A REVISIT TO
DANISH FIRE LOSS DATA

by

Dr. D V S Sastry
Director General

&

Dr. R K Sinha
Deputy Director

Insurance Regulatory & Development Authority, Hyderabad

12th GCA on February 18-19, 2010 at Mumbai

OVERVIEW

- Introduction to Danish Fire Data
- Development of Models
- Descriptive Statistics
- Proposed Model
- Statistical Issues
- Discussion and Conclusions

Introduction to Danish Fire Data

- **A Fire Insurance Claims data**
  - **Of Denmark**
  - **Covering big losses (over 1M DKK)**
  - **Period 1980 – 1990**
  - **2156 records**

- **Why important?**
  - **Globally recognized dataset**
  - **Path breaker for Researchers**
  - **Scarcity of fire data sets**

Development of Models

- **Extreme Value Theory (EVT)**
  - **McNeil (1996)**
  - **A tool for ERM (ST9-new subject)**

[Book: “Quantitative Risk Management” by McNeil, A., R. Frey, P. Embrechts (2005)]

- **Mixture Models**
  - **Frigessi el al (2002)**
  - **Cooray & Ananda (2005)**
  - **Ciumara (2006)**
  - **Scollnik (2007)**
  - **Bengio (2009)**

EVT: Two Approaches

Block Maxima Approach:

- Led by Fisher-Tippett (1928)
- Describes the limiting behaviour
  of
- Appropriately normalized sample maxima
- Generalized Extreme Value (GEV) Distribution

Peaks-Over-Threshold (POT) Approach:

- Led by Pickands (1975)
- Describes that the limiting distribution
  of
- Excesses over a threshold
- Generalized Pareto Distribution (GPD)

McNeil (1996)

Fitted - Lognormal,
- Pareto and,
- GPD

Observations:

- lognormal – reasonable, tail too thin
- Pareto – Overestimates
- GPD – in between

Problem:

- Difficult to select threshold
- Inconsistency across methods

Mixture Models:

Frigessi et al Model (2002):

- **A weighted mixture model**
- **One density – GPD**
- **Other density – a light tailed**

$$f(x) = \pi\, f_1(x) + (1-\pi)\, f_2(x)$$

$f_1(x)$: Heavy tailed density (GPD)
$f_2(x)$: Light tailed density (say, Weibull)

π is the weight function

Cooray & Ananda Model (2005)

- **Mixes lognormal and Pareto**
- **Threshold limitation**

Scollnik Model (2007)

- **Generalizes Cooray & Ananda model**
- **Threshold depends on the underlying data set.**

