Portfolio modelling of operational losses

John Gavin[1], QRMS, Risk Control, UBS, London. April 2004.

- What is operational risk
- Trends over time
- Empirical distributions
- Loss distribution approach
- Compound distributions
- References

[1]The views expressed are those of the author and necessarily those of UBS AG.

What is operational risk

- Operational losses usually mean unexpected losses from the failure of normal business processes.

  There is no universal agreement on the definition of operational risk but, for banks, it usually includes losses from transaction processing error (e.g. fraudulent cheques), security (e.g. virus infection) and liability (e.g. sued for giving bad advice).

- Regulators will require banks to set aside capital to provide a buffer against such losses. Calculating this risk capital is the subject of this paper.

- Operational loss data typically consists of events where the vast majority of losses are small in magnitude, combined with a few losses *several* orders of magnitude higher.

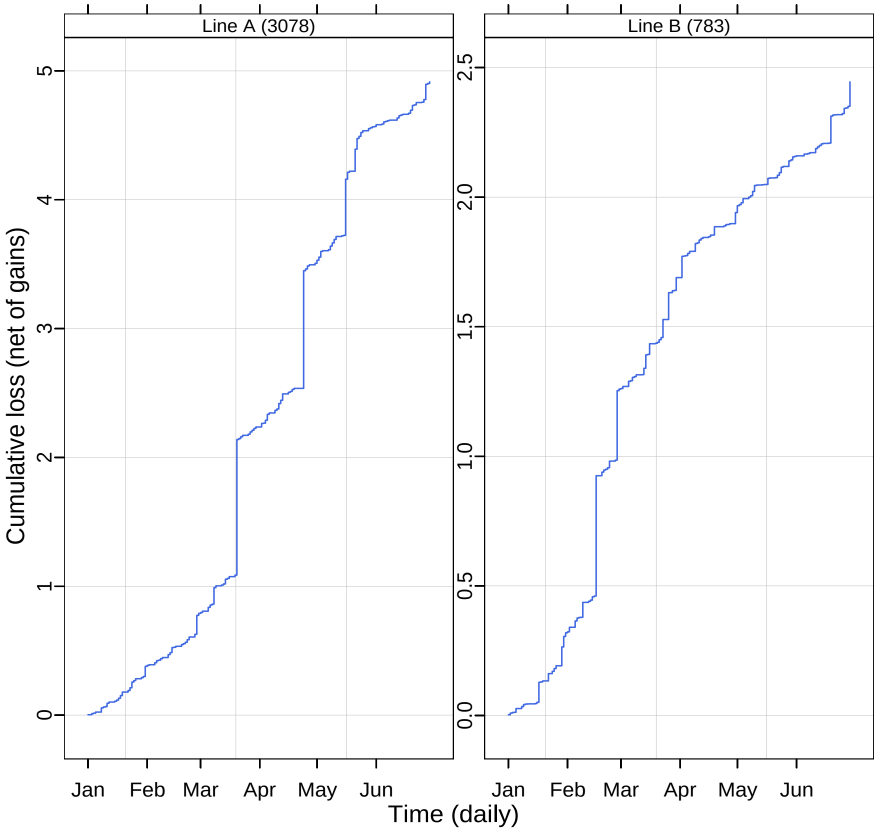
- It is generally believed that the occasional large loss is what impacts the bank's reputation and what most concerns regulators, due to the possibility of systemic risk.

R and UBS

R is being introduced within UBS as a complement to S-Plus for cost reasons.

