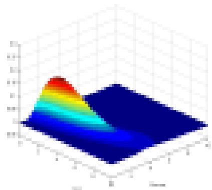
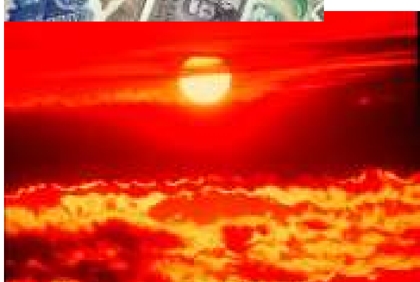
# Trivariate Models for Stochastic Episodes with Applications to Hydrology, Climate and Finance





**Wrocław University of Technology**

**21 February, 2011**

**Anna K. Panorska**
**University of Nevada Reno, USA**

**with**
**Tomasz J. Kozubowski and Fares Qeadan**

***"From the tropics to the arctic, climate and weather have powerful direct and indirect impacts on human life. Extremes of heat and cold can cause potentially fatal illnesses, .... Other weather extremes, such as heavy rains, floods, and hurricanes, also have severe impacts on health. Approximately 600 000 deaths occurred worldwide as a result of weather-related natural disasters in the 1990s."***

**The WHO, 12 November, 2008**

# Outline


- **Motivation**: **Extreme hydroclimatic and weather events, e.g. drought, flood, or heat waves; Financial modeling.**
- **Examples**: **(1) Paris heat wave, August 2003.**
  - **(2) Great Drought of the 1930s in the USA, THE "DUST BOWL",**
  - **Financial modeling – stability "with respect to growth period"**

- **Stochastic Models**: **First joint model for duration N, magnitude X and maximum Y of events: (N, X, Y)**
- **Mathematical Results: The model for (N, X, Y):**
  - **Stochastic representation, pdf;**
  - **Marginals: one dimensional and bivariate; conditionals;**
  - **Correlation structure**
  - **Distribution of the ratio max/sum.**
- **Examples: Back to the "DUST BOWL", heat wave, and financial modeling.**
- **Summary**

# Motivation: Climate and hydrology - drought

**"DUST BOWL" - Great drought of the 1930s in the USA, the setting of John Steinbeck's "Grapes of Wrath".**




**What are the chances that "Dust Bowl" will happen again?**

**What are the chances of a drought with duration exceeding 10 years?**

**What are the chances that a 5 years long drought will have magnitude exceeding "Dust Bowl's" and a large maximum?**

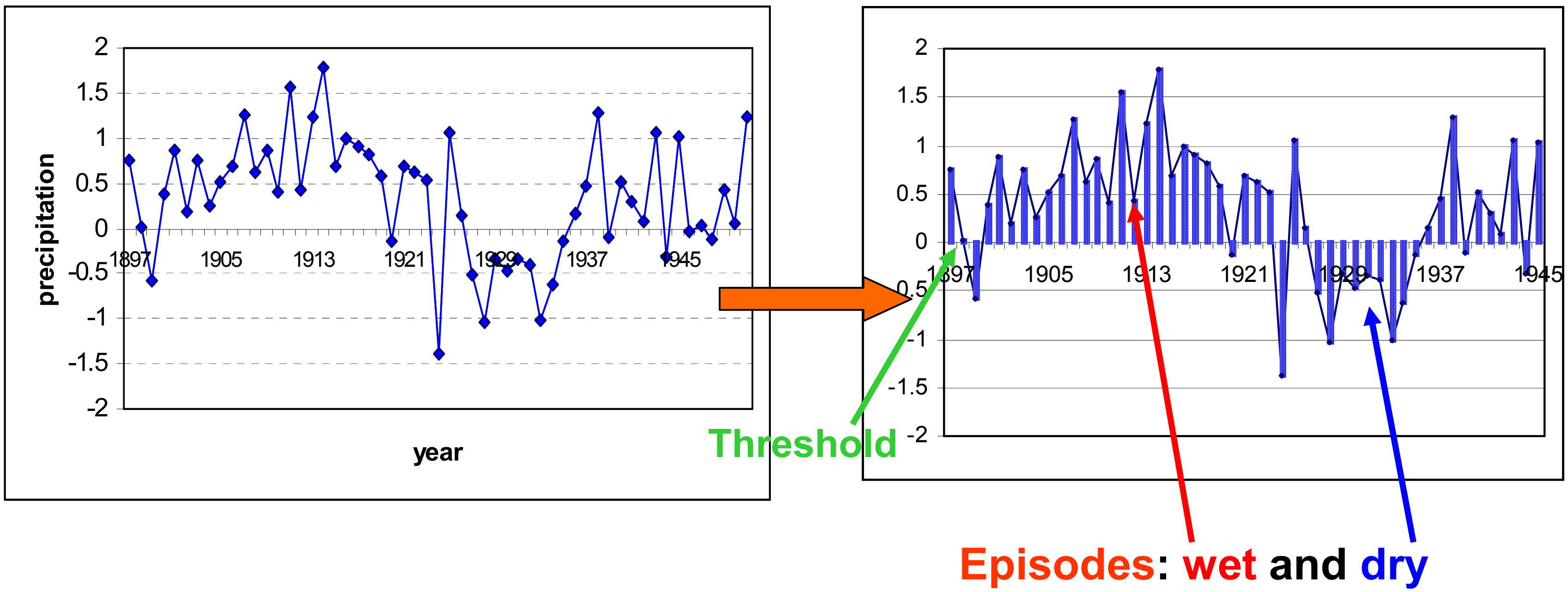


(Kozubowski, Panorska (2005, 2007), Biondi et al. (2007), Saito et al. (2008))

# Motivation – drought, contd.

**Original precipitation data in standard deviation units plotted as difference from a threshold.**

**Precipitation “events” or “episodes”.**

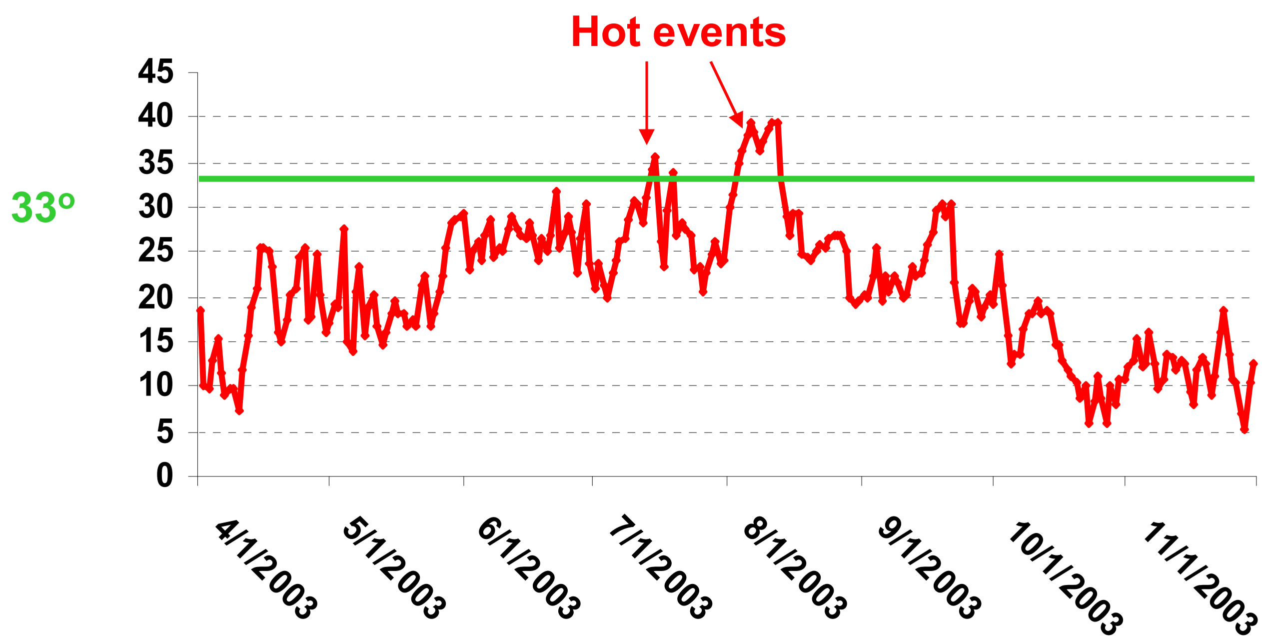



**An episode is a period with the process staying consecutively above/below threshold: e.g. “dry”, “wet” year, drought, flood, etc.**

**Threshold for “dry” or “wet” depends on the definition of the episode (e.g. drought).**

# Motivation: Climate and weather – heat waves

- **In August 2003, France experienced an extreme heat wave, that resulted in an estimated 14,802 deaths*.**
- **Hot event: consecutive observations above the 33°C.**




