ICT and Cities Revisited

Emmanouil Tranos[a,] and Yannis M. Ioannides[b*]

[a]Department of Geography, University of Bristol, UK
e.tranos@bristol.ac.uk

[b]Department of Economics, Tufts University, USA
[*]Corresponding author
yannis.ioannides@tufts.edu

May 11, 2020

ICT and Cities Revisited

Emmanouil Tranos and Yannis M. Ioannides

University of Bristol, UK

and

Tufts University, Medford, MA, USA

May 11, 2020

Abstract

This paper tests whether or not adoption of information and communication technologies (ICT) has offset agglomeration benefits and led to more dispersed spatial structures worldwide. The paper returns to Ioannides et al. (2008) and confirms, first by relying on the Pareto (Zipf) coefficient of the city size distribution as a proxy of spatial dispersion, that the diffusion of fixed telephony has caused more dispersed urban structures worldwide, in other words, greater urban decentralization. Similar causal effects are established for mobile telephony, which are novel, and the internet, which extend previous research. They are confirmed for such alternative measures of dispersion as the Gini coefficient, the Herfindahl index, and the coefficient of variation.

internet, spatial structure, Pareto, Zipf, cities, information and communication technologies

L96, R12, C23

1. Introduction

Substantial research effort has been spent on exploring the spatial incidence of the internet, even much before its massive expansion after the turn of the new millennium. Early research emphasized the internet's aspatial nature (Mitchell, 1995). Geographers, economists, but also engineers, theorized about the spatial impacts that rapid internet penetration might generate on individual cities and the national spatial structure. Much of that literature was speculative and even fanciful and not data-based. Examples abound, especially in the popular literature, like celebrating the emergence of *telecottages* (Toffler, 1980), the rise of a *borderless world* (Ohmae, 1995), the *death of cities* (Drucker, 1998; Gilder, 1995), and, more generally, the *end of geography* (O'Brien, 1992), the *death of distance* (Cairncross, 2001), and the emergence of a new *flat world* (Friedman, 2005). Today, more than 25 years after the commercialization of the internet, we know that the above narratives overstated the potential of the internet and other digital communication technologies to substitute for face-to-face interactions, diminish the cost of distance and, indeed weaken agglomeration economies to such an extent as there would be no benefit for people and businesses to agglomerate in cities. The current global urbanization rates indeed bolster the case for thorough empirical investigations.

The rapid adoption rate and the pervasiveness of new internet-based information and communication technologies, ICT for short, such as online social media and mobile-telephony hosted internet, which have increased rapidly during the last 10-15 years in the developing world, too, prompt us to examine how exactly ICT might have affected agglomeration economies. Answering such research questions is becoming more urgent nowadays when low latency and high density 5G mobile phone networks are deployed in cities around the world. Such networks are speculated to lead to disruptive urban changes including the support of driverless cars (Batty, 2018).

Conflicting technology examples can be illustrated. On the one hand, despite the broader agreement that no digital technology can reach the content richness of face-to-face communications, empirical research from the management field suggests that current digital technologies can effectively facilitate the sharing of knowledge with low to medium *tacitness* and even support knowledge sharing of a high degree of tacitness (Panahi, Watson, & Partridge, 2013). On the other hand, the very same technologies can further enhance what Storper and Venables (2004) termed *buzz,* as the widespread broadcasting of our personal and professional updates and whereabouts enabled by ICT can directly facilitate deliberate as well as unplanned or even unintentional face-to-face meetings (Emmanouil Tranos, 2020).

The above discussion is relevant for theoretical modeling, as it lies within the core of urban agglomeration economies and poses the question of whether physical distance matters. As the urban economics literature suggests, agglomeration externalities are triggered when agents are located in near proximity because the potential for interaction and knowledge spillovers is higher (Rosenthal & Strange, 2004). Hence, face-to-face interactions and the implied knowledge spillovers are facilitated within cities due to the opportunities for decreased transportation and interaction costs (Storper & Venables, 2004). However, ICT have the capacity to directly affect this process by further reducing "transportation" that is interactions cost[1] (McCann, 2008). In essence, the internet and digital communications facilitate some of

[1] Or, more broader, spatial transmission costs to follow McCann's terminology (2008).

the key forces underlying agglomeration economies, as identified by Duranton and Puga (2004), such as *learning* and *matching*. Web-based applications such as Massive Open Online Courses could decrease the need for co-location of participants in order to participate in formal learning activities. Furthermore, online social media such as LinkedIn[2] can enhance the probability of matching and the quality of matches even within cities (e.g. Grant & Grant, 2016; Utz, 2015; Utz & Levordashka, 2017). The critical question is whether the widespread adoption of ICT offsets the benefits of agglomeration economies and result in more dispersed spatial structures, or it further reinforces such externalities and leads to more concentrated spatial structures.

This paper contributes to the above discussion by presenting empirical research on whether the internet and digital communications have affected spatial structure, and more specifically as seen via the size distribution of cities. Contrary to most of the previous empirical studies, which are reviewed in the next section, this paper returns to the same empirical setting of Ioannides, Overman, Rossi-Hansberg, and Schmidheiny (2008) and re-examines their results in view of availability of additional data on fixed (landline) and mobile telephony and internet use, which cover the post-2000 period and until 2017. The enhanced technological maturity of mobile telephony and of the internet are of course key components of ICT diffusion. They thus behoove us to reexamine the question and indeed provide the main motivation in re-examining it.

Interestingly, our results support a substitutability argument. Specifically, the paper examines econometrically whether spatial structures have been affected by the adoption rates of the different communication technologies and establishes a causal effect: ICT adoption reduces urban dispersion, or equivalently, increases decentralization. Our results are also robust against potential endogeneity concerns, as one might claim that the take-up of these technologies could have been affected by spatial structures themselves. That is, individuals living in more dispersed spatial systems might have exhibited greater demand of such technologies in order to overcome the lower level of agglomeration externalities. To do so, we employ an instrumental variable approach and construct instruments which reflect the evolution and the structure of national telecommunications markets. Our findings can directly inform urban policy agendas as they advocate in favor of including digital strategies in policies aiming to enhance the potential of smaller urban centers to in effect avail themselves of agglomeration externalities albeit at a distance and thus improve the position of a city within its national urban hierarchy.

The organization of the paper is as follows: Section 2 provides a brief literature review; Section 3 describes the methods and the data we use; Section 4 presents the results of the multi-country analysis. Section 5 concludes.

2. Literature review

This section provides a brief review of the literature that explores the relationship between agglomeration economies and communication technologies. Gaspar and Glaeser (1998) were the first to model the effect

---

[2] LinkedIn has been identified as "the largest professional matchmaker site in the world" (Van Dijck, 2013, p. 207). See also Gee (2018).

of telecommunication improvements on the intensity of face-to-face interactions and city size. Their results indicate that technological improvements in telecommunications may lead to increased demand for face-to-face interactions, which will in turn increase the importance of larger cities as centers of interaction. Their theoretical model allows for a complementary relation between agglomeration externalities and advances in telecommunications. However, a fundamental assumption of their model is that face-to-face interactions are superior to any technology-mediated interactions. If this assumption were not to hold, then Gaspar and Glaeser (1998) would have predicted the opposite results. As indicated above and as the management literature suggests, the superiority of face-to-face interactions against digitally mediated ones is not always clear. Certain elements of knowledge sharing can also be achieved by via online interactions (Hildrum, 2009; Panahi et al., 2013). Indeed, to a certain extent, this argument is technology dependent. E.g., current teleconferencing capabilities are much more advanced than the ones available in the late 1990s. Online social media may also support such processes. These can be understood as global platforms, which perform an infrastructural role because of their capacity to connect people and support socialization (Barns, 2019; Gillespie & Ananny, 2016; Plantin, Lagoze, Edwards, & Sandvig, 2018). Hence, there is a need to empirically test how the relation between agglomeration externalities and telecommunications might have evolved.

