KU LEUVEN

LStat
Leuven Statistics Research Centre

LRISK

Modelling of long-tail reinsurance data

Jan Beirlant Tom Reynkens

Department of Mathematics, LStat and LRisk
KU Leuven

EAJ Conference Lyon, September 8, 2016

Outline

1. Motor Third Party Liability (MTPL) application
2. Objectives
3. Pareto extreme value analysis (EVA)
4. Solution 1: via interval censoring
5. Solution 2: regressing loss data on nr of development years
6. Conclusion

MTPL application

- **Motor Third Party Liability** (MTPL) insurance in European country between 1995 and 2010.
- 837 claims, 60% not closed in 2011.
- Accident date known in present case study
- All payments, respectively incurred loss data, for a given claim relating to the same development year aggregated in a single claim point, respectively incurred data point
- Average reporting threshold set at $154\,508$
- Indexed cumulative payments and indexed incurreds are available per year

MTPL example

Build upper bounds on total claim amount using incurreds

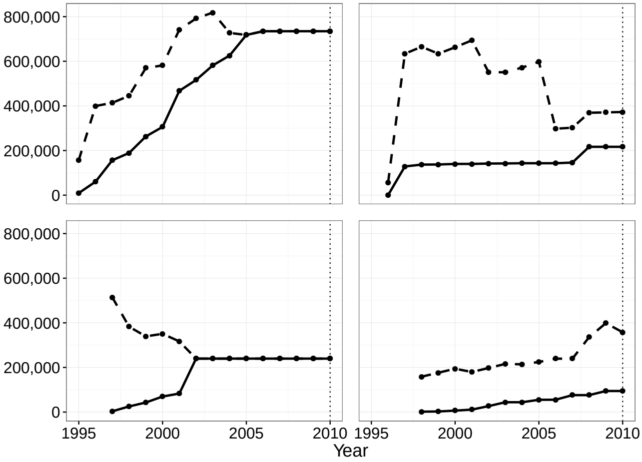
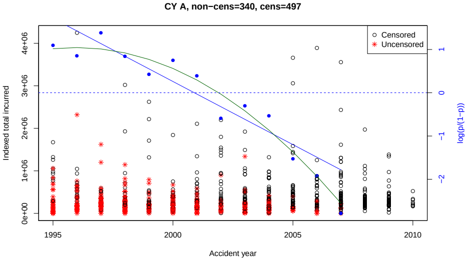


Figure: 4 MTPL claims: cumulative indexed payments (full line) and indexed incurred values (dashed line)

MTPL data; 340 fully developed claims



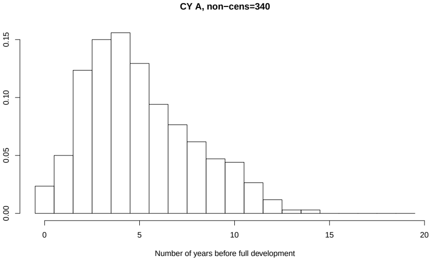


Figure: Incurreds and log-odds p ($1 - p =$ proportion reported but not settled (RBNS) against accident year (left); histogram years before full development for closed claims (right)

Censoring in (re)insurance

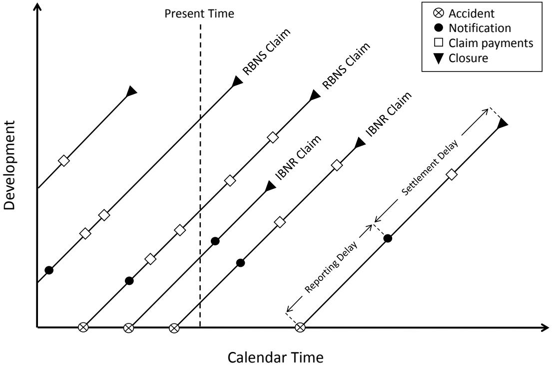


Figure: Claim development scheme

Censoring in MTPL case

Total loss amount X and number of development years nDY at 31/12/2010 are right censored for claims non-developed at 31/12/2010