**What are the chances that a large heat wave will happen again?**

**What are the chances of a hot event with duration equal to this one, 11 days?**

**What are the chances that a heat wave longer than 6 days and larger than current 98$^{th}$ percentile of heat waves' magnitudes?**



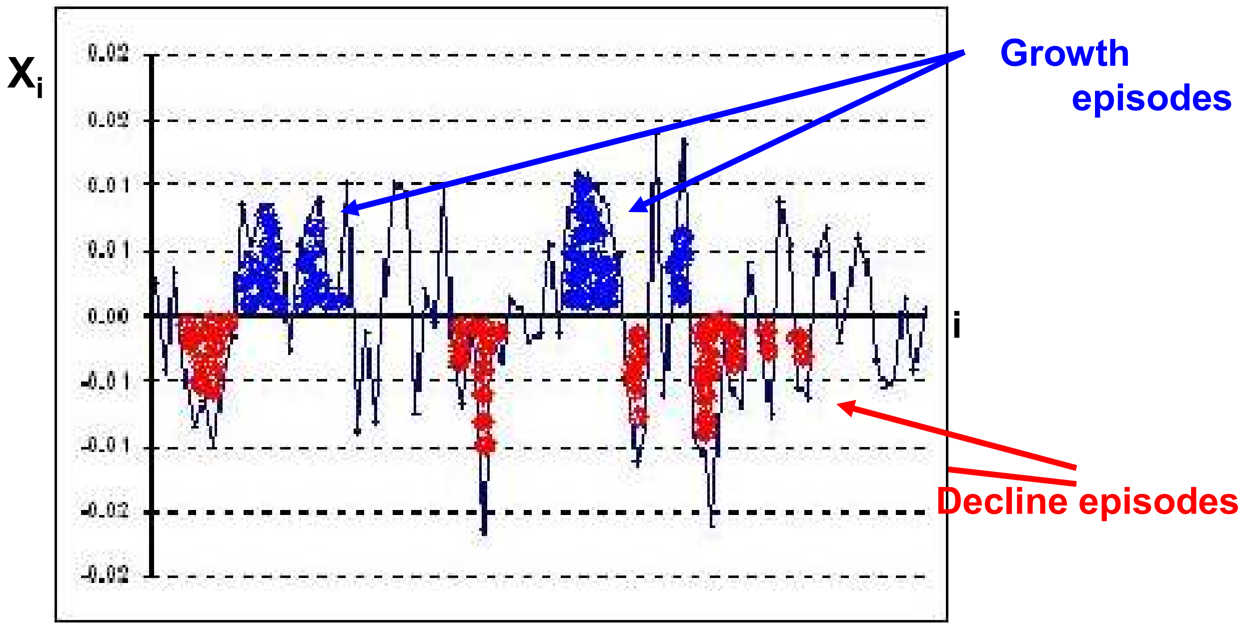
*Dhainaut et al. 2004.

# Motivation: Financial growth/decline episodes

**Daily exchange rates between Japanese Yen and British pound quoted in UK pound, Jan. 2, 1980-May 21, 1996.**

**Process Xi: Daily log returns n=4274,**
**$X_i$=log(Rate_day_i/Rate_day_i-1)**

**Episodes: consecutive days when the exchange rates were growing/declining, i.e.:**
**Growth $X_i$ >0, Decline $X_i$ <0.**




**N = length of a growth period, X=ΣXi=cumulative log return over a growth period, Y=maximum log return over the growth period.**

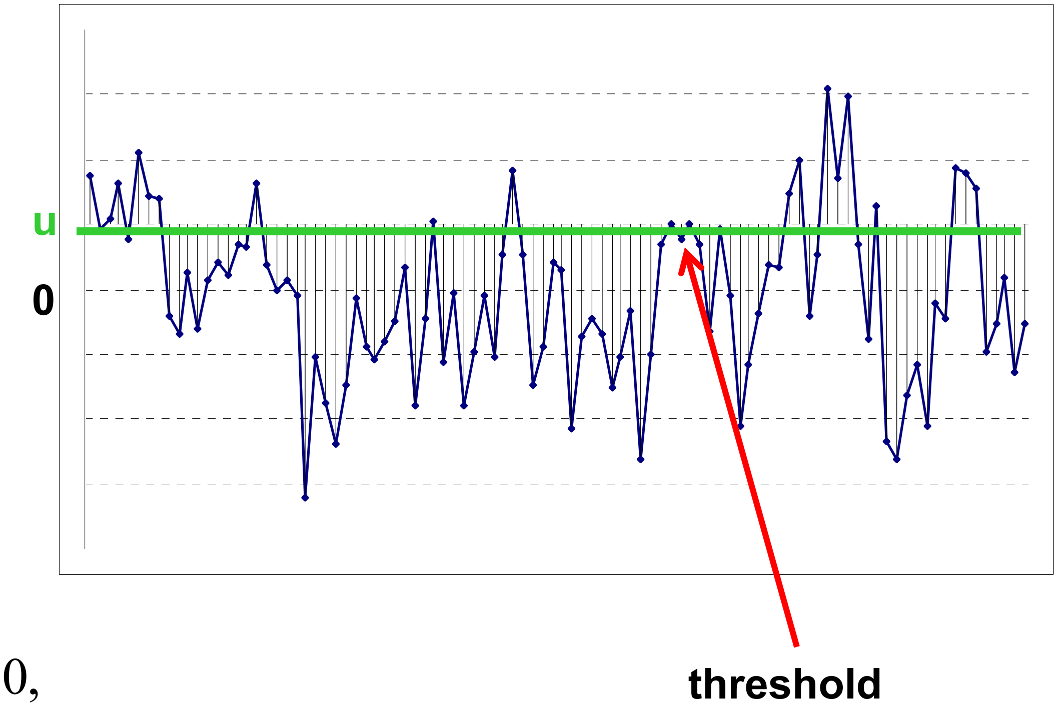


Nolan 2001, Kozubowski and Podgorski 2001, Kozubowski and Panorska 2004, 2007, 2011

# Motivation – theoretical: Peak over Threshold theory

**Take a process $X_i$ , with cdf F, and a threshold u. Compute excesses $X_i$-u. Interested in the distribution of the excesses when the threshold u increases.**

**Theorem: Balkema-deHaan(1974)-Pickands(1975). If the cdf $F^{[u]}(b_u + a_u x)$ of the excesses has a continuous limit as u increases, then the limit is one of three types of distributions: cdf's (in standard forms):**




$$\text{Exponential: } G(x) = 1 - e^{-x},\ x \geq 0,$$

$$\text{Pareto: } G(x) = 1 - \frac{1}{x^{\alpha}},\ x \geq 1,\ \alpha > 0,$$

$$\text{Beta: } G(x) = 1 - \frac{1}{(-x)^{\alpha}},\ -1 \leq x \leq 0,\ \alpha > 0.$$

