CESifo WORKING PAPERS

10008
2022

October 2022

The Impact of the Spatial Population Distribution on Economic Growth: Evidence from the United States

Constantin Bürgi, Nisan Gorgulu

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10008*

The Impact of the Spatial Population Distribution on Economic Growth: Evidence from the United States

Abstract

We look at a part of the spatial angle of economic growth. We introduce a new measure Spatial Population Concentration (SPC) that captures the weighted average population surrounding every person within a geographic area. The weights are a function of the distance between the person in question and everyone else. One special case of the SPC would be to measure how many people live on average within a given radius of every person within a geographic area. We then calculate the SPC measure at the US county level for various radii and identify that the measure has the strongest relationship with subsequent economic growth for a 25km radius. Interacting SPC with various infrastructure measures increases the radius to 50km. This suggests that regional policies which affect density as infrastructure projects should target the 25-50km distance range to maximize the growth impact.

JEL-Codes: O470, O510, R120.

Keywords: spatial population concentration, endogeneous growth, spillover, the United States.

Constantin Bürgi
School of Economics
University College Dublin
Ireland – Belfield, Dublin 4
constantin.burgi@ucd.ie

Nisan Gorgulu
Department of Economics
The George Washington University
USA – Washington, D.C.
nisangorgulu@gwu.edu

October 4, 2022
The authors are thankful to Charles Becker, Remi Jedwab, Timur Kuran, Daniel Pereira, Stephen Smith, Bryan Stuart, Ben Williams, Anthony Yezer and the participants of GWU Development Tea Workshop and the 6th Global Conference on Economic Geography for their helpful comments and suggestions. We also thank Glen MacLachlan and William Bonnett for their help in using the George Washington University's Colonial One High Performance Computing System for our Spatial Population Concentration measure calculations. Opinions expressed in this study are of authors alone and do not necessarily reflect the views of the World Bank, its Board of Directors or members states.

1 Introduction

What drives growth is one of the most important questions in economics. Traditional growth models following Solow (1956); Mankiw et al. (1992) or endogenous growth models Romer (1990, 1994) assume that either all economic activity happens at the same location, or that location is irrelevant. However, an extensive literature on the geography of growth has since shown that these assumptions are not very accurate and geography should be taken into account.[1]

Indeed, an extensive literature that looks at the role of cities for growth has shown that densely populated areas experience faster growth (e.g., Altunbas et al. (2013); Carlino (2001); Charlot and Duranton (2003); Davis and Dingel (2019); de Groot et al. (2007); Perumal (2017); Glaeser et al. (2009)). Cities are places, where people live closely together and have historically been drivers of economic growth (e.g., Chen et al. (2014); Gollin et al. (2016); Jedwab and Vollrath (2015); Jedwab et al. (2017)).[2] However, focusing on cities alone raises the question of the city boundary (i.e., where does a city start and end and to what extent this matters) and the geography of cities themselves. Our research question is at what distance policies (in urban planning, in investing infrastructure, etc.) that affect densities have the strongest relationship with subsequent growth. For example, should one pursue zoning policies that mainly increase the number of people within a 10km radius of every person or one that mainly increases the number of people within a 25km radius to maximize the growth? At what distances do regional policies that impact density as the existence of infrastructure have the strongest impact on growth?

In order to empirically address these questions, we introduce a new measure called the "Spatial Population Concentration" (SPC). This measure is a generalization of the economic density measure in Duranton and Puga (2020); Henderson et al. (2021); Roca and Puga (2017) which measures how many people live within a 10km radius divided by the area. SPC captures the weighted average population surrounding every person within a geographic area. The weights are a function of the distance between the person in question and everyone else. One special case of the SPC would be to measure how many people live on average within a given radius of every person within a geographic area. To get a better intuition about this measure,

[1] For example, Le Van et al. (2002); Rivera-Batiz and Romer (2008); Bottazzi and Peri (2003); Blazek and Sickles (2010); Desmet et al. (2018)

[2] In a cross-country context, Conley and Ligon (2002); Henderson et al. (2018); Moreno and Trehan (1997); Malik and Temple (2009) are examples.

we compare SPC to population density and the urbanization rate. Specifically, the other two measures both capture how close people live together, but they each have important shortcomings relative to our new measure. Population density has the shortcoming that it can be very strongly affected by large areas that are sparsely populated. For instance, Nevada (U.S.) has a very low population density even though people concentrate in the cities due to the surrounding desert. The urbanization rate is more flexible in this regard, as it measures the fraction of people living in cities or otherwise densely populated areas. However, the measure does not differentiate more or less densely populated areas within urban (or rural) areas.[3] In addition to differentiating more or less densely populated urban areas, our measure can be used to construct alternative urbanization rates where a person is deemed to live in an urban area if a threshold number of people live within a particular radius of that person. While we use this measure to estimate whether areas where people live close together coincide with faster growth, there are many other potential applications for this new measure. This new measure of SPC is relevant for almost any application where it matters how close people live together, be it the spread of disease, pollution or infrastructure. As SPC is a more accurate reflection of how close people live together than many other measures, it allows for a more precise analysis and accurate conclusions of potential applications.

While there are many channels through which growth could be affected, we consider two channels through which location might matter for growth. Both channels are closely related to the endogenous growth literature (e.g., Jones (1995), Gong et al. (2004), Aghion and Howitt (1992), Grossman and Helpman (1991), Zeira (2011) or Bloom et al. (2017)) and hence assume that growth is driven by ideas getting implemented. While we focus on the empirical side in this paper, we outlined a simple model of how these channels could be implemented into one model in the appendix (A. Model). Assuming that idea implementation requires high skilled workers that in turn need to commute and are compensated for the commute. The further this commute, the more expensive the commute and the more profitable an idea must be to be worth implementing. In turn, a higher implementation cost due to the geographic spread causes fewer ideas to be implemented and hence slower economic growth. The second channel for how location matters for growth is based on Desmet et al. (2020) and assumes that the final goods need to be distributed to consumers. The more spread out consumers live, the more expensive it becomes to distribute the goods (or the smaller the potential market)

[3] In addition, it is also not standardized across countries.

and hence implementing ideas might not be worthwhile for sparsely populated areas leading to slower growth there. While these channels mainly focus on productivity, we are agnostic that consumption amenities and income sorting could also lead to a positive relationship.

Our paper also has several important policy implications. Aside from reconfirming that growth is faster in areas where people live closely together, we show that policymakers that want to accelerate economic growth should strive to increase the population concentration and avoid too much city sprawling. Given that our results imply the biggest gain for areas with a radius of 20-25km, having several very densely populated hubs should be preferable over a large area of intermediate density population. At the same time, our results suggest that transportation and communication infrastructure also contribute to growth. The peak distance for the growth impact of the interaction of our SPC measure and the infrastructure measures is closer to a radius of 50km. This suggests that the effect we identified can be extended over larger areas, if there is sufficient infrastructure in place. In other words, improving the infrastructure has the largest impact for areas that are not too far from the center itself and should be a priority for policy makers that want to stimulate growth. We also find that education plays a key role and has the highest growth impact in the densely populated areas.

In the remainder of the paper, we discuss our new measure in section 2. We introduce the empirical strategy in section 3 followed by the discussion on data in section 4. After that, sections 5 and 6 present the results and robustness checks. The final section concludes the paper with a discussion on the contribution of our central findings and avenues for future research.

2 A New Measure For How Close People Live Together

We are interested in how the spatial concentration of people impacts GDP growth at different distances. In order to test this, we first need to have a good measure of spatial concentration. Two common measures for the population distribution are population density and the share of urban population. While these measures are easy to calculate and readily available, they have some distinct shortcomings. Population density is affected by large uninhabited areas like deserts or lakes and the share of urban population does not account for different degrees of urbanization. For example, lower Manhattan, NY and Louisville, KY have the same urbanization rate of 100 percent, but lower Manhattan is much more densely populated. In order to overcome these shortcomings, we introduce a new measure; the Spatial Population Concentration (SPC). The SPC intuitively captures the number of people weighted by a function of the radius d of every person within a given area, averaged across all people living within that area. More formally:

$$SPC_d = \frac{\sum_{i=1}^{N} pop_i * n_{di}}{\sum_{i=1}^{N} pop_i} - 1 \tag{1}$$

where SPC_d is the distribution measure for Euclidean distance d. pop_i is the number of people at position i and n_{di} is the number of people weighted by a function of their distance d from position i. This is a generalization of the economic density measure in Duranton and Puga (2020); Henderson et al. (2021); Roca and Puga (2017) which measures the special case of how many people live within a 10km radius divided by the area. As that measure divides by the area in question, it is a weighted average density. While our empirical analysis focuses on the number of people within a given radius, the SPC measure can be adapted to use for example a negative exponential distance weighting. When calculating SPC, we ignore area boundaries and include people living outside the specific area of interest, provided they live within distance d. This means that for some counties, our measure is affected by where people are living across the U.S. border in Canada and Mexico. Our results below use a square shape around each point but the resulting county level metric is virtually unchanged when we use a circle and a diamond shape for robustness.[4]

In order to illustrate how the SPC measure differs from population density, we show two counties with similar population density in Figure 2. Clark County, NV (Las Vegas) includes a substantial desert area around the city which reduces the population density. This causes it to have a similar density to Berrien County, MI, which does not include a major city. Due to not including a major city however, Berrien County has a much lower SPC measure than Clark County. We use George Washington Colonial One High Performance Computing System for our SPC calculations.[5] For the distance d, we use 10km as our baseline results but we also estimated the data for various other distances up to 200km. For higher thresholds, there

[4]This measure might also resemble the measure in Farrokhi and Jinkins (2019), where the distances are weighted by the population contrasting with our measure where the populations is weighted by the distances. In our euclidean space, distances are fixed and hence weighting by distances is superior when comparing the measure across time. This is because the Farrokhi and Jinkins (2019) measure might change for a county even if there is only a change in the population of very distant counties.

