Journal of Urban Economics 144 (2024) 103711

Contents lists available at ScienceDirect

Journal of Urban Economics

journal homepage: www.elsevier.com/locate/jue

Urban and spatial economics after 50 years

J. Vernon Henderson [a,*], Jacques-François Thisse [b]

[a] London School of Economics, UK
[b] UC Louvain, Belgium

> Mills: "Urban economics is a diffuse subject, with more ambiguous boundaries than most specialties. The goal of this Journal is to increase rather than decrease that ambiguity. A wide range of economic phenomena are urban in character, and scholars with diverse backgrounds and interests can shed light on them. Isolation from the work of other economists is a danger to any substantive specialty, and to urban economics more than to others…." Journal of Urban Economics, 1974 Editorial, Vol 1, p iii.

After 50 years, what would we tell Ed Mills? Since the founding of the *Journal of Urban Economics* in 1974, urban economics has gained much higher visibility, as shown by the growing flow of publications in top-5 journals, the increase in quality of papers in top field and other mainstream journals, and the participation of hundreds of researchers in the Urban Economics Association meetings in the USA and Europe each year. Mills' desire for breadth has been met with urban papers on a long list of topics including social networks, scale economies, race, transport, housing, historical and natural amenities, land markets, local labor markets, retailing, urban growth and development, the environment, and so on. Spatial and urban economics are now strongly linked to other fields such as trade theory, local public finance, labor, development, and industrial economics. The remaining obstacle is training of PhD students in urban as a field, rather than being an area into which students wander, sometimes driven by the recent availability of spatially differentiated "big data." Few top departments offer PhD classes in the field, while in 1974 Princeton, Chicago, MIT and Harvard all offered intensive PhD training in urban economics.

In this piece we do not attempt to survey all that has been done and many of Mills' topics will be ignored for lack of space. Instead, we give our perspective on some key developments, retaining the classic division of *within* vs *across* city phenomena, although they overlap. In thinking about key developments, one of the main contributions over the last decades has been the clarification of the main concepts used in our field. Not only has this allowed a sound theory to emerge, but this is required for well-founded empirical work.

We give a rendering of *some* of the key highlights as we perceive them. The bibliography is long but intentionally limited, and many great papers are not even mentioned. There are many reviews and handbook chapters in the Handbook of Regional and Urban Economics series which do that. Nor do we attempt to cover all on-going developments: there are so many important new papers out there in different stages of refereeing.

Before delving into topics, we make a few observations. Some foundations, especially of "spatial economics" were set decades before 1974, culminating in Martin Beckmann's *Location Theory* (1968) and in Walter Isard's *Introduction to Regional Science* (1975). But the urban of urban economics really came into being in the USA in the 1960's and remained largely ignored in Europe where the emphasis was more on regional disparities (Puga, 2002). There was a committee of lead academics (Committee on Urban Public Economics) focused on American urban problems of the 1960's and intent on promoting the development of an understanding of urban phenomenon. The key intellectual development in 10 years before 1974 was the monocentric model of Alonso-Muth-Mills, often referred to as the AMM model, starting with Alonso's 1964 book *Location Theory and Land Use*. The motivation was to understand the physical structure of cities and varying intensity of building investments within cities, as well as modeling the idea that equilibrium in the land markets involved not a single market clearing price as in the paradigm of the day, but a continuum where prices in equilibrium varied multifold over space within a city. Also, already in the 1960's we see a tension in modeling: are we thinking of competitive equilibrium where the land rent is determined by an auction (Alonso, 1964) or strategic competition (Davis and Whinston, 1964).

A crucial observation is that some economists over the decades recognized that cities would not exist without some form of scale economies. In fact, Marshall (1890) set out descriptions for many of today's micro-foundations of agglomeration economies. Mills (1972) noted that with only constant returns to scale in production, a flat featureless plan, and no comparative advantage from human capital or resource endowments, everyone could live in individual homesteads producing at home all products they needed. There would be no trade and transport cost incursion nor agglomeration and associated commuting costs. Agglomeration demanded scale economies. Why was that observation controversial at the time?

In the 1960's and early 1970's, general equilibrium modeling in the Arrow-Debreu tradition focused on showing existence of competitive equilibria under the most general conditions for production and consumption. The history of the relationship between spatial economics and general equilibrium theory is both complex and obscure. It is complex because it is fraught with difficulties that have been put aside for

* Corresponding author.
E-mail address: j.v.henderson@lse.ac.uk (J.V. Henderson).

https://doi.org/10.1016/j.jue.2024.103711

Available online 5 November 2024
0094-1190/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

simplicity. It is obscure because the several attempts made over the last 50 years to clarify this relationship have befuddled the debate with confusing answers. Saying we needed models with scale economies and the resulting non-existence of competitive equilibrium in many regions of parameter space went against the tide and conventional wisdom of the day. Of course, the obvious point existed: without scale economies there would be no cities. But we live in a world of cities.

Housing accounts for the largest share of most consumers' budget, while transport expenditures often rank second or third. Yet, the index in MasColell et al. (1995), a major microeconomics textbook that contains 981 pages, does not include the words "land" and "transport". So, how can it be that the profession put aside the consumption and production of such goods? Many believed that assuming that a good at one location and the same good at another location are two different commodities was sufficient to account for space in general equilibrium theory (see, e.g., Debreu, 1959).

In thinking about how to incorporate transport costs and land in general competitive analysis, Starrett (1978) showed that the introduction of space in the Arrow-Debreu model of general equilibrium has a weird implication*: when individual production and utility functions do not depend on the actions taken by the others, regardless of the technology, total transportation costs in a spatial economy must be zero for a competitive equilibrium to exist*. That observation begged two questions. First, which market structure should we use to model competition across space? The answer has been multiple: oligopolistic competition, competitive developers who internalize the benefits of increasing returns, monopolistic competition, and perfect competition with externalities. Second, as observed by Koopmans as early as in 1957, increasing returns manifest themselves in a nonstandard way, as agents must set up "somewhere". This corresponds to a special form of *indivisibility*, as agents are not ubiquitous but have an "address" in space. For example, a household lives in a very small number of places, usually just one. The same for firms. That agents have an address in space is central to the concept of *spatial equilibrium* in which each agent aims to find its "best" location, which means it must evaluate where those it interacts with are located. For example, when firms exchange goods and those interactions are costly, a firm's profit and relative advantage depend not just on its own location, but the location choices of all other firms. Adding such a requirement violates the traditional competitive paradigm, where the only information agents need to make their choices is the price system given by the market. The concept of spatial equilibrium is, therefore, distinct from that of competitive equilibrium in the Arrow-Debreu model. Instead, it is endowed with a strong Nash equilibrium flavor. We stress that the concept of spatial equilibrium does not preclude assuming competitive markets. In this case, one must add "something" to the model. Ever since the launching of *JUE*, we have made tremendous progress about what these "somethings" can be.

In sum, how to model externalities, transport cost and imperfect competition is a challenge, helped by new analytical tools that became available over the years. This may be one reason why spatial economics remained peripheral in the profession. In struggling with these issues and modeling choices since 1974, urban economists have taken different avenues, although there is now some degree of convergence (Fujita and Thisse, 2013; Koster and Thisse, 2024). At the beginning, the field was fragmented despite valuable contributions made by top scholars such as Ohlin, Koopmans, Samuelson, Solow or Vickrey. These contributions remained unconnected, and were sometimes poorly acknowledged or understood, despite their high potential. As a result, visibility was hindered until the 1990's when there was publication of a number of papers with new perspectives, especially Krugman's (1991) paper. Ever since the pioneering work of Lösch (1940), there is now a plethora of models that confirm the relevance of the trade-off between different types of increasing returns and the cost of moving goods, people, and information to understand the organization of activities from the local to the global through cities and regions.

Finally, we note the spatial spread, or globalization of urban and spatial economics. As we noted above, urban economics was founded in the "urban problems" of the USA in the late 1950 and 1960's. Spatial economics existed and in Europe had a tradition through location theory stretching back decades. The *Journal of Urban Economics* had a distinct American flavor in the early years while its then equal in stature, *Regional Science and Urban Economics*, had a more spatial and regional economics flavor although the overlap was enormous, especially as urban economics spread to European researchers. The field was also well established in Japan because of the influence of Masa Fujita and his numerous students, as well as others such as Yoshi Kanemoto. Beyond that, Greg Ingram and Doug Keare and then Steve Mayo and Steve Malpezzi led efforts at the World Bank to show that developing country cities behaved as the urban models of the day predicted, with work on Brazil, Korea, and parts of Southeast Asia, involving Bank researchers like Lee and Renaud and later Deichmann and Lall and involving many non-Bank researchers. Work then spread to China, Vietnam, South Asia, Indonesia, various countries in sub-Saharan Africa, and the like.

We divide the rest of the paper into four parts: classic urban models and work on the *within* city allocation of resources, the *across* city allocation of resources, quantitative spatial models [QSM], and urbanization in developing countries and the evolution of urban work in addressing developing country issues. We trace some key developments and offer comments on directions.

1. Within city

We divide this section into several parts. First is the discussion of the monocentric city model, dynamics models, agglomeration economies, and polycentric city models. Then, there are the related aspects of the diseconomies and agglomeration economics of cities, with both leading into the second section looking at the across-city allocation of resources.

1.1. The monocentric city model and related topics

The initial foundation of within city work was the AMM monocentric model. As noted above, the original work was to offer theoretical underpinnings for gradients observed in the generally available data of the era: how population density, and housing and land prices all declined, even radically, as one moved away from the center of a city. The initial models were appropriately simple: no congestion, static equilibria, no roads, and identical consumers. Work over the next years focused on relaxing many of these assumptions, although not the one of static equilibria. Fujita (1989) and Duranton and Puga (2015) offer two very nice surveys of what has been accomplished with the monocentric city model.

Having said that, we find it interesting to recall what Solow (1973, p.2) wrote more than 50 years ago about the monocentric city model:

> "To study the locational equilibrium of a city seems almost silly. Buildings, streets, subways, are among the most durable objects we make, and it is very expensive to move them or even to remove them. Existing patterns of location must therefore have been determined in a large part by decisions that were made and events that happened under conditions that ruled long ago. It seems far-fetched to expect that what now exists will bear much relation to what would now be an equilibrium. Nevertheless, it turns out that the equilibrium states of simple models of urban location do actuality reproduce some of the important characteristics of real cities."

As an illustration of Solow's last point, using a dataset that contains gridded data on population densities, rents, housing sizes, and transportation in 192 cities across all continents, Liotta et al. (2022) find that

100% and 87% of the cities exhibit respectively the expected negative population density and rent gradients noted above. For example, on average moving 1 km away from the city center decreases population density by 8.5%.[1] Qualitatively the monocentric city model still makes highly relevant predictions, although it may lose relevance when we come to making detailed quantitative predictions.

