

Tranos, Emmanouil; Ioannides, Yannis

Conference Paper

Ubiquitous digital technologies and spatial structure: a preliminary analysis

55th Congress of the European Regional Science Association: "World Renaissance: Changing roles for people and places", 25-28 August 2015, Lisbon, Portugal

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Tranos, Emmanouil; Ioannides, Yannis (2015) : Ubiquitous digital technologies and spatial structure: a preliminary analysis, 55th Congress of the European Regional Science Association: "World Renaissance: Changing roles for people and places", 25-28 August 2015, Lisbon, Portugal, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/124806>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Ubiquitous digital technologies and spatial structure: a preliminary analysis

Emmanouil Tranos, University of Birmingham, e.tranos@bham.ac.uk

Yannis Ioannides, Tufts University, yannis.ioannides@tufts.edu

Abstract

This paper sheds light on the potential effect that Information and Communication Technologies (ICT) might generate on cities and spatial structure. The extensive theoretical discussion and speculation on how cities and geography might be affected by digital technologies, which took place before the actual adoption of such technologies, have not been coupled by in depth empirical analysis to verify early predictions. The few examples of such studies, which approached such research questions both analytically and empirically, were insightful, but their results were to a certain extent contradictory. Most importantly, these studies took place before digital technologies such as the Internet had matured. Nowadays, these technologies have been adopted widely and we are thus in a better position to approach empirically such a research question and quantify the relation between ICTs and spatial structure. The preliminary empirical analysis presented in this paper suggests significant causal effects that ICT penetration generates on spatial structure. Internet and mobile phone penetration in non-EU/NAFTA countries have led to more spatially dispersed population and more uniform city size distribution. However, such effects are not present in non-EU/NAFTA countries, a phenomenon which might be related to the maturity of urban systems and advanced state of technological adoption in those countries. The proposed methodology, which relies on extensive econometric investigations with a number of models includes 2SLS regressions with instrumented variables, resulted to estimations which are robust against potential endogeneity problems.

1. Introduction

Substantial research effort has been spent on exploring spatial incidence of the Internet, even before its massive penetration, which took place during the last 15 years. Early research emphasized the Internet's non-spatial nature (Mitchell 1995). Geographers, economists but also technologists theorised about the spatial impacts that rapid Internet penetration might generate on cities and the spatial structure. The outcome of various such attempts was rather deterministic declaring the emergence of "telecottages" (Toffler 1980), the rise of a "borderless world" (Ohmae 1995), the death of cities (Gilder 1995; Drucker 1998; Kolko 1999), and, in general, the "end of geography" (O'Brien 1992), the "death of distance" (Cairncross 2001) and the emergence of a new flat world (Friedman 2005). However, such narratives had not been accompanied by empirical investigations and hard evidence. Although we know that urbanisation rates have been steadily increasing, especially for the developing world, and nowadays more than half of earth's population lives in cities, we still do not

know whether and how the Internet and other digital technologies have affected cities and spatial structure. Does the Internet and digital technologies, in general, act as a substitute or as a complement of agglomeration forces? In other words, has the vast penetration of the economy by the Internet affected the tendency of economic activities to be concentrated in urban agglomerations?

Early work by Gaspar and Glaeser (1998) suggested that telecommunication improvements will result in increased demand for face-to-face interactions (see also Brakman and Marrewijk 2008) and therefore, the importance of cities in spatial structure and economic activity, as centres of interaction, will also increase. Sinai and Waldfoegel (2004) wondered about the Internet's role as either a substitute or a complement for cities. Using data concerning the geographic scope of online information, they highlighted that the complementarity of local websites with local agglomeration overcomes any potential Internet substitution effect. In addition, Forman et al. (2005) concluded that despite the fact that the Internet adoption by firms with more than 100 employees was faster in smaller urban agglomerations, the adoption of more sophisticated Internet-based applications was positively related with city size in 2000. At a more aggregated level, Ioannides et al. (2008) examined the impact of ICT and more specifically fixed line telephony on urban structure. Using a panel dataset of spatial dispersion measures they found that contrary to the above presented studies, spatial dispersion of population is positively and causally affected by the fixed telephone penetration.

The main argument behind the above – early – research on the impact of digital technologies on spatial structure is the economic role that distance performs given the vast penetration of digital technologies. As Gaspar and Glaeser (1998) highlighted and as we know from urban economics, agglomeration economies are triggered by physical proximity (Partridge et al. 2008). Face-to-face social interactions are facilitated within cities due to the opportunities for decreased transportation costs. A key question is whether the generalised impact of the vast penetration of digital technologies on interaction cost, which is related with the decrease in spatial transmission costs (McCann 2013), is enough to offset the benefits derived by agglomeration economies and result in a more dispersed spatial structure. The early studies cited above proposed opposing patterns in regards to the dispersion or further concentration effects that the Internet and mobile technologies can generate. Nevertheless, the relevant studies did not incorporate in their analysis the full extent of the potential impact that digital technologies might generate either because of lack of appropriate data or because of the timing of the study, which was characterised by low Internet penetration rates, at least when compared to today's rates.

To summarise the previous research, it is fair to claim that the death of distance discussion, which is related with the weakening of agglomeration economies and the dominance of footloose economic activities, has been proved to be premature (Rietveld and Vickerman 2004). However, the exact impact of ICTs on spatial structure is still under question. Have agglomeration forces lost any of their glory because of the way people interact using digital tools? Have spatial interaction patterns been affected the increased opportunities for virtual interactions?

