Geography of Growth and Development

Esteban Rossi-Hansberg

Subject: International Economics , Macroeconomics and Monetary Economics , Urban, Rural, and Regional Economics
Online Publication Date: Jul 2019 DOI: 10.1093/acrefore/9780190625979.013.273

Summary and Keywords

The geography of economic activity refers to the distribution of population, production, and consumption of goods and services in geographic space. The geography of growth and development refers to the local growth and decline of economic activity and the overall distribution of these local changes within and across countries. The pattern of growth in space can vary substantially across regions, countries, and industries. Ultimately, these patterns can help explain the role that spatial frictions (like transport and migration costs) can play in the overall development of the world economy.

The interaction of agglomeration and congestion forces determines the density of economic activity in particular locations. Agglomeration forces refer to forces that bring together agents and firms by conveying benefits from locating close to each other, or for locating in a particular area. Examples include local technology and institutions, natural resources and local amenities, infrastructure, as well as knowledge spillovers. Congestion forces refer to the disadvantages of locating close to each other. They include traffic, high land prices, as well as crime and other urban dis-amenities. The balance of these forces is mediated by the ability of individuals, firms, good and services, as well as ideas and technology, to move across space: namely, migration, relocation, transport, commuting and communication costs. These spatial frictions together with the varying strength of congestion and agglomeration forces determines the distribution of economic activity. Changes in these forces and frictions—some purposefully made by agents given the economic environment they face and some exogenous—determine the geography of growth and development.

The main evolution of the forces that influence the geography of growth and development have been changes in transport technology, the diffusion of general-purpose technologies, and the structural transformation of economies from agriculture, to manufacturing, to service-oriented economies. There are many challenges in modeling and quantifying these forces and their effects. Nevertheless, doing so is essential to evaluate the impact of a variety of phenomena, from climate change to the effects of globalization and advances in information technology.

Keywords: agglomeration, congestion, quantitative spatial economics, space, development

Introduction

The distribution of population in space is extremely uneven. In 2005, for example, about 70% of the world population was located in 10% of the available land.[1] Economic activity is even more concentrated than people. As Figure 1 indicates, in 2005 about 90% of gross activity was concentrated in 10% of the available land (using market exchange rates).[2] This concentration of economic activity implies that essentially all countries have empty areas and areas with a high density of people and economic activity. Of course, the areas of high density of population and those with high density of economic activity may not coincide. In fact, according to the same data for 2005, the correlation of population density and real income per capita across cells of 1º longitude by 1º latitude in the world is –0.41.[3] That is, throughout the world, densely populated areas tend to be poor while scarcely populated areas are richer. It turns out that most of this negative correlation comes from the correlation of cells across countries. The average correlation within countries is 0.17. Hence, within countries, dense places are rich, while across countries dense countries tend to be poor. The correlation between population density and real income per capita also tends to increase with the level of development: it is –0.11 in Africa but as high as 0.50 in North America and 0.33 in western Europe.[4]

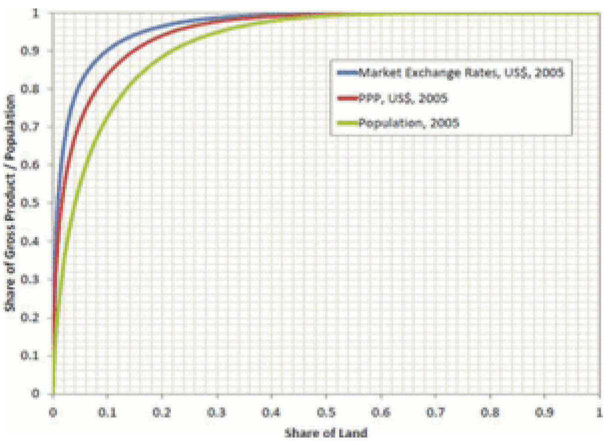

Figure 1. Distribution of local population and gross product (G-Econ 4.0, one-degree cells).

These differences are large, and they affect billions of people. How can they be explained? Has the field of economics come up with an explanation for these and other similar geographic patterns in the process of economic development? This article sets out to describe the state of knowledge on this topic.

A Basic View of the Geography of Development

To avoid suspense, this section starts by putting a tentative answer on the table. At the first stages of development, when all regions have low productivity, people live in areas that provide high living amenities, like good weather or a nice beach. As more people move to the area, the local factors (like land) become scarce, and amenities become con-

date: 24 March 2020

gested, which reduces the marginal value of labor, and therefore wages, and increases the cost of living in the area. These areas attract people until the utility they provide to the marginal agent that lives there is equalized with that of other regions (a spatial equilibrium condition that results from the ability to move across locations). Since the welfare level of individuals combines amenities and the utility from consuming goods, this happens at levels of real income per capita that are lower than in areas with worse amenities. The result is a negative correlation between population density and income per capita.

Now suppose that development happens through a process by which productivity increases in some areas and not in others. Again, these areas will attract individuals until the spatial equilibrium condition equalizes utilities in space for the marginal individuals. Now, however, most individuals will locate in areas that are productive but have to compensate workers to live in dense and not so attractive locations in terms of amenities. So people will tend to live in high real wage locations: a positive correlation between population density and income per capita. Simply put, the process of development makes people move from nice to productive locations, thereby increasing the correlation between population density and real income per capita.

Of course, this simple logic leaves open many questions. Perhaps the most relevant is: Why does productivity increase in some regions and not others? This is a crucial question. If the nicest locations simply become the most productive ones too, the correlation will not change or will change very little. If in contrast the areas that become the most productive ones are not the ones with the highest amenities, the correlation will increase rapidly with development. Reality is somewhere in between. Some cities like Rio de Janeiro are beautiful and fun, with great weather and a beautiful beach. However, they are not necessarily the ones that develop to be the most productive industrial centers. In the case of Brazil, that role is clearly taken by Sao Paolo, not Rio, although Rio is still one of the richest cities in Brazil.

The development of the economy of a region depends on the firms that decide to locate there, the infrastructure and institutions in the region, as well as its natural resources and geography. Well-connected regions have an advantage because firms and individuals located there can more easily trade goods and services with agents in other regions that specialize in a different set of products. Of course, the development of the economy of a region also depends on the size of its population and the characteristics of the individuals who decide to locate there. An obvious feedback is generated between the people that locate in a region, the firms that decide to locate and invest there, and the attractiveness of the region for individuals. Ultimately, local investment decisions by firms are determined by the equilibrium *market size* in the region where they decide to locate.

