DISCUSSION PAPER SERIES

No. 10816

URBAN NETWORKS: SPREADING THE FLOW OF GOODS, PEOPLE, AND IDEAS

Edward L Glaeser, Giacomo AM Ponzetto
and Yimei Zou

INTERNATIONAL TRADE AND REGIONAL ECONOMICS and MACROECONOMICS AND GROWTH

Centre for Economic Policy Research

ISSN 0265-8003

URBAN NETWORKS: SPREADING THE FLOW OF GOODS, PEOPLE, AND IDEAS

Edward L Glaeser, Giacomo AM Ponzetto and Yimei Zou

Discussion Paper No. 10816
September 2015
Submitted 07 September 2015

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL TRADE AND REGIONAL ECONOMICS and MACROECONOMICS AND GROWTH**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Edward L Glaeser, Giacomo AM Ponzetto and Yimei Zou

URBAN NETWORKS: SPREADING THE FLOW OF GOODS, PEOPLE, AND IDEAS[†]

Abstract

Should China build mega-cities or a network of linked middle-sized metropolises? Can Europe's mid-sized cities compete with global agglomeration by forging stronger inter-urban links? This paper examines these questions within a model of recombinant growth and endogenous local amenities. Three primary factors determine the trade-off between networks and big cities: local returns to scale in innovation, the elasticity of housing supply, and the importance of local amenities. Even if there are global increasing returns, the returns to local scale in innovation may be decreasing, and that makes networks more appealing than mega-cities. Inelastic housing supply makes it harder to supply more space in dense confines, which perhaps explains why networks are more popular in regulated Europe than in the American Sunbelt. Larger cities can dominate networks because of amenities, as long as the benefits of scale overwhelm the downsides of density. In our framework, the skilled are more likely to prefer mega-cities than the less skilled, and the long-run benefits of either mega-cities or networks may be quite different from the short-run benefits.

JEL Classification: F15, O18, R10 and R58
Keywords: cities, growth, migration and networks

Edward L Glaeser eglaeser@harvard.edu
Harvard University

Giacomo AM Ponzetto gponzetto@crei.cat
CREI, Universitat Pompeu Fabra, Barcelona GSE and CEPR

Yimei Zou yimei.zou@upf.edu
Universitat Pompeu Fabra

[†] Glaeser acknowledges financial support from the Taubman Center for State and Local Government. Ponzetto acknowledges financial support from the Spanish Ministry of Economy and Competitiveness (RYC-2013-13838 and ECO-2014-59805-P), the Government of Catalonia (2014-SGR-830) and the Barcelona GSE. Zou acknowledges financial support from the La Caixa-Severo Ochoa International Doctoral Fellowship.

1 Introduction

Should rapidly urbanizing nations, like China, build mega-cities or disperse their population into networks of smaller urban centers linked by high-speed transit? Should Europe, which is historically tied to its smaller cities, work to cultivate urban networks as a substitute for large scale agglomeration? What factors determine whether mega-cities dominate urban networks?

While the United State has long accepted the unbridled growth of regionally dominant metropolitan areas, such as New York, Los Angeles and Chicago, Europe and China more seriously question whether networks can be a substitute for mega-city growth. China's interest in urban networks is motivated by scale. Urban population growth is so immense that mega-cities seemed doomed to terrible diseconomies of density, such as crowding and traffic congestion. Europe's interest is motivated by history. Since Europe urbanized when transport costs where high, the continent has an abundance of smaller cities that it does not wish to abandon. Busan—the second largest city in South Korea—would also be the second largest city in Western Europe. Section 2 briefly discusses the history and current state of discussion on networks of cities.

This paper investigate the welfare economics of urban networks, defined as groups of cities that enable the exchange of goods, people and ideas. In our framework, networks enable the movement of skilled workers and ideas. They differ from cities in three ways. While local entrepreneurs can draw inspiration from the wider network, they particularly depend on their home city for intellectual resources. Less skilled workers are able to move readily across employers within a city, but not within a wider network. This distinction follows the U.S. Census definition of a metropolitan area as a single labor market. Finally, amenities are shared within cities but not across cities in an urban network.

We ask two policy-motivated questions. When is it desirable to link cities to form a single large mega-city? When is it desirable to abandon distinct cities altogether and merge populations into a single megalopolis? Our first question imagines a single larger city connected by better transportation and communication technology. Our second question imagines a single denser city with less land. We refer to the first type of city-creation as consolidation and the second type of city-creation as densification. The difference between the two represents the loss of land. Since consolidation brings all of the benefits of densification and delivers more land, the case for consolidation is always better than the case for densification.

We examine urban networks in a model of endogenous amenities and recombinant growth that follows Weitzman (1998) and Jacobs (1969). Different locations are endowed with land and unskilled labor, which can move within a mega-city but not across cities. Locations also begin with initial stocks of firms and amenities. Skilled workers can move across cities within an urban network and must therefore be indifferent between cities within the network. Housing is produced with land and the numeraire good. Individuals privately invest in amenities, but there are spillovers both across time and space. We benefit from amenities built in the past, such as beautiful old buildings, and amenities built by our neighbors, such as beautiful new buildings.

The economic engine of the model is that skilled workers create new firms, which provide value by generating new varieties of intermediate goods following Ethier (1979). The number of new firms in a location is a function of the number of old firms in that place, the number

of skilled workers in that place, and the number of skilled workers elsewhere in the urban network. The number of old firms matter because we assume that new ideas are produced from old ideas. The number of skilled workers in the area matters both because those skilled workers are themselves the entrepreneurs and because they can learn from other entrepreneurs within the area. The number of skilled workers in the network matters because we assume that there are intellectual links that bind the network together. Moreover, the skilled workers in the network may have been educated together and may have learned from each other at that earlier stage. We assume that there are increasing returns in the creation of new firms at the level of the network, but not necessarily at the level of the city.

We first consider the case of locational symmetry, in which every place is endowed with an identical amount of land, unskilled labor, initial firms, productivity levels and initial amenity levels. We consider both the short-run and long-run benefits of urban network expansion, consolidation and densification. We examine the expansion of urban networks through improvements in communications and transportation technology. Within the model, these expansions have no downsides and increase the flow of ideas. Unsurprisingly, larger urban networks mean more firm creation, higher income levels and high utility levels. In reality, these benefits would have to be weighed against the investment costs of linking more cities together into a larger network. Yet, since we do not explicitly model the price of network expansion, which will depend on local technology and costs, our framework is better suited to discuss the trade-offs between networks and cities than the case for network expansion.

Our framework does have a real trade-off when we consider the consolidation of already networked cities into a single mega-city. Even if there is no land lost through the process, city formation can harm entrepreneurship. There can be diminishing returns in new firm creation at the city level, even if there are increasing returns at the level of the network. The model admits the possibility that local culture may be lost when cities are combined, and that creates the possibility that bigger cities may not be better for creativity. If there are local increasing returns to new firm formation, then urban consolidation is always beneficial for incomes and utility. If there are local decreasing returns to new firm formation, urban consolidation may be beneficial for utility, even if it reduces entrepreneurship and income, because more amenities can be enjoyed in a larger city.

The case for densification is always weaker than the case for consolidation because of the loss of land. Even if there are local increasing returns to new idea creation, then densification may be harmful if housing is a large enough share of consumption or if housing supply is sufficiently inelastic. The role of housing supply may explain why mega-cities are more common in the American sunbelt while urban networks, like the Randstad, are more common in Europe. Houston's geography and regulations make it easy to build massive amounts of housing (Glaeser and Tobio 2008). Amsterdam presents a more challenging building environment.

Despite the reduction in the availability of land, it is possible that densification will lead to an increase in the housing stock. This can occur only if densification strongly increases new firm formation and the supply of housing is quite elastic, perhaps as in New York City at the start of the twentieth century. Naturally, housing prices will then be significantly higher after densification.

As the skilled and unskilled have different preferences over amenities, the skilled may

prefer consolidation or densification while the unskilled may prefer separation. With consolidation, this preference divergence can only occur when there are decreasing returns to scale in new idea creation. In the case of densification, it is possible for the unskilled to dislike a bigger city even if incomes go up as long as prices rise sufficiently. The extra amenities generated by a bigger city may be enough to offset the pecuniary losses for the skilled but not for the unskilled, and in that case there may be a political conflict. To us, this conflict between rich and poor resembles the current debate over inequality in cities like New York and London. Urban expansion seems to have been beneficial for skilled, but many less successful urbanites seem convinced that rising housing prices have left them worse off.

We then turn to the long run, first considering the case where the system converges to a steady state. The expansion of urban networks has even larger positive effects in the long run than in the short run, because the process of new firm creation builds up over time. For the same reason, the creation of larger cities through consolidation will also have bigger effects on pecuniary variables in the long run than in the short run, although these effects can be either positive or negative.

Perhaps most interestingly, there could be a reversal of preferences in the long run compared to the short run. If denser cities increase incomes because of increasing returns to scale in new idea creation, but reduce welfare in the short run because of more expensive housing, then it can be that the income gains in steady state are so large that welfare eventually rises with density. The larger lesson is that since densification and consolidation will have a persistent impact on growth, it is important to take a longer term perspective when considering the costs and benefits of changing urban connectivity.

We next allow for heterogeneity: first across regions, assuming that the cities within a region are identical, and then across cities. With asymmetries, it is possible that consolidation or densification can be harmful when done for any one region but helpful for the world as a whole. It is also possible that consolidation or densification can be helpful at the regional level, but harmful for the world as a whole. Increased innovation produces a cross-regional externality. Any change in new firm creation will impact regions elsewhere. When consolidation or densification reduce incomes, then even if local utilities increase because of increases in amenities, global welfare may decline.

We suspect that the more realistic case for a conflict between local and global objectives is that densification raises incomes but lowers welfare at the local level because of higher housing costs. Yet densification at a national or global scale might still be beneficial because of the external benefits of new product creation. There are many successful areas in the United State and Europe, such as Silicon Valley and London, that have restricted densification through land use regulations. These regulations may have created local benefits, but the reduction in density may have also created global harm by reducing the total amount of innovation world-wide.

Finally, we turn to the case where there are heterogeneous locations within a region. We first consider a small set of cities that considers consolidation or densification. If consolidation or densification appeals to the skilled, either by increasing amenities or incomes, then this will attract the skilled to this set of places. The movement of the skilled will make urbanization more appealing on the margin, since the skilled are always more likely to benefit from urban growth than the unskilled, because they care more about the growth in amenities. This result could be reversed if densification created congestion disamenities that are particularly

disliked by the skilled. The larger point in this case is that local choices will impact the migration of entrepreneurs, and this can radically alter the local benefits from urban change.

If land is constant across locations and other factor endowments move together, then a simple condition determines whether densification is beneficial, depending on whether the returns to local scale are decreasing or increasing. If the returns are decreasing then lower density is better, while if there are increasing returns then higher density is better. Consolidation can be unattractive when heterogeneity is low but become appealing if heterogeneity rises enough. If the areas are more different, then there are essentially gains from trade from consolidation, which can increase the returns to forming one large city. This effect gets exacerbated when factor endowment do not perfectly move together. In that case, forming a single city better matches the factor endowments with each other and leads to enhanced welfare.

This result again can help explain why the United States turns to mega-cities while Europe turns to urban networks. American locations can be remarkably different even within a single metropolitan area. The less well educated residents in Oakland benefit from access to the San Francisco job market. European egalitarianism also operates at the city level, which should mean that networks make relatively more sense.

2 Urban Networks in History and Today

In the theory that follows, we will discuss three specific alternatives for spatial organization: larger cities, denser cities, and urban networks. We use "larger" cities to refer to physical locales that may have been once distinct, but are now single labor markets or metropolitan areas. Typically, cities becomes larger because transportation innovations have linked previously disparate locales. The Brooklyn Bridge, for example, turned Manhattan and Brooklyn from distinct cities, with clearly separate central business districts, into a single labor market that would eventually merge into a single political unit. We mean "denser" cities to refer to the case where people have abandoned several locations and come together in a single place. While abandonment is extreme, this densification has been the standard process of urbanization throughout most of human history.