**For "light tailed" distribution of $X_i$, the limiting distribution of the excesses will be exponential.**

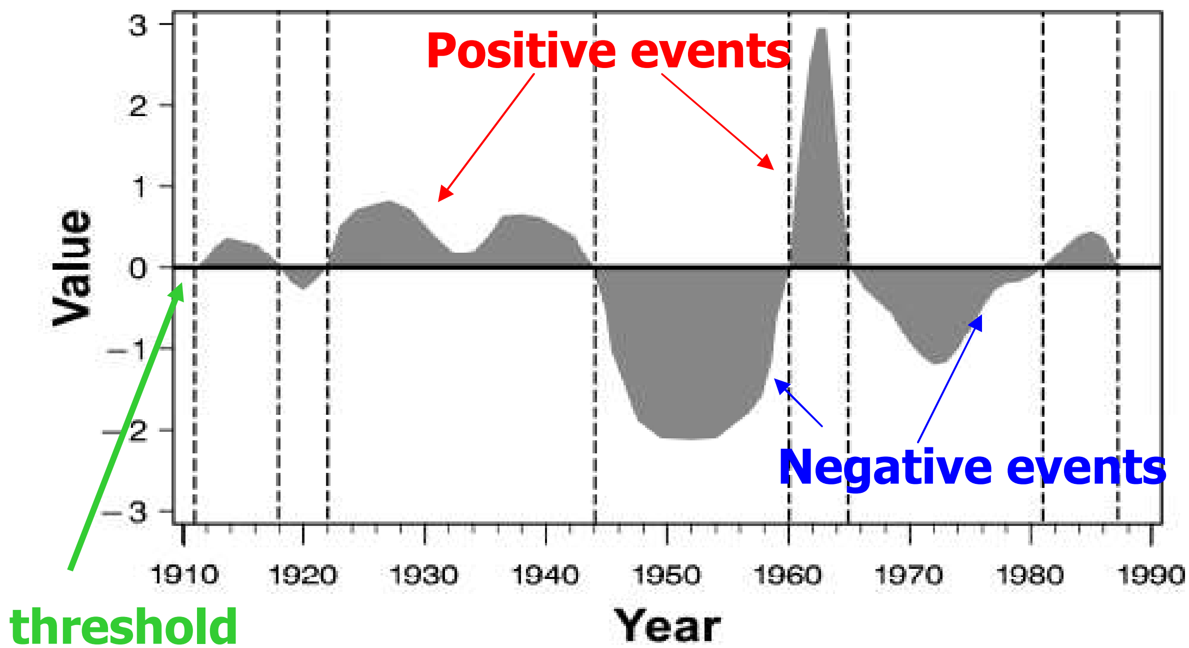
# Summary up to now

**Start with a process:**

$$X_1, X_2, \ldots, X_k,$$

**DURATION = the number of years/time periods in one event : random = N.**

$$\text{Duration} = N$$




**MAGNITUDE = area of the shaded region = random sum of the series values for one event.**

$$\text{Magnitude} = X = \sum_{i=1}^{N} X_i$$

**MAXIMUM = max observation during an event = random maximum of the series values for one event.**

$$\text{Max/Peak} = Y = \max_{i=1\ldots N} X_i$$



**Kozubowski and Panorska (2005), (2008), Biondi, Kozubowski and Panorska (2005), Biondi et al. (2008).**

# Random vector $(N, \sum_{i=1}^{N} X_i, \max_{1 \le i \le N} X_i) = (N, X, Y)$

**GOAL: Construct a mathematically natural model for the JOINT distribution of (duration, magnitude X and maximum Y) of events.**

**Notable properties of the random vector (N, X, Y):**

- **All components are related/dependent, the joint behavior of X and Y is not trivial,**
- **The sum and maximum are of random number of random observations.**

**Hierarchical approach:**

1. **Specify distribution of N**
2. **Given N=n, find conditional distr.** $f(x, y \mid n)$ **of (X, Y|N=n) =**$(\sum_{i=1}^{n} X_i, \max_{i=1,\ldots,n} X_i)$

3. **Get the joint distribution of (N, X, Y) as**

$$f(n, x, y) = f(x, y \mid n) f_N(n)$$

# Joint Distribution of $\left(\sum_{i=1}^{n} X_i, \max_{i=1,\ldots,n} X_i\right) = (X, Y)$

History: Start with X1 , …, Xn iid random variables, and known (nonrandom) n

GOAL: Find the joint distribution of  (X, Y)

*Only asymptotics* for large n were available: $X_i$ iid, X in sum DoA of Normal, Y in max DoA of some EV distribution, then

$$\left(\sum_{i=1}^{n} X_i,\ \max_{1\le i\le n} X_i\right)$$

Appropriate normalization

$$( X_\infty \quad , \quad Y_\infty )$$

Normal

independent

Extreme value

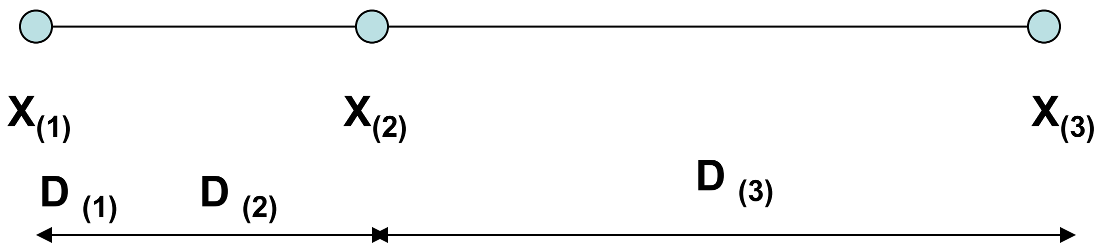
Main tool for limit theory is the hybrid characteristic -distribution function:

$$\psi_{(X,Y)}(t,v) = E\{(e^{itX} I(Y \le v)\} = \left(\frac{1-e^{(it-\beta)v}}{1-it/\beta}\right)^n,\ t \in R,\ v \ge 0.$$

Anderson and Turkman (1991ab, 1992, 1995), Chow, T.L. and Teugels, J.L. (1979), Haas (1992), Ho and Hsing (1996), Hsing (1995ab), Mathew and McCormick (1998), Mori (1981).

# Intuitive understanding for the vector (X, Y)

**$X_1$, $X_2$, $X_3$ iid exp(β)**
**Consider n=3. Take order statistics $X_{(1)}$, $X_{(2)}$, $X_{(3)}$, and spacings**
**$D_{(1)}= X_{(1)}$, $D_{(2)}= X_{(2)} - X_{(1)}$, $D_{(3)}= X_{(3)} - X_{(2)}$.**




**X=ΣXi= 3D(1)+2 D(2)+ D(3), Y=maxXi= X(3)= D(1)+ D(2)+ D(3).**

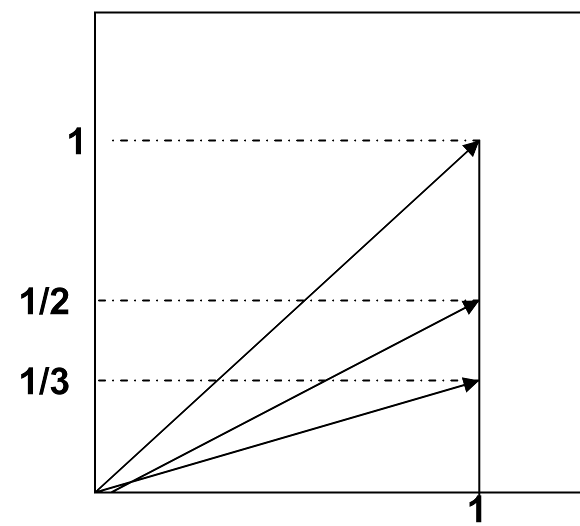
**Take Wi=(n-i+1) Di iid exp(β), and**

**Di=Wi/(n-i+1) are independent exponential,**

**but not identically distributed.**

**X=ΣXi= Σ Wi, Y= maxXi = Σ (Wi/i), so**

**(X, Y)= Σ [Wi(1, 1/i)].**



# Stochastic Representation

**Theorem. For any n ≥ 1 and every i = 1, 2, …, n, let $a_i=(1, 1/i)$. Then a BGGE(β,n) random vector (X, Y) admits stochastic representation**

$$(X,Y)\stackrel{d}{=}\left(\sum_{i=1}^{n} W_i, \sum_{i=1}^{n}(W_i/i)\right)=\sum_{i=1}^{n} a_i W_i,$$

**No max, only sums in bivariate case!**

**where the $\{W_i\}$ are IID exponential random variables with parameter β.**

**This shows that (X, Y) is infinitely divisible.**

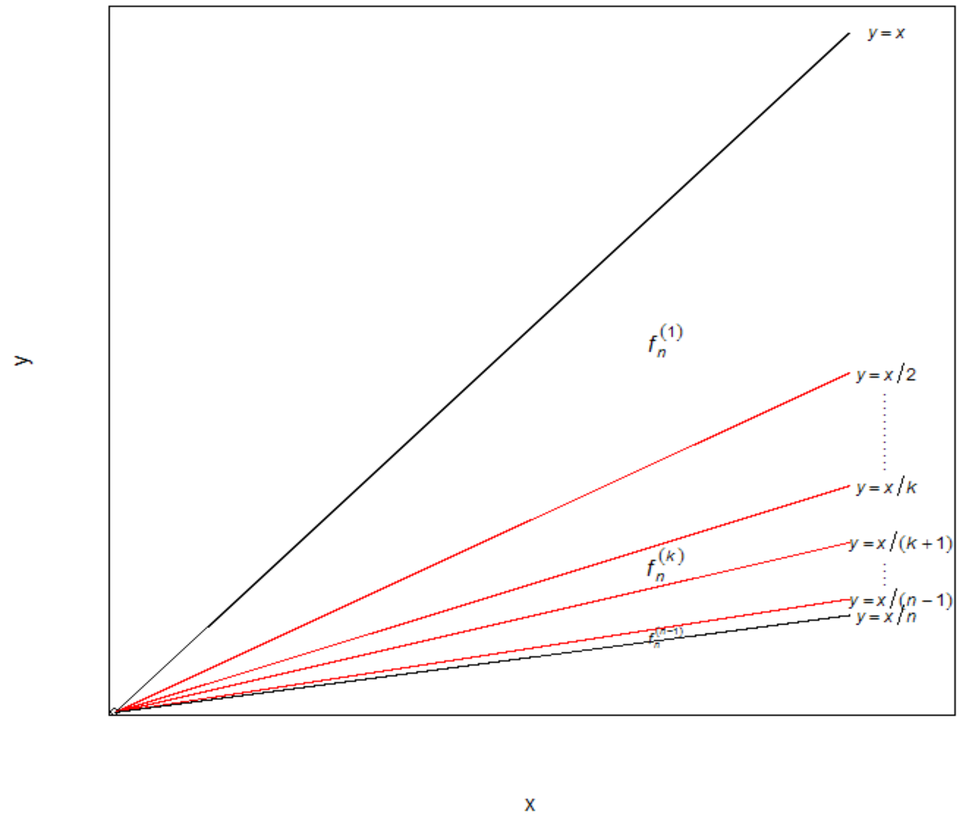
# KEY NEW RESULT: PDF of the Bivariate Distribution of the Random Sum and Max for any n