Gaspar and Glaeser (1998) motivated a number of empirical studies which approached the key theoretical question of how digital communications affect agglomeration forces from various different perspectives. Kolko (2000) uses internet diffusion data and identifies a clear complementary link between internet usage and city size. Interestingly, he identifies higher internet domain densities in remote cities which indicate a substitution effect of the internet for longer-distance non-electronic communications. His results are consistent across different measures of internet diffusion (internet domain density and internet take-up). Sinai and Waldfogel (2004) approach the same question from the consumer's point of view. Specifically, they study the link between market size and locally targeted online content and find that more local content is online for larger markets, which favors a complementary link between the internet and cities. In addition, their analysis also indicates that holding local online content constant, the market size has a negative effect on individual connectivity which is indicative of a substitution effect. Forman et al. (2005) examine whether commercial internet adoption is higher in cities or rural areas. While the former would indicate a complementarity between internet adoption and cities, the latter would reflect a supplementary relation, according to which the internet is used as a means to offset costs and lack of opportunities related to peripherality. Their results indicate that despite internet adoption by firms with more than 100 employees being faster in smaller urban agglomerations, the adoption of more sophisticated internet-based applications is positively related with city size in 2000. Sohn, Kim, and Hewings (2003) compare how information technologies are related to urban spatial structure for Chicago and Seoul. Whereas they report a clear complementary link for Chicago, but not for Seoul, where they find that ICT contributes to a more dispersed spatial pattern. Focusing on the municipalities in the province of Barcelona, Pons-Novell and Viladecans-Marsal (2006) find a complementary link between individual internet take-up and off-line commercial offerings. Bekkerman and Gilpin (2013) focus on the role of locally based information resources using a dataset about the US libraries during the period 2000-2008. Their results suggest that internet access increases the demand and the value of locally accessible information and such complementarities are higher in larger metropolitan areas, which are expected to gain more benefits by internet access. Anenberg and Kung (2015) also identify a complementary relation between the internet and consumption variety in cities by focusing on food truck industry in the US. Craig, Hoang, and Kohlhase (2016) focus on internet take-up rates for the US states during the period 2000-2011. Their analysis provides suggestive evidence of a complementary role that internet connectivity performs on urban living. Focusing on rural areas, Partridge, Rickman, Ali, and Olfert (2008) find no evidence that rural distance penalties in the US have substantially changed since 1970s indicating that technological changes including the internet and digital communications have not managed to alternate spatial structure. Their interpretation of the absence of an increase in relative growth rates in rural counties is that either technological improvements increased the distance related costs, an argument which is aligned with the ideas proposed by McCann (2008), or distance costs have not decreased enough in order to alternate the growth trajectories of rural areas. Regarding distance, E. Tranos and Nijkamp (2013) highlighted the role of physical and relational proximities on the structure of the internet infrastructure. Interesting are also the findings of a recent study by Kim and Orazem (2016) on the economic effects of broadband internet in rural areas. They identify a positive effect on new firm location decisions, but this effect is higher in rural areas with larger population and in those rural areas which are adjacent to a metropolitan area, suggesting a complementarity between the internet and agglomeration economies. At a more aggregated level, Ioannides *et al.* (2008) examine the impact of fixed line telephony on urban structure using country-level data during the period 1980-2000. Using a panel dataset of spatial dispersion measures, they find robust evidence that an increase in the number of telephone landlines per capita encourages the spatial dispersion of population in that they lead to a more concentrated distribution of city sizes. However, whereas by the end of the coverage of their data, the evidence on internet usage is more speculative, they show that it goes in the same direction.

The consensus that emerges from previous research is that the declaration of the "death of distance" might have been premature (Rietveld & Vickerman, 2004). However, the exact impact of digital communications on spatial structure is still an open question. As Leamer and Storper (2001) indicate, the internet can affect both centripetal and centrifugal urban forces. Although most of the above studies support the complementarity argument, the results are not always conclusive and several of the above studies do argue in favor of substitutability.

Interestingly, the only cross country study (Ioannides et al., 2008) supports a clear substitution effect. Indeed, most of the studies discussed above focus either on the US or on some specific cities. Moreover, most of the above studies examine the complementarity/substitutability question for time periods when the internet and other digital communication technologies were still emerging. For instance, internet penetration in the US in 2000, which is the focus for quite a few of the above studies was just above 50 percent, while in 2016 it reached almost 90 percent (Pew Internet, 2016). At a global scale internet penetration increased from 7 to 46 percent during the same period (Internet Live Stats, 2017). In addition, although email and instant messaging technologies were widespread in the developed world in early 2000s, network externalities due to mobile internet and online social media were nowhere close to what we are familiar with today. For instance, Facebook users increased from 1 million in 2004 to more than 1.5 billion in 2015[3] and 65 per cent of adults in the US use it (Gramlich, 2018). Hence, it might have been

[3] http://newsroom.fb.com/company-info/

premature for the spatial economic effects of ICT adoption to have been materialized by the time that most of the above studies were conducted.

3. Methods and data

The main aim of this paper is to estimate the causal impact of ICT adoption on the spatial dispersion of economic activities and consequently population using multi-country data. Section 3.1 takes up some theoretical questions. Having these as a starting point, Section 3.2 discusses the methods and the data we use including the regression models we estimate. Section 4 reports the results.

3.1 Theoretical considerations

In the model of Rossi-Hansberg and Wright (2007), as adapted by Ioannides *et al.* (2008),[4] four parameters are essentially sufficient to express the variance of the long-run city size distribution. Three of these parameters are defined in terms of a specification of total factor productivity (TFP) for aggregate output for a city of type *j* at time *t*,

$$\tilde{A}_{tj} = A_{tj}\tilde{H}_{tj}^{\gamma_j}\tilde{N}_{tj}^{\varepsilon_j}, \tag{1}$$

where $\tilde{H}_{tj}$ and $\tilde{N}_{tj}$ denote industry j specific human capital and total employment for a city of type j, respectively, and $\ln A_{tj}$ is an independent and identically distributed (i.i.d.) productivity shock with mean zero and variance ν across all industries j and time periods t. Eq. (1) specifies, by means of a widely used functional relationship in economics (known as the Cobb–Douglas production function) how industry specific human capital in industry *j* and total employment in industry *j* ($\tilde{H}_{tj}$ and $\tilde{N}_{tj}$) along with a productivity shock, A_{tj}, contribute to define city-specific TFP, $\tilde{A}_{tj}$. Put otherwise, TFP for each city reflects both a random shock, drawn for the same distribution for all cities, and city-specific external effects due to total employment and human capital in that city, which is over and above those factors' direct contribution to aggregate city output. By definition, TFP augments the productivity of all direct inputs, that is over and above the productivity of labor, capital and human capital. It is the most concise and nowadays standard way to express the macroeconomic concept of productivity. Parameters γ_j and ε_j determine the importance of knowledge spillovers from total employment in industry *j*, $\tilde{N}_{tj}$, and industry *j*-specific human capital in the economy, $\tilde{H}_{tj}$, which are external to individual firms in the industry but internal to the urban economy (due to the presence of city developers.) If both parameters γ_j and ε_j are equal to zero, there are no external effects and economic activity has no incentive to agglomerate in cities. In that case, each city's TFP is simply an i.i.d. shock across the economy. The larger both of these parameters are, the more important are a city's total human capital and employment in determining $\tilde{A}_{tj}$ city-specific total factor productivity industry *j.*

[4] See Ioannides *et al.* (2008), Appendix A.

A simple way to introduce the effect of ICT is therefore to let these two parameters vary with the quality of information technology, ι. Namely, let $\gamma_j\left(\iota_t\right)$ where $\partial\gamma_j\left(\iota_t\right)/\partial\iota_t<0$ and, similarly, let $\varepsilon_j\left(\iota_t\right)$ be such that $\partial\varepsilon_j\left(\iota_t\right)/\partial\iota_t<0$. This assumption essentially amounts to ICT's increasing the importance of agglomeration effects since people located far away can now interact at a smaller cost and so people living in the city are less important in determining the city's productivity level. Conversely, we could assume that both $\gamma_j\left(\iota_t\right)$ and $\varepsilon_j\left(\iota_t\right)$ depend positively on the quality of ICT, which would be consistent with arguments that emphasize the greater importance of public goods as a result of changes in ICT. Which effect dominates is, exactly, the empirical question that we try to settle in this paper.

