

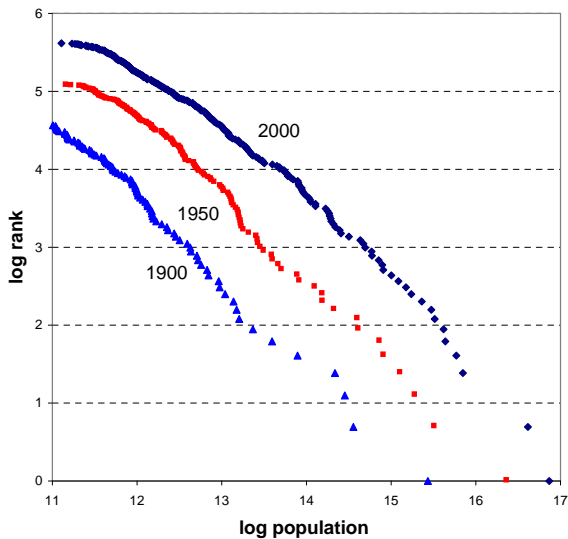
Lecture 8: The Size Distribution of Cities

WWS 582a

Esteban Rossi-Hansberg

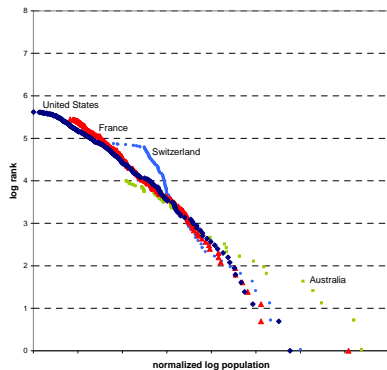
Princeton University

Zipf's Law US

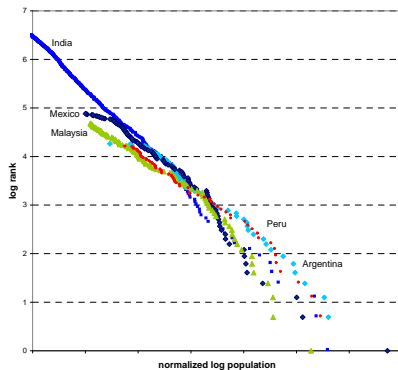


Zipf's Law Across Countries

Developed Countries



Developing Countries



Soo (2005)

- Documents empirically facts that were known since Rosen and Resnick (1980) and first suggested by Auerbach (1913)
- Main empirical regularity is Zipf's Law
- That is, the size distribution of cities is well approximated by a Pareto distribution with coefficient one. So:

$$y = Ax^{-\alpha}$$

or

$$\log y = \log A - \alpha \log x$$

where x is a particular city size and y is the number of cities with population greater than x

- Zipf's Law says that $\alpha = 1$

Use Three Different Estimators

- 1 $\log y = \log A - \alpha \log x + \varepsilon$
- 2 $\log y = (\log A)' + \alpha' \log x + \beta' (\log x)^2 + \varepsilon$
- 3 Use also Hill (1975) estimator which is the maximum likelihood estimator under the null hypothesis of a power law

Results

Table 1
Results of OLS regression of Eqs. (2) and (3) and the Hill estimator, for the sample of cities, for latest year of each country

Country	Year	Cities	OLS				Hill
			α	α'	β'	\log_4	α
Algeria	1998	62	1.351**	-2.3379	0.0408	18.7999**	1.3586*
Egypt	1996	127	0.9958	-2.9116**	0.0781**	15.0635	1.0937
Ethiopia	1994	63	1.0653	-4.3131**	0.1425**	14.2275	1.3341*
Kenya	1989	27	0.8169**	-1.9487**	0.0486**	11.2945**	1.0060
Morocco	1994	59	0.8735**	-1.0188	0.006	13.0697**	0.9295
Mozambique	1997	33	0.859**	1.0146**	-0.0811**	12.1286**	0.8107
Nigeria	1991	139	1.3509**	-0.9491	-0.00375	15.9784**	1.0459
South Africa	1991	94	1.3595**	-1.103	0.01076	19.1221**	1.2679*
Sudan	1993	26	0.9085	-0.2142	-0.0283	13.0723*	1.0066
Tanzania	1988	32	1.01	-1.8169	0.0348	13.6915	0.9089
Australia	1998	131	1.2279**	7.8925**	-0.4055**	17.6039**	0.8012**
Argentina	1999	111	1.0437	2.9939**	-0.1652**	16.1345**	0.9670
Brazil	2000	411	1.1341**	-0.0963**	-0.0418**	18.3681**	1.0607
Canada	1996	93	1.2445**	0.4273	-0.0689	18.0872**	1.2526
Chile	1999	67	0.8669**	-0.6516	-0.00915	13.0195**	0.7908*
Colombia	1999	111	0.9024**	-0.804	-0.00404	14.0252**	0.9345
Cuba	1991	55	1.09	-3.6859**	0.1093**	15.1299	1.3177
Dominican Republic	1993	23	0.8473	-0.6767*	0.0749**	11.6874**	0.8029
Ecuador	1995	42	0.8083**	-1.4086	0.0255	11.6871**	0.9015
Guatemala	1994	13	0.7287**	-3.6578**	-0.1249**	9.71255**	1.2074
Mexico	2000	162	0.9725	1.9514**	-0.1172**	15.8281	0.8127**
Paraguay	1992	19	1.0137	-1.9584	0.0415	13.1465	1.2571
USA	2000	667	1.3781**	-1.9514**	0.0235**	21.3849**	0.9339
Venezuela	2000	91	1.0631**	-0.7249	-0.0139	15.8205**	1.4277**
Azerbaijan	1997	39	1.0347	-5.2134**	0.1812**	13.6575	1.3605
Bangladesh	1991	79	1.0914	-4.1878**	0.1274**	15.6311	1.3545
China	1990	349	1.1811**	1.4338**	-0.1008**	19.5678**	0.9616
India	1991	309	1.1876**	-0.7453	-0.0170**	19.3916**	1.2178**
Indonesia	1990	235	1.1348**	-2.6325**	0.0610**	17.4209**	1.2334**
Iran	1996	119	1.0578**	-1.5539	0.01985	16.2499**	1.0526
Israel	1997	55	1.0892*	1.4982**	-0.1148**	14.8869**	1.0409
Japan	1995	221	1.3169**	-0.6325	-0.02655	20.6491**	1.2249**
Jordan	1994	34	0.8983**	-2.4831**	-0.0699**	12.0845	1.0629
Kazakhstan	1999	33	0.9615	4.8618**	-0.2444**	13.8818	0.8653
Kuwait	1995	28	1.719**	5.8975**	-0.3547**	20.5508**	1.6859*
Malaysia	1991	52	0.8716*	2.8194**	-0.1622**	12.6602**	0.8419
Nepal	2000	46	1.1870**	-2.0959	0.0405	15.5832**	1.2591
Pakistan	1998	136	0.9623	-2.4838**	0.0607**	15.0410**	1.0626
Philippines	2000	87	1.0804	3.4389**	-0.1838**	16.4972**	0.8630
Saudi Arabia	1992	48	0.7824**	0.0246**	-0.0333*	11.9143**	0.7302*
South Korea	1995	71	0.9077**	-0.3178	-0.02251	14.5804**	0.6850**
Syria	1994	10	0.7442*	-1.4709	0.02796	10.8967**	1.0862
Taiwan	1998	62	1.0587**	0.1482**	-0.0487**	15.7536**	0.9294
Thailand	2000	97	1.1864**	-4.9443**	0.1553**	16.6797	1.4184**
Turkey	1997	126	1.0536	-2.6659**	0.0642**	16.1683	1.1850
Uzbekistan	1997	17	1.0488	-8.9535**	0.3048**	14.7941	1.5111*
Vietnam	1989	54	0.9756**	-1.4203	0.0184**	14.1331*	0.8028

