Earthquakes economic costs through rank-size laws

Valerio Ficcadenti, Roy Cerqueti

University of Macerata
Department of Economics and Law

v.ficcadenti@unimc.it

June 2, 2019

Objectives

We use a rank-size approach to design an economic impact measure of the earthquakes.

- We employ two different rank-size laws to estimate the functional relationship between earthquakes and economic cost generated by them.
- We compare the results on different datasets to validate the approach.

Rank-size: the Zipf's kick off

Starting from Zipf (1935, 1949) the rank-size analysis has grown exponentially. Zipf outlined cases in which the rank-frequency relationship is particularly evident:

- Words occurrence within a text (e.g. Ulysses of James Joyce).
- City population distribution within a country (e.g. 100 largest metropolitan districts in the United States in 1940)

Zipf explains it thanks to the principles of Diversification and Unification.

Rank-size: the Mandelbrot's contribution

The Zipf-Mandelbrot's law (ZML, hereafter) is born from two Mandelbrot's landmark studies Mandelbrot (1953, 1961).

$$f(r) = \frac{\alpha}{(r+\beta)^{\gamma}} \tag{1}$$

where r is the rank and $f(r)$ is its respective value. α, β and γ are the parameters to estimate.

Rank-size: the lack of a theoretical model

There exists a link of hyperbolic type between rank and size, why?

Despite many attempts of theoretical explanations like Simon (1955); Hill (1974), the reason why Zipf's law is found to be a valid tool for describing rank-size rule is still an active field of research.

The lack of a theoretical ground associated to such a statistical property is particularly evident for some sets of data, see Fujita et al. (2001); Fujita and Thisse (2000) for the case of economic geography.

Rank-size: applications

Rank-Size analysis shows the distribution of size by rank.
This method has been widely used in:

- Economic geography - Cerqueti and Ausloos (2015);
- Business size field - Bottazzi et al. (2015);
- Biology - Li and Yang (2002);
- Computer Science - Maillart et al. (2008);
- Linguistic - Piantadosi (2014);

Rank-size: weakness

Certain rank-size laws poorly fit dataset when the distributions tails have many outliers, e.g. Ioannides and Skouras (2013); Matlaba et al. (2013). In Ausloos and Cerqueti (2016) a Universal Law (UL hereafter) is proposed:

$$f(r) = k\frac{(N+1-r+\psi)^{\xi}}{[N(r+\phi)]^{\lambda}}, \tag{2}$$

where r is the rank and $f(r)$ is its respective value. k, ψ, ξ, ϕ, λ are the parameters to calibrate on the data and N corresponds to the number of observations.

Rank-size: fit

We implement the best fit procedures through the **Levenberg-Marquardt Nonlinear Least-Squares Algorithm**:

- It minimizes the sum square vector returned by the f(r) considered;
- The Levenberg-Marquardt algorithm needs **starting points** to properly estimate the parameters of the studied functions;
- We use a modification of the basic model. It needs a set of starting points for estimating and the **brute-force algorithm** find the parameters' combination that minimize the error;

Rank-size: applications on earthquakes

- Sornette et al. (1996) where the rank-size approach is used to study transition of the earthquake magnitude distribution between small and large earthquakes.
- Jaumé (2000) shows the application of 'rank-ordering' distribution to further confirm conclusion on 'critical point estimation' in frequency-size behavior.
- Wu (2000) where the 'rank-ordering' approach is used to show the difference between broad-band radiated energy and the scalar seismic moment.
- Newman (2005) shows also the power-law distribution of Richter magnitude of earthquakes occurring in California between January 1910 and May 1992.