Ioannides *et al.* (2008) derive the following expression for the variance of the long-run city size distribution:

$$V_0\left[\ln s_{tj}\right]=4\nu\left(\left(\frac{1}{1-2\left(\gamma_j+\varepsilon_j\right)}\right)^2+\left(\frac{\beta_j}{1-2\left(\gamma_j+\varepsilon_j\right)+\beta_j}\right)^2\right), \tag{2}$$

where parameter β_j denotes the elasticity of physical capital in the city *j* production function. Note from (2) that the variance of the city size distribution is then increasing in the sum of the elasticities of the external effects from human capital and employment, $\gamma_j\left(\iota_t\right)+\varepsilon_j\left(\iota_t\right)$. Therefore, any assumption that we make about the dependence of these elasticities on ICT is reflected on changes on the invariant distribution of city sizes. We note that except fot these parameters, the variance of the long-run city-size distribution also depends on ν, the variance of the i.i.d. shock in the definition of city TFP, Eq. (1).

It is interesting to link the variance of the distribution of city sizes to the Zipf's coefficient. This readily connects the theoretical results in Ioannides *et al.* with the data through this coefficient. The local Zipf coefficient is given by the elasticity of the counter-cumulative of the city size distribution, $P\left(s>S\right)$, with respect to city size,

$$\zeta\left(S\right)=\frac{S}{P\left(s>S\right)}\frac{\partial P\left(s>S\right)}{\partial S}<0. \tag{3}$$

Given the mean of the distribution of city sizes, as the variance increases, mass is shifted to the tails of the distribution. This implies that for large enough city sizes, the term $\left|\zeta\left(S\right)\right|$ will be smaller the larger the variance. As the variance goes to infinity, $\zeta\left(S\right)>-2,\ \ \lim_{S\to\infty}\zeta\left(S\right)$ converges to the Pareto coefficient.[5]

[5] In an evocative passage, Clement (2004) illustrates the process by quoting Rossi-Hansberg and Wright (2007): "It is the size distribution of cities itself, and its evolution through the birth, growth and death of cities that leads to a reconciliation between increasing returns at the local level and constant returns at the aggregate." As he puts it, "

3.2 Multi-country analysis

The multi-country identification strategy is a two-step approach which uses Ioannides *et al.* (2008) as a starting point. It is well known that the Pareto law fits very well the city size distributions world-wide, and therefore, in view of the arguments in the previous section, the Pareto exponent (also known as the Zipf exponent) may be used as a convenient measure of dispersion. This is an appropriate measure of dispersion because of the extreme and characteristic heterogeneity of the city size distribution, and the very good fits normally obtained with such estimations. Other aspects of the city size distribution, such its location, naturally reflect all other parameters that enter the model; see Rossi-Hansberg and Wright (2007), whose modelled is adopted by Ioannides *et al.* (2008) and summrized in Appendix A, 235-239. In the specific estimation model with adopt below, they are reflected in the minimum size, S_o.

The first step of our methodology is to estimate the Zipf coefficient for a broad sample of countries over time. The Zipf coefficient is one of the most widely used measures of spatial dispersion with numerous applications in urban economics and economic geography (see for example Black & Henderson, 2003; Frenken & Boschma, 2007; Giesen & Südekum, 2011; Ioannides & Overman, 2003; Ioannides & Zhang, 2017; Nitsch, 2005; Rauch, 2013). City sizes, *s*, satisfy a Pareto law, if

$$P(s > S_i) = (\frac{S_i}{S_o})^{\zeta} \quad . \tag{4}$$

Parameter ζ is often estimated to be very close to -1, and S_o is a country-specific constant that is equal to the minimum city size; see Gabaix and Ioannides (2004). [6] In other words, the percentage of cities with population greater than S_i equals to a constant divided by the city population size, if ζ is close to *1*. Strictly speaking, eq. (1) implies Zipf's law, if $\zeta = 1$.

An approximation of Zipf's law is the rank-size rule. According to this deterministic rule, twice the population of the second largest city within an urban system equals to the population size of the largest city; similarly, three times the population of the third largest city equals to the population size of the largest city, etc. Therefore, eq. (4) can be approximated by the following equation (Gabaix & Ioannides, 2004):

$$r_i \approx (\frac{S_i}{S_o})^{\zeta}, \tag{5}$$

where, S_o is a constant which is equal to the smallest urban population of the urban system, and r_i is the rank of the city *i*, whose population is denoted by S_i. We are seeking to estimate this equation. The estimation of the logarithmic form of eq. (5) has been extensively used by the literature to obtain an estimate of ζ, known as the Zipf coefficient:

---

a national economy like a living organism, can shape its internal structure so that the national as a whole can expand on a stable course. … the invisible hand of urban growth."

[6] Some researchers in the literature define ζ as a positive parameter, in which case the right hand side of eq. (5) would be written as $(\frac{S_i}{S_o})^{-\zeta}$.

$$lnr_i = -lnS_0 + \zeta lnS_i + e_i\,. \tag{6}$$

The strict Zipf's law and the rank-size rule hold when ζ in (5) is to -1.

More generally, estimations of city size distributions across the world have also considered exponents ζ that are not necessarily equal to 1, in which case the Pareto, or power law, exponent is country- and time-specific, ζ_{ct}. Given that our aim here is to estimate the Zipf coefficient for a number of countries over time as a measure of dispersion, equation (6) describes the rank of city *i* in country *c* in year *t:*

$$lnr_{ict} = -lnS_{0ct} + \zeta_{ct} lnS_{ict} + e_{ict}. \tag{7}$$

The estimation of eq. (7) has been traditionally performed by Ordinary Least Squares (OLS). Gabaix and Ioannides (2004) discuss the downward bias of estimates of (7) using OLS on small samples. Gabaix and Ibragimov (2011) propose a practical correction for this bias, which we (along with many others) adopt in this paper: instead of using the log of rank of a city *i* in a country *c* in year *t*, they propose to use the log of rank-0.5, which is known as the Gabaix-Ibragimov correction.

All parameters that enter the model of urban structure, as articulated by Rossi-Hansberg and Wright (2007) and Ioannides et al. (2008), are also reflected in the minimum size $S_{o.}$ Housing is simplified as consumption of land, and land is homogeneous. For simplicity all industry is located at the Central Business District of each city, and residents commute to it from their homes. The model is premised on spatial equilibrium within each city and across cities. It follows that a more efficient urban transport system increases sizes of all cities and reduces housing rents. However, we assume that the impact of ICT is via the external effects, as expressed on the dispersion of the city size distribution; see discussion of eq. (2) above.

Researchers working in this area must contend with definitional differences as well as differences in availability of different kinds of data sources. Definitions of cities differ across countries for political, administrative and legal reasons. In order to compare our results with previous work from Ioannides *et al.* (2008), our starting point was the data obtained by Thomas Brinkhoff's City Population project (Brinkhoff, 2014)[7], which were also used by Soo (2005).

Table A1 in the Appendix presents the estimated Zipf coefficients for a panel of a sample of countries as near as possible to the one used by Ioannides *et al.* There is considerable variation in the estimated coefficients across different countries. Because of the empirically documented thick upper tail of data for cities and urban agglomerations, the Zipf coefficient constitutes a convenient measure of dispersion. The larger its absolute value, the thinner the upper tail; equivalently, the larger is the coefficient algebraically, the thicker the upper tail. This key observation serves as the basis for the second step of our methodology. It is also verified by reviewing the estimates reported on Table A1. They do not differ very much over time – the city size distribution is quite stable – and do differ considerably across countries. In countries like the UK, Finland, Russia, and Spain, heterogeneity is extreme. In countries like the USA and Denmark, it is much less so.

---

[7] http://www.citypopulation.de

The second step of our methodology involves estimating the following regression equation (Ioannides et al., 2008):

$$\zeta_{ct} = \theta_c + \delta t + X_{ct}\eta + \varphi_{ct} \, . \tag{8}$$

Eq. (8) enables us to estimate the effects of a number of explanatory variables included in the vector X_{ct} which pertain to the spatial structure of country c in year t as depicted on the Zipf coefficient noted as ζ and reported in Table A.1. The main explanatory variables of interest here are measures of ICT adoption, such as landline and mobile telephony and internet penetration, all measured in intensive form per 100 inhabitants (ITU, 2018). We choose these three variables to capture specific technological trends (internet usage and mobile technologies) and also to build comparability with the previous study by Ioannides et al. (2008). ICT capabilities do not remain constant over time. For instance, mobile capabilities and internet speeds were very different during our study period (2000-2017). Moreover, technological updates do not happen at the same time across our cross-section of countries. Hence, we address these heterogeneities by introducing time (yearly) and country fixed effects.