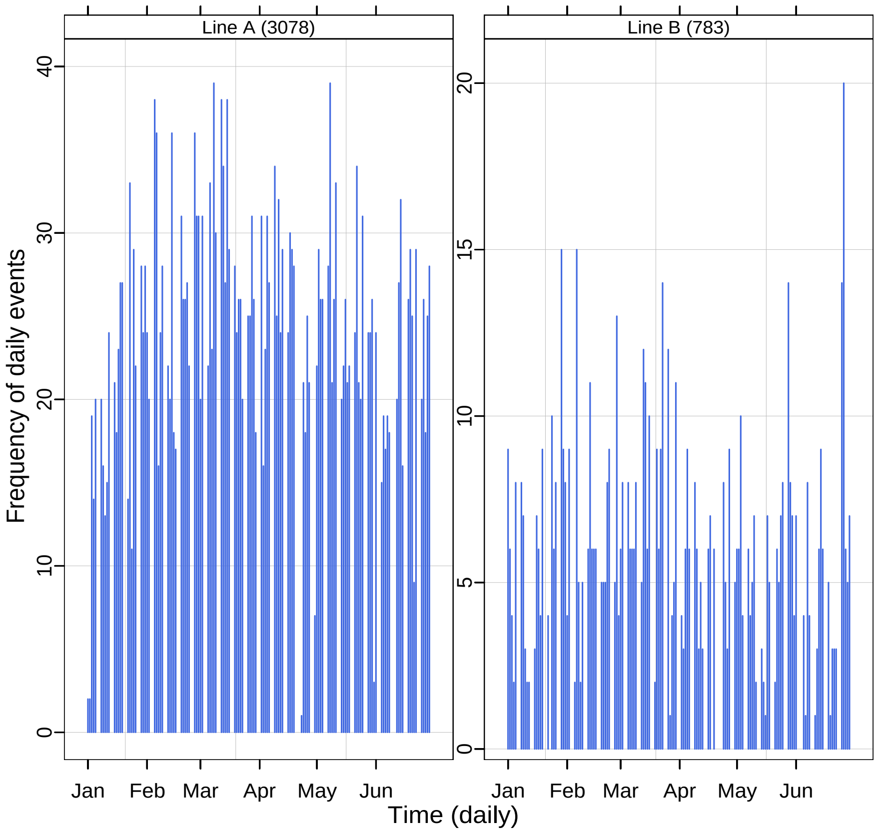
It is an ideal platform for data analysis and statistical modelling for experienced users [1]. For example, this presentation was prepared using R, Sweave, XEmacs and ESS. R (on Windows) needs a GUI for less experienced users, who invariably rely on Excel.

Trends over time



Cumulative daily losses for two business lines (event counts in brackets). To illustrate the method, consider (rescaled) loss events arising from transaction processing of corporate trade settlements.

Small losses steadily accumulate, with a few large losses. The latter drive overall costs and hence the risk.

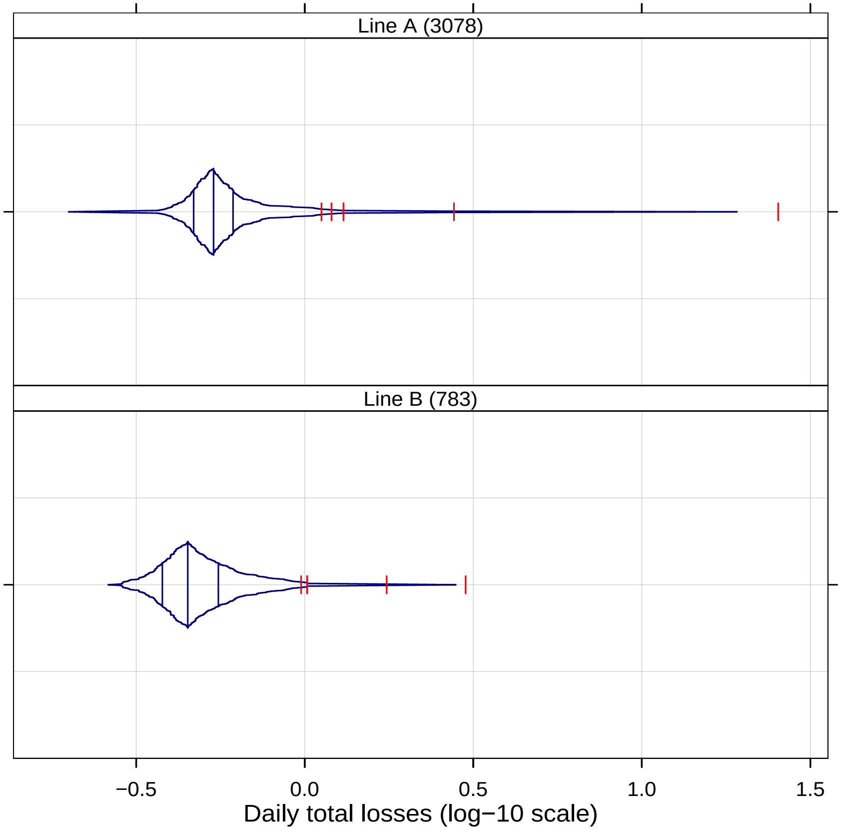


Daily count of events

In general, time trends and clustering are difficult to detect and are complicated by the definition of date (e.g. occurrence, identification or reporting).

Here, daily counts are used because of the limited data. Typically, weekly or monthly figures are used for practical reasons.