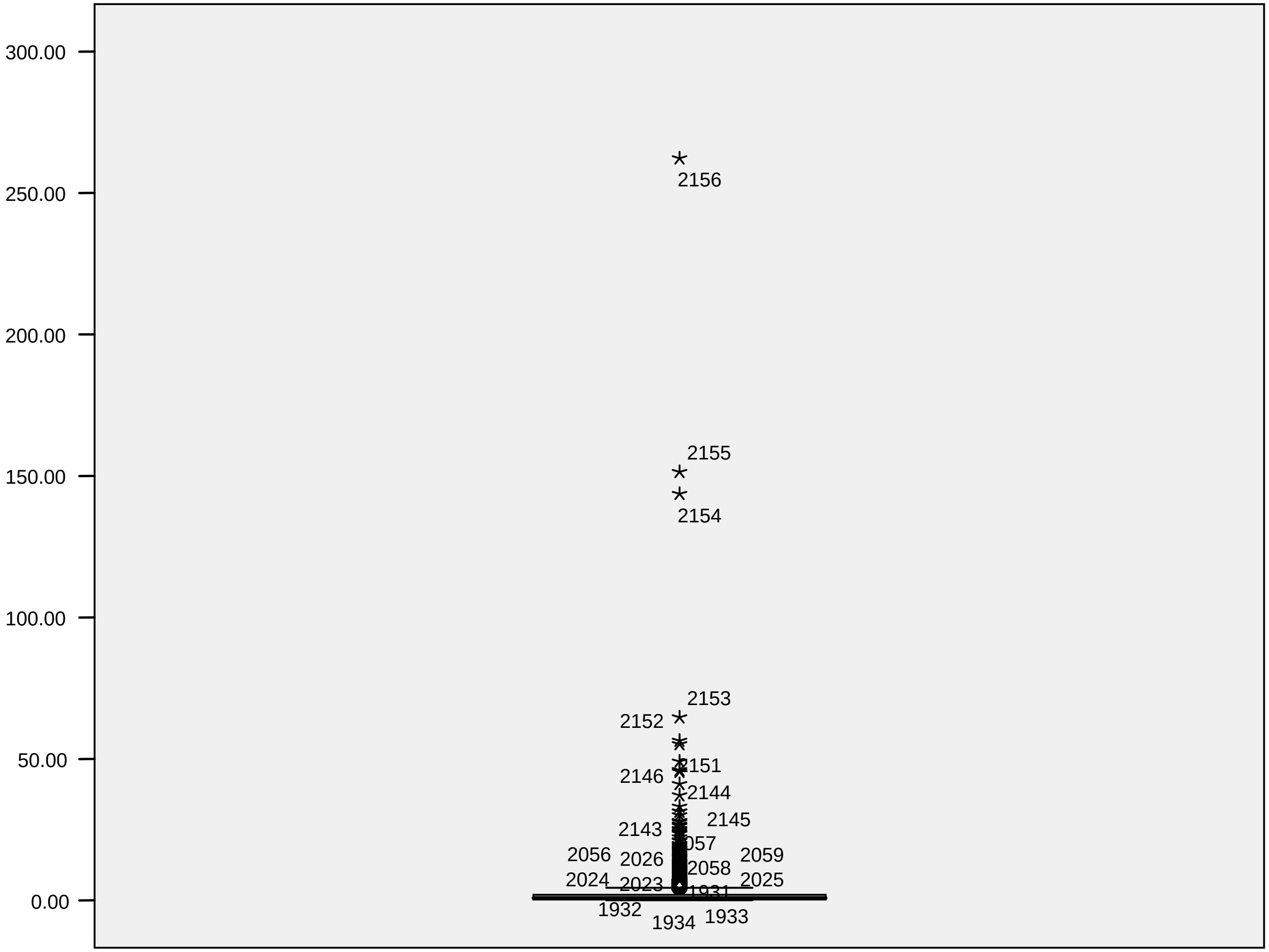
Descriptive Statistics

Data of losses over one million Danish Kroner (MDKK)

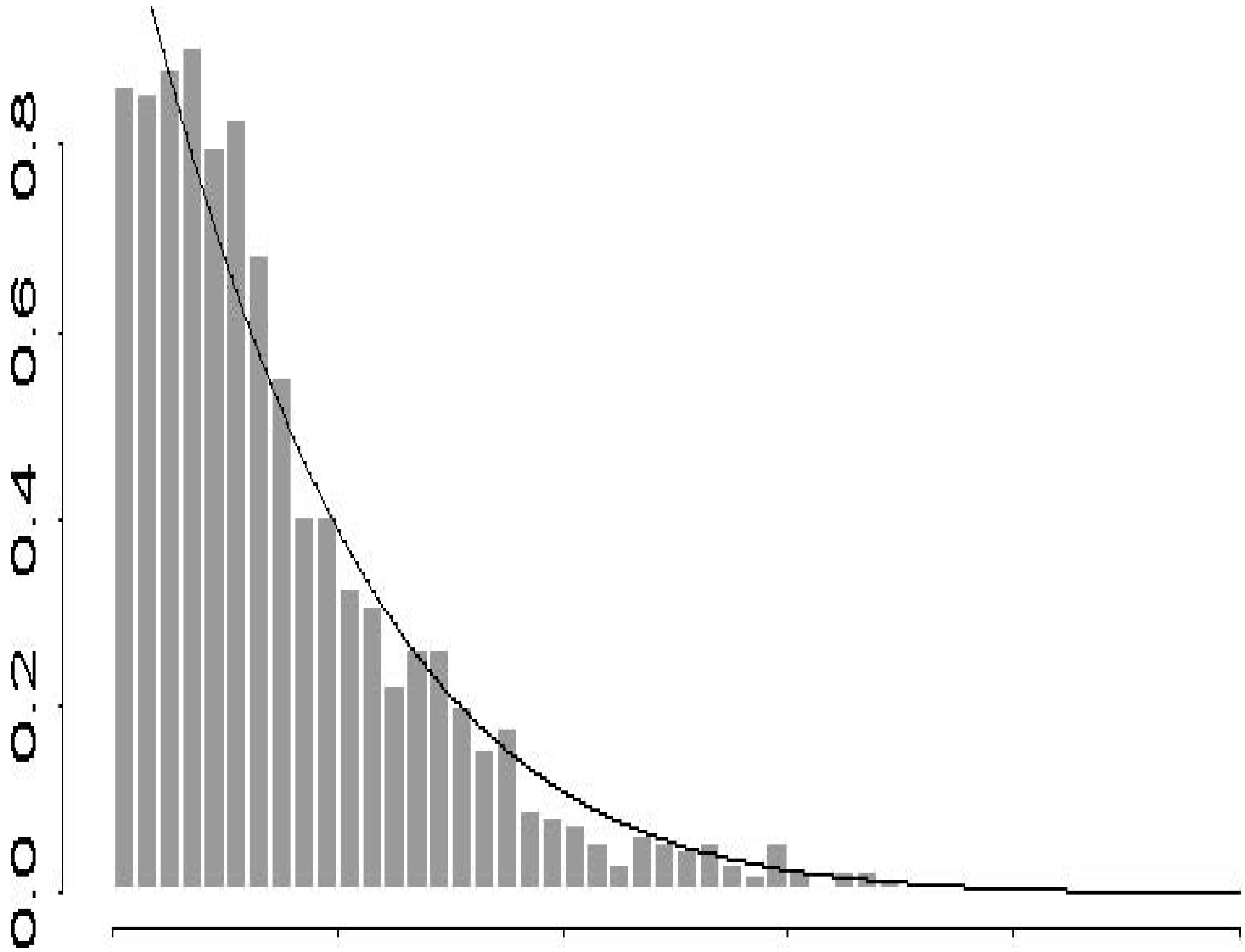
< 1 MDKK: 324 obs (data not available)
= 1 MDKK: 12 obs
> 1 MDKK: 2156 obs