An urban network is defined in the model as a linked set of cities that share ideas and across which skilled labor moves readily. In a sense, a network is just like a city, but with weaker links. Like cities, networks exist to speed the flow of goods, people and ideas.

The first urban networks emerged to facilitate the trade in goods. Networks came before mega-cities because of high transport costs. As argued by Christaller's (1933) Central Place Theory, historically farmers would only travel as far as small market towns to sell, and then merchants brought those goods to regional centers and great national cities. Europe and the eastern United States urbanized when transport costs were high, and so dispersed smaller cities remain.

Yet, as transportation costs have declined, goods-based networks have become far less meaningful, although a few links remain particularly important, such as São Paulo and Santos. Instead, networks are linked by the movement of people and ideas. There has also been a rise in long-distance networked cities, linked typically by air and telecommunications, such as New York and London, and a rise of short-range linkages, such as the towns of

the Brabant. The shortest-range networks can be difficult to distinguish from a standard commuting zone.

In areas that urbanized after 1900, including the Western United States, Latin America, and Asia, transport costs were lower and massive agglomerations became far more standard than networks of smaller cities. Yet even in those newer areas, the debate between networks and agglomerations is not dead. In China, there is a lively urban planning debate about whether to facilitate the increased expansion of the vast agglomerations of Beijing and Shanghai or whether to focus on creating networks of cities that are smaller, albeit still much larger than almost all of the cities of Western Europe. The current government policy favors networks, in the hope that connected smaller cities may be free of the extreme downsides of mass agglomeration, such as extreme congestion, pollution and high housing costs.

2.1 Urban Networks in History

The most ancient urban networks formed along rivers, such as the Indus, the Tigris and the Euphrates, more than six thousand years ago. One plausible view is networks emerged first for trade reasons, but were then consolidated into political units under the military power of kings and empires. This consolidation required military technology and organization to reach the point where the residents of one city were able to control the residents of a second city.

The Greeks developed the first great European urban network. Vermeule (1960) argues that the Mycenaean diaspora was a response to maritime disorder in the thirteenth century B.C.: "Inability to import seems to have demanded emigration as an alternative." Yet it was trade links that turned these scattered settlements into a network that was strong enough to induce Athenians to fight when Persia menaced the Greek cities of the Ionian Coast. Athenian fighting strength was magnified because of the network that came together to form the anti-Persian Delian league. The Phoenicians built a parallel urban network that stretched from the Levant to the straits of Gibraltar.

These networks traded goods, but new technological ideas also moved along the network. Vermeule (1960) notes how iron, the Protogeometric pottery style and "new fashions in body armor and swords" are found in the areas "which maintained Mycenaean traditions the longest, and which lay open to contact with the East by sea." Even more momentously, the Phoenicians spread their alphabet, which became our alphabet, across their network of cities (Howard 2012).

The cities of Rome and China were also connected, but networks seem distinct from empires, because imperial metropolises are connected as much by political administration and the military as by trade. In 400 B.C., Zhuo Zhuan describes a pattern of hierarchical city sizes that has a resemblance to central place theory, but the city sizes are determined entirely by the city leader's place in the aristocratic hierarchy.

In the West, non-imperial networks resume when cities start growing again three centuries after the fall of Rome. In the medieval period, there is a strong trading network of Tuscan cities, including Pisa, Florence and Sienna, which are linked to each other and to the other trading powerhouses of Genoa and Venice, which was in turn closely linked across the Mediterranean to Constantinople. There was a second great urban network in the Low Countries, especially Bruges, Ghent, Leuven and Brussels. These two networks were also

linked by trading and finance, and in the fifteenth century, an artistic revolution was fed by links across these networks. The knowledge of detailed painting in oil moved from Flanders to Italy and the knowledge of linear perspective moved from Italy to Flanders (Skipnes 2007).

The Hanseatic League may be the ultimate example of an historic urban network. Beginning with a simple trading alliance between Hamburg and Lubeck, the League became the master-traders of the North. From Bruges to Novgorod, Hanseatic traders would carry wool, salted fish and knowledge. The core Hanseatic cities may have been part of the Holy Roman Empire, but their network was so strong that they were almost a nation unto themselves. They shared a legal system, the Lubeck Law, a language, Middle Low German, and fought successful wars against nations, like England and Denmark, that threatened to curtail their trading.

The Hanseatic network also helped spread major innovations such as the printing press. The printing press was invented in Mainz, but it naturally spread along the Rhine to great Hanseatic city of Cologne. William Caxton was an English merchant who led the Merchant Adventurers of London's outpost in the Hanseatic city of Bruges, visited Cologne and learned about the press. When he returned to Bruges, he created the first printed book in English and then returned and introduced the printing press to England. Caxton's return to England followed London's return to the Hanseatic urban fold, after an unsuccessful attempt by the English to defeat the Hanseatic League.

America's first urban network was tied by the Atlantic seaboard, which enabled goods to move far more cheaply than any land-based transportation mode. Boston, Providence, New Haven, New York, Philadelphia and Baltimore were linked by sailing vessels during the eighteenth century. People, like the young Benjamin Franklin, travelled from one city to the next, carrying knowledge with them. The trade links between New York City and the south made New York a hotbed of opposition to the Civil War. As Americans moved west, new networks emerged around the Great Lakes and the Mississippi River and then along the western seaboard.

Yet the later the area was urbanized, the more that development took the form of a single large agglomeration rather than a network of smaller cities.

Asia was later to urbanize and its urban networks are still emerging. As China has opened to the world, a network emerged in the south that includes Hong Kong, Guangzhou and Shenzhen. Singapore's network is really global, but it has particular ties to its proximate neighbors, Kuala Lumpur and Jakarta. In the Persian Gulf, Abu Dhabi, Doha and Dubai share close ties despite the political boundaries that separate them.

2.2 Urban Networks Today

No modern network of cities enjoys ties as strong as the Hansa, and modern urban networks do not wage wars. Those who would celebrate the strength of linked twenty-first century cities must still recognize the monopoly on force and political power wielded by the modern nation-state, which is diluted only in Europe by the European Union. Modern urban networks share connections, but they do not exercise legal authority and jealous national governments are unlikely to surrender their political power.

Perhaps more significantly, modern networks in the developed world are not usually focused on the transmission of goods. To make this point, Table 1 documents the ten strongest

"trade" linkages in America's metropolitan areas. We start with the 2012 Commodity Flow Survey, which is a survey of the physical shipments across the United States. We then calculate the total value of flows going into each metropolitan area and the share that comes from every other area in the sample. This creates a matrix with each entry ij being the share of area i's imports that come from area j. The table simply lists the ten largest entries in the matrix. If America's cities were split up into tight networks of goods-connected metropolitan areas, we would expect to see strong reciprocal relationships exchanging goods.

This list shows ten pairs of metropolitan areas with extremely strong one-way links. Every one of the destination metropolitan areas receives at least one fifth of its shipments, by value, from the origin metropolitan area. Somewhat remarkably, in four of those cases Seattle, Washington, is the origin metropolitan area. That fact reflects the enormously high value of the transportation equipment shipped from Seattle, and typically produced by Boeing.

Conversely, however, in nine of the ten cases less than 2.5 percent of shipments to the origin come from the destination area. With the sole exception of Portland, which does indeed originate a significant fraction of shipments to Seattle, all of the linkages are decidedly one way. This doesn't mean that American cities don't ship goods to another. They do. Yet the system is so well linked that it ceases to be a visible network and is instead just one national marketplace where trucks and trains carry goods from one area to the other (Glaeser and Kohlhase 2004).

We still think of Europe as the home of great urban networks, some of which lie within nations while some sprawl across old borders. Germany, the birthplace of the Hansa, has actively promoted urban networks since the 1990s (Mueller 2000). Northern Italy enjoys an urban network that stretches from Turin to Venice. The urban network in Brabant includes both the southern Netherlands and northern Belgium.

2.3 Motivating the Model

In the model that follows, we will be considering networks that allow the mobility of people and ideas, essentially expanding the urbanity of a place to a multitude of urban areas. We will assume that all areas can trade goods with one another. We define a city as an area with perfect mobility across firms and a free flow of ideas.

Between the trading system and the commuting zone lie the smaller groups of cities that represent what is typically meant by an urban network. These mid-level networks enable the flow of people and ideas, and often the two move together. In our model, we will assume that skilled workers can relocate easily within the network, either because the distances are short or because the cultural similarities are high. We are not suggesting that it is legally or physically impossible to relocate anywhere within the European Union or the United States, but it is easy for a Dutchman to move from Rotterdam to Amsterdam and still see his parents once a month (or once a week, if he has such an appetite). A move to Warsaw will be far more wrenching. Similarly, a move from Boston to New York is far less jarring than a move to Las Vegas.

In modern urban networks, as opposed to the Hanseatic League, there is no bright boundary at a network's edge. The difficulty of moving from Boston rises gradually with physical and cultural distance: there is no discontinuous jump at the Hudson River. We assume that

networks have clear borders solely for expositional ease. We also focus on the mobility of the skilled rather than the unskilled partially for expositional reasons, but also to reflect reality.

Inter-metropolitan mobility rates are typically much lower among those with less education. For example, Overman (2015) finds that 52 percent of British workers without formal qualifications work where they were born but only 31 percent of workers with degrees have stayed put. Only 23 percent of Americans with college degrees live in the town where they were born, while 43 percent of those with a high school degree or less have never moved (Cohn and Morin 2008).

Our second assumption is that the network is also marked by the movement of ideas. Following Jacobs (1969) and Weitzman (1998), we assume that new ideas are formed from old ideas. A stock of successful entrepreneurs begets the next generation of entrepreneurs. We assume that knowledge moves more readily over short distances, not because we doubt the efficacy of the internet for transferring basic facts, but because there are still abundant examples of important intellectual exchanges the occur in crowded corridors. It is somewhat ironic that the ideas that created Facebook, which thrives by facilitating electronic friendships, seem to have been formed by face-to-face conversations in Cambridge, Massachusetts. We assume that entrepreneurs draw particularly on the ideas in their own metropolitan area, but can also benefit from the ideas in their larger urban network.

We will use our model to ask both positive questions, about the determinants of city size and income growth, and normative questions about the optimal network size. The model will yield predictions about how the fortunes of cities and ordinary workers change when they are part of a network, and how they change based on the composition of the network.

We focus on expanding the size of urban networks, but not changing the extent of global trade, which has been well studied. Typically easier trade enhances welfare, but the positive effects of increased trade are hardly universal. As in Jefferson (1939), a smaller city that is included into a network with a more productive neighbor could easily lose both business growth and population. The skilled workers in a smaller network might lose out if their network becomes linked to a larger network with a far higher ratio of skilled to unskilled workers. In our model, these forces do not operate at the level of the trading network because trade is not driven by difference in factor proportions, à la Heckscher-Ohlin, but rather by gains from specialization that every location can reap when it has access to a larger market.

We focus above all on the trade-off between network formation and urban growth, either through densification or consolidation. The network is an alternative to a single large city, but it is not exactly the same. The network may have more land available and particular local advantages that show up when there are decreasing returns to scale in new idea formation. Big cities can also have the advantage of more amenities. We now turn to the model.

3 Setup of the Model

We first describe the core assumptions of the model about production, new firm creation, amenities and housing. We then characterize the equilibrium.

3.1 Production

There is a continuum of cities of measure C. Each city c is endowed with unskilled labor L_c and land T_c. Neither of these factors of production are mobile. City c will also host skilled labor $H_{c,t}$ and a measure $N_{c,t}$ of firms in period t. The quantity of skilled labor and the number of firms are endogenously determined equilibrium outcomes.

Firms make differentiated products that can be costlessly traded and then aggregated into a composite good with the constant elasticity of substitution (CES) specification

$$Y_t = \left[\int_0^C \int_0^{N_{c,t}} \left(y_{c,t}^j \right)^{\frac{\sigma-1}{\sigma}} dj dc \right]^{\frac{\sigma}{\sigma-1}} \quad \text{for } \sigma > 1. \tag{1}$$

The composite good serves as our numeraire so its price is normalized to one.

