(x) \sim x^{-(\mu+1)} \quad \Rightarrow \quad \langle R_n \rangle \approx (1+\alpha\mu)\ln(n)$

  Gumbel class: $f(x) \sim \exp[-x^{\beta}] \quad \Rightarrow \quad \langle R_n \rangle \sim \alpha \ln^2(n)$

  Weibull class: $f(x) \sim (x_{\max} - x)^{\delta}, \delta > 0 \quad \Rightarrow \quad \langle R_n \rangle \sim \alpha^{\delta} n^{1/(\delta+1)}$

- Effect of broadening is stronger for fast decaying tails, and generally weaker than effect of drift in the mean value

- Broadening (and drift) generically induces correlations between records

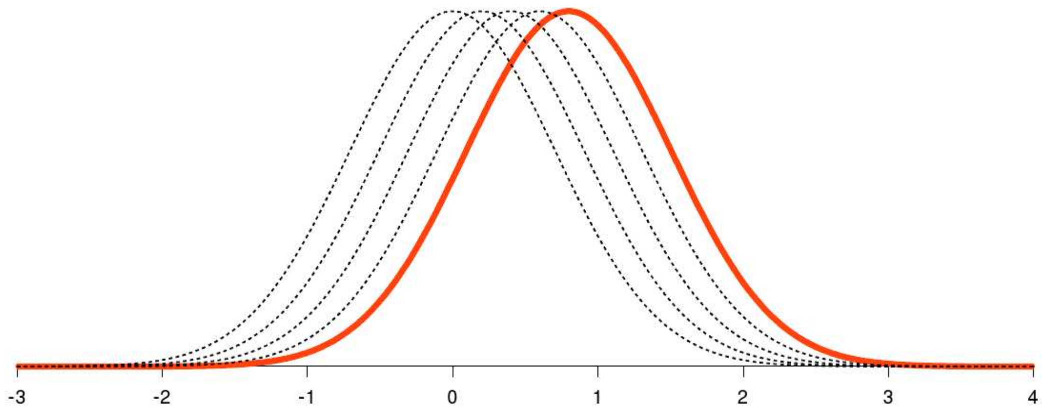
Application to global warming

Record-breaking temperatures and global warming

R.E. Benestad (2003); S. Redner & M.R. Petersen (2006)

- Question: Does global warming significantly increase the occurrence of record-breaking high daily temperatures?

- Model: The temperature on a given calendar day of the year is an independent Gaussian RV with constant standard deviation σ and a mean that increases at speed v



- Typical values: $v \approx 0.03^{\text{o}}$C/yr, $\sigma \approx 3.5^{\text{o}}$C $\Rightarrow v/\sigma \ll 1$

Expansion for small drift speed

- We want to compute the record rate $P_n(v) = \int dx\, p(x) \prod_{k=1}^{n-1} q(x+vk)$ for general $q(x)$ and $p(x) = dq/dx$

- To leading order in v we have $q(x+vk) \approx q(x) + vkp(x)$

$$\Rightarrow P_n \approx \int dx\, p(x) q(x)^{n-1} + \frac{vn(n-1)}{2} \int dx\, p(x)^2 q(x)^{n-2} = \frac{1}{n} + vI_n$$

with $I_n = \frac{n(n-1)}{2} \int dx\, p(x)^2 q(x)^{n-2}$

- Asymptotic behavior of I_n depends on the universality class of p:

Fréchet class: $p(x) \sim x^{-(\mu+1)} \quad \Rightarrow \quad I_n \sim n^{-1/\mu} \to 0$

Weibull class: $p(x) \sim (x_{\max} - x)^{\delta}, \delta > 0 \quad \Rightarrow \quad I_n \sim n^{1/(\delta+1)} \to \infty$

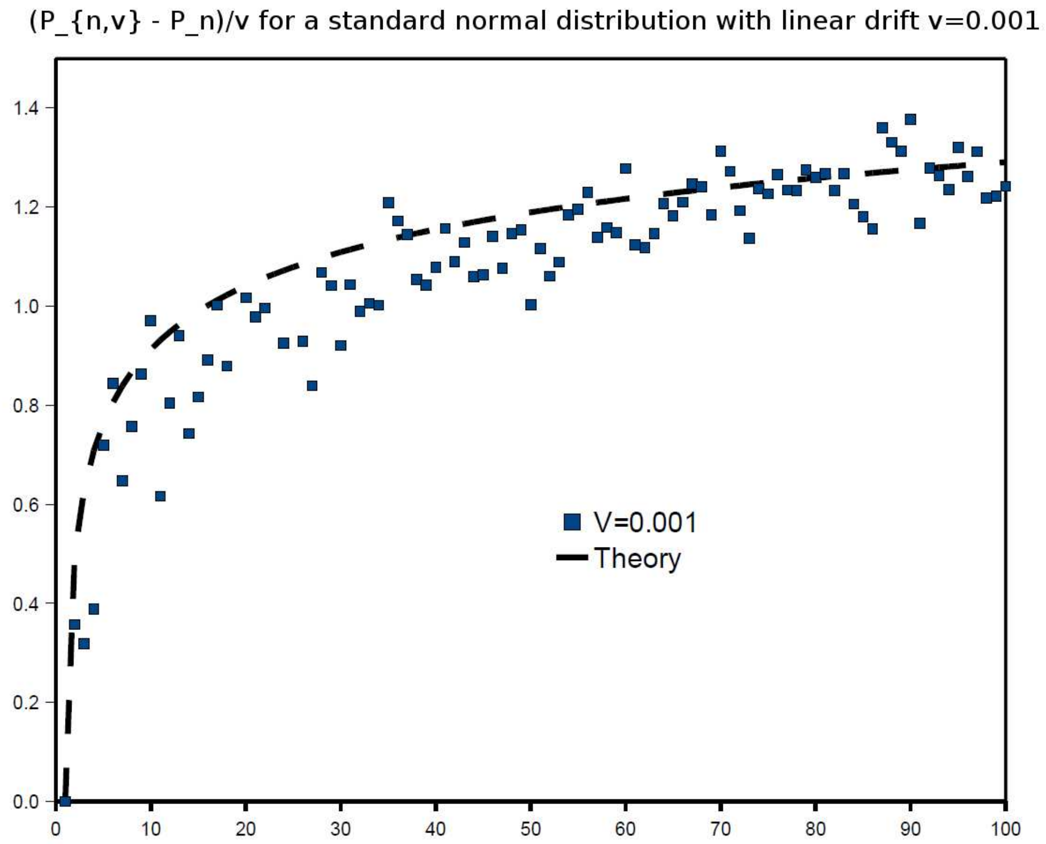
Exponential: $p(x) = e^{-x/a} \Rightarrow \quad I_n = \frac{1}{2a}$