X and nDY are censored or not censored at same time

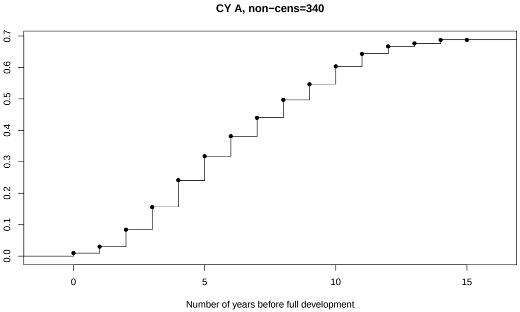


Figure: Kaplan-Meier estimator for the distribution function of the number of development years

MTPL data: use of 'Ultimates'

Ultimates computed by company using 'own' model
Statistical analysis based on Ultimates ?

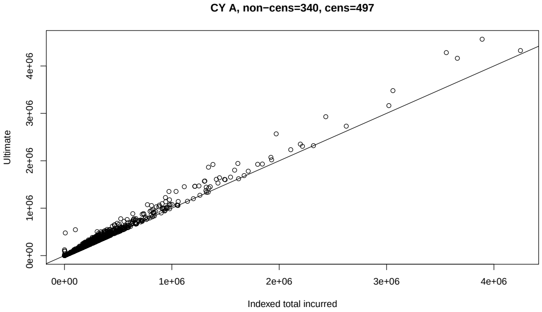


Figure: Ultimates versus incurred losses with unit line

Objectives of this study

- Extreme value analysis of X
- Global fits for risk analysis that does not only focus on extreme events
- Model loss data to
  - Set insurance premiums;
  - Calculate risk measures (VaR, TVaR, ...);
  - Determine capital requirements for solvency regulations;
  - Set optimal retention level for reinsurance.

Classical EVA for non censored data

Notation: sample $X_1, \ldots, X_n$, ordered $X_{1,n} \leq \ldots \leq X_{n-k,n} \leq \ldots \leq X_{n,n}$

Pareto-type distribuations

$$1 - F(x) = x^{-1/\xi} L(x), \; L \text{ slowly varying } L(xt)/L(t) \to 1 \text{ as } t \to \infty$$

Hill estimator of ξ: $H_{k,n} = k^{-1} \sum_{j=1}^{k} \log \frac{X_{n-j+1,n}}{X_{n-k,n}}$

- using $X_{n-k,n}$ for t, and empirical distribution function $\hat{F}$, estimating

$$E(\log X - \log t | X > t) = \frac{\int_t^\infty (1 - F(u))\, u^{-1} du}{1 - F(t)} \to \xi \text{ as } t \to \infty$$

- estimating slope of log-log plot which is ultimately linear near k largest observations
- maximum likelihood estimator maximizing $\Pi_{X_j > t} \left\{ \frac{1}{\xi} \left(\frac{X_j}{t} \right)^{-1/\xi - 1} \right\}$

Random right censoring, Kaplan-Meier

- Observations

$$(Z_i, \Delta_i), i = 1, \ldots, n$$

with

$$\begin{aligned} Z_i &= \min(X_i, C_i), C_i \text{ censoring variables} \\ \Delta_i &= \begin{cases} 1 & \text{if non-censored} \\ 0 & \text{if censored (RBNS)} \end{cases} \end{aligned}$$

- Random right censoring: X_i and C_i independent
- Kaplan-Meier (1958) estimator $\hat{F}^{KM}$

$$1-\hat{F}^{KM}(x) = \Pi_{Z_{i,n} \leq x} \left(1 - \frac{1}{n-i+1}\right)^{\Delta_i} = \Pi_{Z_{i,n} \leq x} \left(1 - \frac{\Delta_{i,n}}{n-i+1}\right)$$