To address a potential omitted variable bias, eq. (8) includes country fixed effects θ_c, as well as a time trend δt; φ_{ct} is the error term. In addition, vector X_{ct} includes a number of control variables derived mostly from The World Bank (2018), the descriptive statistics and data sources of which together with these for the other variables used to estimate (8) are reported in Table 1. Referring to the control variables, total country population is an important measure of the economy's size; GDP per capita and GDP per capita growth are intimately related to urbanization, and so are population density and non-agricultural value-added as a share of GDP. Trade, that is exports and imports as a share of GDP, is an important time varying measure of openness. Government expenditure as a share of GDP may be a proxy of public investment in some countries (and government waste in others). The intensity of economic activity, GDP per capita, and its growth rate are intimately related to an economy's ability to innovate further. An economy's structure in terms of non-traditional sectors, such as agriculture, and of openness to the world economy are known to affect its susceptibility for innovation elsewhere in the world. A country's population and its geographical area are necessary along with the number of cities to proxy, in the most general way, for the density of economy activity, on which urban dispersion Is conditioned. Finally, the magnitude of the public sector as a share of GDP is an important control whose role may be ambiguous. A large public sector may signal a lot of infrastructure spending, which may however coexist with waste and corruption in the provision of public services. Public or private monopolies have been important features of the telecommunications industry and its ability to innovate, and so is the history of liberalization of the telecommunications industry. Whereas many technological improvements in the ICT area originated outside that industry – e.g. Microsoft originated as a software company, and Intel as a computer chip manufacturer both aimed at the general purpose technology market – the widespread adoption and proliferation of ICT has been aided to a remarkable extent by the telecommunications industry. A case in point is the ATT (Bell System) in the United States, prior to its deregulation.

Insert Table 1

Although internet penetration and mobile telephony are very strongly correlated, with a correlation coefficient of 0.846, fixed telephony and internet penetration are positively with a very weak correlation coefficient of 0.154, and fixed and mobile telephony are uncorrelated. This probably highlights the different composition of the population or infrastructure development patterns as mobile telephony can substitute for fixed line infrastructure, and mobile phone networks are also used as the main way to access the internet (Donner, 2008; Hamilton, 2003). Such a pattern of correlation underscores the importance of examining separately the effects of internet and mobile telephony penetration on the city size distributions.

The availability of a panel dataset for city sizes across countries for different years enables us to use country fixed effects, which can address potential endogeneity issues related to unobserved country specific characteristics of city size distributions. However, such a strategy does not address potential simultaneity issues. Simply put, internet penetration might be affected by spatial structure, as reflected in Zipf coefficients, or both internet penetration and spatial structure might be jointly determined by a third variable. E.g., if a country already has a dispersed spatial structure, internet is particularly suitable in facilitating communication. Potential endogeneity in our specification will prevent us from being able to determine the causal impact of internet and digital communication technologies usage on spatial structure, which is the main aim of this paper. In order to address this problem, we will adopt an instrumented variable strategy. Table 1 also includes the descriptive statistics for the instrumental variables we are using, which will be discussed in Section 4.

4. Digital technologies and spatial structure: a cross-country perspective

This section presents the estimation results of eq. (8). The LHS variable is the Pareto coefficient ζ_{ct}, as estimated according to the Gabaix and Ibragimov (2011) correction, at a first stage according to Eq. (7). The main variables of interest are introduced successively on their own in the regressions reported in Table 2. The regressions include country fixed effects to control for unobserved heterogeneity and a time trend. In addition, the observations are weighted with the inverse squared standard error of the estimated Zipf coefficient to address potential noise that is carried over from the first part of our identification strategy.

Regarding the interpretation of the estimated coefficients, given that the Pareto coefficient enters the regression as a real number, a negative coefficient for a RHS variable indicates an impact towards the increase of the spatial dispersion of population. In other words, a negative coefficient indicates an effect towards more uniform city sizes, that is less dispersion of city sizes. That would be evidence in favor of weakened agglomeration economies because of the expansion of digital technologies.

We first estimate eq. (5) using as the LHS variable the Zipf coefficient based on the Thomas Brinkhoff's City Population data (Brinkhoff, 2014). Here we present the results based on the City Population data, which are thus directly comparable with Ioannides *et al.* (2008). Notably, this data source does not have a minimum city size threshold, but we do impose the same threshold as Ioannides *et al.*

In terms of the size of the data, following the sample used in Ioannides et al (2008) the City Population data includes 24 countries (see Table A1 in the Appendix for the estimates of Zipf coefficients), but there are only a handful of observations for each country over the study period. The temporal sparseness of the data reflects the way that the City Population data are constructed as they are mostly based on national censuses.

In a nutshell, our results replicate Table 2, Ioannides et al. (2008) for the different ICT adoption variables, which are treated as endogenous. Table 2 direct replicates Ioannides et al. (2008) for fixed telephony adoption using the expanded data. The estimates for the original data are given in Table A.2. Table 3 reports similar results for internet penetration using the expanded data along and the same instrumental variables approach. Table 4 reports results for the penetration of mobile telephony, again with the same instrumental variables approach. These are completely novel results, unlike Tables 2 and 3, which extend results in Ioannides et al. All are at per capita basis. Indeed, to the best of our knowledge, this is the first such empirical examination of the impact of mobile telephony, Table 4, and the first reliable one for internet, Table 3, for which by the time of the Ioannides et al. (2008) investigation internet penetration was relatively new and their results are tentative.

In addition to fixed effects (country dummies) the following variables are included in the regressions in Tables 2-4: year dummy, country population in logs, log of GDP per capita and the standard deviation of its growth rate, the country's trade share (imports plus exports as share of GDP), the non-agricultural share of the economy, the share of government expenditure, the log of the country's land area, and the log of its number of cities. Relative to Ioannides *et al.*, we do not use information on road density because there are a lot of missing data. All the regressions include a constant. Relative to Ioannides *et al.,* we are also losing observations for some regressions as there are data missing for government expenditure and the share of the non-agricultural sectors.

Overall, there is a strong tendency for a negative effect of the endogenous ICT variables. ICT adoption reduces agglomeration forces and thus increases dispersion. We are losing significance for the early estimations for fixed telephony because we have too few data points. The negative coefficients for all ICT variables agree with the findings of Ioannides *et al*. (2008) results. As before, we turn our emphasis to the 2SLS estimations presented in Table 2, 3, and 4, Columns 5-8. We use the same identification strategy as Ioannides *et al.* (2008).

Insert Table 2: fixed all years

Insert Table 3: internet all years

Insert Table 4: mobile all years

To overcome the potential endogeneity problem, namely that urban structure directly affects ICT adoption, Tables 2-4 report estimates of eq. (8) using two-stage least squares (2SLS) with instrumental variables (IVs). The latter are variables which are correlated with our endogenous variables (internet, fixed and mobile telephony adoption), but are not correlated with the error term, or, in other words, do not directly influence spatial structure. Such an approach enables us to estimate the causal effect – if any – of the ICT variables on spatial structure. At a first stage, our endogenous variables are regressed against the IVs. Then, the predicted values of the endogenous variables based on the IVs and the other control variables are used instead of the endogenous variable to estimate eq. (7). A significant effect will verify the causal impact of the internet and digital communication usage on spatial structure. Following Ioannides et al. (2008), the IVs employed here reflect the structure and the evolution of the telecommunications market. Specifically, we construct two variables which indicate whether or not a telecommunications market is a public or a private monopoly and two variables which indicate how many years have passed since public and private monopoly deregulation. All four variables are time-varying and country specific. The underpinning idea is that market structure directly affects how services like fixed and mobile telephony as well as the internet are used, but it cannot directly affect the urban structure. Indeed, the structure of the telecommunications sector as well as its evolution directly affects the adoption of such ICTs. For instance, the time needed to reach saturation of ICTs adoption can be linked to how long telecommunication markets have been liberalised for and, therefore, the correlation condition can be met. Indeed, the instruments do not appear to be weak (weak identification rule of thumb above 10) for internet, mobile and fixed telephony for all years and they validate the OLS results from Columns 1-4 in Tables 2-4. However, they are weak for fixed telephony for early years (Table A2).