$\Rightarrow$ effect of drift is smaller for fatter tails

- In the Gaussian case I_n can be evaluated in closed form only for $n = 2, 3$

- A saddle point approximation for large n yields the result

$$P_n(v) \approx \frac{1}{n} + \frac{v}{\sigma} \frac{(2\pi)^{3/2}}{e^2} \sqrt{\ln(n^2/8\pi)}$$



Analysis of daily high temperatures

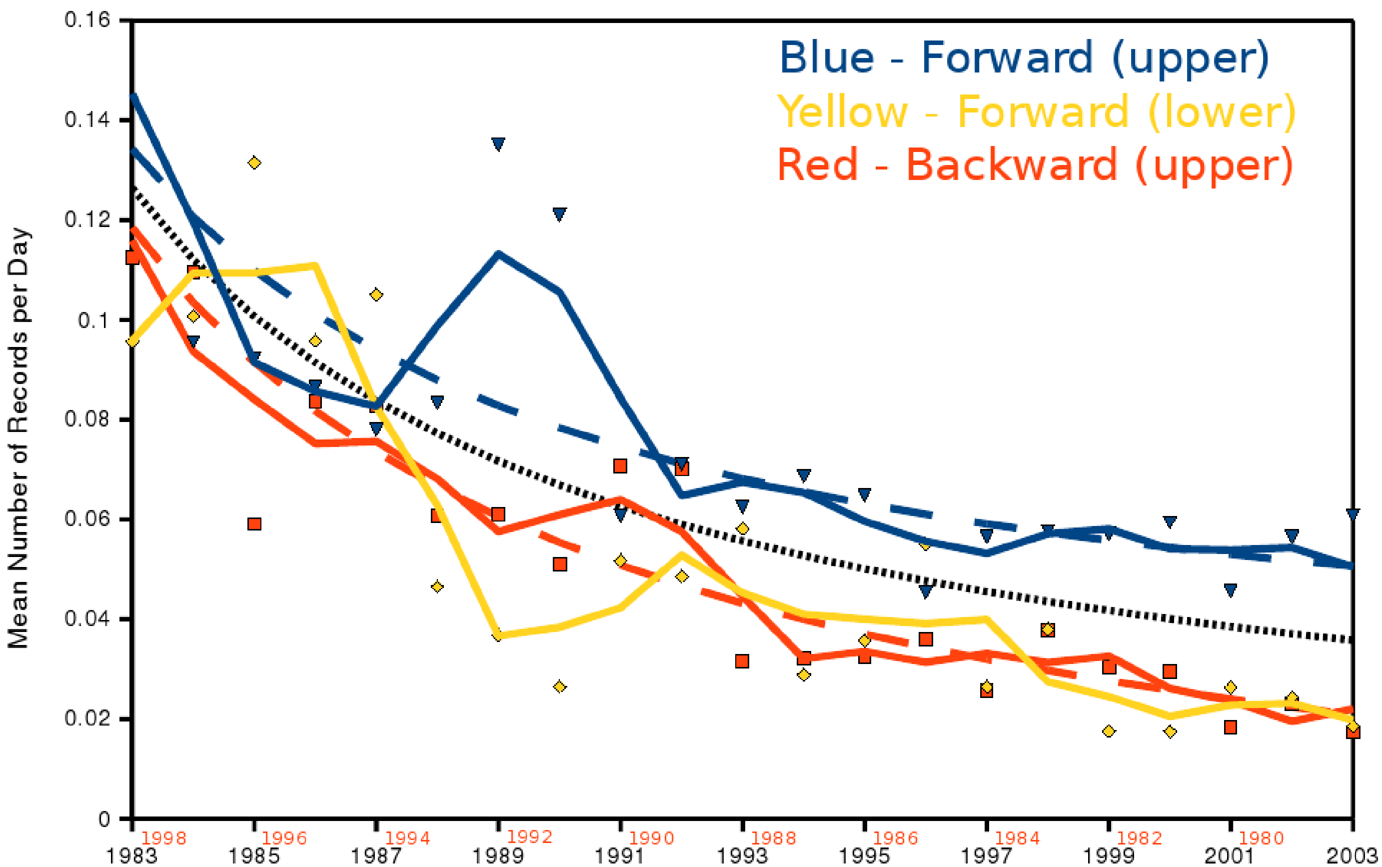
European data

- 43 stations over 100 year period 1906-2005
- 187 stations over 30 year period 1976-2005
- 30 year data: Constant warming rate $v \approx 0.034 \pm 0.01^{\circ}$C/yr, standard deviation $\sigma \approx 3.5 \pm 0.5^{\circ}$C $\Rightarrow v/\sigma \approx 0.01$