What did emerge was a robust literature that relaxed the assumption of homogeneous consumers to have people with different incomes, different races with prejudices, different attitudes toward amenities and local public goods, two commuter families, and the like. But the focus was narrow: who would live nearer the city center given these differentiated consumers. For example, what would a prejudiced city look like in terms of segregation and land allocation to different groups. In the perspective of the time on prejudice, do Black families gain in terms of land allocations and prices if there is "white flight" in a monocentric model?

Moving off the base of the monocentric model, is the literature on sorting within cities per se at a more general level. This is a much broader look than trying to order people by distance to the city center. There is the vast Tiebout literature thinking about segmenting metropolitan areas into Tiebout communities in the context of the USA where larger metropolitan areas can have 100 or 200 local governments within their radius. Since the key public good up for provision is schooling, thinking about school districts and metro areas became one focus. One central strand of that literature is by Dennis Epple and co-authors like Romer, Calabrese, Romano and Sieg to name a few. These are attempts to bring in political economy, voting, and Tiebout sorting, with key innovations in thinking about how to model sorting across space in the face of income and taste differences, well before the development the QSM paper literature reviewed later.

These local public good models are set in contexts without, not just work commuting costs, but also costs of commuting to schools. Abbiasov et al. (2024) use GPS data for 400 urban areas in the United States and find that access to local urban amenities explains 84% of the variation in use of urban amenities across different urban areas. Distance is a major barrier to use of train stations, airports, harbors, schools, hospitals, parks, museums, and the like. Unfortunately, the public economics literature that accounts for both land and transport is still meager, with emerging exceptions like Gaigné et al. (2016), Loumeau (2023) and Pietrabissa (2023).[2]

A recurrent theme in the literature on heterogeneous households is that of spatial segregation and the importance of neighbors. In recent years, the work of Raj Chetty, with coauthors such as Friedman, Henden, and Katz to name a few, has highlighted the importance of the issue. The key is that the initial work in the USA on the moving to opportunity experiment of the 1970's, which incentivized a treatment group of lower income people to move to "better" neighborhoods in the work of that time found few effects on the kids involved in the early years. The shock was to see strong impacts in later years and has motivated a much bigger agenda on how the places people are born and grow up in affects later life outcomes (Chetty and Hendren, 2018). This work which has its controversies points and issues of potential of "social engineering" presents an agenda for urban economists in thinking about sorting and what place and neighborhood attributes matter and to what extent these attributes are observable and measurable.

Other developments include commuting modal choice (McFadden, 1974), with a body of empirical work that continues. Second is a large body of work on congestion pricing, rush hour modelling of decisions of when to commute and staggering of work start hours. Early papers on congestion pricing include Vickrey (1969) and Arnott et al. (1993), while the impact of congestion pricing on cities is now a major topic in transport economics (Small and Verhoef, 2007). That latter morphs today into decisions about how much to work at home (Henderson, 1981; Takayama, 2015).

Despite major contributions, a few important issues remained almost untouched. Emerging from the AMM model was a substantial optimal control literature introducing roads and congestion, with highlighted pieces by Mills and de Ferranti (1971), Solow and Vickrey (1971), and Riley (1974). Congestion modeling was simple: travel speeds declined as one moved towards the city center and the accumulations of people commuting inwards mechanically rose. With that and in the face of rising land prices as one moved towards the city center, what was the optimal allocation of land to roads? Given about 30% of land in cities was devoted to roads in American cities at the time (American Planning Association, 1950), this seemed a crucial question. It also made the obvious point that public land use planning and private markets in cities interacted. This literature fell into oblivion, perhaps because of the tools and complexity of the problem. We note there is a recent paper trying to empirically disentangle congestion from road capacity effects on travel time, pointing out the key role of road quality in determining how quickly people can move through a city (Akbar et al., 2023).

Also falling somewhat into oblivion was the consideration of modelling that accounted for the role of the public sector in allocating 40–50% of the land in a city to public use and in planning the allocation of the other 50–60%, in a context that public land allocations in particular roads drive private demand. Economists recognize that regulation affects private land use in a city; they have many papers cherry-picking examples of poorly designed regulations. There is now a more rigorous literature trying to look at the effect of regulations on urban development: Saiz (2010) on how geography and regulation affect housing supply, Turner et al. (2014) on theory and econometrics of identifying causal impacts of zoning and other policies based on crossing of local jurisdictional boundaries, Duranton and Puga (2023) on restrictive development in high amenity cities, Baum-Snow and Han (2024) on local supply elasticities of housing. Still, what remains unexplored in rigorous modelling is the almost universal role of planning in trying to improve land allocations overall in a city in the face of roads, commuting patterns, externalities among competing types of land use, and political economy.[3]

1.2. Dynamics in the monocentric model

The basic monocentric model is static, or one of long run equilibria. Capital is like Lego-cities—perfectly malleable and moveable. When demand to live in a city increases or decreases, with the shift up or down in the rent gradient, capital intensity changes overnight, with the addition or subtraction of infinitely lived Legos in or out of the city from national capital markets. Since buildings have lives of potentially dozens if not hundreds of years, static models offer a limited perspective when we are thinking of policy and horizons of a generation or two. This is particularly the case in which urbanization in a country is proceeding and cities are growing at a rapid pace. The challenge is that durable capital stock requires irreversible investment decisions in a dynamic model and combining the two dimensions of space and time to investigate the issues is technically challenging. Fujita (1982) and Wheaton (1982) recognized the issue and specified perfect foresight (and myopic)

[1] Working at a much finer geographical scale that covers 85% of the U.S. population and 83% of all single-family homes, Davis et al. (2021) confirm the robustness of the predictions on population density and rents.

[2] On the other hand, the geographical accessibility to public services is at the heart of a rich literature in operations research, which studies the optimal locations of public facilities when users' locations are exogenous (Farahani et al., 2019; Marianov and Eiselt, 2024). There is also an interesting body of literature that focusses on the case of nimby, also called obnoxious, facilities (Church and Drezner, 2022). To the best of our knowledge, no attempt has been made to connect these two strands of literature.

[3] Glaeser, Gyourko, and Saks 2005 outline one political economy process in a non-spatial model.

monocentric models and discussed some key issues. Braid (2001) developed some insights into the process and basics of evolution of the urban capital stock and life of buildings, under perfect foresight.

Henderson et al. (2021) took the Braid specification, added in a second technology for slum versus formal sector housing, and added in costs of converting land to formal sector use in a developing country context. Land can be converted from informal use, requiring sometimes very costly formalization of property rights, or costs can be incurred to overcome poor geography (slope, marsh). Formalization reduces risk of expropriation, where intense investment incentives expropriation. They develop theorems on formal vs. slum development and how long formal sector buildings last before redevelopment, both as a function of time and distance to the city center. They study how investment intensity changes over time and space: how much height changes with tearing down and redevelopment of formal sector building. They then take the model to the data on Nairobi observed with aerial photo and Lidar data for two time periods, estimate all the parameters of the dynamic structural model, and conduct counterfactuals, focused on formalization of slum lands. At the margin formalizing valuable slum lands near the city center would generate gains amounting to about $18,000 per slum household, 30 times typical annual slum rent payments.

Hopefully such work is just at its start. We know little about dynamics in polycentric cities (discussed below) and little about dynamics under imperfect foresight other than the notion that under uncertainty optimal "stopping times" tend to delay irreversible redevelopment (Capozza and Helsley, 1990). There is literature on building heights. For example, Ahlfeldt et al. (2023) use a static QSM to look at how changes in height technology allow for more compact cities. There is a notion in the literature of "creative destruction", where, say, a hurricane destroys capital stock and allows for early renewal and intensified investment. Of course, creation is only to force earlier rebuilding that would occur otherwise.

1.3. The origin and nature of agglomeration economies

1.3.1. The producer city

The AMM model remains silent about why a central business district where people work exists. People were content in saying that an employment center exists because firms are more efficient when they get together. That is, even if individual firms have constant returns there must be increasing returns [IRS] in the aggregate, but the notion was fuzzy in the early literature. Our starting point is Marshall (1890). There are two types of IRS. First, are IRS which are *internal* to firms. Second, are IRS *external* to a firm, where the decisions of other firms (e.g. their location choices) can make the firm more productive. The difference between the two types of IRS is important: internal IRS are generally inconsistent with perfect competition, whereas external returns are consistent because the externalities are treated as exogenous by individual firms (Chipman, 1970).

Internal and external returns may generate the same reduced forms, but they need not. Hoover (1936) went one step further by separating localization (or within-sector) economies and urbanization (or across-sectors) economies, a subject of considerable empirical work over the years. One of the most robust empirical facts about cities is that on average workers earn more in large cities than in small ones, a fact that has been coined the `urban wage premium'. For a long time, agglomeration economies were used as a black box with poor microeconomic foundations. There is now a wide range of "candidate" agglomeration economies that could explain why a higher density of workers raises productivity. Duranton and Puga (2004) proposed organizing this cornucopia into the following trilogy: sharing, matching, and information spillovers/learning.

(i) *Sharing* can refer to input-output linkages, but also to sharing of public goods. On linkages, minimizing transportation cost of physical goods between manufacturing plants was one of the main reasons for the existence of cities during the Industrial Revolution and is a motivation for the formulation of certain new economic geography models (Fujita et al., 1999). Physical transport costs have plummeted, and production of components has globalized. Now linkages may be more relevant in the service sector where firms in large cities share a wide range of business services (IT, advertising, legal, and accounting services) supplied by knowledge-intensive firms, where 'transport costs' (face-to--face communication) are high. Related to sharing are scale economies from greater variety of inputs in cities, formulated by Abdel-Rahman and Fujita (1990) in a model with internal IRS for the intermediates and love of variety by final good producers. On local public goods, there is generally a critical mass below which goods are not supplied and higher density can lead to greater efficiency in provision.

(ii) *Matching* means that the quality of matches between firms and production factors is higher in a thick market than in a thin one because of the greater number of opportunities for agents when they operate in a denser market (Helsley and Strange, 1990). Also, business-to-business service providers have access to a bigger pool of potential customers in a large city. Matching models can be combined with search theory where a large labor market raises the job-seeker's chances of finding a better match and may lead to lower local unemployment rates in bigger cities (Zhang, 2007).

(iii) *Information spillovers* are at the heart of the first microfoundations model in Ogawa and Fujita (1980), discussed later. Information spillovers are hard/impossible to measure, but there are many papers by geographers describing how salespeople and customers spread information of what inputs to use, what competitors are doing, and whom to supply and buy from, improving firm static efficiency. While some might think that IT has reduced the need for face-to-face meetings and density, Gaspar and Glaeser (1998) argue that IT and face-to-face are complements. What has some measure is patents, reflecting knowledge spillovers and learning, where research and development often demand long periods of exchange and discussion, and repeated trial and error. Here the difference between *tacit* and *codified* information or knowledge is noted. Tacit involves the exchange of information which cannot be or is not written down. But even utilizing patents involves spatial proximity of adopters to inventors, to communicate associated tacit information (Henderson et al., 1993; Peri, 2005).