Table 1 (continued)

Country	Year	Cities	OLS				Hill
			α	α'	β'	\log_4	α
Austria	1998	70	0.9876	-3.9862**	0.1358**	13.0823	1.4226**
Belarus	1998	41	0.8435**	0.6492**	-0.0639**	12.2363**	0.7503*
Belgium	2000	68	1.5895**	-2.1862	0.02647	20.5048**	1.8348
Bulgaria	1997	23	1.114	-4.8424**	0.1531**	15.1382	1.2862
Croatia	2001	24	0.9207	-1.7693	0.03769	12.0916**	0.9551
Czech Republic	2001	64	1.1684**	-3.5189**	0.1029**	15.6961**	1.2669
Denmark	1999	58	1.3608**	-2.7601**	0.06274*	17.5639**	1.3753*
Finland	1999	49	1.1924**	-2.468*	0.0569**	15.6367**	1.3462
France	1999	104	1.4505**	-4.1897**	0.1137**	20.2497**	1.6388**
Greece	1998	190	1.238**	-0.3019**	-0.0384**	18.6477**	1.2548**
Germany	1991	43	1.4133**	-6.2010**	0.2036**	18.5979**	1.4804*
Hungary	1999	60	1.124**	-4.0186**	0.1254**	15.1636	1.2789
Italy	1999	228	1.3808**	-3.9073**	0.1064**	19.8143**	1.4967**
Netherlands	1999	97	1.4729**	-0.4333	-0.04491	20.0318**	1.4436**
Norway	1999	41	1.2704**	-4.5945**	0.1481**	16.2593**	1.4026
Poland	1998	180	1.1833**	0.3931**	-0.0679**	17.2931**	1.0908
Portugal	2001	70	1.382**	-4.1362**	0.1241**	17.7945**	1.6703**
Romania	1997	70	1.1092*	-0.0598	-0.0445	15.9369**	1.0598
Russia	1999	165	1.1861**	1.2459**	-0.0942*	18.9423**	1.0344
Slovakia	1998	42	1.3027**	-4.4861**	0.1428**	16.5644**	1.4810*
Spain	1998	157	1.1859**	-0.0658	-0.04697	17.5737**	1.0969
Sweden	1998	120	1.4392**	-1.2181	-0.00991	19.1777**	1.2867**
Switzerland	1998	117	1.4366**	-6.1258**	0.2229**	17.8549**	1.7386**
Ukraine	1998	103	1.0246	1.5787	-0.1058**	15.7615**	1.0197
Yugoslavia	1999	60	1.1827**	-2.2817	0.04839	15.8798**	1.1670
United Kingdom	1991	232	1.4014**	-3.5503**	0.0894**	20.3123**	1.3983**

*Significant at 5%; **significant at 1%; for α , significantly different from 1; for α' , significantly different from -1; for β' , significantly different from 0; for \log_4 , significantly different from the log of the population of the largest city. α is defined as a positive value; to compare the coefficients of $\log x$ in Eq. (2) and $(\log x)^\alpha$ in Eq. (3), we compare $-\alpha$ with α' .

Results

Table 2

Breaking down the results of OLS regressions (2) and (3) and the Hill estimator: statistical significance (5% level) in the latest available observation, for cities and urban agglomerations

Cities				Agglomerations			
Summary results: OLS estimates of α							
Continent	$\alpha < 1$	$\alpha = 1$	$\alpha > 1$	Continent	$\alpha < 1$	$\alpha = 1$	$\alpha > 1$
Africa	3	4	3	Africa	1	1	
N America		1	2	N America	2	1	
S America	4	4	2	S America	3	2	
Asia	5	8	10	Asia	3	2	
Europe	2	3	21	Europe	5	2	2
Oceania			1	Oceania	2		
Total	14	20	39	Total	16	8	2

Cities				Agglomerations			
Summary results: OLS estimates of β							
Continent	$\beta < 0$	$\beta = 0$	$\beta > 0$	Continent	$\beta < 0$	$\beta = 0$	$\beta > 0$
Africa	1	6	3	Africa	1		1
N America		1	2	N America	2	1	
S America	3	4	3	S America		5	
Asia	11	5	8	Asia	2	2	1
Europe	4	7	14	Europe	3	4	2
Oceania	1			Oceania	1	1	
Total	20	23	30	Total	9	13	4