American data

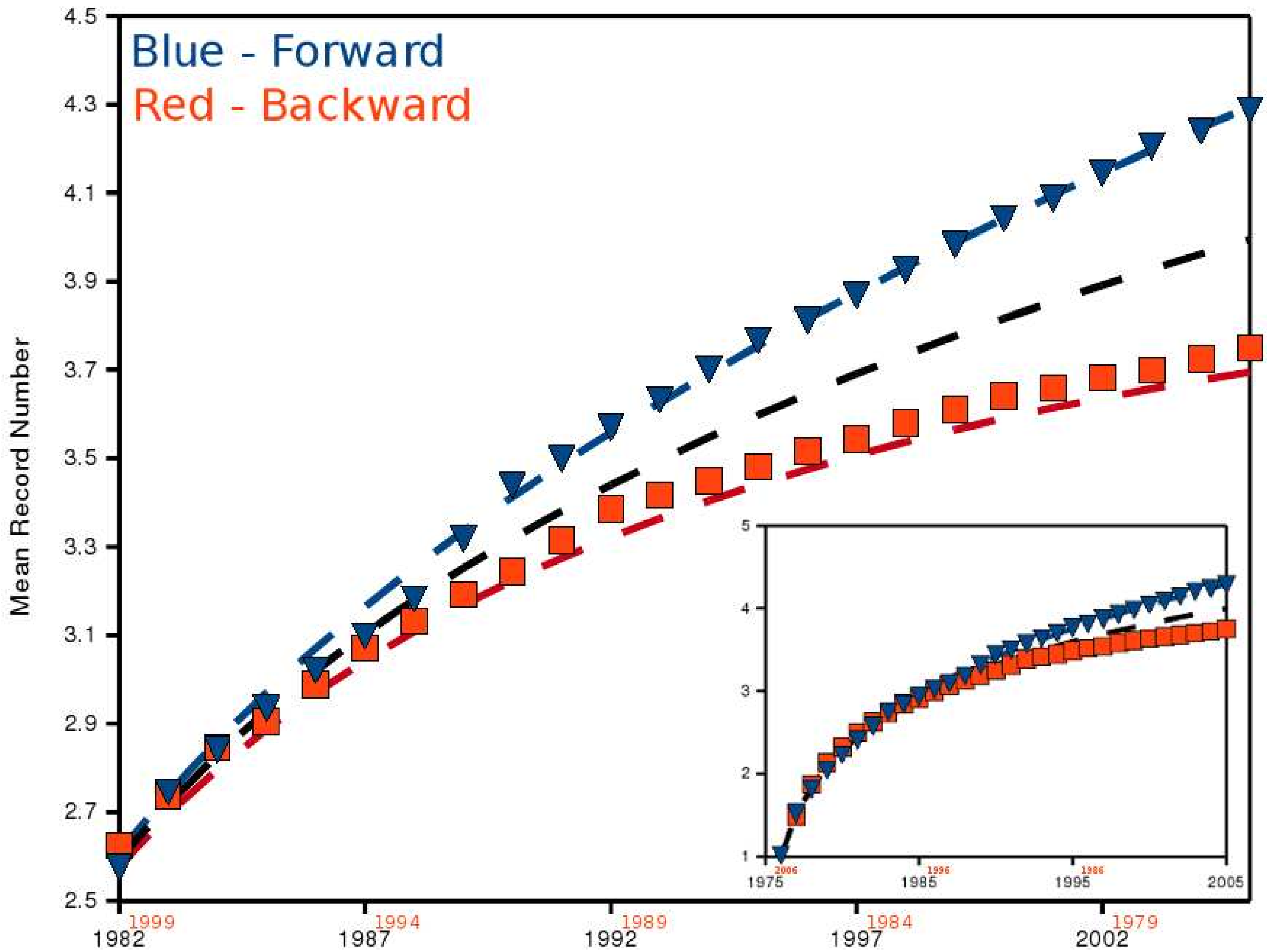
- 87 stations over 100 year period 1906-2005
- 207 stations over 30 year period 1976-2005
- Continental climate implies larger variability: $\sigma = 4.9 \pm 0.1^{\circ}$C, $v = 0.025 \pm 0.002^{\circ}$C/yr $\Rightarrow v/\sigma \approx 0.005$
- Significant effect of rounding to integer degrees Fahrenheit

Record frequency in Europe: 1976-2005



- Expected number of records in stationary climate: $\frac{365}{30} \approx 12$
- Observed record rate is increased by about 50 % $\Rightarrow$ 6 additional records