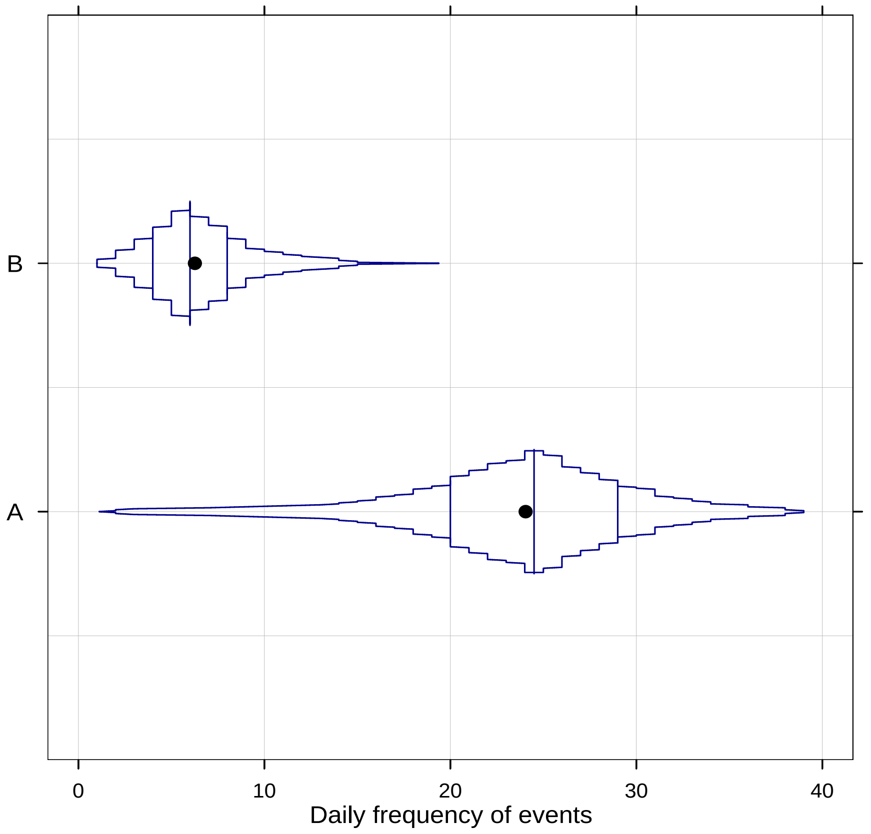
Empirical distributions



Percentile plots of event severity on a log-10 scale. Event severity has a very skewed distribution.

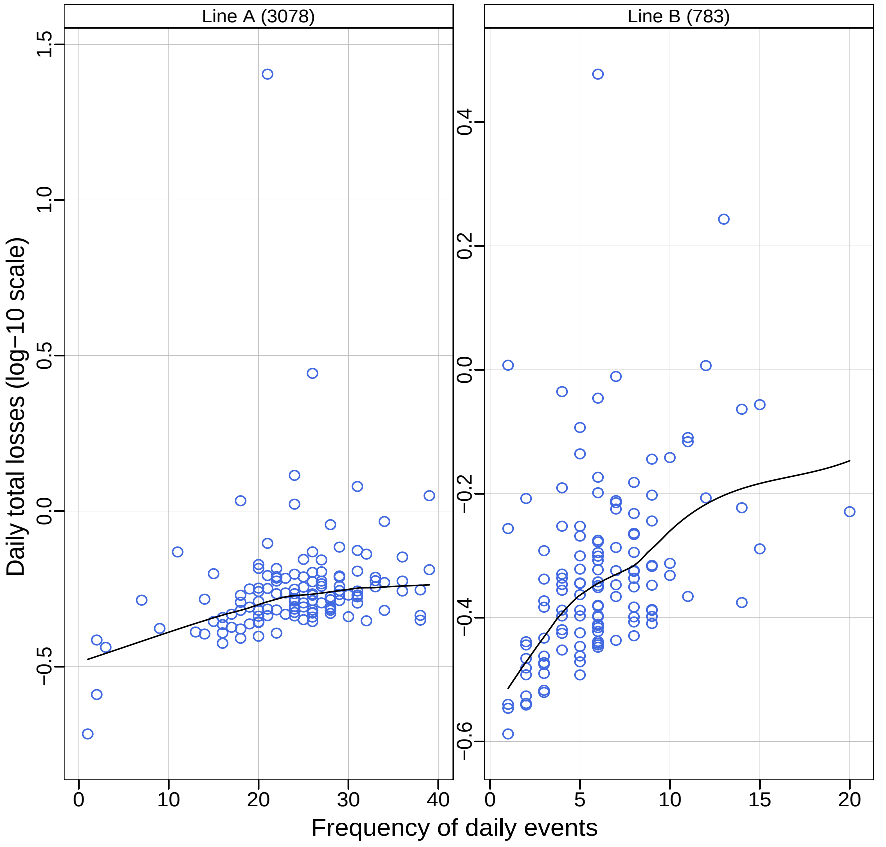
The verticle blue lines are the 25th, 50th and 75th percentiles.

The top 5 individual losses are shown as red tick marks.