Cities				Agglomerations			
Summary results: OLS estimates of A (compared to largest city)							
Continent	Less than	Equal to	Greater than	Continent	Less than	Equal to	Greater than
Africa	3	4	3	Africa	1	1	
N America		1	2	N America	1	2	
S America	5	2	3	S America	5		
Asia	6	7	10	Asia	2	3	
Europe	2	3	21	Europe	5	3	1
Oceania			1	Oceania	2		
Total	16	17	40	Total	16	9	1

Cities				Agglomerations			
Summary results: Hill estimator for α							
Continent	$\alpha < 1$	$\alpha = 1$	$\alpha > 1$	Continent	$\alpha < 1$	$\alpha = 1$	$\alpha > 1$
Africa		7	3	Africa	1		1
N America	1	1	1	N America	1	2	
S America	1	9		S America	1	4	
Asia	2	14	7	Asia		5	
Europe	1	12	13	Europe	1	8	
Oceania	1			Oceania	1	1	
Total	6	43	24	Total	5	21	

Close to One?

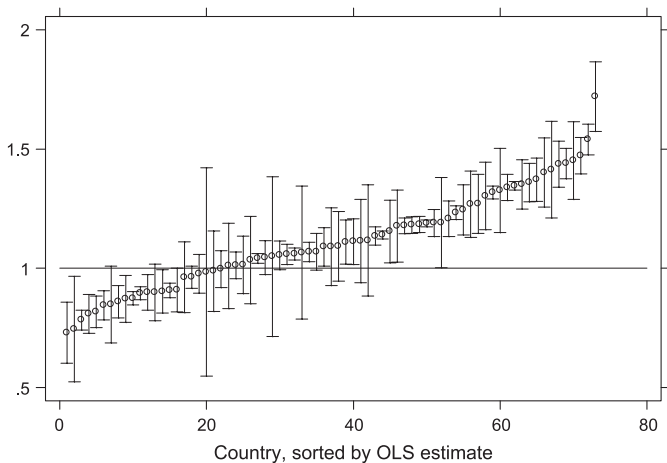


Fig. 1. Values of the OLS estimate of the Pareto exponent with the 95% confidence interval, for the full sample of 73 countries for the latest available period, sorted according to the Pareto exponent.

Agglomerations

Table 4

Results of OLS regression of Eqs. (2) and (3), and the Hill estimator, for the sample of urban agglomerations, for latest year of each country

Country	Year	AGG	OLS			Hill	
			α	α'	β'	$\log A$	α
Morocco	1982	10	1.10466	-14.207**	0.48473**	15.8475	1.5897
South Africa	1991	23	0.6275**	3.8188**	-0.1747**	10.1609**	0.5058**
Australia	1998	21	0.5855**	0.9107	-0.05806*	9.4412**	0.5087**
New Zealand	1999	26	0.7833**	-0.8086	0.0011	10.8562**	0.7830
Argentina	1991	19	0.7025**	-1.1177	0.01527	11.1267**	0.5229**
Brazil	2000	18	0.9904	-1.1245	0.00444	16.5577	0.9737
Canada	1996	56	0.8345**	-0.2635	-0.0225	13.0979**	0.8273
Colombia	1993	16	0.8278**	-0.2378	-0.02141	12.9431**	1.0567
Ecuador	1990	43	0.9046	-2.0169	0.0474	12.7637**	0.9573
Mexico	2000	38	0.9631	-1.3863	0.01501	15.6724	0.8107
Peru	1993	65	0.8295**	-1.5843	0.03171	12.3510**	0.8955
USA	2000	336	0.8847**	3.4992**	-0.1669**	16.1013	0.5225**
Bangladesh	1991	43	0.8068**	-2.9315**	0.08399**	12.1569**	0.9141
India	1991	178	0.9579**	0.1559**	-0.0419**	16.2945	0.9001
Indonesia	1990	193	1.0001	-1.1315	0.00532	15.8411	1.0384
Jordan	1994	10	0.6813**	0.2377	-0.03703	9.7100**	0.7286
Malaysia	1991	71	0.9429	3.3355**	-0.1872**	13.7914	0.8370
Austria	1998	34	0.7501**	-0.6338	-0.0051	10.6591**	0.6778**
Denmark	1999	27	0.8166**	-3.7224**	0.1235**	11.2213**	1.0903
France	1999	114	1.02332	-1.5263	0.02014	15.7905	1.0643
Germany	1996	144	0.8902**	0.5697**	-0.0578**	14.6429**	0.8886
Greece	1991	15	0.6349**	-3.987**	0.1324**	9.2190**	0.9499
Netherlands	1999	21	1.2301*	0.83	-0.08044	17.5350**	0.9703
Norway	1999	19	0.8828*	-1.7724	0.03853	11.7679**	0.9212
Switzerland	1998	48	0.9847	-0.1671	-0.0356**	13.7188	0.9557
United Kingdom	1991	151	1.0303*	-0.9192	-0.0045	16.0465	0.9438

AGG: number of urban agglomerations. *Significant at 5%; **significant at 1%; for α , significantly different from 1; for α' , significantly different from -1; for β' , significantly different from 0; for $\log A$, significantly different from the log of the population of the largest city. α is defined as a positive value; to compare the coefficients of $\log x$ in Eq. (2) and $(\log x)'$ in Eq. (3), we compare $-\alpha$ with α' .