EVA for random right censored data

Both X and C Pareto-type distributed: $\xi_X, \xi_C > 0$

- Likelihood approach: Beirlant et al. (2007), Einmahl et al. (2008)

$$\Pi_{X_j>t} \left\{ \frac{1}{\xi} \left(\frac{Z_j}{t} \right)^{-1/\xi-1} \right\}^{\Delta_j} \left\{ \left(\frac{Z_j}{t} \right)^{-1/\xi} \right\}^{1-\Delta_j}$$

  Hill estimator adapted for right censoring

$$H_{k,n}^{(c1)} = \frac{H_{k,n}}{\hat{p}_k} \text{ with } \hat{p}_k = k^{-1} \sum_{Z_j > Z_{n-k,n}} \Delta_j$$

  $\hat{p}_k$ proportion of non-censored data within top k observations

- Bardoutsos et al. (2016): bias reduced version
- Estimation of $\int_t^\infty (1-F(u))\, u^{-1} du/(1-F(t)$
  Worms and Worms (2014)

$$H_{k,n}^{(c2)} = \frac{\sum_{j=1}^{k} \left(1 - \hat{F}^{KM}(Z_{n-j+1,n}^{-})\right) (\log Z_{n-j+1,n} - \log Z_{n-j,n})}{1 - \hat{F}^{KM}(Z_{n-k,n})}$$

EVA adapted for censoring

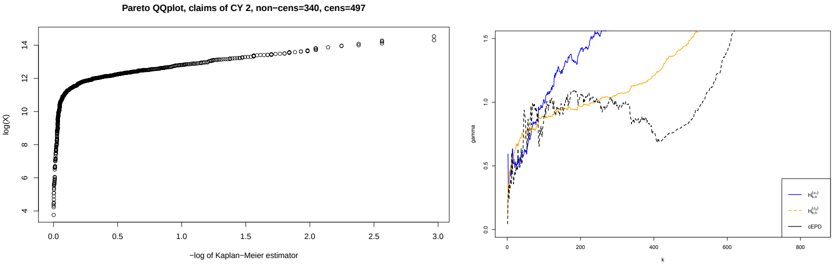


Figure: Pareto QQ-plot adapted for censoring (left) and Hill estimators $H_{k,n}^{(c1)}, H_{k,n}^{(c2)}$ and bias reduced version adapted for censoring (right)

ξ too big! Condition of random right censoring fulfilled? Incurreds ?

Tom Reynkens

- Densities f_1^* and f_2^* and corresponding CDFs F_1^* and F_2^*.
- Lower truncation at t^l, splicing point t
- Transform to valid densities on the intervals $[t^l, t]$ and $[t, \infty]$:

$$f_1(x) = \begin{cases} \frac{f_1^*(x)}{F_1^*(t) - F_1^*(t^l)} & \text{if } t^l \leq x \leq t \\ 0 & \text{otherwise,} \end{cases}$$

$$f_2(x) = \begin{cases} \frac{f_2^*(x)}{1 - F_2^*(t)} & \text{if } t \leq x \\ 0 & \text{otherwise.} \end{cases}$$

Splicing density

$$f(x) = \begin{cases} 0 & \text{if } x \leq t^l \\ \pi f_1(x) & \text{if } t^l < x \leq t \\ (1-\pi) f_2(x) & \text{if } t < x \end{cases}$$

Solution 1: EVA for interval censored data

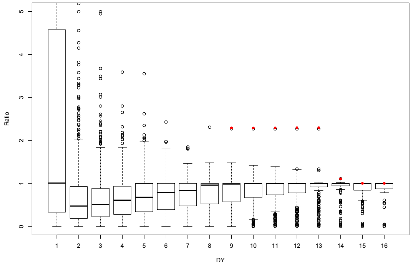
- Turnbull (1976) estimator $\hat{F}^{TB}$ of F_X: adaptation of Kaplan-Meier estimator to interval censoring
- Estimators for $E(X - t | X > t)$ and $E(\log X - \log t | X > t)$:

$$\text{mean excess function } e_n^{TB}(t) := \frac{\int_t^\infty (1 - \hat{F}_n^{TB}(u))du}{(1 - \hat{F}_n^{TB}(t))}$$

$$\text{Hill functional } H_n^{TB}(t) := \frac{\int_t^\infty (1 - \hat{F}_n^{TB}(u))u^{-1}du}{(1 - \hat{F}_n^{TB}(t))}$$