Percentile plots of event frequency are not skewed, compared to severity.

For longer time horizons (e.g. monthly), the empirical frequency distributions tend towards a Poisson for high-frequency data, like transaction processing errors.



Dependence between frequency and severity

In general, the largest total-daily losses are not due to a large number of small losses; typically they consist of one large loss combined with some small losses.

There is no evidence of dependency between frequency and severity for longer time horizons.

Loss distribution approach

The LDA is based on an actuarial approach, fitting separate distributions to the frequency N and severity X of events, then combining them into a compound distribution, $Z = \sum_{i=1}^{N} X_i$.

Event frequency

For comparison, two frequency distributions are fitted: Poisson and negative binomial.

	lambda	size	mu
Line A (3078).estimate	24.0	13.4	24.0
Line B (783).estimate	6.3	9.1	6.3
Line A (3078).sd	0.4	2.9	0.7
Line B (783).sd	0.2	2.8	0.3

Table 1: Parameter estimates of event frequency - for the Poisson (`lambda`) and negative binomial (`size` and `mu`) distributions, by MLE, with standard errors on the bottom two rows.

QQ-plots with (smoothed) confidence envelopes are used as a graphical goodness of fit test. The results are in figures 1 and 2.

Event severity

The severity distribution is based on the semiparametric model of Coles and Tawn (1991, 1994) and Heffernan and Tawn 2003 [2]. (The latter also discusses dependency in the tail between two marginals.)

Univariate extreme value theory provides an asymptotic justification for the generalized Pareto distribution (gpd) as an appropriate model for the distribution of excesses over some sufficiently high threshold u [3]. So the (marginal) tail of the severity distribution, X is

$$\Pr(X > x + u_X | X > u_X) = (1 + \xi x/\beta)^{-1/\xi},$$

where $x > 0$ and β and ξ are the scale and shape parameters, respectively.