[5]https://colonialone.gwu.edu/

is little evidence that there are links between geography and growth (e.g., Bottazzi and Peri (2003)). The larger distances also address concerns that our 1km by 1km resolution maps might not be perfectly accurate and averaging across larger distances counteracts this. In addition, the 10km reference distance is largely preferred in the literature focusing on the scope of knowledge flows to capture the local effects (Bakhtiari and Breunig, 2018; Barrios et al., 2007; Baldwin et al., 2008; Holl et al., 2020). We then follow up by using radii of 15, 20, 25, 50, 75, 100, and 200km to estimate if the results are affected. As we focus on a county level analysis, the radii are also closely related to the county area. With a median county area of 1610 square kilometers, this is roughly equivalent to a radius of 23km, meaning that the determinants of the SPC measure at the 100km and 200km radii mainly reflect the population concentration outside the county.[6]

Conveniently, SPC can easily be aggregated into combined areas (e.g. to get from the county level SPC measure to a state level one). One weights the individual SPC measures according to the population share they have in the combined area.[7]

3 Empirical Strategy

We mainly rely on cross-sectional long-difference regressions as used for example in Jedwab and Vollrath (2015) as a baseline specification. For our sample of 3,119 U.S. counties we estimate the following equation:

$$\Delta LInc_{c,90-15} = \beta_0 + \beta_1 LSPCd_{c,90} + \gamma_1 X_{1c} + \gamma_2 X_{2c} + \mu_s + \epsilon_{c,90-15} \tag{2}$$

where $\Delta LInc_{c,90-15}$ is the log difference in income per capita between 1990 and 2015 for county c. $LSPPd_{c,90}$ is our variable of interest, the natural log of the SPC measure for the specific distance d in 1990 as defined in equation 1. X_{1c} includes our control variables related to the population distribution and output, namely log population in 1990, log Income per capita in 1990 and log area of each county. X_{2c} is our vector of additional infrastructure- and education-related control variables (see Table 7) to ensure that our results are robust and our estimation does not solely capture better infrastructure that might be highly correlated with population density. We also include state dummies μ_s to control for any differences across states like different

[6] There are 66 counties out of over 3000 that have an area smaller than the one implied by the 10km radius (314km^2) and most of them are in Virginia. Our results are robust to excluding them from the analysis.

[7] For example, assume two areas with SPC measures of 3 and 4 and populations of 2 and 1 respectively. Then the combined area will have an SPC measure of (3*2+4*1)/3=10/3.

rates of inflation. While we mainly focus on the SPC measure to test whether the geography matters for growth, the suggested channels also have clear implications for infrastructure and education. The better the infrastructure, the lower the transportation cost and the higher the education, the more high skilled workers there are, reducing the implementation cost also.

3.1 Spatial Dependency

An important additional effect we need to control for are growth spillovers. For example, Le Van et al. (2002); Rivera-Batiz and Romer (2008); Bottazzi and Peri (2003); Blazek and Sickles (2010); Desmet et al. (2018) show that areas which act as innovation growth engines will lead to faster growth in the surrounding area as well. The areas of fast growth might attract people and hence become areas with a high population concentration. One could argue that our SPC measure is merely capturing this phenomenon, rather than independent growth. In order to test this, we run Spatial Auto-Regressive models (SAR) that take the growth of nearby counties into account. For our spatial weighting matrix, we use the inverse of the distances between the geographic centers of counties. To check the spatial dependency, we first calculate Moran's I statistic (Moran, 1950), confirming that the growth rate (and GDP level) is indeed similar in counties that are located close to each other. Table 8 reports the results for this test. Moran's I is 0.409 and statistically significant at 1% level. Given the significance of spatial dependence, we include the spatial lag showing the degree of spatial autocorrelation to our specification (equation 2) throughout our regressions.[8]

4 Data

The data used to calculate the SPC measure come from the Global Human Settlement database.[9] These data provide spatial information on the number of people living with 1 km resolution for the entire world (1 x 1 km $\approx$ 20,004 x 40,004 cells) for the four distinct years: 1975, 1990, 2000 and 2015. The two alternative population distribution measures (population density and urban population) are obtained from the U.S. Census Bureau.[10] We do not explicitly include population density in our regression, but rather include both

[8] We also report the baseline results without spatial dependence in Table 10.

[9] https://ghsl.jrc.ec.europa.eu/

[10] Urban population is defined as people living in urbanized areas and urban clusters; https://www.census.gov/

the population and the area of each geographic area to allow for more flexibility. As all variables are included in logs, this is equivalent to controlling for population density separately. While our measure is available at the global level[11], we restrict our estimation to U.S. counties. This is because U.S. counties are in many ways more homogeneous than the countries of the world, data availability for control variables and state level dummies are available to capture many of the particularities in the U.S.

We calculate the SPC measure for the years 1975, 1990, 2000, and 2015 and get matching personal income data from the Bureau of Economic Analysis. Using income growth as a proxy for GDP growth is also in line with the literature (e.g. see Fulford et al. (2020)).[12] In addition to these geographic data, we also include data on infrastructure and education at the county level. The data sources are described in Table 7 and are mostly compiled by the USDA.[13] Specifically, access to a good transportation infrastructure can make travelling faster and more comfortable, reducing the implementation cost c. In order to measure this transportation infrastructure, we include the average distance of the county population center to the next airport as well as the miles of railroad and highways in a county. Since these data are not available as time-series at the county level, we use the year available to us (usually in the range 2015 - 2019). Since the early 2000s, communication technology also made large jumps with the advance of video-chat and mobile data that proved particularly useful during the COVID-19 pandemic (after our sample period). Being able to make video-calls might mitigate the need to meet in person to some extent and hence reduce the cost associated with the implementation cost. We control for this using the number of cell towers in each county as well as the broadband coverage in each county in percent.

As one channel requires high skilled workers to implement ideas. We want to control for this by including the education level in each county. Specifically, we control for the percentage of the population that has a college degree in each county. An additional factor regarding the implementation cost is the ease of starting and maintaining businesses. We proxy this by including the number of business establishments into our regression.

[11] `https://datacatalog.worldbank.org/search/dataset/0061191/spatial_population_concentration`
[12] Alternative data for GDP are only available 2001 onward at the county level.
[13] `https://www.ers.usda.gov/data-products/county-level-data-sets/`

5 Main Results

Table 1 contains the first set of regressions based on equation 2. The first column only includes geographic variables and the coefficient of 22.76 for the SPC measure implies that an additional 1% of people within a 10km radius of every person increases the growth over 25 years by 0.02%. In other words, doubling the population within a radius 10 km of every person increases the growth over 25 years by 2%. Given the wide range of values the SPC measure can take (e.g., see Figure 3), this is economically meaningful. Once additional controls like infrastructure measures are added, the coefficient drops to about half its size to around 14, but remains significant at the 10% level. Interestingly, while the area of a county has a statistically significant impact on output, the urbanization does not seem to have any impact, once SPC is included. This suggests that the SPC measure captures the population distribution better than urbanization.

Next, we compare the results across several distances as shown in Table 2. We find that the 25km radius has the largest effect, and the coefficient becomes insignificant for higher radii of 100km and has a negative and significant effect at the 10% level for 200km. At larger distances, it is also likely that the infrastructure and the other controls play a larger role. For example, a 10km radius is traversed within a quite reasonable time, while 100km might be just a 1h highway commute, but could also be substantially longer. For example, there might not be a highway available or there could be a mountain range in between, making the most direct route impossible and requiring a substantial detour. In addition, the 10km radius roughly corresponds to the median land area (around 320 square km) of the 150 largest U.S. cities.