Each variety is manufactured using unskilled labor according to a linear production function: $y_{c,t}^j = A_{c,t} l_{c,t}^j$, where y_c^j is the output of firm j in city c in period t, $l_{c,t}^j$ its unskilled workforce, and $A_{c,t}$ the productivity common to all firms in the city.

In equilibrium, city output is

$$Y_{c,t} = \int_0^{N_{c,t}} p_{c,t}^j y_{c,t}^j dj = \left[Y_t \left(A_{c,t} L_c \right)^{\sigma-1} N_{c,t} \right]^{\frac{1}{\sigma}}. \tag{2}$$

Wages for the unskilled equal the marginal product of labor, which is proportional to output per worker:

$$w_{c,t} = \frac{\sigma - 1}{\sigma} \frac{Y_{c,t}}{L_c} \tag{3}$$

Firm profits, i.e., revenues minus labor costs, are identical across firms in the same place and time and are proportional to output per firm. Aggregate output of the composite good integrates output from all firms in all the separate cities:

$$Y_t = \left\{ \int_0^C \left[\left(A_{c,t} L_c \right)^{\sigma-1} N_{c,t} \right]^{\frac{1}{\sigma}} dc \right\}^{\frac{\sigma}{\sigma-1}}. \tag{4}$$

3.2 Entrepreneurship

The cities are partitioned into distinct urban networks, which we refer to as "regions," $r = 1, 2, ..., R$. We order cities so that region r includes cities $c \in [\bar{c}_{r-1}, \bar{c}_r]$ where $\bar{c}_0 = 0$ and $\bar{c}_R = C$. Each region includes a positive measure $C_r = \bar{c}_r - \bar{c}_{r-1} > 0$ of cities.

Region r is endowed with H_r skilled agents. While the unskilled are immobile, the skilled can move within regions, but not across regions. The mobility within regions is meant to capture the cultural connections within the region and the ease of connecting with friends and family in one's place of origin. As we typically assume that regions are identical, the assumption of immobility across regions is largely irrelevant.

These skilled agents are the entrepreneurs in our economy. Firms are created in an entrepreneurial culture through imitation of successful businesses and intellectual spillovers across entrepreneurs within urban networks. Our model of interactive idea creation follows Weitzman's (1998) theory of recombinant growth with four stages.

First, "business concepts" are created by skilled entrepreneurs in a city by looking at existing business models. Second, these concepts are developed into primary ideas by consulting with skilled workers throughout the region. Third, these primary ideas are turned into matched ideas by connecting with other primary ideas in the city. Fourth, these matched ideas are turned into business plans by consulting with other skilled people in the city. The ultimate equations that drive our model do not depend on this particular process, but it does justify why new firm creation depends on skills within the city, skills within the region and the stock of old firms in the city. This four-stage structure is inspired by Weitzman (1998), but we are responsible for assuming that external regional skill matters more at an earlier stage and external local skill matters more at the later stage.

The $N_{c,t-1}$ varieties produced in city c in period $t-1$ and the $H_{c,t}$ skilled entrepreneurs combine to produce $N_{c,t-1}H_{c,t}$ business concepts.

These concepts are developed into primary ideas through regionwide consultation. The productivity of each seed depends on the attention it receives from high-skill agents in the entire regional network, and so the stock of primary ideas is

$$\tilde{N}_{c,t} = n_{c,t} N_{c,t-1} H_{c,t} \left(\frac{H_r}{N_{c,t-1} H_{c,t}} \right)^{\delta} \text{ for } \delta \in [0,1] \tag{5}$$

where $n_{c,t}$ is a productivity shock.

Capturing the assumption of recombinant growth, successful business ventures emerge from the meeting of two entrepreneurs with primary ideas that can be successfully combined. The number of such matches in city c, which we refer to as secondary ideas, is $\tilde{N}_{c,t}^2/2$.

The development of secondary ideas into final business plans depends on the attention given by entrepreneurs in the city, so the total number of varieties that can be successfully produced in city c in period t is

$$N_{c,t} = \tilde{n}_{c,t} \frac{\tilde{N}_{c,t}^2}{2} \left(\frac{H_{c,t}}{\tilde{N}_{c,t}^2/2} \right)^{\tilde{\delta}} \text{ for } \tilde{\delta} \in [0,1], \tag{6}$$

where $\tilde{n}_{c,t}$ is a productivity shock.

We can rewrite this equation more compactly

$$N_{c,t} = \alpha_{c,t} N_{c,t-1}^{\mu} H_{c,t}^{\eta} H_r^{\nu} \tag{7}$$

for an aggregate productivity in firm creation $\alpha_{c,t} \equiv \left(n_{c,t}^2/2\right)^{1-\tilde{\delta}} \tilde{n}_{c,t}$ and elasticities $\mu \equiv 2(1-\delta)\left(1-\tilde{\delta}\right) \in [0,2]$, $\eta \equiv \mu + \tilde{\delta} \in [0,2]$ and $\nu \equiv 2\delta\left(1-\tilde{\delta}\right) \in [0,2]$ such that $\mu+\eta+\nu = 1 + (3-2\delta)\left(1-\tilde{\delta}\right) \in [1,4]$.

Firm profits accrue to the skilled agents who create and manage the firms. Spillovers from pre-existing firms and skilled entrepreneurs in the rest of the regions are pure positive externalities. If $H_{c,t}$ is the equilibrium measure of skilled labor in city c and period t, then each entrepreneur's nominal income is proportional to output per entrepreneur:

$$e_{c,t} = \frac{1}{\sigma} \frac{Y_{c,t}}{H_{c,t}}. \tag{8}$$

3.3 Consumption and Endogenous Amenities

Consumption is a Cobb-Douglas aggregate of housing $h^i_{c,t}$, direct consumption of the numeraire $q^i_{c,t}$, and consumption of amenity-generating goods $g^i_{c,t}$. Man-made amenities yield spillovers across people and over time. The most natural example of such amenities would be the aesthetics of one's residence. We can also think of other amenities that are potentially produced collectively, such as a community club, as long as the club's membership is finite and hence has measure zero.

Resident i of city c at time t has utility

$$u^i_{c,t} = \left(\frac{h^i_{c,t}}{\tau}\right)^{\tau} \left(\frac{q^i_{c,t}}{1-\gamma_i-\tau}\right)^{1-\gamma_i-\tau} \left(\frac{g^i_{c,t} G_{c,t}}{\gamma_i}\right)^{\gamma_i}, \tag{9}$$

where interpersonal and intertemporal spillovers are captured by

$$G_{c,t} = \Gamma_{c,t} G^{1-\phi}_{c,t-1} \left(\int_{\mathcal{I}_{c,t}} g^i_{c,t} di\right)^{\phi} \text{ for } \phi \in (0,1) \tag{10}$$

when each city resident $i \in \mathcal{I}_{c,t}$ spends $g^i_{c,t}$ units of the numeraire on amenities.

In keeping with standard income elasticities of housing consumption (Glaeser, Kahn and Rappaport 2008), all agents have the same budget share for housing τ. We do, however, assume that the budget share of amenities γ_i is higher for skilled than unskilled individuals: $\gamma_H > \gamma_L$. The simplest justification for the assumption that the skilled put more weight on amenities is that education teaches them to appreciate the better things in life. Alternatively, the connection between skills and taste for amenities can be seen as an approximation for assuming that amenities are luxury goods.

The city-wide expenditure on amenities is $\int_{\mathcal{I}_{c,t}} g^i_{c,t} di = \bar{\gamma} Y_{c,t}$, where $\bar{\gamma} = \gamma_H/\sigma + \gamma_L (\sigma - 1)/\sigma$ is the average budget share of amenities, which in turn reflects the equilibrium share of total income going to skilled and unskilled individuals. Amenities therefore equal

$$G_{c,t} = \Gamma_{c,t} G^{1-\phi}_{c,t-1} \left(\bar{\gamma} Y_{c,t}\right)^{\phi}. \tag{11}$$

Amenities will generally be higher in more skilled cities for two reasons in this model. First, the skilled will choose to select into higher amenity cities. Second, the skilled will invest more in local amenities. We believe that both of these forces are also at work in the real world.

3.4 Housing

Housing is built with a Cobb-Douglas production function using land and the numeraire:

$$h^j_{c,t} = \Psi_{c,t} \left(\frac{z^j_{c,t}}{1-\zeta}\right)^{1-\zeta} \left(t^j_{c,t}\right)^{\zeta} \text{ for } \zeta \in (0,1), \tag{12}$$

where $\Psi_{c,t}$ is the productivity of the construction sector, $z^j_{c,t}$ is firm j's input of the numeraire, and $t^j_{c,t}$ the land acquired by the firm from absentee landlords.[1] Somewhat counter-factually,

[1] Our results would be qualitatively unchanged if we assumed instead that landowners consume local housing and local amenities.

for the sake of tractability we assume that housing lasts only one period. Even if we think of a period as a generation, this assumption still surely underestimates the durability of housing.

Demand for housing is $h_{c,t} = \tau Y_{c,t}/r_{c,t}$, which is a function of city output $Y_{c,t}$ and house prices $r_{c,t}$.

The construction industry is competitive. The housing market is in equilibrium when construction firms earn zero profits because housing prices equal construction costs, given land rents, and when all land is used for construction. In equilibrium, the housing supplied will equal

$$h_{c,t} = \Psi_{c,t} T_c^{\zeta} \left(\tau Y_{c,t}\right)^{1-\zeta}, \tag{13}$$

with housing prices

$$r_{c,t} = \frac{1}{\Psi_{c,t}} \left(\tau \frac{Y_{c,t}}{T_c}\right)^{\zeta} \tag{14}$$

and land rents

$$\rho_{c,t} = \zeta \tau \frac{Y_{c,t}}{T_c}. \tag{15}$$

3.5 Spatial Equilibrium

Given their equilibrium incomes $w_{c,t}$ and $e_{c,t}$, amenities $G_{c,t}$, and house prices $r_{c,t}$ as a function of city output $Y_{c,t}$, the utility of unskilled workers in city c at time t is

$$u_{c,t}^{L} = \frac{w_{c,t} G_{c,t}^{\gamma_L}}{r_{c,t}^{\tau}} = \frac{\sigma - 1}{\sigma} \left[\Psi_{c,t} \left(\frac{T_c}{\tau}\right)^{\zeta}\right]^{\tau} \left(\bar{\gamma}^{\phi} \Gamma_{c,t} G_{c,t-1}^{1-\phi}\right)^{\gamma_L} \frac{Y_{c,t}^{1-\zeta\tau+\phi\gamma_L}}{L_c}, \tag{16}$$

while the utility of skilled entrepreneurs is

$$u_{c,t}^{H} = \frac{e_{c,t} G_{c,t}^{\gamma_H}}{r_{c,t}^{\tau}} = \frac{1}{\sigma} \left[\Psi_{c,t} \left(\frac{T_c}{\tau}\right)^{\zeta}\right]^{\tau} \left(\bar{\gamma}^{\phi} \Gamma_{c,t} G_{c,t-1}^{1-\phi}\right)^{\gamma_H} \frac{Y_{c,t}^{1-\zeta\tau+\phi\gamma_H}}{H_{c,t}}. \tag{17}$$

We assume that productivity shocks α_c, A_c, $\Psi_{c,t}$ and Γ_c are realized before skilled workers choose their location, so there is no residual uncertainty. The free mobility of the skilled implies that skilled workers must receive the same utility in all cities within a region where $H_{c,t} > 0$. For all $c \in r$, $u_{c,t}^{H} \leq u_{r,t}^{H}$ and $H_{c,t} \geq 0$ with complementary slackness.

The spatial equilibrium is described by the distribution of skilled agents across cities $H_{c,t}$ that satisfies this indifference condition as well as the labor market clearing condition $\int_{\bar{c}_{r-1}}^{\bar{c}_r} H_{c,t} dc = H_r$.

