



Lancaster University
Management School

Economics Working Paper Series

2015/027

Innovation across cities

Kwok Tong Soo

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

© Authors

All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission, provided that full acknowledgement is given.

LUMS home page: <http://www.lancaster.ac.uk/lums/>

Innovation across cities

Kwok Tong Soo¹
Lancaster University

December 2015

Abstract

This paper examines the distribution of patenting activity across cities in the OECD, using a sample of 218 cities from 2000 to 2008. We obtain three main results. First, patenting activity is more concentrated than population and GDP. Second, patenting activity is less persistent than population and GDP. Third, patenting exhibits mean-reversion, and is positively associated with GDP, the fragmentation of local government, and population density. Our results suggest that policymakers can influence the amount of innovative activity through the use of appropriate policies.

JEL Classification: R1, O3.

Keywords: Patents; Zipf's Law; transition probability; dynamic panel data.

¹ Department of Economics, Lancaster University Management School, Lancaster LA1 4YX, United Kingdom. Tel: +44(0)1524 594418. Email: k.soo@lancaster.ac.uk

“The mysteries of the trade become no mysteries; but are as it were in the air...”

Alfred Marshall, *Principles of Economics*, Chapter 10.

1. Introduction

Since at least Marshall (1890) it has been argued that forces of agglomeration may lead to the formation of industrial clusters, and by extension, cities. As has been discussed in greater detail elsewhere (Krugman, 1991, Fujita et al, 1999), Marshall identified three reasons for the spatial concentration of economic activity: knowledge spillovers, thick markets for specialised skills, and the backward and forward linkages associated with large local markets. Because of the presence of knowledge spillovers, cities are not only the centre of economic activity, but also the focal point of innovative activity. Indeed, if it is argued that innovative activity makes use of all three of Marshall’s external economies, then innovative activity should be even more concentrated than economic activity in general. Anecdotal evidence supports this idea; for instance, Tokyo, the largest city in our sample, in 2008 had 27 percent of Japan’s population, but 32.3 percent of GDP, and 34.3 percent of the number of patents.

This paper explores the distribution of patenting activities across cities, the persistence and growth of patenting in cities, and the determinants of patenting activity. In so doing, we make use of methods developed for the analysis of city populations, and city population thus acts as a useful benchmark to compare with our analysis of patents. We make use of a sample of 218 cities from OECD countries, from 2000 to 2008, and obtain three main results. First, patenting is more unevenly distributed across cities than population or GDP. Second, patenting is less persistent than both population and GDP. Third, even after controlling for the endogeneity of some explanatory variables, the number of patents is positively associated with GDP, population density, and the degree of local government fragmentation. Taken together our results suggest that it may be possible for policymakers to implement policies that encourage innovation in cities.

This paper is related to three strands of the literature. First, the literature on the production of knowledge in cities is discussed in Audretsch and Feldman (1996, 1999) and has been surveyed in Audretsch and Feldman (2004). This line of research is mainly focussed on the impact of industrial concentration and diversity on the productivity of R&D (“spillovers”). A closely related line of work in Glaeser et al

(1992, 1995) investigates the effects of different industrial composition on economic growth in cities. Unlike this literature, our focus is not on R&D spillovers, but rather on the distribution of innovation across cities, and the factors that may explain the distribution.

There is also an associated branch of the literature which examines innovative and creative activities in cities. This includes OhUallachain (1999), Berry and Glaeser (2005), Bettencourt et al (2007, 2010), and Strumsky and Thill (2013). However, much of this literature focuses on US cities, and is primarily interested in describing the distribution of innovative activity across cities. In the present paper, we use an international dataset comprising the largest cities in the OECD, thus allowing us to see whether any trends that we observe operate across national boundaries. In addition, whilst we are also interested in how innovative activity is distributed across cities, we extend the analysis to consider the persistence and evolution of innovative activity over time.

Methodologically, since the paper presents evidence on the distribution and growth of innovation in cities, it is related to the literature on the size distribution and growth of cities, as discussed in Gabaix and Ioannides (2004), Eaton and Eckstein (1997), Black and Henderson (2003), Dobkins and Ioannides (1999, 2001), Ioannides and Overman (2001, 2003, 2004), Soo (2005, 2007), and Bosker et al (2008) . On the distribution of innovation in cities, we make use of the concept of Zipf's Law (Zipf, 1949), that the size of cities follows a Pareto distribution. Gabaix and Ibragimov (2011) develop a simple way of improving the performance of OLS estimates of Zipf's Law. On the persistence of innovation in cities, we make use of the concept of transition probability matrices. Finally, on the growth of innovation over time, we make use of both parametric and non-parametric approaches to describe the growth patterns of innovation, and the determinants of innovative growth.

The next section discusses the data used in this paper. This is followed in Section 3 by the analysis of the distribution of innovative activity, in Section 4 by the persistence of innovative activity, in Section 5 by the growth of innovative activity, and in Section 6 by the determinants of innovative activity. Because of the wide range of methods used, they will be discussed within each section to maximise clarity. The final section concludes.

2. Data

The data is obtained from the OECD Metropolitan Database, which contains data for metro areas with a population of 500,000 or more across OECD countries. Metro areas are defined following a harmonised functional definition developed by the OECD in OECD (2012). There are a total of 275 cities from 28 OECD countries. Patent data is available for 218 metro areas from 16 countries from 2000 to 2008, and represents a count of the number of patent applications by the city of the inventor². The dataset also includes other variables, such as population, geographical and administrative information, labour markets, and GDP (measured in US\$ in constant prices and constant PPPs with a base year of 2005).

Table 1 shows the distribution of cities across countries in the data. Most major OECD countries are represented, with the notable exceptions being Canada, Korea, Spain and the United Kingdom, for which patent data are not available. Table 2 reports the correlation between patenting activity, economic activity as measured by GDP, and population in our sample, for 2008. There is high correlation between all three variables; large cities are also cities with lots of economic activity, and lots of innovative activity. Figures 1 and 2 graphically represent the same information as in Table 2³.

Table 3 presents the ten cities with the largest number of patents in 2008, along with their population and GDP, with their 2000 ranks in parentheses. Although in general the cities with the most patents also have the most population and the highest GDP, there are some anomalies. For instance, San Francisco is associated with Silicon Valley, and has a larger number of patents than would be predicted by its population or GDP. Similarly, Boston is associated with biotechnology and the IT cluster of Route 128, while San Diego is a centre for biotechnology and communications technology. Two other features of Table 3 are noteworthy. First, comparing rankings between 2008 and 2000 shows that populations are persistent over time, whereas GDP and patents are less so; we shall return to this in Section 4 below. Second, cities in the United States dominate the table, occupying seven of the top ten patenting cities in 2008; the equivalent number in 2000 was five of the top ten from the United States. This emphasises the United States' dominance in

² Are patents an input or an output in the knowledge creation process? Griliches (1990) provides an insightful discussion on the use of patent statistics in Economics, and concludes that, in the absence of detailed R&D data, patent data can be used as an indicator of both inventive input and output.

³ A simple regression of the natural log of patents against the natural log of population for any one year yields a coefficient which is always larger than 1; this implies that a 1 percent increase in population has a greater than 1 percent effect on patents. A similar result is obtained for a regression of the natural log of patents against the natural log of GDP.

innovation, although it may be partially driven by cities in countries which have been omitted from our sample due to lack of data, for instance London and Seoul.