1.3.2. The consumer city

Cities are not just efficient production centers but are also great consumption, culture, and leisure places (Glaeser et al., 2001). Starting with Hotelling (1929), there is a rich literature focusses on retailing in cities. Hotelling's main result in a simple context with fixed prices was neat: two retailers selling the same good and competing to attract customers who are spatially distributed along Main Street choose to locate at the street center. This result has been applied in a variety of circumstances including modeling political parties' tendency to move to the center. In general, however, things are not so simple. If firms are located together and free to choose their prices under elastic demands, any retailer has an incentive to undercut its rival because doing so allows it to capture the whole market. By moving away from the center, any firm can separate from its competitor to get a positive markup. Price competition works as a strong dispersion force (d'Aspremont et al., 1979).

The consumer model that lies behind this argument is that of a "robot" which reacts to any infinitesimal change in price or location. McFadden (1986) pleads for a richer model of individual behavior in which consumers are influenced by different factors when they make their choices. In sum, consumers have idiosyncratic tastes about firms and retailers may therefore view these choices as probabilistic. If firms

understand that consumers maximize stochastic utility, consumer behavior can be described by a discrete choice model. In this case, each retailer has a positive probability of being chosen by each consumer. This probability is directly related to the firm's mill price and to its distance from consumers, and inversely related to the prices and locations of competing retailers.

Using the multinomial logit, de Palma et al. (1985) show that $n \geq 2$ firms choose to agglomerate and to price above marginal cost when they sell a sufficiently differentiated product, travel costs are low enough, or both. The result is intuitive. When travelling is cheaper, prices are lower because competition is tougher. Firms, e.g., stores or restaurants, then choose to reconstruct their profit margins by differentiating their products in terms of non-geographical characteristics. As a result, the pro-competitive effects of agglomeration are offset by differentiation in non-geographical product characteristics, with firms striving to be as close as possible to the consumers with whom their matching is best. Since these consumers are spread all over the city, firms minimize their spatial differentiation by forming a shopping street at the city center. Here, agglomeration is caused by the locational interdependence among similar firms that compete for the best locations, and not by the agglomeration economies discussed below.

As examples, Leonardi and Moretti (2023) study the impact of the deregulation of the locations of restaurants in Milan (Italy) after 2005. Their main finding is that restaurants are now geographically more concentrated but more differentiated along other attributes. Furthermore, Koster et al. (2019) find, for Dutch shopping streets, an elasticity of rents with respect to the number of stores of at least 0.25, which is considerably higher than the agglomeration elasticities found on the production side (see below).

1.4. The magnitude of agglomeration economies

Starting with the influential work of Sveikauskas (1975) and Ciccone and Hall (1996), research on city size, employment density, and productivity has progressed enormously during the last two decades, as reviewed in Rosenthal and Strange (2004), Combes and Gobillon (2015) and Grove et al. (2023). Conducting a meta-analysis of the elasticity of different magnitudes with respect to population density, Ahlfeldt and Pietrostefani (2019) find a wage elasticity equal to 0.04, but this number hides a great heterogeneity across cities. Moreover, results in the literature differ by methodology and then by context: different industries, different measures of outcomes such as wages vs. firm or industry productivity, different measures of scale, and different spatial scales of analysis. If one asks urban economists what the degree of agglomeration economies is, one will get a big range of answers, as is appropriate for a result which is so context specific.

Some of the earlier literature looked at the issue of localization versus urbanization economies (Henderson, 1986), finding greater evidence of localization economies. Much of this literature, even today, looks just at manufacturing because of greater availability of economic data in censuses of manufacturing compared to service census data. Starting about 20 years ago, there was a focus on topics like identification of causal effects (Greenstone et al., 2010; de la Rocha and Puga, 2017), formal evidence of clustering of like producers and in different industries (Duranton and Overman, 2005), and the extent to which scale economies are highly localized and extend beyond manufacturing (Arzaghi and Henderson, 2008; Rosenthal and Strange, 2020).

Some recent work has focused on disentangling the microfoundations of agglomeration economies with the issue of so-called `Marshallian equivalence': all the three determinants of agglomeration may lead to similar reduced forms in empirical applications (Koster and Thisse, 2024). This makes it especially hard to measure the contribution of each one and their relative importance is likely to vary with the city size and sectoral specialization (Puga, 2010).

Ellison et al. (2010) provide evidence that Marshallian agglomeration forces are important in explaining the spatial colocation of industries. Examining colocation is a strategy to deal with aspects of causality. All the three forces — input-output sharing, labor pooling, and technological spillovers — together with natural advantages, seem to matter. A recent series of papers have extended and updated these results. Using data from England, Faggio et al. (2017, 2020) find that labor pooling was the most significant factor determining coagglomeration, followed by input-output sharing and technology spillovers, with all three sources being more critical in determining coagglomeration for newer industries. Other work notes that, not only do Marshall's agglomeration forces differ across industries, but they may also change over time (Diodato et al., 2018; Steijn et al., 2022).[4]

We note three challenges this literature still faces. First is identification of causal effects. This is research on which we can't conduct experiments and there are few convincing natural experiments. There is suspicion of panel and fixed effects work, as accentuating the impacts of measurement error and not dealing with shocks correlated with changes in outcome and scale measures. Finally, there is a suspicion that much IV work either fails the exclusion restriction or has weak instruments.

Second is the white elephant in the room: what many think as the prime source of agglomeration economies, information spillovers, is not measured in disentanglement studies (set apart from looking at patenting, a narrow item). Finally, is an omitted variable in many studies: human capital externalities. In panel data typically changes in scale and changes in human capital are strongly correlated: higher skill workers are more mobile and migrate to places and industries that boom. Moretti (2004) makes a strong case for the existence of human capital externalities especially for higher tech industries, but he omits scale changes in his study just as agglomeration studies omit human capital changes. Ciccone and Peri (2006) make a start at trying to disentangle the two, but it is a challenge yet to be really met.

1.5. When are cities monocentric or polycentric?

Cities may be viewed, at least in the first order, as the outcome of the interplay between social interactions and competition for land. Isolation allows an individual to consume more land but makes interactions with others more costly. To study this trade-off, Beckmann (1976) assumes that the utility of an individual depends on the average distance to all individuals and on the amount of land bought on the market. In equilibrium, the city exhibits a bell-shaped population density distribution supported by a similarly shaped land rent curve. In other words, the gregariousness of human beings turns out to be a sufficient motivation for them to gather within *a monocentric and compact area*.

To the best of our knowledge, the first location model that determines different city structures according to parameter values was developed by Ogawa and Fujita (1980) in a fundamental paper that went unnoticed for a long period of time.[5] These authors combine consumers and firms in a full-fledged general equilibrium model in which goods, labor, and land markets are perfectly competitive. Informational spillovers are described by a linear field and act as an agglomeration force because informational spillovers are subject to distance-decay effects. The clustering of firms increases the average commuting distance for workers, which in turn leads workers to pay a higher land rent.

[4] In a more recent study, Steijn et al. (2022) examine U.S. manufacturing data at the metropolitan statistical area level since 1970. They find that labor market pooling and input-output linkages have become less important, while technological spillovers have increased considerably since the 1990s. These changes associate with the rise in import penetration and the decrease in the share of routine employment. In line with Faggio et al. (2017), Steijn *et al.* also find that industries with a higher share of routine employment tend to rely more heavily on labor market pooling.

[5] Only a limited number of papers have tackled the endogenous formation of employment centers. They are surveyed in Duranton and Puga (2004) and Behrens and Robert-Nicoud (2015).

Therefore, firms must pay workers a higher wage as compensation for their longer commutes to work. In other words, the dispersion force stems from the interaction between the land and labor markets in firms' optimization program. The equilibrium distribution of firms and workers is the balance between those opposing forces. Unlike the AMM model in which the CBD is given, interactions among agents make the relative advantage of a given location for an agent dependent on the locations chosen by the other agents. In other words, there are externalities.

Ogawa and Fujita show that, in equilibrium, *the city may display different configurations*. First, when commuting costs are high in relation to the distance-decay effect, the equilibrium involves a full integration of business and residential activities and land use is unspecialized. As commuting costs fall, two employment centers, themselves flanked by a residential area, are formed around an integrated section. Eventually, when commuting costs are low enough, the city becomes monocentric. In this configuration, land use is fully specialized.[6] This directional move seems to concur with the evolution in the spatial organization of cities that has been observed since the beginning of the revolution in intra-city transportation, as discussed below in a QSM modeling London development over 80 years.

In thinking about generalizations, firms or developers may choose to form secondary employment centers, enterprise zones, or edge cities. In this way, firms can pay lower wages and land rents while retaining most of the benefits generated by large urban agglomerations (Henderson and Mitra, 1996). Many firms (e.g., banks or insurance companies, but also industrial firms a long time ago) have moved part of their activities (such as book-keeping, planning, and employee training) to the suburbs. In this case, a firm typically conducts some of its activities (such as communications with other firms) at the front-office located in the CBD while the rest of its activities are carried out at the back-office set up in the suburbs.

With lower intrafirm communication costs, we can have agglomeration of the front-units at the city center, surrounded by a residential area, while back-units are established at the outskirts of the city together with their employees. In other words, *the advancement of communication technologies provides for the emergence of cities having many employment centers* (Ota and Fujita, 1993; Glaeser and Kahn, 2004; Cavailhès et al., 2007; Rossi-Hansberg et al. 2009).

2. The allocation of resources across cities and/or regions

2.1. Hierarchy models

Despite widespread empirical evidence, one of the most enduring problems in spatial economics is to justify the hierarchy that prevails within systems of cities. Just before 1974 and the first issue of the *JUE*, the main approach was central place theory [CPT]. Consider n goods $i = 1 \ldots n$ and let r_{i} denote the radius of the range served by a producer of good i. Lower-order central places have small market areas and provide basic goods that are purchased more frequently than high-order goods. High-order places, which supply high-quality and expensive goods, are fewer than low-order places and supply a larger number of customers. Christaller and Lösch argued that the different systems of market areas can be arranged in such a way that a location where good i is provided also accommodates firms supplying all the goods $j = 1 \ldots i - 1$, but not vice versa. The baseline model of CPT suffers from two main shortcomings: it is not a general equilibrium model and there are no economic forces that ensure firms producing different goods will cluster.