- **Domain of the pdf**

  **$x/n \le y \le x$**

- **PDF formula $f_n^{(k)}$ will depend on the sector k of the plane.**

- **There are (n-1) sectors:**




**$S_0$={(x,y): x=y>0} and**

$$S_k = \{(x,y) \in R^2 : 0 \le \frac{1}{k+1} x \le y < \frac{1}{k} x\},\ k = 1,2,\ldots,n-1.$$

## KEY NEW RESULT: PDF of the Bivariate Distribution of the Random Sum and Max for any n – contd.

**Let $X_i$, i=1, …, n, n ≥ 2, be iid exp(β). The joint pdf of (X, Y)=** $(\sum_{i=1}^{n} X_i, \max_{i=1,\dots,n} X_i)$ **is given explicitly by**

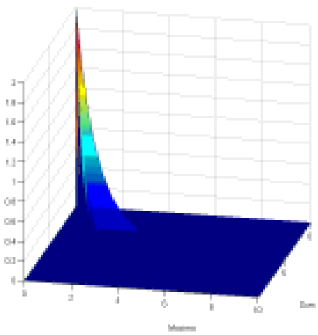
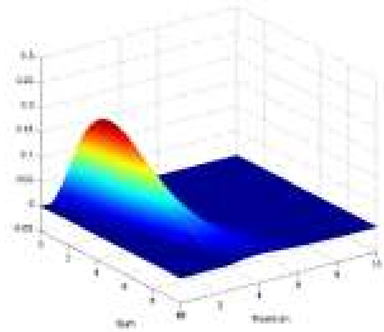
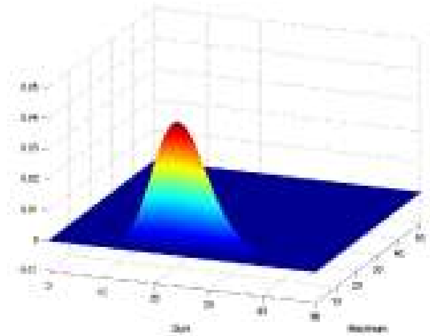
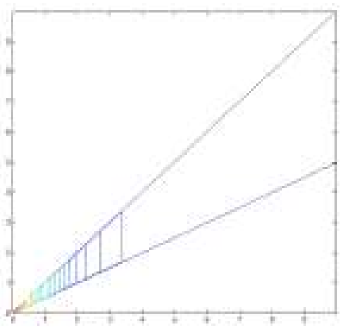
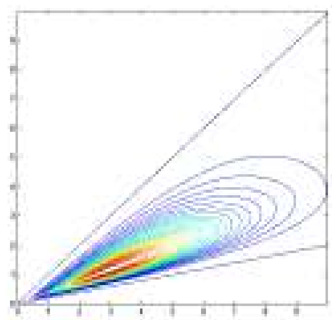
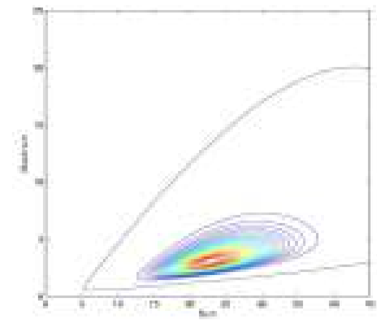
$$f(x, y) = \beta^n e^{-\beta x} H(x, y, n), \text{ where}$$

$$H(x,y,n) = \begin{cases} \sum_{s=1}^{k} \frac{n(n-1)}{(s-1)!(n-s)!}(x-sy)^{n-2}(-1)^{s+1} & \text{for } n \geq 2, (x,y) \in S_k \\ 1 & \text{for } n = 1, (x,y) \in S_0 \\ 0 & \text{otherwise.} \end{cases}$$

**This new distribution is called BGGE(β, n) for bivariate distribution with gamma and generalized exponential marginals.**

**Qeadan, Kozubowski, Panorska (2010), Communication in Statistics: Theory and Methods.**

# Densities

| n=2 | n=5 | n=25 |
|---|---|---|
|  |  |  |
|  |  |  |

# Marginal Distributions

**Univariate**

**Marginal distribution of the sum X is Gamma(n, β).**

**Marginal distribution of the maximum Y is generalized exponential GE(n, β), (Gupta and Kundu, 2007) with cdf**

$$F(y) = (1 - e^{-\beta y})^n, \quad y > 0.$$

**Bivariate**

**Marginal distributions of (N, X) and (N, Y).**

$$(N, \mathrm{X}) = \underbrace{\left( N, \ \sum_{i=1}^{N} \mathrm{X}_i \right)}_{\mathrm{BEG}(\beta,\mathrm{p})}$$

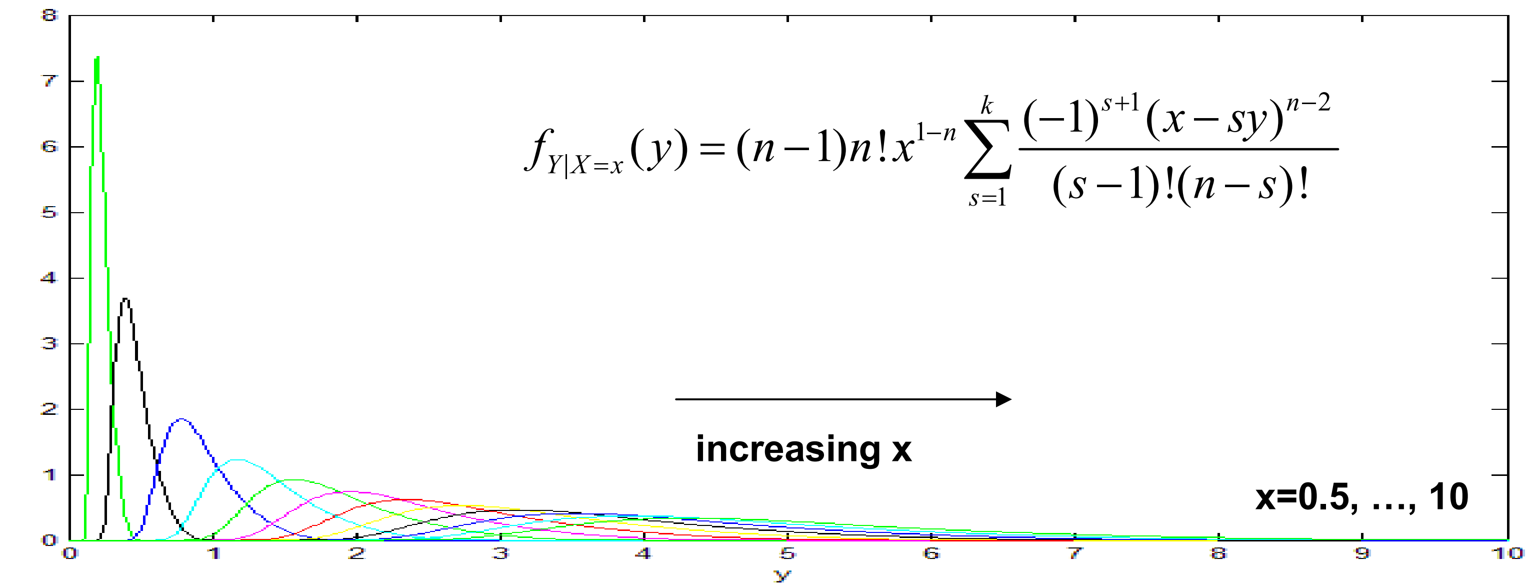
$$(\mathrm{N}, \mathrm{Y}) = \underbrace{\left( N, \ \max_{1 \le i \le n} \mathrm{X}_i \right)}_{\mathrm{BTLG}(\beta,\mathrm{p})}$$