Firm Investment and Market Size

Why is market size so important? The reason is simply that it determines the returns to technological innovation (as well as other forms of local investments). The main characteristic of technology is that it can be used repeatedly. Its use does not deplete it. Economists refer to this property as the “replicability” or “non-rival feature” of technology. Because technology can be used multiple times, the returns to its invention is determined by the level of demand for the good or service it helps produce. This logic applies to all innovations, large and small, from inventing a new computer to improving the presentation of the menu in a restaurant. Of course, some technologies are more easily replicable in a given location than others. A production line in a factory is easily replicable, while restaurant decor might depend on its scale and so might be less scalable. The implication is that firms will innovate more in locations where in equilibrium they face a higher demand for their product or service. This demand depends on the number of customers that they have around them as well as the cost of reaching each of them (as determined by transport costs).

There is plenty of evidence that market size is important for innovation. Carlino, Chatterjee, and Hunt (2007) find that the elasticity in patents per capita increases with density in the United States. Combes, Duranton, Gobillon, Puga, and Roux (2012) find that the distribution of firm productivities is shifted uniformly to the right in larger cities, and Desmet and Rossi-Hansberg (2012) show that more productive cities pay entrepreneurs (who are likely responsible for most innovations) more. All this evidence is suggestive but not causal. Using exogenous variation in trade tariffs, De Loecker (2007) and Bustos (2011) show that firm productivity increases as a result of declines in trade costs. Bustos (2011) and Coelli, Moxnes, and Ulltveit-Moe (2016) actually find evidence that firms spend more in research and development as the result of the decline in trade costs. The latter paper estimates that trade liberalization accounted for up to 7% of knowledge creation during the 1990s. Finally, Sequeira, Nunn, and Qian (2017) find that locations in the United States that received more immigrants had sizable benefits on industrialization and innovation rates. All of these papers—and there are many more—document the effect of market size on local productivity and innovation.[5]

If market size determines innovation and the initial distribution of economic activity in underdeveloped economies is largely determined by local amenities, it seems that the logical conclusion is that places with good amenities should be those that become more productive over time. The reason that this logic, although valid to some extent, is ultimately incomplete is that market size is not only determined by local population but also by the ease of accessing customers in other regions (trade costs) and the ease of attracting new residents (migration costs). Furthermore, the success and profitability of innovations depends heavily on the initial level of technology. There is ample evidence that innovators stand on the shoulders of other innovations, particularly if they locate closer by (Jaffe, Trajtenberg, & Henderson, 2005). In addition, the profitability of an innovation that increases productivity proportionally depends on the initial importance of the idea. Hence, locations that start with good technology because of good institutions, or good natural resources, have an advantage relative to other perhaps initially denser regions with better living amenities. Ultimately, the empirical importance of these channels is an empirical question. The relative importance of transport costs, density, and local productivity in facilitating innovation is a question that requires more empirical research.

Geography and Market Size

Of particular importance in determining local investment is the heterogeneity of locations in terms of their geography. Geography here refers to the particular location of a region relative to other regions and their characteristics. It is not very useful to have a great port that can park large ships if the region is isolated from all other regions with good attributes for economic production. Geographic location is important because it determines the market size of firms in that location. That is, it determines the surrounding distribution of economic activity and the transport and trade costs associated with accessing that purchasing power. This notion is sometimes also referred to as “market access” (as in Donaldson & Hornbeck, 2016; Redding & Sturm, 2008).

Overall, “spatial frictions,” defined as the cost of moving factors and goods and services across space, are an important determinant of the market access of firms. They affect the relative importance for a firm’s demand—and therefore its scale and innovation decisions—of local demand versus demand in other close and far-away locations. If transport costs are high, only local demand matters, and so innovation will mostly happen in locations that have good amenities and good local characteristic for production (like natural resources). In contrast, if transport costs are low, innovation will depend more on the geography of a location, through the cost of reaching other consumers.[6] In fact, this is exactly what Henderson, Squires, Storeygard, and Weil (2018) find. They show, using satellite data of lights at night, that within the set of countries that developed early agricultural variables explain incrementally six times as much variation in nightlights as trade variables. In contrast, for countries that developed later, trade variables explain a much larger part of the variation in nightlights within the country. This is very much in line with the basic model outlined earlier, since countries that developed later face much lower physical and institutional trade barriers due to better transport technology and international institutions, like the World Trade Organization. Hence, for these countries, geography—as determined by trade variables—matters more.

Constant Returns in the Aggregate With Local Decreasing Returns

The feedback loop outlined in the previous section generates a dynamic agglomeration force. A larger population leads to larger market size, which incentivizes firms to innovate more and improve their technology, which in turn increases labor demand and local population size. There are also similar, but static, agglomeration forces where local productivity is a function of local population due, for example, to knowledge spillovers (see the survey of these static forces in Duranton & Puga, 2004; Rosenthal & Strange, 2004). Of course, in equilibrium, there must be strong enough congestion forces that counterbalance these mechanisms. Otherwise, all economic activity would ultimately locate in just one small region with productivity that tends to infinity. This is not a good description of reality. The distribution of economic activity in space has remained fairly stable over time, as Gabaix (1999), Eeckhout (2004), Soo (2005), and Rossi-Hansberg and Wright (2007), among many others, have emphasized. Thus, there does not seem to be a tendency for the largest cities to grow disproportionally relative to the medium and smaller cities.

This stable distribution of city sizes indicates that congestion forces necessarily overwhelm these static and dynamic agglomeration forces. Ultimately, the disadvantages of being too dense overwhelm the advantages. Market sizes might increase, but innovation does not increase further because attracting workers to take advantage of the better technology becomes too expensive. This limits the growth of the denser regions in the world. Eventually, the denser areas in the world do not grow in relative terms. That is, locations ultimately face decreasing returns to scale. The fact that the amount of land in a location is fixed eventually leads to enough congestion so that the marginal return to all other factors starts to decline.