3. The distribution of innovative activity

In this section we compare the distribution of patents across cities with the distribution of population and economic activity. If the idea behind Marshall's external economies is correct, then we would expect that patents are going to be more highly concentrated than economic activity in general, and that economic activity is in turn going to be more highly concentrated than population. To perform this comparison, we make use of Zipf's Law, which states that the size distribution follows a simple Pareto distribution with shape parameter equal to 1. To operationalise this idea, let:

$$R = AS^{-\alpha}, \quad (1)$$

where R is the rank of a city in terms of its size (with the largest city being ranked 1), S is the size of the city used in constructing R , and A and α are parameters. Taking natural logs of equation (1) and adding a random error term ϵ gives:

$$\ln R = \ln A - \alpha \ln S + \epsilon. \quad (2)$$

Thus the Zipf's Law prediction is that there is a linear relationship between the natural log of the rank and the natural log of the size. The parameter α is a measure of the inequality of the distribution; the larger is α , the more equal is the distribution across cities.

Figure 3 plots the scatter diagram of the rank of a city versus its size as measured by population, GDP and number of patents, for 2008, on a log scale with the largest value normalised to 1. The figure shows that, whilst there appears to be a roughly linear relationship between log of rank and log of population, there is pronounced curvature for GDP and especially for patents. Another observation that can be made from Figure 3 is that, overall, population is more equally distributed than GDP, which in turn is more equally distributed than patents. If Marshall's external economies argument is correct, then this is what we would expect; that larger cities are more productive than smaller cities, and this is especially true for innovative activity where proximity to other innovating agents will yield greater external economies than other types of economic activity.

Gabaix and Ioannides (2004) show that OLS estimation of equation (2) leads to biased results, while Gabaix and Ibragimov (2011) show that a simple way to improve OLS estimation of equation (2) is instead to estimate the following equation:

$$\ln\left(R - \frac{1}{2}\right) = \ln A - \alpha \ln S, \quad (3)$$

with the standard error of α being given by $(2/n)^{1/2}\alpha$, where n is the number of cities. The results of estimating equation (3) for each year for population, GDP and patents are presented in Table 4, which reports the values of α . Comparing across the three measures, the coefficients for population are always larger than for GDP, which in turn are always larger than for patents. This confirms the visual inspection of Figure 3 discussed above; population is the most equally distributed across cities, followed by GDP, with patents being the most unequally distributed. Comparing the coefficients across time, the coefficient for population is almost constant over time. The coefficient for GDP shows greater variation over time (although part of the variation is driven by data availability), while the coefficient for patents shows the greatest variation over time. Especially for patents, there appears to be a trend of rising coefficients, which indicates that patenting activity is becoming more dispersed over time. This may indicate that the Marshallian external economies in innovative activity are becoming weaker over time, perhaps in response to developments in communication technology. In terms of Zipf's Law (the hypothesis that $\alpha = 1$), for this sample of cities, Zipf's Law holds for GDP, but not for population and patents. City populations are more equal in size than would be predicted by Zipf's Law, whereas patents are less equally distributed than the Zipf's Law prediction.

As noted above, Figure 3 shows that the Pareto distribution may not be the most appropriate one to describe the size distribution of cities, especially in terms of GDP or patents. Eeckhout (2004) suggested that a lognormal distribution may be a more appropriate distribution. To investigate this, we proceed in the following way. First, following Soo (2012), we make use of the Shapiro-Wilks and Shapiro-Francia tests of normality to see if population, GDP and patents are lognormally distributed. The results of these tests are reported in Table 5, for the years 2000, 2004 and 2008. In all years, the null hypothesis that each of these variables is lognormally distributed can be rejected at any conventional significance level. The second way in which we compare the distribution of the variables to the lognormal distribution is through the use of normal probability plots. These are presented in Figure 4, for 2008. If a variable follows a lognormal distribution, then the observations should lie along the upward-sloping straight line. In none of the three cases do we observe this. Hence we

can conclude that none of the variables: population, GDP and patents, follows a lognormal distribution.

4. The persistence of innovative activity

In this section we examine how persistent is innovative activity, relative to population and GDP. We make use of transition probability matrices first introduced into the economic growth literature by Quah (1993), and used in the city population literature by Eaton and Eckstein (1997), Dobkins and Ioannides (2000), and Black and Henderson (2003). We group the sample of cities into ten cells in each year. Let F_t be a 10×1 vector which denotes the distribution of sizes across cities at time t . Assume that F_t evolves according to:

$$F_{t+1} = MF_t, \tag{4}$$

where M is a 10×10 transition probability matrix, mapping the assignment from period t into an assignment in period $t + 1$. Following Dobkins and Ioannides (2000), we define the vector F_t based on the deciles of the distribution⁴. Since we have data from 2000 to 2008, and since population changes only slowly, we present results for the 8-year transition matrix between 2000 and 2008⁵.

Table 6 presents the results, arranged so as to make the comparison between the three variables (population, GDP and patents) as clear as possible. Overall, patents exhibit less persistence than population and GDP; the diagonal elements of the matrix (in **bold** type) are, on average, smaller for patents than for population and GDP. On the other hand, population and GDP appear to be quite similar in terms of how persistent they are over time⁶. Indeed, the mobility of a city both up and down the distribution of patents is quite large; a city which in the year 2000 was between the 60th and 70th percentiles of the distribution of patents, could by the year 2008 lie anywhere between 30th and 90th percentiles.

⁴ Ioannides and Overman (2001) discuss further the implications of this way of defining F_t as compared to that used by Eaton and Eckstein (1997) and Black and Henderson (2003), which is based on fractions of the contemporaneous mean. In this paper, since we are comparing the distributions of different variables, a decile-based definition seems more appropriate.

⁵ Many of the papers which make use of transition probability matrices on city populations go on to obtain the long run, implied ergodic distribution of city sizes. We do not do so, because the relatively short time period of our sample means there are relatively few off-diagonal elements of the transition matrices, making the calculations sensitive to the choice of cell boundaries. In addition, it would require F_t to be defined based on fractions of the contemporaneous mean (see the previous footnote) as opposed to our decile-based definition.

⁶ Because of the many zero entries in the table, it is not possible to perform a chi-squared test of the similarity between the distributions of the three variables.

Nevertheless, where patenting activity does exhibit considerable persistence, is at both ends of the distribution. Cities in the bottom 10th percentile of the distribution of patents in the year 2000 only had a 13.6 percent chance of moving up to the 20th percentile by 2008, which is a lower likelihood of transition than for both population and GDP. A similar though less pronounced pattern can be observed at the top of the distribution. What this suggests is that cities that start off with low levels of patenting activity, struggle to develop any innovation capacity (or perhaps choose to specialise in non-innovation-intensive activities); cities with lots of patenting activity benefit from Marshallian external economies, while cities in between may end up in either a virtuous or a vicious cycle of innovation.