Relationship with Other Variables

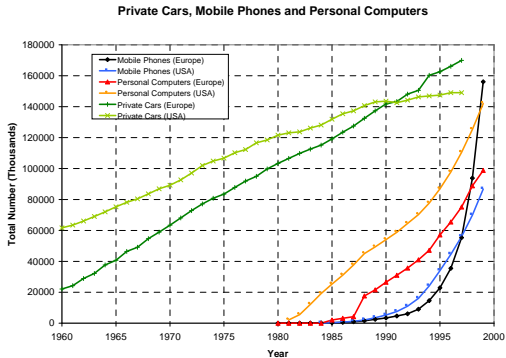
Table 5
Panel estimation of Eq. (5) (dependent variable=OLS coefficient of α)

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
Transport cost	-0.6151 (3.00)***	-0.2763 (1.13)	-0.4064 (1.36)	-0.8702 (3.48)***	-0.5014 (2.56)**	-0.6386 (2.31)**
Trade (% of GDP)	-0.0928 (1.71)*	0.0370 (0.51)	-0.0240 (0.30)	-0.0459 (0.89)	0.0532 (0.81)	-0.0177 (0.25)
Nonagricultural economic activity	-0.2411 (0.73)	-1.0137 (2.37)**	-0.5644 (1.69)*	-0.6002 (1.99)**	-1.4002 (3.37)***	-0.7731 (2.10)**
Scale economies	0.4467 (2.25)**	0.4462 (2.14)**	0.4057 (1.77)*	0.4993 (2.30)**	0.4756 (2.14)**	0.4284 (1.75)*
GASTIL index of dictatorship	-0.0375 (1.96)*	-0.0145 (1.32)	-0.0369 (1.97)**	-0.0307 (1.59)	-0.0028 (0.21)	-0.0284 (1.67)*
Total government expenditure	0.7837 (6.08)***	0.8013 (6.30)***	0.7500 (2.56)**	1.0097 (6.74)***	0.9598 (5.68)***	0.9154 (2.90)***
Timing of independence	-0.0596 (2.36)**	-0.0686 (2.82)***	-0.1429 (3.96)***	-0.0974 (3.80)***	-0.0984 (3.52)***	-0.1692 (4.75)***
War dummy	0.2211 (3.71)***	0.1410 (3.03)***	0.1474 (2.36)**	0.2437 (4.42)***	0.1425 (3.54)***	0.1659 (3.05)***
ln(land area)		0.0066 (0.39)	0.0288 (1.59)		0.0097 (0.64)	0.0239 (1.33)
ln(population)		0.0548 (3.50)***	0.0100 (0.49)		0.0459 (2.81)***	0.0032 (0.16)
ln(GDP per capita)		0.0959 (4.45)***	0.0585 (2.05)**		0.1053 (4.23)***	0.0467 (1.34)
Africa dummy			0.1306 (1.24)			0.0967 (0.97)
Asia dummy			0.2069 (1.85)*			0.1898 (1.92)*
North America dummy			-0.0655 (0.59)			-0.0184 (0.16)
South America dummy			-0.1304 (1.30)			-0.1459 (1.32)
Oceania dummy			-0.0804 (1.02)			-0.0375 (0.50)
Constant	1.1638 (3.96)***	-0.1307 (0.24)	0.3961 (0.69)	1.4082 (5.69)***	0.1885 (0.38)	0.8256 (1.57)
R-squared	0.4702	0.5778	0.6587	0.5403	0.6254	0.7007
Observations	79	79	79	72	72	72
Countries	44	44	44	40	40	40

α statistics in parentheses. *Significant at 10%; **significant at 5%; ***significant at 1% OLS with panel-corrected standard errors results reported.

The Effect of ICT on Urban Structure (Ioannides et. al, 2008)

- Changes in ICT are very clear in the data.
- Use Comin and Hobijn's 'Historical Cross-Country Technological Adoption Dataset' (2004)



Deviations from Zipf's Law

- It is clear that these changes can affect agglomeration effects and therefore the concentration of economic activity in space
- But how?
 - ▶ Increase spatial scope of spillovers
 - ★ Leads to less concentration or a reduction in the variance of city sizes
 - ▶ Could also make public good consumption more important as a share of consumption and more easily available
 - ★ Leads to more concentration

Empirical Methodology

- First obtain Zipf coefficient (ζ_{ct}) from

$$R_{ict} = \lambda_{ct} + \zeta_{ct} P_{ict} + e_{ict}$$

where (R_{ict}) is the rank and (P_{ict}) is population.

- Then run

$$\zeta_{ct} = \theta_c + \delta t + X_{ct}\eta + \varepsilon_{ct}$$

where X_{ct} is a collection of explanatory variables

- We use telephone lines as our measure of ICT, but likely endogenous so use IV
 - ▶ Use market structure: competitive, public or private monopoly

Results: Phone Lines

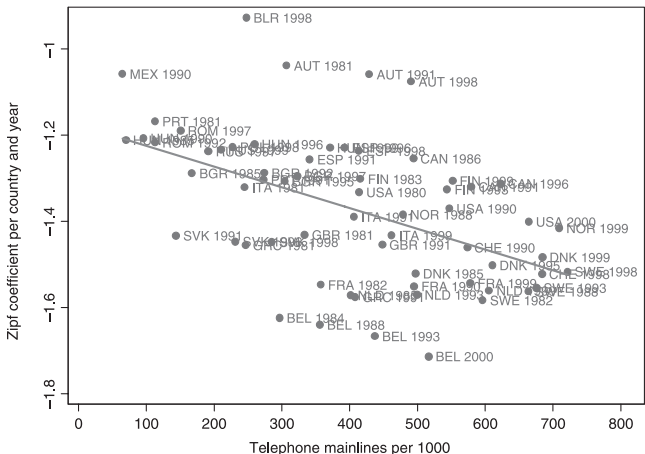


Figure 2. Estimated Zipf coefficients against telephone mainlines, per 1000 inhabitants, different countries and years

Source: Authors' own calculations.

