An extreme value approach for modeling Operational Risk losses depending on covariates

Marius Hofert

(joint work with Valérie Chavez-Demoulin, Paul Embrechts)

2014-03-20

Technische Universität München

The people involved

P. Embrechts

J. Naish

V. Chavez-Demoulin

- Database access granted by John Naish (Willis; `naishj@willis.com`)
- Implementation in the R package `QRM` (R-Forge version 0.4-10)
  $\Rightarrow$ `gamGPDfit()`, `gamGPDboot()`
- Example based on simulated losses: `demo(game)`

Operational Risk (introduced with Basel II ($\leq$ BIS (2004)))

Definition (Operational risk)

Operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk.

Examples (legal risk and strategic risk are difficult to measure)

people:	fraud (internal, external), "fat finger trades"
systems:	ATM, computer (hardware, software)
external events:	Kobe earthquake (1995-01-17), bankruptcy of Barings bank (1995-02-26), 9/11, hurricane Katrina (mortgage default due to lost houses; credit or an OpRisk event?)
reputational risk:	CDOs for UBS

Stylized facts

- data scarcity for companies internally and research (ORX, ORIC)
- loss frequencies vary over time (also: reporting bias)
- loss severities are heavy tailed, often infinite-mean
- losses can be assigned to different business lines (bl; typically 8) or event types (et; typically 7)

. . . and how we model them

- database of 1387 publicly reported events since 1980 (with 950 losses)
- loss frequency: non-homogeneous Poisson process
- loss severities: EVT–POT approach (GPD)
- . . . depending on 10 bl as covariates (and time!)

Goal: Compute Value-at-Risk (VaR) and CIs depending on covariates

EVT based modeling approach

The classical POT approach

- losses $X_{t'_1}, \dots, X_{t'_{n'}} \overset{\text{iid}}{\sim} F$, $\bar{F} \in \mathrm{RV}^{\infty}_{-\frac{1}{\xi}}$ $(\bar{F} = x^{-\frac{1}{\xi}} L, \lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1)$
- $X_{t_1}, \dots, X_{t_n}$ exceedances over u (high enough)
- excesses $Y_{t_i} = X_{t_i} - u > 0$, $i \in \{1, \dots, n\}$

Theorem (Leadbetter (1991))

1) The number of exceedances N_t approximately follows a Poisson process with intensity λ, that is, $N_t \sim \mathrm{Poi}(\Lambda(t))$ with $\Lambda(t) = \lambda t$.
2) The excesses $Y_{t_1}, \dots, Y_{t_{N_t}}$ over u approximately follow (independently of N_t) a $\mathrm{GPD}(\xi, \beta)$ for $\xi \in \mathbb{R}$, $\beta > 0$ with

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\beta), & \text{if } \xi = 0. \end{cases}$$

If $\xi > 0$ (most OpRisk loss models), the approximate likelihood is

$$L(\lambda, \xi, \beta; \boldsymbol{Y}) = \frac{(\lambda T)^n}{n!} \exp(-\lambda T) \prod_{i=1}^{n} g_{\xi,\beta}(Y_{t_i}).$$

Therefore, the log-likelihood splits into the two parts

$$\ell(\lambda, \xi, \beta; \boldsymbol{Y}) = \ell(\lambda; \boldsymbol{Y}) + \ell(\xi, \beta; \boldsymbol{Y}),$$

where

$$\begin{aligned} \ell(\lambda; \boldsymbol{Y}) &= -\lambda T + n \log(\lambda) + \log(T^n/n!), \\ \ell(\xi, \beta; \boldsymbol{Y}) &= \sum_{i=1}^{n} \log g_{\xi,\beta}(Y_{t_i}). \end{aligned}$$

$\Rightarrow$ Maximization can thus be carried out separately for 1) and 2).

A dynamic/smoothing POT approach

- Homogeneity assumptions on λ, ξ, β are often not realistic.
- Assume we have observed vectors $\boldsymbol{z}_i = (t_i, x_i, y_{t_i})$, $i \in \{1, \dots, n\}$ (exceedance time, covariate, excess over u)

The model

1) **Number of exceedances:** a non-homogeneous Poisson process with

$$\lambda = \lambda(x, t) = \exp(f_\lambda(x) + h_\lambda(t))$$

where $f_\lambda(x)$ is a constant for each covariate factor x, $h_\lambda : [0, T] \to \mathbb{R}$ a natural cubic spline. Rewriting leads to

$$\log \lambda = f_\lambda(x) + h_\lambda(t),$$

a generalized additive model (GAM) with logarithmic link function
$\Rightarrow$ Estimate f_λ and h_λ with `mgcv::gam(..., family=poisson)`.