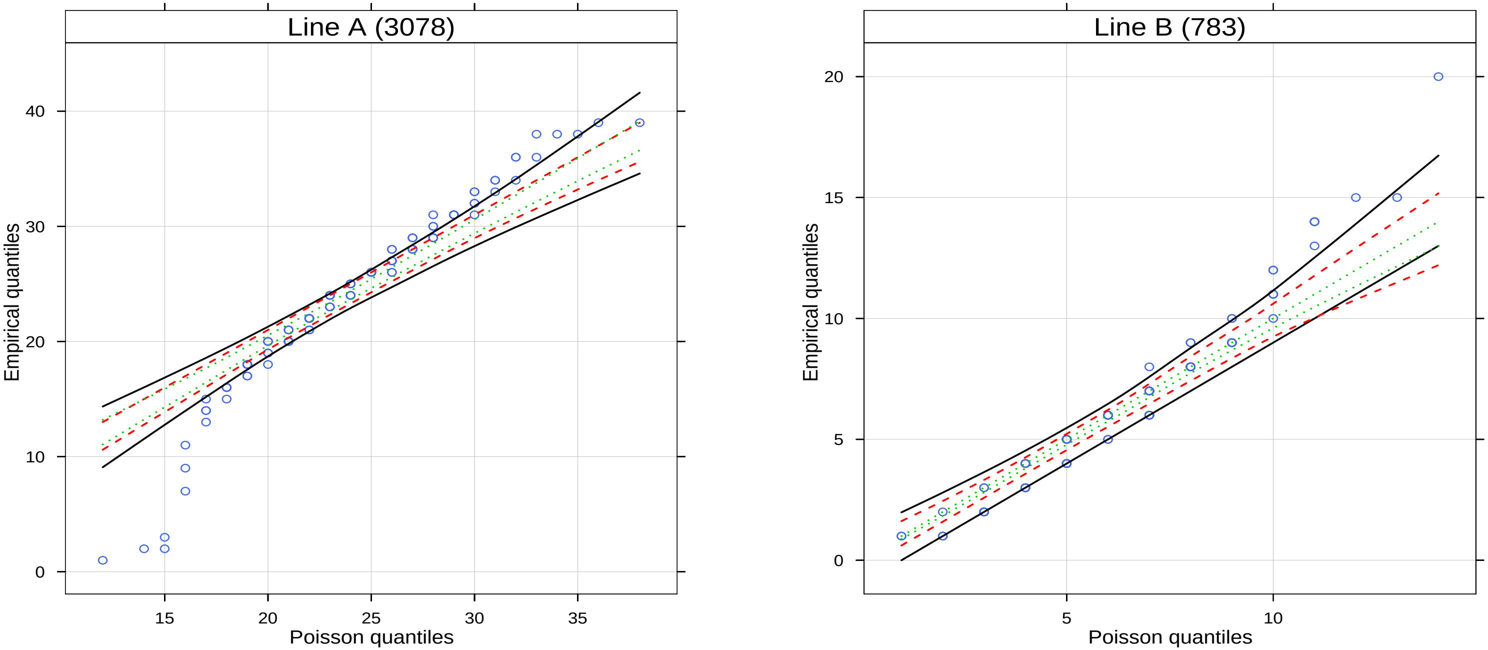


Figure 1: QQ-envelopes for Poisson fits - The lines are smoothed upper and lower 2.5, 12.5 and 25 percent confidence envelopes, for the theoretical distributions, via Monte Carlo. Days when lots of events occur are of interest but the Poisson tends to underestimate those. Days with no events are less of a concern economically.

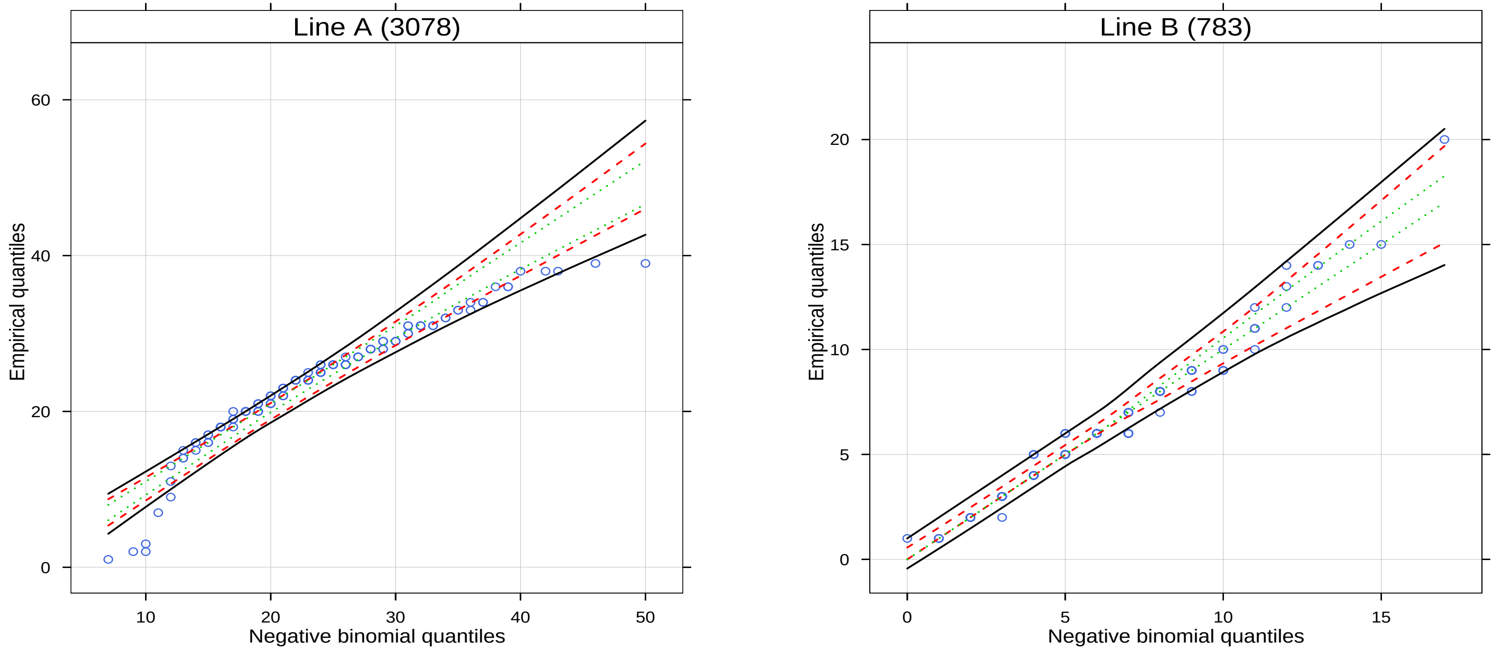


Figure 2: QQ-envelopes for negative binomial fits - The negative binomial is less likely to underestimate the largest number of events that might occur on a day. It is a more conservative choice than a Poisson, from a risk capital viewpoint.