For illustrative purposes, we standardize the indicators by dividing the SPC measure by 1000 and multiplying the population density by 1 million. Indeed SPC has a different distribution than population density (Figure 3) and urbanization (Figure 4). To compare the economic impact of SPC for different distances with the magnitude of alternative measures, we multiply the coefficient by the standard deviation of the underlying variable. Table 9 presents the impacts. A one standard deviation increase in the SPC measure for 10km increases the growth by 2.42% over the period 1990 - 2015 while a one standard deviation increase in population, area, and urban population changes the growth by -28.84%, 1.39%, and -1.41% in order. These numbers suggest that SPC has a stronger relationship with growth than population density or urbanization.[14]

[14] Population density (assuming the area impact) and urbanization both have a negative relationship with growth that is smaller than the positive relationship with SPC in absolute terms across almost all distances.

The economic impact of the SPC is largest at 25km distance (7.63%), decreases with additional distance, and has a negative impact on growth at 200km (-0.50%). The average impact of the SPC measure for distances less than 100km is in the range between 2.07% (75km) and 7.63% (25km).

6 Robustness

6.1 Interaction Effects

SPC and Education. So far, it was assumed that the SPC measure has an effect independent from other variables. In order to distinguish whether the location of high skilled workers matters more for growth or the location of potential consumers, we interact our SPC measure with the share of college educated population in the counties. If the interaction effect is very large, this would be more in line with the former channel, while an insignificant interaction effect is more in line with the latter.

Table 3 shows that the interaction effect of SPC and the share of high skilled workers is significant at all distances except for 200km. This implies that the effect of how closely people live together is stronger if a larger share of a county's population is high skilled. This result also strengthens the argument of the commute for skilled workers as an important spatial component of growth rather than the location of consumers. While experts typically have a higher education than other people in the population a higher density of them should increase innovation. This argument is more difficult to make with regards to the potential market for a new product. While for some products it might be the case that the potential market is larger due to more educated people, it is more difficult to make the case that this is generally the case.

SPC and Digital Measures. Some of the recent literature on growth has concentrated on the role of digital technologies as one important carrier of knowledge spillovers and the importance of digitalization for the economic growth of regions and countries (Batabyal and Nijkamp, 2016; Liu and San, 2006; Vinciguerra et al., 2011). In addition to transport networks such as airline routes and high speed railways, people are well connected to each other through the mobile network and the internet. Hence, we look at the impact of both the total number of cell towers (3G and 4G) in a county and the percentage of broad band coverage in Table 3. While the interaction is positive and significant for cell towers for all distances, it is only significant at distances 25 - 50km for broadband. One potential explanation for why the former one has a lower magnitude could be that cell phone coverage is more flexible. It can be used to substitute broadband access with the added benefit that the area covered by a cell tower is likely larger than the area covered by one broadband outlet. The distance with the most impact is the same as the one for the transportation infrastructure (50km). This is in line with the assumption that technologies like video call can somewhat substitute the need to travel over distances. This result could both be in line with the commute of workers or the potential market. Particularly since COVID, many high skilled jobs have moved online. This means that good digital infrastructure allows to perform some work remotely, removing the need to a commute. Similarly, certain services can be performed over the internet and a good infrastructure might be able to substitute some of the need to physically deliver the service.

SPC and Number of Businesses. There can be other factors that might potentially impacting the relationship between SPC and growth. For instance, having more businesses located in counties can facilitate the rapid exchange of ideas and increase the size of a potential market. Therefore, we also consider the interaction between SPC and the number of business establishments that we use as a proxy decreasing the implementation cost of an idea. As reported in Table 3 having more businesses in populous counties have a positive and significant impact on GDP for all distances, 10 -200km.

SPC and Transport Measures. People living closer reduces the transportation cost. However, the transportation cost is also reduced if there are means of transportation available that allow for faster movement. Indeed, a lower cost of transportation has been found to contributes to economic growth as it allows knowledge transfer over larger distances (Campante and Yanagizawa-Drott, 2016; Dehghan Shabani and Safaie, 2018; Tamura, 2017). Therefore, even if people live in sparsely populated areas but can connect with people through a rapid transportation network, the transportation cost is low. This is the next variable whose interaction with SPC we analyze. Specifically, we use the distance to the closest airport (in km), total miles of highway and railroad in interaction with SPC.

Table 4 reports the interaction effects between SPC and the transport measures. The coefficients show a clear positive and significant effect with both channels. Total highway and railway lengths have positive spatial effect on growth. In addition, living in both populous areas that are close to airport contributes to growth positively. The interaction term is the largest for a distance around 50km which is higher than the individual effect of SPC (20-25km). This suggests that infrastructure helps the most if it connects people in that intermediate range, rather than at the very close range or the very long range. This is in line with both channels.

6.2 Panel Estimation

Aside from the cross-section estimations, we repeat the above analysis in a panel including the years 1975, 1990, 2000, and 2015 where we calculate the SPC measure. As our measure is not just the population divided by a fixed land mass, unlike the population density, we can include our variable into a panel regression. Then we run a panel for the four years according to the equation:

$$\Delta LInc_{c,t} = \beta'_0 + \beta'_1 L\Delta LSPCd_{c,t} + \beta'_2 \Delta Lpop_{c,t} + \theta_c + \eta_t + \mu_{c,t} \tag{3}$$

Our cross-section results suggest that a high SPC measure today leads to faster growth over the next 10-15 years. This means that there is a lead-lag relationship between SPC and income which needs to be reflected in the panel regression as well. We thus include the lagged term for the SPC measure into our regression and run the panel in log differences. Table 6 presents the estimated coefficients for the spatial population concentration measures. Our panel estimation results suggest that how close people live together is closely related to subsequent growth. Similar to the cross-section results, we find very strong significance at the 10km radius. At longer radii, the relationship becomes weaker and at the 200km radius it becomes negative and significant. As mentioned for the cross-section results, the negative coefficient might be due to the area surrounding the county driving our results, rather than the county itself.

7 Conclusion

The geographic distribution of people matters for the knowledge creation hence for the economic growth. We reconfirm empirically that areas in which people live close together grow faster by introducing a new measure, Spatial Population Concentration, to rigorously capture the geographic closeness of people. The estimation results indicate that SPC measure is statistically significant in explaining growth across U.S. counties. Specifically, our results indicate that counties with a low value of the SPC measure in 1990 experienced substantially lower GDP growth over the next 25 years.

Our findings show the biggest relationship between SPC and growth at a radius of 25km. The radius roughly doubles when interacted with physical infrastructure like transportation and communication to around 50km. Hence, U.S. policymakers who want to accelerate economic growth should strive to boost the population density for that range and avoid too much city sprawling. Given the importance of infrastructure, its maintenance and expansion should also be a focus of policymakers, particularly in the 25-50km range. This range also suggests that having several very densely populated hubs should be preferable over a large area of intermediate density population. In addition, we find that education plays a key role and has the highest growth impact in the densely populated areas and should be a priority there.

Findings in this study give avenues for future work. The current estimates of the Spatial Population Concentration measure capture the average effect for the U.S. However, one might be interested to repeat the analysis at a world level. This requires control for additional variables such as internal wars and ethnic or religious segregation at the country or subnational level which impact the knowledge creation and GDP growth in those regions through their impacts on interactions among people. While there is no such concern for the U.S. during the study period, the barriers that can potentially hinder people exchanging ideas can be considered in different settings.

The U.S. and many other countries have experienced the great economic and industrial divergence throughout the time. Since the value-added data at the county level is available for the period 2000-2015, our study period is too short to control for sectoral shifts from manufacturing to high technology services. Yet depending on the availability of granular data, doing the analysis for different time periods would be another possible research topic.

References

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica: Journal of the Econometric Society*, pages 323–351.

Altunbas, Y., Jones, E., and Thornton, J. (2013). Knowledge spillovers and the growth of british cities. *Applied Economics Letters*, 20(1-3):162 – 166.

Bakhtiari, S. and Breunig, R. (2018). The role of spillovers in research and development expenditure in australian industries. *Economics of Innovation and New Technology*, 27(1):14–38.

Baldwin, J., Beckstead, D., Brown, W., and Rigby, D. (2008). Agglomeration and the geography of localization economies in canada. *Regional Studies*, 42:117–132.

Barrios, S., Bertinelli, L., Heinen, A., and Strobl, E. (2007). Exploring the link between local and global knowledge spillovers. *University Library of Munich, Germany, MPRA Paper*.

Batabyal, A. A. and Nijkamp, P. (2016). Digital technologies, knowledge spillovers, innovation policies, and economic growth in a creative region. *Economics of Innovation and New Technology*, 25(5):470–484.