Note that this logic applies to a particular location but not to large regions or countries or to the whole world. The reason, as first proposed in Rossi-Hansberg and Wright (2007), is that aggregate economies can grow by replicating these dense locations. For example, as long as there is enough land in the United States, the aggregate economy can keep growing by replicating the density of New York City in other locations. Of course, some of these locations might not have all the advantages of Manhattan, but eventually the local decreasing returns in Manhattan motivate development elsewhere, perhaps at the outskirts of New York City, or perhaps in a completely different area. The main argument is simple, aggregate growth can happen in the presence of local decreasing returns because, locally, land is fixed. In contrast, in the aggregate, land is abundant and cheap. As stated at the beginning of this article, all countries have land that is not used for economic purposes. The price of that marginal land is essentially zero. The result is a world in which the aggregate world economy, and particularly large countries like the United States, can face close to constant returns to scale and roughly constant long-term growth, even in the presence of local decreasing returns.

The Difficulties of Making the Basic Model Precise

Preliminaries: Static Quantitative Spatial Models

The last two decades have seen significant advances in spatial equilibrium models. Before this last wave of contributions, the literature was dominated by the approached summarized under the label “New Economic Geography.” This literature developed theoretical models of the location of economic activity based on the seminal framework in Krugman

date: 24 March 2020

(1991) that used the endogenous static agglomeration force generated by taste for variety and home-market effects. In these models, in the presence of transport costs, agglomerations happen because having many firms in the same location reduces the cost of accessing the varied basket of goods that agents desire. This reduces the price index in agglomerated locations, which attracts agents to that location. This elegant theoretical mechanism allowed the literature to break away from more ad hoc and exogenous formulations of agglomerations forces where local productivity was simply assumed to depend on density. Furthermore, it connected the strength of the agglomeration forces with the level of transport costs and, more generally, with geography. This represented a great step forward.

The cost of having this theoretically more satisfying framework was that its complexity required focusing on examples with just a few locations that exhibited limited amounts of heterogeneity. As a result, it was hard to link models to specific geographic contexts and data. The connection with the data was mostly made through reduced-form implications of this stylized theory.

In the past 15 years the field has developed a number of frameworks that can connect much more closely with the data. Redding and Rossi-Hansberg (2017) summarize this literature that they label "Quantitative Spatial Economics." The core setup uses a spatial equilibrium framework in which agents can move across locations, in some cases subject to mobility costs. Regions can trade between them, subject to costly transport costs. Regions differ in terms of their productivity, the amenities individuals enjoy while living in them, and their geography. Key to this setup is that it exploits a "gravity structure." Namely, trade flows, migration flows, and potentially commuting flows depend on their iceberg transport, migration, or commuting costs in a log linear fashion. Gravity is not only a good description of the data (as described in Head & Mayer, 2014, for trade flows, or Monte, Redding, & Rossi-Hansberg, 2018, for commuting flows), but it can be easily modeled using Ricardian frameworks where locations differ in terms of their productivity to manufacture differentiated goods (see Allen & Arkolakis, 2014; Eaton & Kortum, 2002). These gravity-based spatial Ricardian models are simple enough to be solved for thousands of locations, multiple industries and local factors (as in Caliendo, Parro, Rossi-Hansberg, & Sarte, 2018), and a variety of other forms of spatial and individual heterogeneity. This yields tractable models that are rich enough to incorporate real-world geography and that can be quantified using detailed local data. In particular, these models can be "inverted" to match the distribution of income and population in space, as well as trade flows between locations, exactly. This is done by finding the local productivities, amenities, and bilateral costs that match these data using an exactly identified model.

The resulting quantified model can then be used to perform a variety of static counterfactual exercises. For example, one can introduce changes in trade costs, local characteristics, local institutions or structures, and so on. Given that the models can be solved for many locations, the framework can be applied to detail regions in a city (as in Ahlfeldt, Redding, Sturm, & Wolf, 2015, for the case of Berlin) or to the whole world (as in Desmet, Nagy, & Rossi-Hansberg, 2018). The key underlying assumption is that the quantified local characteristics are invariant to the policy being analyzed.

One limitation of this approach is that it does not attempt to explain why agglomeration occurs. Agglomeration is simply the result of differences in local characteristics. Even when endogenous agglomeration effects are introduced, the models are quantified for parameter values where multiplicity of equilibria, and therefore the possibility of endogenous agglomeration, does not arise.

Introducing Dynamics

Another important limitation of this literature is that it works with static models in which the patterns in the spatial distribution of growth rates described earlier cannot be analyzed. The justification for this abstraction is simple: introducing dynamics into spatial frameworks is extremely hard. As Desmet and Rossi-Hansberg (2010) underscore, even setting up a dynamic-spatial model consistently is quite difficult. Boucekikine, Camacho, and Zou (2009) show that the basic investment problem with a forward-looking planner and continuous space is an ill-posed dynamic optimization problem where the initial value of the co-state does not determine the whole dynamic path. This essentially means that the current state variables do not determine the future evolution of the economy uniquely, which makes the computation and quantification of these models problematic if not impossible. The implication is that, to make progress, particularly in models that can be complex enough to relate to practical applications, spatial-dynamic problems have to be dramatically simplified.

To understand the type of simplifications that have been proposed in the literature, it is useful to first understand why this is such a complicated problem. Think of a firm that is deciding how much to invest in a given location, if at all. If the investment is durable, and to some degree irreversible, then the firm will need to forecast the return of such an investment contemporaneously and in the future in order to make this decision. In any given period, as argued earlier, the returns to an investment depend on the market size of the firm, which in turn depends on the whole distribution of economic activity in the world, as well as the related "spatial frictions." Thus, accurately evaluating today's returns involves developing a model of the economy that has as a state variable the whole distribution in space of all relevant local characteristics. Evaluating future returns requires the same information in the future, which implies that the firm needs a model of the evolution of theses spatial distributions as well. Hence, understanding the local investment behavior of firms involves a model with at least one distribution of spatial characteristics that evolves over time. This is analogous to the heterogeneous agent literature in macroeconomics where one needs the whole distribution of agent's characteristics as a state variable (e.g., its capital holdings). The main difference is that locations are ordered in space, and so firms care more about certain locations (those close by, for example) than others. Hence, the distribution and its evolution over time cannot be simply summarized using a few general statistics that matter for prices (see Krusell & Smith, 1998, for a prominent example of this methodology). Firms in alternative locations care about dif-

date: 24 March 2020

ferent moments of the distribution of the state variable. This natural characteristic of a spatial-dynamic economy makes the problem much harder and less amenable to existing computational techniques.