Research questions such as the above can now be empirically approached for two reasons. Firstly, ICT such as the Internet and mobile technologies, are no longer exotic technologies. On the contrary, their penetration rates have been stabilised at very high levels (see for example Figure 1), at least for the developed world and therefore it can be safely assumed that such technologies and consequently the potential impacts they might generate, have reached a maturity level. Secondly, because of the

maturity level, we now have enough available time series data which enables us to empirically approach the above research questions

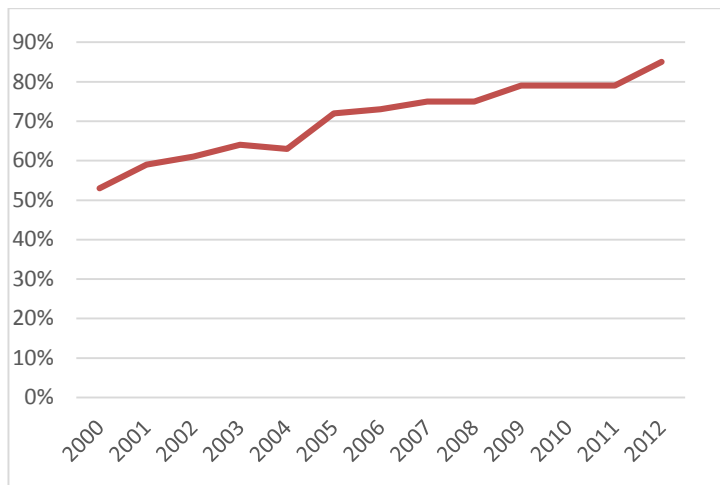


Figure 1: Internet use in the US, 2000-2012

Source: (Pew Internet 2014)

This paper aims to contribute to the above discussion by presenting empirical research on whether ICTs have affected spatial structure and more specifically the distribution of cities. In order to do so, a quantitative approach is adopted here. Section 2 describes the methods employed in this paper and Section 3 discusses the data used. Sections 4 and 5 present the preliminary results and the paper ends with a summary of our findings of the analysis and proposals for further research.

2. Methods

The main aim of this paper is to estimate the potential impact that the vast penetration of ICT might have generated on the spatial dispersion of economic activities and consequently population. In order to do so, we adopt a two-step approach, using the work of Ioannides et al. (2008) as a starting point. First step in our approach is to estimate the Zipf coefficient for a broad sample of countries over time. Zipf coefficient is one of the most widely used measure of spatial dispersion. City sizes satisfy Zipf's law, if

$$P(\text{size} > S) = \frac{\alpha}{S^\zeta}, \quad (1)$$

where $\zeta=1$ and α is a constant that is equal to the minimum city size raised to the power of ζ (e.g. Gabaix and Ioannides 2004). In other words, the percentage of cities with population greater than S equals to a constant α multiplied by the inverse population size S . An approximation of Zipf's law is the rank-size rule. According to this deterministic rule, the population of the second largest city within an urban system equals to the population size of the largest city divided by two. Similarly, the population of the third largest city equals to the population size of the largest city divided by three. Therefore, eq. (1) can be approximated by the following equation (Gabaix and Ioannides 2004):

$$S_i \approx \frac{S_0}{r}, \quad (2)$$

where, S_0 is a constant which is equal to the largest urban population of the urban system and r is the rank of the city i , the population of which S_i we are trying to estimate. The estimation of the logarithmic form of eq. (2) has been extensively used in the relevant literature which yields to the estimation of ζ , known as the Zipf coefficient:

$$\ln r_i = \ln S_0 + \zeta \ln S_i + e. \quad (3)$$

Based on the above discussion, Zipf rule holds when ζ is close to 1. More generally, estimations of city size distributions have also considered exponents ζ that are not necessarily equal to 1, in which case we refer to ζ as the Pareto, or power law, exponent. Given that our aim here is to estimate the Zipf coefficient for a number of countries over time as a measure of dispersion, equation (3) describes the rank of city i in country c at time t :

$$\ln r_{ict} = \ln S_{0ct} + \zeta_{ct} \ln S_{ict} + e_{ct}. \quad (4)$$

The above denotes the estimation of Zipf coefficient ζ for every country c in our sample in year t . The estimation of (4) has been traditionally based on OLS.

Zipf coefficients can also be approximated by the power law coefficient following the methodology proposed by Clauset, Shalizi, and Newman (2009). This methodology, which has been widely utilised in power law coefficient estimations for real world network node degree distributions, estimates a minimum city population size (x_{min}), above which the Pareto distribution applies. It is common that power laws and consequently the Pareto distribution do not apply for small cities. In order to estimate x_{min} , Clauset, Shalizi, and Newman (2009) suggest the use of the Kolmogorov-Smirnov (KS) statistic a commonly used measure of fit, which is the maximum distance between the real data and the Pareto distribution. The KS statistic is calculated for each unique value in the dataset and then x_{min} is defined as equal to the value which yields the smallest KS statistic. Then, the power law coefficient is estimated for $x \geq x_{min}$ using the method of *maximum likelihood* (ML). Following Adamic and Huberman (2002), the absolute value of Zipf coefficient equals the absolute value of the coefficient derived by the Clauset, Shalizi, and Newman (2009) methodology minus one, that is because the KS statistic works with the distribution function. For the purpose of this paper we refer to the estimated coefficient derived from the Clauset, Shalizi, and Newman (2009) method as power law coefficient.

The above alternative estimations of Zipf coefficients leads us to the second and main element of our identification strategy, which involves the regression of the country and year-specific Zipf (or power law) coefficients against a number of right hand-side variables in order to explain the determinants of spatial dispersion. Our main interest here is to estimate whether the ICT-related variables have an effect on Zipf coefficients. Following Ioannides et al. (2008) as well as Soo (2005) and Rosen and Resnick (1980) we are aiming to estimate here the following empirical equation:

$$zipf_{ct} = \theta_c + \delta t + X_{ct}\eta + \varepsilon_{ct}, \quad (5)$$

$zipf_{ct}$ represents the Zipf coefficient for country c in year t , θ_c is a country-specific constant or, in other words, a country fixed effect, δ is a linear time trend, X_{ct} is a vector of the determinants of the city size distribution for country c in year t , and η is a vector of parameters; ε_{ct} is the error term.

The availability of a panel dataset for city sizes across countries 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, ICT penetration might be affected by spatial structure, as reflected in Zipf coefficients, or both ICT penetration and spatial structure might be jointly determined by a third variable. E.g., if a country already has a dispersed spatial structure, ICT is particularly suitable in facilitating communication. Potential endogeneity in our specification will prevent us from being able to determine the causal impact of ICT penetration on spatial structure, which is the main aim of this paper. In order to address this problem, we will adopt an instrumented variable strategy, which will be further described below.

3. Data

Table 1 presents the descriptive statistics of the main variables utilised in this paper. *ols* and *ml* are the two main left hand-side variables and represent the Zipf and power law coefficient, as described above. These left hand-side variables are proxies for the city size dispersion, the spatial structure, as previously discussed and are regressed against a number of explanatory variables. The estimation of these variables was based on data on city population collected from Citypopulation (2014). Following Ioannides et al. (2008) data on city population was used instead of urban agglomeration to achieve higher consistency. As it can be seen in Table 1, the two different measures of population dispersion, which are based on the different estimation methods for Zipf coefficient, vary significantly. The reader is reminded that the absolute value of the power law coefficient (*ml*) equals to the OLS based coefficient (*ols*) plus one. After such a transformation, it becomes apparent that the mean values of these variables are very similar, but *ml* is more dispersed than the OLS based Zipf coefficient.