Blazek, D. and Sickles, R. C. (2010). The impact of knowledge accumulation and geographical spillovers on productivity and efficiency: The case of u. s. shipbuilding during wwii. *Economic Modelling*, 27(6):1484 – 1497.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2017). Are ideas getting harder to find? Technical report, National Bureau of Economic Research.

Bottazzi, L. and Peri, G. (2003). Innovation and spillovers in regions: Evidence from european patent data. *European Economic Review*, 47(4):687 – 710.

Campante, F. and Yanagizawa-Drott, D. (2016). Long-Range Growth: Economic Development in the Global Network of Air Links. NBER Working Papers 22653, National Bureau of Economic Research, Inc.

Carlino, G. A. (2001). Knowledge spillovers: Cities' role in the new economy. *Federal Reserve Bank of Philadelphia Business Review*, pages 17 – 26.

Charlot, S. and Duranton, G. (2003). Communication Externalities in Cities. CEP Discussion Papers dp0592, Centre for Economic Performance, LSE.

Chen, M., Zhang, H., Liu, W., and Zhang, W. (2014). The global pattern of urbanization and economic growth: evidence from the last three decades. *PloS one*, 9(8):e103799.

Conley, T. G. and Ligon, E. (2002). Economic Distance and Cross-Country Spillovers. *Journal of Economic Growth*, 7(2):157–187.

Davis, D. R. and Dingel, J. I. (2019). A spatial knowledge economy. *American Economic Review*, 109(1):153–70.

de Groot, H. L. F., Poot, J., and Smit, M. J. (2007). Agglomeration, innovation and regional development: Theoretical perspectives and meta-analysis.

Dehghan Shabani, Z. and Safaie, S. (2018). Do transport infrastructure spillovers matter for economic growth? evidence on road and railway transport infrastructure in iranian provinces. *Regional science policy & practice*, 10(1):49–63.

Desmet, K., Greif, A., and Parente, S. (2018). Spatial competition, innovation and institutions: The industrial revolution and the great divergence.

Desmet, K., Greif, A., and Parente, S. L. (2020). Spatial competition, innovation and institutions: the industrial revolution and the great divergence. *Journal of Economic Growth*, 25(1):1–35.

Duranton, G. and Puga, D. (2020). The economics of urban density. *Journal of economic perspectives*, 34(3):3–26.

Farrokhi, F. and Jinkins, D. (2019). Wage inequality and the location of cities. *Journal of Urban Economics*, 111:76–92.

Fulford, S. L., Petkov, I., and Schiantarelli, F. (2020). Does it matter where you came from? ancestry composition and economic performance of us counties, 1850–2010. *Journal of Economic Growth*, 25(3):341–380.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (2009). *Growth in Cities.*, pages 451 – 477. Elgar Reference Collection. International Library of Entrepreneurship, vol. 14. Cheltenham, U.K. and Northampton, Mass.: Elgar, Unlisted.

Gollin, D., Jedwab, R., and Vollrath, D. (2016). Urbanization with and without industrialization. *Journal of Economic Growth*, 21(1):35–70.

Gong, G., Greiner, A., and Semmler, W. (2004). Endogenous growth: Estimating the romer model for the us and germany. *Oxford Bulletin of Economics and Statistics*, 66(2):147 – 164.

Grossman, G. M. and Helpman, E. (1991). Quality ladders in the theory of growth. *The review of economic studies*, 58(1):43–61.

Henderson, J. V., Nigmatulina, D., and Kriticos, S. (2021). Measuring urban economic density. *Journal of Urban Economics*, 125:103188.

Henderson, J. V., Squires, T., Storeygard, A., and Weil, D. (2018). The Global Distribution of Economic Activity: Nature, History, and the Role of Trade. *The Quarterly Journal of Economics*, 133(1):357–406.

Holl, A., Peters, B., and Rammer, C. (2020). Local knowledge spillovers and innovation persistence of firms. ZEW Discussion Papers 20-005, ZEW - Leibniz Centre for European Economic Research.

Jedwab, R., Christiaensen, L., and Gindelsky, M. (2017). Demography, urbanization and development: Rural push, urban pull and… urban push? *Journal of Urban Economics*, 100(98):6–16.

Jedwab, R. and Vollrath, D. (2015). Urbanization without growth in historical perspective. *Explorations in Economic History*, 58:1–21.

Jones, C. I. (1995). R & d-based models of economic growth. *Journal of political Economy*, 103(4):759–784.

Le Van, C., Morhaim, L., and Dimaria, C.-H. (2002). The discrete time version of the romer model. *Economic Theory*, 20(1):133 – 158.

Liu, M. and San, G. (2006). Social learning and digital divides: A case study of internet technology diffusion. *Kyklos (Basel)*, 59(2):307–321.

Malik, A. and Temple, J. R. (2009). The geography of output volatility. *Journal of Development Economics*, 90(2):163–178.

Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, 107(2):407–437.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1-2):17–23.

Moreno, R. and Trehan, B. (1997). Location and the growth of nations. *Journal of Economic Growth*, 2(4):399–418.

Perumal, A. (2017). 42 years of urban growth and industry composition. *Atlantic Economic Journal*, 45(2):133 – 147.

Rivera-Batiz, L. A. and Romer, P. M. (2008). *Economic Integration and Endogenous Growth.*, pages 291 – 315. Elgar Reference Collection. Globalization of the World Economy series, vol. 20. Cheltenham, U.K. and Northampton, Mass.: Elgar, Unlisted.

Roca, J. D. L. and Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, 84(1):106–142.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, 98(5, Part 2):S71–S102.

Romer, P. M. (1994). The origins of endogenous growth. *Journal of Economic Perspectives*, 8(1):3 – 22.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics*, 70(1):65–94.

Tamura, R. (2017). The Effect of High-speed Railways on Knowledge Transfer: Evidence from Japanese Patent Citations. *Public Policy Review*, 13(3):325–342.

Vinciguerra, S., Frenken, K., Hoekman, J., and van Oort, F. (2011). European infrastructure networks and regional innovation in science-based technologies. *Economics of innovation and new technology*, 20(5):517–537.

Zeira, J. (2011). Innovations, patent races and endogenous growth. *Journal of Economic Growth*, 16(2):135–156.

A Model

Our model is a mix between the growth model proposed by Romer (1990) and creative destruction models as described for example in Aghion and Howitt (1992) but with a more extensive labor market. We assume that there is a continuum of agents i with $n+1$ types. The first n are experts in field k and the last type is an agent without expertise in any field. In each period t, every agent has an idea in the field x_{it} from the same n possible fields every period. It is assumed that there is no relation between the field of an expert and the field of the idea and there is no ex ante difference in the properties of the ideas experts have and the ones of regular agents. The agent can choose to either become an entrepreneur and implement her idea, or to become a wage worker. If an agent chooses not to become an entrepreneur, the agent earns the competitive wage w_{lt} as a regular worker and the wage w_{kt} as an expert. If there are more experts in a field k than are needed to implement the equilibrium number of ideas, $w_{kt} = w_{lt}$ and the excess number of experts become regular workers. Agents that choose to become an entrepreneurs will implement their idea by starting a company in their current location and produce an output variety. They earn the profits of this company which are given by

$$\pi_{it} = A_{it}(x_{it})l_{it}^{\alpha} - w_{lt}l_{it} - c_{it} - z_{it}A_{it}(x_{it})l_{it}^{\alpha} \tag{4}$$

The productivity term A_{it} is stochastic and reflects the profitability of the idea. l_{it} is the number of workers employed at the market wage w_{lt}. $\alpha < 1$ ensures that the company has decreasing returns to scale, c_{it} is the implementation cost of the idea due to an expert having to commute or move and $z_{it} \in [0, 1)$ is the implementation cost due to the goods having to be transported to the consumers. Entrepreneurs chose the number of workers for a given wage to maximize profits

$$w_{lt} = \alpha A_{it}(x_{it})l_{it}^{\alpha-1}(1 - z_{it}) \tag{5}$$

The choice between becoming a worker and becoming an entrepreneur is determined by the profit they would make if a company was started. If the profit is less than the market wage ($\pi_{it} < w_{lt}$), a regular agent does not become an entrepreneur. If the expert wage with field k is larger than the profit of the potential company ($\pi_{it} < w_{kt}$), the expert does not become an entrepreneur. Both wages are endogenously determined by the labor demand and supply. It is assumed that a percentage p_{kt} of agents are experts in field k that can help implement ideas that match this field. In order to implement an idea, an entrepreneur requires a matching expert. For example, if the idea is a new component for a car, an automobile engineer might be the expert needed.