Results: First Stage

Table 3. First stage regressions for Table 2

Dep. variable:	WLS			
	[5]	[6]	[7]	[8]
log(phone lines per capita)				
Public monopoly	-0.687*** (0.185)	-0.035 (0.147)	0.245*** (0.074)	0.305*** (0.055)
Private monopoly	-0.393* (0.227)	0.07 (0.177)	0.01 (0.177)	0.181 (0.153)
Time since end public monopoly	0.01 (0.027)	-0.001 (0.020)	-0.021 (0.014)	0.004 (0.014)
Time since end private monopoly	0.029 (0.029)	-0.009 (0.021)	0.004 (0.026)	-0.009 (0.024)
EU	0.077 (0.091)	0.113 (0.141)	-0.126 (0.106)	-0.297*** (0.081)
NAFTA	0.211 (0.326)	-0.052 (0.246)	-0.111 (0.148)	-0.222* (0.109)
Inverse road density		-1.054 (1.166)		0.102 (0.514)
log(country population)		-0.265** (0.103)		-1.334 (0.935)
log(GDP per capita), PPP		1.017*** (0.256)		-0.211 (0.315)
Trade, % GDP		0.011 (0.200)		0.472*** (0.150)
Non-agric. sectors, % GDP		2.467 (1.630)		3.804*** (0.877)
Gov. expend., % GDP		1.806** (0.775)		
Std. dev. of GDP growth		15.633* (8.430)		
log(land area)		0.181*** (0.063)		
Number of cities/1000		0.559 (0.849)		
Time	-0.039 (0.027)	0.017 (0.020)	0.047* (0.027)	0.060** (0.026)
Constant	-0.506** (0.223)	-12.352*** (3.843)	country fixed effects	
<i>F</i> -test instruments*	13.77	1.03	6.48	9.04
<i>R</i> ²	0.660	0.899	0.943	0.978
Obs country-year	63	63	63	63
Obs countries	24	24	24	24

Notes: Standard errors in brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively. Within *R*² is reported for fixed effects models. Weighted least squares (WLS) is weighted by the inverse standard error of the estimated Zipf coefficient.

* Jointly tests public and private telephony monopoly; EU/EEC-, NAFTA-membership.

Results: Phone Lines

Table 2. Phone lines and the city size distribution

Dep. variable:	WLS				IV			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Estimated Zipf coefficient								
log(phone lines per capita)	-0.146*** (0.027)	-0.109* (0.056)	-0.054* (0.029)	-0.100** (0.038)	-0.124*** (0.035)	0.151 (0.193)	-0.091** (0.039)	-0.117** (0.047)
Inverse road density		-0.335 (0.457)		0.168 (0.171)		-0.023 (0.588)		0.165 (0.172)
log(country population)		-0.021 (0.031)		-0.111 (0.092)		0.031 (0.052)		-0.131 (0.098)
log(GDP per capita), PPP		-0.103 (0.110)		-0.044 (0.093)		-0.368 (0.228)		-0.041 (0.094)
Trade, % GDP		-0.161** (0.076)		0.083* (0.045)		-0.166* (0.091)		0.085* (0.045)
Non-agric. sectors, % GDP		-0.408 (0.570)		0.320 (0.330)		-0.982 (0.791)		0.388 (0.348)
Gov. expend., % GDP		-0.828*** (0.280)				-1.221*** (0.433)		
Std. dev. of GDP growth		-1.101 (3.155)				-4.270 (4.374)		
log(land area)		0.003 (0.024)				-0.041 (0.041)		
Number of cities/1000		-0.171 (0.301)				-0.179 (0.359)		
Time	0.002 (0.001)	0.003 (0.002)	-0.002** (0.001)	0.001 (0.002)	0.001 (0.002)	0.002 (0.003)	-0.001 (0.001)	0.001 (0.002)
Constant	-1.495*** (0.033)	0.683 (1.603)	country fixed effects		-1.473*** (0.039)	3.995 (3.006)	country fixed effects	
R^2	0.344	0.675	0.807	0.845	0.336	0.537	0.798	0.844
Obs country-year	63	63	63	63	63	63	63	63
Obs countries	24	24	24	24	24	24	24	24

Notes: Standard errors in brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively. Within R^2 is reported for fixed effects models. Instruments are variables for public and private telephony monopoly, EU/EEC-, NAFTA-membership (see Table 3). Weighted least squares (WLS) is weighted by the inverse standard error of the estimated Zipf coefficient.

Results: Internet

Table 4. Internet users, phone lines and the city size distribution

Dep. variable:	WLS			
	[1]	[2]	[3]	[4]
Estimated Zipf coefficient				
log(internet users per capita)	-0.035*** (0.012)	-0.013 (0.018)	0.006 (0.016)	-0.015 (0.017)
log(phone lines per capita)			-0.185*** (0.056)	-0.312** (0.127)
Inverse road density		-0.886 (0.635)		-0.998 (0.587)
log(country population)		-0.006 (0.046)		-0.053 (0.046)
log(GDP per capita), PPP		-0.217 (0.135)		0.061 (0.168)
Trade, % GDP		-0.12 (0.105)		-0.067 (0.099)
Non-agric. sectors, % GDP		0.067 (0.800)		-0.389 (0.761)
Gov. expend., % GDP		-1.058** (0.399)		-0.874** (0.376)
Std. dev. of GDP growth		-1.481 (4.632)		-0.432 (4.294)
log(land area)		0.011 (0.027)		0.046 (0.029)
Number of cities/1000		-0.357 (0.410)		-0.27 (0.379)
Time	0.013** (0.006)	0.007 (0.009)	-0.002 (0.007)	0.01 (0.008)
Constant	-1.694*** (0.122)	1.07 (1.963)	-1.450*** (0.131)	-1.311 (2.052)
R^2	0.199	0.686	0.381	0.742
Obs country-year	41	41	41	41
Obs countries	23	23	23	23

Notes: Standard errors in brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively. Weighted least squares (WLS) is weighted by the inverse standard error of the estimated Zipf coefficient.

Results: Alternative Measures of Urban Concentration

Table 5. Effect of phone lines using alternative measures of urban concentration

Dep. variable	WLS OLS ^c				IV ^d			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Zipf coefficient ^a	-0.146*** (0.027)	-0.109* (0.056)	-0.054* (0.029)	-0.100** (0.038)	-0.124*** (0.035)	0.151 (0.193)	-0.091** (0.039)	-0.117** (0.047)
Gini coefficient	-0.051*** (0.016)	-0.084*** (0.028)	-0.017*** (0.006)	-0.032** (0.013)	-0.063** (0.027)	-0.208* (0.104)	-0.020*** (0.006)	-0.026 (0.018)
Herfindahl index	-0.038*** (0.011)	-0.036* (0.018)	-0.015*** (0.003)	-0.022** (0.008)	-0.068*** (0.019)	-0.114* (0.067)	-0.015*** (0.004)	-0.015 (0.011)
Coefficient of variation	-0.358** (0.162)	-0.699** (0.281)	-0.083** (0.039)	-0.173 (0.104)	-0.376 (0.272)	-1.975* (1.043)	-0.064 (0.045)	-0.024 (0.147)
Country fixed effects	no	no	yes	yes	no	no	yes	yes
Control variables ^b	no	yes	no	yes	no	yes	no	yes

Notes: Standard errors in brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

^a Results from Table 2.