Table 1: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max	Source
<i>ml</i>	240	-2.187	0.344	-4.145	-1.640	Own calculation
<i>ols</i>	240	-0.945	0.221	-1.662	-0.258	Own calculation
ln(bbuser)	176	0.115	3.088	-8.537	3.686	(ITU 2014)
ln(netuser)	238	2.060	2.310	-6.948	4.554	(The World Bank 2014)
ln(mobile)	211	3.585	1.617	-3.982	5.243	(ITU 2014)
ln(pop)	240	16.566	1.425	12.547	21.014	(The World Bank 2014)
ln(gdppc)	230	8.961	1.294	5.743	10.786	(The World Bank 2014)
trade	235	79.738	37.272	16.750	220.407	(The World Bank 2014)
gov_exp	235	16.113	5.618	4.506	34.759	(The World Bank 2014)
nonagri	217	87.473	12.736	32.738	99.307	(The World Bank 2014)
ln(area)	240	12.506	1.697	7.616	16.612	(The World Bank 2014)
press	187	41.861	23.261	8.000	94.000	(Freedom House 2014)
polrights	238	2.866	2.035	1.000	7.000	(Freedom House 2014)
eu	240	0.338	0.474	0.000	1.000	Own calculation
nafta	240	0.046	0.210	0.000	1.000	Own calculation

To visually represent how ml differs from ols , the OLS based Zipf coefficient, Figure 2 presents the probability distribution function for city size in the US in 2000. The continuous red line represent the real data, the dotted red line ml coefficient when all the data is included and the dotted blue line represents the ml coefficient for these cities the population of which was higher or equal to x_{min} in 2000. This approach allow us to estimate the power law coefficient only for the upper tail of the distribution, which includes the cities with the highest population which usually are the cities which are known to follow closer a Pareto distribution. For example, as the blue line in Figure 2 indicates, the tail of the city size distribution in the US, which includes cities with less than 99,250 habitants, do not follow a Pareto distribution and therefore their inclusion in the sample would result in a significant change of the Zipf coefficient. Having said that, the estimation of x_{min} does not come without limitations.

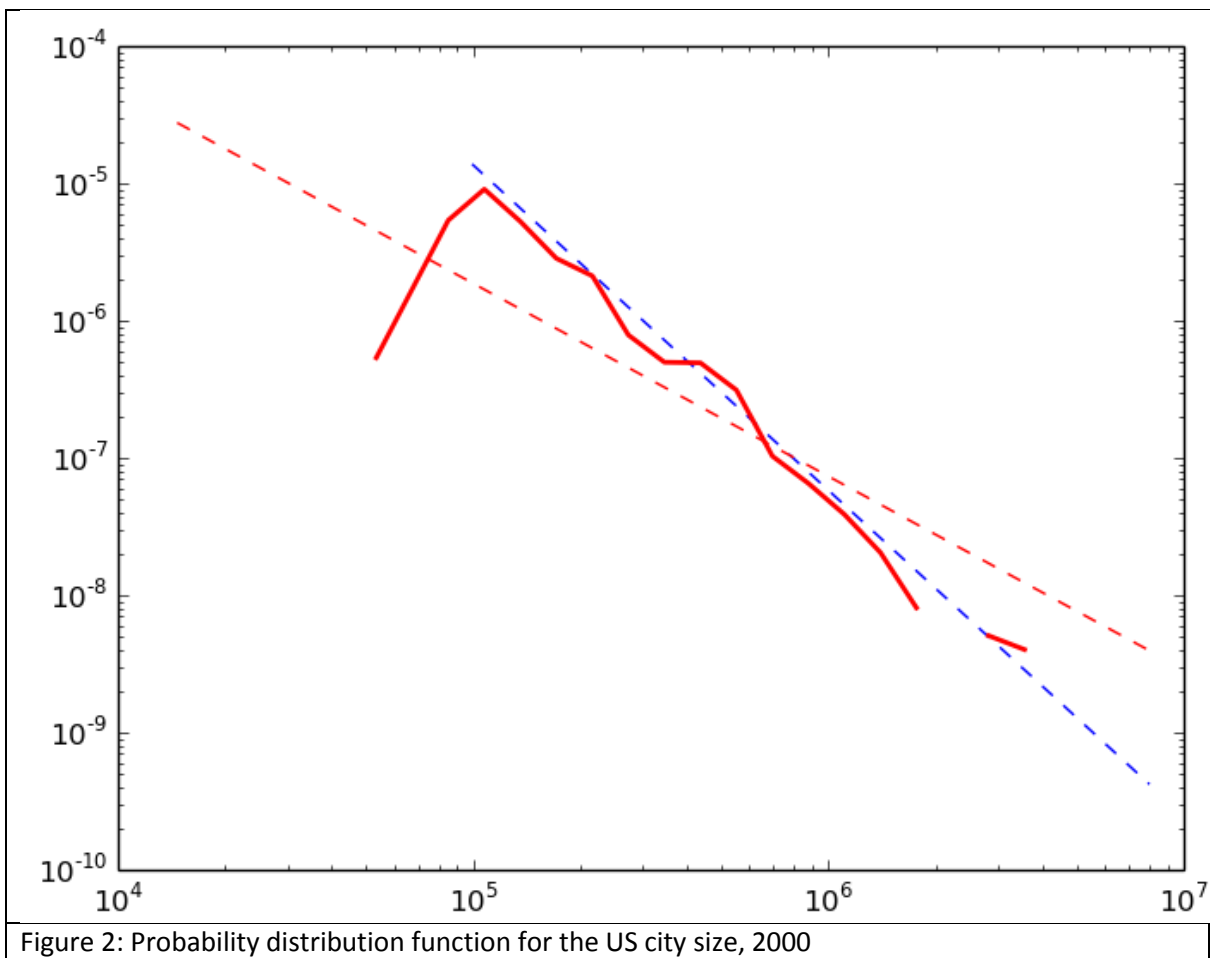


Table 1 also includes descriptive statistics for the explanatory variables included in equation 5 via X_{ct} . Among them, *netuser* and *mobile* represent the number of internet users and the number of mobile phone users per 100 habitants.