When skilled agents choose their location for period t, each city c is characterized by eight pre-determined characteristics: its time-invariant endowment of unskilled labor L_c and land T_c; its stock of pre-existing firms $N_{c,t-1}$ and amenities $G_{c,t-1}$; and its productivity in entrepreneurship $\alpha_{c,t}$, production $A_{c,t}$, amenities $\Gamma_{c,t}$ and construction $\Psi_{c,t}$.

Given the joint distribution of these characteristics across cities, the equilibrium sorting of skilled agents $H_{c,t}$ yields a stock of firms described by equation (7). The number of firms determines city output by equation (2) and output in turn determines the city's wages (eq. 3), housing stock (eq. 13), house prices (eq. 14) and land rents (eq. 15). We refer to

these outcomes, along with profits per entrepreneur, as the pecuniary outcomes. Output also determines amenities (eq. 11) and unskilled workers' utility (eq. 16). The profits and the utility of each skilled entrepreneur are determined both by output and directly by the number of his peers and competitors $H_{c,t}$ (eq. 17). Solving this system of equations implies that the skilled entrepreneurs' utility in city c in region r is

$$u_{c,t}^{H} = \frac{1}{\sigma}\left[\alpha_{c,t} N_{c,t-1}^{\mu} H_r^{\nu} Y_t \left(A_{c,t} L_c\right)^{\sigma-1}\right]^{\frac{1-\zeta\tau+\phi\gamma_H}{\sigma}} \left[\Psi_{c,t}\left(\frac{T_c}{\tau}\right)^{\zeta}\right]^{\tau} \left(\Gamma_{c,t}\bar{\gamma}^{\phi} G_{c,t-1}^{1-\phi}\right)^{\gamma_H} \times H_{c,t}^{\eta\frac{1-\zeta\tau+\phi\gamma_H}{\sigma}-1}. \quad (18)$$

We adopt the standard heuristic notion of stability of the spatial equilibrium: an equilibrium is stable if transferring a marginal amount of skilled agents from one city to another makes the sending city better off and the receiving city worse off. Stability in this sense requires that the centrifugal forces arising from the scarcity of land and unskilled labor are sufficient to offset the centripetal forces resulting from spillovers in entrepreneurship and man-made amenities.

Assumption 1 *The budget share of housing (τ) is sufficiently high, while the housing supply elasticity ($1/\zeta - 1$), the appeal of amenities to the skilled (γ_H), within-city spillovers in the creation of firms (η) and amenities (ϕ) and productivity gains from firm creation ($\sigma/(\sigma-1)$) are sufficiently low that a stable equilibrium exists:*

$$\sigma > \eta\left(1 + \phi\gamma_H - \zeta\tau\right).$$

When this condition is satisfied, there is a unique stable equilibrium with a positive number of skilled agents and firms in every city. Equilibrium sorting in region r is given by the equilibrium number of skilled people in each city c:

$$H_{c,t} = \frac{H_r\left\{\left[\alpha_{c,t} N_{c,t-1}^{\mu}\left(A_{c,t}L_c\right)^{\sigma-1}\right]^{\frac{1-\zeta\tau+\phi\gamma_H}{\sigma}}\left(\Psi_{c,t}T_c^{\zeta}\right)^{\tau}\left(\Gamma_{c,t}G_{c,t-1}^{1-\phi}\right)^{\gamma_H}\right\}^{\frac{\sigma}{\sigma-\eta(1-\zeta\tau+\phi\gamma_H)}}}{\int_{\bar{c}_{r-1}}^{\bar{c}_r}\left\{\left[\alpha_{s,t} N_{s,t-1}^{\mu}\left(A_{s,t}L_s\right)^{\sigma-1}\right]^{\frac{1-\zeta\tau+\phi\gamma_H}{\sigma}}\left(\Psi_{s,t}T_s^{\zeta}\right)^{\tau}\left(\Gamma_{s,t}G_{s,t-1}^{1-\phi}\right)^{\gamma_H}\right\}^{\frac{\sigma}{\sigma-\eta(1-\zeta\tau+\phi\gamma_H)}} ds}, \quad (19)$$

The numerator is a function solely of a multiplicative combination of parameters that determine the utility level of the skilled, holding the number of skilled entrepreneurs constant. The denominator simply averages that combination across cities in the region. A city attracts more than the average amount of talent (H_r/C_r) if and only if it has more than the average appeal. Factors that increase utility of the skilled holding the number of skilled residents constant, i.e., the stock of old firms, productivity, unskilled labor, land and past amenities, also increase the number of skilled entrepreneurs that come to the city.

4 Urban Networks with Spatial Symmetry

We begin by deriving closed-form solutions for the fully symmetric case. The economy consists of R equal-sized regions, each comprising measure C_r of cities. All cities and thus

all regions have the same factor endowments and common productivity shocks. If the initial endowments of firms and amenities are also identical, then in the dynamic equilibrium all cities (and thus all regions) are always identical, with equal shares of skilled entrepreneurs.

Then the aggregate number of firms is

$$N_t = C_r R N_{c,t} = \alpha_t C_r^{1-\mu-\eta} R^{1-\mu-\eta-\nu} H^{\eta+\nu} N_{t-1}^{\mu}, \tag{20}$$

so aggregate output is

$$Y_t = C_r R Y_{c,t} = A_t L N_t^{\frac{1}{\sigma-1}}. \tag{21}$$

The basic equations of our model then apply directly, without city-specific indices, to all pecuniary variables. Eq. (3) yields wages and eq. (8) profits per entrepreneur; eq. (13) the aggregate housing stock ($h_t = C_r R h_{c,t}$), eq. (14) house prices and eq. (15) land rents; eq. (11) the aggregate amount of amenities ($G_t = C_r R G_{c,t}$). As a result, unskilled workers have utility

$$u_t^L = \frac{\sigma-1}{\sigma}\frac{1}{L}\left[\Psi_t\left(\frac{T}{\tau}\right)^{\zeta}\right]^{\tau}\left(\frac{\Gamma_t \bar{\gamma}^{\phi} G_{t-1}^{1-\phi}}{C_r R}\right)^{\gamma_L} Y_t^{1-\zeta\tau+\phi\gamma_L} \tag{22}$$

and skilled entrepreneurs have utility

$$u_t^H = \frac{1}{\sigma}\frac{1}{H}\left[\Psi_t\left(\frac{T}{\tau}\right)^{\zeta}\right]^{\tau}\left(\frac{\Gamma_t \bar{\gamma}^{\phi} G_{t-1}^{1-\phi}}{C_r R}\right)^{\gamma_H} Y_t^{1-\zeta\tau+\phi\gamma_H}. \tag{23}$$

The spatial structure of the economy affects welfare through two channels. First, amenities generate local externalities within a commuting zone. As a consequence, the smaller and more fragmented such commuting zones, the less access to man-made amenities each resident enjoys. The importance of this effect is measured by the share of amenity-generating consumption in an individual's budget (γ_i). Second, human-capital spillovers in innovation at the region and the city level determine incomes. Welfare is naturally increasing with incomes. The relationship between income and welfare is stronger when the output share that is absorbed by the construction sector ($\zeta\tau$) is smaller and when the higher the amenity spillovers from consumption ($\phi\gamma_H$) are larger.

The number of firms, and therefore total output, increases with the number of cities if and only if $\mu+\eta$ is less than one. This condition determines whether the local ingredients in new firm creation (human capital and the stock of old ideas) display increasing or decreasing returns. When $\mu+\eta$ is greater than one, then there are local increasing returns and more division leads to less creativity. When $\mu+\eta$ is less than one, then there are locally decreasing returns, and fragmentation increases creativity.

The following proposition formalizes these points.

Proposition 1 *Suppose the economy consists of identical cities divided into identical regions.*

Fragmentation of the economy into separate idea networks reduces firm creation ($\partial N_t/\partial R < 0$), output ($\partial Y_t/\partial R < 0$), wages ($\partial w_t/\partial R < 0$), profits per entrepreneur ($\partial e_t/\partial R < 0$), the housing stock ($\partial h_t/\partial R < 0$), house prices ($\partial r_t/\partial R < 0$), land rents ($\partial \rho_t/\partial R < 0$), and everyone's utility ($\partial u_t^i/\partial R < 0$ for $i \in \{H, L\}$).

Fragmentation of each idea network into separate commuting zones reduces all pecuniary variables ($\partial x_t/\partial C_r < 0$ for $x \in \{N, Y, w, e, h, r, \rho\}$) if and only if $\mu + \eta > 1$. It reduces individual i's utility ($\partial u_t^i/\partial C_r < 0$) if and only if

$$\gamma_i + (1 - \zeta\tau + \phi\gamma_i)\frac{\mu + \eta - 1}{\sigma - 1} > 0.$$

The first part of the proposition merely makes the point that idea networks are unambiguously good. There is no downside in a symmetric world to increasing the flow of ideas. More new firms are created which means higher incomes, higher amenities and higher welfare levels. Our model of recombinant growth implies that creativity comes from connection, and so network connections are good (formally, $\mu + \eta + \nu > 1$).

The second part of the proposition notes that the once networks are in place, the advantages of expanding commuting zones are more ambiguous, and depend on whether $\mu + \eta > 1$. If firm creation is primarily determined by local variables, then agglomeration into larger commuter zones increases new firm creation and wealth by allowing fruitful matches between existing firms and young talent within a single commuting zone. The proposition shows that all of the pecuniary variables, including incomes and housing costs, rise or fall along with new firm creation.

Conversely, if region-level spillovers are high enough relative to purely local spillovers ($1 - \nu < \mu + \eta < 1$), then fragmentation can be beneficial. Firm creation is liveliest if entrepreneurs are separated in different cities with different local cultures that can profit from interactions within the idea network but develop distinctive ideas that would be lost in the homogeneity of a single commuting zone.

Empirically, there is evidence suggesting local increasing returns in creativity. Bettencourt, Lobo and Strumsky (2007) show that the logarithm of patenting increases more than one-for-one with the logarithm of metropolitan area population, because there are more inventors in larger cities. Glaeser (2007) finds that the self-employment rate is higher in big cities. Glaeser and Kerr (2009) regress the log of employment in new establishments within an industrial cluster of a city on the log of population in that city and that cluster. The sum of the coefficients is greater than one, which also suggests local increasing returns. If this evidence actually means that $\mu + \eta > 1$, then the model suggests the even within urban networks, further integration into common commuting zones will speed new firm creation and increase average incomes.

If greater integration does correspond to greater creativity, then it surely increases welfare as well. However, the final part of Proposition 1 reminds us that the converse is not necessarily true, because efficiency in production does not coincide with optimality. Amenities also matter. The residents of a more connected commuting zone get to enjoy the benefits of all the amenities within the area. Anyone living in greater Paris can visit the Louvre on the weekend. An additional cost of fragmentation is that individuals are prevented from enjoying man-made amenities that spillover within a city, but not outside the commuting range.

This additional source of local spillovers unambiguously favors larger commuting zones. It implies that welfare may decline with fragmentation even as output rises. This ranking reversal is especially likely when amenities are more important (higher γ_H); when greater firm creation raises income less (higher σ); when housing has a higher income share (τ) and

more rigid supply (higher ζ), so income gains accrue less to consumers and more to the housing sector; and when amenities are more persistent (lower ϕ), so they react less to a decline in income.

A final implication of our result is that high- and low-skill workers may have opposite preferences over the separation of commuting zones, because they have different tastes for amenities. High-skill workers may benefit from larger commuting zones, even if the reduce incomes, because amenities also go up. They are willing to forgo income to live in a consumer city. Low-skill workers, who care relatively more about their wages, however, may oppose amenity-enhancing integration if local spillovers in entrepreneurship are low enough that integration lowers incomes.

Proposition 1 presents the consequences of improvements in information technology and transportation technology that make idea exchange and commuting possible over greater distances. We now ask a different question: when or whether it would be preferable to concentrate people, firms and amenities in a smaller, denser city relative to leaving them in a more far-flung ideas network. This is exactly the public policy question that countries like China face when they are deciding between expanding a single mega-city, like Shanghai, or building a second urban center that will be connected to the mega-city by tight transportation and communication linkages. In a European context, where networks are already the norm, the question is whether to replace the network with a single large city.

