Boosting, sorting and urban scaling of university cities in Europe

Štefan Rehák
Juraj Šimčisko

February 2026

Abstract

A well-established empirical regularity in urban science is that innovation outputs such as patents scale superlinearly with city size. While this relationship is typically summarized by a single scaling exponent, its structural interpretation remains unclear. This paper examines whether urban scaling operates directly through population size or indirectly through mediating mechanisms. Using data on 606 European Functional Urban Areas in 2020, we test whether inventor sorting and collaboration-network structure mediate the population–innovation relationship. Innovation is measured using fractional patent counts from the European Patent Office, and collaboration structure is captured by the External–Internal (E–I) index derived from co-inventor networks. We estimate a structural equation model to decompose total scaling effects into sorting and network components. Results show strong mediation through the number of inventors, consistent with a sorting mechanism. Larger cities also exhibit more internally oriented collaboration structures. Once inventor concentration and collaboration orientation are accounted for, population size no longer has a statistically significant direct effect on patenting. These findings suggest that urban scaling of innovation reflects the joint effect of inventor accumulation and collaboration structure rather than a direct impact of population size.

0.1 Introduction

A robust empirical regularity in urban science and economic geography is that knowledge-intensive outputs—such as patents and scientific publications—scale superlinearly with city size. In its canonical form, this relationship is described by a power law,

$$Y_i = Y_0 N_i^{\beta} e^{\xi_i},$$

where Y_i denotes an innovation outcome in city i, N_i is population, and $\beta > 1$ is typically found for innovation-related measures [Bettencourt et al., 2007]. While this pattern has been documented across countries, periods, and innovation indicators, its interpretation remains contested. In particular, the estimated scaling exponent represents a reduced-form relationship that conflates several underlying mechanisms rather than identifying a single causal channel.

The literature points to two mechanisms as particularly important. First, larger cities attract and retain a disproportionate share of skilled workers, researchers, and innovative firms, generating a sorting effect whereby innovation increases due to the composition of human capital rather than population scale per se [Broekel et al., 2023]. Second, cities differ systematically in their position within inter-urban collaboration networks. Innovation is increasingly produced through team-based and geographically distributed research, implying that cities are embedded in broader systems of intercity knowledge exchange rather than functioning as isolated production units [Yao et al., 2020].

Building on these insights, this essay decomposes the urban scaling coefficient into two mediating mechanisms. Rather than treating population size as exerting a direct effect on innovativeness, urban scaling is modeled as operating through (i) the sorting of researchers into cities, measured by the number of inventors, and (ii) the structure of collaboration networks, captured by an EI index reflecting the balance between internal and external collaboration ties. By jointly mediating these channels, the analysis distinguishes between sorting and collaboration effects while accounting for the original urban scaling relationship.

0.2 Conceptual framework

In interaction-based models of urban scaling, superlinear returns arise because larger populations generate a more than proportional increase in potential interactions [Bettencourt et al., 2007]. In interaction-based models of urban scaling, superlinear returns arise because larger populations generate a more than proportional increase in potential interactions [Bettencourt et al., 2007]. In these models, cities are effectively treated as closed interaction basins: the probability of productive encounters rises with local density. However, contemporary innovation systems are characterized by multi-scalar interaction structures. While local density facilitates frequent encounters, many knowledge flows occur across cities through formal collaboration networks. Cities therefore operate simultaneously as dense local interaction platforms and as nodes within larger interurban systems. Recognizing this dual structure motivates a decomposition of urban scaling into within-city and between-city interaction channels. This interaction-based logic is largely silent about between cities interactions. In Bettencourt-type models, innovation is implicitly assumed to emerge from interactions within cities. In practice, however, scientific and technological innovation is increasingly produced through intercity collaboration, with research teams spanning multiple urban areas.Put differently, urban scaling theory and network theory are conceptually linked through the role of interactions. In scaling models, superlinear outcomes emerge because the number of potential pairwise interactions increases more than proportionally with population size. Network theory, by contrast, focuses precisely on the structure of interactions—who connects with whom, and whether ties are concentrated locally or extend externally.

We therefore conceptualize population size as shaping urban innovativeness indirectly through two channels. The first is a sorting channel, whereby larger cities host more researchers and inventors. The second is a collaboration-network channel, whereby cities differ in the extent to which inventive activity relies on internal versus external collaboration ties.

To capture the latter, $E - I$ denote the external-internal index measuring the relative prevalence of inventor ties connecting a city to external partners versus internal collaborators. Following Krackhardt and Stern [1988], the $E - I$ index ranges from -1 (all ties internal) to $+1$ (all ties external). Higher values indicate a more outward-oriented collaboration structure, while lower values reflect predominantly local inventive activity.