The modern economic literature devoted to CPT is meager. Two valuable contributions are worth mentioning. First, emerging from the new economic geography literature reviewed below, Fujita et al. (1999a) proposed a general equilibrium model in which dimensionless cities produce differentiated manufactured goods which are shipped to dispersed farmers producing the agricultural good. These authors show that a regular hierarchical central place system emerges as an equilibrium outcome as the population size rises. Higher-order cities provide wider ranges of manufactured goods and export more varieties than lower-order cities. Unfortunately, the model must be solved numerically. Second, Hsu (2014) proposes a spatial competition model in which goods are differentiated by the value of their fixed production costs, so that firm's market areas are determined by the free-entry condition. Hsu shows that large and small central places formed by nesting market areas is an equilibrium. The distance between cities sharing the same order in the urban hierarchy is inversely related to their size. Unfortunately, demands are perfectly inelastic, so that the model fails to capture individual consumption choices. We may thus conclude that there exists no full-fledged general equilibrium model that encompasses what some consider key aspects of urban systems, including costs of trade across an endogenous number of cities.

2.2. Systems of cities

Henderson (1974, 1988) designed the first general equilibrium model of the allocation of resources across cities. Henderson posited a world with cities where people clustered due to external economies of scale, but are subject to diseconomies, for example in the form of commuting. There is an unexhausted set of identical urban sites on which cities formed. The paper dealt with two cases, one with "large agents" or developers who each own a potential urban site and could act in competition to achieve efficient city sizes and one where there are no large agents and multiple equilibria under what is now called "self-organization". What makes Henderson's approach unique are these large agents—land developers, kings or rulers, factory towns, or local governments—which act as coordinating agents in the development of cities. Note that the term `large agent' is somewhat misleading because these agents do not behave strategically. They are large in the following sense: they understand how competitive markets interact with agglomeration (dis)economies and internalize these effects, but they do so at the city level only. Apart from that, they treat the rest of the economy as a given.

Under the regions of parameter space where utility of the representative worker is a bell-shaped ("inverted U-shaped") curve and stability conditions in (costless) worker movements across cities are met, Henderson solved for a general equilibrium with different types of cities of different sizes that specialized in different goods (each subject to localization economies of potentially different degrees) that are costlessly traded. Homogeneous workers are paid different nominal wages in different size cities but equal real wages. Empirical work decades later traced out bell-shaped real income curves for different types of cities (Au and Henderson, 2006).[7]

2.2.1. Some subsequent developments

In the years that followed there were some key developments of this

[6] Fujita and Ogawa (1982) extended their model to an exponential decay function. When the distance decay effect is very small, the spatially discounted accessibility can be well approximated by the linear one, so that the three possible configurations described above still prevail. However, things change dramatically when the decay parameter takes intermediate values. The following may happen: (i) there exist equilibria with two or more employment centers; (ii) there is often multiplicity of equilibria; and (iii) the transition from an equilibrium to another may be catastrophic. Lucas and Rossi-Hansberg (2002) allow for substitution between land and the numéraire in a two-dimensional space. In both papers, the authors appeal to numerical analysis to characterize the spatial equilibria (see Dong and Ross, 2015, for corrections).

[7] Migration restrictions in China pre-2000 allowed the tracing, given in general one might not expect to find observations to the left of the maximum point.

model with many papers. Early on was to introduce pure public goods and the Henry George theorem. Flatters et al. (1974) formulated this in a regional model but allowed "optimal" region size. Arnott and Stiglitz (1979) did this for a system of cities and Henderson (1988) more generally. At efficient city/region size, the efficient level of public goods can be financed out of just land rents. Then, various papers brought more precision to the role of large agents. For example, Helsley and Strange have a set of papers exploring the role of developers, city governments, commitment and the micro-foundations of agglomeration economies (Helsley and Strange, 1990, 1994, 2014). One focus in thinking about developers and governments is whether equilibrium cities were too big or too small, an on-going topic (e.g., Albouy et al., 2019).

Other work tried to explore heterogeneity in city production patterns beyond Henderson's simple specification of specialization. Duranton and Puga (2001) specified a model mimicking aspects of the product cycle hypothesis, explaining the greater diversity of larger (nursery) cities and the movement of innovated techniques to specialized smaller cities. Duranton (2007) looked at product innovation in cities, to explain, as seen in the data, the churning of cites in terms of production patterns in the face of stability in their sizes. Duranton et al. (2014) explore theoretically and empirically how investment in intercity transport costs affects what products cities tend to specialize in, with heavier investment leading affected cities to engage more in production of heavier products.

Gabaix (1999) specified a model of cities facing amenity shocks in which Zipf's Law was generated (conditional on a lower limit on how far any city could deteriorate). This generated a wave of papers focused on the distribution of city sizes. Indeed, the urban hierarchy seems to follow the Zipf Law, also known as the Rank-Size Rule, which says the population of a city ranked n in the city distribution will host $1/n$ times the population of the largest city (Gabaix and Ioannides, 2004). This requires qualification. Using US Census 2000 data, Eeckhout (2004) studies the entire population size distribution and finds that the non-truncated distribution is lognormal, rather than Pareto. This work triggered a succession of papers offering alternative models for Zipf's Law. And for some period, models of systems of cities took as a benchmark whether they could generate Zipf's Law at least locally (Duranton, 2007; Rossi-Hansberg and Wright, 2007).[8]

More recent work has focused on the *heterogeneity* of firms, people, and city amenities across urban systems. There are so many published and on-going papers exploring these topics; we illustrate with a few of the earlier ones. There are several key issues. First, what happens to city size equilibria when cities are on sites of different qualities, or face different levels of natural amenities? Better quality cities will tend to be bigger but have an incentive to restrict population, to not dissipate their advantage through increased commuting and other diseconomies.

Next, how do heterogeneous workers and firms sort across cities? In general papers find that better quality workers (who can also invest in skills) benefit more from being in bigger cities than less talented workers and thus sort disproportionately into bigger cities (Behrens et al., 2014; Davis and Dingel, 2019). What makes the latter paper unique is that spillovers are endogenous as they stem from the individual choices of costly exchange of ideas. In the aggregate, the level of productivity depends on a social technology that may be more or less efficient, thus reflecting a variety of attitudes toward innovation that characterize different social environments. Skill sorting is responsible in part for the increased inequality between high and low skill workers in the USA (Moretti, 2012; Diamond, 2016). Some of these papers combine aspects of systems of cities and quantitative spatial models, reviewed below, and estimate fully structural models.

Then, besides the spatial sorting of workers is the so-called selection of firms. Competition is usually tougher in large markets than in small ones because the former attract more firms than the latter. When firms are differentiated by their productivity, one may expect those which stay in business to be sufficiently efficient to survive in a competitive environment. By contrast, less efficient firms can survive in a small market where local competition is softer. Under these circumstances, the larger cities would be more efficient because they host the most efficient firms (Syverson, 2004; Melitz and Ottaviano, 2008). This relationship between market size and productivity is called the spatial selection of firms.

While worker sorting is important in explaining the density elasticity, there is no agreement yet on whether firm selection is important. In a detailed study, Combes et al. (2012) use the linear-quadratic model of monopolistic competition, which captures competitive effects, and develop a model that nests both agglomeration economies and heterogeneous firms. They find that firm selection is unlikely to be the main reason why French firms cluster in cities. Gaubert (2018) workers with CES preferences in a setting in which city developers compete to attract firms. She finds, in contrast, that half of the density elasticity in France would be due to firm selection.

Empirically, what spatially separated markets are, or what spatial units are used, is important because the effects of agglomeration economies are often very localized whereas the selection of firms could operate over much larger areas. For example, using Chinese provinces, which are known to be relatively fragmented, Ding and Niu (2019) find that market size matters for the selection of firms for 15 out of 29 sectors.

2.2.2. *Growth in urban systems*

The work reviewed above generally involves static systems of cities. What happens as an economy grows in terms of both population and capital accumulation? Henderson and Ioannides (1981) formulated an exogenous growth model with exogenous population growth and technological change, worrying about the malleability of capital. Black and Henderson (1999) formulated an endogenous growth model for systems of cities, with endogenous human capital formulation and increases in city sizes and potentially numbers of each type. Human capital accumulation enhances static agglomeration economies (Marshall's 'mysteries of the air' become more valuable) leading to increased city sizes. Rossi-Hansberg and Wright (2007) introduced a version with stochastic elements which generated the Zipf Law.

Duranton and Puga (2023) go further by allowing heterogeneity across cities in amenities that evolve over time and again endogenous human capital investment by people. Higher amenity cities can potentially pay higher real incomes and potentially will be of larger sizes. While they do study the evolution of a system of cities, they focus on the idea that incumbent "owner-occupier" residents of high amenity cities have an incentive to restrict entry to the city, so that in-migration does not dissipate the net agglomeration benefits through escalating commuting costs. Not only is the result that utility levels differ across cities by amenity levels, but the outcome is inefficient. As they show in counterfactuals based on estimated parameters of the model, welfare gains are to be made by having more entry into high amenity cities, rather than those people being shunted into low amenity sites.

A final comment concerns papers with cities and durable capital looking at the growth process when cities have non-malleable capital. Glaeser and Gyourko (2005) with durability in mind look at growth versus decline of cities subject to shocks. In declining cities, population is held in place by sharp declines in rents/housing prices, as they show in the data, avoiding large fluctuations in population that would occur with reversible investment. Related, Henderson and Venables (2009) model an economy with population growth, growing from one city to many over time. Cities form sequentially without population fluctuations of existing cities. Each new city grows from scratch to an equilibrium final size (absent on-going technological change), with existing

[8] One must keep in mind that the Zipf Law can be found in a wide variety of phenomena, e.g., word usage in human languages, webpage visits, and earthquakes, which can hardly be micro-founded in economics (Li, 2002; Cominas-Murtra and Solé, 2010).

cities maintaining their sizes. This equilibrium is maintained with free mobility of people across cities by cycles in housing rents.

2.3. New economic geography

The Industrial Revolution exacerbated regional disparities within countries, revealing new forces driving regional decline and rise. It was the great merit of Krugman (1991) to show that regional imbalance may emerge as an equilibrium outcome of the trade-off between increasing returns and transportation costs. By marrying the CES model of monopolistic competition with iceberg transport technology, Krugman developed what became to be known as "new economic geography" [NEG]. Krugman considers a setting with two dimensionless regions, two production factors and two sectors (the core-periphery [CP] model). One sector produces a differentiated good under monopolistic competition and increasing returns, using only skilled labor; the second sector supplied a homogeneous good produced under perfect competition and constant returns, using only unskilled labor. Shipping the homogeneous good is costless, but shipping the differentiated good requires scarce resources, so that the demand for this good varies with workers' locations. A key ingredient in Krugman is that skilled workers are mobile between the two regions, whereas unskilled workers are immobile and evenly distributed. Turning next to the specific conditions for agglomeration or dispersion to arise, Krugman asks when the uniform pattern ceases to be a stable spatial equilibrium.

