SIZE DISTRIBUTION OF CITIES: EVIDENCE FROM A LAB

Rocco Rante[1] Federico Trionfetti[2] Priyam Verma[2]

EEA-ESEM 2023, Barcelona

August 28, 2023

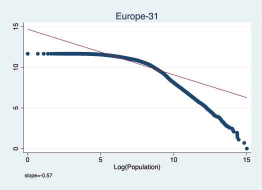
[1] The Louvre Museum, France.
[2] Aix Marseille Univ, CNRS, AMSE.

Introduction

- Many factors determine population size differences across cities
  - Amenity differential (eg. Next to a beach).
  - Productivity differential (eg. Silicon valley for tech).
  - Spatial connectedness.
- Aggregation of these factors determines city size distribution.
  - For instance, agglomeration can result in fatter tails (NY pop is 8 mil).
  - Factors hard to measure and often time-varying.
- What is the distribution when most factors are almost uniform?
  - Use 9^{th} century archaeological data from the oasis of Bukhara.
  - Examine the role of geography in city size distribution.

What is the distribution today?

- Standard tests of log-normality and pareto don't fit the data.
  - Zipf's law: Linear slope of (log) population rank and size is -1.
  - If data is Pareto with shape parameter 1, Zipf's law should hold
  - For cities in most of Europe and US, this slope is $>$ -1.



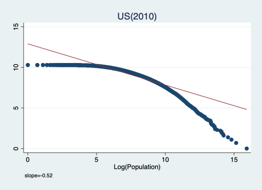


Table 1, Table 2

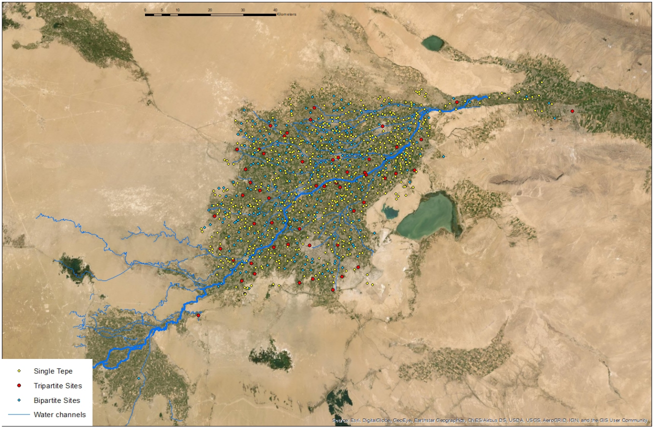
In this paper

- Using 9^{th} century archaeological data from Bukhara (Uzbekistan)
  - Construct cities using statistical methods.
  - Which distribution best describes the data? Any stylized facts?
- Write a discrete choice model of locations
  - Estimate three key parameters while fitting the data to our model
  - Show how distribution changes with the three shocks:
    - Amenity, Agglomeration, and Geography.
- Re-estimate the parameters with 21^{st} century data from Uzbekistan
  - See how the distribution changed?
  - Which of three contributed more?

The Oasis of Bukhara

- Extends over delta of the Zerafshan in southeastern Uzbekistan.
  - Irrigates a surface of land whose area measures about 5,100 sq km.
  - Flat surface (Maximum difference in altitude is 200 meters)
  - Outside the delta lays a desert made of clay giving natural boundary.
- Consists of 618 sites:
  - Manufacturing cities (53)
  - Agricultural cities (284)
  - Hamlets (266)
  - Forts (15)

Oasis of Bukhara



Urban System: Suitable Aggregation

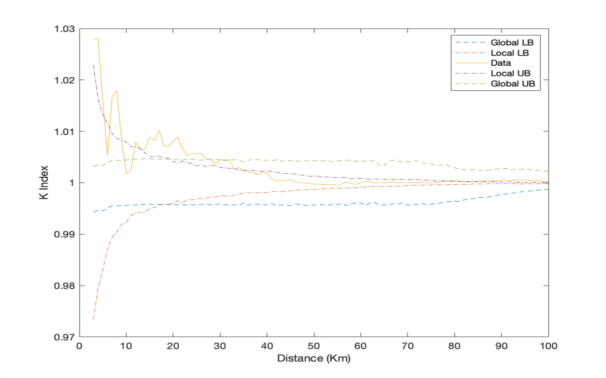
- Non-manufacturing cities surround manufacturing cities.
  - We count them together as one urban unit.
- Test if this is a valid assumption by calculating

$$K(r) = \left(\frac{\sum_{i \in M} v_i(r)}{\sum_{i \in M} n_i(r)} \right) \left(\frac{V}{N-1} \right)^{-1} \tag{1}$$

  where

  - $v_i(r)$: # of non-manufacturing cities around manu. city i within radius r.
  - $n_i(r)$: # of neighboring cities around manu. city i within radius r.
  - V: Total non-manufacturing cities
  - N: Total cities (man. + non-man.)
- Also calculate 95% CI based on simulated city distribution.