The main aim of the empirical analysis is to estimate the impact of these variables on spatial structure. As discussed above some of these variables, may be endogenous to the determination of city sizes, and thus to equation 5, and therefore instrumenting for these variables is necessary in order to conduct causal inference on the impact that ICT penetration might have on spatial structure. Then, a number of other covariates are also introduced to control for other determinants of spatial structure. These covariates include population (*pop*), Gross Domestic Product per Capita expressed in Purchased Power Units (*gdppc*), international trade (*trade*) which is the sum of exports and imports of goods and services measured as a share of gross domestic product, the government expenditure as a percentage of GDP (*gov_exp*) and the share of non-agricultural activities in the economy (*non_agri*).

4. Internet penetration and spatial structure

The next step in the analysis is the estimation of (5). The estimation results are reported in Table 2 using the Zipf coefficient, which has been estimated using the Zipf regression (eq. 4), as the LHS variable. All the available observations have been used for this estimation as no city size threshold was introduced. Column 1 in Table 2 presents the results of an OLS regression where the impact of ICT, proxied by $\ln(\textit{netuser})$, the number of Internet users per 100 inhabitants, is estimated. The regression also includes a year trend to control for changes in city size distribution over time. In this case, the Internet penetration does not have a significant impact on the city size distribution. The regression in Column 2 also includes a number of control variables which affect city size distribution. Given that the Zipf coefficient is inserted in the regression not as an absolute value but 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. In this context, wealthy large countries in terms of GDP per capita and population are related with more uniform city size distributions, while the opposite effect is observed for government expenditure and the share of non-agricultural activities. Two more control variables included in (2) are notable: *obs_avgpan* captures the potential effect of countries with high number of cities as it expresses the mean number of cities per country over time; in addition, *area* represents the areas of the country in square kilometres. Columns (3) and (4) report regressions as in (1) and (2) with the addition country fixed effects. Nevertheless, the main variable of interest, $\ln(\textit{netuser})$, remains statistically insignificant. In addition, most of the control variables lose their significance when country fixed effects are included.

However, the main challenge of estimating (5) is the potential endogenous nature of the share of Internet users which might prevent us from being able to infer a truly causal effect. Endogeneity might be an issue here as spatial structure, which is represented by the Zipf coefficient, might be affected by another source, which also affects Internet penetration. For instance, economic development might affect the concentration of population in large cities and at the same time enable more people to go online. If we do not address this issue, the coefficient for the main variable of interest will capture potential effects that Internet penetration has on spatial structure, but also potential reverse causality effects that spatial structure might generate on Internet penetration. To overcome this potential problem, column (5) reports estimates of eq. (5) using two-

stage least squares (2SLS) with instrumental variables (IVs). The latter refer to variables which are correlated with Internet penetration, but are not correlated with spatial structure. Such variables will enable the estimation of the causal effect – if any – of the Internet penetration on spatial structure. At a first stage, internet penetration is regressed against the IVs. Then, the predicted values of the Internet penetration based on the IVs and the other control variables are used instead of the endogenous variable from eq. (5). A significant effect will verify the causal impact of Internet penetration on spatial structure as the IVs are correlated with Internet penetrations, but are not themselves affected by changes in urban structure.

The main challenge for such an exercise is to find relevant IVs. Using the work of Jha and Sarangi (2014) as a starting point, we utilize here variables reflecting political rights (*polrights*) and freedom of press (*press*) as IVs in our regressions (see descriptive statistics in Table 1). The proposed underpinning mechanism is that the internet penetration is directly affected by freedom of press and political rights, but these variables are not related with spatial structure. So, the regression in column (5) mirrors the regression in column (1), but the endogenous variable is instrumented by the IVs described above.

The significant endogeneity test flags endogeneity as a problem, and therefore 2SLS estimations are preferred than the OLS ones. In addition, the significant underidentification test (LM test) enables the rejection of the test's null hypothesis according to which the model is under-identified. Also, the significant F test of the excluded instruments in the first stage indicates the relevance of the IVs. However, the Sargan test is significant for all 2SLS regressions (col. 5-8) posing doubts on the validity of the IVs. So, although endogeneity is a problem, our identification strategy did not manage to unpack the underlying relation.

Table 2: Internet users and city size distribution using OLS-based Zipf coefficient

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(netuser)	0.00102 (0.00990)	0.0110 (0.0133)	0.00149 (0.00429)	-0.00477 (0.00555)	-0.0342 (0.0211)	0.119 (0.101)	-0.0185 (0.0173)	-0.0532*** (0.0150)
year	-0.00580 (0.00392)	-0.00261 (0.00373)	-0.00179 (0.00122)	-0.000805 (0.00169)	-0.00667 (0.00567)	-0.0186 (0.0141)	0.00116 (0.00269)	-0.00407*** (0.000863)
ln(pop)		-0.126*** (0.0155)		-0.0585 (0.102)		-0.108*** (0.0227)		0.352*** (0.0631)
ln(gdppcpt)		-0.106*** (0.0274)		-0.0192 (0.0501)		-0.203** (0.0886)		0.0805 (0.0537)
trade		0.00115** (0.000473)		0.000605 (0.000423)		0.00180*** (0.000621)		0.00190*** (0.000335)
gov_exp		-0.00981*** (0.00316)		-0.00350 (0.00312)		-0.00737* (0.00383)		0.00154 (0.00173)
nonagri		0.00656** (0.00316)		0.00725* (0.00420)		0.00256 (0.00524)		0.0164*** (0.00382)
ln(area)		0.0868*** (0.00955)		-0.179 (1.478)		0.0832*** (0.0125)		-0.0219 (0.180)
obs_avgpan		0.0295 (0.146)				-0.0139 (0.192)		
FE			yes	yes			yes	yes
Constant	10.66 (7.848)	5.716 (7.545)	3.011 (2.440)	4.046 (19.92)	12.53 (11.34)	38.33 (28.95)	-2.893 (5.381)	0 (0)

Observations	206	206	206	206	156	156	156	156
R-squared	0.016	0.439	0.984	0.986	0.002	0.431	0.995	0.996
sargan					22.70	28.21	26.83	20.29
sargan_p					0.00005	0.00000	0.00000	0.00001
under					71.08	10.97	9.523	21.34
under_p					0	0.0269	0.00855	2.32e-05
endogeneity					6.963	0.537	3.318	59.11
endogeneity_p					0.00832	0.464	0.0685	0
F test of excluded instruments					31.39	2.704	1.593	3.408

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

The next step in the analysis is to use as the LHS variable the power law exponent (m) instead of the Zipf coefficient. Table 3 presents these regressions following the structure of Table 2. Again, our specifications identified endogeneity as a problem, but the IVs did not pass the relevant statistical tests.