5. The growth of innovative activity

In this section we make use of both parametric and nonparametric approaches to examine the growth of innovative activity. Perhaps a natural starting point is to assume that city growth and city size are independently distributed; that is, that city growth obeys Gibrat's Law. We follow Black and Henderson (2003) in estimating the following equation:

$$\ln(S_{it+1}) - \ln(S_{it}) = \beta_i + \delta_t + \gamma \ln(S_{it}) + \epsilon_{it}, \quad (5)$$

where β_i are city fixed effects and δ_t are time fixed effects. The null hypothesis implied by Gibrat's Law is that $\gamma = 0$. Given the null hypothesis of Gibrat's Law, the error term cannot be serially correlated, so we use a conventional fixed-effects model to estimate equation (5)⁷. Here, unlike in the previous section, we make use of data on an annual basis.

The estimated values of γ for population, GDP and patents are reported in Table 7. Standard errors are clustered by city to allow for heteroskedasticity and within-city correlation in the residuals, and all results reported include both year and city fixed effects. For all three variables of interest, the Gibrat's Law null hypothesis that $\gamma = 0$ is rejected in favour of the alternative that $\gamma < 0$. That is, rather than random growth, we find evidence of mean-reversion; large cities grow more slowly than small ones. The coefficient is largest in absolute terms (hence mean reversion is the quickest) for patents, followed by GDP and population. Similarly to the results of

⁷ Equation (5) is of course just the equation that is estimated in a panel unit root test. Conventional panel unit root tests cannot be used for our data because of the limited time dimension and the fact that we have an unbalanced panel for GDP. See for instance Bosker et al (2008) for an application of tests of this type to German city sizes.

the previous section, patents exhibit less persistence than GDP and especially population.

However, parametric models such as equation (5) do not give a complete picture of the relationship between size and growth of cities. Therefore, we supplement equation (5) with a non-parametric, local linear estimator. This is an extension of the Nadaraya-Watson local mean smoothing approach used by Ioannides and Overman (2003); instead of using mean values, we use fitted values from a local linear regression to determine the conditional expectation. Fan (1992) shows that the local linear estimator suffers from less bias than the Nadaraya-Watson approach. To implement this estimator, we standardise the size and growth of cities by subtracting the annual mean from the raw data and dividing by the standard deviation. This allows us to pool observations across years. We use the Epanechnikov kernel, and the bandwidth chosen is obtained using the rule-of-thumb method of bandwidth selection.

Figure 5 reports the results of the nonparametric estimates, for the three variables population, GDP and patents, together with a 95 percent confidence interval; the scatterplot of data points has been omitted for clarity. From this figure it can be seen that GDP most closely follows the Gibrat's Law null hypothesis of no relationship between GDP and GDP growth. Even here there is some evidence that cities with larger GDP exhibit slower growth than cities with smaller GDP. For population, cities in the middle of the population distribution grow faster than those at both ends of the distribution. For patents, the confidence bands are much narrower than for the other two variables, and, consistently with the parametric results in Table 5, it is cities with the fewest patents that experience the fastest patent growth rates. However, cities between 1 and 2 standard deviations below the mean experience slower patent growth rates on average. This may indicate the presence of multiple equilibria; whilst it may be relatively easy for cities with few patents to rapidly increase their patenting rate, it may be more difficult to step up to the next level and join the ranks of the major innovating centres.

6. The determinants of innovative activity

In the previous section, one general conclusion that emerged was that cities with relatively fewer patents, experience more rapid growth in patenting activity. In this section we explore this further, and investigate the possible determinants of innovative activity in a city. Similarly to Black and Henderson (2003), we rearrange

and extend equation (5) in the previous section to include additional explanatory variables:

$$\ln(S_{it+1}) = \beta_i + \delta_t + (1 + \gamma) \ln(S_{it}) + \psi X_{it} + \epsilon_{it}, \quad (6)$$

where X_{it} may include both time-varying and time-invariant variables. Including the lagged dependent variable in equation (6) means that conventional OLS, fixed- and random-effects estimates are all biased. We therefore use the Blundell and Bond (1998) system GMM method in its asymptotically efficient, two-step form. The method estimates a system of two equations; the equation in levels, and in orthogonal deviations (each observation is subtracted from the average of all future available observations). Because of the inclusion of the levels equation, it is possible to recover the coefficients on time-invariant explanatory variables. The reported standard errors are clustered by city so are robust to heteroskedasticity and arbitrary serial correlation within panels, and are corrected for downward bias using the Windmeijer (2005) correction. Time dummies are included in all regressions to reduce the contemporaneous correlation across cities.

The lagged dependent variable is assumed to be endogenous and needs to be instrumented. Under standard system GMM, the variables in the levels equation are instrumented with lags of their own first differences, while the variables in the orthogonal transformed equation are instrumented with lags of the variables in levels. However, this results in the number of instruments being quadratic in the time dimension. To avoid the problem of too many instruments in system GMM (see Roodman (2009b)), we follow the recent literature (Mehrhoff (2009), Kapetanios and Marcellino (2010), Bai and Ng (2010)) and replace the GMM instruments with their principal components. Principal components analysis is run on the correlation matrix of the GMM instruments, and the principal components with the largest eigenvalues are selected as instruments. Additional statistics reported in Table 8 show that in each specification the principal components explain most of the variation in the instruments, and that they perform well based on the Kaiser-Meyer-Olkin measure of sampling adequacy.

The results of estimating equation (6) are presented in Table 8. As a benchmark for comparison, column (1) reports the results of estimating equation (6) using conventional fixed effects estimation, including both city and year fixed effects, but no additional controls. The coefficient on the lagged dependent variable is 0.201, which shows strong evidence of mean reversion. However, as noted above, this estimate may be biased. Column (2) reports the results for the same equation using the Blundell and Bond (1998) system GMM method described above. The coefficient

increases dramatically, to 0.921, and providing much weaker evidence of mean reversion. That controlling for variable endogeneity often changes the magnitude if not the direction of the result, suggests that endogeneity is an important issue.

The remaining columns of Table 8 include additional controls. In columns (3) and (4), population and GDP are included. In column (3), these two variables are assumed to be exogenous, while in column (4) they are assumed to be endogenous and are instrumented in the same way as the lagged dependent variable. Including population and GDP reduces the size of the coefficient on lagged patents, although it is still positive and highly significant. Controlling for the other two variables, population has no significant effect on patents, while GDP has a positive and significant effect. This suggests that economic activity is more strongly associated with innovative activity than the mere presence of a larger population.

Columns (5) and (6) include additional controls. This includes the number of local governments per 100,000 inhabitants of the metropolitan area (capturing the fragmentation of local government), the number of non-contiguous core areas in the metro area (the polycentricity of the city), the share of the total metropolitan population living in the core areas of the city, the population density, and an indicator for whether there is a top-100 university in the city. By including core population, population density and polycentricity, we seek to explore whether the concentration of people (Marshall's knowledge spillovers) affects the degree of innovative activity. The fragmentation of local government may affect the coordination of government policies across local governments, which again may influence innovation. In column (5), all these additional variables are assumed to be exogenous, whereas in column (6), population density and the share of the total metropolitan population living in the core areas of the city are treated as endogenous and are instrumented in the usual way.

Including the presence of a top-100 university as an explanatory variable comes from the idea that knowledge spillovers from university research and research collaborations with local universities may spur private sector research. Early research on such relationships includes Jaffe (1989), and more recently Abramovsky et al (2007). There are three major global university rankings: the Academic Ranking of World Universities (ARWU or the Shanghai Ranking), the Times Higher Education World University Rankings, and the QS World University Rankings. The QS World University Rankings were not available for our sample period, and the other two rankings are available only since 2003 (ARWU) and 2004 (Times). The results

reported below make use of the ARWU rankings in 2008, and we code all cities with a top-100 university according to this ranking equal to 1, and all other cities equal to zero. A total of 32 cities in our sample includes at least one top-100 university according to this measure⁸.