Italian seismic events 2016 - 2017

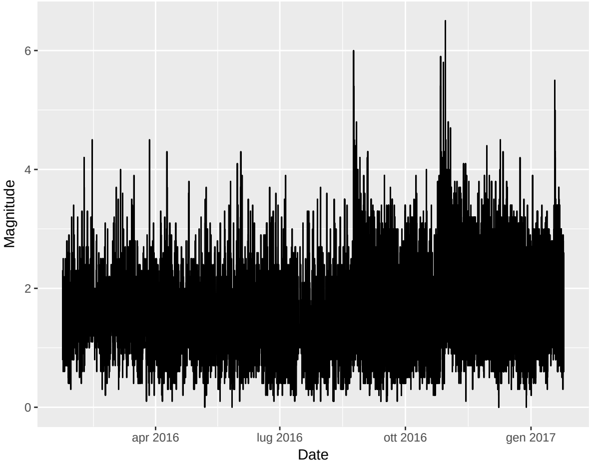

Figure: Time series of the Italian seismic events between 24^{th} January 2016 and 24^{th} January 2017, according to the INGV data. The number of registered shocks is 59190.

Dataset

- Source: Italian National Institute of Geophysics and Vulcanology (INGV);
- Timeframe: 24^{th} January 2016 - 24^{th} January 2017;
- Area: Italy;
- Richter magnitude range: [3.1-6.5];

Note: We include the cases of August, 24^{th} 2016 in Accumuli and October, 26^{th} and 30^{th} 2016 in Ussita and Norcia (maximum magnitude reached).

Dataset: catalog incompleteness

We adopt a cutoff at magnitude equal to 3.1 for 2 reasons:

1. The earthquakes with magnitude inferior to 3.1 has law chances of creating observable problems (United State Geological Survey);
2. To avoid the catalog incompleteness problem. See Marchetti et al. (2016); Chiaraluce et al. (2017)

Definition of aftershocks: smaller earthquakes that occur after a previous large earthquake, in the same area of the main shock.

Dataset: features

Stats	Val
N	978
Max.	6.50
Min.	3.10
μ	3.42
m	3.30
σ	0.39
Skewness	2.67
Kurtosis	14.36

Table: Statistical features of the Italian shocks dataset - no catalog incompleteness. Period: 24/01/2016 - 24/01/2017

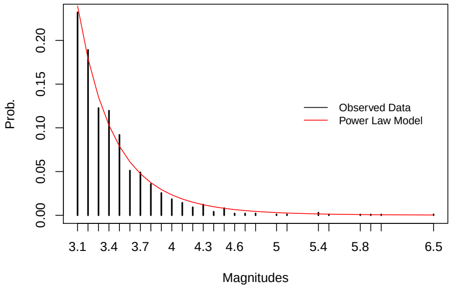

Figure: Probability of the earthquakes occurred during 24/01/2016 - 24/01/2017. Fit of a power law of the type $y = ax^b$. $\hat{a} = 7428.58$, $\hat{b} = -9.14$, R^2 of 0.99.

Our rank-size approach

We use the magnitude to sort the dataset. At $r = 1$ there is the highest registered magnitude.
We use the ZML (Eq. 1) and UL (Eq. 2) to explore the relation between the size-magnitude z and the respective rank.

To do so, the Levenberg-Marquardt Non-linear Least-Squares Algorithm is used, see Levenberg (1944); Lourakis (2005); Marquardt (1963).

Assumption: no cumulative effects from foreshocks

z is the magnitude of seismic event occurred in t and n foreshocks of $z_1, ..., z_n$ are occurred in $[t - \Delta t, t]$.

$$z \Rightarrow \bar{z} = \eta(n, z_1, ..., z_n, \Delta t) \times z$$

where $\eta(n, z_1, ..., z_n, \Delta t)$ is a parameter that increases with respect to $z_1, ..., z_n$ and n decreases with respect to the time window $[t - \Delta t, t]$, with Δt not smaller than 1.
We assume:

$$\eta(n, z_1, ..., z_n, \Delta t) = 1 \, \forall \, n, z_1, ..., z_n, \Delta t$$

It means no cumulative effects from foreshocks, namely each earthquake is treated as if it is unique and isolated. **We underestimate the cost.**

The Economic Cost Indicator

The economic cost aggregation is defined as $\Gamma_{\diamondsuit}$, $\diamondsuit = ZML, UL$.