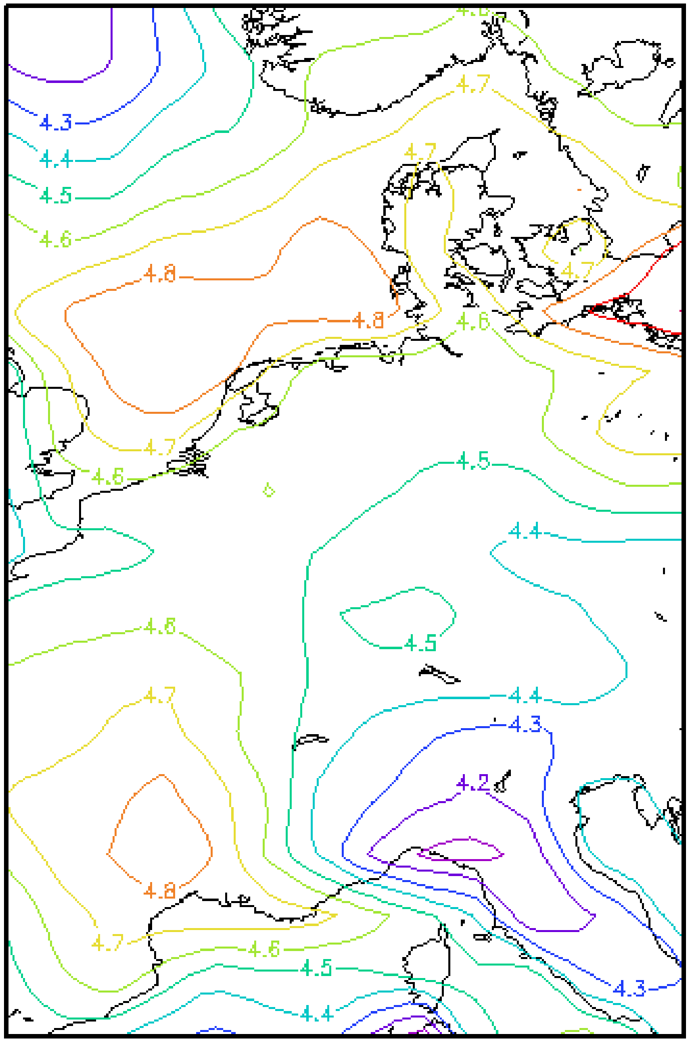
Mean record number: 1976-2005



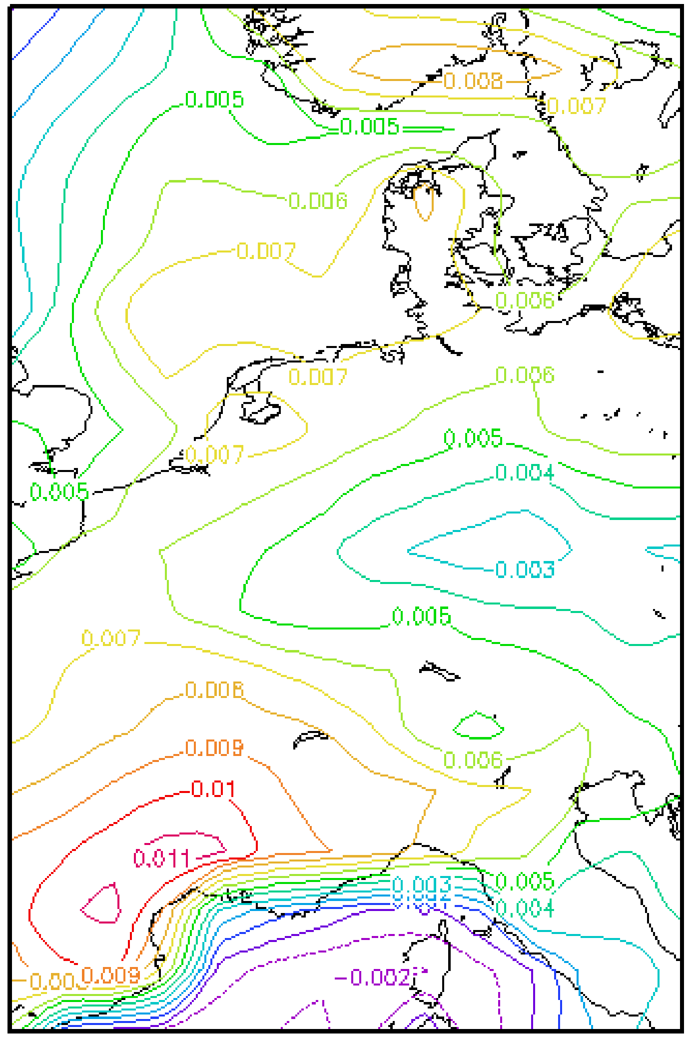
Inset: American data

Re-analysis data: Record maps

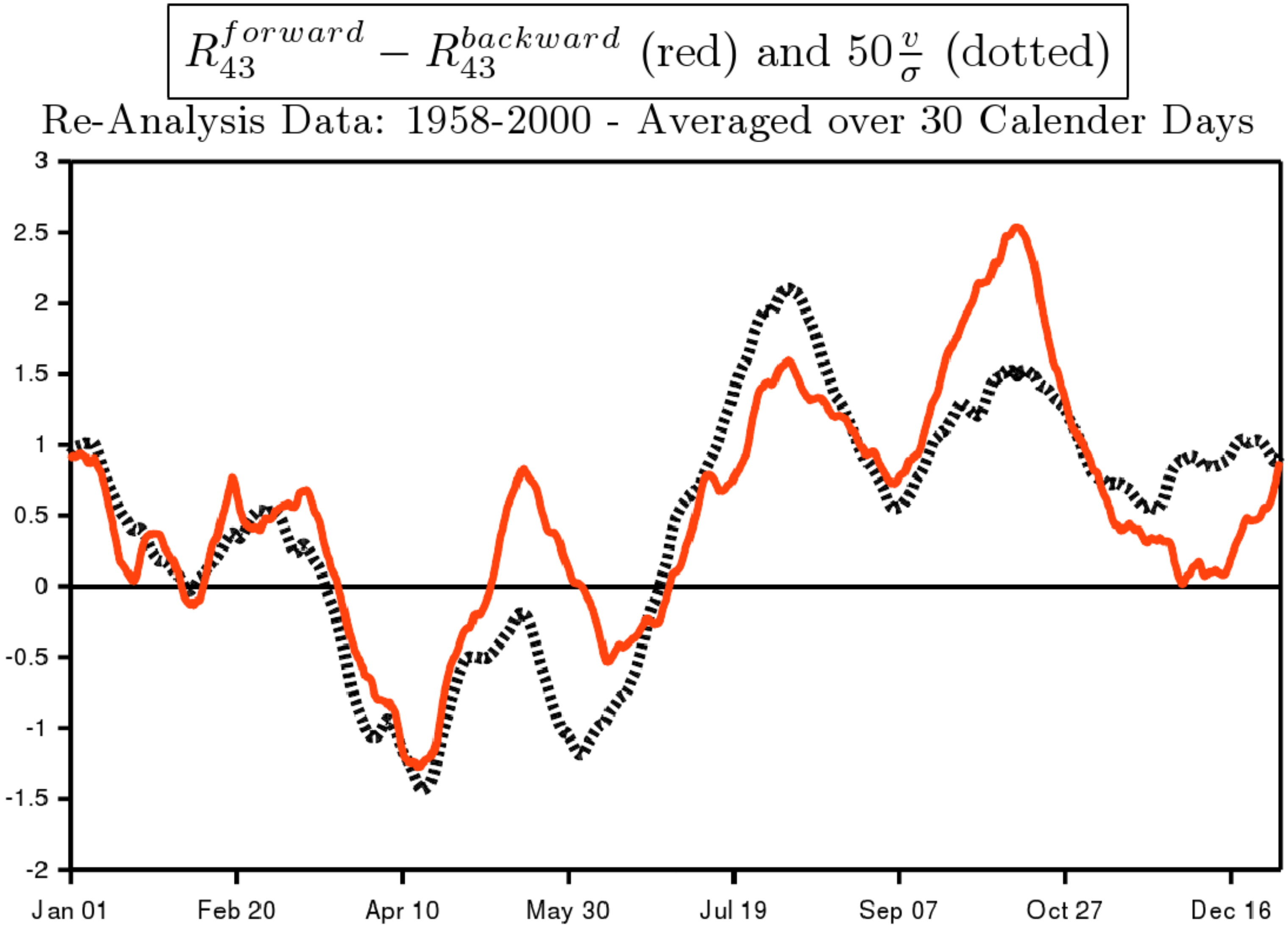
number of records 1957-2000



normalized warming rate v/σ



Re-analysis data: Seasonal variation



Random walks & market fluctuations

Records of random walks

S.N. Majumdar & R.M. Ziff, PRL **101**, 050601 (2008)

- Let X_n be an unbiased random walk:

$$X_n = \sum_{k=1}^{n} \eta_k$$

with i.i.d. RV's η_k drawn from a symmetric, continuous distribution $\phi(\eta)$

- The probability of having m records in n steps is given by

$$P(m,n) = \binom{2n-m+1}{n} 2^{-2n+m-1} \rightarrow \frac{1}{\sqrt{\pi n}} \exp[-m^2/4n]$$

- Mean number of records: $\langle R_n \rangle \approx \sqrt{4n/\pi}$

- This result does not require $\phi(\eta)$ to have finite variance
$\Rightarrow$ valid also for Lévy flights!