These relationships are estimated within a mediation framework, which allows us to decompose the total effect of population size on patenting activity into distinct indirect components. Rather than interpreting urban scaling as a single aggregate parameter, this approach explicitly models the pathways through which population influences innovation. The first indirect channel captures sorting. Larger cities host more inventors, reflected in the elasticity δ_1, and a greater number of inventors in turn increases patent output, captured by γ_3. The product $\delta_1\gamma_3$ therefore quantifies the contribution of sorting to the overall scaling relationship. The second channel operates through collaboration structure. Population size may alter the balance between internal and external ties, and this network orientation can independently shape innovative performance. Following Fleming et al. [2007], external collaboration may provide access to novel knowledge and increase innovative output.

We estimate a structural equation model consisting of the following system:

$$EI_i = \alpha_0 + \alpha_1 \ln N_i + \varepsilon_{1i},$$

$$\ln(\text{Inventors}_i) = \delta_0 + \delta_1 \ln N_i + \varepsilon_{2i},$$

$$\ln(Y_i) = \gamma_0 + \gamma_1 \ln N_i + \gamma_2 EI_i + \gamma_3 \ln(\text{Inventors}_i) + \varepsilon_{3i}.$$

Where N_i denotes the population, Y_i denotes the number of patents and (Inventors_i) denotes the number of researchersin a given city.

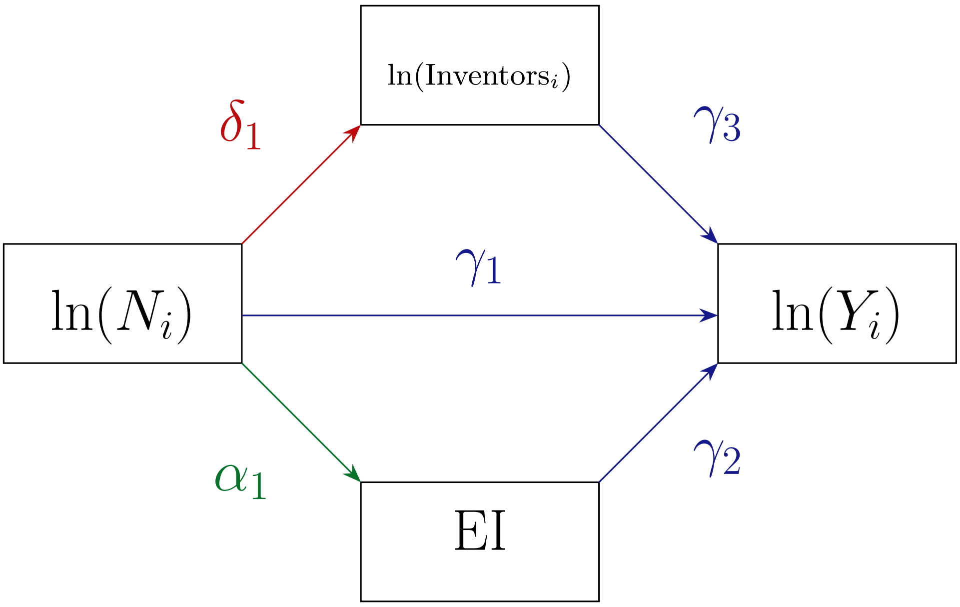

Figure 1: Conceptual mediation framework.

In this system, urban scaling operates through the joint mediation of sorting and collaboration-network structure. The indirect effects of population are given by $\delta_1\gamma_3$ for the sorting channel and $\alpha_1\gamma_2$ for the network channel, while the total effect corresponds to the sum of these mediated components and any remaining direct effect. Inference for indirect effects relies on nonparametric bootstrap procedures [Imai et al., 2010, Tingley et al., 2014].

0.3 Data

Our empirical analysis covers 606 European Functional Urban Areas (FUAs) observed in 2020. Innovation output is measured using fractional counts of patents from the European Patent Office (EPO). Population is measured as the resident count of each FUA, expressed in logarithmic form, providing the main scaling variable of interest. Collaboration structure is captured by the External–Internal (E–I) index constructed from patent co-inventor networks. The index is defined as $EI = (EL - IL)/(EL + IL)$, where EL denotes the number of external links connecting co-inventors located in different cities, and IL denotes the number of internal links between co-inventors located within the same city. The index ranges from -1 to $+1$. Higher values indicate a greater prevalence of intercity collaboration relative to within-city ties, while negative values reflect predominantly local inventive activity.