Co-concentration Index



- Non-Man. cities concentrate around Manu. cities at 7 kilometers.

Log-normality vs Pareto

A: Skwenes-Kurtosis and Shapiro-Wilkinson tests for log-normality

		Sk-Kurt			SW
Measure of ‘size’	Type of unit	Sk.	Kurt	Joint	
Population	Urban System	0.6428	0.8190	0.8748	0.5597

- The urban systems pass the test of normality.

B: OLS Regression of log rank on log size

Measure of ‘size’	Type of unit	Slope	s.e.	№ of obs.
Population	Urban Systems	-1.25	.071	53

- Don’t conform to Zipf’s law and hence not Pareto with parameter 1.

Rank-Size Relation

C: Rank-size slope by quartiles

	Q1	Q2	Q3	Q4
Urban System	-0.27	-0.74	-1.69	-2.04
	(.02)	(0.03)	(0.12)	(0.13)

- Falling slope show presence of concavity in rank size relationship.
  - For smaller cities, concentration forces are active.
  - For larger cities, dispersion forces are active.

Model

- **Location**: Utility derived from being located in i:

$$U_i = \mathbb{E}[u_i] + \epsilon_i \quad \text{Amenity shocks} \tag{2}$$

  - Here location shocks ϵ_i is $i.i.d. \sim \text{G}(0, \mu_L)$.
- **Consumption**: Utility from consumption is

$$c_i = \alpha \bar{l} + \xi_i \quad \text{Agglomeration} \tag{3}$$

  - $\bar{l}$ is time endowment, α of which spent of production.
  - Preference for traders in market i ξ is $i.i.d. \sim G(L_i, \mu_c)$
- **Travel across cities**: Utility from trip b/w i and j is

$$t_{ij} = \tau_{ij} + \zeta_j \tag{4}$$

  - Preference for travel ζ is $i.i.d. \sim \text{G}(0, \mu_z)$.
  - Remainder of time net of travel cost: $\tau_{ij} = (1 - \alpha)\bar{l} - 2\delta_{ij}/s$.

Spatial Equilibrium

- The expected indirect utility from location i is

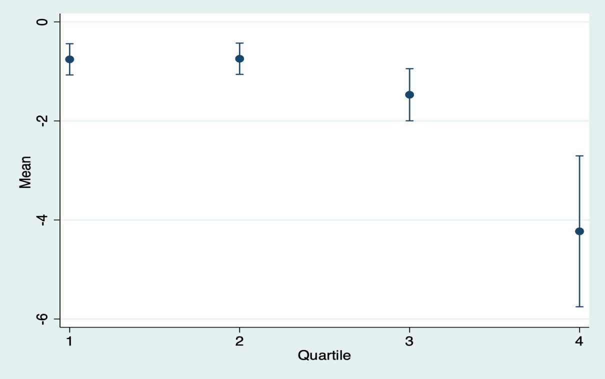
$$\mathbb{E}\left[u_i\right] = \bar{l} + \underbrace{\mu_C \ln L_i^m}_{\text{Market Size}} + \underbrace{\mu_Z \ln\left(\sum_{j=1}^{N} \exp\left(-\delta_{ij}/\mu_Z\right)\right)}_{\text{centrality}} \tag{5}$$

- In equilibrium, population share of location i is

$$\ln \lambda_i^* = \underbrace{\frac{\mu_C}{\mu_L - \mu_C} \ln s_i}_{\text{Silk Dummy}} + \underbrace{\frac{\mu_Z}{\mu_L - \mu_C} \ln \sum_{j=1}^{N} e^{\frac{-\delta_{ij}}{\mu_Z}} - \ln \sum_{k=1}^{N} s_i \lambda_i^* \left(\sum_{j=1}^{N} e^{\frac{-\delta_{kj}}{\mu_Z}}\right)^{\frac{\mu_Z}{\mu_L}}}_{\text{Relative centrality}} \tag{6}$$

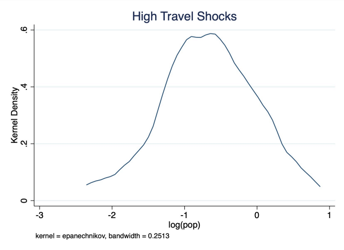
- Geographical advantage is a key determinant of population.