With $\tilde{F}_X$ as the empirical distribution of X, the unconditional distribution is

$$\hat{F}_X(x) = \begin{cases} 1 - \{1 - \tilde{F}_X(u_X)\}\{1 + \xi(x - u_X)/\beta\}^{-1/\xi}, & \text{for } x > u_X \\ \tilde{F}_X(x) & \text{for } x \leq u_X. \end{cases}$$

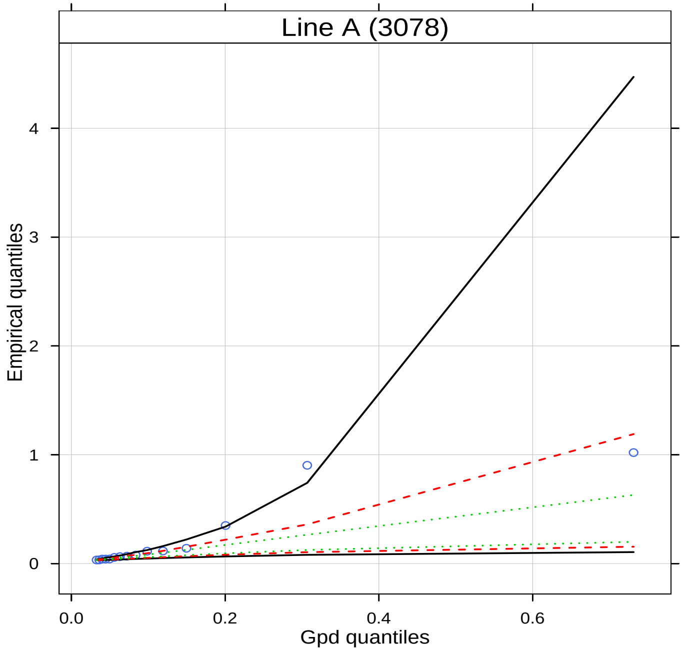
Fitting is based on probability weighted moments. (MLE and method of moments are also considered but not shown in detail.)
Given specific numbers of exceedences above a threshold, sample threshold values are:

	15	30	60
Line A (3078)	0.031	0.012	0.006
Line B (783)	0.035	0.012	0.005

Given different threshold levels, parameter estimates of event severity for both lines are:

	subPort	numExceed	xi	beta	threshold	probLtThres	worstLosses
1	Line A (3078)	60	0.84	0.01	0.006	0.981	1,0.9,0.3,0.1
2	Line A (3078)	29	0.77	0.02	0.012	0.991	1,0.9,0.3,0.1
3	Line A (3078)	15	0.74	0.05	0.031	0.995	1,0.9,0.3,0.1
4	Line B (783)	15	0.67	0.02	0.035	0.981	0.5,0.2,0.1,0.1
5	Line B (783)	60	0.65	0.01	0.005	0.923	0.5,0.2,0.1,0.1
6	Line B (783)	30	0.49	0.02	0.012	0.962	0.5,0.2,0.1,0.1

Estimates are via probability weighted moments.



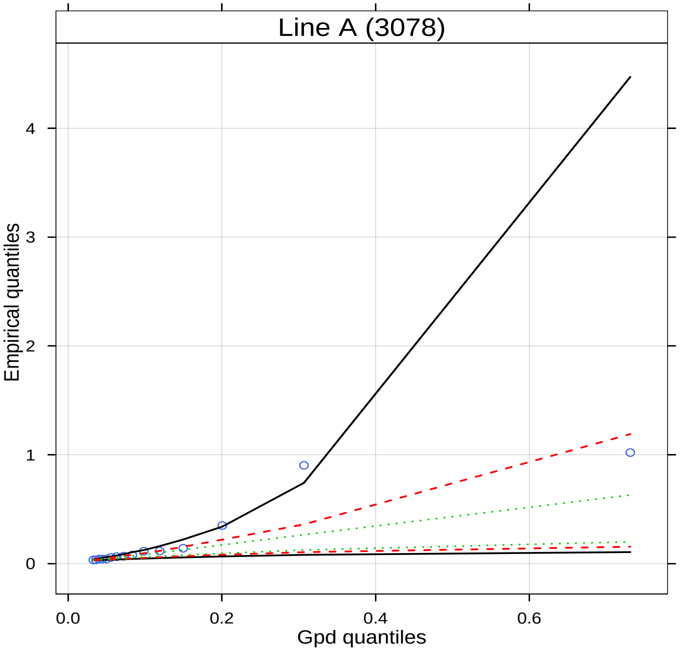


Figure 3: QQ-envelopes of Gpd fits for line A - with 15 and 30 points in the tails, respectively. The level of uncertainty generated by fat-tailed distributions is very material. (These gpd confidence envelopes are unsmoothed.)