Regarding the exclusion restriction, the structure of the telecommunications market and how they evolved cannot directly affect spatial structure as such liberalisation policies are not spatial policies. Moreover, we cannot think of a way that such policies can lead to the relocation of individuals and businesses and, therefore, to changes in spatial structure. One might say though that the liberalisation of telecommunications markets may have increased the popularity of liberalisation policies in general, something which led to the privatisation of transportation markets. However, it is difficult to imagine a causal path running from the liberalisation of telecommunications market to the liberalisation of transportation market and, through this, to changes in the location choices of individuals and business and, therefore, to shifts in spatial patters. On the contrary, changes in both of the above markets are linked to a general privatisation trend that peaked in the late 1980s and early 1990s. Hence, we believe that our IVs can only affect spatial structure through the shifts they generate on ICTs usage (i.e. our endogenous variables).

Regarding the control variables, only a few of them have significant effects on spatial structure probably because the fixed effects (within) estimation masks the between-country variation. These effects are in agreement with previous research (Ioannides et al., 2008). Namely, GDP per capita has a consistent negative effect which indicates that wealthier countries tend to have more dispersed urban systems. The same applies for trade openness. The latter is in accordance with New Economic Geography models which indicate that international trade openness might weaken agglomeration forces (Fujita, Krugman, & Venables, 1999). In addition, the significant and negative coefficient of the time trend indicates that over time agglomeration forces weaken.

Finally, while the extreme heterogeneity of city size distributions make the Zipf coefficient an appropriate measure of dispersion, we also examine the robustness of our results by means of similar estimations of the effect of ICT but in terms of alternative measures of dispersion, that is, the Gini coefficient, the Herfindahl index, and the coefficient of variation. That is, Eq. (8) is estimated with the alternative measures as left-hand side variables, instead of ζ_{ct}. Table 5 reports estimated coefficients for ICT with these measures as endogenous variables, performing identical regressions to those reported earlier, respectively for the penetration of fixed telephony, of internet and of mobile telephony. The results support those obtained for the Zipf coefficient.

Specifically, using the full set of controls and country fixed effects, the IV estimates, reported in column 8, for fixed telephony, Table 5.1, for the internet, Table 5.2, and for mobile telephony, Table 5.3, agree with the previous results for all alternative measures of dispersion and are significant at high levels of significance, and the WLS ones do not only for the Herfindahl index. For the internet penetration, Table 5.2, all of the results are significant and in agreement with earlier ones.

Insert Table 5: robustness checks using alternative measures of dispersion

5. Discussion and Conclusions

This paper approaches empirically a research question which lies at the heart of urban economics and economic geography. Specifically, this paper employs data from a variety of sources in order to test econometrically whether or not the widening adoption of information and communication technologies has offset the benefits of agglomeration economies and resulted in more dispersed spatial population structures, or further reinforced such urban externalities and led to more concentrated spatial structures. Previous studies have produced conflicting results regarding whether ICT adoption and urban agglomeration externalities are complements or substitutes. As Leamer and Storper (2001) indicated, the internet can affect both centripetal and centrifugal forces.

The present study revisits Ioannides *et al.* (2008) with a completely open mind in view of the availability of several years of additional data on all ICT variables of interest. We examine first the effects of ICT adoption on Pareto (Zipf) coefficients for city size distributions, as measures of dispersion for thick-tailed data. We demonstrate that the evidence of a causal effect of ICT penetration in reducing city size dispersion and leading to more spatially dispersed population structures is also present when dispersion is measured alternatively in terms of the Gini coefficient, the Herfindahl, and the coefficient of variation. Our panel regressions include controls for cross-country socio-economic characteristics, which the literature indicated that can affect spatial structure. We also present 2SLS regressions using IVs previously introduced in the literature to address potential endogeneity and reverse causality issues. Our results remain qualitatively unchanged across the different specifications and measures of spatial of urban concentration, something which increases our confidence in the research strategy adopted here.

The main limitation of our study is linked to its very nature: a cross-country analysis of a relatively new phenomenon that is modern ICT. Regarding the latter we are limited to rather short panels as the take-up of commercial internet and mobile telephony were limited before 2000. Moreover, cross-country analysis is restricted to missing data, which leads to unbalanced panels like ours. Nevertheless, such cross-country analysis enables us to draw a broader picture regarding the relationship between ICT and spatial structure. Such an approach needs to be followed-up by more granular studies, which will not have to overcome the burden of cross-country heterogeneity.

We believe that our results, apart from their potential value in helping to resolve the theoretical ambiguity, have the capacity to inform urban policy. Improvements in internet speed is not a trivial policy instrument as it involves numerous complexities. Apart from infrastructure installation costs and engineering challenges, governance issues regarding the ownership of such digital networks, the provision of state subsidies and the design of public-private partnerships poses obstacles for the inclusion of such strategies in the urban growth agenda. To further inform such policies, more research at granular scales is needed. Such research may shed light on the micro-mechanisms behind such urban processes. For instance, such granular studies could also control for residential effects and housing prices, which also affect the structure of cities, but cannot be directly controlled for in such a cross-country setting. Examining the same set of questions for a greater number of countries also deserves attention in future research. In view of the fact that the urban structures of most of the countries studied here are likely much more settled than those of the emerging and developing economies strengthens the significance of our findings. The above questions become more relevant nowadays given the heated discussions regarding the rollout of 5G mobile networks. There is no doubt that such infrastructure will support the diffusion of radical urban innovations. What is unclear though is how such technologies are going to affect the structure of cities and the dispersion of the national population over them.

Acknowledgements

We are grateful to many audience members at the 2016 Annual Winter Seminar of the German speaking section of the European Regional Science Association, the North American Regional Science Council meetings, and Fall 2018 Opportunity and Inclusive Growth Conference, Federal Reserve Bank of Minneapolis, MN. We are also grateful to Max Nathan for very helpful comments, and to Esteban Rossi-Hansberg and anonymous reviewers, whose suggestions have not been yet fully incorporated in this version. Naturally, all errors are ours.

References

Anenberg, E., & Kung, E. (2015). Information technology and product variety in the city: The case of food trucks. *Journal of Urban Economics, 90*, 60-78.

Barns, S. (2019). Negotiating the platform pivot: From participatory digital ecosystems to infrastructures of everyday life. *Geography Compass, 13*(9), e12464.

Batty, M. (2018). *Inventing future cities*: Mit Press.

Bekkerman, A., & Gilpin, G. (2013). High-speed Internet growth and the demand for locally accessible information content. *Journal of Urban Economics, 77*, 1-10.

Black, D., & Henderson, V. (2003). Urban evolution in the USA. *Journal of Economic Geography, 3*(4), 343-372.

Brinkhoff, T. (2014). City Population. Retrieved from http://www.citypopulation.de/

Cairncross, F. (2001). *The death of distance 2.0*. London: Texere Publishing Limited.

Craig, S. G., Hoang, E. C., & Kohlhase, J. E. (2016). Does closeness in virtual space complement urban space? *Socio-Economic Planning Sciences*.

Donner, J. (2008). Research approaches to mobile use in the developing world: A review of the literature. *The information society, 24*(3), 140-159.

Drucker, P. F. (1998). From Capitalism to Knowledge Society. In D. Neef (Ed.), *The knowledge economy* (pp. 15-34). Woburn, MA: Butterworth-Heinemann.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. *Handbook of regional and urban economics, 4*, 2063-2117.

Forman, C., Goldfarb, A., & Greenstein, S. (2005). How did location affect adoption of the commercial Internet? Global village vs. urban leadership. *Journal of Urban Economics, 58*(3), 389-420.

Freedom House. (2014). Freedom in the World. Retrieved from https://www.freedomhouse.org/report-types/freedom-world#.VIa85DGsURo

Frenken, K., & Boschma, R. A. (2007). A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process. *Journal of Economic Geography, 7*(5), 635-649.

Friedman, T. L. (2005). *The world is flat*. New York: Farrar, Straus and Giroux.

Fujita, M., Krugman, P., & Venables, A. J. (1999). *The spatial economy*. Cambridge: MIT Press.

Gabaix, X., & Ibragimov, R. (2011). Rank− 1/2: a simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics, 29*(1), 24-39.

Gabaix, X., & Ioannides, Y. M. (2004). The evolution of city size distributions. *Handbook of regional and urban economics, 4*, 2341-2378.