2) **Excess distribution:** Similarly, but for convergence it is crucial that ξ and β are orthogonal in the Fisher information metric
$\Rightarrow$ Replace β by $\nu = \log((1+\xi)\beta)$ (see Cox and Reid (1987)).
The reparametrized log-likelihood is

$$\ell^r(\xi, \nu; \boldsymbol{Y}) = \ell\Big(\xi, \frac{\exp(\nu)}{1+\xi}; \boldsymbol{Y}\Big).$$

Assume that ξ and ν are of the form

$$\xi = \xi(x,t) = f_\xi(x) + h_\xi(t),$$
$$\nu = \nu(x,t) = f_\nu(x) + h_\nu(t),$$

Simultaneously estimating ξ and ν is not possible with `mgcv::gam`.

- What we in fact have are vectors $\boldsymbol{\xi}$ and $\boldsymbol{\nu}$ in $\mathbb{R}^n$ with ith components:

$$\xi_i = f_\xi(x_i) + h_\xi(t_i),$$
$$\nu_i = f_\nu(x_i) + h_\nu(t_i).$$

- To obtain reasonably smooth functions h_ξ, h_ν, we use a penalized log-likelihood approach. The penalized loglikelihood is

$$\ell^p(f_\xi, h_\xi, f_\nu, h_\nu; \boldsymbol{z}.) = \ell^r(\boldsymbol{\xi}, \boldsymbol{\nu}; \boldsymbol{y}) - \gamma_\xi \int_0^T h_\xi''(t)^2\,dt - \gamma_\nu \int_0^T h_\nu''(t)^2\,dt$$

where $\gamma_\xi, \gamma_\nu \geq 0$ are smoothing parameters (larger $\Rightarrow$ smoother curves).

- Let $0 = s_0 < s_1 < \dots < s_m < s_{m+1} = T$ denote the (ordered) distinct values among $\{t_1, \dots, t_n\}$. For a natural cubic spline h,

$$\int_0^T h''(t)^2\,dt = \boldsymbol{h}^\top K \boldsymbol{h}$$

where $\boldsymbol{h} = (h(s_1), \dots, h(s_m))$ and K is a symmetric $m \times m$ matrix of rank $m-2$ only depending on the knots $s_1, \dots, s_m$.

$$\Rightarrow \ell^p(f_\xi, h_\xi, f_\nu, h_\nu; \boldsymbol{z}.) = \ell^r(\boldsymbol{\xi}, \boldsymbol{\nu}; \boldsymbol{y}) - \gamma_\xi \boldsymbol{h}_\xi^\top K \boldsymbol{h}_\xi - \gamma_\nu \boldsymbol{h}_\nu^\top K \boldsymbol{h}_\nu \quad \text{with}$$

$$\ell^r(\boldsymbol{\xi}, \boldsymbol{\nu}; \boldsymbol{y}) = \sum_{i=1}^n \ell\Big(\xi_i, \frac{\exp(\nu_i)}{1+\xi_i}; y_{t_i}\Big)$$

The backfitting algorithm for estimating (ξ, β)

Algorithm (Updater; `gamGPDfitUp()`)

Let $\boldsymbol{\xi}^{(k)} = (\xi_1^{(k)}, \dots, \xi_n^{(k)})$ and $\boldsymbol{\nu}^{(k)} = (\nu_1^{(k)}, \dots, \nu_n^{(k)})$ be given.

1) **Setup:** Specify formulas `xi.formula` and `nu.formula` for `gam()` for fitting $\xi_i = f_\xi(x_i) + h_\xi(t_i)$ and $\nu_i = f_\nu(x_i) + h_\nu(t_i)$.