^b Includes the same set of control variables as in Table 2.

^c Weighted least squares (WLS) estimation is used for the Zipf coefficient; ordinary least squares (OLS) for the Gini coefficient, Herfindahl index and coefficient of variation.

^d Includes the same set of instruments as in Table 3.

Rossi-Hansberg and Wright (2007)

- Most economic activity occurs in cities
 - ▶ 80% of the US population lives in cities; they earn 85% of personal income
- Aggregate economic activity is *urban* economic activity
- This creates a tension:
 - ▶ Urban agglomerations are evidence of increasing returns
 - ▶ It is difficult to find evidence of increasing returns at the aggregate level:
Balanced Growth
- We argue that urban structure is the margin that eliminates local increasing returns to yield constant returns to scale in the aggregate
- Using this margin has strong implications on urban structure: Zipf's Law and its deviations

Key Issues

- This paper presents an urban growth model that illustrates this mechanism using a particular specification that we can solve in closed form
- In the model cities arise endogenously: Trade-off between agglomeration forces, due to production externalities, and congestion forces, due to commuting costs
- Each city produces at an optimal scale given technology, productivity levels and the level of factors in the industry
- Given productivity levels and factor proportions industries behave as if using a linear technology by varying the number of cities
 - ▶ Constant returns to scale in the aggregate

Key Issues

- This mechanism has strong implications for the size distribution of cities once we include factor accumulation and productivity shocks
 - ▶ Our specification allows for exogenous and endogenous growth (via linearity in human capital accumulation)
- Scale independent city growth has been proven to generate Zipf's law (Gabaix (1999), Cordoba (2002))
- The model produces a scale independent city growth process for two polar cases
 - ▶ Fixed factors with permanent productivity shocks
 - ▶ *AK* model with temporary productivity shocks

Key Issues

- Diminishing returns to factor accumulation add persistence to the effect of transitory shocks
 - ▶ In general this implies scale dependent growth processes
- Urban structure exhibits some of the observed robust deviations
 - ▶ Growth process exhibits reversion to the mean
 - ★ Not enough small cities and large cities are not 'large enough'
 - ▶ Cross-country variation in the dispersion of city sizes
 - ★ Model points to industry standard deviation as the key parameter

Urban Growth Model

- Households:

- ▶ Order preferences over consumption according to

$$(1 - \delta)E_0 \left[\sum_{t=0}^{\infty} \delta^t N_t \left(\sum_{j=1}^J \theta_j \ln (C_{tj} / N_t) \right) \right]$$

- ▶ And accumulate physical and human capital according to

$$K_{t+1j} = K_{tj}^{\omega_j} X_{tj}^{1-\omega_j}$$

$$H_{t+1j} = H_{tj} \left[B_j^0 + (1 - u_{tj}) B_j^1 \right]$$

- Households grow at rate g , and allocate their members across industries satisfying

$$\sum_j N_{tj} \leq N_t$$

Urban Growth Model

- Technology:

- ▶ J industries in which a representative firm in city i produces using

$$A_{tj}^i \left(K_{tj}^i\right)^{\beta_j} \left(H_{tj}^i\right)^{\alpha_j} \left(u_{tj}^i N_{tj}^i\right)^{1-\alpha_j-\beta_j}$$

- ▶ where total factor productivity differs by city and industry according to

$$A_{tj}^i = A_{tj} \left(H_{tj}^i\right)^{\gamma_j} \left(N_{tj}^i\right)^{\varepsilon_j}$$

- ▶ Stochastic productivity A_{tj} is industry, but not city, specific
- ▶ The production externality is urban in scope

Urban Growth Model

- Technology:

- ▶ We identify different goods as different industries in the model
- ▶ Most definitions of industries include a variety of goods in the same industry
- ▶ Goods or industries in the model can be grouped according to technology
- ▶ Within groups, industries have the same parameters, but face idiosyncratic productivity shocks
- ▶ Across groups, all technological parameter may differ

The City Model

- Agents living in a circular city commute to work at the center.
- Commuting is costly, so rents are given by

$$R(z) = \tau (\bar{z} - z)$$

- Agents live in one unit of land, which implies city size

$$n = \bar{z}^2 \pi$$

- Total commuting costs are

$$TCC = \int_0^{\bar{z}} \tau z (2\pi z) dz = bn^{\frac{3}{2}}$$

City Developer Problem

$$\left\{ \frac{N_{tj}}{\mu_{tj}}, \frac{K_{tj}}{\mu_{tj}}, \frac{H_{tj}}{\mu_{tj}}, T_{tj}, \tau_{tj}^k, \tau_{tj}^h \right\} \max \left[\frac{b}{2} \left(\frac{N_{tj}}{\mu_{tj}} \right)^{\frac{3}{2}} - T_{tj} \frac{N_{tj}}{\mu_{tj}} - \tau_{tj}^k \frac{R_{tj}}{P_{tj}} \frac{K_{tj}}{\mu_{tj}} - \tau_{tj}^h \frac{S_{tj}}{P_{tj}} \frac{H_{tj}}{\mu_{tj}} \right],$$

subject to

$$(1 - \tau_{tj}^k) R_{tj} / P_{tj} = \beta_j Y_{tj} / K_{tj},$$

$$(1 - \tau_{tj}^h) S_{tj} / P_{tj} = \alpha_j Y_{tj} / H_{tj},$$

$$I_{tj} = (1 - \alpha_j - \beta_j) \frac{Y_{tj}}{N_{tj}} + T_{tj} - \frac{3b}{2} \left(\frac{N_{tj}}{\mu_{tj}} \right)^{1/2}.$$

Competition from other developers ensures that profits are zero, so

$$T_{tj} = \frac{b}{2} \left(\frac{N_{tj}}{\mu_{tj}} \right)^{\frac{1}{2}} - \tau_{tj}^k \frac{R_{tj}}{P_{tj}} \frac{K_{tj}}{N_{tj}} - \tau_{tj}^h \frac{S_{tj}}{P_{tj}} \frac{H_{tj}}{N_{tj}}.$$

Welfare Theorems

Proposition 1: There exists a unique Pareto efficient allocation for this economy.