Table 3: Internet users and city size distribution using ML-based Zipf coefficient

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(netuser)	0.00133 (0.0103)	0.0115 (0.0156)	-0.00671 (0.00557)	-0.0124* (0.00734)	-0.0323 (0.0233)	0.0772 (0.0964)	-0.0395** (0.0177)	-0.141 (0.0987)
year	-0.00690 (0.00423)	-0.00493 (0.00469)	0.000258 (0.00178)	0.00145 (0.00241)	-0.00154 (0.00716)	-0.0152 (0.0160)	0.00715** (0.00344)	0.0103 (0.00638)
ln(pop)		-0.121*** (0.0174)		-0.0282 (0.133)		-0.113*** (0.0225)		0.217 (0.366)
ln(gdppcpt)		-0.0995*** (0.0373)		0.0232 (0.0597)		-0.158* (0.0907)		0.177 (0.138)
trade		0.000255 (0.000507)		-0.000313 (0.000615)		-0.000144 (0.000739)		0.00124 (0.00138)
gov_exp		-0.0131*** (0.00356)		-0.00706 (0.00434)		-0.0100** (0.00415)		-0.0135** (0.00556)
nonagri		0.00900** (0.00364)		0.00310 (0.00509)		0.00603 (0.00508)		0.0283 (0.0188)
ln(area)		0.101*** (0.0127)		0.453 (1.900)		0.103*** (0.0162)		-2.194*** (0.544)
obs_avgpan		0.119 (0.151)				0.114 (0.190)		
FE			Yes	yes			yes	yes
Constant	11.79 (8.468)	8.821 (9.496)	-2.365 (3.567)	-10.63 (25.36)	1.142 (14.33)	29.77 (32.76)	-16.17** (6.888)	0 (18.14)
Observations	206	206	206	206	156	156	156	156
R-squared	0.019	0.339	0.975	0.976	-0.036	0.343	0.987	0.648
sargan					23.84	14.71	1.374	320.2
sargan_p					2.69e-05	0.00209	0.241	0
under					61.37	16.19	23.89	14.91
under_p					0	0.00277	6.50e-06	0.000577
endogeneity					5.902	0.275	0.813	14.72
endogeneity_p					0.0151	0.600	0.367	0.000125

F test of excluded instruments

24.32 4.141 4.430 2.273

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

As an effort to overcome the above misspecification of our IV strategy, we are proposing here a sample split. Arguably, the initial sample of countries was very heterogeneous including countries with very different levels of urban and technological maturity as far as it concerns their urban systems. Therefore, the initial sample is split in two parts: one which includes EU/NAFTA countries and one which includes non-EU/NAFTA countries. Table 4 presents regression (4) and (8) from the above tables for the two sub-samples and also for the two different LHS variables.

Table 4: Regressions for EU & NAFTA countries and non-EU, non-NAFTA countries

LHS variable:	(1) ml	(2) ml	(3) ml	(4) ml	(5) ols	(6) ols	(7) Ols	(8) ols
ln(netuser)	-0.00907 (0.0112)	-0.112 (0.0742)	-0.00716 (0.0135)	-0.121*** (0.0263)	0.00804 (0.00922)	-0.193*** (0.0516)	-0.0173* (0.00906)	-0.0272* (0.0145)
Year	0.00322 (0.00299)	0.00247 (0.00553)	0.000532 (0.00653)	0.0280*** (0.00510)	-0.000639 (0.00211)	0.0185*** (0.00303)	0.00224 (0.00500)	-0.0120*** (0.00258)
ln(pop)	-0.202 (0.209)	0.107 (0.577)	0.0493 (0.272)	-0.661*** (0.248)	-0.271 (0.173)	-0.866*** (0.323)	-0.0228 (0.229)	0.643*** (0.127)
ln(gdppcppt)	-0.0757 (0.118)	0.253 (0.183)	0.0233 (0.0917)	0.0406 (0.0962)	-0.0209 (0.0924)	-0.191 (0.133)	-0.0725 (0.0819)	0.0897* (0.0531)
trade	-0.00123 (0.000938)	0.00248 (0.00226)	-0.000283 (0.000891)	0.00136* (0.000744)	-0.000160 (0.000699)	0.00106 (0.00125)	0.000670 (0.000647)	0.000634* (0.000358)
gov_exp	-0.0105 (0.00901)	0.00419 (0.0141)	-0.00632 (0.00578)	-0.0142*** (0.00387)	-0.00581 (0.00632)	-0.00578 (0.00873)	-0.00201 (0.00451)	0.00404** (0.00175)
nonagri	0.00705 (0.00890)	0.0292 (0.0288)	-0.000445 (0.00739)	0.0174** (0.00692)	0.000960 (0.00779)	0.0361** (0.0169)	0.0126* (0.00633)	0.00942** (0.00451)
ln(area)	-22.02*** (5.839)	-0.310 (1.380)	1.551 (2.390)	-3.599*** (0.624)	0.484 (3.411)	-2.533*** (0.153)	-0.654 (1.861)	0.824** (0.331)
FE	yes	yes	yes	yes	yes	yes	Yes	yes
Constant	220.1*** (60.68)	-9.754 (0)	-24.57 (33.08)	0 (0)	-0.248 (36.02)	0 (6.658)	3.657 (25.88)	0 (0)
Observations	78	60	128	96	78	60	128	96
Sample	EU & NAFTA		non EU & NAFTA		EU & NAFTA		non EU & NAFTA	
R-squared	0.994	0.953	0.954	0.981	0.997	0.999	0.950	0.991
sargan		85.12		0.529		3.435		6.476
sargan_p		0		0.467		0.0638		0.0109
under		2.187		23.24		8.692		24.09
under_p		0.335		8.97e-06		0.0130		5.88e-06

endogeneity	81.27	13.27	-1.747	7.656
endogeneity_p	0	0.000270	1	0.00566
F test of excluded instruments	0.303	2.875*	1.355	3.015*