Testing the Model: Rank Size Relation

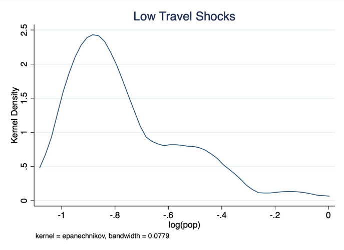


- Model replicates the rank size concavity as seen in the data.

Low Travel Shocks → Fatter Tails + Smaller Size



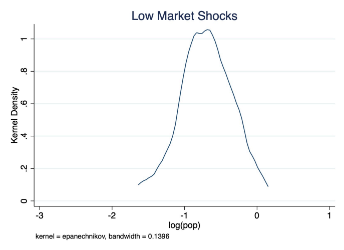
(a) Baseline



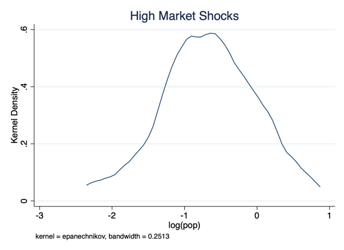
(b) Low μ_Z

- The importance of geography increases as travel shocks fall.
- Makes the distribution more skewed towards central places.

Higher Market Shocks → Increase Range and Size



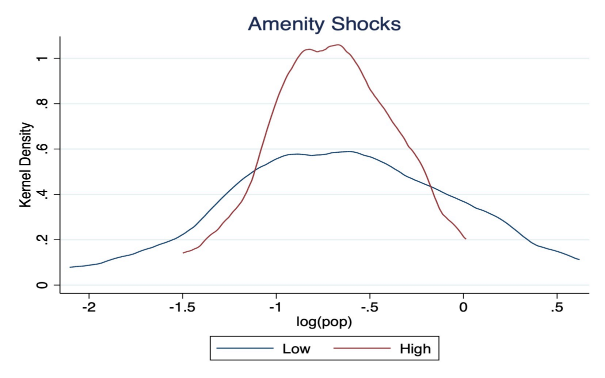
(a) Baseline



(b) High μ_c

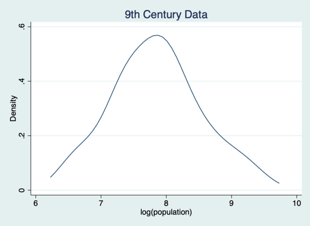
- Agglomeration Forces: Higher market-size shocks make
  - Big cities bigger.
  - Small cities smaller.

Higher Amenity Shocks → Reduces Range

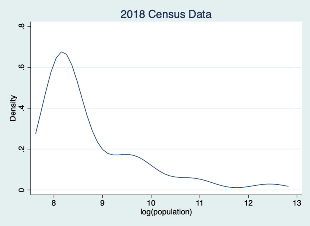


- Higher amenity shocks reduce
  - Importance of geography
  - Population range

Data and Estimates: City Size Distribution



(a) Historical



(b) Current

	Market size (μ_c)	Amenity (μ_l)	Travel (μ_z)
Historical	0.18	1.74	3.3
Current	3.15	3.28	1.63

- Agglomeration forces much more important today
- Geography explains most of the skewness in the data.

