





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

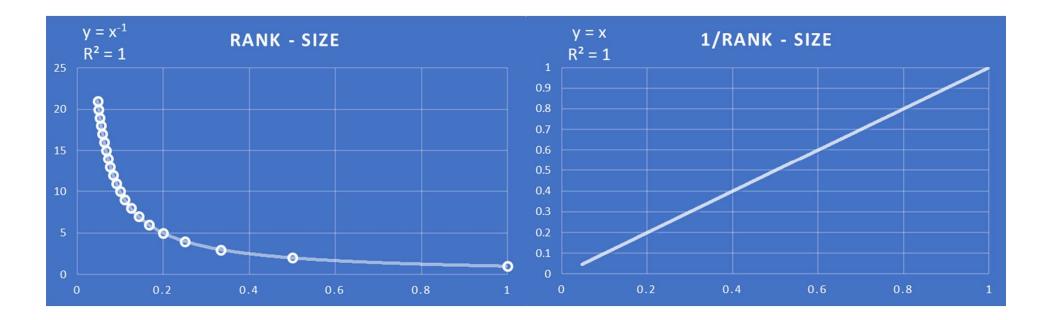
### Rolf Bergs<sup>a</sup> & Uwe Neumann<sup>b</sup>

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 $\alpha \approx 1$





#### 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		
	RESEARCHARTCLE Spatial dependence in the rank-size	
	distribution of cities - weak but not negligible	
	Rolf Bergs ()*	
	PRAC, Bad Stoder, Germany	
	* Bottli #proc.de	
	Abstract	
Check for updates	Power law destructions characterise several natural and social phenomene. Zipf's law for obes is one of hour. The study views the question of whether that global regularity is inde- pendent of others point distributions of other. For the propose, a hydraic Zipfann mini- size distribution of others is generated with random numbers. This distribution is then cast the host different strating of basis is contrained. The similar test and the host different strating of basis is contrained. The similar test and the host different strating of basis is contrained. The similar test and the host different test and test and test and the similar test and the host different test.	
GOPEN ACCESS	size are supplemented by considerations of spatial dependence within a spatial econometric	
Gitation: Deeps R (2021) Spatial dependence in the stank-size databation of cline - weak but not regisples. PLoS CRIE: 16(2): 452(47)(8): Millsr (Ha), cop 11: (2115)(sur to Lanes COVE726)	approach. Results suggest that distance potentially matters. This finding is further corrobo- rated by four country analyses even though estimates reveal only modest effects.	
Editor: Yunnis Islamides, Tufts University, UNITED STATES		
Received June 3, 2020		
Accepted: January 26, 2021	1 Introduction	
Published: Nonury 3, 2021	Zipf's law of the rank-size distribution of cities is regarded as an extinuliment of rare social	
Peer Texteen Hology: PLES recognises the benefits of transparency in the peer makes process; theretains, we enable the publication of all of the center of peer review and author responses strongiste that published address. The editorial hology of this addres a service horse. https://doi.org/10.1071/journal.pore.2000/36	Physics, Koupman 11] has does therefore this phenometers or sets in an embarrament for economic theory (p. 6: 46: 46), thin figures, the relationship betteress in size and suggests trainingiant relationship because size disordy determinant rank and sets versus. The independent variable is in the predupts in travel predictor but could gate bug in of a simple size and any size of the pre- ton. The properties between the predictor but could be prior of a simple size and any size of the prior to be empower trajectory that the could be prior of a simple size of the prior to be empower the prior to be any size of the prior to be prior to be prior to be prior to be any size of the prior to be any size of the prior to be	
Depyright: 0.0001 Full Tarrys, This is an open access active disbibilited under the terrors of the Constant Common Ambridge Constant with permits amendated uses distributions and improductions in an endown, provided the ongost author and source are credited.	the exploration of bidden explorativey factors below. Various authors, inch as Gobiat, Fujita, et al., Brokinna et al., Registeri and Nijkamy and Insweidels have done this $[1-v]_1$ . In contrast to the Zprf distribution of frequency of words in language [1], the rank-one distributions of etilities appears algibility more varied between constitution as $g_i$ shows by Rosen and Researk [1] and loss studie event runs $w_i$ , found by the linkame et al. [10] but there is a wordshow correspond to the	
Data Availability Statement Cuta are available from Hansed Datawaya Style - Statement Januard, eductorizati, eductorizati (Statement Januard, Diskowa	is explained by Ghraf's law and its resulting steady state [3]. The fact that this happens in all countries, regardless of their economic structures and histories, will lack a turly sufficient explanation. When comparing such power law distributions for different types of data score cri- terion avaid perhaps add misser insight. A spatial reverse nan-spatial context, spatial depen-	
Funding: The author is affiliated as partner with PRAC PRAC is a pointer include argument as a	dence in terms of contiguity or distance between cities of different rank or size may affect Zipf's law for cities; in contrast, space can more predict the rank-distribution of words.	

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



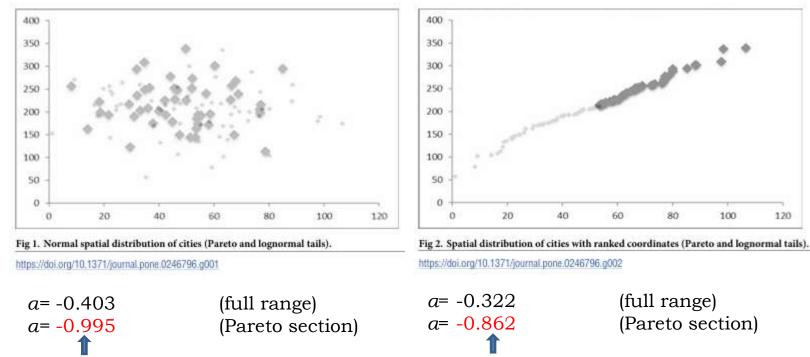
2021 study: a tedious tautology or is there more?

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)  $Ln(R-0.5) = \rho Wln(R)+ln(C)-aln(S)+\varepsilon$  (SAR) Ln(R-0.5) = ln(C)-aln(S)+v (SEM)  $v = \lambda Wv + \varepsilon$ 

#### PRAC PLANUNG & FORSCHUNG POLICY RESEARCH & CONSULTANCY

#### Simulation: upper tail Pareto; lower tail lognormal



No spatial dependence effect

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 ( $\lambda$  and  $\rho$  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?

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

8



### 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

### PLANUNG & FORSCHUNG POLICY RESEARCH & CONSULTANCY

#### Belgium: LISA

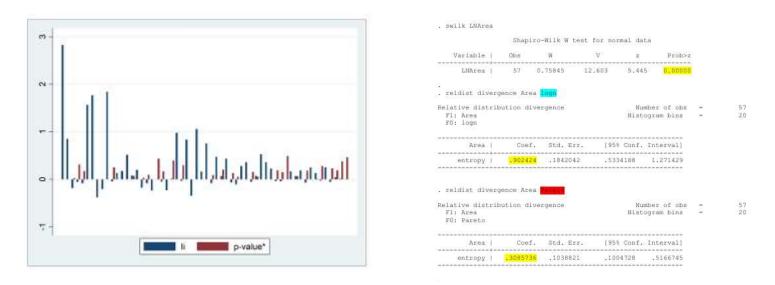


Figure 1: Belgium: Local Moran I coefficients (LISA) and related p-values for the upper 57 observations

10



#### 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