Solving dynamic problems with so many state variables can be quite complicated. Going to thousands of locations, as many applications require, is virtually impossible with today's numerical methods unless attention is restricted to very simple setups that can be linearized without much loss of accuracy.[7] What to do?

One option is to forget about general equilibrium linkages across locations and focus solely on the problem of an individual firm, as in the industrial organization literature. However, such an approach gives up completely on many of the questions of interest. The effect of changes in transport costs, trade tariffs, or migration restrictions are all the result of the type of reallocation decisions that result from general equilibrium linkages (see Dixit & Pindyck, 1994, and the many papers that followed).

A second option is to use perfect foresight and assume that firms and agents understand exactly the evolution of the economy moving forward. As long as this evolution is invariant to the counterfactuals to be analyzed, and the state variables influence decisions in a log-linear way, these future states can be differentiated out. This allows a full characterization of forward-looking transition dynamics in spatial models. Such techniques have been used, for example, in Dix-Carneiro (2014), Traiberman (2017), and Caliendo, Dvorkin, and Parro (IN PRESS), among others. For example, Caliendo et al. proposes a model to analyze the dynamic effect of import competition on local outcomes. The problem is inherently dynamic because agents face moving costs. Evaluating moving decisions then involves forecasting the benefits of being in a different location in the future and the potential paths and decisions triggered by being there. The simplified forward-looking problems of agents in different locations can be solved with these techniques, as population location decisions do not affect local productivities, amenities, or spatial frictions. To be clear, the framework can accommodate any change in these fundamentals, but they have to be exogenous to the equilibrium effects of the policies or events analyzed.

A third option is to limit the extent to which firms and individuals are "forward-looking." This can be done simply with a behavioral assumption about the myopic behavior of agents, or with a more careful design of an economy in which agents' returns do not depend on what happens in the future. In either case, although agents choose their actions only looking at the present characteristics of the economy, their actions can affect the future in complex ways. This is appealing because current economic circumstances motivate micro-founded changes in the economic environment in the future. The resulting frameworks can speak to a number of endogenous growth effects of a variety of changes in the economy, including changes in any spatial friction or geographic characteristic. Furthermore, these setups can make precise and quantify the logic of the basic model proposed at the start of this article. Of course, the drawback is that they omit any preemptive reaction to future policies or shocks.

Desmet and Rossi-Hansberg (2014) propose such a framework. Firms innovate in particular locations and their innovations lead to random technological advances that are spatially correlated in space. These local innovations can be interpreted as local worker expertise in an industry, local institutions, some local investments in structures and infrastructure, and the technology of firms for whom its costly to move. These advances are therefore "embedded" in the location of the firm and diffuse to other nearby locations. Perfect competition for land then implies that the value of innovations will be capitalized fully in land rents. Thus, firms obtain some returns from innovations in the period when they make them but not in the future. All future returns are capitalized in the value of land, so firms obtain no returns in the future. The result is a much-simplified firm investment problem where the dynamic, forward-looking investment problem of firms collapses to a static investment problem. Even though innovations do not provide gains in the future (firms benefit from the better technology but have to pay higher land rents), firms do invest because this allows them to win their bid for land in the current period. Hence, the economy invests, although at suboptimal rates. These investments determine productivity in the future and leads to growth in the economy. Desmet and Rossi-Hansberg (2014) show how such a model with two industries can generate industrial growth patterns that are consistent with the evidence of the structural transition of the U.S. economy in the later part of the 20th century between manufacturing and services. In particular, it adds a spatial dimension to the standard productivity-based narrative where local innovation in the service industry, caused by collocation of services in manufacturing centers, plays an important role.

The spatial setup in Desmet and Rossi-Hansberg (2014) is a continuum of locations in a line. This can be useful to talk abstractly about the role of space, as well as spatial collocation of industries, but it cannot be matched to actual geography in a map. Desmet et al. (2018) extends this dynamic investment model to a spatial model in the class of quantitative spatial models discussed in Redding and Rossi-Hansberg (2017). It proposes a model with dynamics and two-dimensional geography that includes migration and trade costs. This model can be used to understand quantitatively the evolution of a variety of spatial phenomena and their implied effect on growth rates. Desmet et al. show that the distribution of economic activity as quantified and then forecasted by the model eventually stabilizes and, absent any exogenous changes, grows at a constant rate. Namely, the model converges to a balanced growth path. In the balanced growth path, the distribution of economic activity is fixed. Naturally, the balanced-growth distribution of economic activity, as well as the long-run growth rate, depends on the initial characteristics of the economy. Therefore, changes in these characteristics will have permanent growth impacts. The model can be used to measure these effects, once it is properly quantified.

The Need for Data

Quantifying the family of models in the quantitative spatial economics literature requires detailed spatial data (as does discovering patterns in the spatial distribution of growth rates). For static analysis, at a minimum, it requires data on population counts at the lev-

date: 24 March 2020

el of spatial resolution of the analysis as well as measures of output (or factor prices), land, and structures. It also demands data either directly on transport costs or on trade flows. In addition, it involves estimating, or somehow determining, a number of model parameters (e.g., the elasticity of substitution in the consumption of different goods).

When dynamics is added, the data requirements naturally go up substantially. First, one needs to quantify bilateral migration costs. Second, one must determine the costs associated with investment or innovation decisions. This second set of parameters can be obtained, using the specification provided by the model, from the cross-sectional relationship between growth and density across regions. In Desmet et al. (2018), the model is quantified using a grid of 1º longitude by 1º latitude using Geographic Based Economic Data (G-Econ) data. Migration costs are assumed to consist only of an origin and a destination fixed effect, and no bilateral term. This implies that bilateral migration costs cannot depend, say, on distance. These restrictions, together with an assumption that migration costs are experienced as a flow cost on utility, imply that an agent's migration decision is also a static decision that is taken every period. They also imply that the model has predictions on population counts in a given period, but not on bilateral migration flows. Hence, data on population counts in space in two periods, together with the cross-sectional data described earlier, can be used to infer the migration costs that rationalize observed migration flows according to the model. Given initial conditions that make the model match cross-sectional data, the dynamics in the theory yield investment rates and, therefore, a new set of productivity levels in all locations. This then implies population counts next period given migration costs. One can adjust these migration costs until the model and data population counts in the second period match exactly.