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Regressions in columns (1)-(4) use the power law coefficient (*ml*) as the LHS variable, while regression (5)-(8) use the OLS-based Zipf coefficient (*ols*). Columns (1), (2), (5) and (6) include only the EU and NAFTA countries. Coefficients in columns (1), (3), (5) and (7) are estimated using OLS, while the other regressions are based on 2SLS. Starting with the regressions using *ml* as the LHS variable, we observe again a non-significant effect for columns (1)-(3). However, the effect of the internet penetration appears to be significant and negative only when we work with the non-EU, non-NAFTA countries and 2SLS is utilised. Moreover, our instrumentation variables strategy seems to work for this case: Sargan test is not significant and the null hypothesis of the validity of the instruments cannot be rejected. Underidentification is not a problem and endogeneity can be detected in our specification. Moreover, the F test of the excluded instruments is marginally significant indicating rather relevant IVs. The use of OLS-based Zipf coefficients as the LHS variable (columns 5-8) did not result to any fruitful results. In a nutshell, the above analysis results to a significant and causal negative effect of the Internet penetration on the power law coefficient, when only the non-EU and non-NAFTA countries are included in the analysis. In other words, Internet penetration appears to have decreased the variance of city sizes and result to an increase of the spatial dispersion of population in non-EU/NAFTA countries.

In order to address the potential critique against the endogenous nature of the estimated threshold utilised for the *ml* coefficient, we exogenously impose a cut-off value for the estimation of the Zipf coefficient. Table 5 below replicates the basic regressions (e.g. columns 4 and 8 from Table 2), but in this case the LHS variable is the OLS-based coefficient estimated for the cities which have population equal or higher than 50,000. As before, both the results of WLS and 2SLS are presented for all the countries in our sample, for the EU and NAFTA countries and for non-EU and non-NAFTA countries. According to the relevant test, endogeneity is not a problem here and therefore we can use the WLS results. What is interesting here is that the results for the specifications where the LHS variable is based on an exogenously imposed population threshold matches the previously presented results from Table 4, where *ml* was the LHS variable. Again, the only significant effect that was diagnosed was the negative effect that the share of Internet users has on the Zipf coefficient for the non-EU and non-NAFTA countries.

Table 5: Regressions for Zipf coefficient for cities > 50.000

	(1) all countries	(2)	(3) EU & NAFTA	(4)	(5) non-EU & non-NAFTA	(6)
ln(netuser)	-0.00447 (0.00428)	0.628*** (0.0960)	-0.00641 (0.00556)	-0.0148 (0.0299)	-0.0159** (0.00694)	-0.00884 (0.0103)
year	-0.00216 (0.00166)	0.0316*** (0.00663)	0.00194 (0.00129)	-6.11e-05 (0.000985)	0.00995* (0.00523)	0.00200 (0.00170)

ln(pop)	-0.125 (0.107)	-6.238*** (0.513)	-0.158 (0.111)	-0.0546 (0.159)	-0.547** (0.216)	0.0132 (0.0598)
ln(gdppcppt)	-0.0499 (0.0394)	-0.278 (0.538)	-0.0420 (0.0586)	0.137*** (0.0519)	-0.170** (0.0706)	-0.0341 (0.0298)
trade	0.00161*** (0.000412)	0.0187*** (0.00359)	0.000191 (0.000438)	0.000715 (0.000813)	0.00208*** (0.000587)	0.000465** (0.000212)
gov_exp	-0.00528 (0.00352)	0.0961*** (0.0298)	0.0230*** (0.00404)	0.0143*** (0.00351)	-0.00609 (0.00527)	-0.00102 (0.00124)
nonagri	0.0115** (0.00442)	-0.155*** (0.0466)	0.000425 (0.00557)	-0.00365 (0.00837)	0.0174** (0.00666)	0.000608 (0.00363)
ln(area)	-0.118 (1.018)	4.285*** (0.796)	1.020 (2.358)	0.0553 (0.286)	0.697 (1.238)	-0.349 (0.537)
FE	yes	yes	yes	yes	Yes	yes
Constant	6.922 (13.56)	0 (0)	-12.39 (24.59)	-1.412 (0)	-20.66 (18.62)	0 (6.790)
Observations	198	151	73	56	125	95
R-squared	0.993	0.998	0.999	0.993	0.970	0.998
sargan		0.124		294.4		0.735
sargan_p		0.725		0		0.391
under		9.183		6.085		13.30
under_p		0.0101		0.0477		0.00130
endogeneity		-5.686		-0.973		-0.0259
endogeneity_p		1		1		1
F test of excluded instruments		1.327		0.914		1.383

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

5. Mobile telephony and spatial structure

In order to further investigate the effect of ICT on spatial structure, the share of mobile phone users is utilised as the main RHS variable of interest. Table 6 replicates the regressions presented in Table 4, but the RHS variable *netuser* has been replaced by the variable *mobile*, which indicates the number of mobile phone users per 100 habitants. As before, the sample is split to EU/NAFTA and non-EU/NAFTA countries. Two different LHS variables are utilised as above: the power law coefficient based on ML (*ml*) and the OLS-based Zipf coefficient (*ols*). Starting with the former, a significant negative coefficient is estimated using the 2SLS for the non-EU/NAFTA countries, which is actually the only significant coefficient estimated for the *ml* RHS variable. As discussed before, endogeneity is the main concern, but according to the relevant tests this specification addresses this issue. As indicated by the endogeneity test, endogeneity is indeed a problem here and therefore 2SLS is preferred than OLS. Our IVs are valid and relevant as indicated by the non-significant Sargan

test and the significant F test. Moreover, underidentification is not a problem according to the relevant test.

The results are more difficult to interpret when the LHS variable is the OLS-based Zipf coefficient. When the focus is on the EU/NAFTA countries (col. 5 and 6 in Table 6), the coefficient for the variable indicating mobile phone penetration is significant and negative. Endogeneity is again a problem, as indicated by the relevant test, but the IVs included in regression (6) address this problem and enable us to argue that there is a causal negative effect of mobile phone penetration on the OLS-based Zipf coefficient and thus on the spatial structure. However, when only the non-EU/NAFTA countries are included in the analysis, then our IV strategy does not result to valid instruments (significant Sargan test). Therefore, the significant negative coefficient of the variable *mobile* in columns (7) and (8) in Table 6 might be biased because of endogeneity issues. In a nutshell, although the regressions using the power law coefficient as a measure of spatial structure result in similar effects for both Internet and mobile phone penetration, this is not the case for the regression using the OLS-based Zipf coefficient, which result in opposing effects.