Conclusion

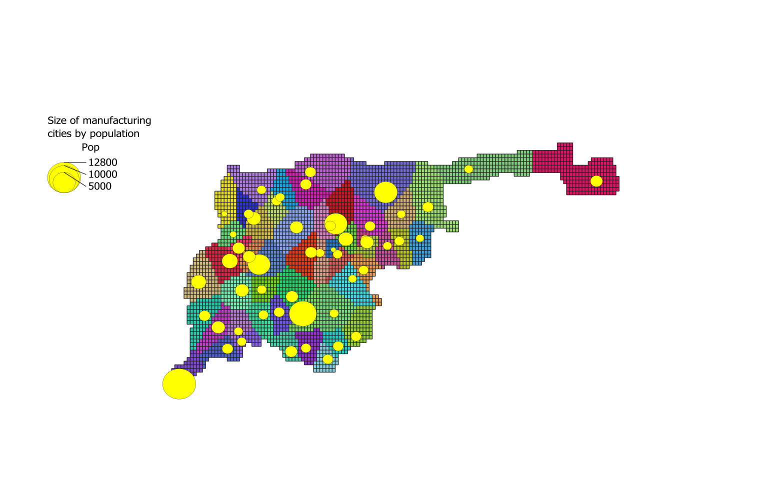
- New evidence on city size distribution using archaeological data.
- Log-normality is a good fit under homogeneous conditions.
  - Rank size relation is concave and hence Zipf's law doesn't hold.
- As agglomeration forces increase, distribution shifts to right.
- Geographic centrality important for explaining fat tails.

Archeological Survey

Theory

- Models of city size has focused on stochastic TFP and amenities.
  - But geography plays no role in determining city size.
  - Some spatial models with implications for city size.
    - Caliendo, Dvorkin, and Parro (2019), Allen and Arkolakis (2014), Redding and Sturm (2008).
- Can size distribution in Bukhara be explained by a model of
  - traveling people across cities where
  - some cities are more central than others?

- With 618 sites, create 53 clusters around manufacturing sites.



Skewness-Kurtosis test for log-normality

Country	Sk.	Kurt.	Joint	Country	Sk.	Kurt.	Joint
Austria	.0000	.0000	.0000	Liecht.	.0607	.3459	.1034*
Belgium	.3323	.0000	.0000	Lithuania	.0000	.0006	.0000
Bulgaria	.0000	.0020	.0000	Luxembourg	.0004	.0125	.0005
Croatia	.0000	.0000	.0000	Malta	.0408	.3460	.0820*
Cyprus	.0191	.8962	.0632*	Norway	.0000	.0617	.0001
Czech Rep.	.0000	.0000	.0000	Poland	.0000	.0000	.0000
Denmark	.1059	.0001	.0009	Portugal	.0000	.1412	.0000
Estonia	.0000	.0000	.0000	Romania	.0000	.0000	.0000
Finland	.0036	.0082	.0012	Slovakia	.0000	.0000	.0000
France	.0000	.0000	.0000	Slovenia	.1550	.0374	.0465
Germany	.0000	.0002	.0000	Spain	.0000	.1962	.0000
Greece	.0023	.3875	.0101	Sweden	.0000	.0313	.0001
Hungary	.0000	.0000	.0000	Switzerland	.2033	.0349	.0488
Iceland	.1448	.5410	.2739*	UK	.0008	.7839	.0034
Ireland	.0000	.4217	.0000	Europe-31	.0000	.0000	.0000
Italy	.0000	.0000	.0000	US (2010)	.0000	.1011	.0000
Latvia	.0000	.0014	.0000	US (2000)†	.019	.000	.0000

- Except for 4 out of 34 countries, tests rejects normality.

back

Zipf regressions for all cities

$$\ln(rank) = c_0 + c_1 \ln(Population) + \epsilon$$

Country	c_1	s.e.	№ cities	Country	c_1	s.e.	№ cities
Austria	-1.019	.0078	2357	**Liecht.**	-.7125	.1889	11
Belgium	-1.043	.0179	589	Lithuania	-.9100	.0092	540
Bulgaria	-.8571	.0161	264	**Luxemb.**	-.9802	.0221	116
Croatia	-.9730	.0101	556	Malta	-.8365	.0623	68
Cyprus	-.4799	.0090	395	Norway	-.7928	.0123	430
Czech Rep.	-.7871	.0027	6251	Poland	-1.195	.0038	2479
Denmark	-.9678	.0572	99	Portugal	-.7043	.0041	4260
Estonia	-.8336	.0168	226	Romania	-1.018	.0083	3181
Finland	-.7565	.0143	336	Slovakia	-0.782	.0005	2926
France	-.7047	.0011	36700	Slovenia	-.8684	.0247	192
Germany	-.6128	.0022	11355	Spain	-.5172	.0019	7517
Greece	-.7206	.0237	316	**Sweden**	-.9792	.0162	290
Hungary	-.6855	.0046	3154	Switzerland	-.6826	.0061	2515
Iceland	-.5331	.0197	76	UK	-1.315	.0065	10292
Ireland	-.9297	.0049	3409	Europe-31	-.5742	.0007	109111
Italy	-.6795	.0031	8092	US (2010)	-.5081	.0012	29494
Latvia	-.9196	.0216	119	US (2000)	-.5258	0.0014	25358

We find evidence of Zipf's conformity in only 4 out of 34 countries.