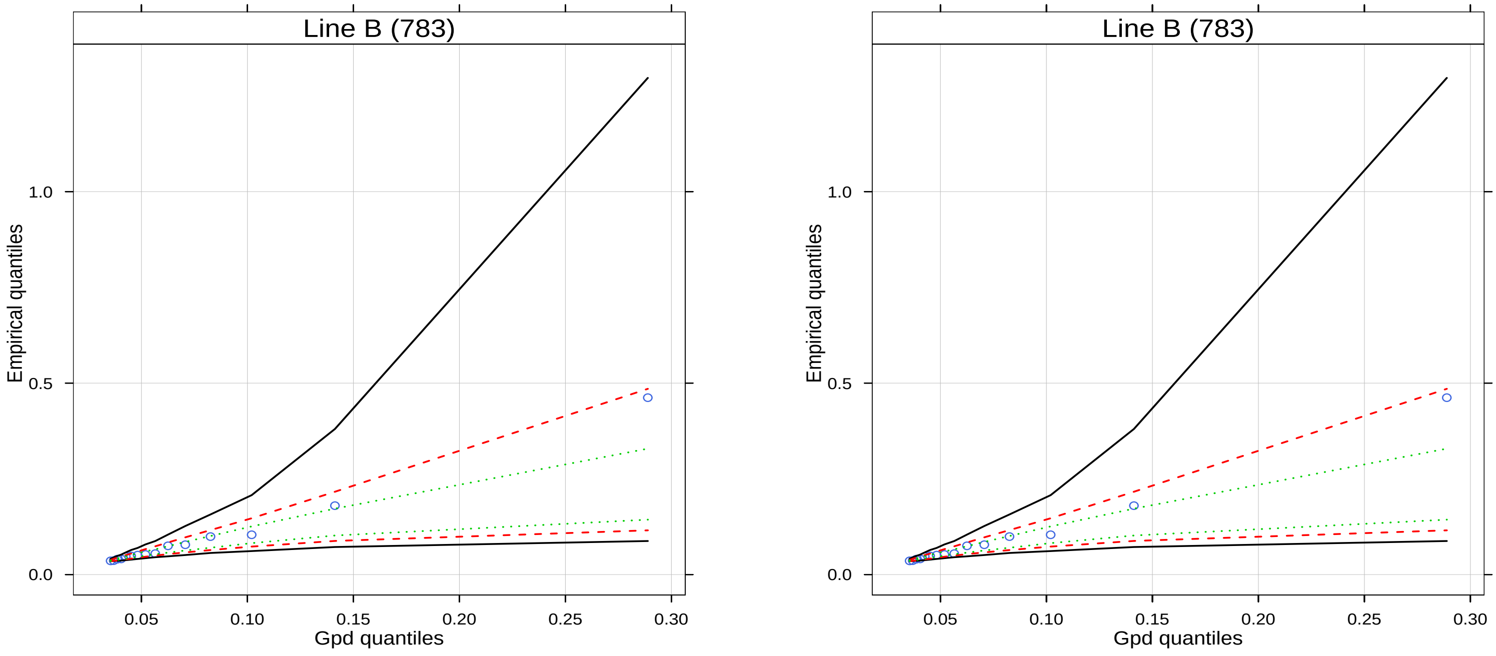
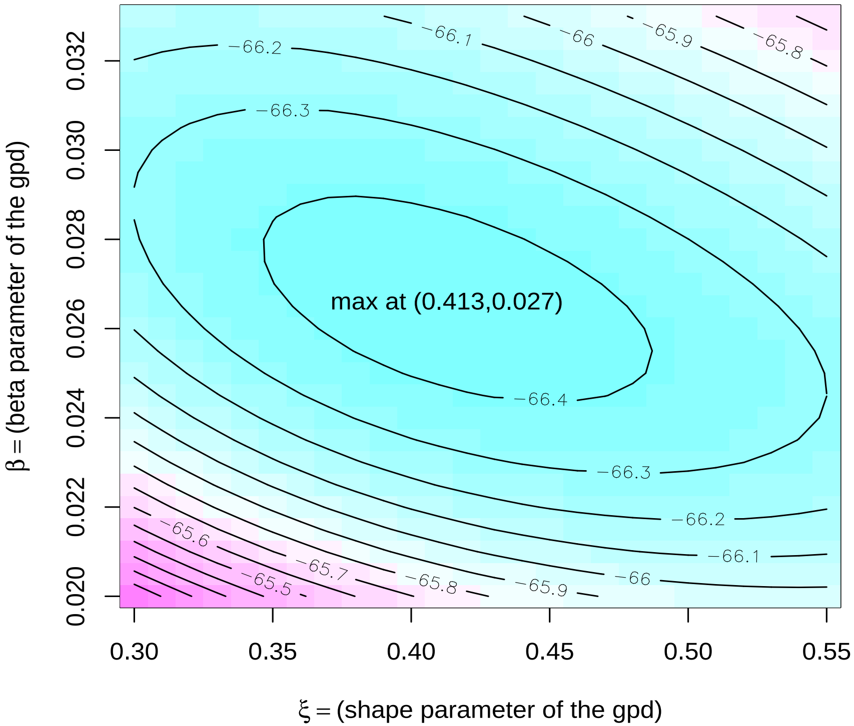
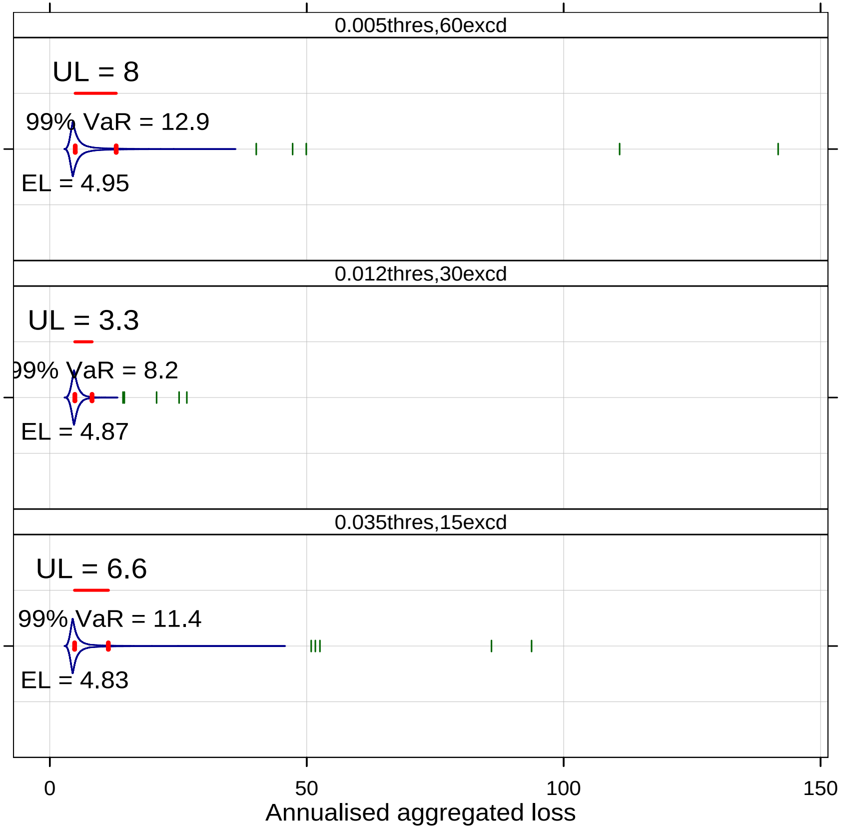