Table 6: Regressions for EU & NAFTA countries and non-EU, non-NAFTA countries for mobile phone users

LHS variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ml	ml	ml	ml	ols	ols	ols	ols
ln(mobile)	0.00178 (0.0240)	0.0339 (0.0458)	-0.0192 (0.0164)	-0.0650*** (0.0136)	-0.0936*** (0.0167)	-0.667*** (0.111)	-0.0335*** (0.0114)	-0.0165* (0.0100)
Year	0.00262 (0.00385)	0.00442 (0.00299)	2.45e-05 (0.0101)	0.0166*** (0.00457)	0.00464** (0.00177)	0.0460*** (0.00688)	0.000865 (0.00584)	-0.0141*** (0.00242)
ln(pop)	-0.309 (0.310)	-0.580 (0.374)	0.0338 (0.513)	-0.623** (0.242)	-0.344* (0.169)	-0.113 (0.456)	0.231 (0.308)	0.644*** (0.125)
ln(gdppcpt)	-0.151 (0.154)	-0.0687 (0.186)	0.0537 (0.153)	0.177 (0.108)	0.166** (0.0770)	0.492* (0.299)	-0.132 (0.107)	0.115** (0.0568)
Trade	0.000911 (0.00112)	-0.000933 (0.000924)	-0.000200 (0.00104)	0.000716 (0.000657)	-0.000562 (0.000486)	-0.00482*** (0.00141)	0.000197 (0.000665)	0.000527 (0.000334)
gov_exp	-0.00627 (0.00928)	-0.00859 (0.0106)	-0.00349 (0.00799)	-0.0147*** (0.00375)	-0.00371 (0.00441)	-0.0521*** (0.0165)	-0.00401 (0.00484)	0.00392** (0.00169)
Nonagri	0.00766 (0.0103)	-0.0101 (0.0114)	0.00575 (0.0114)	0.0105* (0.00639)	0.00434 (0.00534)	0.0632*** (0.0220)	0.0314*** (0.00787)	0.00778* (0.00409)
ln(area)	23.12*** (5.505)	0.411 (0.336)	-2.916 (7.407)	-1.961*** (0.433)	-2.104 (2.253)	-9.541*** (1.504)	-2.812 (3.500)	1.143 (1.698)
FE	yes	yes	yes	yes	yes	yes	yes	yes
Constant	234.6*** (57.07)	-4.801 (5.748)	35.87 (102.0)	0 (0)	15.19 (22.96)	0 (0)	30.10 (48.19)	0 (22.02)
Observations	71	60	109	96	71	60	109	96
Sample	EU + NAFTA		non EU + NAFTA		EU + NAFTA		non EU + NAFTA	
R-squared	0.995	0.993	0.964	0.982	0.999	0.999	0.973	0.992
Sargan		37.14		0.123		0.240		8.525
sargan_p		1.10e-09		0.725		0.624		0.00350

Under	53.54	30.95	42.96	17.91
under_p	0	1.91e-07	4.68e-10	0.000129
endogeneity	-527.0	13.71	111.0	2.861
endogeneity_p	1	0.000214	0	0.0908
F test of excluded instruments	66.27	4.281**	20.18	2.064

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

To further address this issue, Table 7 utilises the OLS-based Zipf coefficient estimated only for these cities with population above 50,000 habitants. In essence, Table 7 reproduces the regressions presented in Table 5, but the main RHS variable of interest here is the mobile phone penetration. When all the countries are included in the analysis, endogeneity does not appear to be a problem given the non-significant endogeneity test (col. 2) and therefore we can accept the OLS estimation (col. 1), according to which mobile phone penetration has a negative effect on Zipf coefficient. However, such an effect is not apparent when only the EU/NAFTA countries are included in the analysis (col. 3 and 4 in Table 7), but it is present when sample only includes non-EU/NAFTA countries (col. 5 and 6). While the latter agrees with the previous results, this is the only time that a significant effect, which is not potentially biased by endogeneity issues, becomes evident for *all* the countries. What is interesting though is that no significant effect has been observed for the EU and NAFTA countries, a finding which is in agreement with the previous results.

VARIABLES	(1) all countries	(2)	(3) EU & NAFTA	(4)	(5) non-EU & non-NAFTA	(6)
ln(mobile)	-0.0336*** (0.00575)	-0.00930** (0.00419)	0.00665 (0.0199)	-0.00191 (0.0103)	-0.0355**: (0.00698)	-0.00237 (0.00459)
Year	-0.00121 (0.00153)	-0.000444 (0.00114)	-0.000748 (0.00185)	-0.000307 (0.000586)	0.00892* (0.00477)	0.000986 (0.00137)
ln(pop)	0.146 (0.113)	-0.0436 (0.0363)	-0.161 (0.207)	-0.0629** (0.0263)	-0.179 (0.239)	0.00581 (0.0549)
ln(gdppcppt)	-0.0269 (0.0435)	0.0659* (0.0398)	0.0614 (0.0776)	0.123*** (0.0331)	-0.176** (0.0777)	-0.0297 (0.0304)
Trade	0.00139*** (0.000406)	0.000509** (0.000255)	0.000722 (0.000550)	0.000465** (0.000196)	0.00123** (0.000530)	0.000546*** (0.000190)
gov_exp	-0.00717* (0.00379)	-0.00345* (0.00202)	0.0189*** (0.00527)	-0.0164*** (0.00162)	-0.0108* (0.00537)	-0.000883 (0.00132)
Nonagri	0.0237*** (0.00495)	-0.000808 (0.00338)	0.00166 (0.00635)	-0.00689** (0.00285)	0.0309*** (0.00643)	-0.000848 (0.00325)
ln(area)	-0.977 (1.474)	0.0487 (0.144)	3.653 (2.602)	0.0717 (0.0957)	-0.443 (1.773)	-0.182 (0.969)
FE	yes	yes	yes	yes	yes	yes
Constant	10.94 (19.84)	0 (0)	-34.96 (25.80)	-0.568 (0)	-10.43 (25.07)	0 (13.12)

