UDC 330.4

Uladzimir Valetka
Belarus State Economic University
e-mail: uladzimir.valetka@gmail.com

Explanation of the variation of the city size distribution among transition countries

ABSTRACT. *The purpose of the present paper is to explain the determinants of the variation of city size distribution in time and across CEE and CIS transition economies (namely Belarus, Hungary, Poland and Russia) in 1970-2007. We use a unified database for CEE and CIS countries concerning city dynamics. In order to explain the differences in the city distributions and obtain valid statistical inference, we estimate, using cross-section dependence robust standard errors, a panel data fixed effects model to control for unobserved country specific determinants.*

ACKNOWLEDGMENTS. This research was supported by a grant from the CERGE-EI Foundation under a program of the Global Development Network. All opinions expressed are those of the author and have not been endorsed by CERGE-EI or the GDN.

Introduction

The demise of the socialist economic system and its subsequent restructuring has led to profound changes in the spatial patterns of urban economies in cities of CEE and CIS. The most important and visible trend of urban development during the transition period has been the decentralization of economic activities, a process which has played a major part in the transformation of the post-socialist city. The privatization of assets and the introduction of land rent have been the two determinant factors governing the process of urban spatial readjustments within the reality of a new market-oriented social environment (Stanilov, 2007).

To identify main drivers of city size distribution differences among examined countries and sequential policy implications we use panel data modeling to explain the determinants of the Pareto exponent variability. It is expected this should help us to understand better the earlier results of studying cities distribution Pareto and non-Pareto behavior and their "within" movements.

2 The data and the model

In order to explain the differences in the city distributions, we will estimate a panel data fixed effects model. To ensure valid statistical inference we will employ cross-section dependence robust standard errors as explained in section 4.5 of Necula et al (2010).

Variables of the panel for Belarus, Hungary, Poland and Russia for 1970-2007 annual data are presented in the Table 1.

Table 1. Description of the variables

pareto_cons	ζ_{it} consensus estimate of the Pareto exponent for the country i at time t
Gdpa	Real 2005 GDP ($ths) per country area (sq km)
Raila	Rail lines (total route-km) per country area (sq km)
mobpc	Mobile cellular subscriptions per 100 people
Telpc	Telephone lines per 100 people
Fri	Freedom index. It is an average of Political Rights and Civil Liberties indices measured on a one-to-seven scale, with one representing the highest degree of Freedom and seven the lowest.
prim1	Ratio of the lagest city population to the country population
prim5	Ratio of the 5 lagest city population to the country population
birthpc	Live births per 1000 people
abortion ratio	Abortions per 1000 live births
pop_log	Log of country population
gdppc_log	Log of country real 2005 GDP per capita ($)

Descriptive statistics for these variables are given in the Table 2.

Table 2. Summary statistics of the variables

Variable		Mean	Std. Dev.	Min	Max
Gdpa	overall	387,1828	347,815	29,50352	1168,422
	between		367,2832	39,88817	790,977
	within		138,3204	61,55127	897,8645
raila	overall	4,822252	3,386811	0,462357	8,694053
	between		3,860494	0,494237	8,234675
	within		0,467875	3,114926	5,598575
telpc	overall	14,75578	10,61307	2,812716	37,75789
	between		1,458964	13,18703	16,05529
	within		10,53709	1,67023	36,57452
mobpc	overall	11,58132	27,41879	0	115,5061
	between		4,849984	5,671009	17,3746
	within		27,09302	-5,79328	116,4641
fri	overall	4,842105	2,112264	1	7
	between		1,467838	3,552632	6,368421
	within		1,68376	1,973684	7,289474
prim1	overall	0,109544	0,062161	0,040217	0,203554
	between		0,069467	0,043094	0,188427
	within		0,014861	0,05976	0,147687
prim5	overall	0,194024	0,080886	0,105446	0,340832
	between		0,088904	0,116721	0,282678
	within		0,023985	0,103625	0,252178
ab_ratio	overall	1033,031	721,8916	0,34	2541,2
	between		759,8259	149,9337	1922,903
	within		291,9902	28,72814	1651,328
birthpc	overall	13,34557	3,389471	8,134464	19,70818
	between		0,988537	12,34449	14,69424
	within		3,278834	7,74579	19,42145
pop_log	overall	17,11243	1,099001	16,01575	18,81603

	between		1,263957	16,09978	18,7726
	within		0,040979	16,98827	17,16126
gdppc_~g	overall	8,38544	0,459095	7,428048	9,298145
	between		0,45203	7,761562	8,843453
	within		0,237708	7,881959	9,01591

