Insurance

Data

Science

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Truncated regression models for the analysis of operational losses due to fraud:

A high performance computing implementation in R

About the speaker

- **Alberto Glionna**
- Senior Risk Analyst
- Alberto has a MSc in actuarial science and has 4 years of working experience gained in the non life insurance industry

---

- **Assicurazioni Generali**
- The 3rd largest insurer in Europe
- Over 70 thousands employees around the World
- Operating in more than 100 countruies

Agenda

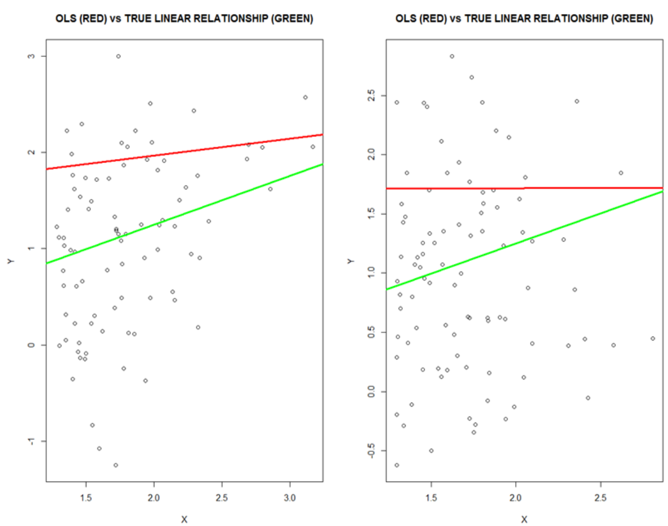
- Truncation Issues in Insurance Fraud Modeling
- AGLM
- Computational Challenges
- High Performance Computing in R
- Q&A
- Back Up

Truncation Issues in Insurance Fraud Modeling

Standard Truncation

$$Y_i^* = X_i\beta + \varepsilon_i \quad \leq L \qquad included$$

$$Y_i^* = X_i\beta + \varepsilon_i \quad \geq L \qquad excluded$$



$$L(\beta,\sigma^2 \,|\, \boldsymbol{y},\boldsymbol{X}) = \prod_{i=1}^{n} \frac{\frac{1}{\sigma}\varphi\left(\frac{y_i - \boldsymbol{x}_i\beta}{\sigma}\right)}{1-\phi\left(\frac{L-\boldsymbol{x}_i\beta}{\sigma}\right)}$$

Fraud Truncation

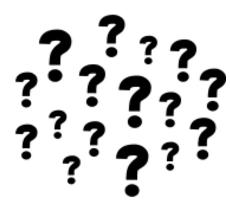
$$Y_i^* = \sum_{j=1}^{N_i^*} Z_j^*$$

$$Y_i^* = X_i\beta + \varepsilon_i \qquad Z_j^* \leq L \quad excluded$$

$$Y_i^* = X_i\beta + \varepsilon_i \qquad Z_j^* \leq L \quad included$$

The truncation has an impact on both Frequency and Severity

Standard Normal Truncated OLS may not be accurate to model these effects



AGLM (1/2)

Augmented Linear Models (AGLM) assume the following:

- $Y_i^* \left(= \sum_{j=1}^{N_i^*} Z_j^* \right)$ independent variables $\forall\, i$
- $Z_{i,1}^* | N_i^* = n_i, \dots, Z_{i,n_i}^* | N_i^* = n_i$ are i.i.d. random variables $\forall\, i$ following a lognormal distribution with parameters $Ln(\beta_0^\mu + \beta_1^\mu x_i; \sigma)$
- $\forall i\ N_i^* \sim Pois(e^{\beta_0^\lambda + \beta_1^\lambda x_i})$

The conditional expected value (with a logarithmic link) may be written as:

$$log(\lambda_i) + \left(\mu_i + \frac{\sigma^2}{2}\right) = \beta_0{}^\mu + \beta_0{}^\lambda + (\beta_1{}^\mu + \beta_1{}^\lambda)x_i$$

$$log(\lambda_i) + \left(\mu_i + \frac{\sigma^2}{2}\right) = \beta_0 + \beta_1 x_i$$

Linear Model → there are two nested linear model: one for the Frequency and one for the Severity

AGLM (2/2)