Once the endogeneity of population density and population share of the core are controlled for in column (6), the degree of local government fragmentation and population density are positively and significantly associated with patenting activity. The first result may suggest that competition among local governments may spur them to implement innovation-friendly policies to attract firms. The second result suggests that if Marshall's knowledge spillovers are active, one channel via which they operate is through increased interaction because of greater population density. The other three additional variables – the degree of polycentricity, the share of metropolitan population living in the core and the presence of a top-100 university – do not have statistically significant effects on patenting, although there is a suggestion that polycentricity is negatively related to patenting, while population share in the core and the presence of a top-100 university are both positively related to patenting. Inclusion of these additional variables reduces the size of the previous coefficients, but they retain their level of statistical significance.

We also include a set of diagnostic statistics in Table 8. First, we report the number of instruments used, which ranges from 21 to 37 instruments. These are fairly low, which should mitigate the problem of having too many instruments (see Roodman (2009b)), and as discussed above, is because we have used the principal components of the GMM instruments; if we had not done so, column (6) of Table 8 would have had over 150 instruments. Second, we report the Sargan and Hansen tests of over-identification; we report both tests since the Sargan test, whilst not robust to heteroskedasticity, is not weakened by increasing numbers of instruments, whereas the Hansen test is robust to heteroskedasticity, but is weakened by having many instruments. We can see that the Sargan test becomes insignificant as we include more variables and instruments, whereas the opposite occurs for the Hansen test, which is insignificant in the baseline column (2), but becomes significant as more variables and instruments are included. Overall these tests suggest cautious confidence in our instruments. A third set of statistics is the Arellano and Bond (1991) tests for first- and second-order serial correlation in the first-differenced residuals. We find evidence of first-order serial correlation, but not second-order

⁸ Results using the Times ranking are qualitatively similar, with the exception that the degree of fragmentation loses significance.

serial correlation, across all specifications in Table 8. First-order serial correlation is expected in a dynamic panel; that we do not find second-order serial correlation provides evidence that our use of lags as instruments is valid.

7. Conclusions

Competition among firms drives innovation in a capitalist economy, as firms seek to gain a competitive edge over their rivals. Hence as urbanisation proceeds and economic activity becomes increasingly concentrated in cities, so too does innovative activity. What this paper has set out to do, is to describe and explain the distribution of innovative activity across OECD cities. Although there has been much research on innovation in cities, to our knowledge this is the first paper to compare the distribution of innovation to the distribution of population and economic activity across cities.

Our first main result is that innovation is more highly concentrated than both population and general economic activity. This is suggestive of the role of Marshall's knowledge spillovers as a key driver of innovation. Our second main result is that innovation is less persistent than population or economic activity. Even in the relatively short time period in our sample, cities can become much more (or less) innovative. This gives policymakers hope, that government policy can influence how innovative a city is. Our third main result is that, even after controlling for the endogeneity of some explanatory variables, innovation is positively related to general economic activity, population density, and the degree of fragmentation of local government. Again this gives policymakers a handle on what types of policies may be more effective at promoting innovation.

From the policymaker's perspective, the present paper's focus on cities as centres of innovative activity is a double-edged sword. On the one hand, cities are undoubtedly important; in the OECD, the vast majority of the population lives and works in cities. So thinking about government policies in terms of cities may be the more natural unit of analysis. On the other hand, precisely because cities have not historically been the default unit of analysis, our analysis suffers from data limitations that not only restrict our sample, but also prevent us from digging deeper into the determinants of innovative activity as in Audretsch and Feldman (1996, 1999). Such data is available for different geographical units, and analysis using this data should serve as an important next step in this line of research.

Acknowledgements

Thanks to Geraint Johnes and seminar participants at The Work Foundation for helpful suggestions. The author is responsible for any errors and omissions.

References

Abramovsky, Laura, Harrison, Rupert and Simpson, Helen (2007), "University research and the location of business R&D", *Economic Journal* 117(519), C114-C141.

Arellano, Manuel and Bond, Stephen (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies* 58(2), 277-297.

Audretsch, David B. and Feldman, Maryann P. (1996), "R&D spillovers and the geography of innovation and production", *American Economic Review* 86(4), 253-273.

Audretsch, David B. and Feldman, Maryann P. (1999), "Innovation in cities: Science-based diversity, specialization and localized competition", *European Economic Review* 43(2), 409-429.

Audretsch, David B. and Feldman, Maryann P. (2004), "Knowledge spillovers and the geography of innovation", in Henderson, J. Vernon and Thisse, Jacques-Francois (eds.), *Handbook of Regional and Urban Economics* Volume 4, 2713-2739.

Bai, Jushan and Ng, Serena (2010), "Instrumental variable estimation in a data rich environment", *Econometric Theory* 26(6), 1577-1606.

Berry, Christopher R. and Glaeser, Edward L. (2005), "The divergence of human capital levels across cities", *Papers in Regional Science* 84(3), 407-444.

Bettencourt, Luis M.A., Lobo, Jose, Strumsky, Deborah and West, Geoffrey B. (2010), "Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities", *PLoS ONE*, 5(11): 1-9.

Bettencourt, Luis M.A., Lobo, Jose and Strumsky, Deborah (2007), "Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size", *Research Policy* 36(1), 107-120.

Black, Duncan and Henderson, J. Vernon (2003), "Urban evolution in the USA", *Journal of Economic Geography* 3(4), 343-372.

Blundell, Richard and Bond, Stephen (1998), "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics* 87(1), 115-143.

Bosker, Maarten, Brakman, Steven, Garretsen, Harry and Schramm, Marc (2008), "A century of shocks: The evolution of the German city size distribution 1925-1999", *Regional Science and Urban Economics* 38(4), 330-347.

Dobkins, Linda Harris and Ioannides, Yannis M. (2000), "Dynamic evolution of the US city size distribution", in Huriot, Jean-Marie and Jacques-Francois Thisse (eds.), *The Economics of Cities*, Cambridge, Cambridge University Press, 217-260.

Dobkins, Linda Harris and Ioannides, Yannis M. (2001), "Spatial interactions among US cities: 1900-1990", *Regional Science and Urban Economics* 31(6), 701-731.

Eaton, Jonathan and Eckstein, Zvi (1997), "Cities and growth: Theory and evidence from France and Japan", *Regional Science and Urban Economics* 27(4-5), 443-474.

Eeckhout, Jan (2004), "Gibrat's Law for (all) cities", *American Economic Review* 94(5), 1429-1451.

Fan, Jianqing (1992), "Design-adaptive nonparametric regression", *Journal of the American Statistical Association* 87(420), 998-1004.

Fujita, Masahisa, Krugman, Paul R. and Venables, Anthony J. (1999), *The Spatial Economy*, Cambridge, MA, MIT Press.

Gabaix, Xavier and Ibragimov, Rustam (2011), "Rank – α : A simple way to improve the OLS estimation of tail exponents", *Journal of Business Economics and Statistics* 29(1), 24-39.