The result of this procedure is a quantified model that performs fairly well. For example, Desmet et al. (2018) show that if time is revered and a calibration based on data from 2000 and 2005 is used, the correlation between data and model-implied country population levels in 1950 is 0.965—almost perfect. This is partly the result of population levels being very persistent. If changes in data and model-implied country population counts are correlated, the correlation is still an impressive 0.742. Note that the model only includes basic investment forces. It includes no policy shocks, natural disasters, geopolitical changes, or general-purpose technological innovations. Of course, the world experienced all of these types of shocks between 1950 and 2000.

The quantified version of the model can make precise the explanation of the data advanced at the beginning of this article. A core mechanism behind the evolution of the spatial distribution of output per capita and population is that the correlation between these two variables increases over time. The data puts it at around –0.4 in 2000. According to the model, that correlation will stay fairly constant for about 100 years than increase to –0.2 over the following century. In the balance growth path, the correlation becomes 0.6. Thus, in the balance growth path, dense places are rich places. This is the ultimate global "triumph of the city," paraphrasing Ed Glaeser.

Data Sources

The main source of spatially detailed data described here is G-Econ 4.0, compiled at Yale University by William Nordhaus and Xi Chen.[8] This data provides production data (as well as population and geographical data), at the 1º by 1º level, for virtually all cells with positive land mass in the world. This level of spatial resolution is somewhat coarse; it amounts to 100 km by 100 km in the equator. The advantage is that the coverage is global and the authors compiled for four cross-sections in 1990, 1995, 2000, and 2005. This permits comparisons over time, which is, of course, of great help when studying the spatial distribution of growth rates.

Of course, for studies that focus on a particular country, region, or city, such data might be too aggregate, and the authors might want to go to more specialized country-specific data sources. In the United States, for example, there is good census data at the county and zip-code levels and in some cases the census tract level. Population data can be easily found at a spatially disaggregated level for the whole world and for a number of periods. For example, LandScan has provided population counts at the 30-arc second level (about 1 square km at the equator) yearly since 1998.[9] Caution is recommended. Although all of these data make detailed use of official country sources, virtually all of it also utilizes approximations and extrapolations based on geographic and other local characteristics.

Obtaining global and spatially disaggregated economic data is much harder. For most countries output (or factor price), data is simply not available at levels of spatial detail below the 1º by 1º level. Even then, one can argue that for some countries the data is probably somewhat imprecise. In response to this challenge, researchers have turned to remote sensing data, in particular satellite data on luminosity of nightlights. This is data based on satellite images of the world surface.[10] Donaldson and Storeygard (2016) provide a nice review of this source of data and applications in economics. Henderson, Storeygard, and Weil (2012) argue that this data can be used to complement official growth statistics and that it is informative about the spatial distribution of output. Of course, the great advantage of this data is that it is available for many periods in time and it covers the whole earth at a fine level of spatial resolution. Some of the drawbacks are that images from different satellites are not directly comparable, the quality of satellite sensors deteriorates over time, and, perhaps most important, very luminous places at night (all major cities) are in most cases top-coded. Nevertheless, in this field local data is imperfect anyway, so this is certainly a useful source of information if used with sufficient caution.[11]

Why a Good Quantitative Dynamic-Spatial Model Is Necessary: Applications

Perhaps the best way to illustrate the usefulness of an empirically accurate dynamic spatial model is by highlighting the set of applications where such a model could be put to a good use. One reason for using spatial-dynamic general equilibrium models is that it is important to account for spatial interactions, both static and dynamic, across locations.

date: 24 March 2020

Depending on the spatial disaggregation, trade in goods and services, commuting, migration, and technology diffusion tightly link locations. As Monte et al. (2018) forcefully argue for the case of commuting, the effect of local productivity changes, as well as most other local shocks, depends importantly on these goods' and factors' spatial networks. Hence, local observations cannot simply be treated as independent observations in a reduced-form empirical analysis. Furthermore, because these links are endogenous and react to the shocks and policy changes being evaluated, any such reduced-form analysis is subject to the Lucas Critique (Lucas, 1976). Hence, in the list of topics where quantitative general equilibrium models can be useful, spatial economic applications rank very high.

The first obvious set of areas where the insights and quantification of a spatial-dynamic theory is needed is to understand the dynamic effect of changes in spatial frictions. Understanding the implications of changes in trade and migration policy for the world economy is essential to understand the role of globalization in economic development. Most static quantitative analysis of the effect of trade frictions conclude that the welfare gains from trade liberalization are modest. Costinot and Rodriguez-Clare (2014) estimate that a 40% worldwide tariff would reduce real income by less than 3.5%—not negligible, but also not a transforming policy for the world economy. This evaluation does not include any dynamic effects. Changes in trade policy will in general change market size and the incentives to invest, potentially leading to much larger gains from trade (as argued in Desmet, Nagy, & Rossi-Hansberg, 2017). Desmet et al (2018) also argue that these dynamic effects can lead to large effects from reductions in migration restrictions. As for infrastructure, Nagy (2018) contends that one of the important effects of the creation of the railroad network in the United States was the creation of cities that promoted innovation and technological growth. The paper argues that the creation of the railroad network was responsible for 27% of aggregate growth between 1790 and the Civil War. Delventhal (2017) also uses a spatial-dynamic general equilibrium model to argue that the spatial distribution of growth rates in the world can be largely rationalized using improvements in transportation technology over time. In related work, Trew (2017) proposes a quantification of the role of infrastructure investments on the takeoff in England and Wales between 1710 and 1881, associated with the Industrial Revolution. In particular, he argues that the time and location of the takeoff was determined by infrastructure investments and growth in the transport sector.

Spatial-dynamic frameworks can also be useful to understand the location of industries in space and how this distribution is related to technological innovation. Desmet and Rossi-Hansberg (2009) present evidence of a related empirical regularity. They show that, across U.S. counties between 1980 and 2000, growth rates in manufacturing employment declined with manufacturing employment in 1980. That is, locations that had many workers in the manufacturing sector saw lower growth rates during the next 20 years than locations that had little manufacturing employment in 1980. In sum, during this period manufacturing employment became more dispersed in space. In contrast, service sectors exhibit a very different pattern. While counties that had very little or a lot of service-sector employment also experienced increasing dispersion, counties with intermediate levels of employment exhibited a distinct pattern. Among these counties, the ones with higher service-sector employment in 1980 were the ones that grew faster. So intermediate-sized counties exhibited concentration.