When the skilled move, they bring with them both their production and consumption capabilities. As a result, their movements affect the size of labor and product markets in both the origin and destination regions. These effects have the nature of pecuniary externalities because skilled workers do not take them into account. Pecuniary externalities, which differ from the agglomeration economies, are relevant in imperfectly competitive markets because prices do not reflect the true social value of individual decisions. If transportation costs are sufficiently high, then interregional shipments of goods are expensive. As a result, the economy displays a symmetric regional pattern of production in which firms focus on regional markets. By contrast, if transportation costs are sufficiently low, skilled firms and workers will concentrate in a single core region, whereas the peripheral region supplies only the homogeneous good. In this way, firms can exploit increasing returns by selling more in the larger market without losing much business in the smaller. Krugman's model thus allows for the possibility of agglomeration or dispersion between regions. Note the difference between NEG and urban economics: the former assume imperfect competition and internal increasing returns whereas the latter often relies on perfect competition and external increasing returns.

One of the main criticisms of the CP model is that it ignores the fact that a growing spatial concentration of people intensifies competition for land and, therefore, leads to higher housing prices and longer commutes. In other words, even when nominal wages increase with employment density, rising urban costs make large agglomerations less attractive. In this context, transportation and commuting costs have opposite impacts on the space-economy. More specifically, when commuting costs are high, with an endogenous number of cities until the CP model, the economy involves a dispersed pattern of small cities because urban costs are too high when the number of cities is low. When commuting costs decrease, cities are fewer and larger because market size still matters, if transportation costs are not too low. By contrast, when commuting costs are not too high, decreasing transportation costs leads to more and smaller cities. Indeed, since the level of urban costs is unaffected when the population distribution remains the same, dispersion forces offset agglomeration forces (Murata and Thisse, 2005).

From Krugman to Baldwin et al. (2003) through Fujita et al. (1999b), the CP model was the origin of one of the liveliest research topics of the 1990s. Since then, it is fair to say that NEG has been fading, as shown by the small number of papers published in the top-5. One reason for this decreasing lack of interest is the poor analytical tractability of Krugman's model.[9] Another, more fundamental reason is the dimensionality problem. The new critical ingredient that a multiregional setting brings about is that the accessibility to spatially dispersed markets varies across regions. When there are more than two regions, any global or local change in the regional economy is likely to trigger complex effects that vary in non-trivial ways with the structure and shape of the transportation network. This is why existing multi-regional NEG models typically rely on simple geometries such as a circle.

3. Quantitative spatial models

We turn to the big development of the last 10 years in urban and spatial work: the formulation of quantitative spatial models. We start with models that look at the allocation of resources across a fixed number of regions of a country, before turning to those looking at the internal allocation of resources within a city. Note the key difference between the QSM models of the allocation of resources across regions and the hierarchy and systems of cities models is that in one the number of spatial entities is fixed, while in the others a key focus concerns the endogenous formation and numbers of spatial entities.,

3.1. The allocation of resources across regions of a country

Eaton and Kortum (2002) may be viewed as the baseline origin of QSM. EK developed a general equilibrium model of trade across countries in which sectors produce a unit mass of differentiated goods under constant returns and perfect competition. Consumers are distributed across a finite number of locations and have CES preferences so that each good is consumed everywhere and produced somewhere. By assuming that locations have access to different technologies while goods are costly traded, EK can show that each location is good at producing some things but is poorly efficient in others. Investment in transport infrastructure lowers the costs of trade and allows locations to specialize more by exporting goods for which they have high productivity draws and importing those for which they have low draws. What makes this paper unique is that technologies are drawn from the Fréchet distribution, a close relative to the multinomial logit. As a result, EK's setting has the nature of a friendly discrete choice model, which allows them to describe the equilibrium outcome through explicit structural gravity equations. The same basic idea lies at the foundation of all quantitative spatial models.

In implementation of these models, papers calibrate, estimate via regression analysis (like gravity equations), or estimate moments to either recover or impose parameters of the model. The point then is to conduct counterfactuals, to evaluate welfare gains and/or losses from, say, policy reforms or transport investments. The big advantage of this general equilibrium approach is that it accounts for all the direct and indirect effects associated with the provision of, for example, new transport infrastructure. For example, a location that is not directly affected by a new link can be indirectly affected through the redistribution of workers associated with the decrease in transport costs along some least-cost routes. A prominent example of this approach is Allen and Arkolakis (2014) who develop a continuous location model for the U.S. and use a rich, in-depth treatment of transport costs, which are modeled by considering all geographic details, including the least-cost routes across all modes. Production is assumed to be perfectly competitive, while the Armington assumption explains why trade occurs. Allen and Arkolakis use their model to assess the effects of the interstate highway system by recomputing all bilateral transport costs without the interstate highway option, which would reduce total welfare by 1.1–1.4%.

[9] It was not until Robert-Nicoud (2005) that a complete characterization of the equilibrium became available. A much more analytically tractable specification of the CP model has been proposed by Ottaviano et al. (2002).

In two rich, meticulous papers, Donaldson (2018) and Donaldson and Hornbeck (2016) study the effect of the development of railroads in colonial India (1870–1930) and in the U.S. from 1870 to 1890, respectively. Railroads in India decreased transport costs and increased agricultural output in the connected districts by 17%. Since railroads allowed the different regions to exploit the gains from trade, there was also an overall increase in income for India. For the U.S., Donaldson and Hornbeck (2016) find a strong increase in agricultural output in districts with railroad access.

While NEG assumes a mechanical relationship between migration flows and relative real wages, the reality that so many people live in places where they earn relatively low real incomes suggests there is so much more to migration costs and how to model them and think about them (Gollin et al., 2014; Zabek, 2024). This has led Diamond (2016) and Redding (2016) to introduce individual mobility into these regional models by means of a logit-like discrete choice model. As a result, particular workers are sorted out according to their productivity whereas others are gathered along their preferences for local amenities, including being at one's birthplace. Papers then investigate the inferred extent (from inter-regional flows) of migration costs with different assumptions about costs of exit (loss of social capital) and entry (migration restrictions) for origins and destinations. Tombe and Zhu (2019) find that the average cost of intra-provincial migration in China is around 51% of the annual income, whereas the average cost of inter-provincial migration ranges from 94 to 98% of annual income. To provide context for this figure, note that Bryan and Morten (2019) have estimated that removing all migration barriers in Indonesia would result in a 22% increase in labor productivity. What is apparent from these two papers is that inter-regional migrations costs can be astonishingly high in developing countries (even if units of time migration costs are incurred over are not always clear). That may explain why there are such large wages gaps between cities and rural areas for seemingly identical workers in some development contexts.

3.2. QSM and the internal allocation of resources within cities

On resource allocation within cities, the big revolution came with the "Berlin-Wall paper" by Ahlfeldt et al. (2015). This presented a tractable discrete choice model with city blocks/places/neighborhoods, endogenous floor space that may be residential or commercial, workers in an open city model. Besides housing, there is one consumption good that is produced by a sector operating under a constant returns Cobb-Dougals technology; this good is costlessly traded. There are travel/commuting times between each pair of blocks. There are two critical parts. First each block is endowed by a residential and a production amenity. Second, each potential worker in the city receives idiosyncratic preferences on each possible workplace-residential place pair. Workers draw from a "friendly" (us/(us + them)) distribution, the Fréchet, which ensures that all places have both residents and workers, in opposition to even the polycentric city models.[10] Expected fractions of workers and residences by block each have closed-form expressions and intuition, as do the expressions for flows across pairs of residence-work places. For example, a block has more workers if it has higher residential amenities, lower floor prices, a higher average idiosyncratic component, and better access to jobs. Ahlfeldt et al. use the exogenous event of the fall of the Berlin Wall to help identify parameters.

Since Ahlfeldt et al. (2015), dozens more using this basic framework have been written and there is an on-going flow, as PhD students writing on intra-city phenomena strive to embed their work in a QSM. A concern in these applications is that it is often unclear whether the comparative statics that are usually the object of the model are like "theorems" that teach us something about equilibrium outcomes, or if they reflect details of the specific real-world problem to which the model is applied. There have been many improvements: allowing for different types of workers, an endogenous component to production amenities (such as local scale) or consumption amenities such as share of high skill in the local area or racial composition. We note one recent work out of many.

Heblich et al. (2020) examine the impact on the development of London over almost a century of the innovation of the building of the commuter and underground rail system. The modelling is based on the "Berlin-Wall" paper, but the adjustments to the modelling make excellent use of the available but limited data over a century ending in 1921. The paper goes back to a key point of the monocentric and polycentric city papers: reductions in commuting costs will have huge impacts on the shape and nature of cities, both commuting patterns and density gradients. In 1840 in London, modal commuting distance was about a kilometre, what a person would walk; and the day and night populations of the different districts of London were about the same, implying that most people lived near their workplaces. The reduction in commuting costs given by the construction of over and underground rails created the separation of living and workplaces. Modal commuting distance rose to 5–6 kms by 1921. The city of London (despite overall population growth of the metro area) lost all but 10–15% of its residential population, while its daytime population (work force primarily) more than doubled from about 1865 to 1921. What is the insight? As developing country cities invest in improved commuting modes whether rails, bus rapid transit or highways, major cities will transform from a "collection of villages", where people walk to work and employment activity is dispersed, to modern cities with more distinct and heightened centers of employment agglomeration.

How does this relate to the monocentric or polycentric models that existed before 2015? A key issue concerns amenities which are "residuals" that explain the "unmodelled" components of blocks. So, for example blocks in central business districts (which still exist in most cities) will have high production amenities as a residual to explain why so much employment is concentrated there. These "amenities" are determined to make up *any* gap between the actual and estimated/calibrated city sizes, rather than dependent on spillovers as in the Ogawa-Fujita model. More problematic, regressing the log of the actual city size on the log of the computed amenity parameters, Berliant and Mori (2017) find that most of the variation is explained by amenities.

Equally important, the way heterogeneity is introduced in many papers has an implication that has been disregarded: if wages and rents were equalized across locations, people would still prefer to locate at the space center, implying a *propensity for centrality* because average commuting costs are minimized with centrality. In the absence of more familiar agglomeration effects operating through production, there are two 'agglomeration' forces, the propensity for central work and the propensity for central residence. These propensities are not agglomeration forces in the conventional sense. They do not incentivize geographic concentration, rather they incentivize concentration in the central location. Even under constant returns and in the absence of first-nature advantage, these forces may be sufficient for the unique equilibrium outcome to be given by a monocentric city (Thisse et al., 2024).

3.3. The immediate future

Today, if one is hiring on the job market in the urban field, there are lots of applicants and most feel compelled to have a QSM, so as in principle to measure general equilibrium and welfare effects. Many students may have little urban training but are driven to the field because of the increased availability of geo-coded historical data and modern "big data", including satellite data, aerial photo data, housing sales and financing data, travel time data, cell phone data and some social media data, as well as increased spatial resolution in census data of many different countries.