The Log-Likelihood function is:

$$\pi_i = \left(1 - F_Y{}^{LogNorm}\left(c, \beta_0{}^{\mu} - \frac{\sigma^2}{2} + \beta_1{}^{\mu} x_i; \sigma\right)\right)$$

$$l\left(\underline{y}, \underline{x}; \underline{\beta}\, \sigma^2\, c\right)$$

$$= \sum_{i=1}^{n} \left[n(\beta_0{}^{\lambda} + \beta_1{}^{\lambda} x_i) + n log(\pi_i) - \left(e^{\beta_0{}^{\lambda} + \beta_1{}^{\lambda} x_i}\right)\pi_i - log(n_i!) \right.$$

$$\left. + \sum_{j=1}^{n_i} log\left(f_Y{}^{Lognotrm}\left(z_j; \beta_0{}^{\mu} - \frac{\sigma^2}{2} + \beta_1{}^{\mu} x_i; \sigma\right)\right) - log(\pi_i) \right]$$

This is possible thanks to a property of the Poisson distribution and, in general, of the Panjer Class:

$$N_D = \sum_{i=1}^{N} I_{\{Z_i > c\}} \qquad Pr(Z > c) = \pi$$

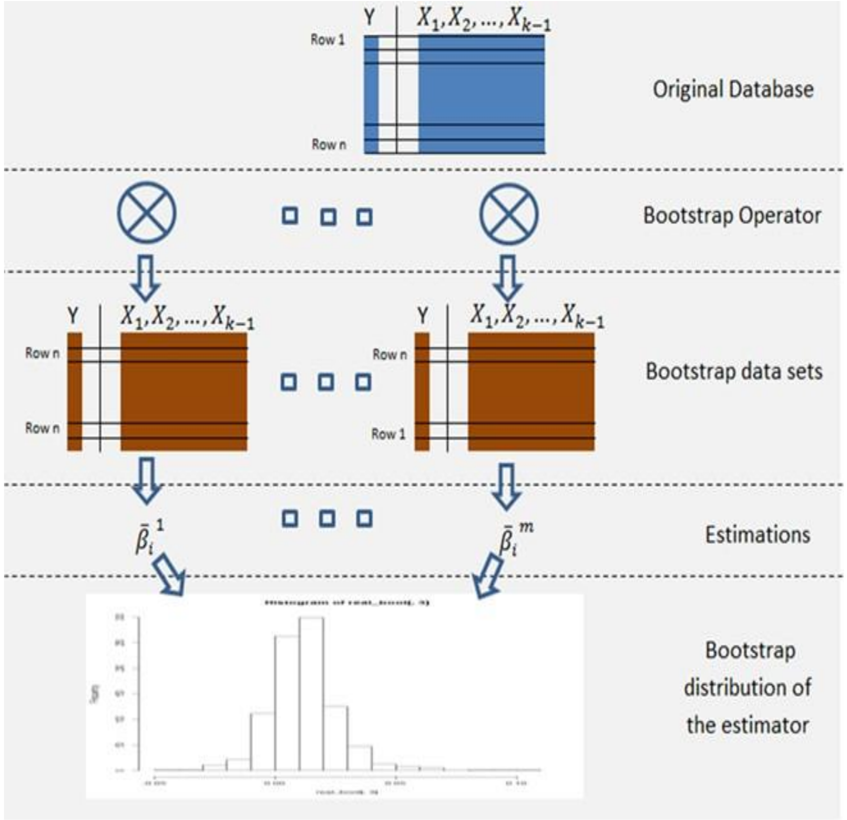
$$N \sim Pois(\lambda) \quad => \quad N_D \sim Pois(\pi\lambda)$$

Computational Challenges

The numerical calculation of both confidence intervals and *p*-values are very time-consuming exercises given the complexity associated with the numerical optimization of the Log-Likelihood function

Luckily, bootstrap simulations, being independent calculations, are often eligible for parallelization

The package *parallel* allowed for a practical implementation of the AGLMs overcoming one of their main limitations



Insurance

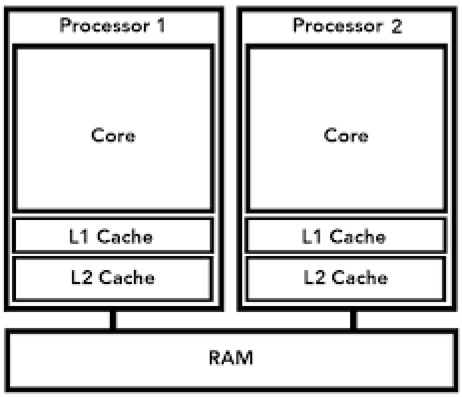
Data

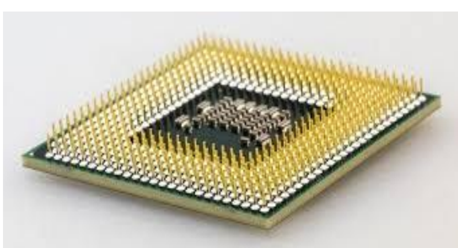
Science

High Performance Computing in R (1/2)