Desmet and Rossi-Hansberg (2009) argue that these patterns caused by the diffusion of general-purpose technologies (GPT) in the sector. When a new GPT arises (like electricity or the Internet), the sector that uses it intensively agglomerates to use it and benefit from its diffusion. For the counties with the largest employment levels, this process is hindered by congestion, at the bottom, by the willingness of some firms to move to the cheapest locations. As the GPT becomes older and more dispersed, the importance of the diffusion force declines yielding dispersion throughout the distribution. In the 1980s the Internet had this effect on service sectors but not on manufacturing. Of course, if growth at the turn of the 20th century is considered, where electricity was a GPT for the manufacturing sector, a similar pattern as for services between 1980 and 2000 emerges. Perhaps surprisingly, this is exactly the case. The patterns for manufacturing at the turn of the century and services in the last two decades of the 20th century look virtually identical.

This type of industrial spatial collocation can also explain some of the patterns of relative sectoral growth that are referred to as the structural transformation. The first part of the structural transformation involved a large decline of the agricultural employment share as the manufacturing employment share grew substantially. In some developing countries, and in particular in China and other parts of Asia, some of this process of structural transformation is ongoing. General equilibrium models of spatial growth can help explain where and at what time such transitions can happen. Desmet and Rossi-Hansberg (2014) show how the collocation of service industries close to manufacturing clusters led to productivity growth in the service industry close to these manufacturing clusters, which started the structural transformation between manufacturing and services.

As a final application consider the economic impact of climate change. The environmental impact of climate change is a protracted, slow-moving phenomenon that has, and will have, heterogeneous impacts across locations, so modeling its effect with a dynamic spatial model is essential. Furthermore, because it affects many regions simultaneously, it will also have general equilibrium effects that can be sizable. Furthermore, since it affects regions differentially depending on their latitude, location relative to the coasts, and other geographic characteristics, it is likely to generate important shifts in economic activity in the world. Desmet and Rossi-Hansberg (2015) argue that global warming could shift economic activity in the northern hemisphere by about 10º north in the next 100 years. Importantly, because agents can adapt to new climates by trading, migrating, and other forms of economic adaptation, the cost of climate change will depend crucially on the magnitude of these spatial frictions. Absent any mobility costs, adapting by moving can be very effective, thereby leading to small costs. In contrast, if migration costs are large or prohibitive (as in the case of institutional restrictions), the resulting costs can be catastrophic. Put in context, the cost of climate change for the people in northern locations depends fundamentally on their ability to migrate to Europe and other countries that might be less affected by this phenomenon.

Future Directions and Challenges

Although the literature discussed here has made important contributions to the understanding of the geography of growth and development, this important topic remains underexplored. On the theoretical side, these frameworks are still missing fully forward-looking dynamics in contexts where agents naturally care about the impact of their actions in the future. Modeling forward-looking behavior in dynamic-spatial models with growth is essential when aiming to characterize optimal government policy, for example. Introducing irreversible capital investments would also require modeling forward-looking dynamics. Progress on this front is urgent.

On the more applied side, the use of satellite nightlights data together with population count data at a high level of spatial resolution over time can uncover many patterns that are still not well understood. As an example, consider Figures 2 and 3. These figures depict local patterns in the growth of nightlights and population density. Using a single satellite for consistency, the growth in nightlights for all 30-second squared pixel of the world surface between 2000 and 2007 was calculated, as was the growth in population density over this period. Locations where there is no data or where the change in both variables was zero (or close to zero using a 0.5% band) are shown in white. Locations where one of these variables did not change are depicted in yellow. In most cases, these are cells that have very little economic activity and population. The areas of large cities where nightlights are top-coded (like Paris in Figure 2) are depicted in maroon. The rest of the cells are depicted according to the pattern of the sign of the growth rates in these two measures. Red cells show growth in both population density and nightlights. These are areas of mixed growth where both residences and, probably, businesses in the service sector are concentrating. Blue areas are where population density is declining but lights are growing. These areas are most likely specializing in industrial production. Green areas are where population density is growing but lights are declining—areas that are probably specializing as residential neighborhoods. Finally, pink areas are where both measures are in decline. Of course, these interpretations are tentative, at best.

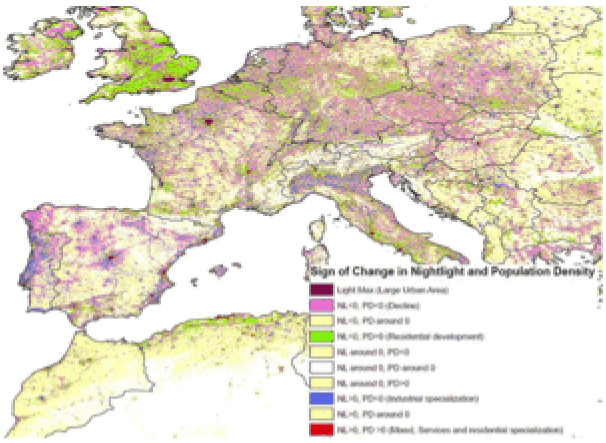

Figure 2. Nightlight and population density growth, Europe (LandScan and NOAA, 2000–2007).

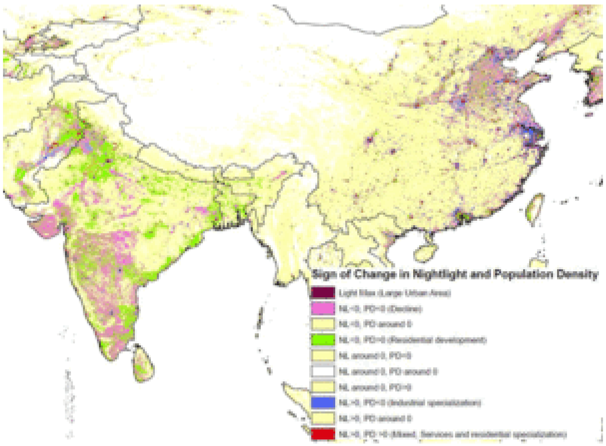

Figure 3. Nightlight and population density growth, Asia (LandScan and NOAA, 2000–2007).