A.1 Movement Cost c

While the production of output variety using the production function

$$y_{it} = A_{it}(x_{it})l_{it}^{\alpha} \tag{6}$$

can use any other agent as a worker, it is necessary to get one expert with field k to implement the idea and pay him the market wage w_{kt}. It is assumed that experts need to physically move to the entrepreneur in order to implement an idea, the cost of which is paid by the entrepreneur. Depending on how close the expert is to this entrepreneur, the cost for moving is higher or lower. The movement cost can either be interpreted as a commuting cost for small distances or as a relocation cost for long distances. The movement cost does not only include the wage w_{kt} and the transportation cost but also the opportunity cost. For example, a worker relocating might live further from relatives, making visits to them more expensive. We cap the cost of moving the expert at the market wage w_{lt}. This implies that in the worst case, an entrepreneur needs to pay a cost $c_{it} = w_{lt} + w_{kt}$ to implement an idea and in the best case pays $c_{it} = w_{kt}$. For most entrepreneurs, the cost is somewhere inbetween and depends directly on the distance between agents at the aggregate level. If the entire population of a county lives in the same place, the cost is the going wage w_{kt}. There are several distributional aspects that can affect the cost c_{it}. Specifically, this cost for implementing an idea can depend on the percentage of experts for the specific idea in the population (the more, the lower the cost), how concentrated they are (the less concentrated, the lower the cost), and how many people live within a certain radius d_t of every agent (the more people, the lower the cost).[15]

[15] In order to simulate this economy, one might make further assumptions about the model for tractability like that productivity A_{it} is geographically independently distributed from the implementation cost c_{it}. Also, one might restrict that only non-experts can have ideas that can potentially be implemented and that the number of ideas is equal to the number of experts.

A.2 Potential Market Cost z

Each output

$$y_{it} = A_{it}(x_{it})l_{it}^{\alpha} \tag{7}$$

that has been produced, needs to be distributed to consumers. The further away the potential consumer lives, the more expensive the distribution. If the consumer lives far enough, it might not be worthwhile to produce for this consumer. As a result, the closer people live together, the more potential consumers there are and the more profitable it is to implement an idea. This is very similar to the model in Desmet et al. (2020) which has a potential market for each good produced that depends on the population distribution. The cost z_{it} for implementing an idea can depend on the percentage of consumers for the specific good in the population (the more, the lower the cost), how concentrated they are (the less concentrated, the lower the cost), and how many people live within a certain radius d_t of every agent (the more people, the lower the cost). This is very similar to the cost of moving labor c_{it}. Indeed, both costs depend on the spatial population concentration and a higher concentration results in more ideas being implemented.

One can simplify the model by only including one of the two sources for the spatial distribution of the population to matter and setting the other equal to zero, or one can include both.

A.3 Growth

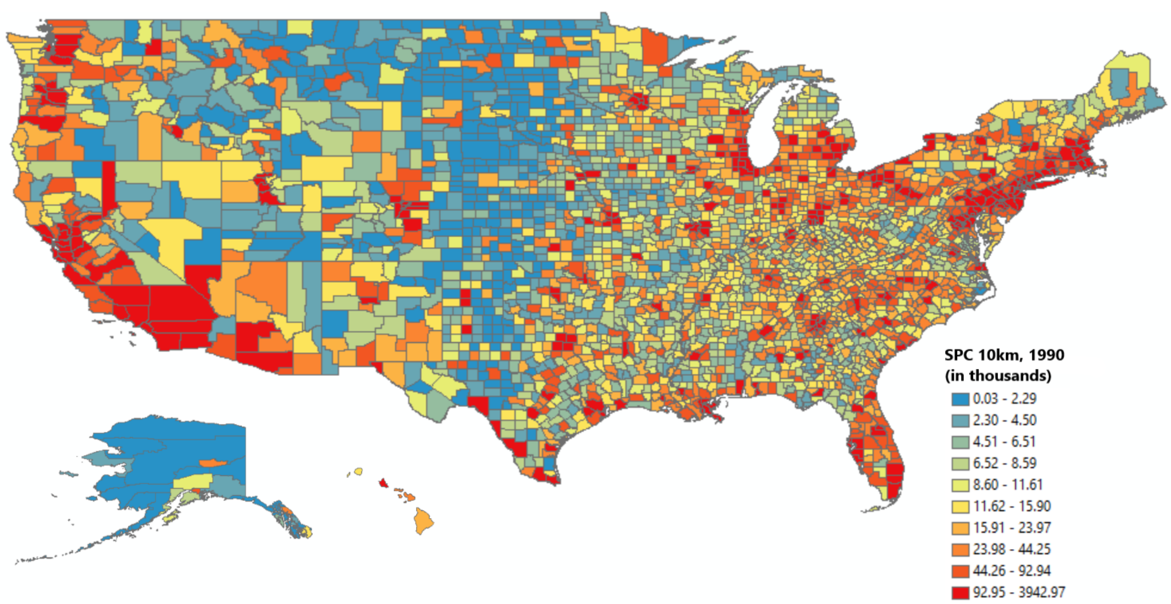
So far, the model can compare of how developed regions are, depending on the geographic distribution of their population. This leads to different wage and migration patterns as well. To obtain the impact of growth over time, it is necessary to specify how growth enters this model. In line with the literature endogenous growth (e.g. Romer (1990)), we choose that research and development increases the productivity of firms. In our model, the closest to this is the implementation cost c. We assume that only the wage of the specialist is an expense in research and development rather than the movement cost. This way, two counties with the same number of ideas implemented but different movement costs experience the same productivity growth. If the movement cost was instead included, the county with the higher movement cost would experience faster productivity growth. As a result, we assume that the expected value of productivity $E(A_t)$ evolves as follows:

$$E(A_{t+1}) - E(A_t) = f\left(W_{kt}\right) \tag{8}$$

where W_{kt} is the sum of all wages w_{kt} paid to experts as part of the implementation cost c and $f(\cdot)$ is some positive and increasing function. This setup ensures that the more ideas are implemented, the faster growth is achieved. Because more ideas implemented means more experts are hired by companies, the expenditure is directly related to the number of people working in research and development, which is endogenously determined in the model. Specifically, the growth depends on the distribution of $A_{it}(x_{it})$ as well as the distribution of experts and consumers (c_{it} and z_{it}) and together, they determine the wage w_t as well as the number of ideas implemented h_t.