Most QSM papers have no government behavior and financing nor dynamics involving investment decisions. On government, one attempt

[10] Discrete choice models had already proven to be a powerful tool to study spatial competition models (Anderson et al., 1992).

has local government budget constraints and differing local objective functions for the government in China (Henderson et al., 2022). On dynamics, papers like Cruz and Rossi-Hansberg (2024) have forward looking behavior and migration costs within and across countries (as well as trade costs), given the idea of climate refugees looking to the future. In fact, there is a new, large literature on climate change and the effect on population flows, emissions and climate of tariffs, migration, regulation, and/or transport policies, with dynamic elements. For example, Balboni (2024) looks at the allocation of road investments in Vietnam from 2000–2010, given flooding of coastal areas in the future, allowing for forward looking behavior up to 2100. However, she has no capital, government budget and financing, nor road construction beyond 2010.

In general, dynamics involving investment decisions are missing from QSM's. Kleinman et al. (2023) is an exception. While workers do not save and invest, location-specific landowners make individually optimal savings and investment decisions. These authors examine adjustment to a steady-state and study the decline in the rate of income convergence across USA states. While there is capital depreciation, investments are reversible, removing the durability component that is essential to some applications.

4. Urban economics in developing countries

There is now an emerging urban economics literature focused on developing countries. There are papers that use particular settings and policy contexts (like China), that provide plausibly exogenous policy shocks that allow us to better quantify and identify aspects of typical urban models. We note countries like China now have enough academic urban economists to have an urban economic association and annual meetings. We don't discuss these types of papers here; the list is too long, and the objectives so diffuse. Second is work trying to distinguish urban issues in developing countries from those in developed countries, unlike early work which tried to show that developing country cities experienced the same economic forces and patterns as developed countries. This literature says there are issues that are particular to developing countries that demand different models and questions (Glaeser and Henderson, 2017).

Such work stretches back to the Arthur Lewis model of the static allocation of people between a formal urban sector and an informal rural one, in a context where countries are undergoing structural and corresponding spatial transformation, as people leave the countryside for cities and agriculture for manufacturing or service sector jobs. What makes this different than the economic history of structural transformation in developed countries is that the process is so rapid, taking a few decades as opposed to a century or more. That speed means low-income countries need to try to develop institutions to support the building and financing of cities in a comparatively short space of time, an enormous challenge.

Structural transformation has a long tradition in development economics and is covered in major development texts today, with some more dynamic 2-sector versions of the basic model. There is a review of the urban aspects of that literature in Desmet and Henderson (2015) and of current issues in urbanization in Henderson and Turner (2020). We do not attempt another review but note one item already partially covered in this paper. Gollin et al. (2013) observe the stark differences in urban-rural wages particularly in sub-Sharan Africa, while other papers focus on the substantial urban wage premium in many developing countries. Given this gap, why don't more people move to cities, than already have; why isn't the pace of urbanization faster? There are two sets of answers. First, as noted earlier, is that migration costs, broadly defined, are high. Second, as explored in Henderson and Turner there are downsides to developing country cities in terms of considerations like health and crime, as well as cost of living differences (Chauvin et al., 2017). While this is an important question and long-standing strand of work, there are so many other issues of cities in developing countries for which urban economics is only starting to scratch the surface.

Our models traditionally are based on a developed country world with strong institutions. We think of two key dimensions, apart from governance of cities with weak and corrupt political processes supporting capture by elites at the *local* level, a topic on which there is essentially no economics literature of which we are aware. The first dimension concerns institutions governing land, housing and construction markets. The second concerns urban financing and fiscal capacity of cities, where typically cities in many countries collect only a small fraction of the assessed property taxes, a major policy issue.[11] We focus on the first dimension with the example of land markets.

In developed countries, urban land is typically owned privately with public records and clear titling and records of encumbrances. This is not the case in much of the world. In Latin America, aspects of land market failure led to a literature on "squatting" that prevailed in the 1980′s and 1990′s (Brueckner and Lall, 2015). In Asia and Africa, many developing countries operate in theory with a dual system such as informal communal rights in rural areas and formal private ownerships in parts of cities, especially, in relevant contexts, in the tiny former colonial sections of cities. As cities expand out into rural areas under communal rights, typically there is not a straightforward process to convert to formal rights and much land in cities falls into informal usage with various degrees of rights (use right, possessory right, quasi-illegal use and the like). Moreover, even then formalized land recordings and disputes can be subject to a high degree of corruption.

As alluded to in Section 1.2, land with poor rights generates limited investment due to concerns about expropriation, as well as issues of obtaining financing and insurance for projects. That inhibits city development and raises costs of housing a population. Given frictions in converting informal land to formal usage that in many cities results in a very slow transition to formality and intense investment, NGO's and international agencies like the World Bank and UN have many programs attempting to improve living conditions in slum, or informal areas of developing country cities. There are now papers emerging that analyze aspects of these issues.

Michaels et al. (2021) document the demand by the middle class in developing country cities for titled and well-planned communities, with road access and regular layout of neighborhoods, as opposed to the hodge-podge that prevails in many contexts. That paper looks at the historical sites and services project of World Bank program in Tanzania, a program that operated in many countries. Sites and service provided regularly laid-out neighborhoods with city blocks and universal road access for plots upon which families self-build. Today in urban Tanzania, these neighborhoods which were once on the outskirts of rapidly growing cities have become part of the urban core. They command high price premiums, compared to adjoining neighborhoods that were not part of the program.

Gechter and Tsivanidis (2023) in a paper in process examine the impact of releasing large tracts of unused land in the main part of Mumbai (former mill land) for formal development. Their model has informal housing and commercial uses on either formal or informal land, two types of consumers, and an analysis of the frictions in converting informal to formal land and in residents moving. They have a 2-period model, before the release of the land and after. In this complex setting they have two key findings: areas adjoining the newly, formally developed former mill lands gentrify, and this gentrification then leads to welfare losses for poorer residents of the former informal areas who are evicted from their housing.

Harari and Wong (2024) in a paper in process look at the dimension of slum upgrading projects. Their paper is set in Jakarta looking at slum improvement programs that occurred 30 or more years ago. They find that slum improvement indeed improved the life of residents at the time. But there is a twist. They argue that upgrading shored up informal rights

[11] See the policy note by Manwaring and Regan (2019) on Kampala.

and enshrined fragmented land use. Areas subject to this program compared to similar neighboring areas that were excluded from the program today have worse outcomes in terms of quality and intensity of housing investment. The excluded areas were able to redevelop and upgrade earlier, rather than being stuck in an old regime.

5. Concluding remarks

We return to the original question: fifty years after the founding of the JUE, what would we tell the senior economists of that day who were defining urban work through their own work and training students, in particular Ed Mills but also Kain, Rothenberg, Tolley, and the like? While the paper has covered a small selection of topics, it is clear from just that selection that the field has moved far beyond the monocentric model. Urban economists have developed models and empirics on many aspects of intra-city allocation of resources and the micro foundations of agglomeration economies, dealing with crime, congestion pricing, the spatial fragmentation of firms from the city center to suburbs to Asia and Latin America, social interactions and externalities, race, retailing, housing, regulation, pollution, education, financing, and so on. We are starting to develop analytical models of irreversible capital investment in cities, so relevant to developing countries with their rapidly growing cities and waves of redevelopment and intense construction. We have QSM's that allow for better defined welfare analysis. They retain the centrality driven by commuting but allow for a rich development of cities with neighborhoods of both work and residence, commuting to work and schools, sorting of heterogeneous residents and firms, endogenous amenities and the like. Many challenges remain; and, in 50 years from now, we are sure we would have been amazed at what will have transpired. Desperately needed today especially in thinking about the urbanizing world in Asia and Africa are models that deal with land use planning and regulation in a more complete way, that deal with modal choice that covers not just rails but the array of common modes like roads, and that start to think about local political economy, to name a few aspects. Simple things we don't know are an ever-evolving challenge. For example, is building tall cities good or bad for the environment, and what could be other effective instruments for cities to face the challenges raised by a less friendly environment?

Fifty years ago, we had just invented the first general equilibrium model of the allocation of resources across cities which started us thinking about different types of cites and how they interact and what are the drivers of whether we have an economy with a few large (perhaps over-sized) cities versus many smaller ones, albeit in a world with no intercity transport costs. Again, we have taken that framework and developed a whole array of innovations: selection of heterogeneous workers and firms, products cycles, the role of developers and city governments in determining city sizes with the regulation tools at their disposal, urban innovation, and dynamic models of growth in a system of cities.

With the advent of NEG in the early 1990's, we developed a parallel framework with a fixed number of regions, that incorporates trade-costs across regions and allows us to think about the core and periphery of countries, relevant to many contexts. QSM models of the inter-regional allocation of resources dealt with a whole new array of issues and are in some sense more tractable than NEG and, of course, oriented to welfare analysis. These models have been useful in thinking about impacts of policies to do with inter-city transport and environmental issues, as well as sorting and the role of migration costs in developing countries in explaining huge urban wage and income premiums. The future of urban economics is bright: our world changes fast in different dimensions and there will always be more issues to study, and we will develop new tools to deal with these.

Last, as a final observation, unlike in location theory and industrial organization, in most spatial models firms play a minor role. More specifically, firms move with workers and things work as if workers were snails that carry their firms on their backs. Instead, one expects firms' and workers' mobility to be driven by different forces as households care about utility differentials while firms focus on profit gaps. Consequently, the spatial distribution of economic activity is determined by the location decisions made by very different types of agents. In this case, spatial adjustments are much more complex than those used in the extant literature because firms and households feature different spatial mobility frictions. That many regions strive to attract foreign direct investments is evidence of this dichotomy.

References

Abbiasov, T., Heine, C., Sabouri, S., Salazar-Miranda, A., Sant, P., Glaeser, E., Ratti, C, 2024. The 15-minute city quantified using human mobility data. Nat. Hum. Behav. 8 (3), 1–11.

Abdel-Rahman, H., Fujita, M., 1990. Product variety, Marshallian externalities, and city sizes. J. Reg. Sci. 30 (2), 165–183.

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

Ahlfeldt, G.M., Pietrostefani, E., 2019. The economic effects of density: a synthesis. J. Urban. Econ. 111, 93–107.

Ahlfeldt, G., Baum-Snow, N., Jedwab, R., 2023. The Skyscraper Revolution: Global Economic Development and Land Savings. CESifo. WP No. 10784.

Akbar, P., Couture, V., Duranton, G., Storeygard, A., 2023. Mobility and congestion in urban India. Am. Econ. Rev. 113 (4), 1083–1111.

Albouy, D., Behrens, K., Robert-Nicoud, F., Seegert, N., 2019. The optimal distribution of population across cities. J. Urban. Econ. 110, 102–113.

Allen, T., Arkolakis, C., 2014. Trade and the topography of the spatial economy. Q. J. Econ. 129 (3), 1085–1140.