Number of observations	**: 2156**
Mean	**: 2.3973 MDKK**
Standard Deviation	**: 8.5274 MDKK**
First Quartile	**: 0.33081 MDKK**
Median	**: 0.78178 MDKK**
Third Quartile	**: 1.97249 MDKK**
Skewness	**: 18.722**
Kurtosis	**: 484.58**

BOX-PLOT


300.00
250.00
200.00
150.00
100.00
50.00
0.00
2156
2155
2154
2153
2152
2151
2146
2144
2145
2143
057
2056
2026
2059
2058
2024
2023
2025
1931
1932
1934
1933

HISTOGRAM
(on logarithmic scale)



Proposed Model

4P-Burr Distribution:

$f(x) = \alpha\, k\, \{(x-\gamma)/\beta\}^{\alpha-1} / \beta[1+\{(x-\gamma)/\beta\}^{\alpha}]^{k+1}$

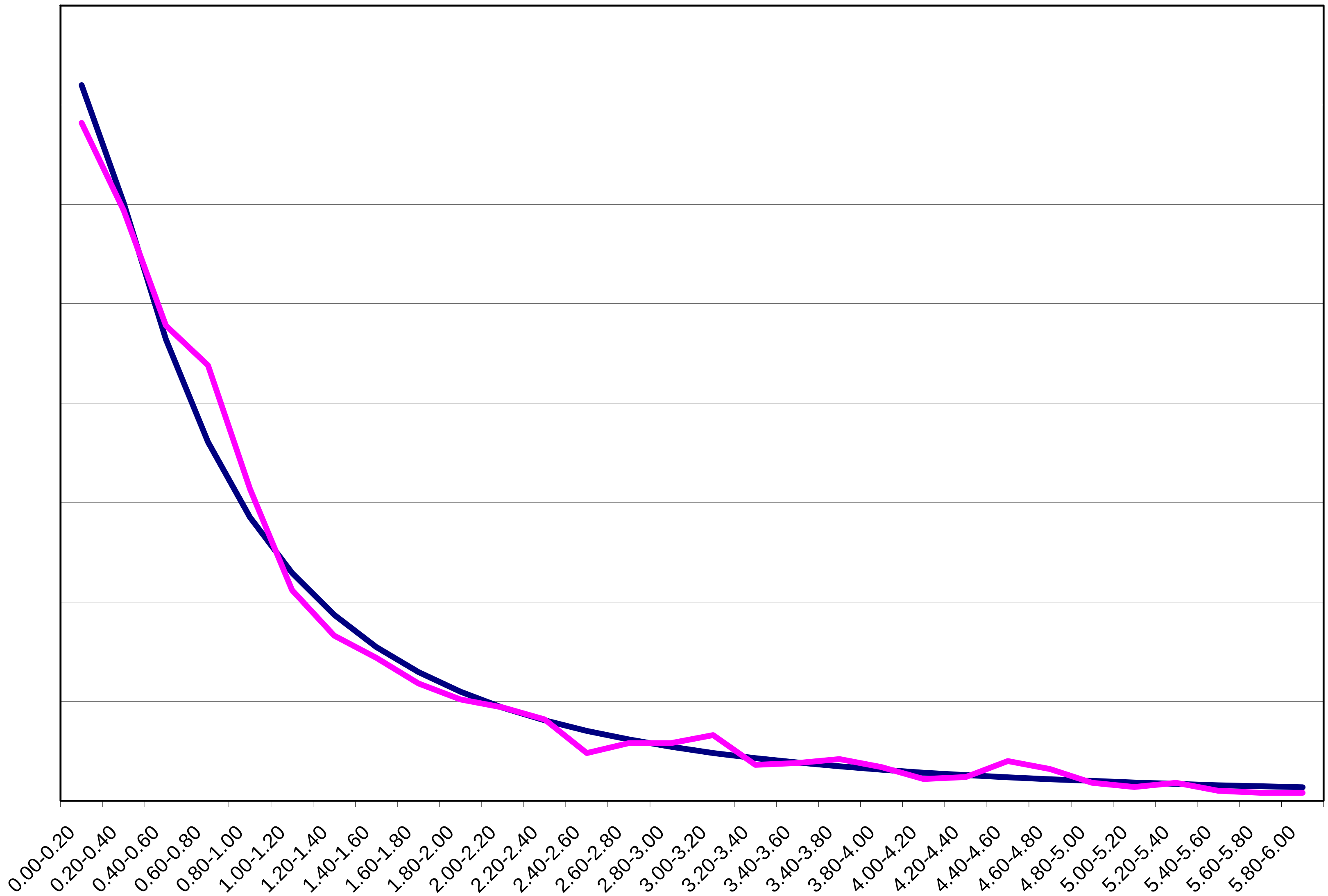
$F(x) = 1 - 1/[1+\{(x-\gamma)/\beta\}^{\alpha}]^{k}$

