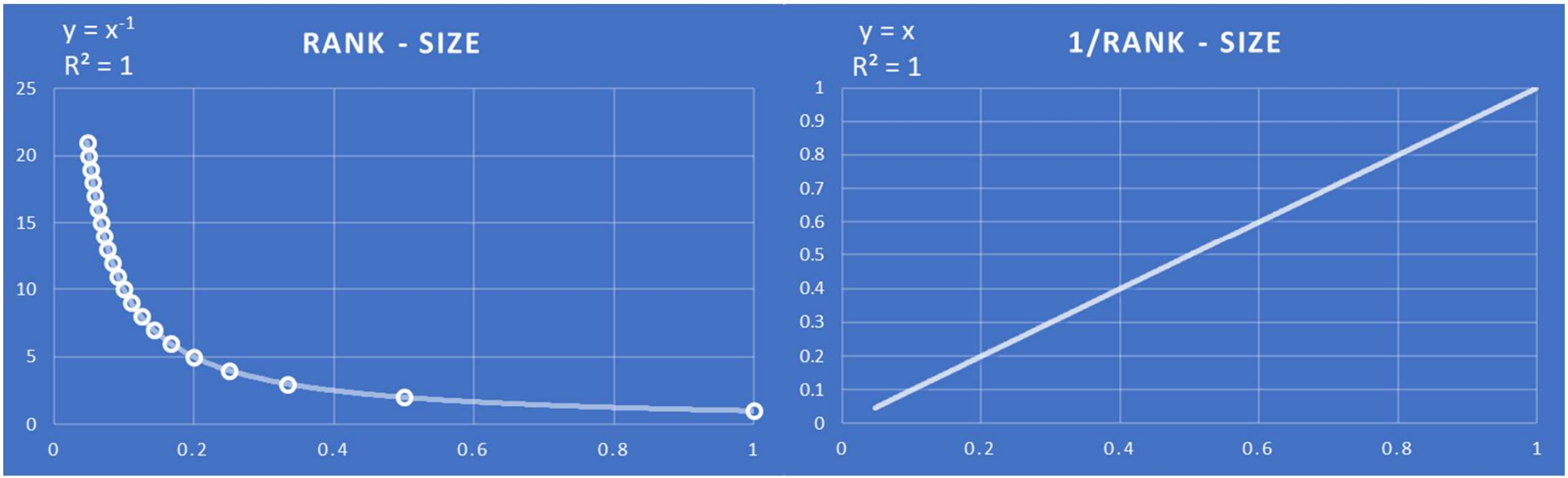
Stability of Zipf's law for cities and occult spatial dependence effects: A study on the OECD countries

Rolf Bergs[a] & Uwe Neumann[b]

a: PRAC – Bergs & Issa Partnership Co. b: RWI – Leibniz Institute for Economic Research

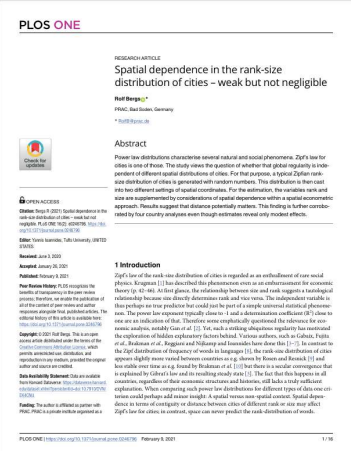
GfR 2025 meeting, NIERS, Krefeld, 26-27 June 2025

Zipf's law for cities = Pareto distribution with α≈1



Starting point: Zipf's law – an enthrallment?

- Spooky: hard to explain by economic theory (Paul Krugman)
- A second – hidden - level of spatial heterogeneity (Bin Jiang)
- A hybrid Pareto+lognormal distribution function (e.g. Skouras S & Ioannides Y 2013)
- But: a tedious statistical tautology without further relevance (Gan L et al. 2006)


PLOS ONE

Spatial dependence in the rank-size distribution of cities – weak but not negligible

Abstract

1 Introduction

Bergs R (2021) Spatial dependence in the rank-size distribution of cities – weak but not negligible. Plos One 16(2): e0246796

2021 study: a tedious tautology or is there more?