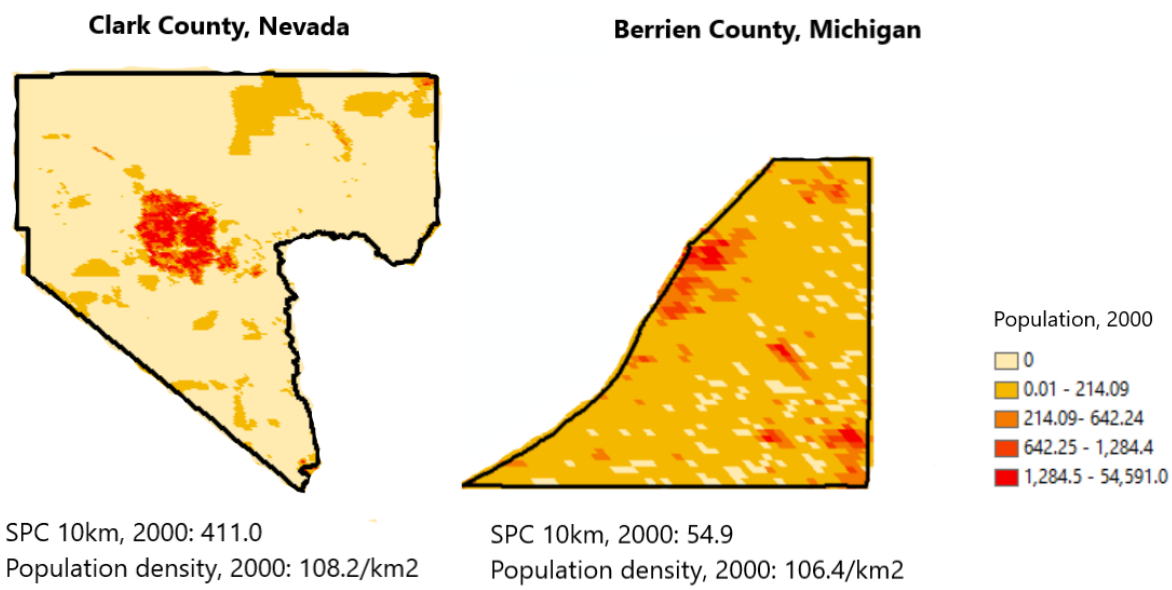
Figures and Tables

Figure 1: Distribution of SPC measure for 10km distance in 1990



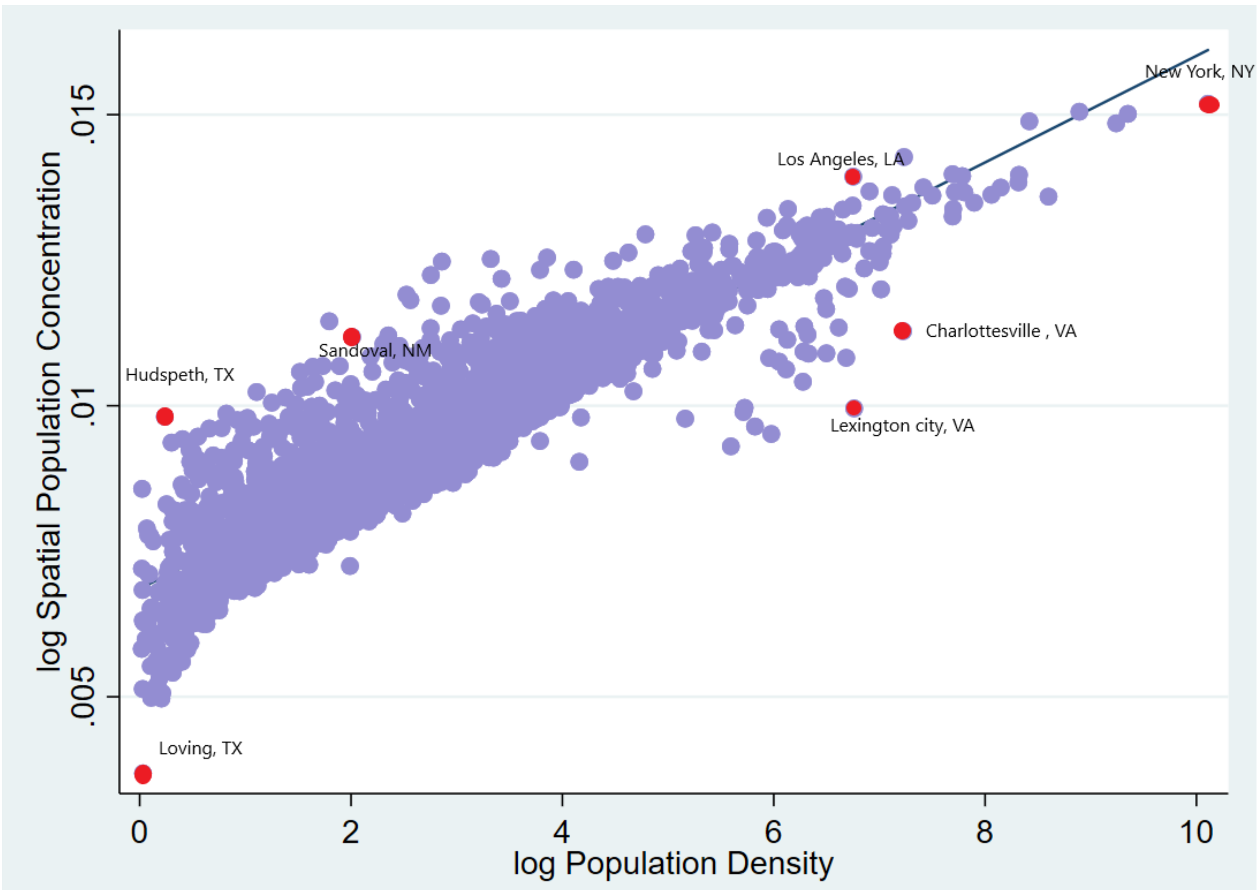
Note: This map shows the SPC measure for 10 km distance in 1990. It is drawn based on quantiles. The counties shown in blue color are the ones people more spread out compared to the countries shown in red where people concentrate (due to geographic or some other factors).

Figure 2: Comparison of Spatial Population Concentration and Population Density



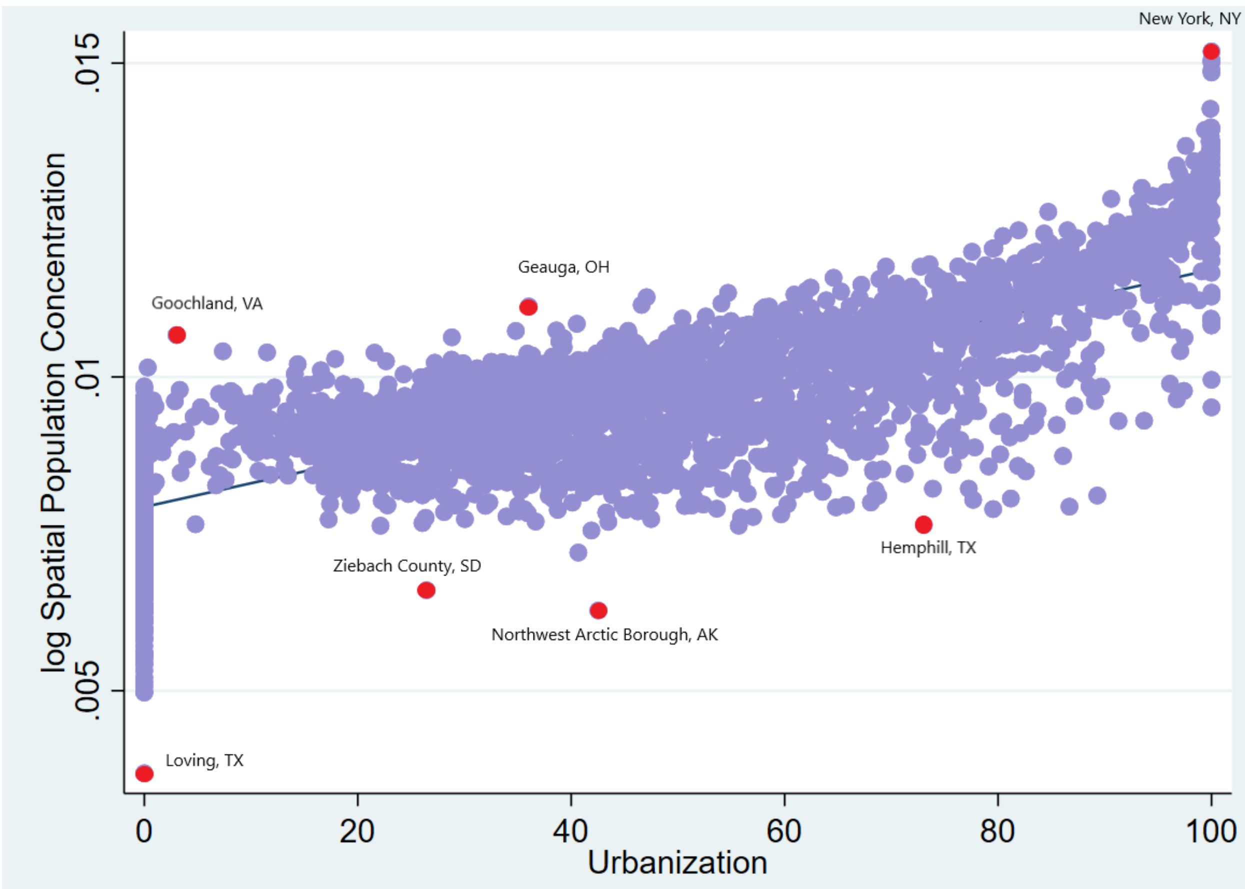
Note:This figure shows the comparison of SPC measure for 10km and the population density in two counties. As shown in the maps light yellow areas are uninhabited whereas red color represents the most populous places. In Berrien, Michigan people are spread out across the county. On the other hand, in Clark county, Nevada people mostly concentrated.

Figure 3: Comparison of Spatial Population Concentration and Population Density across U.S. counties



Note: This figure shows the comparison of SPC and Population density measures. SPC measure captures the spatial population concentration in 1990 for 10km (baseline distance). SPC measure is less skewed whereas population density increases quickly. This implies that mean of SPC is better for where the average country is. Selected counties are labelled for illustrative purposes.

Figure 4: Comparison of Spatial Population Concentration and Urban Population across U.S. counties



Note: This figure shows the comparison of SPC and Population density measures. SPC measure captures the spatial population concentration in 1990 for 10km (baseline distance). Selected counties are labelled for illustrative purposes.