Solution 1: building upper bounds for X_i based on incurreds

- Cumulative payment claim i in 2010: X_i
- Incurred value of claim i at development year $DY = j$: $I_{j,i}$
- $R_{j,i} = \frac{X_i}{I_{j,i}}$, $i = 1 \ldots, n_j$, $j = 1, \ldots, 15$



- Estimate endpoints M_j of censored data sets $R_{j,i}$, $i = 1, \ldots, n_j$, per DY (Beirlant et al. (2007), Einmahl et al. (2008))
- Upper bounds for censored X_i given $nDY_{2010,i} = j$: $M_j I_{j,i}$

Solution 1: EVA for MTPL with interval censoring

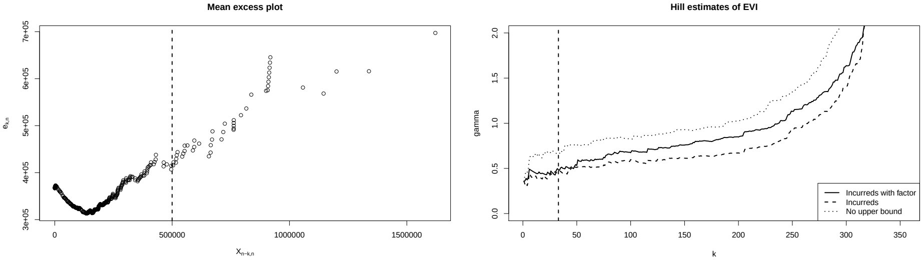


Figure: mean excess plot based on interval censored data (left); Hill plots adapted for interval censoring with upper bounds by incurreds , incurred $\times M_j$, no upper bounds (right)

Mixed Erlang distribution

- Mixture of Erlangs (ME) with common scale parameter $\theta > 0$.

- Density of the Mixed Erlang distribution

$$f_1^*(x; \boldsymbol{\alpha}, \boldsymbol{r}, \theta) = \sum_{j=1}^{M} \alpha_j \frac{x^{r_j - 1} e^{-x/\theta}}{\theta^{r_j} (r_j - 1)!} = \sum_{j=1}^{M} \alpha_j f_E(x; r_j, \theta) \quad \text{for } x > 0.$$

  - Shape parameters $\boldsymbol{r} = (r_1, \ldots, r_M)$: positive integers with $r_1 < \ldots < r_M$.
  - Mixing weights $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$.

Solution 1: MTPL splicing model with interval censoring (incurreds directly)

Splicing	ME	Pareto
$\hat{\pi} = 0.873$	$\hat{\alpha} = (0.171, 0.829)$	$\hat{\xi} = 0.438$
$t^l = 0$	$\hat{r} = (1, 4)$	
$t = 500000$	$\hat{\theta} = 55227$	

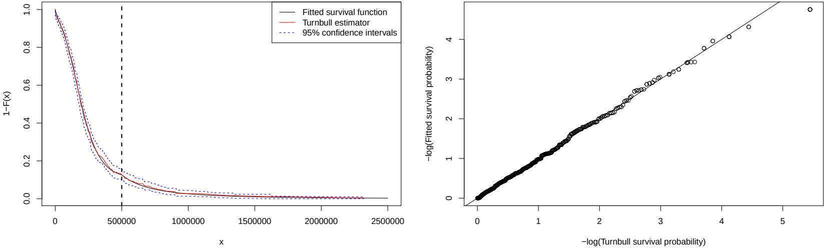


Figure: Survival plot and PP-plot on $-\log$-scale of fitted model using incurreds as upper bound (1995-2005)

Solution 1: MTPL splicing with interval cens.(incurreds $\times M_j$)