Table 1: Descriptive statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
E–I index	606	0.467	0.468	-1.000	1.000
log(Population)	606	12.592	0.942	9.815	16.418
log(Patents)	601	4.817	2.034	0.000	10.345
log(Inventor share)	606	3.084	2.157	-2.485	8.998

Table 1 reports descriptive statistics for the main variables. Patent output exhibits a highly skewed distribution, with a small number of large metropolitan Functional Urban Areas accounting for a disproportionate share of inventive activity. The E-I index has a mean of 0.47 and a standard deviation of 0.47, indicating substantial heterogeneity in collaboration structures across cities. The inventor share varies widely across cities, suggesting large differences of their concentration across European urban systems.

0.4 Results

Table 2: Structural equation model estimates

	Coefficient	Std. error	p-value
Outcome: ln(**InvShare**)			
EI index	-0.384^{***}	(0.084)	0.000
ln(Number of inventors)	0.956^{***}	(0.021)	0.000
ln(Population)	0.018	(0.035)	0.612
Constant	-1.537^{***}	(0.404)	0.000
Outcome: EI index			
ln(Population)	-0.153^{***}	(0.017)	0.000
Constant	2.394^{***}	(0.221)	0.000
Outcome: ln(**Number of inventors**)			
ln(Population)	1.383^{***}	(0.069)	0.000
Constant	-12.605^{***}	(0.871)	0.000
Observations		601	
Log-likelihood		-2941.73	
Bootstrap replications		1,000	

Notes: Table reports structural equation model estimates estimated by maximum likelihood. Bootstrap standard errors (1,000 replications) are reported in parentheses. The model includes three structural equations for patenting in European FUA, collaboration orientation (EI index), and the number of inventors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The first set of results examines the relationship between population size, the number of inventors, and collaboration orientation. Population is strongly associated with the number of inventors in a given city, with an estimated elasticity of 1.383 ($p < 0.001$). This finding points to a pronounced sorting mechanism, whereby larger cities host a disproportionately large share of inventive human capital.

Population size is also systematically related to collaboration structure. The coefficient linking population to the E–I index is negative and highly significant (-0.153, $p < 0.001$), indicating that inventive collaboration becomes increasingly internally oriented as cities grow. Larger urban systems rely more heavily on local inventor communities, while dependence on external collaborative ties declines.

Turning to the determinants of locally recorded patenting activity, two effects stand out. First, the number of inventors exhibits a strong and positive association with local patent outcomes. Conditional on collaboration structure and population size, a 1% increase in the number of inventors is associated with an approximately 0.96% increase in local outputs ($p < 0.001$). Second, collaboration orientation plays a significant role: a higher E–I index—reflecting a greater reliance on external collaboration—is associated with a significantly lower local patent share. The estimated coefficient is -0.384 ($p < 0.001$).

Once these two mediated channels are taken into account, population size itself no longer has a statistically significant direct effect on locally recorded inventive activity. The coefficient on population in the final equation is small and statistically indistinguishable from zero ($p = 0.612$). This result indicates that the conventional population-patents scaling relationship operates primarily through indirect mechanisms rather than a direct effect of population size.

Overall, these findings underscore that urban scaling exponents should not be interpreted as monolithic parameters. Instead, they reflect the aggregate outcome of multiple, overlapping mechanisms. Population growth increases innovation mainly by expanding the inventor base, which exhibits strong superlinear scaling. At the same time, city size reshapes collaboration networks: larger cities not only host more inventors but also provide dense local platforms for collaboration. Taken together, the results indicate that urban scaling of innovation arises from the joint accumulation of inventive capacity and the structure of collaboration networks.

References

Luis M. A. Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West. Growth, innovation, scaling, and the pace of life in cities. *PNAS*, 104(17):7301–7306, 2007.

Tom Broekel, Pierre-Alexandre Balland, and David L. Rigby. Boosting, sorting and complexity: urban scaling of innovation around the world. *Journal of Economic Geography*, 23(5):979–1006, 2023.

Lee Fleming, Santiago Mingo, and David Chen. Collaborative brokerage, generative creativity, and creative success. *Administrative science quarterly*, 52(3):443–475, 2007.

Kosuke Imai, Luke Keele, and Teppei Yamamoto. A general approach to causal mediation analysis. *Psychological Methods*, 15(4):309–334, 2010. doi: 10.1037/a0020761.

David Krackhardt and Robert N. Stern. Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly*, 51(2):123–140, 1988. doi: 10.2307/2786835.

Dustin Tingley, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59(5):1–38, 2014. doi: 10.18637/jss.v059.i05.

Li Yao, Jun Li, and Jian Li. Urban innovation and intercity patent collaboration: A network analysis of china's national innovation system. *Technological Forecasting and Social Change*, 160:120185, 2020.