Gabaix, Xavier and Ioannides, Yannis M. (2004), "The evolution of city size distributions", in Henderson, J. Vernon and Thisse, Jacques-Francois (eds.), *Handbook of Regional and Urban Economics* Volume 4, 2341-2379.

Glaeser, Edward L., Kallal, Hedi D., Scheinkman, Jose A. and Shleifer, Andrei (1992), "Growth in cities", *Journal of Political Economy* 100(6), 1126-1152.

Glaeser, Edward L., Scheinkman, Jose A. and Shleifer, Andrei (1995), "Economic growth in a cross-section of cities", *Journal of Monetary Economics* 36(1), 117-143.

Griliches, Zvi (1990), "Patent statistics as economic indicators: A survey", *Journal of Economic Literature* 28(4), 1661-1707.

Ioannides, Yannis M. and Overman, Henry G. (2004), "Spatial evolution of the US urban system", *Journal of Economic Geography* 4(2), 131-156.

- Ioannides, Yannis M. and Overman, Henry G. (2003), "Zipf's Law for cities: An empirical examination", *Regional Science and Urban Economics* 33(2), 127-137.
- Ioannides, Yannis M. and Overman, Henry G. (2001), "Cross-sectional evolution of the US city size distribution", *Journal of Urban Economics* 49(3), 543-566.
- Jaffe, Adam B. (1989), "Real effects of academic research", *American Economic Review* 79(5), 957-970.
- Kapetanios, George and Marcellino, Massimiliano (2010), "Factor-GMM estimation with large sets of possibly weak instruments", *Computational Statistics and Data Analysis* 54(11), 2655-2675.
- Krugman, Paul R. (1991), *Geography and Trade*, Cambridge, MA, MIT Press.
- Marshall, Alfred (1890), *Principles of Economics*, London, MacMillan.
- Mehrhoff, Jens (2009), "A solution to the problem of too many instruments in dynamic panel data GMM", Deutsche Bundesbank Discussion Papers 200931.
- OECD (2012), *Redefining "Urban": A new way to measure metropolitan areas*, OECD Publishing.
- OhUallachain, Breandan (1999), "Patent places: Size matters", *Journal of Regional Science* 39(4), 613-636.
- Quah, Danny (1993), "Empirical cross-section dynamics in economic growth", *European Economic Review* 37(2-3), 426-434.
- Roodman, David (2009a), "How to do xtabond2: An introduction to 'difference' and 'system' GMM in Stata", *Stata Journal* 9(1), 86-136.
- Roodman, David (2009b), "A note on the theme of too many instruments", *Oxford Bulletin of Economics and Statistics* 71(1), 135-158.
- Soo, Kwok Tong (2005), "Zipf's Law for cities: A cross-country investigation", *Regional Science and Urban Economics* 35(3), 239-263.
- Soo, Kwok Tong (2007), "Zipf's Law and urban growth in Malaysia", *Urban Studies* 44(1), 1-14.
- Soo, Kwok Tong (2012), "The size and growth of state populations in the United States", *Economics Bulletin* 32(2), 1238-1249.

Strumsky, Deborah and Thill, Jean-Claude (2013), "Profiling US metropolitan regions by their social research networks and regional economic performance", *Journal of Regional Science* 53(5), 813-833.

Windmeijer, Frank (2005), "A finite sample correction for the variance of linear efficient two-step GMM estimators", *Journal of Econometrics* 126(1), 25-51.

Zipf, George K. (1949), *Human behaviour and the principle of least effort*, Cambridge, MA, Addison-Wesley.

Table 1: Distribution of cities across countries in the sample.

Country	Number of cities
Austria	3
Belgium	4
Denmark	1
Estonia	1
Finland	1
France	15
Germany	24
Italy	11
Japan	36
Mexico	31
Netherlands	5
Norway	1
Portugal	2
Spain	8
Sweden	3
United States	72
Total	218

Table 2: Correlation between patents, GDP and population, 2008 (N = 218).

	Patents	GDP	Population
Patents	1.000		
GDP	0.833	1.000	
Population	0.797	0.939	1.000

Table 3: Top 10 cities with the largest number of patents in 2008.

City	Population	Rank	Patents	Rank	GDP (million US\$)	Rank
Tokyo	34,482,744	1(1)	8,727.0	1(2)	1,316,049	1(-)
San Francisco	6,778,659	10(10)	5,138.2	2(1)	463,435	7(5)
Osaka	17,211,140	4(4)	4,451.1	3(4)	534,747	5(-)
San Diego	3,036,850	35(37)	2,689.3	4(10)	160,635	23(18)
Paris	11,529,670	7(7)	2,467.6	5(7)	575,983	4(3)
Boston	3,616,814	29(28)	2,207.5	6(3)	241,083	12(8)
New York	16,453,331	6(5)	2,001.7	7(6)	977,119	2(1)
Los Angeles	16,742,427	5(6)	1,957.7	8(5)	768,032	3(2)
Minneapolis	3,212,176	34(34)	1,672.5	9(11)	174,234	18(16)
Houston	5,363,803	16(17)	1,590.1	10(16)	323,819	9(7)

Notes: Figures in parentheses are ranks in 2000. (-) indicates that data was not available in the year 2000.

Table 4: Zipf regressions for population, GDP and patents, by year.

Year	(1) Population	(2) GDP	(3) Patents
2000	1.246 (0.119)**	1.041 (0.121)	0.330 (0.032)***
2001	1.247 (0.119)**	1.026 (0.107)	0.350 (0.033)***
2002	1.248 (0.119)**	1.025 (0.107)	0.375 (0.036)***
2003	1.247 (0.119)**	0.927 (0.089)	0.358 (0.034)***
2004	1.248 (0.119)**	0.927 (0.089)	0.370 (0.035)***
2005	1.247 (0.119)**	0.926 (0.089)	0.384 (0.037)***
2006	1.246 (0.119)**	0.928 (0.089)	0.387 (0.037)***
2007	1.245 (0.119)**	0.928 (0.089)	0.397 (0.038)***
2008	1.244 (0.119)**	0.932 (0.089)	0.426 (0.041)***

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Statistical significance is in terms of the null hypothesis that the coefficient is equal to 1. N = 218 for all years in columns (1) and (3); N = 148 in 2000, N = 184 in 2001 and 2002, N = 217 in 2003 to 2007, and N = 218 in 2008 in column (2). The values reported are the values of α estimated using the Gabaix and Ibragimov (2011) approach in equation (3). Standard errors in parentheses are calculated using the Gabaix and Ibragimov (2011) approach.

Table 5: The Shapiro-Wilks and Shapiro-Francia tests for lognormality.

Year	Variable	Shapiro-Wilks test		Shapiro-Francia test	
		z	p-value	z	p-value
2000	Population	7.147	0.000	6.617	0.000
	GDP	3.761	0.000	3.523	0.000
	Patents	6.104	0.000	5.614	0.000
2004	Population	7.182	0.000	6.645	0.000
	GDP	4.314	0.000	4.114	0.000
	Patents	5.779	0.000	5.339	0.000
2008	Population	7.111	0.000	6.581	0.000
	GDP	4.348	0.000	4.140	0.000
	Patents	4.911	0.000	4.542	0.000

Notes: The null hypothesis is that each variable follows a lognormal distribution. N = 218 for all years for population and patents; N = 148 in 2000, N = 217 in 2004, and N = 218 in 2008 for GDP.