- Truncate distribution tail to match upper tail c_1 to -1.
- Fit upper tail rank data to population size and its square.

$$log(Rank_i) = \alpha + \beta_1 \log(Population_i) + \beta_2 \log(Population_i)^2 + \epsilon_i$$

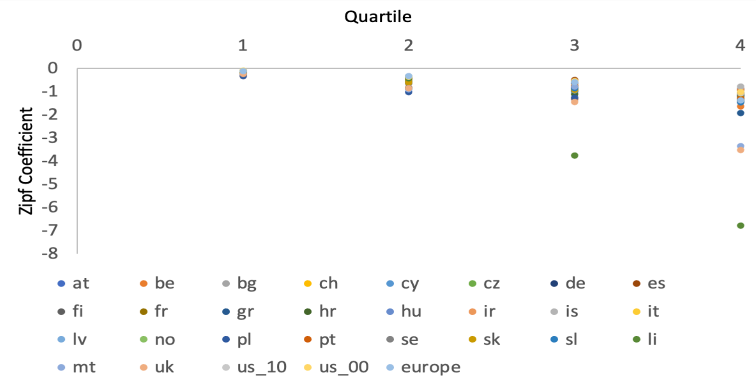
	Upper Tail		Quadratic regr. within upper tail				Upper tail's
Country	c_1	Std	B1	Std	B2	Std	Quadratic
Bulgaria	-1.004	0.01	0.43	0.09	-0.07	0.004	Conc
Croatia	-1.003	0.008	0.37	0.06	-0.07	0.003	Conc
Cyprus	-1.004	0.03	2.7	0.52	-0.19	0.02	Conc
Cz.Rep.	-0.999	0.001	-0.52	0.006	-0.031	0.0004	Conc
Estonia	-0.994	0.01	-1.16	0.101	0.009	0.005	Lin
Finland	-1.004	0.007	0.56	0.05	-0.07	0.002	Conc
France	-1.0006	0.0008	0.04	0.003	-0.064	0.0001	Conc
Germany	-0.9995	0.002	0.64	0.01	-0.08	0.0006	Conc
Greece	-1	0.02	4.95	0.14	-0.29	0.007	Conc
Hungary	-0.999	0.002	-0.18	0.016	-0.04	0.0009	Conc
Irland	-1.006	0.017	-1.68	0.16	0.03	0.008	Conv
Italy	-0.999	0.002	1.19	0.01	-0.12	0.0006	Conc
Latvia	-1.008	0.02	-0.84	0.24	-0.008	0.01	Lin
Liecht.	-1.03	0.25	17.25	6.25	-1.16	0.39	Conc
Lithuania	-1.001	0.007	0.58	0.06	-0.09	0.003	Conc
Malta	-1.04	0.06	7.88	0.65	-0.52	0.03	Conc

Concavity in Upper Tail

	Upper Tail		Quadratic regr. within upper tail				Upper tail's
Country	c_1	Std	B1	Std	B2	Std	Quadratic
Norway	-1.004	0.01	0.94	0.07	-0.1	0.003	Conc
Portugal	-1.001	0.004	1.84	0.02	-0.17	0.001	Conc
Slovakia	-0.999	0.003	-0.1	0.021	-0.05	0.001	Conc
Slovenia	-1.009	0.02	1.71	0.16	-0.14	0.009	Conc
Spain	-1.003	0.004	1.07	0.02	-0.103	0.001	Conc
Sweden	-1.007	0.015	2.3	0.11	-0.16	0.005	Conc
Switzerland	-1.003	0.005	1.43	0.025	-0.14	0.001	Conc
US(2000)	-1.001	0.001	1.02	0.007	-0.1	0.0003	Conc
US(2011)	-1.0003	0.001	1.11	0.006	-0.1	0.0003	Conc

- We can get an upper tail Zipf in 24/34 countries.
- Find concavity in upper tails for 21 countries.

Rank Size Relation across city sizes



- Zipf coefficient falls with increasing city size.
    - In line with Duranton (2007) predictions.

Simulated Clusters within Oasis