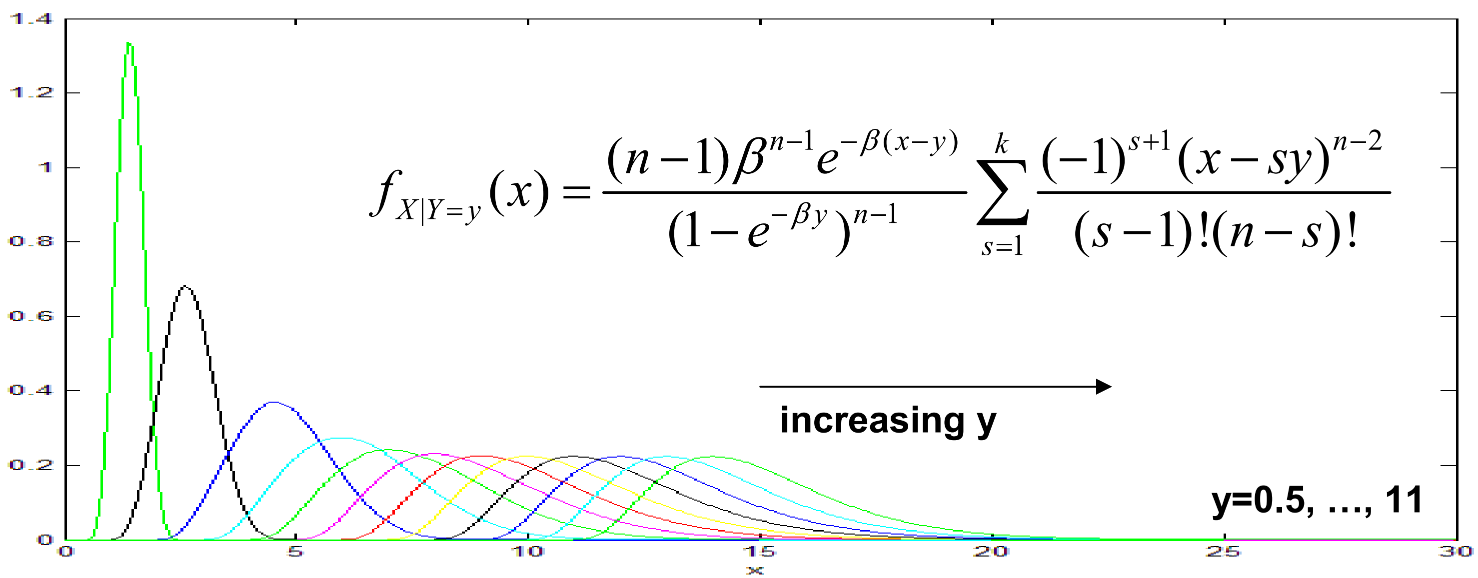
**BEG: Bivariate distribution with exponential and geometric marginals**

**BTLG: Bivariate distribution with truncated logistic and geometric marginals**

Kozubowski, Panorska (2005) (2008), Biondi, Kozubowski, Panorska (2005), Saito, Biondi, Kozubowski, Panorska (2008)

# Conditional Distributions

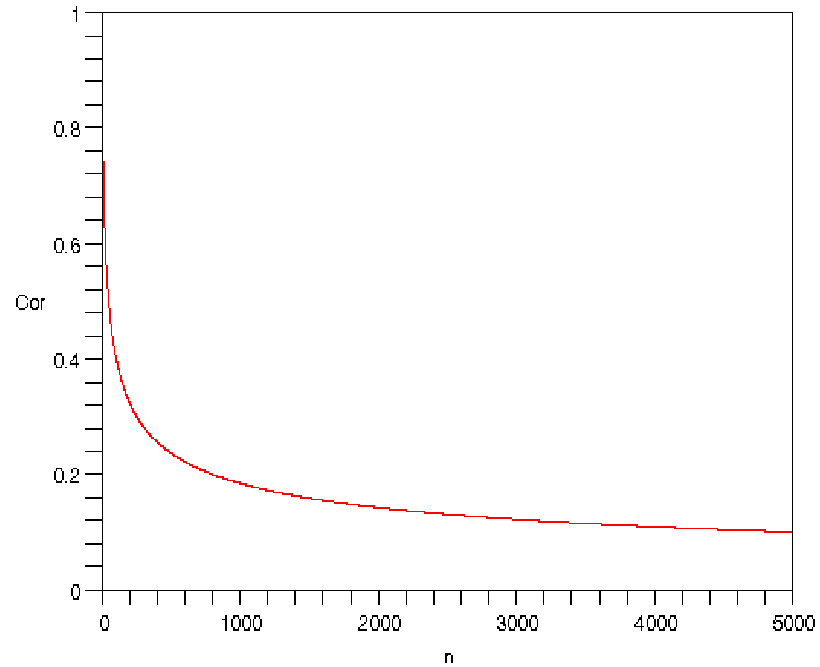


# Covariance Structure

**(X, Y)~ BGGE(β, n). EX=n/ β, EY=(ψ(n+1)+γ)/ β, corr(X, Y) is**

$$\rho_n = \frac{\psi(n+1)+\gamma}{\sqrt{\dfrac{n\pi^2}{6} - n\psi'(n+1)}},$$




**where ψ and ψ' are the digamma and polygamma functions, respectively, and γ is the Euler constant.**

**NOTE: $\rho_n \to 0$, but very slowly.**

**Thus, approximation of (X, Y) with independent X and Y is not good even for large n. E.g. $\rho_{500}$ =0.24**

# Distribution of the Ratio Y/X

**History. Only limiting results were done for the ratio**

**New Results.**

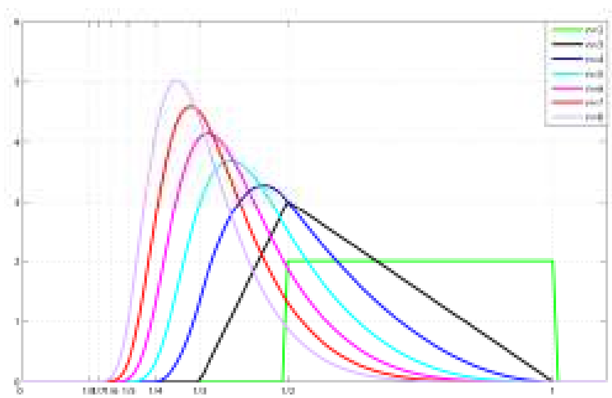
**Theorem. If (X, Y) is BGGE(β,n), then**

$$(X, Y) \stackrel{d}{=} (X, RX),$$

**where X and R are independent random variables, X has Gamma(n, β) distribution, and R has pdf of the ratio Y/X.**

**Theorem. Distribution of the ratio is parameter free, with density:**

$$h_n(t) = n(n-1)\sum_{s=1}^{k}\binom{n-1}{s-1}(1-st)^{n-2}(-1)^{s+1}, \quad \text{for } \frac{1}{k+1} \le t \le \frac{1}{k}, \quad \text{k=1, 2, …, n-1.}$$



Morrison and Tobias 1965, O'Brien 1980, Maller and Resnick 1984, Haas 1992, Kesten and Maller 1994.

# Trivariate Model: Joint Distribution of (X, Y, N)

N-random N-Geo(p), Xi's iid exp(β) ind. of N.

$$(\mathrm{X}, \mathrm{Y}, \mathrm{N}) = \underbrace{\left( \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{X}_{\mathrm{i}}, \max_{1 \le i \le N} Y_i, N \right)}_{\mathrm{TETLG}(\beta,\mathrm{p})}$$

**T**rivariate distribution with **e**xponential, **t**runcated **l**ogistic and **g**eometric marginals[1].

The pdf of (X, Y, N) is $f(x, y, n) = \beta^n e^{-\beta x} p(1-p)^{n-1} H(x, y, n),$

where function H is given in the pdf of the (X,Y|N=n)

[1] Kozubowski, Panorska, Qeadan (2011), JSPI, in press.