This question is slightly different from the question of simply expanding a commuting zone, discussed above, because we assume that there are congestion costs from expanding the city. In our model, the single city will have less land than the network and that will mean higher housing prices. In the real world, the larger city could indeed sprawl further, which would ease pressure on housing prices, but there would also be costs from congestion in transportation, which we do not model.

Proposition 2 *Suppose the economy consists of identical cities divided into identical regions. Concentrating people, firms and man-made amenities within an idea-sharing region into denser commuting zones with less land (such that $\partial \ln T / \partial \ln C_r = 1$) increases firm creation, output, wages and profits per entrepreneur ($dx_t/dC_r|_{T \propto C_r} < 0$ for $x \in \{N, Y, w, e\}$) if and only if $\mu + \eta > 1$. It increases individual i's utility ($du_t^i/dC_r|_{T \propto C_r} < 0$) if and only if*

$$\gamma_i - \zeta\tau + (1 - \zeta\tau + \phi\gamma_i)\frac{\mu + \eta - 1}{\sigma - 1} > 0.$$

The proposition's requirement that $\partial \ln T / \partial \ln C_r = 1$ simply means that the land availability scales down with the number of cities. So as the population as consolidated in ten percent fewer commuting zones, it must adapt itself to live on ten percent less land.

The proposition shows that the condition for generating positive changes in income and firm creation from consolidation into fewer cities with less land is the same as the condition for generating positive changes in income and firm creation from consolidation into fewer cities with the same amount of land, given in Proposition 1. This should not be surprising, for land plays no role in our production process. More density might be unpleasant, but it will not reduce productivity following the assumptions of our model.

The condition for improving utility, however, is different and more stringent when density is increased as a result of consolidation. Improvements in commuting technology raise welfare

if, but not only if, they raise output. Instead, increases in density may raise output yet lower welfare, just as they still can lower output but increase welfare. Intuitively, changes in density have not one but two effects independent of income. On the positive side, they still enable greater access to amenities. On the negative side, however, they reduce the availability of land and thereby drive up house prices. This downside is greater the larger the budget share of housing (τ) and the lower the housing supply elasticity (the higher ζ).

Accordingly, the larger the value of $\zeta\tau$ the more likely it becomes that density decreases welfare even though it raises output. This ranking reversal is especially likely when greater firm creation raises income less (higher σ) and when amenities are less important (low γ_H) and more persistent (lower ϕ), so they react less to an increase in income.

This result suggests that urban networks might be optimal in the Netherlands, while bigger cities might be optimal in Texas. The Netherlands has significant land use controls which limit the elasticity of housing supply. Texas does not. Consequently, it is relatively easy for Houston to expand its housing stock, even within a given geographic footprint, which is much harder in Amsterdam. According to this view, the reliance on networks in Europe is a natural means of adapting to the constraints that history and geography place on the growth of that continent's older cities.

Incidentally, it is not automatic that consolidation will lead to higher housing costs or less housing production. It is possible that consolidation will lower productivity by so much that housing costs and land rents will drop ($\partial \ln r_t / \partial \ln C_r + \partial \ln r_t / \partial \ln T > 0$ if and only if $\mu + \eta + \sigma < 2$). Conversely, if incomes increase, not only do house prices increase: if the rise in output and the housing supply elasticity are high enough, housing production may also rise ($\partial \ln h_t / \partial \ln C_r + \partial \ln h_t / \partial \ln T < 0$ if and only if the supply elasticity is $1/\zeta - 1 > (\sigma - 1) / (\mu + \eta - 1) > 0$).

We can also consider the impact of consolidating the population into fewer regions with less land. In this case, we assume that land per city is constant (so $T = C_r R T_c$ for given T_c), but that the size of the regions shrink as the number of regions shrink. Conversely, in this experiment a decline in density is equivalent to considering the creation of a new region, not linked to the old, but with new land. Perhaps the most natural interpretation of these results is that they capture the potential benefits of expanding the frontier of an economy into previously empty space that is distant from the original populations.

Corollary 1 *Suppose the economy consists of identical cities divided into identical regions. Concentrating people, firms and man-made amenities into denser regions with greater idea sharing but less land ($\partial \ln T / \partial \ln R = 1$) increases firm creation, output, wages and profits per entrepreneur, house prices and land rents ($dx_t/dR|_{T \propto R} < 0$ for $x \in \{N, Y, w, e, r, \rho\}$) . It increases individual i's utility ($du_t^i/dR|_{T \propto R} < 0$) if and only if*

$$\gamma_i - \zeta\tau + (1 - \zeta\tau + \phi\gamma_i) \frac{\mu + \eta + \nu - 1}{\sigma - 1} > 0.$$

As in Proposition 1, more populous regions always lead to more productivity and higher incomes. As in Proposition 2, increasing density has costs that technological progress does not. While the effects on firm creation, incomes and amenities are unchanged, increasing density now pushes house prices up, and this can lead to a reduction in utility. The condition

for consolidation at the region level weighs the benefits of greater connections against the loss of physical space.

Again, whether this loss translates in a reduction in the supply of housing depends on the elasticity of housing supply. The increase in income could be large enough that a denser region builds up so much that the housing stock increases (if the elasticity is $1/\zeta - 1 > (\sigma - 1)/(\mu + \eta + \nu - 1)$). More generally, density is again more likely to be desirable if the supply of housing is elastic or if housing is a small share in total consumption. If there are local increasing returns to new firm creation, then denser regions create a horse race between two positive forces and a negative one: density raises incomes and amenities, but it makes housing scarcer.

In the case of denser commuting zones or denser regions, the skilled and the unskilled could have opposite preferences. The skilled would typically rather have greater density, more entrepreneurship and more amenities at the cost of smaller and more expensive houses. The unskilled would prefer fewer amenities but cheaper real estate in a less dense economy with lower idea spillovers.

4.1 Urban Networks in the Long Run

We now turn to the dynamics of this system of cities. Again, our focus will be on the connection over cities into networks of ideas and larger commuting zones. With complete symmetry, the growth path of the economy is characterized exactly by a log-linear first-order vector autoregression for the number for firms and the amount of amenities ($\ln N_t$ and $\ln G_t$).

If $\mu < 1$, this stochastic process is stationary. Denote the steady-state value of each variable by $\bar{x} = \exp(\mathbb{E} \ln x_t)$. In the steady state, the number of firms is

$$\bar{N} = \left(\bar{\alpha} C_r^{1-\eta-\mu} R^{1-\eta-\mu-\nu} H^{\eta+\nu}\right)^{\frac{1}{1-\mu}}, \tag{24}$$

so aggregate output and the aggregate amount of amenities are respectively

$$\bar{Y} = \bar{A} L \bar{N}^{\frac{1}{\sigma-1}} \text{ and } \bar{G} = \bar{\gamma} \bar{\Gamma}^{\frac{1}{\phi}} \bar{Y}. \tag{25}$$

All other pecuniary variables follow from our basic equations: wages (eq. 3) and profits per entrepreneur (eq. 8); the aggregate housing stock (eq. 13), house prices (eq. 14) and land rents (eq. 15). All these variables are monotonically increasing in the number of firms. Naturally, they depend directly on other factors as well, but they do not depend directly on the number of cities or regions.

Aggregate amenities depend only on aggregate output and a productivity shock. However, the effective enjoyment of amenities decreases directly with the number of commuting zones, because amenity spillovers operate through a commuting zone. As a result, in the steady state unskilled workers have utility

$$\bar{u}^L = \frac{\sigma - 1}{\sigma} \frac{1}{L} \left[\bar{\Psi} \left(\frac{T}{\tau}\right)^{\zeta}\right]^{\tau} \left(\frac{\bar{\gamma} \bar{\Gamma}^{\frac{1}{\phi}}}{C_r R}\right)^{\gamma_L} \bar{Y}^{1-\zeta\tau+\gamma_L} \tag{26}$$

and skilled entrepreneurs have utility

$$\bar{u}^H = \frac{1}{\sigma} \frac{1}{H} \left[\bar{\Psi} \left(\frac{T}{\tau}\right)^{\zeta}\right]^{\tau} \left(\frac{\bar{\gamma} \bar{\Gamma}^{\frac{1}{\phi}}}{C_r R}\right)^{\gamma_H} \bar{Y}^{1-\zeta\tau+\gamma_H}. \tag{27}$$

We can now state our first proposition about urban dynamics.

Proposition 3 *If the economy consists of identical cities divided into identical regions, its growth path converges to a steady state if and only if* $\mu < 1$.

Consolidation into larger idea networks has larger positive effects on all outcomes in the steady state than in the short run ($\partial \bar{x}/\partial R < \partial x_t/\partial R < 0$ *for* $x \in \left\{N, Y, w, e, h, r, \rho, u^L, u^H\right\}$). *If* $\mu+\eta > 1$, *so does consolidation of each idea network into larger commuting zones* ($\mu+\eta > 1 \Rightarrow \partial \bar{x}/\partial C_r < \partial x_t/\partial C_r < 0$ *for* $x \in \left\{N, Y, w, e, h, r, \rho, u^L, u^H\right\}$).

If instead $\mu + \eta < 1$, *consolidation of each idea network into larger commuting zones has larger negative effects on all pecuniary variables in the steady state than in the short run* ($\mu + \eta < 1 \Rightarrow \partial \bar{x}/\partial C_r > \partial x_t/\partial C_r > 0$ *for* $x \in \{N, Y, w, e, h, r, \rho\}$). *Then there is a non-empty range of gains from variety such that consolidation into larger commuting zones increases individual utility in the short run and decreases it in the steady state* ($\partial u_t^i/\partial C_r < 0 < \partial \bar{u}^i/\partial C_r \Leftrightarrow \sigma \in (\underline{\sigma}_s^i, \bar{\sigma}_s^i)$). *When gains from variety are higher, consolidation into larger commuting zones decreases individual utility in the steady state more than it does in the short run* ($\partial \bar{u}^i/\partial C_r > \partial u_t^i/\partial C_r > 0 \Leftrightarrow \sigma < \underline{\sigma}_s^i$). *When gains from variety are lower, consolidation into larger commuting zones increases individual utility in the steady state less than it does in the short run* ($\partial u_t^i/\partial C_r < \partial \bar{u}^i/\partial C_r < 0 \Leftrightarrow \sigma > \bar{\sigma}_l^i$). *The range of gains from variety that generate such a preference reversal expands when amenities have greater persistence* ($\partial\left(\bar{\sigma}_s^i - \underline{\sigma}_s^i\right)/\partial\phi < 0$), *when housing has greater supply elasticity* ($\partial\left(\bar{\sigma}_s^i - \underline{\sigma}_s^i\right)/\partial\zeta < 0$) *and a lower budget share* ($\partial\left(\bar{\sigma}_s^i - \underline{\sigma}_s^i\right)/\partial\tau < 0$), *and when human capital generates lower local spillovers* ($\partial\left(\bar{\sigma}_s^i - \underline{\sigma}_s^i\right)/\partial\eta < 0$).

An economy of identical firms in identical regions will converge to a steady state as long as $\mu < 1$. Otherwise increasing returns are so strong that the growth is perpetual. Larger idea networks are always good for firm creation and welfare in both the short run and in steady state. The effects are magnified in the steady state, relative to the short run, because changes in firm creation accumulate over time. The changes in the pecuniary variables build on accumulating new firm creation and gradually grow to their steady-state size.

This result may be significant when considering the evaluation of investments that expand urban networks. The proposition suggests that the full effects of these investments will appear only after a long period of time. This suggests that cost-benefit analysis of network formation needs to internalize longer-term impacts on the growth of the economy.