The *fixed effects model* allows the intercept to vary across countries, while keeping the slope coefficients the same for all 4 countries. The model can be made explicit for our application by inserting a 0-1 covariate for each of the countries except the one for which comparisons are to be made. The estimated equation is:

$$\zeta_{it} = \beta_1 + \beta_2 EcGeo_{it} + \beta_3 ICT_{it} + \beta_4 SocPolit_{it} + \beta_5 YEAR_t + \beta_6 CONTR_{it} + \varepsilon_{it} \qquad (1)$$

where ζ_{it} is the Pareto exponent, *EcGeo* is the vector of economic geography variables (real 2005 GDP ($ths) per country area (sq km), rail lines (total route-km) per country area (sq km)), ICT is the vector of *information and communication technologies (*mobile cellular subscriptions per 100 people, telephone lines per 100 people), SocPolit is a group of political and social variables (Freedom index defined as an average of Political Rights and Civil Liberties indices measured on a one-to-seven scale, with one representing the highest degree of Freedom and seven the lowest, Primacy index1 defined as a Ratio of the lagest city population to the country population, Primacy index1 defined as a Ratio of the 5 lagest city population to the country population, Abortions per 1000 live births). CONTROL is a set of variables controlling for the size of the country; here the control variables used are the log of the real 2005 GDP per capita in constant US dollars and the log of population.

3 The results and discussion

Table 3 presents the results using the OLS estimate of the Pareto exponent as the dependent variable. Column (1) is the model without country controls. Both economic geography variables, real GDP per sq km of the country area and rail lines density, appear to facilitate the more even distribution of the cities. We cannot say the same about the influence of the information and communication technologies: proxy variable illustrating a popularity of mobile cellular services provided to be a factor explaining the bigger agglomerations development. Again primacy measured as a dominance of the 5 biggest cities has a negative effect on Pareto exponent thus contributing to less even development of urban systems.

Table 3. Panel estimation of the model (dependent variable - pareto_cons)

Dependent variable	(1)	(2)
Gdpa	.00036626	.00011472
	(5.19) ***	(1.48)
Raila	.06593139	.00897641
	(4.17) ***	(0.61)
Telpc	.00108669	-.00468902
	(1.03)	(-4.25) ***
mobpc	-.00079857	-.00153218
	(-3.56) ***	(-7.49) ***
Fri	-.00590168	.0021019
	(-1.08)	(0.46)
prim1	.86097608	1.3577834
	(0.45)	(0.86)
prim5	-3.012506	-3.7829106
	(-2.61) *	(-3.89) ***
abortion ratio	-.00004309	-2.226e-06
	(-2.30)*	(-0.13)
pop_log		-1.1784986
		(-7.90) ***
gdppc_log		.13604305
		(3.97) ***
Year	.0004134	.0100561
	(0.26)	(5.84) ***
Constant	.5110595	.84262033
	(0.17)	(0.32)
R-squared	0.7406	0.8289

t statistics in parentheses. * Significant at 5%; ** significant at 1%; *** significant at 0,1% level.

Index of political freedom enters with the theoretically predicted sign but is not significant at 5% level. It is interesting to note that the sign of the coefficient which held such a sensitive variable as abortion ratio (illustrating abortions per 1000 live births) confirms its connection with uneven urbanization.

Including controls for country size (column (2)) shows that the results of the economic geography variables are not robust. The same is stressed by Soo (2005) in his analysis of 44 countries panel. This contrasts with the strong robustness of the information and communication technologies variables. The only robustly significant variable from the social and political group is the level of primacy of the 5 biggest cities, and this enters with the sign we would expect from theoretical reasoning. Thus, these results suggest that political factors play a more important role than economic geography variables in driving variation in the Pareto exponent across countries.

The signs of all significant variables remain unchanged in both equations. Intraclass correlation (rho) suggests that almost all the variation in Pareto exponent is related to inter countries

differences (see Tables A.1-2 in the Annex). The F tests indicate that there are significant individual (country level) effects implying that pooled OLS would be inappropriate. Nevertheless we have run OLS and can see that the fixed effects estimates of the panel are considerably lower than the OLS estimates, suggesting that the OLS estimates were inflated by unobserved heterogeneity. The Hausman test rejects the null hypothesis that the coefficients estimated by the efficient random effects estimator are the same as the ones estimated by the consistent fixed effects estimator.

Comparing our results to previous findings, one can see that our results are quite in line with findings of Soo (2005). At the same time, we have to some extent different results from those of Soo (2005) and Rosen and Resnick (1980), as they find that the Pareto exponent is positively related to total population. Our specification demonstrates larger R-squared compared to those of both Soo (2005) and Rosen and Resnick (1980) papers.

4 Concluding remarks

To answer the question about the sources of cities distribution differences among countries we use panel data techniques. Urban and regional policy implications could be based on derived conclusions.

Fixed effects model estimations controlling for country size show that economic geography variables are not robust what is in agreement with Soo (2005). This contrasts with the strong robustness of the information and communication technologies variables. The only robustly significant variable from the social and political group is the level of primacy of the 5 biggest cities which enters with the negative sign. This result confirms that political factors play a more important role than economic geography variables in driving variation in the Pareto exponent across countries (assuming this variable is a good proxy for the level of centralization and state intervention). The sign of the primacy variable coefficient indicates that the lower political intervention means the more even population distribution. Our general conclusion thus is that political intervention with significant probability takes the form of the expansion of the largest cities and the size distribution becomes more unequal.