# Estimation of Parameters

**MLEs for β and p exist in explicit forms.**

$(X_1, Y_1, N_1),\ \ldots,\ (X_m, Y_m, N_m)$ **random sample from TETLG(β, p) distribution.**

**The MLE's of β and p are:** $\hat{\beta}_m = \overline{N}_m / \overline{X}_m$ and $\hat{p} = \dfrac{1}{\overline{N}_m}.$

**The vector MLE is consistent, asymptotically normal, and asymptotically efficient.** **The asymptotic covariance matrix is known.**

**Note: The MLEs do not depend on $Y_i$'s.**

# Application 1: Financial Data

**For the exchange rates data constructed data set of episodes.**

**Growth Xi >0, decline Xi <0**

**For analysis, considered episodes of growth: Xi >0.**

**Our data were 1902 triples (X, Y, N) of growth episodes.**

**We checked that:**

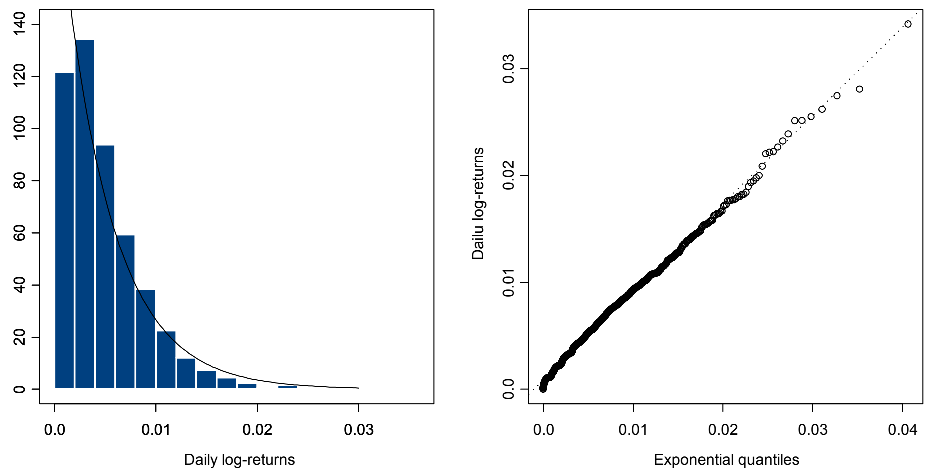
- **Positive returns come from exponential distribution;**
- **Magnitudes of growth periods also come from exponential distribution;**
- **The fit of all bivariate marginals is quite reasonable;**
- **Fit of conditional distributions is reasonable**

***Stability of the returns wrt growth periods*: Cumulative positive returns have the same distribution as the returns.**

Nolan 2001, Kozubowski and Podgorski 2001, Kozubowski and Panorska 2004, 2008, 2011.

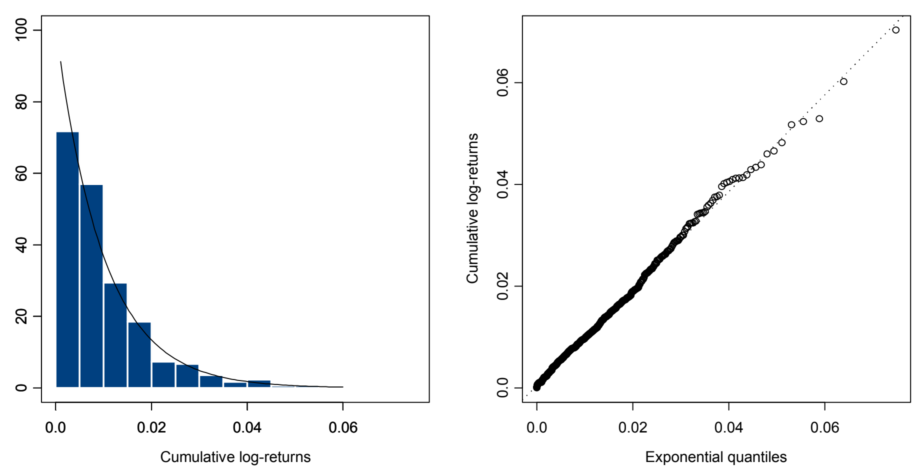
# Modeling financial data

**Daily positive log-returns $X_i$ fit with exponential distribution**




**Cumulative positive log-returns**

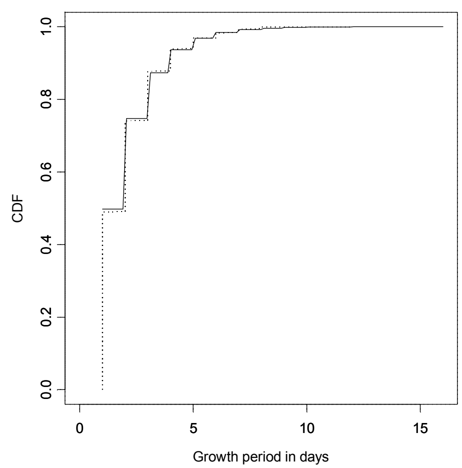
**$X=\Sigma X_i$ fit with exponential model**



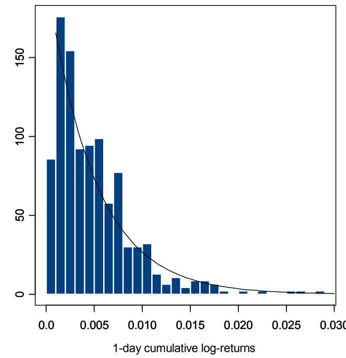
# Modeling financial data: Fit of BEG model to $(N, X=\Sigma X_i)$