Proposition 2: There exists a competitive equilibrium that attains the Pareto efficient allocation.

Proposition 3: Every competitive equilibrium in this economy is Pareto efficient.

- We use the social planner problem to solve for the unique equilibrium in this economy

Social Optimum

Choose state contingent sequences $\{C_{tj}, X_{tj}, N_{tj}, \mu_{tj}, u_{tj}, K_{tj}, H_{tj}\}_{t=0, j=1}^{\infty, J}$ so as to maximize

$$(1 - \delta)E_0 \left[\sum_{t=0}^{\infty} \delta^t N_t \left(\sum_{j=1}^J \theta_j \ln (C_{tj} / N_t) \right) \right]$$

subject to, for all t and j ,

$$C_{tj} + X_{tj} + b \left(\frac{N_{tj}}{\mu_{tj}} \right)^{\frac{3}{2}} \mu_{tj} \leq A_{tj} K_{tj}^{\beta_j} H_{tj}^{\alpha_j + \gamma_j} N_{tj}^{1 - \alpha_j - \beta_j + \varepsilon_j} u_{tj}^{1 - \alpha_j - \beta_j - \varepsilon_j - \gamma_j} \mu_{tj}$$

$$N_t = \sum_{j=1}^J N_{tj},$$

$$K_{t+1j} = K_{tj}^{\omega_j} X_{tj}^{1 - \omega_j},$$

$$H_{t+1j} = H_{tj} \left[B_j^0 + (1 - u_{tj}) B_j^1 \right]$$

City Size

- The problem of choosing the number of cities is static. The first order condition implies

$$\frac{N_{tj}}{\mu_{tj}} = \left[\frac{2(\varepsilon_j + \gamma_j) Y_{tj}}{b N_{tj}} \right]^2$$

and so

$$TCC_{tj} \equiv b \left(\frac{N_{tj}}{\mu_{tj}} \right)^{\frac{3}{2}} \mu_{tj} = 2(\varepsilon_j + \gamma_j) Y_{tj}$$

- City size is determined by

$$\frac{d(Y_{tj}/N_{tj})}{d(N_{tj}/\mu_{tj})} = \frac{d(ACC)}{d(N_{tj}/\mu_{tj})}$$

- No positive production externalities ($d(Y_{tj}/N_{tj})/d(N_{tj}/\mu_{tj}) \leq 0$) implies no cities
- No commuting costs ($TCC = 0$) imply only one city

City Size

- City developers choose this city size and impose subsidies on employment and human capital, but not on physical capital
 - ▶ This is the result of production externalities on labor and human capital
- As is evident from the previous equations the number of cities in an industry is not restricted to be an integer
- As the externality is industry specific, cities specialize
 - ▶ This can be easily generalized if we specify externalities across well defined groups of industries

City Growth

- Imposing optimal city size results in constant returns to scale production function

$$F_j \hat{A}_{tj} H_{tj}^{\hat{\alpha}_j} K_{tj}^{\hat{\beta}_j} N_{tj}^{1-\hat{\alpha}_j-\hat{\beta}_j} u_{tj}^{\hat{\phi}_j} = Y_{tj} - TCC_{tj}$$

- Tests at the aggregate level will find constant returns even though the economy produces with increasing returns to scale technology
- Industry productivity is determined in part by

$$F_j = (1 - 2(\gamma_j + \varepsilon_j)) \left[\frac{2(\gamma_j + \varepsilon_j)}{b} \right]^{\frac{2(\gamma_j + \varepsilon_j)}{1 - 2(\gamma_j + \varepsilon_j)}}$$

- ▶ Differences in the way production is organized in cities will be reflected in this term: A theory of differences in TFP

Differences in TFP

- Suppose cities are organized at a sub-optimal size, captured by a parameter $\kappa_j \neq 1$, such that

$$\frac{N_{tj}}{\mu_{tj}} = \kappa_j \left[\frac{2(\gamma_j + \varepsilon_j)}{b} \frac{Y_{tj}}{N_{tj}} \right]^2.$$

- Output net of commuting costs would be given by the same equation with modified F_j given by

$$F_j = (1 - 2(\gamma_j + \varepsilon_j) \kappa_j^{1/2}) \left[\frac{2(\gamma_j + \varepsilon_j)}{b} \kappa_j^{1/2} \right]^{\frac{2(\gamma_j + \varepsilon_j)}{1 - 2(\gamma_j + \varepsilon_j)}}.$$

- This has a global optimum at $\kappa_j = 1$.
- Inefficiently organized cities (too small *or* too large), implies lower total factor productivity.