Table 6: Transition probability matrices for population, GDP and patents, 2000-2008.

2000	2008	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
0.10	Population	71.43	25.00	3.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	GDP	85.71	14.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Patents	86.36	13.64	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.20	Population	25.93	51.85	18.52	3.70	0.00	0.00	0.00	0.00	0.00	0.00
	GDP	15.00	65.00	15.00	5.00	0.00	0.00	0.00	0.00	0.00	0.00
	Patents	13.64	63.64	22.73	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.30	Population	0.00	17.86	53.57	28.57	0.00	0.00	0.00	0.00	0.00	0.00
	GDP	0.00	20.00	55.00	15.00	10.00	0.00	0.00	0.00	0.00	0.00
	Patents	0.00	13.64	36.36	31.82	9.09	9.09	0.00	0.00	0.00	0.00
0.40	Population	0.00	0.00	25.93	48.15	25.93	0.00	0.00	0.00	0.00	0.00
	GDP	0.00	0.00	25.00	55.00	20.00	0.00	0.00	0.00	0.00	0.00
	Patents	0.00	4.55	18.18	18.18	22.73	27.27	4.55	4.55	0.00	0.00
0.50	Population	3.57	0.00	0.00	17.86	67.86	10.71	0.00	0.00	0.00	0.00
	GDP	0.00	0.00	5.00	25.00	65.00	5.00	0.00	0.00	0.00	0.00
	Patents	0.00	4.76	14.29	38.10	28.57	14.29	0.00	0.00	0.00	0.00
0.60	Population	0.00	3.70	0.00	0.00	7.41	66.67	22.22	0.00	0.00	0.00
	GDP	0.00	0.00	0.00	0.00	4.76	85.71	9.52	0.00	0.00	0.00
	Patents	0.00	0.00	9.09	4.55	27.27	31.82	27.27	0.00	0.00	0.00
0.70	Population	0.00	0.00	0.00	0.00	0.00	21.43	64.29	14.29	0.00	0.00
	GDP	0.00	0.00	0.00	0.00	0.00	10.00	65.00	25.00	0.00	0.00
	Patents	0.00	0.00	0.00	9.09	9.09	18.18	36.36	22.73	4.55	0.00
0.80	Population	0.00	0.00	0.00	0.00	0.00	0.00	14.81	81.48	3.70	0.00
	GDP	0.00	0.00	0.00	0.00	0.00	0.00	25.00	55.00	20.00	0.00
	Patents	0.00	0.00	0.00	0.00	0.00	0.00	31.82	54.55	13.64	0.00
0.90	Population	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.57	89.29	7.14
	GDP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	20.00	75.00	5.00
	Patents	0.00	0.00	0.00	0.00	0.00	0.00	0.00	18.18	68.18	13.64
1.00	Population	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.41	92.59
	GDP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.00	95.00
	Patents	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	14.29	85.71

Notes: N = 275 for population, N = 202 for GDP, and N = 218 for patents. The number in each cell shows the probability of transitioning from one decile in 2000 to the corresponding decile in 2008. The values in **bold** are the percentages of cities that remain in the same decile between the two years.

Table 7: Test of Gibrat's Law.

Variable	(1) Population	(2) GDP	(3) Patents
$\ln(\text{population})_{t-1}$	-0.053 (0.023)**		
$\ln(\text{GDP})_{t-1}$		-0.108 (0.018)***	
$\ln(\text{patent})_{t-1}$			-0.814 (0.058)***
R^2	0.18	0.22	0.44
N	2,475	2,235	1,744
City fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Estimation is via OLS with city and year fixed effects, with standard errors clustered by city.

Table 8: The determinants of patenting activity (dependent variable: $\ln(\text{patent})_t$).

Estimation method	(1) FE	(2)	(3)	(4)	(5)	(6)
			Blundell-Bond (1998) System GMM			
$\ln(\text{patent})_{t-1}$	0.201 (0.057)***	0.921 (0.071)***	0.739 (0.095)***	0.588 (0.079)***	0.665 (0.125)***	0.451 (0.103)***
$\ln(\text{pop})_t$			-0.321 (0.630)	-0.329 (0.740)	1.802 (1.428)	-0.103 (0.583)
$\ln(\text{GDP})_t$			1.105 (0.398)***	1.534 (0.360)***	0.198 (0.613)	1.085 (0.609)*
$\ln(\text{CorePop})_t$					-1.385 (4.009)	1.621 (1.673)
$\ln(\text{PopDens})_t$					0.003 (0.710)	0.647 (0.207)***
$\ln(\text{Poly})_t$					-2.954 (6.529)	-3.228 (2.665)
$\ln(\text{Fragment})_t$					0.386 (0.802)	0.643 (0.356)*
Top 100 university					3.144 (3.072)	2.053 (2.154)
R^2	0.33					
N	1,968	1,968	1,879	1,879	1,671	1,671
Number of cities	246	246	246	246	218	218
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Instruments		21	23	27	31	37
Sargan Test p-value		0.00	0.00	0.30	1.00	0.99
Hansen Test p-value		0.20	0.02	0.00	0.00	0.00
AB AR(1) Test p-value		0.00	0.00	0.00	0.00	0.00
AB AR(2) Test p-value		0.26	0.38	0.46	0.42	0.57
PCA R2		0.97	0.97	0.89	0.90	0.93
Kaiser-Meyer-Olkin		0.89	0.89	0.94	0.94	0.96

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Estimation is via fixed effects with standard errors clustered by city in column (1), and via the two-step Blundell-Bond (1998) System GMM with Windmeijer (2005) corrected standard errors in columns (2) to (6). *CorePop* is the concentration of population in the metropolitan core. *PopDens* is population density. *Poly* is the degree of polycentricity of the city. *Fragment* is the degree of fragmentation of local government. Top 100 university is an indicator for whether there is a top-100 university in the city, as ranked by ARWU. The Sargan and Hansen tests are the tests of over-identifying restrictions. The Sargan test is not robust to heteroskedasticity but is not weakened by many instruments, whereas the Hansen test is robust to heteroskedasticity but is weakened by many instruments. The Arellano and Bond tests (AB) are tests for serial correlation in the first-differenced errors, of orders 1 and 2. PCA R2 is the fraction of the variance explained by the principal components, and Kaiser-Meyer-Olkin is a measure of the sampling adequacy of the principal components.

Figure 1: Scatterplot of patent applications and population, 2008 (N = 218).