2) **Update $\boldsymbol{\xi}^{(k)}$:**

   2.1) Newton step: Compute (componentwise)

$$\boldsymbol{\xi}^{\text{Newton}} = \boldsymbol{\xi}^{(k)} - \frac{\ell^r_\xi(\boldsymbol{\xi}^{(k)}, \boldsymbol{\nu}^{(k)}; \boldsymbol{y})}{\ell^r_{\xi\xi}(\boldsymbol{\xi}^{(k)}, \boldsymbol{\nu}^{(k)}; \boldsymbol{y})}.$$

   2.2) Fitting: Compute $\boldsymbol{\xi}^{(k+1)}$ via

$$\texttt{fitted(gam(}\boldsymbol{\xi}^{\texttt{Newton}}\texttt{\textasciitilde xi.formula,...,weights=}-\ell^r_{\xi\xi}\texttt{)).}$$

3) **Given $\xi^{(k+1)}$, update $\nu^{(k)}$:**

3.1) Newton step: Compute (componentwise)

$$\nu^{\text{Newton}} = \nu^{(k)} - \frac{\ell^r_\nu(\xi^{(k+1)}, \nu^{(k)}; y)}{\ell^r_{\nu\nu}(\xi^{(k+1)}, \nu^{(k)}; y)}.$$

3.2) Fitting: Compute $\nu^{(k+1)}$ via

fitted(gam(ν^{Newton}~nu.formula,...,weights=$-\ell^r_{\nu\nu}$)).

- `gamGPDfit()` iterates over this algorithm until convergence
- `gamGPDboot()` additionally computes (post-blackend) bootstrapped confidence intervals
- For more details, use `demo(game)`

Descriptive analysis of the loss data

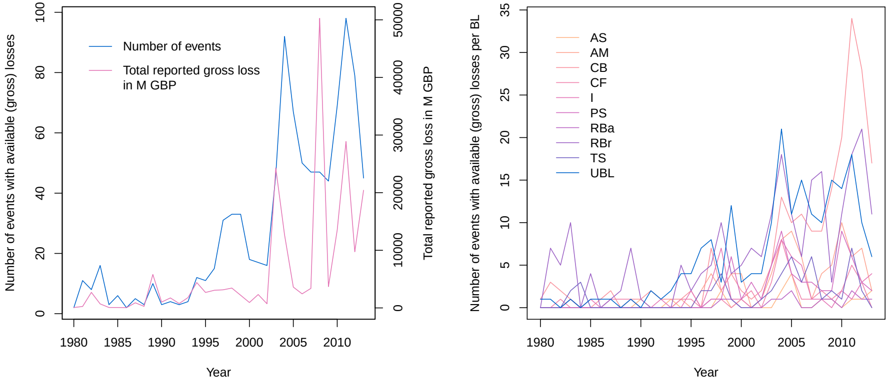
- 1387 OpRisk events collected from public media since 1980 (for the loss severity, we use the 950 reported losses)
- For each event, the following information is given:

  used: business line, event type, year of the event, (gross) loss in GBP (31.51% missing)

  unused: reference number, organization affected, country of head office, country of event, type of insurance, net loss (97.55% missing), regulator involved, source (newspapers, databases, press releases, webpages), loss description

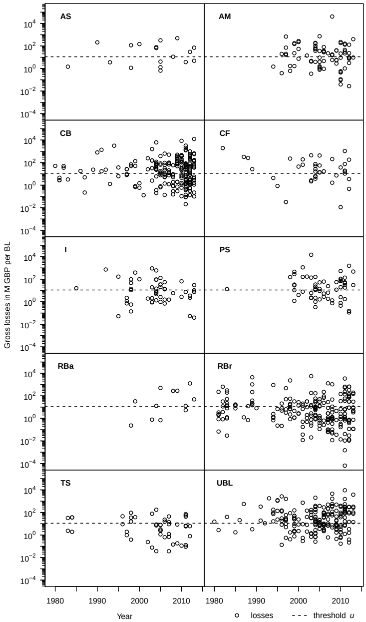
  Not available is the company size.
- Most events happened in USA (44.34%), UK (26.03%), Japan (5.05%), Australia (2.31%), and India (2.02%); China?
- 63.95% were (partially) insured; insurance cover unclear.

Number of available losses and total available loss aggregated per year over time (left). For each business line (right).



- Increasing frequency probably due to reporting bias.
- Frequency depends on the business line.

⇒ Both features our model can take into account.



⇒ Losses are not identically distributed. We will take this into account by interpreting business lines and time as covariates.

⇒ Data pooling

⇒ Pooling is also suggested by the Basel matrix/vector:

	IF	EF	EPWS	CPBP	DPA	BDSF	EDPM		
AS	2	1	0	12	0	0	3	18	AS
AM	12	3	4	55	0	0	6	80	AM
CB	60	54	4	77	1	0	11	207	CB
CF	12	4	0	23	0	0	2	41	CF
I	13	2	2	32	0	0	4	53	I
PS	10	3	0	38	2	0	9	62	PS
RBa	3	0	0	4	0	2	3	12	RBa
RBr	71	62	5	73	1	0	14	226	RBr
TS	13	3	2	28	0	1	2	49	TS
UBL	60	2	20	107	0	3	10	202	UBL

Note: This is aggregated since 1980!