Hypothesis: a universal statistical tautology necessitates absence of spatial disturbance in the Pareto regression ($a \approx -1$, free of spatial dependence)

$Ln(R-0.5) = ln(C)-aln(S)+\varepsilon$ (non-spatial)

$Ln(R-0.5) = \rho W ln(R)+ln(C)-aln(S)+\varepsilon$ (SAR)

$Ln(R-0.5) = ln(C)-aln(S)+\upsilon$ (SEM)
$\upsilon = \lambda W\upsilon+\varepsilon$

Simulation: upper tail Pareto; lower tail lognormal

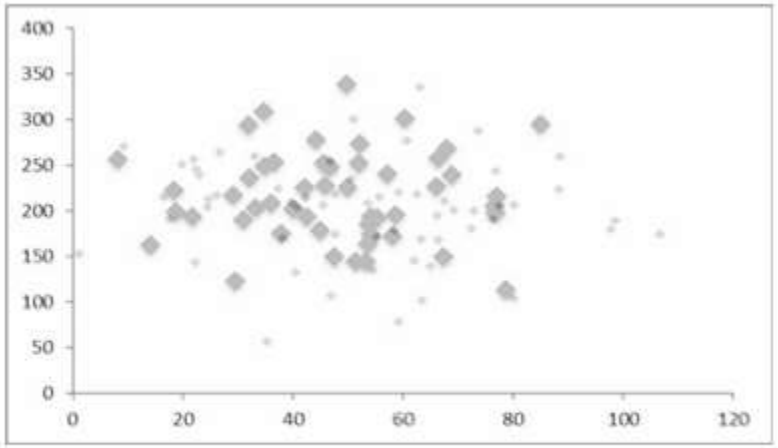


Fig 1. Normal spatial distribution of cities (Pareto and lognormal tails).

https://doi.org/10.1371/journal.pone.0246796.g001

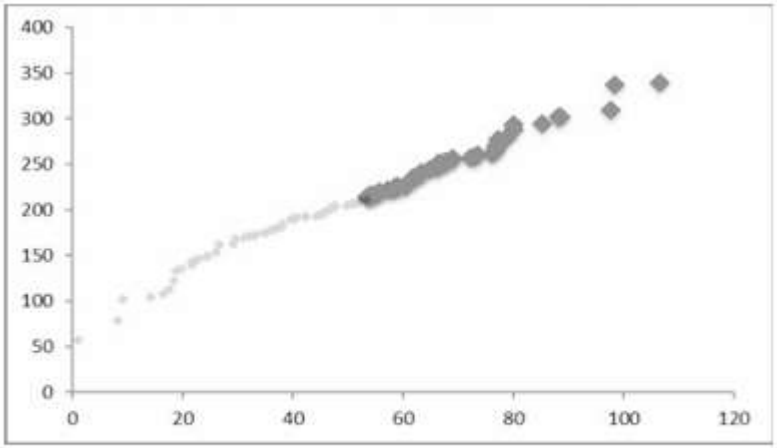


Fig 2. Spatial distribution of cities with ranked coordinates (Pareto and lognormal tails).

https://doi.org/10.1371/journal.pone.0246796.g002

a= -0.403 (full range)
a= -0.995 (Pareto section)

No spatial dependence effect

a= -0.322 (full range)
a= -0.862 (Pareto section)

Significant spatial dependence effect

2021 real world results for *a* (Pareto tail)

Based on population data:

- USA: -1.005 versus -1.004 (λ=-0.781 ***)
- Germany: -0.939 versus -0.948 (λ=-3.520 ***)
- UK: -1.056 versus -1.059 (λ and ρ are both insignificant)

Estimation based on night light segmentation (VIIRS):

- Slovenia: -0.983 versus -0.860 (ρ=-0.549*)

Result: In some countries, spatial dependence effects on Zipf's law are detectable. They are weak but not negligible.

Spatial Zipf study: OECD with 2024 data and a more precise truncation

- Meanwhile, recent studies confirm the spatial relevance in Zipf's law for cities (Griffiths 2022, Xiao & Gong 2022)
- We were curious to understand the spatially augmented Pareto model by applying it to a larger group of countries with global coverage and a larger variation in economic development

- Data set: NGIA, US Geological Survey, US Census Bureau, and NASA (2024) World Cities Database. Link: https://simplemaps.com/data/world-cities

Truncation to better isolate the Pareto tails

- Log-transformation of S
- Rank of 50 percent of the population (Malevergne 2011); if n is too small we perform a stepwise increase of ranks
- Visual inspection and Shapiro-Wilk tests of the log-transformed population data to differentiate between Pareto and lognormal tails
- Back-up of the tests with Kullback-Leibler divergence: do observations in the respective section fit a Pareto or rather a lognormal distribution?