Random walks and stock market fluctuations

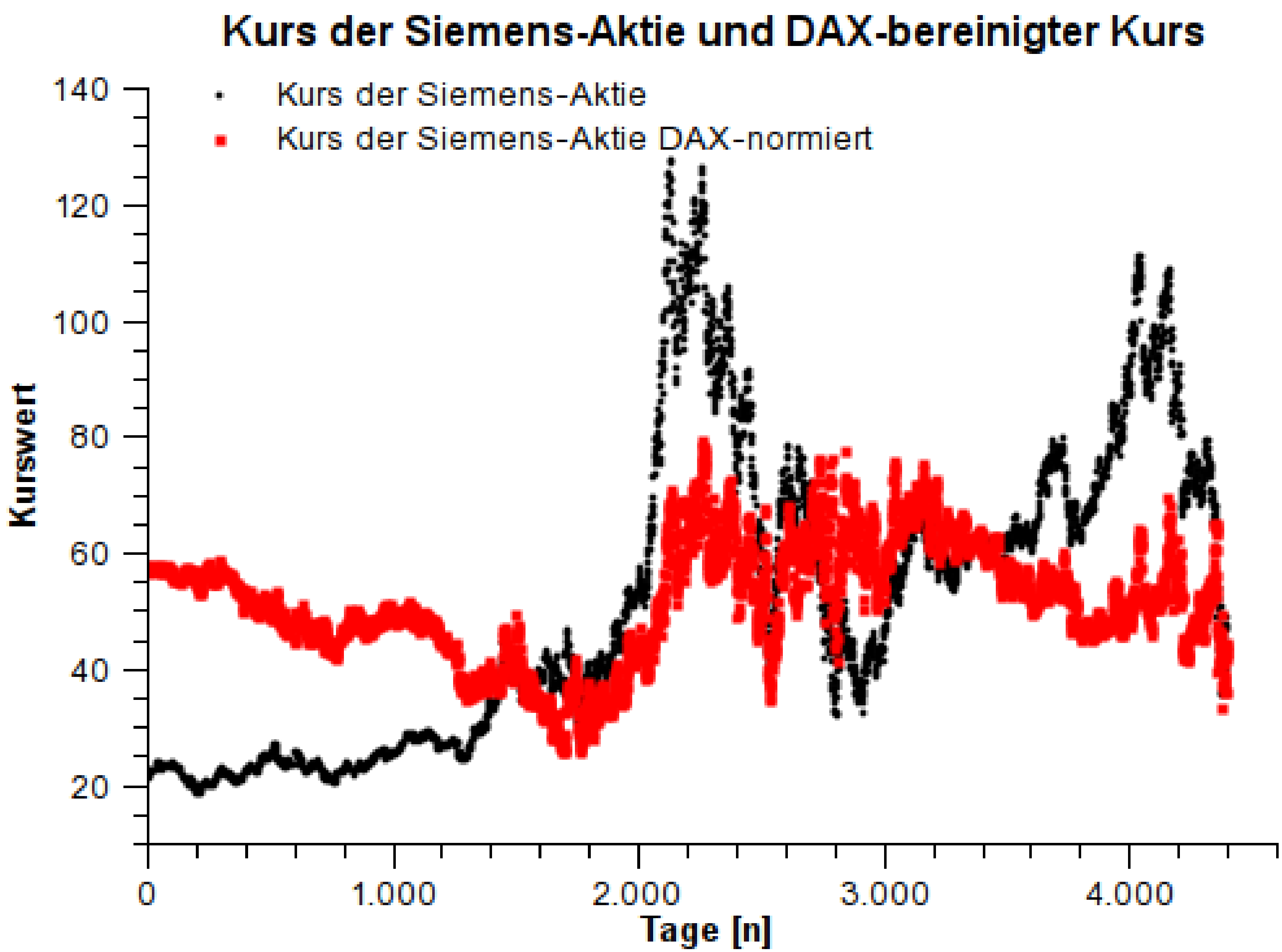
L. Bachelier, *Théorie de la spéculation* (1900)

- Basic model of a fluctuating stock price S_n is the exponential random walk

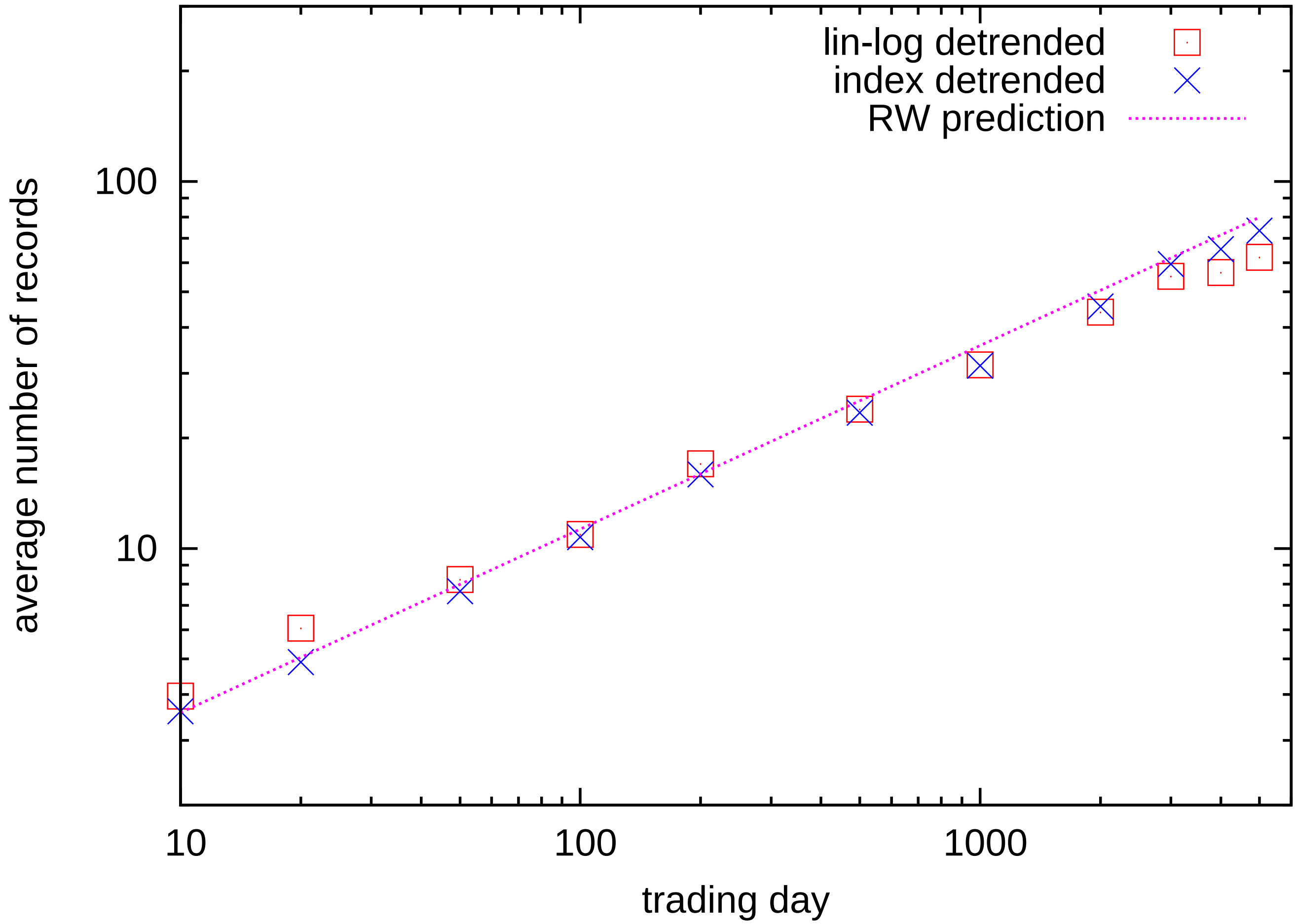
$$S_n = e^{X_n} = \exp[\sum_{k=1}^{n} \eta_k] \Rightarrow \eta_k = \ln(S_n/S_{n-1})$$