ANNEX

Table A1. *Fixed effects model (1) results*

N of obs = 152; N of groups = 4

pareto_cons	Coef	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdpa	0,000366	7,05E-05	5,19	0	0,000227	0,000506
raila	0,065931	0,015812	4,17	0	0,034668	0,097195
telpc	0,001087	0,001053	1,03	0,304	-0,001	0,003169
mobpc	-0,0008	0,000224	-3,56	0,001	-0,00124	-0,00036
fri	-0,0059	0,00548	-1,08	0,283	-0,01674	0,004934
prim1	0,860976	1,907311	0,45	0,652	-2,91012	4,632068
prim5	-3,01251	1,156043	-2,61	0,01	-5,29821	-0,7268
ab_ratio	-4,3E-05	1,87E-05	-2,3	0,023	-8E-05	-6.04e-06
year	0,000413	0,001561	0,26	0,792	-0,00267	0,0035
_cons	0,51106	3,058127	0,17	0,868	-5,5354	6,55752
R-sq:	-within	0.7406		sigma_u	0,423641	
	-between	0.2170		sigma_e	0,042469	
	-overall	0.1920		rho	0,99005	
	corr(u_i, Xb)	-0.9630		F u_i=0	75.20	
	F(9,139)	44.09				
	Prob > F	0,0000		Prob > F	0.0000	

Table A2. *Fixed effects model (2) results*

N of obs = 152; N of groups = 4

pareto_cons	Coef	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdpa	0,0001147	0,0000775	1,48	0,141	-0,0000386	0,000268
raila	0,0089764	0,0147515	0,61	0,544	-0,0201936	0,0381464
telpc	-0,004689	0,0011027	-4,25	0	-0,0068695	-0,0025086
mobpc	0,0021019	0,0046139	0,46	0,649	-0,0070217	0,0112255
fri	1,357783	1,570498	0,86	0,389	-1,74778	4,463334
prim1	1,357783	1,570498	0,86	0,389	-1,747767	4,463334
prim5	-3,782911	0,9720792	-3,89	0	-5,70513	-1,860691
ab_ratio	0,1360431	0,034285	3,97	0	0,0682469	0,2038392
year	0,0100561	0,001723	5,84	0	0,0066489	0,0134633
_cons	0,8426203	2,627306	0,32	0,749	-4,352696	6,037937
R-sq:	-within	0.8289		sigma_u	1.27206	
	-between	0.1176		sigma_e	0,0347403	
	-overall	0.0859		rho	0,9992547	
	corr(u_i, Xb)	-0.9951		F u_i=0	21,26	
	F(11,137)	60,34				
	Prob > F	0,0000		Prob > F	0.0000	

References

1. Alperovich, G, (1993), "An Explanatory Model of City-Size Distribution: Evidence From Cross-Country Data," *Urban Studies* 30 (9): 1591-1601.

2. Baltagi, B.H., (2005), *Econometric Analysis of Panel Data*, 3rd Edition, John Wiley & Sons

3. Black, D., and J.V. Henderson, (1999), "A theory of urban growth", *Journal of Political Economy*, 107, 252–284.

4. Buckley, R. M. and F. Mini, (2000), *From Commissars to Mayors. Cities in the Transition Economies*, Washington, DC: World Bank

5. Cornia, G. A. and R. Paniccià (1998). The Transition's Population Crisis: Nuptiality, Fertility and Mortality Changes in Severely Distressed Economies in *Population and Poverty in the Developing World,* eds. G. de Santis and M. Livi Bacci, Oxford: Oxford University Press, 361-393.

6. Eeckhout, J., (2004), "Gibrat's Law for (All) Cities," *American Economic Review*, 94(5): 1429-1451.

7. Horváth, G. Regional and cohesion policy in Hungary in M. Brusis (Ed.) Central and Eastern Europe on the Way into the European Union: Regional Policy-Making in Bulgaria, the Czech Republic, Estonia, Hungary, Poland and Slovakia, CAP Working Paper, Munich, December 1999, 90–130.

8. Hsiao, C., (2003), *Analysis of Panel Data*, 2nd Edition, Cambridge University Press

9. Krugman, P., (1991), "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483-99.

10. Krugman, P., (1996), *The Self-Organizing Economy*, Blackwell, Cambrige, MA.

11. Pumain, D. (2010), "Urban systems", in Hutchison R. (ed.), *Encyclopedia of Urban Studies*, Sage Publications.

12. Quah, D., (1993), "Empirical cross-section dynamics in economic growth," *European Economic Review*, 37:426–434.

13. Rosen K., and M. Resnick, (1980), "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy", *Journal of Urban Economics,* 8, 165-186.

14. Soo, K.T., (2005), "Zipf's Law for cities: a cross-country investigation," *Regional Science and Urban Economics*, 35(3), 239-263.

15. Stanilov, K. (2007), "The restructuring of non-residential uses in the post-socialist metropolis", in K. Stanilov (ed.) *The post-socialist city: urban form and space transformations in Central and Eastern Europe after socialism*. Dordrecht: Springer, pp. 73–97.

16. White, H., (1980), „A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity" *Econometrica* 48(4): 817–838.

17. Wooldridge, J. M., (2001), *Econometric Analysis of Cross Section and Panel Data,* MIT Press.