$$\Gamma_{ZML} = \int_{\bar{z}}^{Z_{MAX}} C_{ZML}(z)dz = \int_{1}^{\bar{r}_{ZML}} H\left(\hat{\alpha}(r+\hat{\beta})^{-\hat{\gamma}}\right) dr, \tag{3}$$

$$\Gamma_{UL} = \int_{\bar{z}}^{Z_{MAX}} C_{UL}(z)dz = \int_{1}^{\bar{r}_{UL}} H\left(\hat{k}\frac{(N+1-r+\hat{\psi})^{\hat{\xi}}}{[N(r+\hat{\phi})]^{\hat{\lambda}}}\right) dr, \tag{4}$$

Where $Z_{MAX} = 6.5$.

The Economic Cost Indicator

(*i*)

$$H(z) = \begin{cases} \exp(z), & \forall z \in [\bar{z}, Z_{MAX}]; \\ 0, & \forall z \in [0, \bar{z}); \end{cases}$$

(*ii*)

$$H(z) = \begin{cases} z, & \forall z \in [\bar{z}, Z_{MAX}]; \\ 0, & \forall z \in [0, \bar{z}); \end{cases}$$

(*iii*)

$$H(z) = \begin{cases} \ln(z), & \forall z \in [\bar{z}, Z_{MAX}]; \\ 0, & \forall z \in [0, \bar{z}); \end{cases}$$

Rank-size fits: results

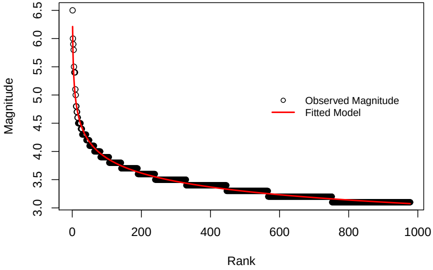

Figure: Seismic events with $m \geqslant 3.1$ registered in Italy between: 24/01/2016 and 24/01/2017. ZML model is reported. See eq. (1).

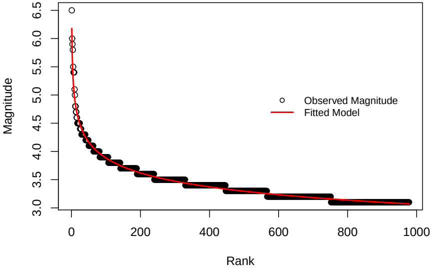

Figure: Seismic events with $m \geqslant 3.1$ registered in Italy between: 24/01/2016 and 24/01/2017. UL model is reported. See eq. (2).

Rank-size fits: results

Eq. (1)	**Calibrated parameter**	**Value**
	$\hat{\alpha}$	6.21
	$\hat{\beta}$	0.00
	$\hat{\gamma}$	0.10
	R^2	0.98

Eq. (2)	**Calibrated parameter**	**Value**
	$\hat{k}$	8.63
	$\hat{\phi}$	0.00
	$\hat{\lambda}$	0.10
	$\hat{\psi}$	6972.72
	$\hat{\xi}$	0.04
	R^2	0.98

Table: Estimations realized by the fit of the eqs. (1) and (2) for the Italian earthquakes catalog that covers the period: 24/01/2016 - 24/01/2017 (N $=$ 978, $m \geqslant 3.1$).

Cost Indicator

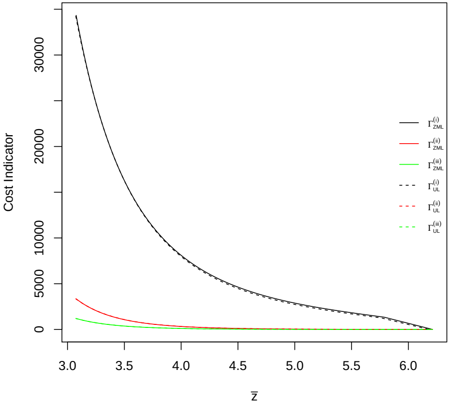

Figure: Comparison among eqs. (3) and (4) for the cases $(i), (ii), (iii)$ as $\bar{z}$ varies. Italian earthquakes catalog that covers the period: 24/01/2016 - 24/01/2017 (N $=$ 978, $m \geqslant 3.1$).