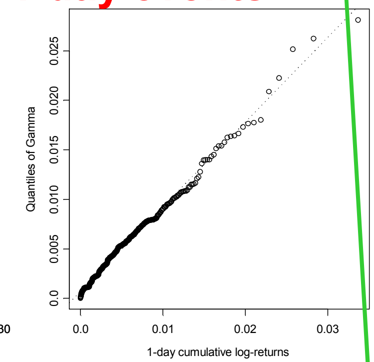
**Geometric fit to durations**

**Gamma fit to the conditional distributions**

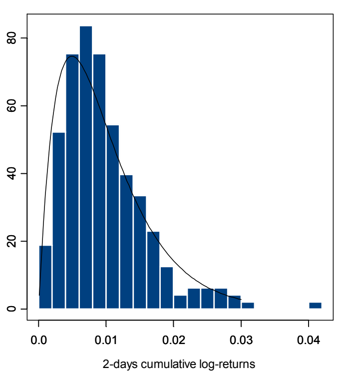


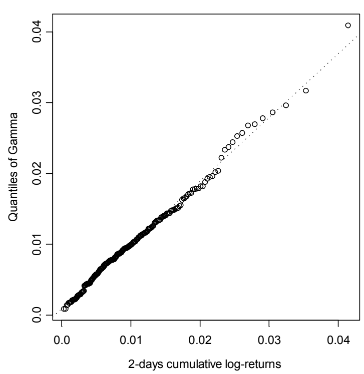

**1-day events**



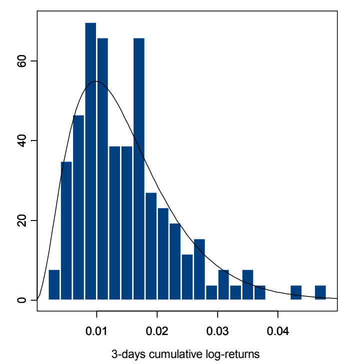



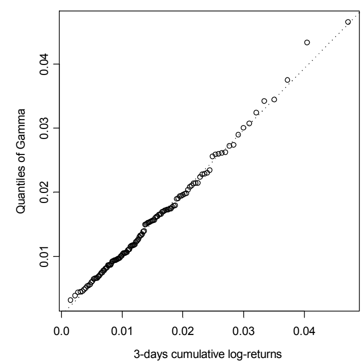

**2-day events**




**3-day events**

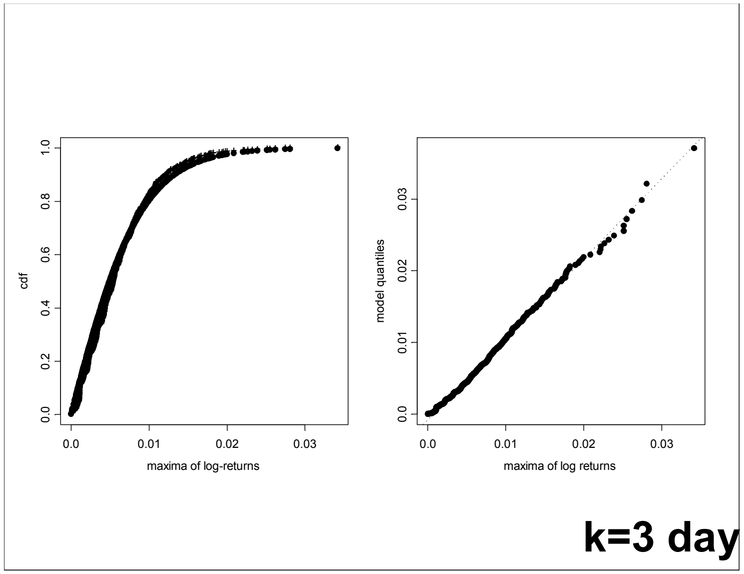


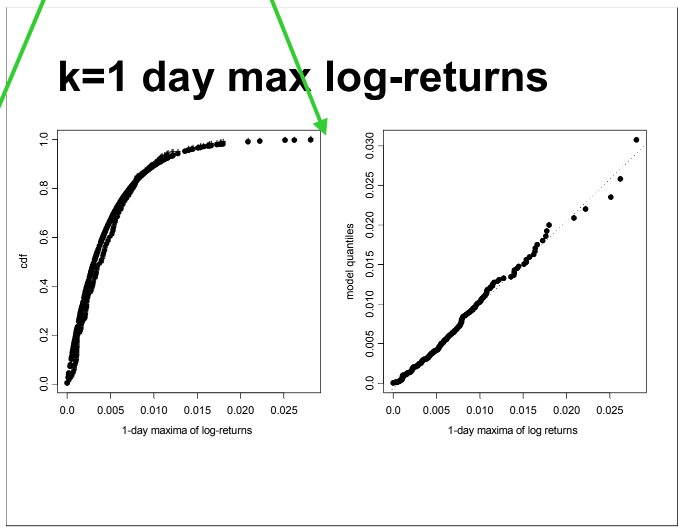



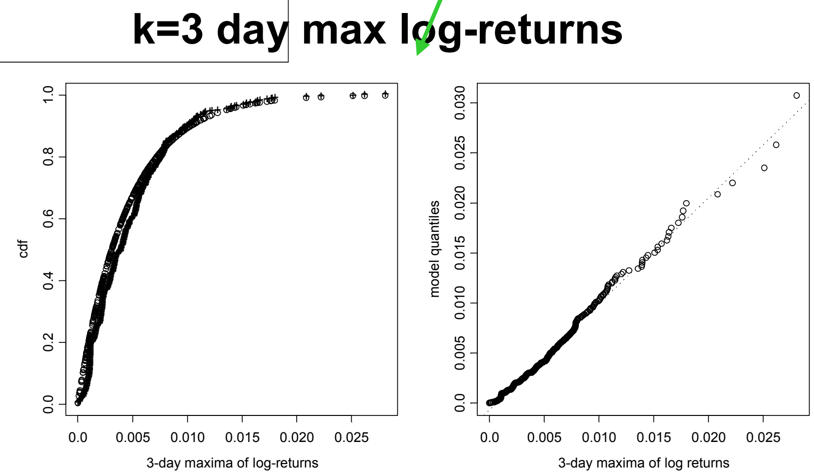
# Modeling financial data: Fit of BTLG model to (N, $Y=\max X_i$)

**Marginal distribution of the maximum fit.**

**Fit of conditional distr. X|N=k**



# Fit of TETLG model to Exchange data

**All tests on 0.05 significance level. Checking bivariate fit.**

**Is the ratio independent from the sum?**

**Checked Pearson's correlation between ratio and sum;**

**Ho: $\rho=0$ Ha: $\rho\neq0$**

**r = 0.179, p-value = 0.1749**

**Conclusion: Ratio and sum are not correlated, so BGGE model OK**

**Bivariate Kolmogorov-Smirnov goodness of fit test.**

**Ho: data comes from BGGE distribution Ha: data does not come from BGGE**

**Test stat=1.346, p-value=0.45**

**Conclusion: BGGE model OK**

**Bivariate Kolmogorov-Smirnov goodness of fit tests for conditional distributions of (X, Y) GIVEN n=N.**

**Test results below.**

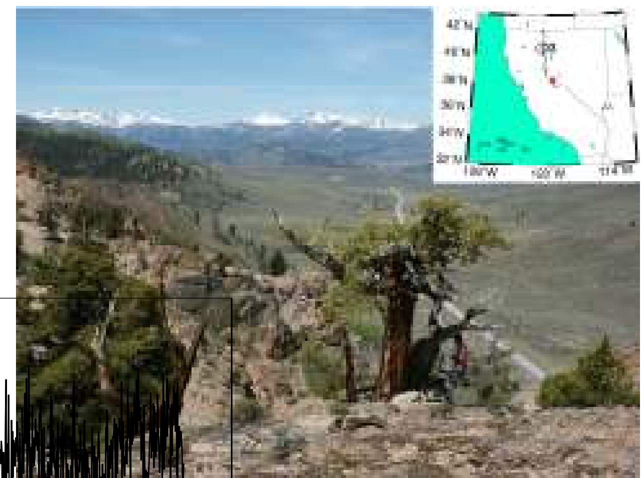
| N=2 | 3 | 4 | 5 |
|---|---|---|---|
| P-value=0.293 | 0.806 | 0.335 | 0.770 |

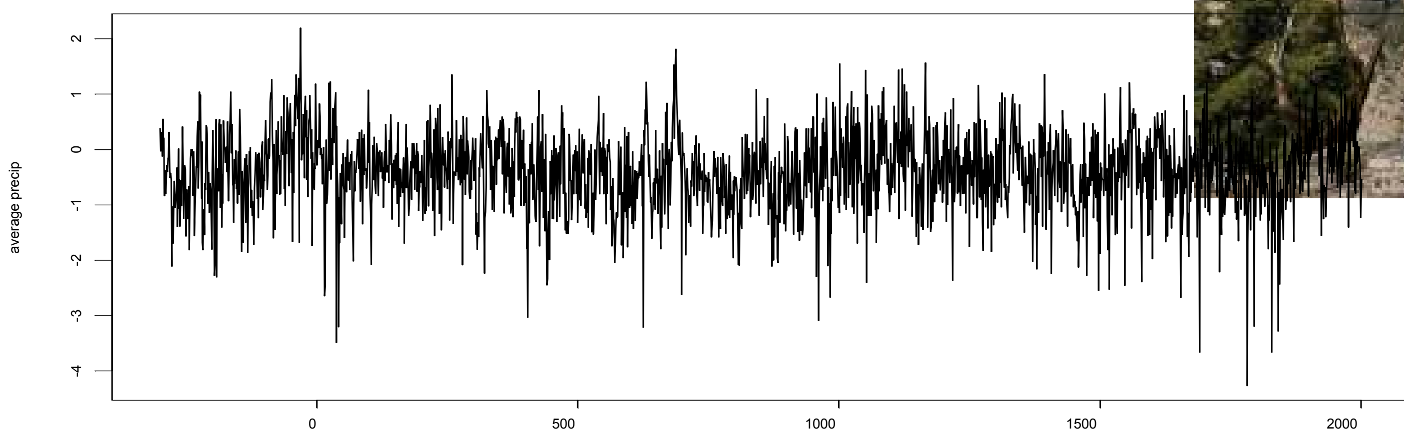
**OVERALL CONCLUSION: REASONABLE FIT of all the bivariate models (BEG, BTLG, and BGGE) for the growth episodes**

# Example 2: Climate and Hydrology: The “Dust Bowl”

The great drought of the 1930s in the US (and Canada), the 'Dust Bowl' impacted millions of people in many states. It had terrifying effects on the economy and the natural environment.

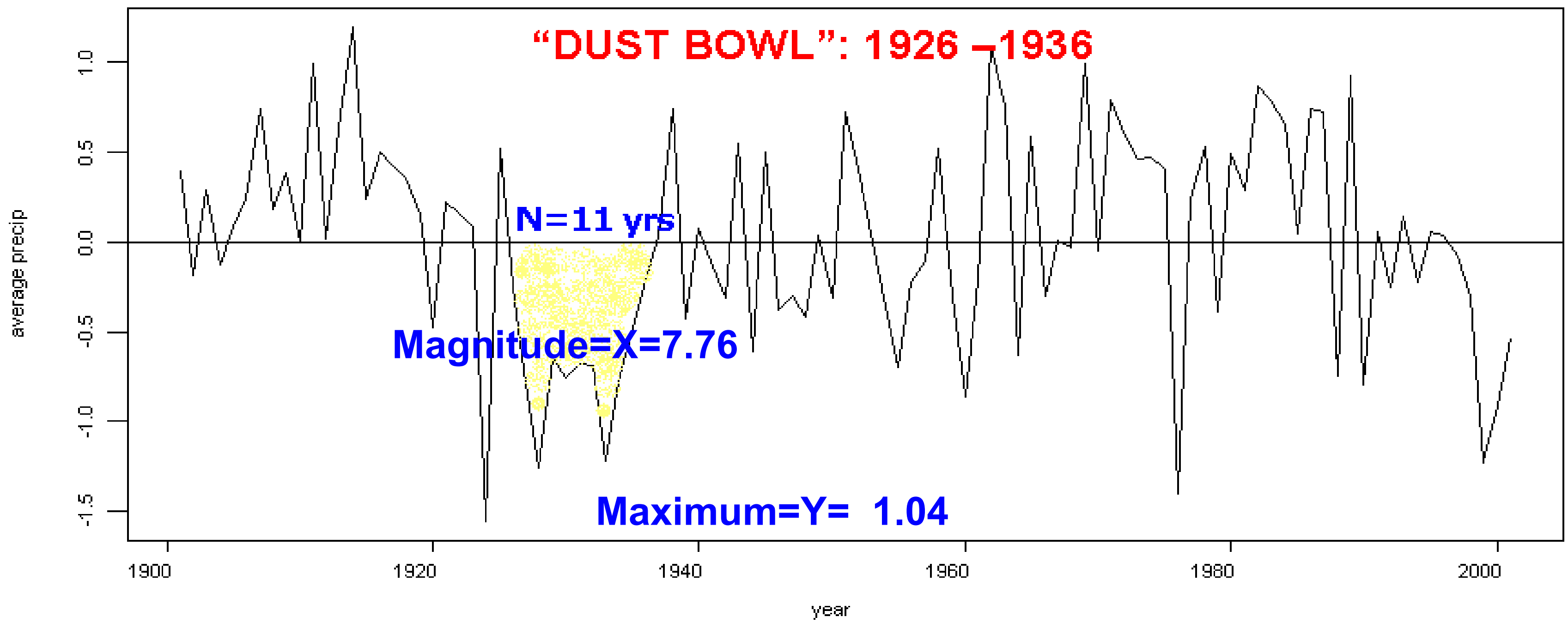
**The data:** Dendroclimatic (western juniper) reconstruction of precipitation from 300 BC to AD 2001 in the Walker River watershed (California/Nevada). California Climate Division 3.