Observations	173	151	66	56	107	95
R-squared	0.992	0.998	0.998	0.999	0.986	0.998
Sargan		44.22		18.00		0.865
sargan_p		0		2.20e-05		0.352
Under		30.39		40.46		24.18
under_p		2.52e-07		1.64e-09		5.63e-06
endogeneity		0.514		444.5		0.0330
endogeneity_p		0.474		0		0.856
F test of excluded instruments		5.165		19.53		2.901

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

6. Discussion and further research

The above analysis resulted to some preliminary, but still interesting patterns. ICTs penetration, as reflected in Internet and mobile phone penetration appears to have a significant effect on the spatial structure of countries which are not part of what one could call the 'Western World'. This effect is negative which indicates an impact towards the increase of the spatial dispersion of population. In other words, this negative effect results to more uniform city size distribution. Although the above results are preliminary, the effects appears to be robust against different specifications and different proxies of spatial pattern. In addition, the above analysis appears to be robust against potential endogeneity problems as the proposed IVs managed to address this issue in most of the cases.

Conceptually, the above preliminary results add to the existing literature in various ways. Firstly, the econometric analysis indicates that ICTs had an effect on spatial structure. Although Internet and mobile phone access are almost universal nowadays, their penetration rates appeared to have had a significant causal effect towards more balanced urban systems. However, this effect is not universal. On the contrary, no such effect was apparent for western countries, while such effected was flagged for non-EU/NAFTA countries. A possible explanation of the spatial heterogeneity of the effects of ICTs on spatial structure might be the combination of the level of maturity of urban systems and ICTs in western countries. Urbanisation and agglomeration forces are historically much more a western story as they played more pivotal role in the growth of western countries (see discussion in McCann 2013). Therefore, agglomeration forces have been carving cities and urban systems in western countries for centuries. On the contrary, very high urbanisation rates and strong agglomeration forces are a much more recent phenomenon in non-western countries. This difference in the levels of maturity of urbanisation and agglomeration forces might be part of the explanation of the patterns emerged from the above analysis.

In the same vein, technological improvements are historically rooted in western countries. First long haul communication systems, such as the telegraph and fixed line telephony, but also more recent infrastructural systems such as the submarine cable, which form the core of the Internet, were firstly established in the western world. Hence, it is not surprising that the recent introduction of ubiquitous technologies, such as the Internet and mobile telephony, resulted to more radical effects in non-western countries.

More research is necessary in order to further validate the above analysis. For instance, different measures of spatial dispersion should also be tested as LHS variables as a sensitivity analysis. In addition, more effort should be spent in unpacking the spatial heterogeneity of the effect of ICTs on spatial structure either in terms of interactions but also in terms of different sub-sampling of our dataset. It might also be useful to test whether such heterogeneity is present on the potential effect that older technologies, such as fixed telephony, had on spatial structure.

References

- Adamic, Lada A. and Bernardo A. Huberman. 2002. "Zipf's law and the Internet." *Glottometrics* 3, 143-150.
- Brakman, Steven and Charles van Marrewijk. 2008. "It's a big world after all: on the economic impact of location and distance." *Cambridge Journal of Regions, Economy and Society* 1, 411-437.
- Cairncross, Frances. 2001. *The death of distance 2.0*. London: Texere Publishing Limited.
- Citypopulation. 2014. "City Population," in.
- Clauset, Aaron, Cosma Rohilla Shalizi and M. E. J. Newman. 2009. "Power-law distributions in empirical data." *SIAM Review* 51, 661-703.
- Drucker, Peter F. 1998. "From Capitalism to Knowledge Society," in D. Neef (ed.), *The knowledge economy* Woburn, MA: Butterworth-Heinemann. pp. 15-34.
- Forman, Chris, Avi Goldfarb and Shane Greenstein. 2005. "How did location affect adoption of the commercial Internet? Global village vs. urban leadership." *Journal of Urban Economics* 58, 389-420.
- Freedom House. 2014. "Freedom in the World," in.
- . 2014. "Freedom of the Press," in.
- Friedman, T. L. 2005. *The world is flat*.
- Gabaix, Xavier and Yannis M Ioannides. 2004. "The evolution of city size distributions." *Handbook of regional and urban economics* 4, 2341-2378.
- Gaspar, Jess and Edward L. Glaeser. 1998. "Information Technology and the Future of Cities." *Journal of Urban Economics* 43, 136-156.
- Gilder, George. 1995. *Forbes ASAP*. pp. 56.
- Ioannides, Yannis M, Henry G Overman, Esteban Rossi-Hansberg and Kurt Schmidheiny. 2008. "The effect of information and communication technologies on urban structure." *Economic Policy* 23, 201-242.
- ITU. 2014. "Statistics," in.
- Jha, Chandan and Sudipta Sarangi. 2014. "Social Media, Internet and Corruption." *Working Paper*.
- Kolko, Jed. 1999. "The Death of Cities? The Death of Distance? Evidence from the Geography of Commercial Internet Usage," paper presented at Selected Papers from the Telecommunications Policy Research Conference 1999 Newcastle.
- McCann, Philip. 2013. *Modern urban and regional economics*: Oxford University Press.
- Mitchell, William J. 1995. *City of bits: space, place and the infobahn*. Cambridge, MA: MIT Press.
- 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.
- Partridge, Mark D., Dan S. Rickman, Kamar Ali and M. Rose Olfert. 2008. "Employment Growth in the American Urban Hierarchy: Long Live Distance." *The B.E. Journal of Macroeconomics* 8, Article 10.
- Pew Internet. 2014. "Pew Internet & American Life Project." Pew Internet.
- Rietveld, Piet and R. W. Vickerman. 2004. "Transport in regional science: The "death of distance" is premature." *Papers in Regional Science* 83, 229-248.

- Rosen, Kenneth T. and Mitchel Resnick. 1980. "The size distribution of cities: An examination of the Pareto law and primacy." *Journal of Urban Economics* 8, 165-186.
- Sinai, Todd and Joel Waldfoegel. 2004. "Geography and the Internet: is the Internet a substitute or a complement for cities?" *Journal of Urban Economics* 56, 1-24.
- Soo, Kwok Tong. 2005. "Zipf's Law for cities: a cross-country investigation." *Regional Science and Urban Economics* 35, 239-263.
- The World Bank. 2014. "Indicators," in.
- Toffler, Alvin. 1980. *Third way*. New York: William Morrow.