The impact of larger commuting zones on the number of new firms and the other pecuniary firms remains ambiguous, but the sign of the effect depends on exactly the same condition in the long run and the short run. If $\mu + \eta > 1$, so that local variables display increasing returns, then larger commuting zones increase new firm creation in the short run and the long run, and all other pecuniary variables follow the same rule. If $\mu + \eta < 1$, then larger commuting zones reduce new firm creation and incomes, and do so more in the long run than in the short run. Just as in the case of larger idea networks, larger commuting zones have stronger effects on output in the long run than in the short run, although these effects can be either positive or negative.

Moreover, it is possible in the long run, as well as in the short run, for consolidation to lower income levels and still raise welfare, because larger metropolitan areas expand the range of amenity spillovers. It is also possible to see a reversal of fortune such that consolidation

increases welfare in the short run by increasing the effective level of amenities, but reduces welfare in the long run by reducing new firm formation: formally, that

$$\frac{\partial \ln u_t^i}{\partial \ln C_r} = (1 - \zeta\tau + \phi\gamma_i)\frac{1-\mu-\eta}{\sigma-1} - \gamma_i < 0$$
$$< \frac{\partial \ln \bar{u}^i}{\partial \ln C_r} = (1 - \zeta\tau + \gamma_i)\frac{1-\mu-\eta}{(1-\mu)(\sigma-1)} - \gamma_i. \quad (28)$$

Such reversals can only occur when entrepreneurship spillovers are mostly regional rather than local and as a consequence fragmentation increases firm creation ($\mu + \eta < 1$). In the short run amenities are largely predetermined, so lower fragmentation is attractive because it allows wider enjoyment of existing amenities. In the long run, however, not only does the cost of reduced firm creation increases by a factor $1/(1-\mu)$ as lower creativity is compounded. The ensuing slowdown in income growth also translates into lower levels of amenities because amenities are generated with income, and their income elasticity is higher in the long run than in the short run due to their persistence ($\phi < 1$).

This reversal occurs when love of variety occupies a middle range. If σ is too high, then the pecuniary effects of new firm creation become modest and the amenity effect will always dominate. At the extreme, as the overall production function becomes close to linear in output across firms, new firms would have no value at all, and expanding the city size would always be beneficial because of the increased availability of amenities. If σ is too close to one, then the pecuniary effects of new firm creation become extremely strong, and the welfare consequences of consolidation will depend completely on the impact on new firm creation, which always has the same sign in the short and long run. But when σ occupies a middle ground, then it is possible that access to amenities matters most in the short run, but the pecuniary effect dominates in the long run, leading to a preference reversal.

The comparative statics on the size of the interval $\bar{\sigma}_s^i - \underline{\sigma}_s^i$ can be interpreted as comparative statics on the likelihood of a welfare reversal occurring. For example, the comparative statics on amenity persistence can be interpreted as suggesting that a reversal is more likely when amenities persist longer over time. More persistent amenities make the early welfare gains from consolidation stronger because of the legacy of past amenities. These amenities will eventually decrease because of the lower level of firm creation after consolidation. A lower housing supply elasticity or a larger budget share of housing make reversals less likely because more of the income losses from consolidation are absorbed by lower housing prices. Likewise, greater human-capital spillovers imply smaller income losses from consolidation.

The range $(\underline{\sigma}_s^i, \bar{\sigma}_s^i)$ depends on an individual's preferences and shifts down as amenities become more important ($\partial\bar{\sigma}_s^i/\partial\gamma_i < \partial\underline{\sigma}_s^i/\partial\gamma_i < 0$). As a result, both groups may experience a preference reversal, or either group alone might. Yet, preference reversals by the skilled always bring them into closer alignment with the unskilled, if initially the skilled like consolidation but later come to dislike it. Preference reversals by the unskilled are more likely to break the alignment, since the unskilled will move from preferring consolidation, like the skilled, to disliking consolidation. This result suggests that short-sightedness by unskilled workers may be a thornier political problem. Consensus-building politicians would have an incentive to dispel myopic reasoning by the skilled in order to reach a result that is in everyone's long-run interest. Conversely, they may have a perverse incentive to cultivate

myopia in the unskilled so they support a consolidation that brings them short-run gains but long-run losses.

We now return to increases in density, just as in Proposition 2 and Corollary 1, instead of improvements in technology. This generates the scope for a different preference reversal, where densification is unattractive in the short run but attractive in the long run. As before, we begin with the trade-off between an idea network and a mega city by examining the effects of consolidating the population of an idea network into fewer, denser cities with less land.

Proposition 4 *Suppose the economy consists of identical cities divided into identical regions and that* $\mu < 1$.

If $\eta > 1 - \mu$, *consolidation of each idea network into denser commuting zones has larger positive effects on all pecuniary variables in the steady state than in the short run* ($\mu + \eta > 1 \Rightarrow d\bar{x}/dC_r|_{T \propto C_r} < dx_t/dC_r|_{T \propto C_r} < 0$ *for* $x \in \{N, Y, w, e, h, r, \rho\}$). *Then if and only if* $\gamma_i < \zeta\tau$ *there is a non-empty range of gains from variety such that consolidation into denser commuting zones reduces individual utility in the short run and increases it in the steady state* ($du_t^i/dC_r|_{T \propto C_r} > 0 > d\bar{u}^i/dC_r|_{T \propto C_r} \Leftrightarrow \sigma \in \left(\underline{\sigma}_{d^+}^i, \bar{\sigma}_{d^+}^i\right)$). *When gains from variety are higher, consolidation into denser commuting zones increases individual utility in the steady state more than it does in the short run* ($d\bar{u}^i/dC_r|_{T \propto C_r} < du_t^i/dC_r|_{T \propto C_r} < 0 \Leftrightarrow \sigma < \underline{\sigma}_{d^+}^i$). *When gains from variety are lower, consolidation into denser commuting zones decreases individual utility in the steady state less than it does in the short run* ($du_t^i/dC_r|_{T \propto C_r} > d\bar{u}^i/dC_r|_{T \propto C_r} > 0 \Leftrightarrow \sigma > \bar{\sigma}_{d^+}^i$). *The range of gains from variety that generate such a preference reversal expands when amenities have greater persistence* ($\partial\left(\bar{\sigma}_{d^+}^i - \underline{\sigma}_{d^+}^i\right)/\partial\phi < 0$) *and when local spillovers are higher* ($\partial\left(\bar{\sigma}_{d^+}^i - \underline{\sigma}_{d^+}^i\right)/\partial\mu > \partial\left(\bar{\sigma}_{d^+}^i - \underline{\sigma}_{d^+}^i\right)/\partial\eta > 0$).

If instead $\eta < 1 - \mu$, *consolidation of each idea network into denser commuting zones has larger negative effects on all pecuniary variables in the steady state than in the short run* ($\mu + \eta < 1 \Rightarrow d\bar{x}/dC_r|_{T \propto C_r} > dx_t/dC_r|_{T \propto C_r} > 0$ *for* $x \in \{N, Y, w, e, h, r, \rho\}$). *Then if and only if* $\gamma_i > \zeta\tau$ *there is a non-empty range of gains from variety such that consolidation into denser commuting zones increases individual utility in the short run and decreases it in the steady state* ($du_t^i/dC_r|_{T \propto C_r} < 0 < d\bar{u}^i/dC_r|_{T \propto C_r} \Leftrightarrow \sigma \in \left(\underline{\sigma}_{d^-}^i, \bar{\sigma}_{d^-}^i\right)$). *When gains from variety are higher, consolidation into denser commuting zones decreases individual utility in the steady state more than it does in the short run* ($d\bar{u}^i/dC_r|_{T \propto C_r} > du_t^i/dC_r|_{T \propto C_r} > 0 \Leftrightarrow \sigma < \underline{\sigma}_{d^-}^i$). *When gains from variety are lower, consolidation into denser commuting zones increases individual utility in the steady state less than it does in the short run* ($du_t^i|_{T \propto C_r} < d\bar{u}^i/dC_r|_{T \propto C_r} < 0 \Leftrightarrow \sigma > \bar{\sigma}_{d^-}^i$). *The range of gains from variety that generate such a preference reversal expands when amenities have greater persistence* ($\partial\left(\bar{\sigma}_{d^-}^i - \underline{\sigma}_{d^-}^i\right)/\partial\phi < 0$) *and when human capital generates lower local spillovers* ($\partial\left(\bar{\sigma}_{d^-}^i - \underline{\sigma}_{d^-}^i\right)/\partial\eta < 0$).

Proposition 4 again shows that the condition for consolidation into denser cities to increase wealth and new firm creation is the same in the long run as it is in the short run. As in Proposition 3, the long-run effects on all pecuniary variables are larger than the short-run effects because new firm creation accumulates over time. Also as in Proposition 3, it is possible to have a preference reversal, such that people initially prefer densification (or not) but eventually prefer the opposite.

In the case where $\mu + \eta > 1$, densification increase incomes, but as in Proposition 2,

welfare may go down because the cost of housing increases. This can only happen if welfare is more responsive to higher housing costs than to greater access to amenities ($\zeta\tau > \gamma_i$). However, it is possible that even if higher housing costs overwhelm the benefits of density in the short run, they no longer do so in the long run, when output rises much more and translates into even greater man-made amenities: formally, that

$$\frac{\partial \ln u_t^i}{\partial \ln C_r} + \frac{\partial \ln u_t^i}{\partial \ln T} = \zeta\tau - \gamma_i - (1 - \zeta\tau + \phi\gamma_i)\frac{\mu + \eta - 1}{\sigma - 1} > 0$$
$$> \frac{\partial \ln \bar{u}^i}{\partial \ln C_r} + \frac{\partial \ln \bar{u}^i}{\partial \ln T} = \zeta\tau - \gamma_i - (1 - \zeta\tau + \gamma_i)\frac{\mu + \eta - 1}{(1-\mu)(\sigma - 1)}. \quad (29)$$

As in Proposition 3, preference reversals are possible when gains from variety are intermediate, so the short-run increase in firm creation does not raise income enough to offset the welfare loss from higher housing costs, but the long-run increase does. The range of returns to diversity consistent with such a preference reversal is higher when density induces a larger increase in firm creation because local spillovers in firm creation are stronger; and when amenities are more persistent, so the wedge between the long-run and short-run welfare impacts of a change in incomes is larger.

In this case, preference reversals for a single group display the opposite pattern as in Proposition 3. The less skilled spend relatively more on housing and less on amenities than the skilled. Thus, their preference reversal tends to generate consensus: initially they dislike density and the ensuing increase in house prices, but eventually they recognize its overwhelming productivity benefits. Conversely, when the skilled have a preference reversal of this sort, alignment across groups may be broken: in the short run everyone prefers less density, but in the long run the skilled alone find density attractive.

The second part of Proposition 4 shows that when density reduces productivity ($\mu+\eta < 1$) preference reversals are analogous to the case of larger commuting zones. As in Proposition 2, the benefit of greater access to amenities might offset the cost of both income losses and the loss of real estate (provided that $\gamma_i > \zeta\tau$). Just as in Proposition 3, this might be true in the short run but not in the long run, when any income losses are magnified. Given that densification entails a reduced endowment of land, if it reduces productivity it is naturally less likely that it should prove attractive in the short run, and conversely more attractive that it should be welfare-reducing in the long run.

A similar set of results occur when we consider densification across rather than within regions.

Corollary 2 *Suppose the economy consists of identical cities divided into identical regions and that $\mu < 1$. If and only if $\gamma_i < \zeta\tau$ there is a non-empty range of gains from variety such that crowding into denser regions reduces individual utility in the short run and increases it in the steady state ($du_t^i/dR|_{T\propto R} > 0 > d\bar{u}^i/dR|_{T\propto R} \Leftrightarrow \sigma \in (\underline{\sigma}_D^i, \bar{\sigma}_D^i)$). When gains from variety are higher, crowding into denser regions increases individual utility in the steady state more than it does in the short run ($d\bar{u}^i/dR|_{T\propto R} < du_t^i/dR|_{T\propto R} < 0 \Leftrightarrow \sigma < \underline{\sigma}_D^i$). When gains from variety are lower, crowding into denser regions decreases individual utility in the steady state less than it does in the short run ($du_t^i/dR|_{T\propto R} > d\bar{u}^i/dR|_{T\propto R} > 0 \Leftrightarrow \sigma > \bar{\sigma}_D^i$). The range of gains from variety that generate such a preference reversal expands when amenities*

have greater persistence ($\partial\left(\bar{\sigma}_D^i - \underline{\sigma}_D^i\right)/\partial\phi < 0$) *and when innovation spillovers are higher* ($\partial\left(\bar{\sigma}_D^i - \underline{\sigma}_D^i\right)/\partial\mu > \partial\left(\bar{\sigma}_D^i - \underline{\sigma}_D^i\right)/\partial\eta = \partial\left(\bar{\sigma}_D^i - \underline{\sigma}_D^i\right)/\partial\nu > 0$).