Splicing	ME	Pareto
$\hat{\pi} = 0.777$	$\hat{\alpha} = (0.155, 0.845)$	$\hat{\xi} = 0.506$
$t^l = 0$	$\hat{r} = (1, 4)$	
$t = 500000$	$\hat{\theta} = 63410$	

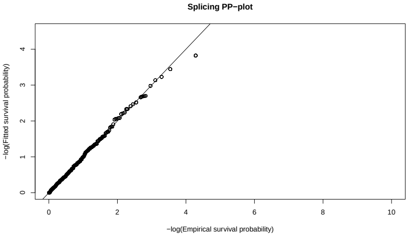


Figure: PP-plot on $-\log$-scale of fitted model using incurreds $\times M_j$

Solution 1: comparing three splicing models

Comparing random right censored model and interval censoring using incurred, and incurred $\times M_j$

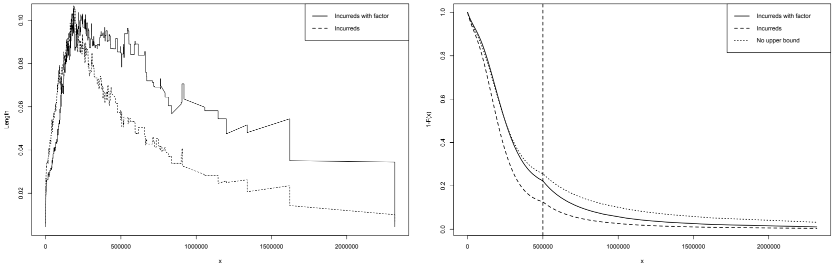


Figure: Size of confidence intervals (left) and estimated survival functions (right) using incurreds, incurreds $\times M_j$, and no upper bounds (KM)

Solution 2: Nonparametric regression with censoring

Both response X and covariate DY censored:

$$\begin{aligned}
Z_i &= \min(X_i, C_i^{(2)}) \\
nDY_{2010,i} &= \min(nDY_i, C_i^{(1)}) \\
\Delta_i^{(1)} &= \Delta_i^{(2)} = \Delta_i,\ i = 1, \ldots, n
\end{aligned}$$

Akritas and Van Keilegom (2003):
if X and $C^{(1)}$ are conditionally independent given nDY

$$1 - \hat{F}_{X|nDY}(x|d) = \prod_{Z_i \leq x} \left(1 - \frac{W_{n,i}(d; h_n)}{\sum_{j=1}^n W_{n,j}(d; h_n) I\{Z_j \geq Z_i\}} \right)^{\Delta_i}$$

with

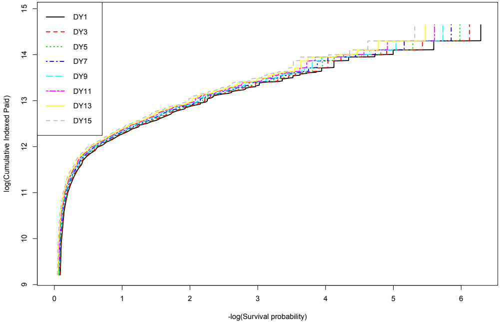
$$W_{n,i}(d; h_n) = \begin{cases} K\left(\frac{d - nDY_{2010,i}}{h_n}\right) / \sum_{\Delta_j = 1} K\left(\frac{d - nDY_{2010,j}}{h_n}\right) & \text{if } \Delta_i = 1 \\ 0 & \text{if } \Delta_i = 0. \end{cases}$$

Solution 2: Pareto QQ-plots with censoring given $nDY = d$

Setting $d = 1, 3, 5, \ldots, 15$, K bi-weight kernel, and $h = 15$

Pareto QQ-plots adapted for censoring per chosen d value

$$\left(-\log\left(1-\hat{F}_{X|nDY}(Z_{n-j+1,n}|d)\right), \log Z_{n-j+1,n}\right), \; j = 1, \ldots, n$$