Robustness Check

To robustness check has been run through the analysis of 2 different dataset to explore respectively time and space effect:

- A wider set of seismic event detected between April 16^{th}, 2005 and March 31^{st}, 2017.
- A local catalog made by earthquakes occurred between January 24^{th}, 2016 and January 24^{th} 2017, with epicenter in Provinces of Macerata, Perugia, Rieti, Ascoli Piceno, L'Aquila, Teramo, Terni and Fermo (see Natale et al., 1988).

Robustness Check

Global Dataset

Stats	Value
N	13239
Max	6.50
Min	2.50
μ	2.88
m	2.80
σ	0.42
Skewness	1.89
Kurtosis	8.24

Table: Magnitudes not smaller than 2.5 of the earthquakes occurred from 16/04/2005 to 31/03/2017.

Local Dataset

Stats	Value
N	849
Max	6.50
Min	3.10
μ	3.42
m	3.30
σ	0.39
Skewness	2.75
Kurtosis	15.05

Table: Magnitudes of the earthquakes with epicenters in the provinces of Macerata, Perugia, Rieti, Ascoli Piceno, L'Aquila, Teramo, Terni and Fermo from 24/04/2016 to 24/01/2017.

Robustness Check – Fits

Eq. (1)	Estimation	Global data	Local Data	Original Data
	$\hat{\alpha}$	9.48	6.07	6.21
	$\hat{\beta}$	68.80	0.00	0.00
	$\hat{\gamma}$	0.14	0.10	0.10
	R^2	0.98	0.98	0.98

Eq. (2)	Estimation	Global data	Local Data	Original Data
	$\hat{k}$	0.88	9.50	8.63
	$\hat{\phi}$	9.52	0.00	0.00
	$\hat{\lambda}$	0.11	0.10	0.10
	$\hat{\psi}$	36951.95	6749.18	6972.72
	$\hat{\xi}$	0.30	0.02	0.04
	R^2	0.99	0.98	0.98

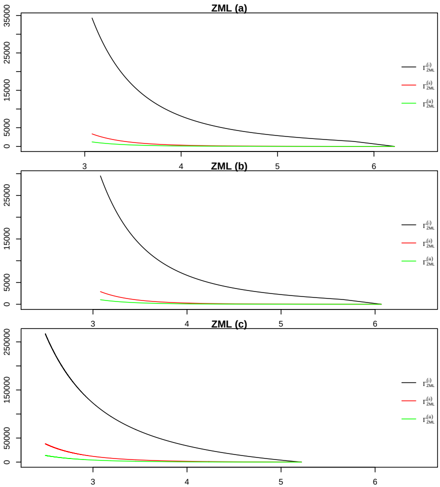

Figure: Eq. (3) for the cases $(i), (ii), (iii)$ as $\bar{z}$ varies.
(a) Original Data.
(b) Local Data.
(c) Global Data

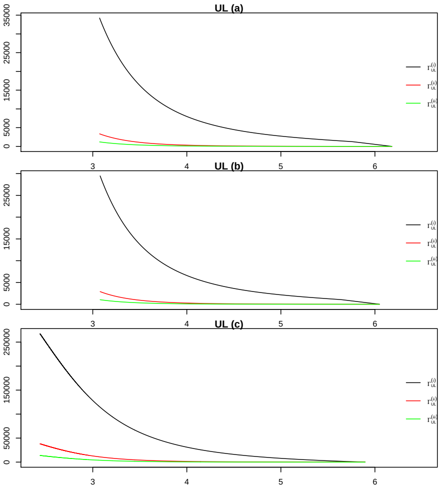

Figure: Eq. (4) for the cases $(i), (ii), (iii)$ as $\bar{z}$ varies.
(a) Original Data.
(b) Local Data.
(c) Global Data