Gaspar, J., & Glaeser, E. L. (1998). Information Technology and the Future of Cities. *Journal of Urban Economics, 43*(1), 136-156.

Giesen, K., & Südekum, J. (2011). Zipf's law for cities in the regions and the country. *Journal of Economic Geography, 11*(4), 667-686.

Gilder, G. (1995, February 27). *Forbes ASAP,* p. 56.

Gillespie, T., & Ananny, M. (2016). Exceptional Platforms Paper presented at the The Internet, Politics and Policy Conference. Retrieved from https://blogs.oii.ox.ac.uk/ipp-conference/2016/programme-2016/track-b-governance/platform-studies/tarleton-gillespie-mike-ananny.html

Gramlich, J. (2018). *8 facts about Americans and Facebook*. Retrieved from http://www.pewresearch.org/fact-tank/2018/10/24/facts-about-americans-and-facebook/

Grant, S. B., & Grant, S. B. (2016). Classifying emerging knowledge sharing practices and some insights into antecedents to social networking: a case in insurance. *Journal of Knowledge Management, 20*(5), 898-917.

Hamilton, J. (2003). Are main lines and mobile phones substitutes or complements? Evidence from Africa. *Telecommunications Policy, 27*(1), 109-133.

Hildrum, J. M. (2009). Sharing Tacit Knowledge Online: A Case Study of e-Learning in Cisco's Network of System Integrator Partner Firms. *Industry and Innovation, 16*(2), 197-218.

Internet Live Stats. (2017). Internet Users. Retrieved from http://www.internetlivestats.com/internet-users/

Ioannides, Y. M., & Overman, H. G. (2003). Zipf's law for cities: an empirical examination. *Regional Science and Urban Economics, 33*(2), 127-137.

Ioannides, Y. M., Overman, H. G., Rossi-Hansberg, E., & Schmidheiny, K. (2008). The effect of information and communication technologies on urban structure. *Economic Policy, 23*(54), 201-242.

Ioannides, Y. M., & Zhang, J. (2017). Walled cities in late imperial China. *Journal of Urban Economics, 97*, 71-88.

ITU. (2018). *World Telecommunication/ICT Indicators Database*. Retrieved from: www.itu.int/ITU-D/ict/statistics/

Kim, Y., & Orazem, P. F. (2016). Broadband Internet and New Firm Location Decisions in Rural Areas. *American Journal of Agricultural Economics, 99*(1), 285-302.

Kolko, J. (2000). The death of cities? The death of distance? Evidence from the geography of commercial Internet usage. In I. Vogelsang & B. M. Compaine (Eds.), *The internet upheaval: Raising questions, seeking answers in communications policy* (pp. 73-98). Cambridge, MA and London, UK: The MIT Press.

Leamer, E. E., & Storper, M. (2001). The Economic Geography of the Internet Age. *Journal of International Business Studies, 32*(4), 641-665.

McCann, P. (2008). Globalization and economic geography: the world is curved, not flat. *Cambridge Journal of Regions, Economy and Society, 1*(3), 351-370.

Mitchell, W. J. (1995). *City of bits: space, place and the infobahn*. Cambridge, MA: MIT Press.

Nitsch, V. (2005). Zipf zipped. *Journal of Urban Economics, 57*(1), 86-100.

O'Brien, R. (1992). *Global Financial Integration: The End of Geography*. London: Pinter.

Ohmae, K. (1995). *The Borderless World: Power and Strategy in an Interdependent Economy*. New York: Harper Business.

Panahi, S., Watson, J., & Partridge, H. (2013). Towards tacit knowledge sharing over social web tools. *Journal of Knowledge Management, 17*(3), 379-397.

Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2008). Employment Growth in the American Urban Hierarchy: Long Live Distance. *The B.E. Journal of Macroeconomics, 8*(1 (Contributions)), Article 10.

Pew Internet. (2016). *Pew Internet & American Life Project*. Retrieved from http://www.pewinternet.org/data-trend/internet-use/internet-use-over-time/

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New media & society, 20*(1), 293-310.

Pons-Novell, J., & Viladecans-Marsal, E. (2006). Cities and the Internet: The end of distance? *Journal of Urban Technology, 13*(1), 109-132.

Rauch, F. (2013). Cities as spatial clusters. *Journal of Economic Geography*, lbt034.

Rietveld, P., & Vickerman, R. W. (2004). Transport in regional science: The "death of distance" is premature. *Papers in Regional Science, 83*(1), 229-248.

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. *Handbook of regional and urban economics, 4*, 2119-2171.

Sinai, T., & Waldfogel, J. (2004). Geography and the Internet: is the Internet a substitute or a complement for cities? *Journal of Urban Economics, 56*(1), 1-24.

Sohn, J., Kim, T. J., & Hewings, G. J. D. (2003). Information technology and urban spatial structure: A comparative analysis of the Chicago and Seoul regions. *Annals of Regional Science, 37*(3), 447-462.

Soo, K. T. (2005). Zipf's Law for cities: a cross-country investigation. *Regional Science and Urban Economics, 35*(3), 239-263. doi:http://dx.doi.org/10.1016/j.regsciurbeco.2004.04.004

Storper, M., & Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography, 4*(4), 351-370.

The World Bank. (2018). *World Bank Indicators*. Retrieved from: http://data.worldbank.org/indicator

Toffler, A. (1980). *Third way*. New York: William Morrow.

Tranos, E. Social Network Sites and Knowledge Transfer: An Urban Perspective. *Journal of Planning Literature, in press*. doi:10.1177/0885412220921526

Tranos, E. (2020). Social Network Sites and Knowledge Transfer: An Urban Perspective. *Journal of Planning Literature, in press*. doi:10.1177/0885412220921526

Tranos, E., & Nijkamp, P. (2013). The death of distance revisited: cyber-place, physical and relational proximities. *Journal of Regional Science, 53*(5), 855-873.

Utz, S. (2015). Is LinkedIn making you more successful? The informational benefits derived from public social media. *New media & society*, 1461444815604143.

Utz, S., & Levordashka, A. (2017). Knowledge networks in social media. In *The Psychology of Digital Learning* (pp. 171-186): Springer.

Van Dijck, J. (2013). 'You have one identity': performing the self on Facebook and LinkedIn. *Media, Culture & Society, 35*(2), 199-215.

Table 1: Descriptive statistics

Variables	Obs.	Mean	Std. Dev.	Min	Max	Source
Zipf	114	-1.300	0.187	-1.849	-0.907	Own calculation using data from Brinkhoff (2014)
ln(fixed)	100	3.505	0.613	1.708	4.292	ITU (2018)
ln(net)	81	2.370	2.469	-2.996	4.552	ITU (2018)
ln(mobile)	82	3.129	2.248	-2.952	5.113	ITU (2018)
ln(population)	100	16.805	1.148	15.260	19.550	World Bank (2016)
ln(GDPpc)	104	10.117	0.815	7.879	11.382	World Bank (2016)
GDPpc growth (st. dev.)	114	2.113	1.626	0.085	7.722	World Bank (2016)
trade (GDP %)	95	75.170	34.009	19.761	185.747	World Bank (2016)
non agriculture (GDP %)	77	95.847	3.826	80.594	99.397	World Bank (2016)
gov. expenditure (GDP %)	81	33.011	9.753	12.478	54.168	World Bank (2016)
ln(area)	108	12.702	1.777	10.318	16.612	World Bank (2016)
n. of cities in stage 1	114	114.386	128.014	18.000	825.000	Own calculation using data from Brinkhoff (2014)
monopub	114	0.325	0.470	0.000	1.000	Ioannides et al. (2008)
monopriv	114	0.070	0.257	0.000	1.000	Ioannides et al. (2008)
time_after_public_monopoly	114	9.842	10.321	0.000	37.000	Ioannides et al. (2008)
time_after_private_monopoly	114	20.421	11.830	0.000	38.000	Ioannides et al. (2008)