As in Corollary 1, denser regions generate higher output and offer greater access to amenities at the cost of a loss of land. Again, if real estate affects welfare more than access to amenities ($\zeta\tau > \gamma_i$) there is a range of returns to variety that generate a preference reversal. When gains from variety are intermediate, long-run income gains from denser idea networks overcome greater scarcity of land in the steady state even though they are insufficient to do so in the short run.

The range is larger than in Proposition 4 because it is increasing in innovation spillovers, and a denser idea network enables more human-capital spillovers than a denser commuting zone. For the same reason, in this case only one type of preference reversal is possible: short-term losses followed by longer-term gains. Density unambiguously increases firm creation and output when densification occurs throughout the network and reaps all gains from idea exchange, rather than the more limited gains from consolidation within a commuting zone (recombinant growth implies that $\mu + \eta + \nu > 1$). When densification can only cause new firm formation to rise, then the long run can only be better than the short run.

4.2 Urban Networks and Balanced Growth

While we will not consider the case of explosive growth, we can address the knife-edge case where $\mu = 1$ and the economy has a balanced growth path. Then the mean growth rate of the aggregate number of firms is

$$\mathbb{E}\left(\ln N_t - \ln N_{t-1}\right) = (\eta + \nu)\left(\ln H - \ln R\right) - \eta \ln C_r + \mathbb{E} \ln \alpha_t \text{ for all } t. \tag{30}$$

The mean growth rate of aggregate output, wages, profits per entrepreneur, land rents and aggregate amenities is $1/(\sigma - 1)$ times that of firms. The mean growth rate of the aggregate housing stock is $(1 - \zeta)/(\sigma - 1)$ times that of firms and the mean growth rate of house prices is $\zeta/(\sigma - 1)$ times that of firms. The mean growth rate of utility is

$$\mathbb{E}\left(\ln u_t^i - \ln u_{t-1}^i\right) = \frac{1 - \zeta\tau + \gamma_i}{\sigma - 1}\mathbb{E}\left(\ln N_t - \ln N_{t-1}\right) \text{ for } i \in \{H, L\}. \tag{31}$$

These equations illustrate that along a balanced growth path, fragmentation reduces the mean growth rate. The ambiguous earlier results disappear in this case, because a balanced growth rate can only occur when $\mu = 1$, which also guarantees that consolidation is beneficial in the earlier cases as well. When the impact of past local ideas is strong enough to generate balanced growth, then they are also strong enough to ensure that larger networks and larger commuting zones are also beneficial.

Naturally, it is quite possible to imagine different idea generation processes that generate balanced growth without strong agglomeration economies at the city level. For example, if all past idea were shared within the entire idea network, not just a single city, then it would be possible to have balanced growth with much weaker local agglomeration effects. In that case, it would be possible for consolidation to reduce the growth rate.

Proposition 5 *Suppose the economy consists of identical cities divided into identical regions and that firm creation is difference stationary ($\mu = 1$). Then the economy converges to a balanced path of endogenous growth.*

The mean growth rate of each variable is higher when the economy is less fragmented into separate idea-sharing networks and when these are less fragmented into separate commuting zones. The separation of idea-sharing networks is more harmful than that of commuting zones ($\partial \mathbb{E} \ln (N_t/N_{t-1}) / \partial R < \partial \mathbb{E} \ln (N_t/N_{t-1}) / \partial C_r < 0$).

A balanced growth path exists only in the knife-edge case of a difference-stationary process of firm creation. Such a process implies that firms are perfectly persistent at the local level. This immediately implies that any fragmentation is bad in the short run and even worse in the long run. Fragmentation reduces steady-state growth relative to the case of a single megalopolis, which can better combine the stock of human capital with the aggregate stock of persistent firms.

This result emphasizes, again, that agglomerations need to be judged for their long-run effects as well as their short-run effects. While the short-run costs of connecting a metropolis or a network may not be offset by the short-run benefits, if those connections lead to a higher growth rate, then eventually the benefits will offset the costs. The decision to invest in connections will then depend on the discount rate.

5 Urban Networks and Regional Heterogeneity

We now relax the assumption of complete symmetry across regions. We now assume that all cities in region r have the same endowments of factors, firms and amenities and the same productivity levels, but that these endowments and productivities can differ across regions. This assumption is akin to assuming that Europe is made up of different urban networks and those networks might be very different, but within each network, every city is the same.

Given this symmetry, the total number of firms in each region r is

$$N_{r,t} = C_r N_{c,t} = C_r^{1-\eta-\mu} H_r^{\eta+\nu} \alpha_{r,t} N_{r,t-1}^{\mu}. \tag{32}$$

The evolution of the number of firms in each region then follows a logarithmic first-order autoregressive process independent of conditions in all other regions. The simultaneous but disjoint process of firm creation in each region drives aggregate output, which equals

$$Y_t = \left\{ \sum_{r=1}^{R} \left[(A_{r,t} L_r)^{\sigma-1} N_{r,t} \right]^{\frac{1}{\sigma}} \right\}^{\frac{\sigma}{\sigma-1}}. \tag{33}$$

The basic equations of our model then apply directly, with region- instead of city-specific indices, to all pecuniary variables. Eq. (2) yields total regional output ($Y_{r,t} = C_r Y_{c,t}$), eq. (3) wages, eq. (8) profits per entrepreneur; eq. (13) the region's total housing stock ($h_{r,t} = C_r h_{c,t}$), eq. (14) house prices and eq. (15) land rents; eq. (11) total amenities in the region ($G_{r,t} = C_r G_{c,t}$). As a result, in region r unskilled workers have utility

$$u_{r,t}^L = \frac{\sigma - 1}{\sigma} \frac{1}{L_r} \left[\Psi_{c,t} \left(\frac{T_r}{\tau} \right)^{\zeta} \right]^{\tau} \left(\frac{\Gamma_{r,t} \bar{\gamma}^{\phi} G_{r,t-1}^{1-\phi}}{C_r} \right)^{\gamma_L} Y_{r,t}^{1-\zeta\tau+\phi\gamma_L} \tag{34}$$

and skilled entrepreneurs have utility

$$u_{r,t}^H = \frac{1}{\sigma}\frac{1}{H_r}\left[\Psi_{c,t}\left(\frac{T_r}{\tau}\right)^{\zeta}\right]^{\tau}\left(\frac{\Gamma_{r,t}\bar{\gamma}^{\phi}G_{r,t-1}^{1-\phi}}{C_r}\right)^{\gamma_H} Y_{r,t}^{1-\zeta\tau+\phi\gamma_H}. \tag{35}$$

Gains from trade imply that firm creation in one region entails positive spillovers for all its trading partners:

$$\frac{\partial \ln Y_{r,t}}{\partial \ln N_{s,t}} = \frac{1}{\sigma(\sigma-1)}\frac{Y_{s,t}}{Y_t} > 0 \text{ for all } s \neq r. \tag{36}$$

From the point of view of one region, these spillovers are externalities, which raise the possibility of coordination failures. It is certainly possible that one region might pursue policies that increase utility within that region at the expense of firm creation, perhaps by reducing the density of cities. These policies would harm the rest of the country.

Proposition 6 *Suppose the economy consists of R regions, each divided into measure C_r of identical cities.*

If $\mu+\eta>1$ and $\gamma_i<\zeta\tau$ there is a non-empty range of gains from variety such that if a region represents a sufficiently small share of aggregate output then its residents' utility decreases if commuting zones become denser in the region but increases if commuting zones become denser in all regions ($Y_{r,t}/Y_t < \hat{y}^i_{d^+} \Leftrightarrow \partial \ln u^i_{r,t}/\partial \ln C_r + \partial \ln u^i_{r,t}/\partial \ln T_r > 0 > \sum_{s=1}^R \partial \ln u^i_{r,t}/\partial \ln C_s + \partial \ln u^i_{r,t}/\partial \ln T_r$). The threshold is lower when amenities are more important ($\partial \hat{y}^i_{d^+}/\partial \gamma_i < 0$) and less persistent ($\partial \hat{y}^i_{d^+}/\partial \phi < 0$), housing has a higher supply elasticity ($\partial \hat{y}^i_{d^+}/\partial \zeta > 0$) and a lower budget share ($\partial \hat{y}^i_{d^+}/\partial \tau > 0$), and local spillovers or gains from variety are greater ($\partial \hat{y}^i_{d^+}/\partial \eta = \partial \hat{y}^i_{d^+}/\partial \mu < 0 < \partial \hat{y}^i_{d^+}/\partial \sigma$).

If $\mu+\eta<1$ there is a non-empty range of gains from variety such that if a region represents a sufficiently small share of aggregate output then its residents' utility decreases if commuting zones become smaller in the region but increases if they become smaller in all regions ($Y_{r,t}/Y_t < \hat{y}_s \Leftrightarrow \partial \ln u^i_{r,t}/\partial \ln C_r < 0 < \sum_{s=1}^R \partial \ln u^i_{r,t}/\partial \ln C_s$). The threshold is lower when amenities are less important ($\partial \hat{y}_s/\partial \gamma_i > 0$) and less persistent ($\partial \hat{y}_s/\partial \phi < 0$), housing has a higher supply elasticity ($\partial \hat{y}_s/\partial \zeta > 0$) and a lower budget share ($\partial \hat{y}_s/\partial \tau > 0$), gains from variety are greater ($\partial \hat{y}_s/\partial \sigma > 0$) and local spillovers are smaller ($\partial \hat{y}_s/\partial \eta = \partial \hat{y}_s/\partial \mu > 0$).

It $\mu+\eta<1$ and $\gamma_i>\zeta\tau$ there is a non-empty range of gains from variety such that if a region represents a sufficiently small share of aggregate output then its residents' utility increases if commuting zones become denser in the region but decreases if they become denser in all regions ($Y_{r,t}/Y_t < \hat{y}_{d^-} \Leftrightarrow \partial \ln u^i_{r,t}/\partial \ln C_r + \partial \ln u^i_{r,t}/\partial \ln T_r < 0 < \sum_{s=1}^R \partial \ln u^i_{r,t}/\partial \ln C_s + \partial \ln u^i_{r,t}/\partial \ln T_r$). The threshold is lower when amenities are less important ($\partial \hat{y}_{d^-}/\partial \gamma_i > 0$) and less persistent ($\partial \hat{y}_{d^-}/\partial \phi < 0$), housing has a lower supply elasticity ($\partial \hat{y}_{d^-}/\partial \zeta < 0$) and a higher budget share ($\partial \hat{y}_{d^-}/\partial \tau < 0$), gains from variety are greater ($\partial \hat{y}_{d^-}/\partial \sigma > 0$) and local spillovers are smaller ($\partial \hat{y}_{d^-}/\partial \eta = \partial \hat{y}_{d^-}/\partial \mu > 0$).

The proposition details three possibilities in which there can be conflict between local benefits and national benefits. In the first part of the proposition, we assume that $\mu+\eta>1$, which means that firm creation increases as cities become denser. In this case, there is a positive externality from denser cities and there will be a potential conflict if the density reduces utility within the region.