Conclusions

- The UL fits the data slightly better than ZML. It manifests increased sensitivity to data variation.
- No space effect in both the cases (ZML and UL).
- The ZML has poorer fit at low ranks with the global dataset. In such a case the UL acts better thanks to ϕ and ψ.
- The most expensive case emerges with exponential transformation of the magnitude, $\Gamma^{(i)}$. This has a change in decay after a $\bar{z}$ of $\sim$3.7 and in concavity after $\sim$5.7.
- The reduction of the impact of the earthquake on infrastructure should be pursued by increasing the infrastructures' resistance to seismic events with the highest magnitudes.

Catalog's feature: Queen - Harem & King - Viceroys effect

Stats	Val
N	59191
Max.	6.50
Min.	0.1
μ	1.57
m	1.50
σ	0.54
Skewness	0.95
Kurtosis	5.19

Table: Statistical features of the Italian shocks dataset - with catalog incompleteness. Period: 24/01/2016 - 24/01/2017

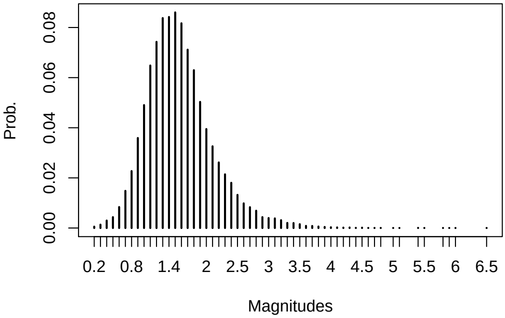

Figure: Probability of the earthquakes occurred during 24/01/2016 - 24/01/2017

Fits Comparison

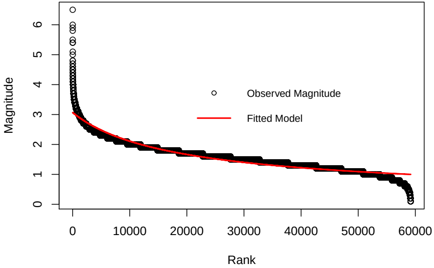

Figure: All the seismic events occurred in Italy during 24/01/2016 - 24/01/2017. They are sorted in decreasing order according to their magnitude and the ZML model is reported.

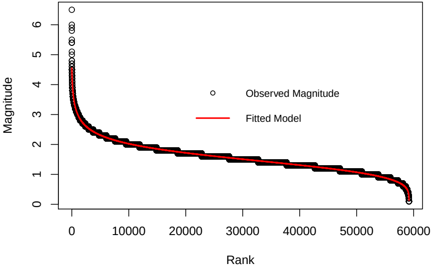

Figure: All the seismic events occurred in Italy during 24/01/2016 - 24/01/2017. They are sorted in decreasing order according to their magnitude and the UL model is reported.

Robustness Check – Fits

Eq. (1)	Estimation	Global Data	Local Data	Original Data	Complete Data
	$\hat{\alpha}$	9.48	6.07	6.21	1450.52
	$\hat{\beta}$	68.80	0.00	0.00	12879.57
	$\hat{\gamma}$	0.14	0.10	0.10	0.65
	R^2	0.98	0.98	0.98	0.93

Eq. (2)	Estimation	Global Data	Local Data	Original Data	Complete Data
	$\hat{k}$	0.88	9.50	8.63	4.85
	$\hat{\phi}$	9.52	0.00	0.00	88.48
	$\hat{\lambda}$	0.11	0.10	0.10	0.16
	$\hat{\psi}$	36951.95	6749.18	6972.72	0
	$\hat{\xi}$	0.30	0.02	0.04	0.22
	R^2	0.99	0.98	0.98	0.99

Thank You

References I

Ausloos, M. and R. Cerqueti
2016. A universal rank-size law. *PloS one*, 11(11):e0166011.

Bottazzi, G., D. Pirino, and F. Tamagni
2015. Zipf law and the firm size distribution: a critical discussion of popular estimators. *Journal of Evolutionary Economics*, 25(3):585–610.