Evaluate *F(x)* *n* times

n/2 on core A n/2 on core B





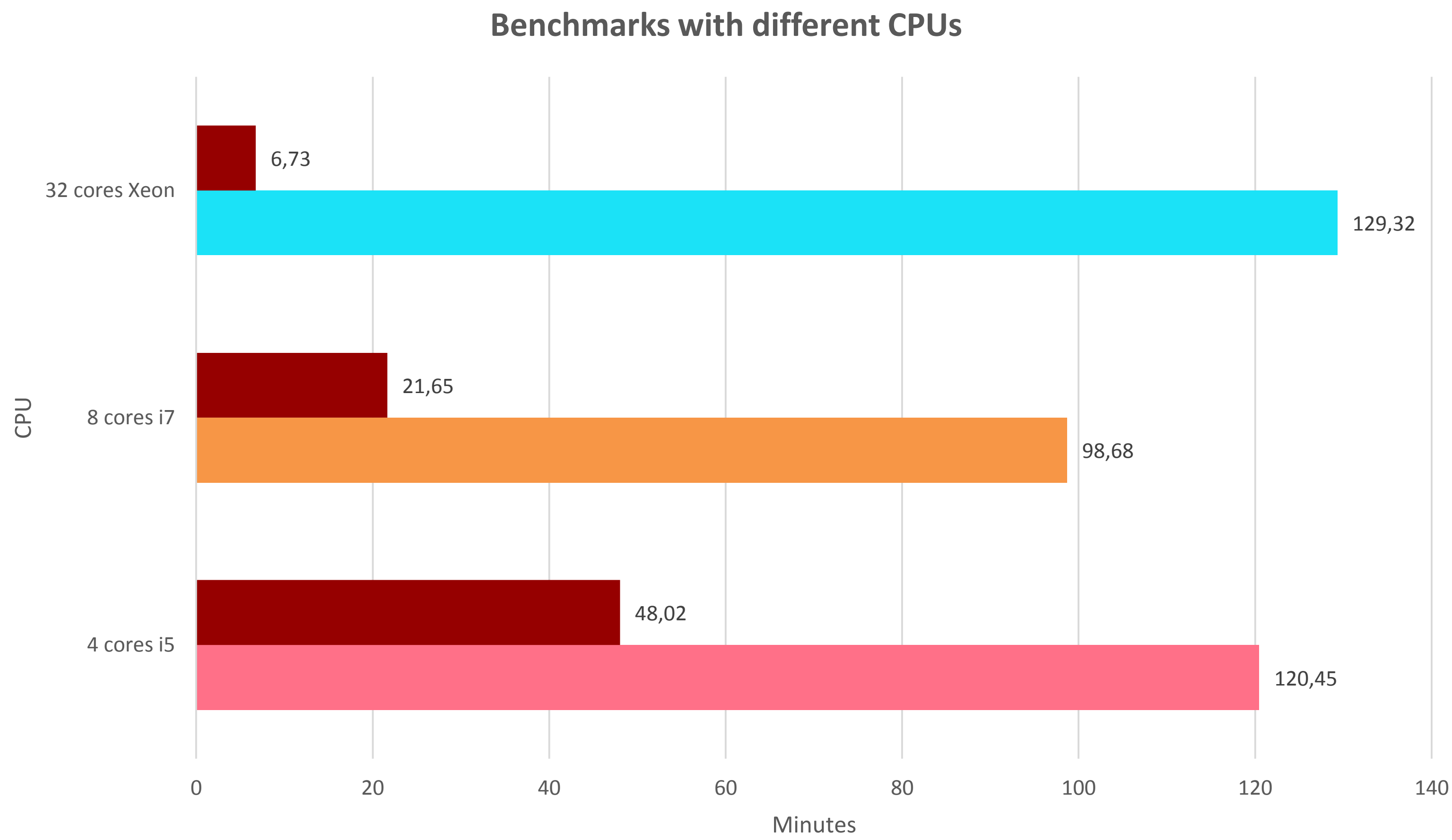
Socket Vs **Forking**

```
cl = makeCluster(4)

clusterExport(cl, ls(), envir = environment())
 system.time(parLapply(cl,1:100,function(x)
Parametric_bootstrap_sample(a,b,c,d,sigma,Regr
essore,truncation))
 )


stopCluster(cl)
```

High Performance Computing in R (2/2)