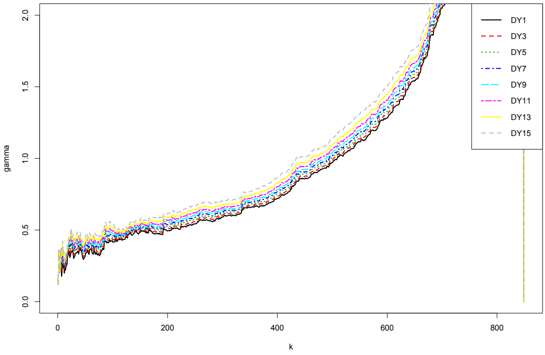


Solution 2: Hill type estimator with censoring given $nDY = d$

$$
\begin{aligned}
&H_{k,n}^{(c)}(nDY = d) \\
&= \frac{\int_{Z_{n-k,n}}^{\infty} \left(1 - \hat{F}_{X|nDY}(y|d)\right) \frac{dy}{y}}{1 - \hat{F}_{X|nDY}(Z_{n-k,n}|d)} \\
&= \frac{\sum_{j=1}^{k} \left(\prod_{i=1}^{n-j} \left[\left(1 - \frac{W_{i,n}(d;h_n)}{1-\sum_{l=1}^{i-1} W_{l,n}(d;h_n)}\right)^{\Delta_i} \right] \log \frac{Z_{n-j+1,n}}{Z_{n-j,n}} \right)}{\prod_{i=1}^{n-k} \left[\left(1 - \frac{W_{i,n}(d;h_n)}{1-\sum_{l=1}^{i-1} W_{l,n}(d;h_n)}\right)^{\Delta_i} \right]}
\end{aligned}
$$

(1)

Solution 2: Hill type estimator with censoring given $nDY = d$



To be continued

- Bias reduction for censored regression
- Splicing for full modelling under censored regression
- See also approach by Pigeon and Denuit (2014)
- ...

References I

Akritas, M.G. and Van Keilegom, I. (2003). Estimation of bivariate and marginal distributions with censored data. J. R. Stat. Soc. Ser. B. Stat. Methodol., 65(2), 457–471.

Albrecher, H., Beirlant, J. and Teugels, J. (2016). Reinsurance: Actuarial and Statistical Aspects. John Wiley Sons, Ltd, Chichester, UK. To appear.

Beirlant, J., Bardoutsos, A., de Wet, T. and Gijbels, I. (2016). Bias reduced tail estimation for censored pareto type distributions. Statistics and Probability Letters, 109, 78–88.

Beirlant, J., Guillou, A., Dierckx, G. and Fils-Villetard, A. (2007). Estimation of the Extreme Value Index and Extreme Quantiles Under Random Censoring. Extremes, 10(3), 151–174.
1–38.

Einmahl, J. H., Fils-Villetard, A. and Guillou, A. (2008). Statistics of Extremes Under Random Censoring. Bernoulli, 14(1), 207–227.

References II

Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. Ann. Statist., 3(5), 1163–1174.

Kaplan, E. L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. J. Amer. Statist. Assoc., 53(282), 457–481.

Pigeon, M. and Denuit, M. (2014). Multivariate Skew-Normal Individual Excess-of-Loss Reserving. No. ISBA Discussion Paper (2014/29). UCL.

Reynkens, T., Verbelen, R., Beirlant, J. and Antonio, K. (2016). Modeling censored losses using splicing: a global fit strategy with mixed Erlang and extreme value distributions, arXiv:1608.01566

Turnbull, B. W. (1976). The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. J. R. Stat. Soc. Ser. B. Stat. Methodol., 38(3), 290–295.

References III

Verbelen, R., Gong, L., Antonio, K., Badescu, A. and Lin, S. (2015). Fitting Mixtures of Erlangs to Censored and Truncated Data Using the EM Algorithm. Astin Bull., 45(3), 729–758.

Worms, J. and Worms, R. (2014). New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. Extremes 17 337-358.