Table 2: Fixed telephony regressions

VARIABLES	(1) ZipfZipf	(2) Zipf	(3) Zipf	(4) Zipf	(5) Zipf	(6) Zipf	(7) Zipf	(8) Zipf
ln(fixed)	-0.0978***	-0.0734*	-0.0037	-0.0638***	-0.0774***	-0.0015	-0.0023	-0.1012***
	(0.0190)	(0.0395)	(0.0117)	(0.0190)	(0.0286)	(0.0814)	(0.0122)	(0.0258)
Year	-0.0016	0.0011	-0.0030***	-0.0026**	-0.0019*	0.0024	-0.0030***	-0.0033***
	(0.0010)	(0.0020)	(0.0003)	(0.0010)	(0.0010)	(0.0023)	(0.0003)	(0.0008)
ln(population)		-0.0457*		0.1055		-0.0461*		0.1696**
		(0.0265)		(0.0954)		(0.0251)		(0.0822)
ln(GDPpc)		-0.0514		0.0752*		-0.0894		0.0933***
		(0.0436)		(0.0419)		(0.0562)		(0.0339)
GDPpc growth (st. dev.)		-0.0239		0.6412		-0.0222		101.7910
		(0.0168)		(0.3840)		(0.0160)		(78.1181)
Trade (GDP %)		0.0002		-0.0007		0.0002		-0.0007*
		(0.0008)		(0.0005)		(0.0008)		(0.0004)
non agriculture (GDP %)		-0.0100		0.0058		-0.0105		0.0083**
		(0.0101)		(0.0043)		(0.0096)		(0.0036)
gov. expenditure (GDP %)		-0.0057***		0.0031***		-0.0053***		0.0028***
		(0.0018)		(0.0010)		(0.0018)		(0.0008)
ln(area)		0.0389**		-4.0203*		0.0379***		-2.6930
		(0.0154)		(2.3828)		(0.0146)		(1.9883)
n. of cities in stage 1		-0.0002*		-0.0007***		-0.0002*		-0.0008***
		(0.0001)		(0.0001)		(0.0001)		(0.0001)
Country FE			Yes	Yes			Yes	Yes
Constant	2.2444	-1.2116	4.9280***	45.9584*	2.7482	-3.5874	4.9498***	-111.1068
	(1.9512)	(3.5989)	(0.6357)	(26.3979)	(2.0048)	(4.1565)	(0.5587)	(87.2169)
Observations	100	72	100	72	100	72	100	72
R-squared	0.2736	0.6148	0.9452	0.9915	0.2651	0.5938	0.9452	0.9907
Sargan					5.257	9.624	1.673	2.942
Chi-sq(1) P-val					0.154	0.0221	0.643	0.230
Weak identification					18.08	3.864	36.15	5.833

Standard errors in parentheses, Columns 1-4 are WLS regressions and 5-8 2SLS

IV for 5-8: monopub, monopriv, time_after_public_monopoly, time_after_private_monopoly

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Internet regressions

VARIABLES	(1) Zipf	(2) Zipf	(3) Zipf	(4) Zipf	(5) Zipf	(6) Zipf	(7) Zipf	(8) Zipf
ln(net)	-0.0201*	-0.0145	-0.0022	-0.0104***	-0.0250*	-0.0195	0.0017	-0.0144***
	(0.0106)	(0.0122)	(0.0035)	(0.0037)	(0.0146)	(0.0139)	(0.0045)	(0.0035)
Year	0.0023	0.0045*	-0.0022**	0.0003	0.0034	0.0052*	-0.0031***	0.0010
	(0.0028)	(0.0026)	(0.0009)	(0.0012)	(0.0036)	(0.0027)	(0.0011)	(0.0009)
ln(population)		-0.0409		-0.0112		-0.0424		-0.0171
		(0.0288)		(0.1010)		(0.0265)		(0.0727)
ln(GDPpc)		-0.0759*		0.1010**		-0.0765*		0.1145***
		(0.0433)		(0.0479)		(0.0396)		(0.0353)
GDPpc growth (st. dev.)		-0.0200		0.6899		-0.0201		153.2378*
		(0.0180)		(0.4941)		(0.0164)		(86.2043)
Trade (GDP %)		0.0005		-0.0002		0.0006		-0.0000
		(0.0008)		(0.0006)		(0.0008)		(0.0004)
non agriculture (GDP %)		-0.0110		0.0006		-0.0091		0.0016
		(0.0134)		(0.0063)		(0.0126)		(0.0046)
gov. expenditure (GDP %)		-0.0057***		0.0024**		-0.0060***		0.0019**
		(0.0020)		(0.0012)		(0.0018)		(0.0009)
ln(area)		0.0408**		-4.1594		0.0413***		-3.8856*
		(0.0158)		(3.0545)		(0.0144)		(2.2026)
n. of cities in stage 1		-0.0002*		-0.0006***		-0.0002*		-0.0006***
		(0.0001)		(0.0001)		(0.0001)		(0.0001)
Country FE			Yes	Yes			Yes	Yes
Constant	-5.9077	-8.0304	3.3888*	43.3946	-8.1224	-9.5915*	5.1627**	-175.3057*
	(5.6482)	(5.2587)	(1.7741)	(34.0861)	(7.2235)	(5.4751)	(2.1363)	(95.9495)
Observations	81	66	81	66	81	66	81	66
R-squared	0.0689	0.6282	0.9501	0.9916	0.0663	0.6271	0.9489	0.9913
Sargan					5.810	7.096	2.335	1.012
Chi-sq(1) P-val					0.121	0.0288	0.311	0.315
Weak identification					19.29	31.46	12.56	21.82

Standard errors in parentheses, Columns 1-4 are WLS regressions and 5-8 2SLS
IV for 5-8: monopub, monopriv, time_after_public_monopoly, time_after_private_monopoly *** p<0.01, ** p<0.05, * p<0.1

Table 4: Mobile telephony regressions

VARIABLES	(1) Zipf	(2) Zipf	(3) Zipf	(4) Zipf	(5) Zipf	(6) Zipf	(7) Zipf	(8) Zipf
ln(mobile)	-0.0259**	-0.0173	-0.0023	-0.0145***	-0.0147	-0.0284*	0.0014	-0.0156***
	(0.0114)	(0.0150)	(0.0036)	(0.0039)	(0.0157)	(0.0168)	(0.0037)	(0.0034)
year	0.0031	0.0048*	-0.0022**	-0.0000	0.0008	0.0062**	-0.0030***	0.0001
	(0.0028)	(0.0028)	(0.0008)	(0.0010)	(0.0035)	(0.0029)	(0.0008)	(0.0007)
ln(population)		-0.0417		-0.0181		-0.0459*		-0.0195
		(0.0291)		(0.0936)		(0.0269)		(0.0668)
ln(GDPpc)		-0.0854*		0.1135**		-0.0893**		0.1173***
		(0.0431)		(0.0450)		(0.0397)		(0.0327)
GDPpc growth (st. dev.)		-0.0197		0.8792*		-0.0201		207.0368***
		(0.0179)		(0.4538)		(0.0164)		(78.3142)
trade (GDP %)		0.0003		-0.0002		0.0003		-0.0002
		(0.0008)		(0.0005)		(0.0008)		(0.0004)
non agriculture (GDP %)		-0.0082		0.0046		-0.0046		0.0050
		(0.0130)		(0.0056)		(0.0123)		(0.0040)
gov. expenditure (GDP %)		-0.0053***		0.0026**		-0.0056***		0.0025***
		(0.0019)		(0.0010)		(0.0017)		(0.0008)
ln(area)		0.0382**		-5.2602*		0.0378***		-5.2738***
		(0.0157)		(2.8055)		(0.0144)		(2.0010)
n. of cities in stage 1		-0.0002*		-0.0005***		-0.0002*		-0.0005***
		(0.0001)		(0.0001)		(0.0001)		(0.0001)
Country FE			Yes	Yes			Yes	Yes
Constant	-7.4011	-8.7359	3.3025*	55.9028*	-2.8358	-11.8304**	4.8045***	-233.6414***
	(5.4901)	(5.5074)	(1.6635)	(31.1881)	(7.0037)	(5.7139)	(1.6368)	(87.2983)
Observations	82	67	82	67	82	67	82	67
R-squared	0.0866	0.6232	0.9492	0.9925	0.0754	0.6196	0.9482	0.9925
Sargan					8.067	6.951	3.970	0.560
Chi-sq(1) P-val					0.0446	0.0735	0.265	0.756
Weak identification					20.21	27.93	25.30	22.61

Standard errors in parentheses, Columns 1-4 are WLS regressions and 5-8 2SLS

IV for 5-8: monopub, monopriv, time_after_public_monopoly, time_after_private_monopoly