Thanks for your attention

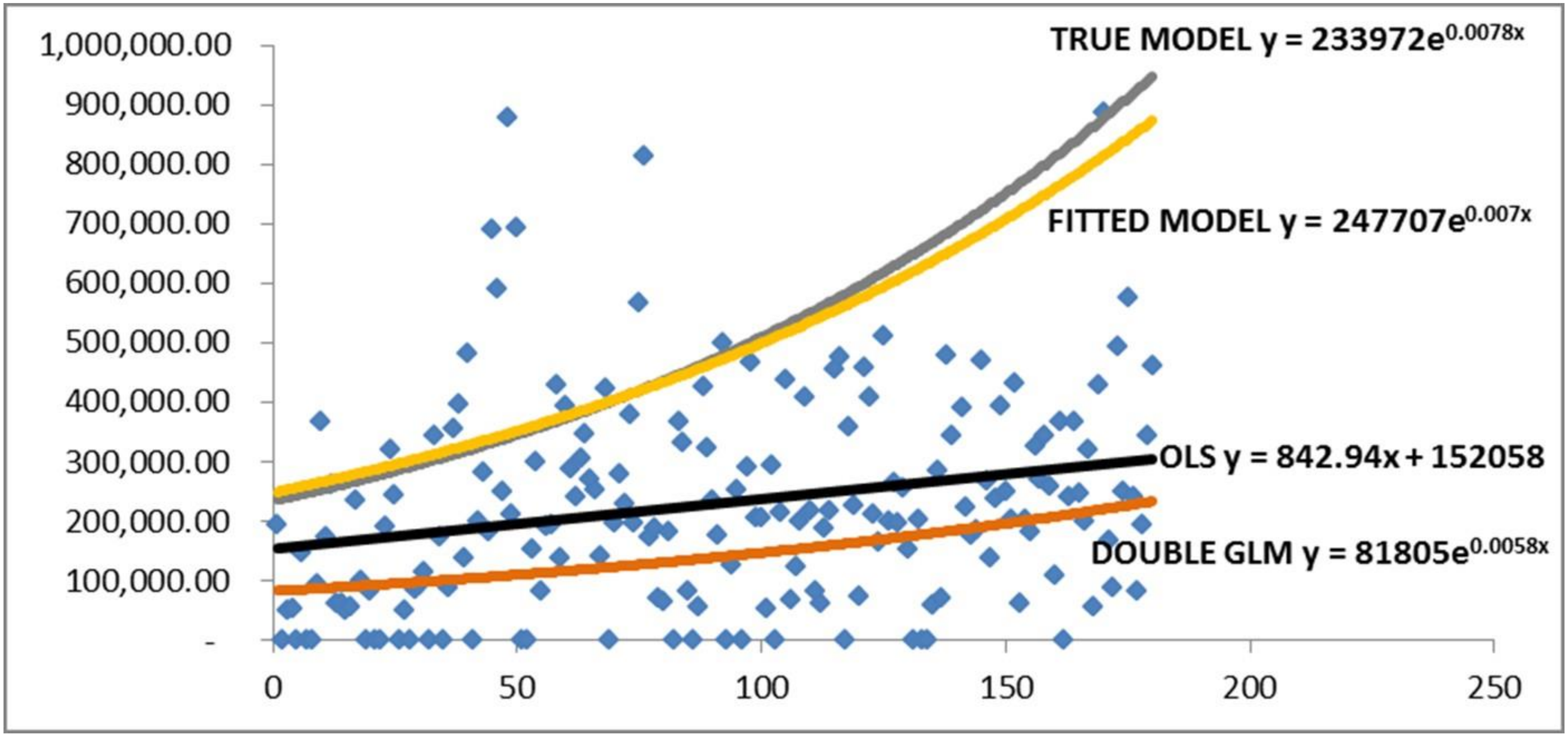
Contacts:

albertgl@me.com (Alberto Glionna personal) Assicurazioni Generali S.p.a
alberto.glionna@generali.com (Alberto Glionna work) Università Degli Studi Di Trieste

nic.torelli@gmail.com (Nicola Torelli personal) Università Degli Studi Di Trieste