Figure 4: QQ-envelopes of Gpd fits for line B - with 15 and 30 points in the tails, respectively.



Profile of the MLE – for business line B, with 30 exceedences in the tail. Typically, the MLE surface if fairly flat, especially for the shape parameter, ξ. This large standard error can result in materially different capital charges, see [4] for more. Compare estimates to probability weighted moments in table .

Compound distributions



Percentile plots of annualised loss distributions for line B with different thresholds.

Expected loss (`EL`), value at risk (`VaR`) and capital charge (`Unexpected Loss`) are annotated. UL is very sensitive to the shape (ξ) parameter.

(Scaling from daily to annual (252) and 99% VaR are *heroic* assumptions.)

References

References

[1] R Development Core Team (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3. URL: Comprehensive R Archieve Network http://www.R-project.org. 3

[2] A conditional approach for multivariate extreme values, J. E. Heffernan and J. A. Tawn, RSS Research Section Ordinary Meeting, Wednesday, October 15th, 2003. 10

[3] Pickands, J. (1975) Statistical inference using extreme order statistics. Ann. Statist., 3, 119–31. 10

[4] Using Loss Data to Quantify Operational Risk, Patrick de Fontnouvelle, Virginia DeJesus-Rueff, John Jordan, Eric Rosengren, Federal Reserve, Bank of Boston, April, 2003. See section VI (page 19) and Table 5 (page 32). 17

Resource information

Portfolio modelling of operational losses
John Gavin

```
rwnFile     =  opsPres.Rnw
infoRnwFile =  modified = 2004-05-24 16:57:34, size = 69kb.
projectDir  =  S:/QRMS/Stats/R/research/ops/
[1] "User, sys, tot elap time current process, cum sum user, system times child processes"
[1]  90.16   6.90 115.11     NA     NA
[1] "R version 1.9.0, 2004-05-06"
   OS.type   file.sep dynlib.ext        GUI     endian
 "windows"        "/"     ".dll"    "RTerm"   "little"

                 _
platform i386-pc-mingw32
arch     i386
os       mingw32
system   i386, mingw32
status   Patched
major    1
minor    9.0
year     2004
month    05
day      06
language R
```