k (>0) and α (>0): shape parameters
β (>0): scale parameter
γ ($0 \leq x < \infty$): location parameter

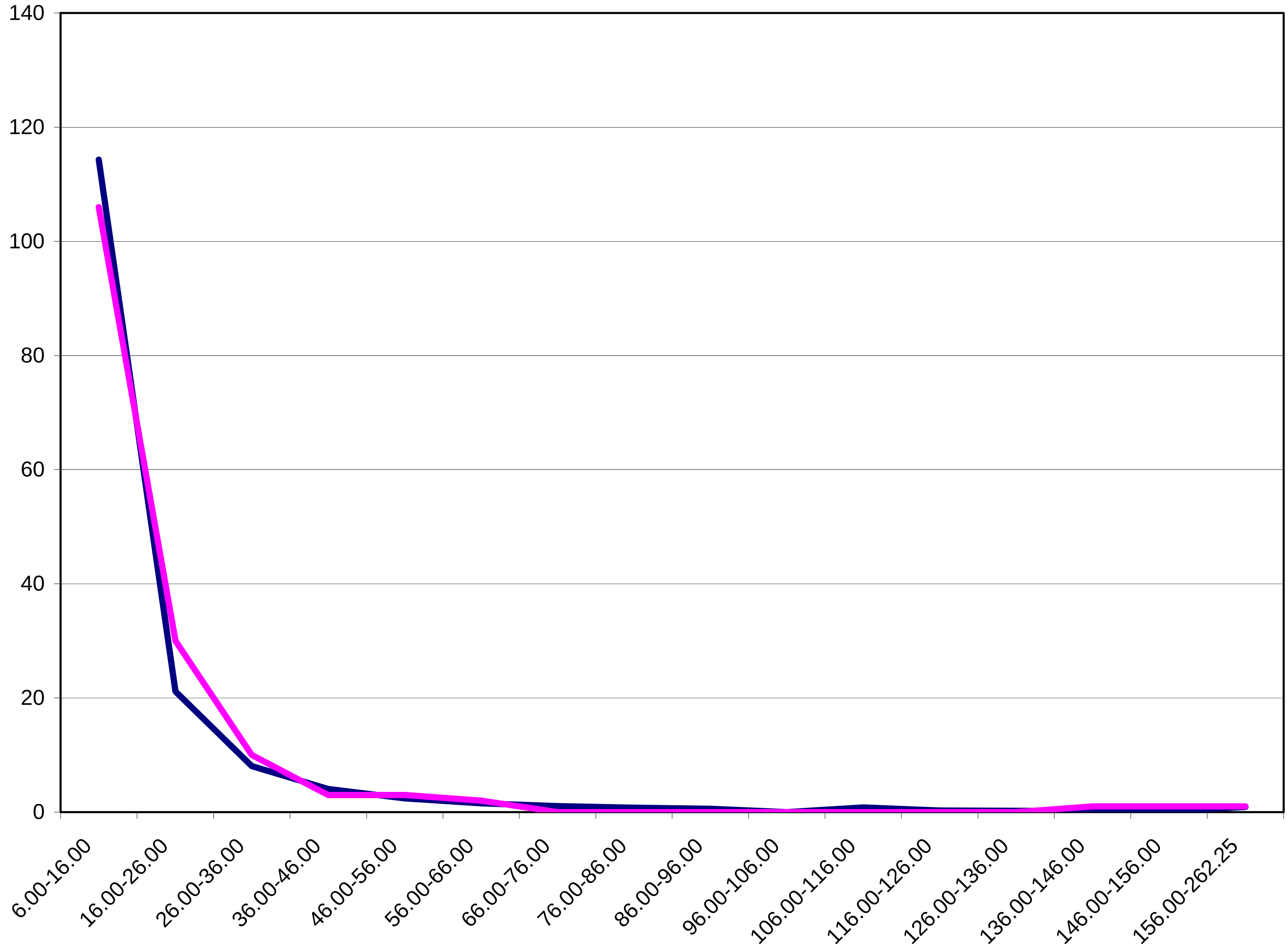
Estimated parameters:
$k = 1.3264$, $\alpha = 1.0944$, $\beta = 1.1329$ and $\gamma = 0.0027$