Alonso, W., 1964. Location and Land Use: Towards a General Theory of Land Rent. Harvard University Press.

American Planning Association, 1950. Urban Land Use. American Society of Planning Officials, Chicago, Illinois.

Anderson, S.P., de Palma, A., Thisse, J.-F., 1992. Discrete Choice Theory of Product Differentiation. MIT press.

Arnott, R.J., Stiglitz, J.E., 1979. Aggregate land rents, expenditure on public goods, and optimal city size. Q. J. Econ. 93 (4), 471–500.

Arnott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak-period congestion: a traffic bottleneck with elastic demand. Am. Econ. Rev. 83 (1), 161–179.

Arzaghi, M., Henderson, J.V., 2008. Networking off Madison Avenue. Rev. Econ. Stud. 75 (4), 1011–1038.

Au, C.-C., Henderson, J.V., 2006. Are Chinese cities too small? Rev. Econ. Stud. 73 (3), 549–576.

Balboni, C., 2024. In harm's way? infrastructure investments and the persistence of coastal cities. Am. Econ. Rev. forthcoming.

Baldwin, R.E., Forslid, R., Martin, P., Ottaviano, G.I.P., Robert-Nicoud, F., 2003. Economic Geography and Public Policy. Princeton University Press.

Baum-Snow, N., Han, L., 2024. The micro-geography of housing supply. J. Polit. Economy 132 (6), 1897–1946.

Beckmann, M., 1968. Location Theory. Random House.

edited by Beckmann, M.J., 1976. Spatial equilibrium in the dispersed city. In: Papageorgiou, Y.Y. (Ed.), Mathematical Land Use Theory. Lexington Books, pp. 117–125.

Behrens, K., Duranton, G., Robert-Nicoud, F, 2014. Productive cities, sorting, selection and agglomeration. J. Polit. Economy 122 (3), 507–553.

Behrens, K, Robert-Nicoud, F, 2015. Agglomeration theory with heterogeneous agents. In: Handbook of Regional and Urban Economics. In: Duranton, G., Henderson, J.V., Strange, W.C. (Eds.), 5. Elsevier, pp. 171–245.

Berliant, M., Mori, T., 2017. Beyond urban form: how Masahisa Fujita shapes us. Int. J. Econ. Theory 13 (1), 5–28.

Black, D., Henderson, J.V., 1999. A theory of urban growth. J. Polit. Economy 107 (2), 252–284.

Braid, R.M., 2001. Spatial growth and redevelopment with perfect foresight and durable housing. J. Urban. Econ. 49 (3), 425–452.

edited by Brueckner, J., Lall, S., 2015. Cities in developing countries. In: Duranton, G., Henderson, J.V., Strange, W.C. (Eds.), Handbook or Regional and Urban Economics. Elsevier, pp. 1399–1455.

Bryan, G., Morten, M., 2019. The aggregate productivity effects of internal migration: evidence from Indonesia. J. Polit. Economy 127 (5), 2229–2268.

Capozza, D., Helsley, R.W., 1990. The stochastic city. J. Urban. Econ. 28 (2), 187–203.

Cavailhès, J., Gaigné, C., Tabuchi, T., Thisse, J.-F., 2007. Trade and the structure of cities. J. Urban. Econ. 62 (3), 383–404.

Chetty, R., Hendren, N., 2018. The impacts of neighborhoods on intergenerational mobility I: childhood exposure effects. Q. J. Econ. 133 (3), 1107–1162.

Chipman, J.S., 1970. External economies of scale and competitive equilibrium. Q. J. Econ. 84 (3), 347–385.

Chauvin, J.P., Glaeser, E., Ma, Y., Tobio, K., 2017. What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. J. Urban. Econ. 98, 17–49.

Church, R.L., Drezner, Z., 2022. Review of obnoxious facilities location problems. Comput. Oper. Res. 138, 105468.

Ciccone, A., Hall, R.E., 1996. Productivity and the density of economic activity. Am. Econ. Rev. 86 (1), 54–70.

Ciccone, A., Peri, G., 2006. Identifying human-capital externalities: theory with applications. Rev. Econ. Stud. 73 (2), 381–412.
Combes, P.-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.
edited by Combes, P.P., Gobillon, L., 2015. The empirics of agglomeration economies. In: Duranton, G., Henderson, J.V., Strange, W.C (Eds.). Elsevier, pp. 247–348.
Corominas-Murtra, B., Solé, R.V., 2010. Universality of Zipf's law. Phys. Rev. E 82 (1), 011102.
Cruz, J.-L., Rossi-Hansberg, E., 2024. The economic geography of global warming. Rev. Econ. Stud. 91 (2), 899–939.
d'Aspremont, C., Gabszewicz, J.J., Thisse, J.-F., 1979. On Hotelling's "Stability in competition". Econometrica 47 (5), 1145–1150.
Davis, O.A., Whinston, A.B., 1964. The economics of complex systems: the case of municipal zoning. Kyklos. 17 (3), 419–446.
Davis, D.R., Dingel, J.I., 2019. A spatial knowledge economy. Am. Econ. Rev. 109 (1), 153–170.
Davis, M.A., Larson, W.D., Oliner, S.D., Shui, J., 2021. The price of residential land for counties, ZIP codes, and census tracts in the United States. J. Monet. Econ. 118, 413–431.
Debreu, G., 1959. The Theory of Value: An Axiomatic Analysis of Economic Equilibrium. Yale University Press.
Desmet, K., Henderson, J.V., 2015. Duranton, Henderson, and Strange. In: The Geography of Development within Countries. Handbook of Regional and Urban Economics, 5. Elsevier, Amsterdam, pp. 1457–1517 edited by.
De la Rocha, J., Puga, D., 2017. Learning by working in big cities. Rev. Econ. Stud. 84 (1), 106–142.
de Palma, A., Ginsburgh, V., Papageorgiou, Y.Y., Thisse, J.-F., 1985. The principle of minimum differentiation holds under sufficient heterogeneity. Econometrica 53 (4), 767–781.
Diamond, R., 2016. The determinants and welfare implications of US workers diverging location choices by skill: 1980-2000. Am. Econ. Rev. 106 (3), 479–524.
Ding, C., Niu, Y., 2019. Market size, competition, and firm productivity for manufacturing in China. Reg. Sci. Urban. Econ. 74, 81–98.
Diodato, D., Neffke, F., O'Clery, N., 2018. Why do industries coagglomerate? how Marshallian externalities differ by industry and have evolved over time. J. Urban. Econ. 106, 1–26.
Donaldson, D., Hornbeck, R., 2016. Railroads and American economic growth: a 'market access' approach. Q. J. Econ. 131 (2), 799–858.
Donaldson, D., 2018. Railroads of the Raj: estimating the impact of transportation infrastructure. Am. Econ. Rev. 108 (4–5), 899–934.
Dong, X., Ross, S.L., 2015. Accuracy and efficiency in simulating equilibrium land-use patterns for self-organizing cities. J. Econ. Geogr. 15 (4), 707–722.
Duranton, G., 2007. Urban evolutions: the fast, the slow, and the still. Am. Econ. Rev. 97 (1), 197–221.
Duranton, G., Overman, H.G., 2005. Testing for localization using micro-geographic data. Rev. Econ. Stud. 72 (4), 1077–1106.
Duranton, G., Murrow, P., Turner, M., 2014. Roads and trade: evidence from the US. Rev. Econ. Stud. 81 (2), 681–724.
Duranton, G., Puga, D., 2001. Nursery cities: urban diversity, process innovation, and the life cycle of product. Am. Econ. Rev. 91 (5), 1454–1477.
edited by Duranton, G., Puga, D., 2004. Micro-foundations of urban increasing returns: theory. In: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics. Cities and Geography, North-Holland, pp. 2063–2117.
edited by Duranton, G., Puga, D, 2015. Urban land use. In: Duranton, G., Henderson, J. V., Strange, W.C. (Eds.), Handbook of Regional and Urban Economics. Elsevier, pp. 467–560.
Duranton, G., Puga, D., 2023. Urban growth and its aggregate implications. Econometrica 91 (6), 2219–2259.
Eaton, J., Kortum, S., 2002. Technology, geography, and trade. Econometrica 70 (5), 1741–1779.
Eeckhout, J., 2004. Gibrat's law for (all) cities. Am. Econ. Rev. 94 (5), 1429–1451.
Ellison, G., Glaeser, E.L., Kerr, W.R., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. Am. Econ. Rev. 100 (3), 1195–1213.
Farahani, R.Z., Fallah, S., Ruiz, R., Hosseini, S., Asgari, N, 2019. OR models in urban service facility location: a critical review of applications and future developments. Eur. J. Oper. Res. 276 (1), 1–27.
Faggio, G., Silva, O., Strange, W.C., 2017. Heterogeneous agglomeration. Rev. Econ. Stat. 99 (1), 80–94.
Faggio, G., Silva, O., Strange, W.C., 2020. Tales of the city: what do agglomeration cases tell us about agglomeration in general? J. Econ. Geogr. 20 (5), 1117–1143.
Flatters, F., Henderson, J.V., Mieszkowski, P., 1974. Public goods, efficiency, and regional fiscal equalization. J. Public Econ. 3 (2), 99–112.
Fujita, M., 1982. Spatial patterns of residential development. J. Urban. Econ. 12 (1), 22–52.
Fujita, M., 1989. Urban Economic Theory. Land Use and City Size. Cambridge University Press, Cambridge.
Fujita, M., Ogawa, H., 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. Reg. Sci. Urban. Econ. 12 (2), 161–196.
Fujita, M., Krugman, P., Mori, T., 1999a. On the evolution of hierarchical urban systems. Eur. Econ. Rev. 43 (2), 209–251.
Fujita, M., Krugman, P., Venables, A.J., 1999b. The Spatial Economy. Cities, Regions and International Trade. MIT Press.
Fujita, M., Thisse, J.-F., 2013. Economics of Agglomeration. Cities, Industrial Location and Globalization. Cambridge University Press, 2nd ed.Cambridge.
Gabaix, X., 1999. Zipf's law for cities: an explanation. Q. J. Econ. 114 (3), 739–767.
edited by Gabaix, X., Ioannides, Y.M., 2004. The evolution of city size distributions. In: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics. Cities and Geography. Elsevier, pp. 2341–2378.
Gaigné, C., Rioux, S., Thisse, J.-F., 2016. How to make the metropolitan area work? Neither big government, nor laissez-faire. J. Public Econ. 134, 100–113.
Gaspar, J., Glaeser, E.L., 1998. Information technology and the future of cities. J. Urban. Econ. 43 (1), 136–156.
Gaubert, C., 2018. Firm sorting and agglomeration. Am. Econ. Rev. 108 (11), 3117–3153.
Gechter, M., Tsivanidis, N., 2023. Spatial spillovers from high-rise developments: evidence from the Mumbai mills. R&R Econom.
Glaeser, E.L., Kolko, J., Saiz, A., 2001. Consumer city. J. Econ. Geogr. 1 (1), 27–50.
edited by Glaeser, E.L., Kahn, M.E., 2004. Sprawl and urban growth. In: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics. Cities and Geography. Elsevier, pp. 2481–2527.
Glaeser, E.L., Gyourko, J., Saks, R.E., 2005. Why have housing prices gone up? Am. Econ. Rev. 95 (2), 329–333.
Glaeser, E.J., Gyourko, J., 2005. Urban decline and durable housing. J. Polit. Economy 113 (2), 345–375.
Glaeser, E.J., Henderson, J.V., 2017. Urban economics for the developing world: an introduction. J. Urban. Econ. 98, 1–5.
Gollin, D., Lagakos, D., Waugh, M.E., 2014. The agricultural productivity gap. Q. J. Econ. 129 (2), 939–993.
Greenstone, M., Hornbeck, R., Moretti, E., 2010. Identifying agglomeration spillovers: evidence from winners and losers of large plant openings. J. Polit. Economy 118 (3), 536–598.
Grover, A., Lall, S., Timmis, J., 2023. Agglomeration economies in developing countries: a meta-analysis. Reg. Sci. Urban. Econ. 101, 103901.
Harari, M., Wong, M., 2024. Slum upgrading and long-run urban development: evidence from Indonesia. R&R Rev. Econ. Stud.
Heblich, S., Redding, S.J., Sturm, D.M., 2020. The making of the modern metropolis: evidence from London. Q. J. Econ. 135 (4), 2059–2133.
Helsley, R.W., Strange, W.C., 1990. Matching and agglomeration economies in a system of cities. Reg. Sci. Urban. Econ. 20 (2), 189–212.
Helsley, R.W., Strange, W.C., 1994. City formation with commitment. Reg. Sci. Urban. Econ. 24 (3), 373–390.
Helsley, R.W., Strange, W.C., 2014. Coagglomeration, clusters, and the scale and composition of cities. J. Polit. Economy 122 (5), 1064–1093.
Henderson, J.V., 1974. The sizes and types of cities. Am. Econ. Rev. 64 (4), 640–656.
Henderson, J.V., 1981. The economics of staggered work hours. J. Urban. Econ. 9 (3), 349–364.
Henderson, J.V., 1988. Urban Development. Theory, Fact and Illusion. Oxford University Press.
Henderson, J.V., Ioannides, Y.M., 1981. Aspects of growth in a system of cities. J. Urban. Econ. 10 (1), 117–139.
Henderson, J.V., 1986. Efficiency of resources usage and city size. J. Urban. Econ. 19 (1), 47–70.
Henderson, V., Mitra, A., 1996. New urban landscape: developers and edge cities. Reg. Sci. Urban. Econ. 26 (6), 613–643.
Henderson, J.V., Turner, M.A., 2020. Urbanization in the developing world. The J. Econ. Perspect. 34 (3), 150–173.
Henderson, J.V., Venables, A.J., 2009. The dynamics of city formation. Rev. Econ. Dyn. 12 (2), 233–254.
Henderson, J.V., Regan, T., Venables, A.J., 2021. Building the city: from slums to the modern metropolis. Rev. Econ. Stud. 88 (3), 1157–1192.
Henderson, J.V., Su, D., Zhang, Q., Zheng, S., 2022. Political manipulation of urban land markets: evidence from China. J. Public Econ. 214, 104730.
Henderson, R., Jaffe, A., Trajtenberg, M., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Q. J. Econ. 108 (3), 577–598.
Hoover, E.M., 1936. Location Theory and the Shoe and Leather Industries. Harvard University Press, Cambridge, MA.
Hotelling, H., 1929. Stability in competition. Econ. J. 39 (153), 41–57.
Hsu, W.T., 2014. Central place theory and city size distribution. Econ. J. 122 (563), 903–932.
Isard, W., 1975. Introduction to Regional Science. Prentice Hall, Englewood Cliffs.
Kleinman, B., Liu, E., Redding, S.J., 2023. Dynamic spatial general equilibrium. Econometrica 91 (2), 385–424.
Koopmans, T.C., 1957. Three Essays on the State of Economic Science. McGraw-Hill.
Koster, H.R.A., Pasidis, I., van Ommeren, J., 2019. Shopping externalities and retail concentration: evidence from Dutch shopping streets. J. Urban. Econ. 114, 103194.
Koster, H.R.A., Thisse, J.-F., 2024. Understanding spatial agglomeration: increasing returns, land, and transportation costs. Annu Rev. Econom. 16, 55–78.
Krugman, P., 1991. Increasing returns and economic geography. J. Polit. Economy 99 (3), 483–499.
Leonardi, M., Moretti, E., 2023. The agglomeration of urban amenities: evidence from Milan restaurants. Am. Econ. Rev.: Insight 5 (2), 141–157.
Li, W., 2002. Zipf's Law everywhere. Glottometrics 5, 14–21.
Liotta, C., Viguié, V., Lepetit, Q., 2022. Testing the monocentric standard urban model in a global sample of cities. Reg. Sci. Urban. Econ. 97, 103832.
Lösch, A., 1940. Die Räumliche Ordnung der Wirtschaft. Jena: Gustav Fischer. English translation: The Economics of Location, 1954. Yale University Press, New Haven (CN).
Loumeau, G., 2023. Locating public facilities: theory and micro evidence from Paris. J. Urban. Econ. 135, 103544.
Lucas, R.E., Rossi-Hansberg, E., 2002. On the internal structure of cites. Econometrica 70 (4), 1445–1476.