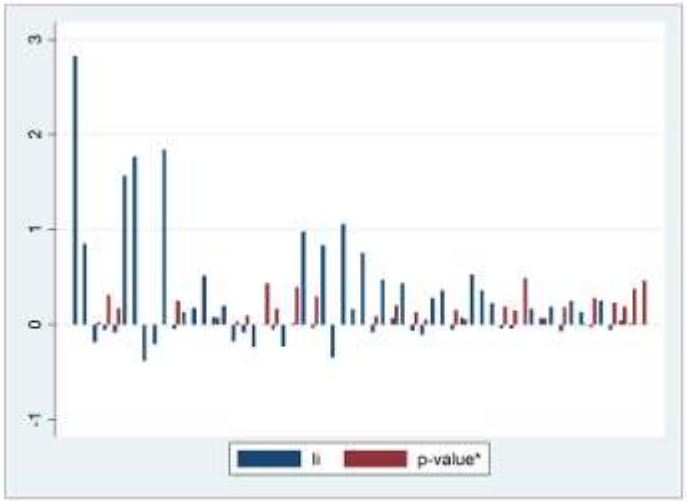
$$D_{KL}(p \parallel q) = \sum_{x \in X} p(x_i) log \frac{p(x_i)}{q(x_i)}$$

2024 Significant results for OECD countries

	Non-spatial coefficient	Spatial coefficient
Belgium	-1.293***	-1.343*** (SEM)
France	-1.082***	-1.072*** (SEM)
Germany	-1.110***	-1.094** (SAR)
Italy	-1.369***	-1.374** (SAR)
Sweden	-0.729***	-0.685** (SAR)
USA	-1.076***	-1.078* (SEM)

Significance: ***=0.01 level; **=0.05 level; *=0.1 level

Belgium: LISA




```
. swilk LNArea

                    Shapiro-Wilk W test for normal data

    Variable |   Obs        W           V          z       Prob>z
-------------+--------------------------------------------------
      LNArea |    57     0.75845     12.603      5.445     0.00000

.
. reldist divergence Area logn

Relative distribution divergence                 Number of obs   =        57
  F1: Area                                       Histogram bins  =        20
  F0: logn

------------------------------------------------------------------------------
        Area |      Coef.   Std. Err.      [95% Conf. Interval]
-------------+----------------------------------------------------------------
     entropy |    .902424   .1842042       .5334188    1.271429
------------------------------------------------------------------------------

. reldist divergence Area Pareto

Relative distribution divergence                 Number of obs   =        57
  F1: Area                                       Histogram bins  =        20
  F0: Pareto

------------------------------------------------------------------------------
        Area |      Coef.   Std. Err.      [95% Conf. Interval]
-------------+----------------------------------------------------------------
     entropy |   .3085736   .1038821       .1004728    .5166745
------------------------------------------------------------------------------
```

Figure 1: Belgium: Local Moran I coefficients (LISA) and related p-values for the upper 57 observations

Conclusion

- Among the OECD countries there is a larger variation of spatial dependence in Zipf's law for cities, but basically it appears modest.
- Smaller countries seem less affected than medium and larger countries
- Significant influence of λ or ρ on a can be confirmed by inspecting local spatial autocorrelation (LISA coefficients): The mean of the LISA coefficients is then significantly different from zero

However, …

- any negative LISA values originate from the close spatial assembly of big and smaller cities exclusively within the Pareto tail. The truncated lognormal tail of small municipalities is not covered. Hence, the specific function of minor settlements is disregarded.
- In other words: There is a dilemma between the necessary isolation of the Pareto tail and the essential neglect of the large number of small cities in the spatial weight matrix. It is thus imaginable that negative spatial autocorrelation in the Pareto tail could be also over-emphasised.

Link to arXiv

https://doi.org/10.48550/arXiv.2503.22463