Test statistics:
KS : 0.01865
AD : 0.90398
Chi^2 : 17.8980

Fitting curve (0 to 6M DKK)



Fitting curve (over 6M DKK)



Comparison with other Models

Quantile (%)	Observed	Expected				
		Cooray & Ananda (2005) Model	Scollnik (2007) 1st Model	Scollnik (2007) 2nd Model	Present (Burr) Model
50.00	1.634	1.586	1.570	1.611	1.617
75.00	2.645	2.569	2.651	2.716	2.693
90.00	5.080	4.863	5.297	5.169	5.104
95.00	8.406	7.879	8.943	8.245	8.213
97.50	14.395	12.765	15.099	13.024	13.195
99.00	24.614	24.157	30.175	23.628	24.698
99.50	32.431	39.139	50.945	36.929	39.709
99.90	105.893	120.008	171.889	103.446	119.871
99.95	150.509	194.437	290.207	160.957	193.060
99.99	235.641	596.182	979.157	448.579	582.988

Statistical Issues

- **Assumption of IID?**
- **Correlation? Dependency?**
- **Season Effect?**

The biggest 6 claims occurred on:

- ❖ **56.225426 - May 29, 1981**
- ❖ **57.410636 - August 23, 1985**
- ❖ **65.707491 - October 24, 1982**
- ❖ **144.657591 - October 08, 1990**
- ❖ **152.413209 - August 04, 1989**
- ❖ **263.250366 - July 15, 1980**

Example: The Climate of Denmark

Winter : November to April
Summer : May to October

Question: Is it a Coincidence?

Answer: Statistical tests do not say so?

A quote from McNeil (1996) paper:

"....In practice we may be confronted with clustering, trends, seasonal effects and other kinds of dependencies. When we consider fire losses in Denmark it may seem a plausible first assumption that individual losses are independent of one another: however, it is also possible to imagine that circumstances conducive or inhibitive to fire outbreaks generate dependencies in observed losses. Destructive fires may be greatly more common in the summer months; buildings of a particular vintage and building standard may succumb easily to fires and cause high losses...."

Descriptive Statistics according to Season

Statistics	Summer Data	Winter Data
Observations	1104	1052
Mean	3.7745	3.0013
SD	11.157	4.2608
25% Quartile	1.3372	1.3226
Median	1.8113	1.7638
75% Quartile	3.1107	2.8917
90%	5.818	5.2094
95%	11.024	8.6071
Minimum	1.002893	1.002893
Maximum	263.250366	50.065531
Range	262.247473	49.062638
Skewness	15.949	5.8464
Kurtosis	320.83	47.531

Parameter Estimates and Test Statistics

Statistics	Summer Data	Winter Data
First shift parameter (k)	1.1824	1.5401
Second shift parameter (α)	1.1181	1.0664
Scale parameter (β)	1.0094	1.3296
Location parameter (γ)	1.0027	1.0028
Kolmogorov - Smirnov test	0.01721	0.02829
Anderson - Darling test	0.36272	0.68856
Chi square test	10.184	14.780

Discussion and Conclusions

- **The 4-Parameter Burr distribution reasonably explains the underlying Danish data set.**
- **Need to explore all statistical distributions to assess whether a distribution could explain the data set appropriately.**
- **For the insurance industry, such a protocol is important.**
- **We find that the Danish fire losses are seasonal in nature. Such knowledge will be helpful for the actuary.**

Thank You