- Distribution of returns η_k display fat tails when viewed at high temporal resolution
- Key problem: How to distinguish trends and fluctuations?
- Standard approach removes a linear trend from $\ln(S_n)$
- Alternative: Normalize stock prices within an index by the index itself

Siemens stock normalized by DAX

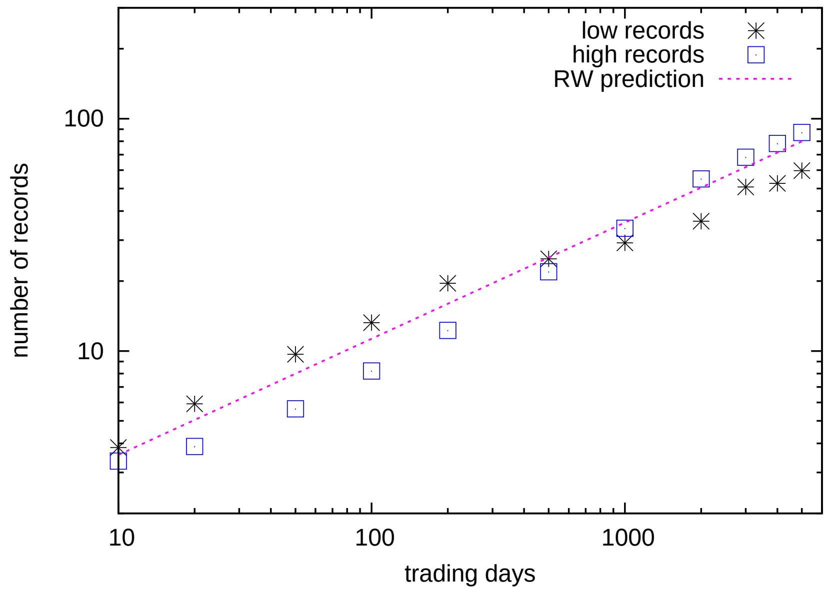


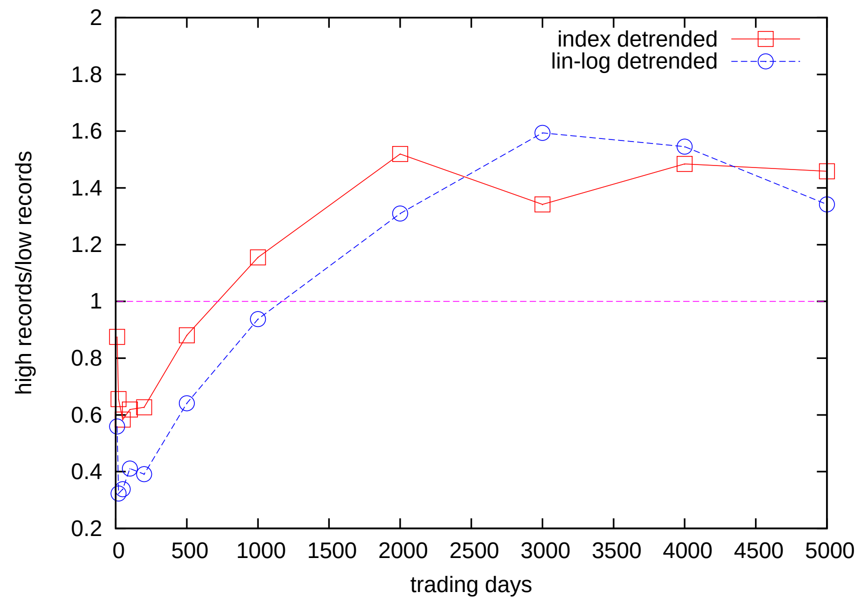
Average number of records in the S&P 500



- 366 stocks from S&P 500, 1.1.1990-31.3.2009
- Detrending by index seems to work better

High and low records of the index-detrended S&P 500





- Excess of lower (upper) records for short (long) times
- Ratio $\langle R_n \rangle_{\text{high}} / \langle R_n \rangle_{\text{low}}$ tends to a constant limit
- Asymmetry may reflect different market reactions to positive and negative price changes

Beyond the simple random walk

- Financial markets show volatility correlations, i.e. the variance of price increments depends on past market history.