*** p<0.01, ** p<0.05, * p<0.1

Table 5.1: Fixed telephony regressions for using alternative measures of urban concentration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Zipf	-0.0978***	-0.0734*	-0.0037	-0.0638***	-0.0774***	-0.0015	-0.0023	-0.1012***
	(0.0190)	(0.0395)	(0.0117)	(0.0190)	(0.0286)	(0.0814)	(0.0122)	(0.0258)
Gini	-0.0079	-0.0083	-0.0168**	-0.0363***	0.0315	0.0643	-0.0234***	-0.0521***
	(0.0132)	(0.0295)	(0.0072)	(0.0098)	(0.0309)	(0.1076)	(0.0084)	(0.0122)
HHI	0.0089	0.0058	-0.0027	-0.0154	-0.0211	0.0187	-0.0060	-0.0383**
	(0.0109)	(0.0220)	(0.0050)	(0.0131)	(0.0256)	(0.0765)	(0.0059)	(0.0163)
CV	0.0564	-0.0081	-0.1433***	-0.1519**	0.4200	0.1784	-0.1432***	-0.1860***
	(0.1208)	(0.2643)	(0.0389)	(0.0569)	(0.2843)	(0.9221)	(0.0455)	(0.0687)

Table 5.2: Internet regressions using alternative measures of urban concentration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Zipf	-0.0201*	-0.0145	-0.0022	-0.0104***	-0.0250*	-0.0195	0.0017	-0.0144***
	(0.0106)	(0.0122)	(0.0035)	(0.0037)	(0.0146)	(0.0139)	(0.0045)	(0.0035)
Gini	-0.0017	-0.0055	-0.0046**	-0.0046**	0.0002	-0.0057	-0.0075***	-0.0061***
	(0.0069)	(0.0110)	(0.0018)	(0.0019)	(0.0110)	(0.0156)	(0.0020)	(0.0019)
HHI	0.0017	-0.0069	-0.0032**	-0.0050**	-0.0097	-0.0072	-0.0027*	-0.0044**
	(0.0056)	(0.0079)	(0.0013)	(0.0020)	(0.0092)	(0.0112)	(0.0014)	(0.0020)
CV	-0.0205	-0.0588	-0.0266**	-0.0186*	0.0179	-0.1315	-0.0440***	-0.0309***
	(0.0651)	(0.1013)	(0.0105)	(0.0105)	(0.1041)	(0.1445)	(0.0117)	(0.0107)

Table 5.3: Mobile telephony regressions using alternative measures of urban concentration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Zipf	-0.0259**	-0.0173	-0.0023	-0.0145***	-0.0147	-0.0284*	0.0014	-0.0156***
	(0.0114)	(0.0150)	(0.0036)	(0.0039)	(0.0157)	(0.0168)	(0.0037)	(0.0034)
Gini	-0.0039	-0.0094	-0.0036**	-0.0058**	-0.0006	-0.0158	-0.0057***	-0.0080***
	(0.0065)	(0.0116)	(0.0016)	(0.0023)	(0.0127)	(0.0167)	(0.0017)	(0.0023)
HHI	0.0037	-0.0084	-0.0011	-0.0034	-0.0087	-0.0128	-0.0022*	-0.0050**
	(0.0053)	(0.0083)	(0.0012)	(0.0025)	(0.0106)	(0.0119)	(0.0013)	(0.0024)
CV	-0.0045	-0.1035	-0.0273***	-0.0264**	0.0037	-0.2268	-0.0339***	-0.0350***
	(0.0620)	(0.1062)	(0.0093)	(0.0125)	(0.1207)	(0.1538)	(0.0095)	(0.0122)
For all regressions								
Country FE			Yes	Yes			Yes	Yes
Control variables	No	Yes	No	Yes	No	Yes	No	Yes

Standard errors in parentheses, Columns 1-4 are WLS regressions for the Zipf coefficient and OLS for the Gini coefficient, Hefrindahl index and coefficient of variation. 2SLS is used for columns 5-8 with the following IVs: monopub, monopriv, time_after_public_monopoly, time_after_private_monopoly. The Zipf coefficient estimates are duplicates from Table 5

*** p<0.01, ** p<0.05, * p<0.1

Table A1: Zipf coefficients based on the Gabaix-Ibragimov correction

Countries	2000	2001	2002	2005	2006	2007	2008	2009	2010	2011	2014	2015	2016	2017
Austria		-1.120								-1.116				
Belarus								-0.925						
Belgium	-1.714								-1.708					-1.702
Bulgaria		-1.272								-1.186				-1.136
Canada		-1.178			-1.168					-1.159			-1.143	
Denmark	-1.098				-1.113					-1.132				
Finland	-1.294								-1.261					-1.215
France						-1.559						-1.559		
Greece		-1.715								-1.849				
Hungary		-1.230								-1.232				-1.229
Italy		-1.433								-1.474			-1.471	
Mexico	-1.073			-1.100					-1.128		-1.145			
Netherlands		-1.292					-1.295			-1.286				
Norway	-1.426								-1.412					
Poland			-1.222							-1.245				-1.251
Portugal		-1.467								-1.410				
Russia			-1.232						-1.211					-1.185
Slovakia		-1.454								-1.456			-1.436	
Spain		-1.261								-1.296				-1.299
Sweden	-1.194			-1.189					-1.176			-1.167		-1.163
Switzerland	-1.510								-1.559					-1.560
UK		-1.443								-1.424			-1.413	
USA	-1.409								-1.444					-1.437

Table A2: Fixed telephony regressions for 1980-2000

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Zipf	Zipf	Zipf	Zipf	Zipf	Zipf	Zipf	Zipf
ln(fixed)	-0.0932***	-0.0667	-0.0057	-0.1307	-0.0669*	-0.0126	0.0176	-0.0480
	(0.0202)	(0.0930)	(0.0454)	(0.1102)	(0.0359)	(0.1136)	(0.0380)	(0.0620)
year	-0.0007	0.0034	-0.0038**	-0.0011	-0.0021	0.0022	-0.0044***	-0.0053
	(0.0018)	(0.0062)	(0.0014)	(0.0057)	(0.0024)	(0.0052)	(0.0012)	(0.0035)
ln(population)		-0.2228***		0.5692		-0.2111***		0.3120
		(0.0540)		(0.4696)		(0.0464)		(0.2417)
ln(GDPpc)		-0.3621***		-0.0001		-0.3819***		0.1200
		(0.0911)		(0.1903)		(0.0785)		(0.1121)
GDPpc growth (st. dev.)		-0.0027		-0.1385		0.0066		-0.0496
		(0.0343)		(0.1106)		(0.0309)		(14.2799)
trade (GDP %)		-0.0050**		-0.0002		-0.0050***		0.0007
		(0.0018)		(0.0013)		(0.0014)		(0.0007)
non agriculture (GDP %)		0.0392**		0.0129		0.0408***		0.0124***
		(0.0167)		(0.0087)		(0.0134)		(0.0040)
gov. expenditure (GDP %)		-0.0007		0.0027		-0.0012		0.0032***
		(0.0033)		(0.0024)		(0.0027)		(0.0012)
ln(area)		-0.0082		0.1459		-0.0093		0.2214
		(0.0319)		(0.3501)		(0.0252)		(0.4032)
n. of cities in stage 1		0.0007**		-0.0050**		0.0007**		-0.0041***
		(0.0003)		(0.0018)		(0.0003)		(0.0010)
Country FE			Yes	Yes			Yes	Yes
Constant	0.4155	-3.7361	6.4340**	-9.8539	3.0499	-1.7542	7.6888***	0.0000
	(3.6296)	(11.4485)	(2.6755)	(11.6292)	(4.6654)	(9.5555)	(2.1881)	(17.5689)
Observations	50	28	50	28	50	28	50	28
R-squared	0.4255	0.7865	0.9680	0.9983	0.4048	0.7823	0.9677	0.9980
Sargan					2.724	0.606	1.948	5.741
Chi-sq(1) P-val					0.436	0.895	0.583	0.0567
Weak identification					4.900	2.478	14.29	6.514

Standard errors in parentheses, Columns 1-4 are WLS regressions and 5-8 2SLS

IV for 5-8: monopub, monopriv, time_after_public_monopoly, time_after_private_monopoly

*** p<0.01, **
p<0.05, * p<0.1