Table 1: Effect of Spatial Population Concentration on Income in the U.S.: Cross-Section (SAR)

Dependent variable:	log change in income per capita				
Periods:	1990-2015	1990-2015	2000-2015	1975-2015	1975-2000
log SPC: 10km	17.31***	11.56**	8.17	27.68***	13.74***
	(6.11)	(5.71)	(5.26)	(6.32)	(5.05)
log Δ Population	0.01	-0.13***	-0.15***	-0.26***	-0.19***
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
log Population	-0.03***	-0.20***	-0.17***	-0.34***	-0.19***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log Income	-0.13***	-0.36***	-0.28***	-0.67***	-0.58***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
log Area	1.60	3.05	8.55*	-12.27**	-26.53***
	(5.66)	(5.26)	(4.82)	(5.95)	(4.76)
log Rail		0.60	2.38	-0.62	-2.26
		(1.79)	(1.63)	(2.05)	(1.63)
log Dist.Airport		-0.03	0.01	4.07	5.69**
		(3.18)	(2.92)	(3.63)	(2.90)
log N.Businesses		0.15***	0.11***	0.28***	0.20***
		(0.01)	(0.01)	(0.01)	(0.01)
log Highway		-0.62	-0.04	-0.29	-0.34
		(0.57)	(0.52)	(0.65)	(0.52)
log Cell Towers		0.02***	0.02***	0.03***	0.01***
		(0.001)	(0.001)	(0.001)	(0.001)
Broadband Coverage		0.13***	0.13***	0.19***	0.08***
		(0.02)	(0.02)	(0.02)	(0.02)
Education		4.08***	3.05***	4.62***	5.58***
		(0.55)	(0.45)	(0.86)	(0.69)
log Urban Population	-2.22**	-3.49***	-1.68**	-5.98***	-5.00***
	(0.93)	(0.88)	(0.81)	(0.99)	(0.79)
Spatial Lag	-0.23***	-0.12***	-0.34***	-0.02	0.11***
	(0.03)	(0.03)	(0.05)	(0.01)	(0.01)
Observations	3,078	3,072	3,075	3,057	3,057
Wald χ^2 (p)	42.57 (0.00)	13.49 (0.00)	34.77 (0.00)	1.70 (0.19)	38.47 (0.00)
Spatial FE	Yes	Yes	Yes	Yes	Yes

This table shows the long-difference cross-section estimation results for the impact of Spatial Population Concentration (SPC) on the change in income per capita for various periods for the U.S. counties. Each column is a separate regression for different time period. Columns (1) is the baseline results. In columns (2) - (5) we add control variables. The Wald test χ^2 (p-val in parentheses) measures the joint significance of the excluded instruments.Robust standard errors are shown in parentheses. The details on control variables are presented in Table 7.
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Effect of Spatial Population Concentration on Growth in the U.S.: Various Distances (SAR)

Dependent variable:	log change in income per capita (1990 - 2015)						
SPC Distance (d):	(15 km)	(20 km)	(25 km)	(50 km)	(75 km)	(100 km)	(200 km)
log SPC 1990	23.46***	36.34***	50.93***	31.36***	15.99***	8.00**	-4.63
	(5.06)	(4.69)	(4.61)	(3.78)	(3.65)	(3.77)	(4.61)
log Δ Population	-0.14***	-0.18***	-0.18***	-0.17***	-0.15***	-0.13***	-0.13***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log Population	-0.21***	-0.22***	-0.25***	-0.22***	-0.20***	-0.19***	-0.19***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
log Income	-0.37***	-0.37***	-0.39***	-0.38***	-0.37***	-0.37***	-0.36***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
log Area	10.32*	19.00***	28.02***	16.53***	6.61	1.65	-3.21
	(5.30)	(5.29)	(5.26)	(5.06)	(4.99)	(4.92)	(4.69)
log Urban Population	-3.51***	-3.21***	-2.57***	-2.31***	-2.74***	-2.90***	-3.09***
	(0.85)	(0.84)	(0.83)	(0.84)	(0.85)	(0.85)	(0.85)
Spatial Lag	-0.12***	-0.12***	-0.14***	-0.16***	-0.15***	-0.14***	-0.11***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Observations	3,072	3,072	3,072	3,072	3,072	3,072	3,072
Wald χ^2	13.79	15.81	20.05	25.52	21.25	18.30	12.16
(p - value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Spatial FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

This table shows the long-difference cross-section estimation results for U.S. counties for various distances. Each column is a separate regression where the impact of Spatial Population Concentration (SPC) on the change in personal income per capita for the period 1990 - 2015 is estimated. The table reports the coefficients of the variables in baseline estimation. The Wald test χ^2 (p-val in parentheses) measures the joint significance of the excluded instruments.Robust standard errors are shown in parentheses. All regressions include control variables presented in Table 7.
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Effect of Spatial Population Concentration on Growth in the U.S.: Interactions 1 (SAR)

Dependent variable:	log change in income per capita (1990 - 2015)					
SPC Distance (d):	(10 km)	(25 km)	(50 km)	(75 km)	(100 km)	(200 km)
log SPC 1990	-12.87*	21.64***	5.54	-6.83	-12.19**	-19.01***
	(6.62)	(5.58)	(5.06)	(5.20)	(5.51)	(6.51)
Education	-1.34	-4.61***	-6.31***	-7.49***	-7.87***	-6.42*
	(1.74)	(1.79)	(2.08)	(2.39)	(2.78)	(3.59)
SPC x Education	737***	901***	984***	1033***	1018***	833***
	(163)	(149)	(160)	(175)	(195)	(233)
Pseudo R-squared	0.47	0.49	0.48	0.47	0.47	0.47
log SPC 1990	-29.69***	10.25	-11.72*	-28.07***	-35.89***	-40.18***
	(8.05)	(6.83)	(6.91)	(7.52)	(8.22)	(9.74)
log N.Businesses	0.09***	0.10***	0.09***	0.07***	0.06***	0.06***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
SPC x N.Businesses	6.64***	6.87***	7.04***	7.09***	7.07***	5.98***
	(0.92)	(0.86)	(0.95)	(1.06)	(1.18)	(1.45)
Pseudo R-squared	0.44	0.46	0.45	0.44	0.44	0.43
SPC and Digital Measures						
log SPC 1990	-9.90	29.54***	8.21	-5.92	-12.83**	-18.06***
	(6.68)	(5.44)	(5.06)	(5.28)	(5.59)	(6.47)
log Cell Towers	-0.04***	-0.05***	-0.06***	-0.06***	-0.06***	-0.04**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
SPC x Cell Towers	5.55***	6.06***	6.24***	5.79***	5.57***	3.89***
	(0.91)	(0.84)	(0.92)	(1.01)	(1.11)	(1.32)
Pseudo R-squared	0.44	0.46	0.45	0.44	0.44	0.43
log SPC 1990	3.95	34.53***	11.32	0.44	-1.74	-3.14
	(8.84)	(7.34)	(7.00)	(7.30)	(7.68)	(8.80)
Broadband Coverage	0.01	-0.18*	-0.29**	-0.21	-0.09	0.17
	(0.11)	(0.11)	(0.12)	(0.14)	(0.16)	(0.20)
SPC x Broad. Cover.	12.27	28.12***	34.09***	26.27**	16.59	-2.63
	(10.88)	(9.79)	(10.04)	(10.69)	(11.40)	(13.22)
Pseudo R-squared	0.43	0.45	0.45	0.44	0.43	0.43
Observations	3,072	3,072	3,072	3,072	3,072	3,072

This table shows the interaction effect of SPC measure with different interactions for the U.S. counties for various distances. Each panel and column is a separate regression. Robust standard errors are shown in parentheses. The details on control variables are presented in Table 7. The results of the Wald test is statistically significant at 1 percent for all estimations. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Effect of Spatial Population Concentration on Growth in the U.S.: Interactions 2 (SAR)

Dependent variable:	log change in income per capita (1990 - 2015)					
SPC Distance (d):	(10 km)	(25 km)	(50 km)	(75 km)	(100 km)	(200 km)
			SPC and Transport Measures			
log SPC 1990	6.67	44.86***	24.48***	10.69***	2.98	-8.96*
	(6.09)	(4.88)	(4.12)	(4.05)	(4.21)	(5.00)
log Highway	-9.31**	-15.77***	-19.20***	-15.40***	-15.28***	-16.06**
	(3.85)	(4.04)	(4.52)	(5.02)	(5.51)	(6.96)
SPC x Highway	939**	1384***	1530***	1142***	1081***	1024**
	(411)	(375)	(371)	(384)	(402)	(458)
Pseudo R-squared	0.43	0.46	0.45	0.44	0.43	0.43
log SPC 1990	43.04***	86.18***	77.35***	62.09***	48.86***	24.33**
	(8.67)	(7.98)	(8.20)	(8.58)	(9.15)	(10.36)
log Dist. Airport	-80.29***	-101.78***	-134.79***	-138.30***	-125.64***	-96.45***
	(16.98)	(18.02)	(21.12)	(23.46)	(25.93)	(31.36)
SPC x D.Airport	7892***	8545***	10797***	10699***	9372***	6535***
	(1640)	(1582)	(1711)	(1804)	(1912)	(2093)
Pseudo R-squared	0.44	0.46	0.45	0.44	0.44	0.43
log SPC 1990	1.81	40.75***	22.09***	8.83*	-0.23	-10.45*
	(6.75)	(5.38)	(4.86)	(4.85)	(5.04)	(5.90)
log Rail	-24.91***	-34.96***	-33.44***	-27.58**	-33.86**	-27.80
	(9.61)	(9.93)	(11.41)	(12.77)	(14.13)	(18.11)
SPC x Railway	2924**	3514***	2925***	2242**	2597**	1909
	(1082)	(962)	(968)	(1004)	(1055)	(1211)
Pseudo R-squared	0.43	0.46	0.45	0.44	0.43	0.43
Observations	3,072	3,072	3,072	3,072	3,072	3,072