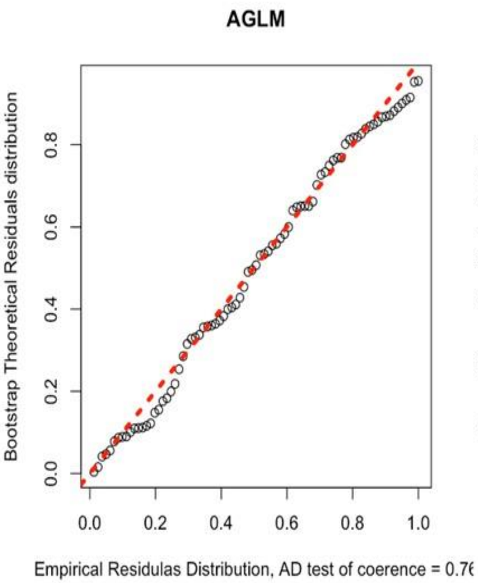
Giovanni.millo@generali.com (Giovanni Millo work) Università Degli Studi Di Trieste
Assicurazioni Generali Trieste

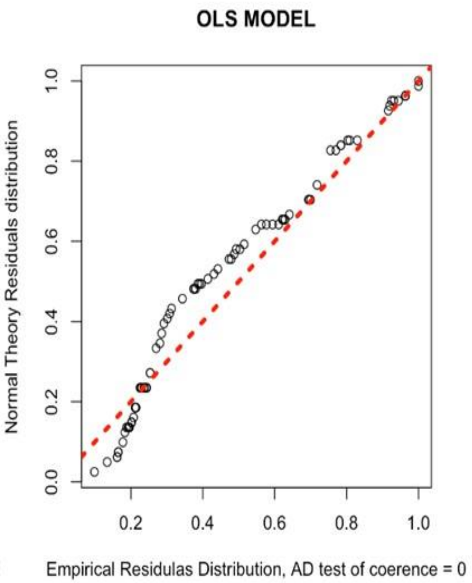
Back Up

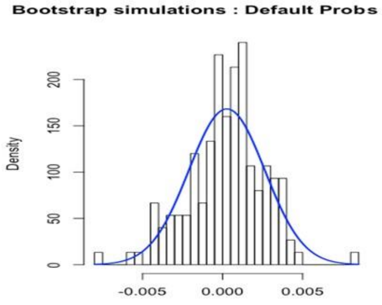

TRUE MODEL y = 233972e^0.0078x
FITTED MODEL y = 247707e^0.007x
OLS y = 842.94x + 152058
DOUBLE GLM y = 81805e^0.0058x
1,000,000.00
900,000.00
800,000.00
700,000.00
600,000.00
500,000.00
400,000.00
300,000.00
200,000.00
100,000.00
-
0
50
100
150
200
250

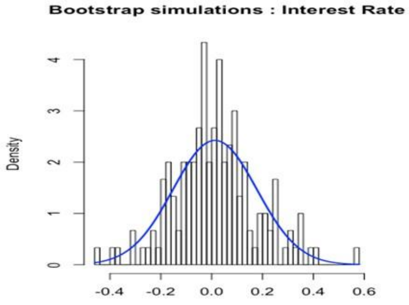
Back Up

	Estimate	Boot Std.errors	5% Boot P	95% Boot Upper Percentile	P value
sigma	3.5227	0.0396	3.4526	3.5829	0%
freq_intercept	6.2982	0.0642	6.1894	6.4008	0%
freq_GDP.g	0.0001	0.0000	0.0000	0.0001	17%
freq_EU.28.core.cpi	1.2749	0.1372	1.0932	1.5446	28%
freq_X10Y.bund..price.	-0.5243	0.0362	-0.5859	-0.4670	1%
freq_Ftse.Europe	0.0002	0.0007	-0.0008	0.0014	47%
freq_usd.eur	-62.4089	0.8765	-63.5263	-60.6429	0%
freq_Non.Performing.Loans	2.1563	0.0834	2.0303	2.3048	21%
freq_Default_Probabilities	-0.0035	0.0001	-0.0037	-0.0033	7%
sev_intercept	2.2461	0.0581	2.1720	2.3631	0%
sev_GDP.g	-0.0001	0.0000	-0.0001	0.0000	2%
sev_EU.28.core.cpi	-1.6322	0.3023	-2.0253	-1.0307	23%
sev_X10Y.bund..price.	0.5807	0.0384	0.5276	0.6539	1%
sev_Ftse.Europe	0.0001	0.0011	-0.0018	0.0018	50%
sev_usd.eur	73.2776	1.9904	71.2863	77.8341	0%
sev_Non.Performing.Loans	-1.5149	0.1037	-1.6406	-1.2995	10%
sev_Default_Probabilities	0.0023	0.0013	-0.0005	0.0039	0%









Frequenza Media mensile	5,494	Frequenza Media Annua	65,929
Importo medio per evento	€ 23,119.32	Totale medio annuo	€ 1,524,244,758.27
Totale medio mesile	€ 127,020,396.52		