Manwaring, P., Regan, T., 2019. Property taxes in Kampala. In: ICG policy note, p. 43448.
Marianov, V., Eiselt, H.A., 2024. Fifty years of location theory - a selective review. Eur. J. Oper. Res. 318 (3), 701–771.
Marshall, A., 1890. Principles of Economics, 8th ed. Macmillan. published in 1920.
MasColell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic Theory. Oxford University Press.
McFadden, D., 1974. The measurement of urban travel demand. J. Public Econ. 3 (4), 303–328.
McFadden, D., 1986. The choice theory of market research. Mark. Sci. 5 (4), 275–297.
Melitz, M., Ottaviano, G.I.P., 2008. Market size, trade, and productivity. Rev. Econ. Stud. 75 (1), 295–316.
Mills, E.S., de Ferranti, D.M., 1971. Market choices and optimum city size. Pap. Proc. Am. Econ. Assoc. 61 (2), 340–345.
Mills, E.S., 1972. Studies in the Structure of the Urban Economy. The Johns Hopkins Press.
Moretti, E., 2004. Workers' education, spillovers and productivity: evidence from plant-level production functions. Am. Econ. Rev. 94 (3), 656–690.
Moretti, E., 2012. The New Geography of Jobs. Houghton Mifflin Harcourt.
Murata, Y., Thisse, J.-F., 2005. A simple model of economic geography à la Helpman–Tabuchi. J. Urban. Econ. 58 (1), 137–155.
Ogawa, H., Fujita, M., 1980. Equilibrium land use patterns in a non-monocentric city. J. Reg. Sci. 20 (4), 455–475.
Ota, M., Fujita, M., 1993. Communication technologies and spatial organization of multi-unit firms in metropolitan areas. Reg. Sci. Urban. Econ. 23 (6), 695–729.
Ottaviano, G., Tabuchi, T., Thisse, J.-F., 2002. Agglomeration and trade revisited. Int. Eco. Rev. 43, 409–435.
Peri, G., 2005. Determinants of knowledge flows and their effect on innovation. Rev. Econ. Stat. 87 (2), 308–322.
Pietrabissa, G., 2023. School access and city structure. In: CEMFI. Mimeo.
Puga, D., 2002. European regional policies in light of recent location theories. J. Econ. Geogr. 2 (4), 373–406.
Puga, D., 2010. The magnitude and causes of agglomeration economies. J. Reg. Sci. 50 (1), 203–219.
Redding, S.J., 2016. Goods trade, factor mobility and welfare. J. Int. Econ. 101, 148–167.
Riley, J.G., 1974. Optimal residential density and road transportation. J. Urban. Econ. 1 (2), 230–249.
Rossi-Hansberg, E., Wright, M.L.J., 2007. Urban structure and growth. Rev. Econ. Stud. 74 (2), 597–624.
Rossi-Hansberg, E., Sarte, P.-D., Owens III, R, 2009. Firm fragmentation and urban patterns. Int. Econ. Rev. (Philadelphia) 50 (1), 143–186.
edited by Rosenthal, S.S., Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, J.V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics. Cities and Geography, North-Holland.
Rosenthal, S.S., Strange, W.C., 2020. How close is close? The spatial reach of agglomeration economies. J. Econ. Perspect. 34 (3), 27–49.
Saiz, A., 2010. The geographic determinants of housing. Q. J. Econ. 125 (3), 1253–1296.
Small, K., Verhoef, E.T., 2007. The Economics of Urban Transportation. Routledge.
edited by Solow, R.M., 1973. On equilibrium models of urban locations. In: Parkin, J.M. (Ed.), Essays in Modern Economics. Longman, pp. 2–16.
Solow, R.M., Vickrey, W.S., 1971. Land use in a long narrow city. J. Econ. Theory 3 (4), 430–447.
Starrett, D., 1978. Market allocations of location choice in a model with free mobility. J. Econ. Theory. 17 (1), 21–37.
Steijn, M.P.A., Koster, H.R.A., Van Oort, F.G., 2022. The dynamics of industry agglomeration: evidence from 44 years of coagglomeration patterns. J. Urban. Econ. 130, 103456.
Syverson, C., 2004. Market structure and productivity: a concrete example. J. Polit. Economy 112 (6), 1181–1222.
Takayama, Y., 2015. Bottleneck congestion and distribution of work start times: The economics of staggered work hours revisited. TR-Part B or Transp. Res. B Methodol. 81, 830–847.
Tombe, T., Zhu, X., 2019. Trade, migration, and productivity: a quantitative analysis of China. Am. Econ. Rev. 109 (5), 1843–1872.
Turner, M., Haughwout, A., van der Klaauw, W., 2014. Land use regulation and welfare. Econometrica 32 (4), 1341–1403.
Vickrey, W.S., 1969. Congestion theory and transport investment. Am. Econ. Rev. 59 (2), 251–260.
Wheaton, W., 1982. Urban spatial growth under perfect foresight. J. Urban. Econ. 12 (1), 1–21.
Zabek, M., 2024. Local ties in spatial equilibrium. Am. Econ. J. Macroecon. 16 (2), 287–317.
Zhang, Q., 2007. A micro-foundation of local business cycles. Reg. Sci. Urban. Econ. 37 (5), 568–601.