Urban Structure

- City size in industry j is given by

$$\frac{N_{tj}}{\mu_{tj}} = \left[\frac{2(\varepsilon_j + \gamma_j)}{b} \frac{Y_{tj}}{N_{tj}} \right]^2$$

- So city growth rates satisfy

$$\begin{aligned} \ln \left(\frac{N_{t+1j}}{\mu_{t+1j}} \right) - \ln \left(\frac{N_{tj}}{\mu_{tj}} \right) &= 2 [\ln(A_{t+1j}) - \ln(A_{tj})] - 2(\hat{\alpha}_j + \hat{\beta}_j) g_N \\ &\quad + 2\hat{\alpha}_j \ln(B_j^0 + (1 - u_j^*)B_j^1) \\ &\quad + 2\hat{\beta}_j [\ln(K_{t+1j}) - \ln(K_{tj})] \end{aligned}$$

Urban Structure

- Or, solving for the dynamics of capital, in the long run

$$\begin{aligned} & \ln \left(\frac{N_{t+1j}}{\mu_{t+1j}} \right) - \ln \left(\frac{N_{tj}}{\mu_{tj}} \right) \\ = & \frac{2\hat{\alpha}_j}{1 - \hat{\beta}_j} \left[\ln \left(B_j^0 + (1 - u_j^*) B_j^1 \right) - g_N \right] \\ & + 2 \left[\ln(A_{t+1j}) - \ln(A_{tj}) \right] \\ & + 2(1 - \omega_j) \hat{\beta}_j \left[\ln(A_{tj}) - \sum_{s=1}^{\infty} \frac{\left(\omega_j + (1 - \omega_j) \hat{\beta}_j \right)^{-s}}{\left(1 - \left(\omega_j + (1 - \omega_j) \hat{\beta}_j \right) \right)^{-1}} \ln(A_{t-sj}) \right] \end{aligned}$$

Urban Structure

Proposition 4 (Exact Gibrat's Law and Zipf's Law): The growth process of city sizes satisfies Gibrat's Law, and therefore the invariant distribution for city sizes satisfies Zipf's Law, *if and only if* one of the following two conditions is satisfied:

- ① (No physical capital) There is no physical capital ($\beta_j = \hat{\beta}_j = 0$ or $\omega_j = 1$) and productivity shocks are permanent;
- ② (AK model) City production is linear in physical capital and there is no human capital ($\hat{\alpha}_j = 0, \hat{\beta}_j = 1$), depreciation is 100% ($\omega_j = 0$), and productivity shocks are temporary

Gibrat's Law and Zipf's Law

Proposition (Gabaix 1999): Suppose that city sizes S_t are determined by Gibrat's Law $S_{t+1}^i = \gamma_{t+1}^i S_t^i$, for some γ_{t+1}^i iid with distribution $f(\gamma)$. Then there exists an invariant distribution of city sizes satisfying Zipf's Law

Sketch of Proof: Normalize city sizes so that average size stays constant; then normalized growth rates satisfy $E[\gamma] = 1$. Then

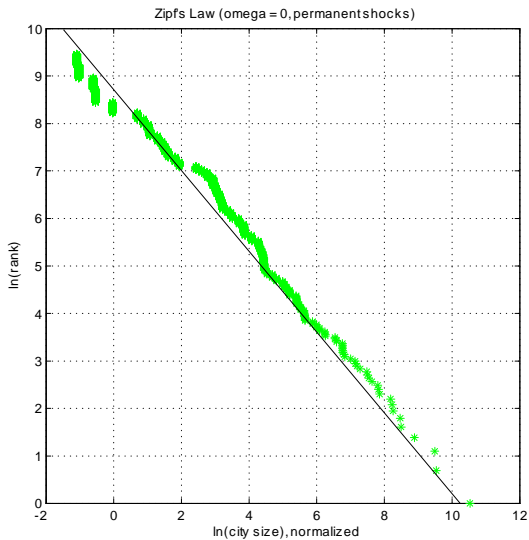
$$\begin{aligned} G_{t+1}(s) &= P(S_{t+1} > s) = P(\gamma_{t+1} S_t > s) \\ &= E \left[\mathbf{1}_{S_t > s/\gamma_{t+1}} \right] = E \left[G_t \left(\frac{s}{\gamma_{t+1}} \right) \right] = \int_0^\infty G_t \left(\frac{s}{\gamma} \right) f(\gamma) d\gamma. \end{aligned}$$

If there exists an invariant distribution G , we must have

$$G(s) = \int_0^\infty G \left(\frac{s}{\gamma} \right) f(\gamma) d\gamma,$$

which is obviously satisfied by a distribution of the form $G(s) = a/s$.

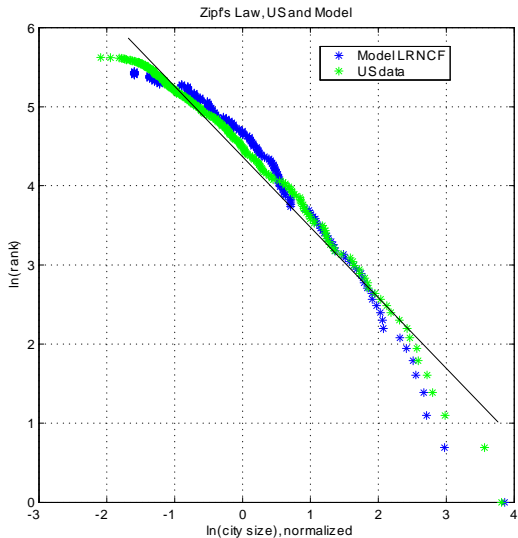
Exact Zipf's Law



Urban Structure

- The conditions placed on the model by the previous proposition are restrictive
- Reality surely lies between these two extremes: capital is one, but not the only, factor of production
- Between these two extremes, how close are the predictions of our model to observed urban structures?
- Recall that there are systematic deviations from Zipf's Law in the data

Deviations from Zipf's Law



Deviations from Zipf's Law

Proposition 5 (Concavity): The growth rate for cities exhibits (weak) reversion to the mean

- That is, given a common distribution of all future shocks,
 - ▶ Small cities grow faster
 - ★ Small past productivity shocks imply smaller capital stocks which increase the rate of return on capital given today's shock
 - ▶ Large cities grow more slowly
- In the data plots of log-rank against log-population are concave: small cities under-represented and big cities not 'big enough'

Deviations from Zipf's Law

Proposition 6: The standard deviation of city sizes increases with the standard deviation of industry shocks

- Given capital stocks, a larger standard deviation of shocks implies a larger standard deviation of city sizes and a larger standard deviation of investments, which in turn implies more dispersed capital stocks
- Zipf's coefficients estimated to lie between 0.8 and 1.6, the coefficients are positively correlated with income (Soo 2005)
- This explains the positive correlation between Zipf's coefficient and income *if* high income countries experience less volatile shocks

The Effect of Volatility