- This effect is taken into account by ARCH and GARCH models
  ARCH = Autoregressive Conditional Heteroskedasticity

  Engle 1982, Bollerslev 1986

- In the simplest GARCH, the variance σ_n^2 of the n'th increment η_n is determined recursively through

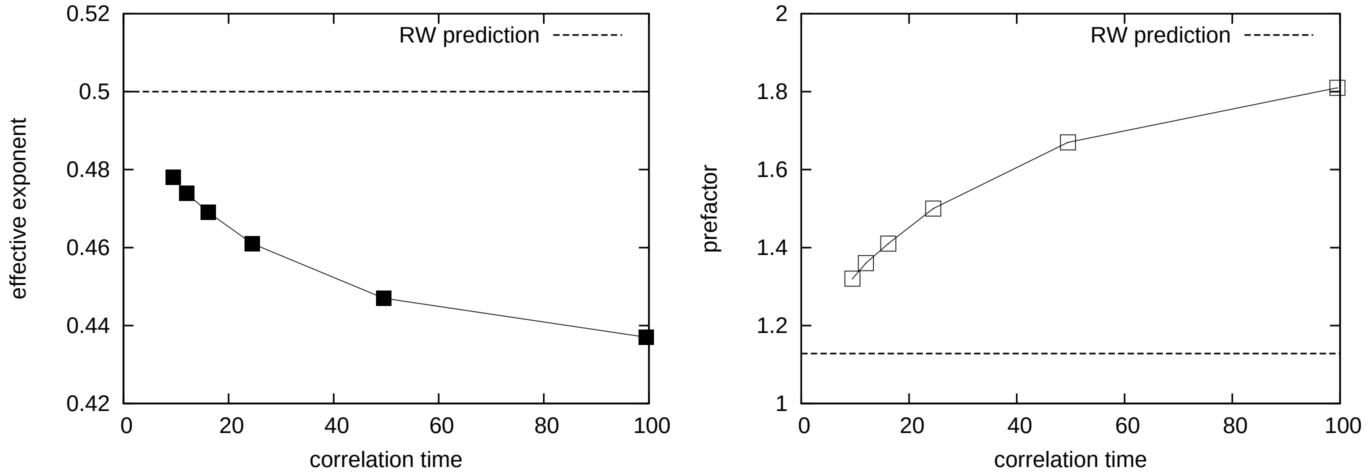
$$\sigma_n^2 = \alpha_0 + \alpha_1 \eta_{n-1}^2 + \beta_1 \sigma_{n-1}^2$$

  with constants $\alpha_0, \alpha_1, \beta_1 > 0$; the ARCH has $\beta_1 = 0$.

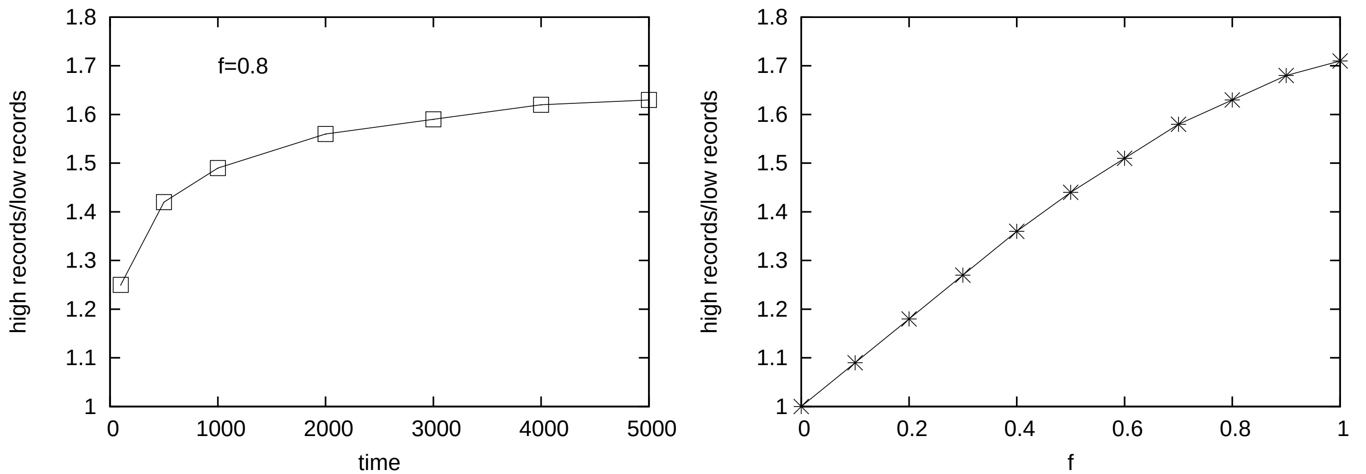
- The GARCH volatility correlation function is given by

$$\langle \eta_k^2 \eta_{k+n}^2 \rangle - \langle \eta_k^2 \rangle^2 = \frac{\alpha_0 (\alpha_1 + \beta_1)^n}{1 - \alpha_1 - \beta_1} \sim e^{-n/\tau} \quad \text{with} \quad \tau = |\ln(\alpha_1 + \beta_1)|^{-1}$$

- Volatility correlations reduce the exponent but increase the prefactor in the effective power law $\langle R_n \rangle \approx A n^{\nu}$



- An asymmetry between positive and negative increments is introduced in the QGARCH: $\sigma_n^2 = \alpha_0 + (\alpha_1 \eta_{n-1} - f)^2 + \beta_1 \sigma_{n-1}^2$ Sentana 1995



Conclusions

- Records statistics as a paradigm of non-stationary dynamics of rare events
- Global warming affects the rate of record-breaking temperatures in a moderate but significant way
- Records in financial data conform to the basic random walk model, with some additional features
- Theory of processes beyond i.i.d. RV's largely remains to be developed