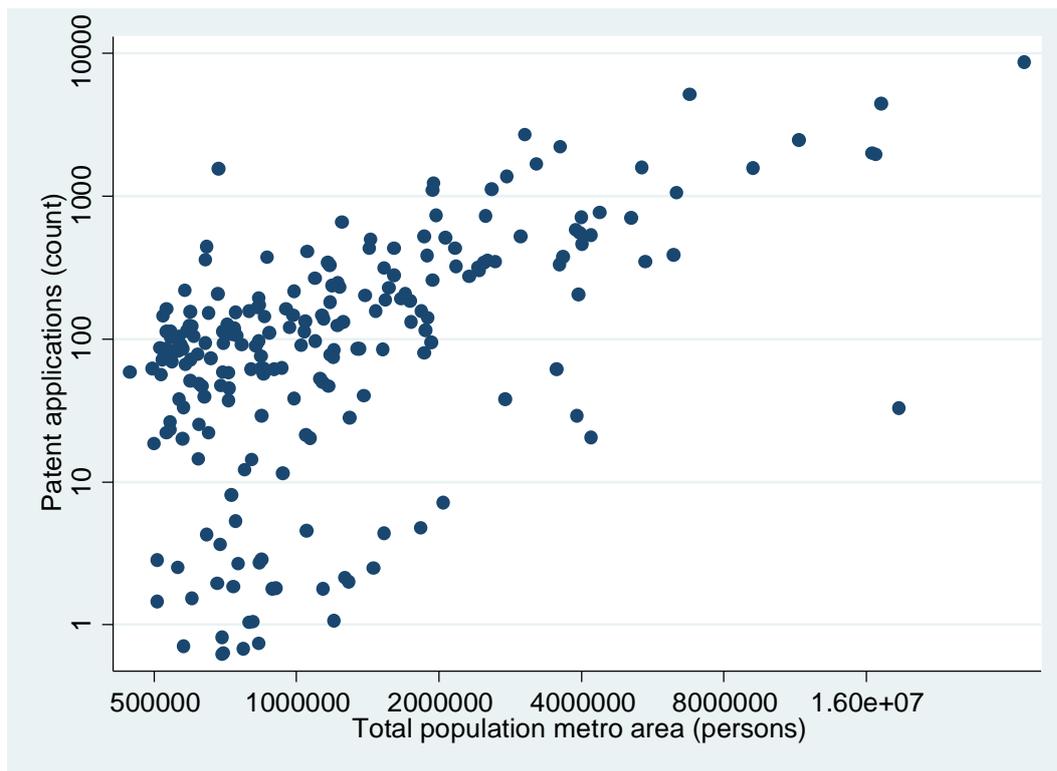


Figure 2: Scatterplot of patent applications and real GDP, 2008 (N = 218).

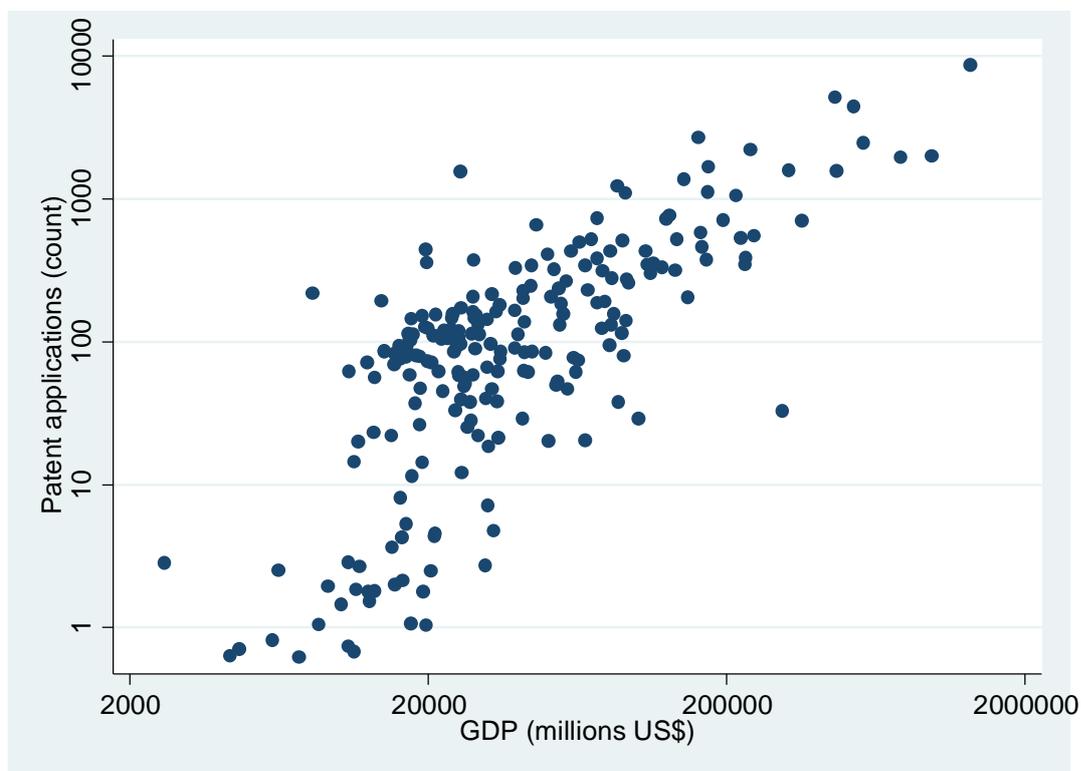


Figure 3: Zipf plots of population, patents and GDP, for 2008, log scale, normalised to the size of the largest city.

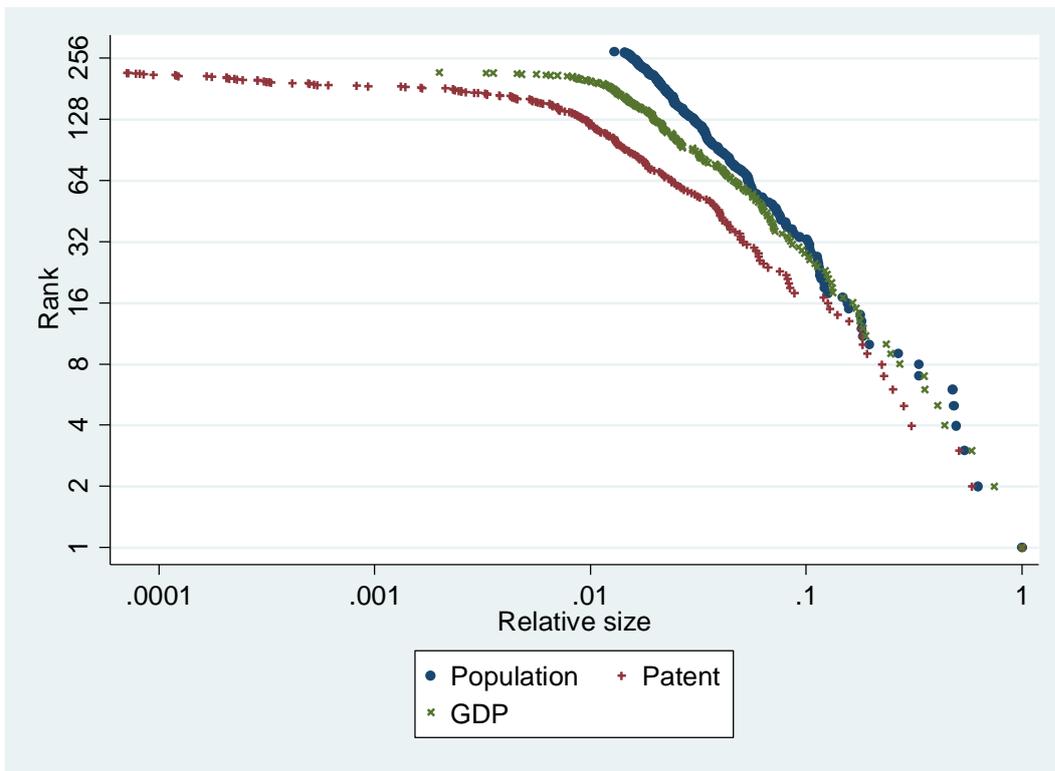
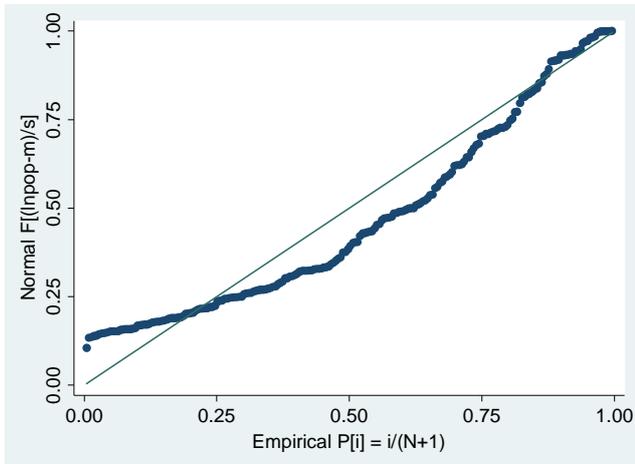
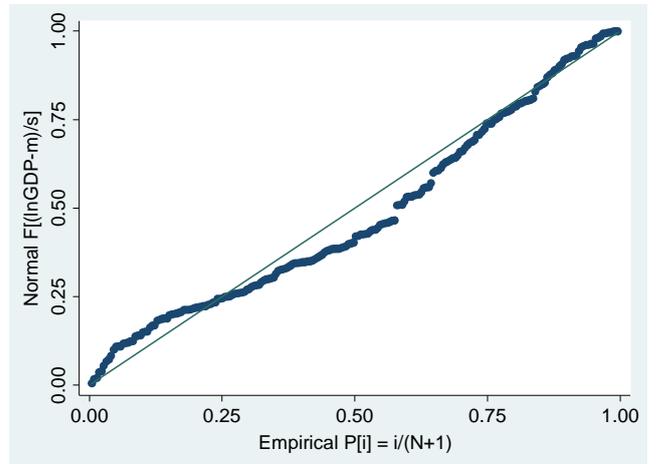