Dynamic POT analysis

Goal: Use all losses from 1980 to 2013 which exceed the threshold u of 11.02 M GBP (median) and compute the risk measure $\mathrm{VaR}_{0.999}$ including 95% bootstrapped confidence intervals.

- Graphical GoF test for the GPD model: If the model is correct,

$$R_i = -\log(1 - G_{\hat{\xi}_i, \hat{\beta}_i}(Y_{t_i})) \overset{\text{approx.}}{\sim} \mathrm{Exp}(1), \quad i \in \{1, \dots, n\}$$

$\Rightarrow$ check with a Q-Q plot $\Rightarrow$ threshold choice

- Given $\hat{\lambda}$, $\hat{\xi}$, $\hat{\beta}$ (evaluated at x_i's and t_i's), an estimate of VaR_α is

$$\widehat{\mathrm{VaR}}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-\alpha}{\hat{\lambda}}\right)^{-\hat{\xi}} - 1\right)$$

- Confidence intervals can be constructed with the post-blackend bootstrap of Chavez-Demoulin and Davison (2005).

Loss frequency

- We fit the following models for λ using `gam(..., family=poisson)`:

  $\log\lambda(x,t) = c_\lambda$ (constant/classical model)

  $\log\lambda(x,t) = f_\lambda(x)$ (bl as covariate)

  $\log\lambda(x,t) = f_\lambda(x) + c_\lambda t$ (bl and time [parametrically] as covariate)

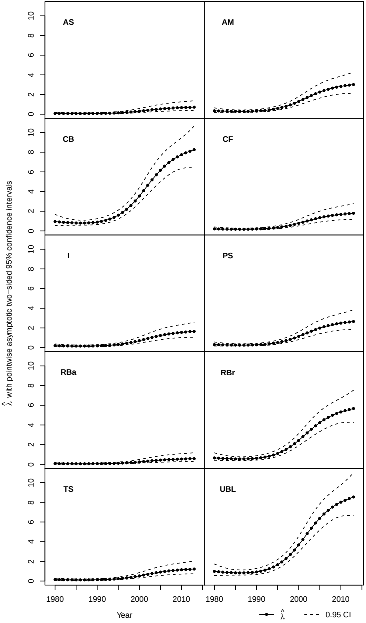
  Likelihood-ratio tests $\Rightarrow$ dependence on bl and time.

- We then compare $\log\lambda(x,t) = f_\lambda(x) + c_\lambda t$ with models of the form

  $$\log\lambda(x,t) = f_\lambda(x) + h_\lambda^{(\mathrm{Df})}(t),\ \mathrm{Df}\in\{1,\dots,8\} \quad \text{(non-parametric)}$$

  $\mathrm{AIC} \Rightarrow$ selected model: $\log\hat{\lambda}(x,t) = \hat{f}_\lambda(x) + \hat{h}_\lambda^{(3)}(t)$

- The selected model shows that considering a homogeneous Poisson process for the occurrence of losses (classical approach) is not adequate.



- Final model for λ:

$$\hat{\lambda}(x,t) = \exp(\hat{f}_\lambda(x) + \hat{h}_\lambda^{(3)}(t))$$

(depends on business line and time)

- 95% confidence intervals (bootstrapped)

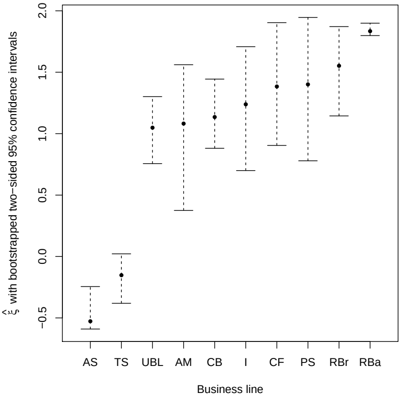
Loss severity

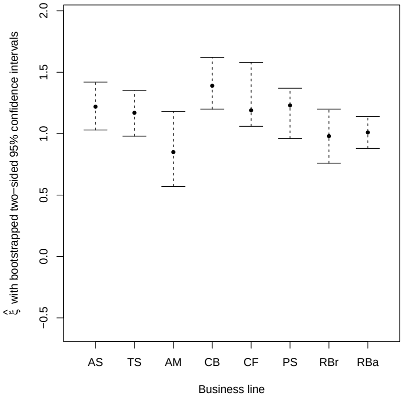
- We fit the following models for (ξ, ν) using `gamGPDfit()`:

$$
\begin{aligned}
\xi(x,t) &= c_\xi, & \nu(x,t) &= c_\nu,\\
\xi(x,t) &= f_\xi(x), & \nu(x,t) &= c_\nu,\\
\xi(x,t) &= f_\xi(x) + c_\xi t, & \nu(x,t) &= c_\nu,\\
\xi(x,t) &= f_\xi(x), & \nu(x,t) &= f_\nu(x),\\
\xi(x,t) &= f_\xi(x), & \nu(x,t) &= f_\nu(x) + c_\nu t,\\
\xi(x,t) &= f_\xi(x), & \nu(x,t) &= f_\nu(x) + h_\nu(t),
\end{aligned}
$$

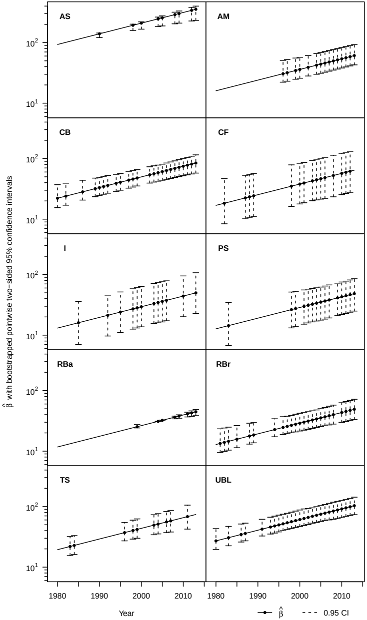
$\Rightarrow$ selected model: $\hat{\xi}(x,t) = \hat{f}_\xi(x)$, $\hat{\nu}(x,t) = \hat{f}_\nu(x) + \hat{c}_\nu t$

- Results about $\hat{\xi}(x,t) = \hat{\xi}(x)$ are similar to Moscadelli (2004) (right)





$\Rightarrow$ Hints at infinite-mean models (in 80% of the cases).

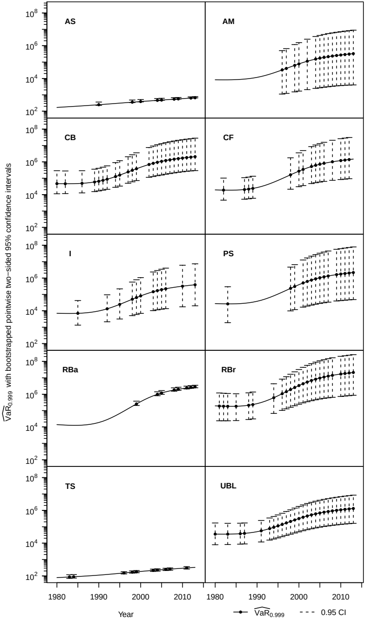


- Final model for β:

$$\hat{\beta}(x, t) = \frac{\exp(\hat{f}_\nu(x) + \hat{c}_\nu t)}{1 + \hat{\xi}(x)}$$

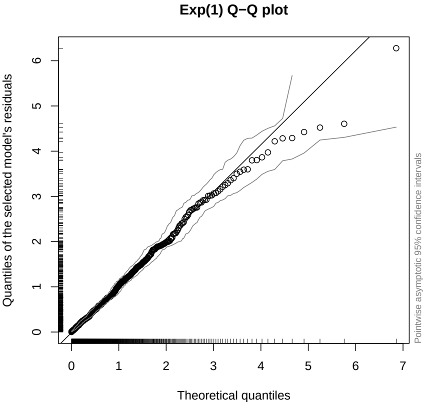
(depends on business line and time)

- 95% confidence intervals (bootstrapped)



- $\widehat{\text{VaR}}_{0.999}$ estimates (depending on time and business line)
- 95% confidence intervals (bootstrapped)

. . . and the residuals are. . .



- Overall fine (asymptotically)
- Depends on the choice of the threshold u (bias–variance trade-off)
- Higher u (e.g. 90%) not possible given the sample size

Thank you for your attention

Technische Universität München