Cerqueti, R. and M. Ausloos
2015. Evidence of economic regularities and disparities of Italian regions from aggregated tax income size data. *Physica A: Statistical Mechanics and its Applications*, 421:187–207.

References II

Chiaraluce, L., R. Di Stefano, E. Tinti, L. Scognamiglio, M. Michele, E. Casarotti, M. Cattaneo, P. De Gori, C. Chiarabba, G. Monachesi, et al.
2017. The 2016 central italy seismic sequence: A first look at the mainshocks, aftershocks, and source models. *Seismological Research Letters*, 88(3):757–771.

Fujita, M., P. R. Krugman, and A. J. Venables
2001. *The spatial economy: Cities, regions, and international trade*. MIT press.

Fujita, M. and J.-F. Thisse
2000. The formation of economic agglomerations: old problems and new perspectives. *Economics of cities: Theoretical perspectives*, Pp. 3–73.

References III

Hill, B. M.
1974. The rank-frequency form of zipf's law. *Journal of the American Statistical Association*, 69(348):1017–1026.

Ioannides, Y. and S. Skouras
2013. Us city size distribution: Robustly pareto, but only in the tail. *Journal of Urban Economics*, 73(1):18–29.

Jaumé, S. C.
2000. Changes in earthquake size-frequency distributions underlying accelerating seismic moment/energy release. *Geocomplexity and the Physics of Earthquakes*, Pp. 199–210.

Levenberg, K.
1944. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168.

References IV

Li, W. and Y. Yang
2002. Zipf's law in importance of genes for cancer classification using microarray data. *Journal of Theoretical Biology*, 219(4):539–551.

Lourakis, M. I.
2005. A brief description of the Levenberg-Marquardt algorithm implemented by levmar. *Foundation of Research and Technology*, 4(1).

Maillart, T., D. Sornette, S. Spaeth, and G. Von Krogh
2008. Empirical tests of ZipfâĂŹs law mechanism in open source Linux distribution. *Physical Review Letters*, 101(21):218701.

Mandelbrot, B.
1953. An informational theory of the statistical structure of language. *Communication theory*, 84:486–502.

References V

Mandelbrot, B.
1961. On the theory of word frequencies and on related markovian models of discourse. *Structure of language and its mathematical aspects*, 12:190–219.

Marchetti, A., M. G. Ciaccio, A. Nardi, A. Bono, F. M. Mele, L. Margheriti, A. Rossi, P. Battelli, C. Melorio, B. Castello, et al.
2016. The italian seismic bulletin: strategies, revised pickings and locations of the central italy seismic sequence. *Annals of Geophysics*, 59.

Marquardt, D. W.
1963. An algorithm for least-squares estimation of non-linear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441.

References VI

Matlaba, V. J., M. J. Holmes, P. McCann, and J. Poot
2013. A century of the evolution of the urban system in brazil. *Review of Urban & Regional Development Studies*, 25(3):129–151.

Natale, G., F. Musmeci, and A. Zollo
1988. A linear intensity model to investigate the causal relation between calabrian and north-aegean earthquake sequences. *Geophysical Journal International*, 95(2):285–293.

Newman, M. E.
2005. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351.

Piantadosi, S. T.
2014. ZipfâĂŹs word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5):1112–1130.

References VII

Simon, H. A.
1955. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440.

Sornette, D., L. Knopoff, Y. Kagan, and C. Vanneste
1996. Rank-ordering statistics of extreme events: Application to the distribution of large earthquakes. *Journal of Geophysical Research: Solid Earth*, 101(B6):13883–13893.

Wu, Z.
2000. Frequency–size distribution of global seismicity seen from broad-band radiated energy. *Geophysical Journal International*, 142(1):59–66.

Zipf, G. K.
1935. The psycho-biology of language.

References VIII

Zipf, G. K.
1949. Human behaviour and the principle of least-effort. Cambridge MA edn.