Figure 4: Normal probability plots for population, GDP and patents, in 2008.

Population



GDP



Patents

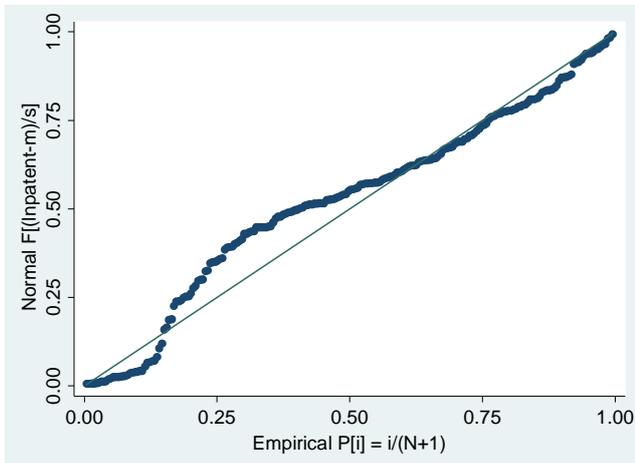
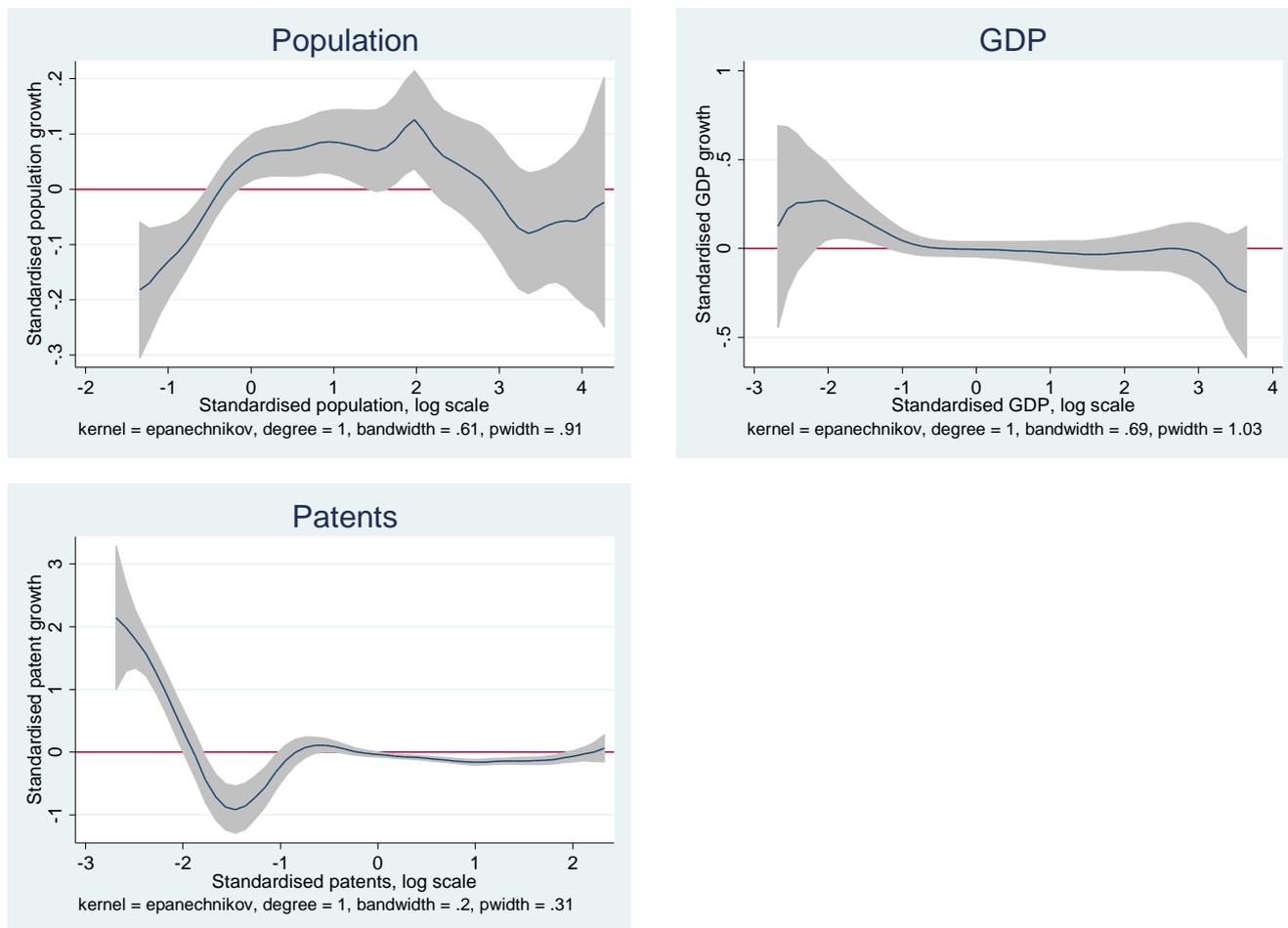


Figure 5: Nonparametric estimates of the relationship between city size and city growth.



Notes: The shaded area indicates the 95% confidence interval. Bandwidth indicates the bandwidth used for the smoothing, while pwidth indicates the bandwidth used for the confidence interval.