Compare the difference in patterns between England, France and Spain in Figure 2. The southern part of England is (apart from London, which is all top-coded) painted in green and red dots. Namely, under this interpretation, the area is developing as a group of towns specialized in services and mixed land use surrounded by mostly residential areas where population is growing. France and Spain look very different. They have much more pink, indicating that population and economic activity are leaving these areas and concentrating in space. Most of the periphery cities are becoming less dense but brighter at night, indicating more industrial specialization. The differences in these patterns is striking in countries that are, roughly, at similar levels of development.

Figure 3 shows the same graph for China and India. Although both countries exhibit spatial concentration (evidenced in the ubiquity of the color pink), the difference in their spatial patterns is striking. China seems to exhibit industrial concentration in the north, together with residential development only around its largest cities (including Hong Kong). Periphery cities seem to be developing as both business and residential hubs (depicted in red). India, in contrast, exhibits growth in nightlights only in a few selected locations.

New data brings a variety of new questions. Making sense of these and many other patterns embedded in these and other spatially disaggregated data can hold the key to understanding why some economies grow rapidly and other stagnate.

Further Reading

Review Articles

Carlino, G., & Kerr, W. (2015). Agglomeration and innovation. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 349–404). Amsterdam, The Netherlands: Elsevier.

Davis, M., & Van Nieuwerburgh, S. (2015). Housing, finance, and the macroeconomy. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 753–811). Amsterdam, The Netherlands: Elsevier.

Desmet, K., & Henderson, V. (2015). The geography of development within countries. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 1457–1517). Amsterdam, The Netherlands: Elsevier.

Desmet, K., & Rossi-Hansberg, E. (2009). Spatial growth and industry age. *Journal of Economic Theory*, *144*(6), 2477–2502.

Desmet, K., & Rossi-Hansberg, E. (2010). On spatial dynamics. *Journal of Regional Science*, *50*(1), 43–63.

Lewis, E., & Peri, G. (2015). Immigration and the economy of cities and regions. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 625–685). Amsterdam, The Netherlands: Elsevier.

Redding, S., & Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, *9*, 21–59.

Other Central Papers

Ahlfeldt, G. M., Redding, S., Sturm, D. M., & Wolf, N. (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica*, *83*(6), 2127–2189.

Allen, T., & Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, *129*(3), 1085–1140.

Caliendo, L., Dvorkin, M., & Parro, F. (In press). Trade and labor market dynamics: General equilibrium analysis of the China trade shock. *Econometrica*.

Desmet, K., Nagy, D. K., & Rossi-Hansberg, E. (2018). The geography of development. *Journal of Political Economy*, *126*(3), 903–983.

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, *70*(5), 1741–1779.

References

Ahlfeldt, G. M., Redding, S., Sturm, D. M., & Wolf, N. (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica*, *83*(6), 2127–2189.

Allen, T., & Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, *129*(3), 1085–1140.

Anderson, J., Larch, M., & Yotov, Y. (2017). *Growth and trade with frictions: A structural estimation framework*. Working Paper. Drexel University, Philadelphia, PA.

Bleakley, H., & Lin, J. (2012). Portage and path dependence. *The Quarterly Journal of Economics*, *127*, 587–644.

Boucekkine, R., Camacho, C., & Zou, B. (2009). Bridging the gap between growth theory and the new economic geography: The Spatial Ramsey Model. *Macroeconomic Dynamics*, *13*, 20–45.

Burchfield, M., Overman, H. G., Puga, D., & Turner, M. A. (2006). Causes of sprawl: A portrait from space. *The Quarterly Journal of Economics*, *121*(2), 587–633.

Bustos, P. (2011). Trade liberalization, exports and technology upgrading: Evidence on the impact of MERCOSUR on Argentinean firms. *American Economic Review*, *101*(1), 304–340.

Caliendo, L., Dvorkin, M., & Parro, F. (In press). Trade and labor market dynamics: General equilibrium analysis of the China trade shock. *Econometrica*.

Caliendo, L., Parro, F., Rossi-Hansberg, E., & Sarte, P. D. (2018). The impact of regional and sectoral productivity changes on the U.S. economy. *The Review of Economic Studies*, *85*(4), 2042–2096.

Carlino, G., Chatterjee, S., & Hunt, R. (2007). Urban density and the rate of invention, *Journal of Urban Economics*, *61*(3), 389–419.

Carlino, G., & Kerr, W. (2015). Agglomeration and innovation. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 349–404). Amsterdam, The Netherlands: Elsevier.

Coelli, F., Moxnes, A., & Ulltveit-Moe, K. H. (2016). *Better, faster, stronger: Global innovation and trade liberalization*. NBER Working Paper 22647. Cambridge, MA: National Bureau of Economic Research.

Combes, P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, *80*(6), 2543–2594.

Costinot, A., & Rodriguez-Clare, A. (2014). Trade theory with numbers: Quantifying the consequences of globalization. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.), *Handbook of international economics* (Vol. 5, pp. 197–261). Amsterdam, The Netherlands: Elsevier.

Davis, D. R., & Weinstein, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. *American Economic Review*, *92*(5), 1269–1289.

Davis, M., & Van Nieuwerburgh, S. (2015). Housing, finance, and the macroeconomy. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of Regional and Urban Economics* (Vol. 5, pp. 625–685). Amsterdam, The Netherlands: Elsevier.

De Loecker, J. (2007). Do exports generate higher productivity? Evidence from Slovenia. *Journal of International Economics*, *7*(1), 69–98.

Delventhal, M. (2017). *The globe as a network: Geography and the origins of the world income distribution*. Working Paper. Universidad Autonoma de Barcelona, Barcelona, Spain.

Desmet, K., & Henderson, V. (2015). The geography of development within countries. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 1457–1517). Amsterdam, The Netherlands: Elsevier.

Desmet, K., Nagy, D. K., & Rossi-Hansberg, E. (2017). Asia's geographic development. *Asian Economic Review*, *34*(2), 1–25.

Desmet, K., Nagy, D. K., & Rossi-Hansberg, E. (2018). The geography of development. *Journal of Political Economy*, *126*(3), 903–983.