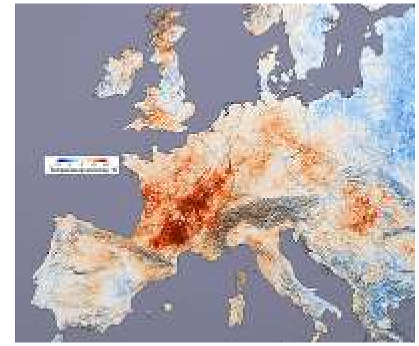

**Total annual precipitation (full reconstruction) plotted as deviation from the overall median in standard deviation units.**

# EXAMPLE 2: THE “DUST BOWL” contd.



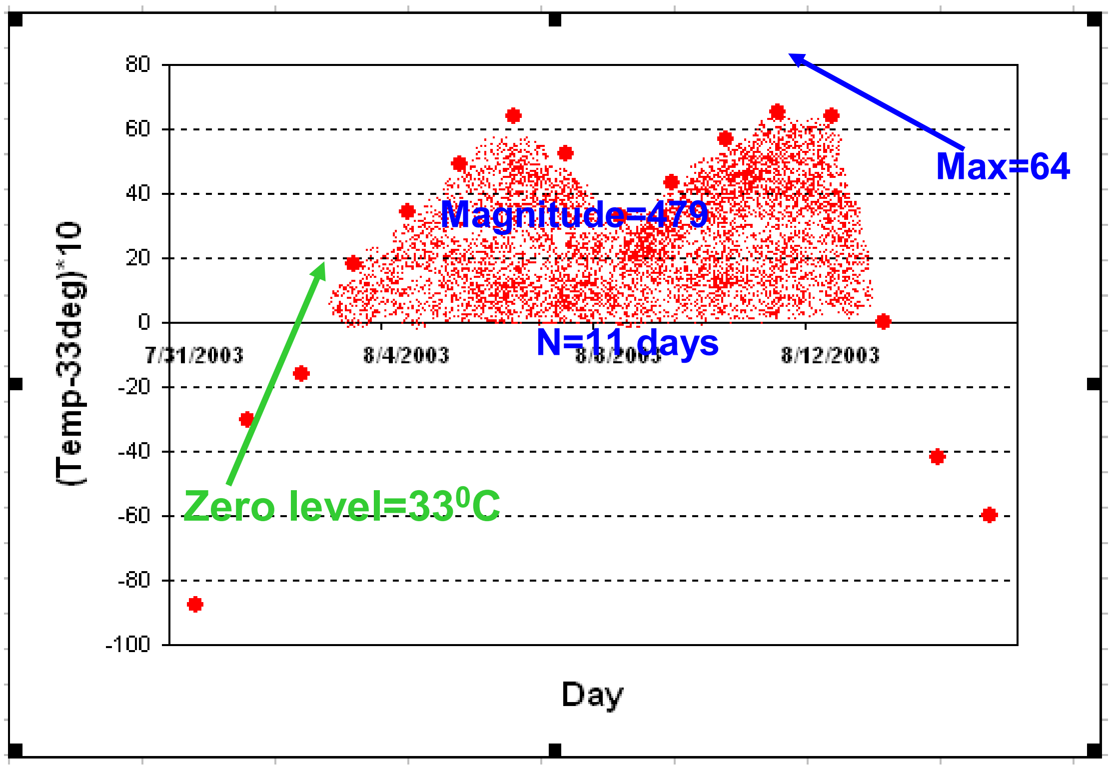

- **Probability of a drought longer or larger than the 'Dust Bowl' is 0.08;**
- **Probability of a drought longer and larger than 'Dust Bowl' is 0.06.**
- **Conditional probability of a drought with at least 'Dust Bowl''s magnitude given that duration is 11 years is 0.46**

# EXAMPLE 3: Paris heat wave of the 2003



**In August 2003, France experienced an extreme heat wave, that resulted in an estimated 14,802 deaths*.**

**Definition of a hot event: consecutive observations above the 33$^o$C.**




- **Probability of a hot event with the same duration of 11 days is 0.000075;**
- **Conditional probability of a heat wave with at least that magnitude given that duration is 11 days is about 5.5e-4.**
- **Probability of a heat wave longer than 6 days and larger than 100 (98th percentile of magnitudes) is 0.005.**

**Note: Maximum Y = 64deg C*10 (really 39.4oC)**

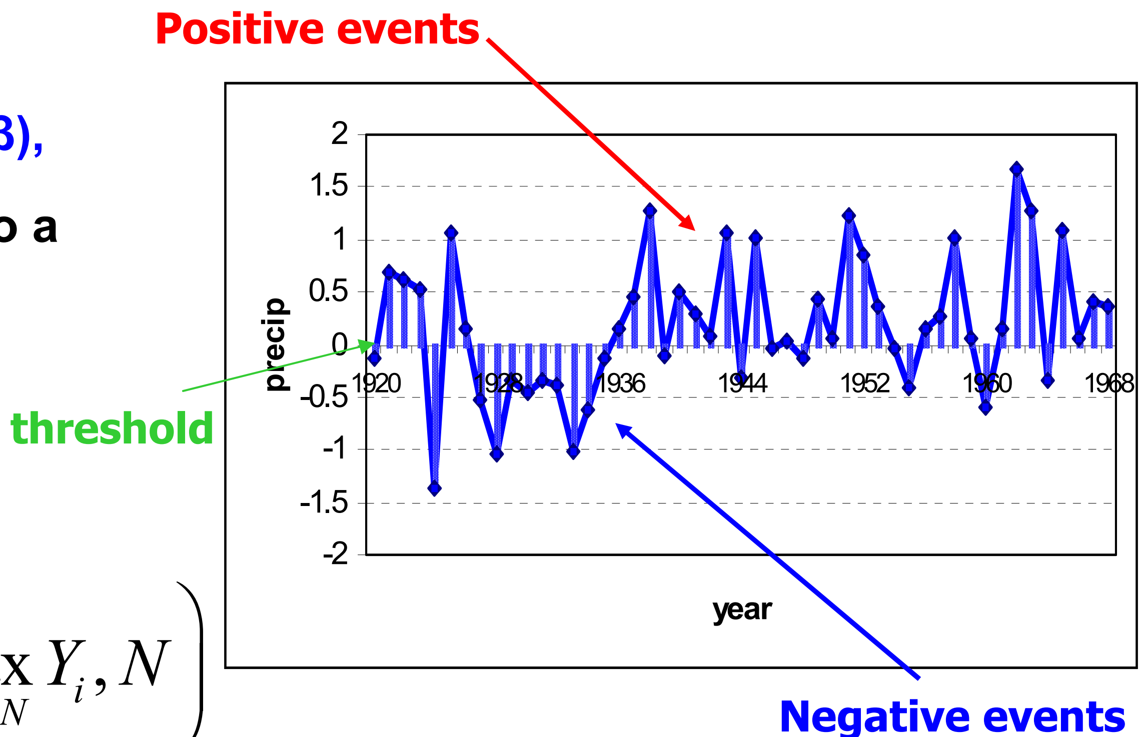


***Dhainaut et al. 2004., Data from http://eca.knmi.nl/, station ID 104969**

# SUMMARY

Given a process X1, X2, X3, …. fluctuating around a threshold.

Suppose Xi's are iid exp(β), and the process fluctuates according to a geometric model.




$$(\mathrm{X}, \mathrm{Y}, \mathrm{N}) = \underbrace{\left( \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{X}_{\mathrm{i}}, \max_{1 \le i \le N} Y_i, N \right)}_{\mathrm{TETLG}(\beta,\mathrm{p})}$$

NEW stochastic model for the joint distribution of duration, magnitude, and maximum of the events.

WHY IMPORTANT FOR NEVADA and OTHER WESTERN STATES ?

We have to deal with lots of extremes.

# NEVADA: A STATE OF EXTREMES

Thank You