This table shows the interaction effect of SPC measure with different interactions for the U.S. counties for various distances. Each panel and column is a separate regression. Robust standard errors are shown in parentheses. The details on control variables are presented in Table 7. The results of the Wald test is statistically significant at 1 percent for all estimations. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Effect of Spatial Population Concentration on Income per capita in the U.S.: Spatial Autoregressive Model (SAR) - excluding populous counties

Dependent variable:	log change in income per capita				
Periods:	1990-2015	1990-2015	2000-2015	1975-2015	1975-2000
log SPC: 10km	21.55***	13.99**	8.24	31.62***	18.64***
	(6.24)	(5.84)	(5.44)	(6.36)	(5.03)
Spatial Lag	-0.20***	-0.10***	-0.30***	-0.02	0.09***
	(0.03)	(0.03)	(0.05)	(0.01)	(0.01)
Observations	2,922	2,917	2,920	2,902	2,902
Wald χ^2 (p)	34.36 (0.00)	10.36 (0.00)	26.94 (0.00)	2.04 (0.15)	30.26 (0.00)
Spatial FE	Yes	Yes	Yes	Yes	Yes
Control Variables	Spatial only	All	All	All	All

This table shows the long-difference cross-section estimation results for the impact of Spatial Population Concentration (SPC) on the change in income per capita for various periods for the U.S. counties. Each column is a separate regression for different time period. Columns (1) is the baseline results. In columns (2) - (5) we add control variables. Robust standard errors are shown in parentheses. The details on control variables are presented in Table 7. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Effect of Spatial Population Concentration on Income per capita in the U.S.: Panel

Dependent variable:	log change in income per capita					
SPC Distance (d):	(10 km)	(25 km)	(50 km)	(75 km)	(100 km)	(200 km)
log SPC	850.98***	10.23**	506.42***	208.13	33.72	-644.28***
	(118.96)	(4.90)	(170.06)	(154.35)	(154.52)	(189.28)
log Population	-0.12	-0.74***	-0.53***	-0.68***	-0.74***	-0.87***
	(0.12)	(0.08)	(0.11)	(0.10)	(0.10)	(0.09)
Education	-0.50	-0.03	-0.10	-0.07	-0.00	0.49
	(1.82)	(1.88)	(1.86)	(1.87)	(1.88)	(1.89)
Observations	6,163	6,158	6,163	6,163	6,163	6,163
R-squared	0.25	0.23	0.24	0.23	0.23	0.24
Spatial & Year FE	Yes	Yes	Yes	Yes	Yes	Yes

This table shows panel estimation results for the U.S. counties for various distances. Each column is a separate regression where the impact of Spatial Population Concentration (SPC) on the change in personal income per capita for the period 1990-2015 is estimated. Table reports the coefficients of the variables in baseline estimation. All regressions include control variables presented in Table 7. *** p<0.01, ** p<0.05, * p<0.1

Table 7: List of Variables

U.S. County Data	
Variables	**Explanation-Source**
Spatial Population Concentration (SPC)	Authors' calculation based on Global Human Settlement (GHS) data
Income	Bureau of Economic Analysis
Population	Authors' calculation based on GHS data
Area	Authors' calculation based on U.S. county shapefile from U.S. Census Bureau
Rail	Total miles of railroad based on Bureau of Transportation Statistics
Highway	Total miles of highway, authors' calculation based on GRIP (Global Roads Inventory Project)
Distance to Airport	Distance to nearest airport in km from the population center of each county, Authors' calculation based on Global Airport Database
Number of Businesses Establishments	Bureau of Transportation Statistics
Cell Towers	3G and 4G cell phone towers per county; calculated by the authors' based on Opencellid.org
Broadband Coverage	Percentage of Internet use in U.S. counties based on Tolbert & Mossberger (2020)
Education	Percent of adults with a bachelor's degree or higher from USDA
Urban Population	Log total population of the county represented by urban population, authors' calculation based on U.S. Census Bureau Urban Area Shape files

Table 8: Spatial Independence: Moran's I

	log change in Income	log Income
	(1)	(2)
Moran's I	0.420***	0.654***
	(0.020)	(0.020)
Observations	3080	3,135

This table shows the Moran's I results for the spatial independence test of growth. The exponential spatial weight matrix is calculated based on the centroid of each county. Standard errors are shown in parentheses. Results stay robust when we calculate Moran's I using power function type spatial weight matrix.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9: Economic impact of SPC relative to other population measures

Measure	**Coefficient**	**Standard Deviation**	**Impact (%) 1990 - 2015**
SPC, 10 km	17.31***	0.0014	2.42
SPC, 15 km	23.46***	0.0014	3.28
SPC, 20 km	36.34***	0.0016	5.81
SPC, 25 km	50.93***	0.0015	7.63
SPC, 50 km	31.36***	0.0014	4.39
SPC, 75 km	15.99***	0.0013	2.07
SPC, 100 km	8.00	0.0013	1.04
SPC, 200 km	-4.63	0.0011	-0.50
Population	-0.21***	1.3738	- 28.84
Area	11.502 (a)	0.0009	1.39
Urban Population	-3.22***	0.0044	-1.41

This table shows the economic impact of the Spatial Population Concentration (SPC) coefficient on explaining the income growth considering alternative distances as well as alternative measures. The coefficients of population, area, and log population area the average of estimated coefficients in Tables 1 and 2. The Statistical significance of area changes across estimations. The impact is calculated by multiplying each coefficient with its standard deviation. (a) indicates the significant changes based on regression. *** $p<0.01$, ** $p<0.05$, *$p<0.1$

Table 10: Effect of Spatial Population Concentration on Income in the U.S.: Cross-Section (OLS)

Dependent variable:	log change in income per capita				
Periods:	1990-2015	1990-2015	2000-2015	1975-2015	1975-2000
log SPC: 10km	22.76***	13.89*	11.42	28.64***	10.10*
	(7.10)	(7.99)	(7.20)	(7.44)	(6.09)
log Δ Population	0.01	-0.14***	-0.15***	-0.26***	-0.18***
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
log Population	-0.04***	-0.21***	-0.18***	-0.34***	-0.18***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
log Income	-0.12***	-0.37***	-0.29***	-0.67***	-0.58***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
log Area	12.96**	9.19	16.99***	-9.71	-36.30***
	(6.33)	(6.07)	(5.63)	(6.72)	(5.77)
log Rail		0.74	2.53	-0.55	-2.52
		(2.21)	(1.94)	(2.41)	(1.90)
log Dist.Airport		1.83	2.78	4.81	2.89
		(3.22)	(2.88)	(3.59)	(3.09)
log N.Businesses		0.15***	0.11***	0.28***	0.20***
		(0.02)	(0.01)	(0.01)	(0.01)
log Highway		-0.82	-0.36	-0.37	-0.05
		(0.55)	(0.51)	(0.61)	(0.50)
log Cell Towers		0.02***	0.02***	0.03***	0.01**
		(0.01)	(0.00)	(0.01)	(0.01)
Broadband Coverage		0.13***	0.13***	0.19***	0.08***
		(0.03)	(0.03)	(0.03)	(0.02)
Education		4.19***	3.20***	4.70***	5.28***
		(0.69)	(0.62)	(1.20)	(1.06)
log Urban Population	-2.28**	-3.53***	-1.74*	-6.00***	-4.93***
	(1.02)	(0.95)	(0.91)	(1.05)	(0.83)
Observations	3,078	3,072	3,075	3,057	3,057
R-squared	0.32	0.43	0.52	0.54	0.65
State FE	Yes	Yes	Yes	Yes	Yes

This table shows the long-difference cross-section estimation results for the impact of Spatial Population Concentration (SPC) on the change in income per capita for various periods for the U.S. counties. Each column is a separate regression for different time period. Columns (1) is the baseline results. In columns (2) - (5) we add control variables. Robust standard errors are shown in parentheses. The details on control variables are presented in Table 7.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$