Desmet, K., & Rossi-Hansberg, E. (2010). On spatial dynamics. *Journal of Regional Science*, *50*(1), 43–63.

Desmet, K., & Rossi-Hansberg, E. (2012). Innovation in space. *American Economic Review*, *102*(3), 447–452.

Desmet, K., & Rossi-Hansberg, E. (2014). Spatial development. *American Economic Review*, *104*(4), 1211–1243.

Desmet, K., & Rossi-Hansberg, E. (2015). On the spatial economic impact of global warming. *Journal of Urban Economics*, *88*, 16–37.

Dix-Carneiro, R. (2014). Trade liberalization and labor market dynamics. *Econometrica*, *82*(3), 825–885.

Dixit, A., & Pindyck, R. (1994). *Investment under uncertainty*. Princeton, NJ: Princeton University Press.

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*, 899–934.

Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. *The Quarterly Journal of Economics*, *131*(2), 799–858.

Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, *30*(4), 171–198.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In J. V. Henderson & J. F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, pp. 2063–2117). Amsterdam, The Netherlands: Elsevier.

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, *70*(5), 1741–1779.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, *94*(5), 1429–1451.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *The Quarterly Journal of Economics*, *114*(3), 739–767.

Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.), *Handbook of international economics* (Vol. 4, pp. 131–195). Amsterdam, The Netherlands: Elsevier.

Henderson, V., Storeygard, A., & Weil, D. (2012). Measuring economic growth from outer space. *American Economic Review*, *102*(2), 994–1028.

Henderson, V., Squires, T., Storeygard, A., & Weil, D. (2018). The global spatial distribution of economic activity: Nature, history, and the role of trade. *The Quarterly Journal of Economics*, *133*(1), 357–406.

Jaffe, A., Trajtenberg, M., & Henderson, R. (2005). Geographic localization of knowledge spillovers as evidenced by patent citations. Ed. John Cantwell. Edward Elgar.

Krugman, P. R. (1991). Increasing returns and economic geography. *Journal of Political Economy*, *99*(3), 483–499.

Krusell, P., & Smith, A., Jr. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, *106*(5), 867–896.

Lee, S. & Lin, J. (2018). Natural amenities, neighbourhood dynamics, and persistence in the spatial distribution of income. *Review of Economic Studies*, *85*, 663–694.

Lewis, E., & Peri, G. (2015). Immigration and the economy of cities and regions. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (pp. 625–685). Amsterdam, The Netherlands: Elsevier.

Lucas, R. (1976). Econometric policy evaluation: A critique. In K. Brunner & A. Meltzer (Eds.), *The Phillips curve and labor markets. Carnegie-Rochester Conference Series on Public Policy* (pp. 19–46). New York, NY: Elsevier.

Marshall, A. (1920). *Principles of economics*. London, U.K.: Macmillan.

Monte, F., Redding, S., & Rossi-Hansberg, E. (2018). Commuting, migration and local employment elasticities. *American Economic Review*, *108*(12), 3855–3890.

Nagy, D. K. (2018). *City location and economic development*. CREI Working Paper. Barcelona, Spain: Centre de Recerca en Economia Internacional.

Ravikumar, B., Santacreu, A. M., & Sposi, M. (2017). *Capital accumulation and dynamic gains from trade*. St. Louis Federal Reserve Bank Working Paper. St. Louis, MO: Federal Reserve Bank of St. Louis.

Redding, S. J., & Sturm, D. M. (2008). The costs of remoteness: Evidence from German division and reunification. *American Economic Review*, *98*(5), 1766–1797.

Redding, S., & Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, *9*, 21–59.

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economics. In In J. V. Henderson & J. F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, 2119–2171). Amsterdam, The Netherlands: Elsevier.

Rossi-Hansberg, E., & Wright, M. (2007). Urban structure and growth. *Review of Economic Studies*, *74*(2), 597–624.

Sequeira, S., Nunn, N., & Qian, N. (2017). *Migrants and the making of America: The short and long-run effects of immigration during the age of mass migration*. Working Paper. Harvard University, Cambridge, MA.

Soo, K. (2005). Zipf's law for cities: A cross country investigation. *Regional Science and Urban Economics*, *35*(3), 239–263.

Traiberman, S. (2017). *Occupations and import competition*. Working Paper. New York University, New York, NY.

Trew, A. (2017). *Endogenous infrastructure development and spatial takeoff*. Working Paper. University of St. Andrews, St. Andrews, U.K.

Notes:

(1.) Using data of G-Econ 4.0 from Yale University.

(2.) The number is around 83% using purchasing power parity. Again, the data comes from G-Econ 4.0 from Yale University.

(3.) See Desmet, Nagy, and Rossi-Hansberg (2018) for details.

(4.) As shown by Desmet, Nagy, and Rossi-Hansberg (2017), this increasing correlation with the level of a region's income is also presence across counties and metropolitan statistical areas in the United States.

(5.) Lee and Lin (2018) show that natural amenities resulted in persistent higher income across neighborhoods in U.S. cities between 1880 and 2010. Bleakley and Lin (2012) also show the persistence of development to natural characteristics like the presence obstacles to river navigation. David and Weinstein (2002) also provide evidence of persistence on regional development.

(6.) Actually, the relationship should be U-shaped since geography is irrelevant for prohibitive or zero transport costs but positive for intermediate values. However, at empirically relevant levels of transport costs, reductions in trade costs probably lead to increases in the importance of geography (as evidenced by the large increases in world trade flows in the last several decades, for example).

(7.) See also Anderson, Larch, and Yotov (2017) and Ravikumar, Santacreu, and Sposi (2017) for examples of dynamic numerical approaches with capital accumulation in trade frameworks.

(8.) See G-Econ for full data description and downloads.

(9.) See LandScan for data description and downloads.

(10.) Data is provided by the National Center for Environmental Information; see Version 4 DMSP-OLS Nighttime Lights Time Series for data descriptions and downloads.

(11.) Other studies use more detailed satellite imagery to measure changes in developed land. See for example Burchfield, Overman, Puga, and Turner (2006), who measured the extent of U.S. urban sprawl between 1976 and 1992.

Esteban Rossi-Hansberg

Department of Economics and Woodrow Wilson School of Public and International Affairs, Princeton University