As in Proposition 2, the condition $\gamma_i < \zeta\tau$ ensures that housing supply is sufficiently important and sufficiently inelastic that it is possible that the increased land made available by expanding the number of cities offsets the losses generated by a reduction in new firm formation and the losses from reductions in amenity spillovers, which is captured by γ_i. Yet that extra land provides no benefit for the other regions of the country, and collectively they may lose if all regions reduce density. Formally, the condition is

$$\frac{\partial \ln u^i_{r,t}}{\partial \ln C_r} + \frac{\partial \ln u^i_{r,t}}{\partial \ln T_r} = \zeta\tau - \gamma_i - (1 - \zeta\tau + \gamma_i\phi)\frac{\mu+\eta-1}{(\sigma-1)\sigma}\left(\sigma - 1 + \frac{Y_{r,t}}{Y_t}\right)$$
$$> 0 > \sum_{s=1}^{R} \frac{\partial \ln u^i_{r,t}}{\partial \ln C_s} + \frac{\partial \ln u^i_{r,t}}{\partial \ln T_r} = \zeta\tau - \gamma_i - (1 - \zeta\tau + \gamma_i\phi)\frac{\mu+\eta-1}{\sigma-1}. \quad (37)$$

As before, we have assumed that the change in the amount of available land is proportional to the change in the number of cities ($\partial \ln T_r / \partial \ln C_r = 1$). The first expression in the condition is the change in utility within a network from increasing the number of cities and the amount of land. The third term, which reflects firm creation, is negative since $\mu + \eta < 1$. The overall derivative can be positive only if $\zeta\tau - \gamma_i$ is positive and large enough to offset the last term, which means that the gains from extra land minus the losses from amenity spillover reductions are large enough to offset the lost income. Crucially, the change in income from an increase in firm creation reflects both a direct effect ($\sigma - 1$) and a general-equilibrium effect that is proportional to the size of region r relative to the whole economy ($Y_{r,t}/Y_t$).

The last expression in the condition reflects the net impact on national utility if all regions increase their number of cities in the same proportion. The direct effect of increasing land availability but reducing access to amenities is unchanged. The third term, which again reflects the welfare loss from reduced new firm formation, is always larger in magnitude and hence more negative than the comparable term on the first line. This increase in magnitude reflects the internalization of gains from variety across regions.

The increase in magnitude also makes it possible for densification to be bad at the region level but good at the national level. Once again, this can occur only if gains from variety are neither so small that increasing density in the whole economy fails to compensate the loss of land, nor so large that the direct income gain from greater density dominates even before the general-equilibrium benefits are taken into account. For intermediate gains from variety, some regions may be tempted to free-ride on the innovation provided by the others, as in Epifani and Gancia (2009).

This result may well reflect real-world conditions. Many areas in the developing world limit densification, which plausibly causes an increase in the total number of urban areas. Land use restrictions may well restrict the growth of Greater London, which in turn increases the number of people living elsewhere in England. Land use restrictions in Silicon Valley similarly push some would-be entrepreneurs to locate elsewhere. These restrictions are often locally popular, which is presumably why they are so common. Yet it may well be that restricting the densification of London and Silicon Valley hurts the overall economies of the United Kingdom and the United States.

Naturally, the temptation to free-ride is greater for smaller regions. A region with a lower share of output has less weight in global productivity. If its rate of innovation declines,

aggregate income is going to be less affected. This perhaps explains why larger cities, like New York, have more aggressively tried to allow increase in density despite neighborhood opposition to building. In larger regions, the region itself benefits internalizes more of its own productivity gains—while all regions fully internalize changes to land availability and amenities.

The comparative statics follow intuitively from this tug of war. When output gains are large, internalizing a small share of the global economy is enough to dispel coordination failure. Hence, free-riding is less of a problem when local spillovers and gains from variety are large. Greater love of amenities has the same effect, both because it makes density itself more attractive, and because it makes individuals keener on raising output, as does lower persistence of amenities. Conversely, a greater share of housing and a lower elasticity of housing supply both make density locally costlier and reduce the utility value of income gains, a greater fraction of which is absorbed by the housing sector. In that case, free-riding becomes more tempting even for larger regions.

The second part of Proposition 6 examines the reverse case, where densification is bad for new firm creation because $\mu + \eta < 1$. In this case, densification can be attractive only because it increases the level of amenities. We first consider the case of shrinking the size of commuting zones, holding total land constant. We did not consider that case when $\mu+\eta > 1$ because in that case larger commuting zones are good for both productivity and amenities and have no impact on land availability, so they are always good for both region and nation.

But when $\mu + \eta < 1$, smaller commuting zones can be good because there is more entrepreneurship. However, smaller commuting zones have lower access to amenities. In this case, it is possible to have local utility gains from larger cities, and global utility losses. If cities expand primarily because of amenities and this creates a loss in productivity, then the country as a whole suffers as a result.

Perhaps the closest real-world counterpart to this case occurs when large capital cities grow as places of consumption, rather than production. This may have been true of London or Paris in particular epochs and it may be true in some developing world countries today. If these large cities reduce productivity, even as they raise their own residents' welfare, then they are harming the rest of the country.

In this case, free riding is a greater problem when amenities are more important, because now increased access to amenities rather than increased availability of land is the reason why the region's preferences depart from the nation's. Instead, the role of persistence in amenities and of the housing parameters remains the same because they are identically reflected in the welfare value of changes in income. Again, free riding is more easily solved when income changes are larger, namely when gains from variety are larger but, in this case, when local spillovers are smaller and so the disadvantages of concentration are starkest.

The final part of Proposition 6 discusses densification when $\mu + \eta < 1$. In this case, the increase in amenities needs to offset both the loss in productivity and the loss in land. Again, if these amenity benefits are large enough, it is possible that the network may benefit from having fewer commuting zones with less total land and that the nation as a whole may lose out because of the reduction in firm formation. Naturally, in this last case a greater importance of housing makes a preference for income-reducing density more unlikely, and thus tends to eliminate the incentive to free-ride.

In the long run, if $\mu < 1$ then $\ln N_{r,t}$ follows a stationary first-order autoregressive process

and the number of firms, amenities and welfare levels converge to their steady-state levels. A version of Proposition 6 applies again because in the steady state

$$\frac{\partial \ln \bar{u}_r^i}{\partial \ln C_r} = (1 - \zeta\tau + \gamma_i) \frac{1 - \eta - \mu}{(1 - \mu)(\sigma - 1)\sigma} \left(\sigma - 1 + \frac{\bar{Y}_r}{Y_t} \right) - \gamma_i \tag{38}$$

while

$$\sum_{s=1}^{R} \frac{\partial \ln \bar{u}_r^i}{\partial \ln C_s} = (1 - \zeta\tau + \gamma_i) \frac{1 - \eta - \mu}{(1 - \mu)(\sigma - 1)} - \gamma_i. \tag{39}$$

As in Proposition 3, income effects loom larger in the long run because they are amplified and fully impact on amenity levels.

This magnification implies that when a coordination failure exists in the short run, foresight tends to make it disappear. Foresight brings income effects in sharper focus, and makes changes in access to amenities of the endowment of land less important. Even smaller regions, then, are less likely to free ride: formally, the three thresholds $\hat{y}^i$ are all decreasing in the magnitude of income changes.

It does not follow, however, that foresight unambiguously reduces the possibility of coordination failures. On the contrary, it can also create free riding where none existed in the short run. A coordination failure requires not only that a single region should care more about amenities or real estate than innovation and output, but also that output gains should be large enough to dominate welfare flosses at the aggregate level. If this latter condition is satisfied only in the steady state, then coordination failures emerge only in the long run. Long-run national welfare may rise when density increases in all regions, accelerating innovation and yielding productivity gains. Yet the scarcity of real estate may suffice both to offset the short-run national gains, and to make densification in one's own region welfare-reducing at any time horizon.

6 Urban Networks and Urban Heterogeneity

We now introduce heterogeneity and look at the incentives of a small set of cities to combine into a larger commuting zone or even into a smaller denser city. These incentives reflect not only the direct impact of this consolidation but also the impact that city size will have on migration to and from the cities. We assume that these cities form a small open economy whose decisions have no impact on the overall level of output in the system. For simplicity, we also retain our assumption that the rest of the economy includes no heterogeneity within regions, so we can refer to the representative city in each region.

Formally, we assume that a full measure C_r of cities in every region r are identical. This assumption captures the homogeneity of the outside economy that is not considering consolidation. We focus on a finite, measure-zero set of atomistic cities in region r that may differ from the representative city. This assumption guarantees that the cities that are consolidating are small (measure-zero), while allowing them to be different from the rest of the economy. As long as these different cities have a finite endowment $(L_c, T_c; N_{c,t-1}, G_{c,t-1})$, they will have no influence on aggregates at the regional level, nor on the global economy.

In period t each city c in region r attracts a number of high-skill entrepreneurs equal to

$$H_{c,t} = \frac{H_r}{C_r} \times \left\{ \left[\left(\frac{C_r N_{c,t-1}}{N_{r,t-1}} \right)^{\mu} \left(\frac{C_r L_c}{L_r} \right)^{\sigma-1} \right]^{\frac{1-\zeta\tau+\phi\gamma_H}{\sigma}} \left(\frac{C_r T_c}{T_r} \right)^{\zeta\tau} \left(\frac{C_r G_{c,t-1}}{G_{r,t-1}} \right)^{(1-\phi)\gamma_H} \right\}^{\frac{\sigma}{\sigma-\eta(1-\zeta\tau+\phi\gamma_H)}}. \tag{40}$$

The city would attract an even share H_r/C_r of the region's aggregate amount of talent if it were exactly as attractive as the representative city. It attracts more or fewer skilled entrepreneurs depending on their relative attractiveness, which is an increasing function of their relative endowments. We maintain the assumption that productivity shocks are instead common to all cities in the region.

Firm creation in city c is then

$$N_{c,t} = \Omega^N_{r,t} \left\{ N^{\mu}_{c,t-1} \left[L_c^{\frac{\sigma-1}{\sigma}(1-\zeta\tau+\phi\gamma_H)} T_c^{\zeta\tau} G^{(1-\phi)\gamma_H}_{c,t-1} \right]^{\eta} \right\}^{\frac{\sigma}{\sigma-\eta(1-\zeta\tau+\phi\gamma_H)}}, \tag{41}$$

where to simplify notation we let $\Omega^N_{r,t}$ (which is defined in the appendix) denote a composite of all regional variables, which are independent of conditions in any finite set of atomistic cities. Locally, the number of firms in a city is an increasing function of its endowment of firms last period, labor, land, and historic amenities.

City output is determined by eq. (2), wages by eq. (3) and profits by eq. (8); the housing stock by eq. (13), house prices by eq. (14) and land rents by eq. (15); and city amenities by eq. (11). The spatial equilibrium requires that skilled entrepreneurs' utility $u^H_{c,t} = u^H_{r,t}$ is equalized throughout each region. Since we are considering a non-measurable set of cities, these cities will have no impact on the aggregate welfare of the mobile skilled. However, since the unskilled are fixed geographically, their welfare will change. Unskilled workers' utility is

$$u^L_{c,t} = \Omega^u_{r,t} \frac{T_c^{\zeta\tau} G^{(1-\phi)\gamma_L}_{c,t-1}}{L_c} \left\{ N^{\mu}_{c,t-1} L_c^{\sigma-1} \left[T_c^{\zeta\tau} G^{(1-\phi)\gamma_H}_{c,t-1} \right]^{\eta} \right\}^{\frac{1-\zeta\tau+\phi\gamma_L}{\sigma-\eta(1-\zeta\tau+\phi\gamma_H)}}, \tag{42}$$

where again $\Omega^u_{r,t}$ is independent of local conditions and reported explicitly in the appendix. Endowments of land, amenities and labor have a direct effect on workers' utility, and also an indirect effect because they attract more skilled entrepreneurs and make them more productive. This influx of skilled entrepreneurs will increase wages and amenities, which is good for the unskilled, but more skilled workers will also increase home prices.

This double-edged impact of the skilled on the unskilled can explain why the unskilled and politicians that cater to them try to attract the skilled into commercial zones but not into residential zones. Less skilled urbanites can sometimes support local investments that appeal to the more skilled, such as local universities or innovation districts, like Boston's Innovation Zone. Simultaneously, they can still oppose gentrification of higher poverty areas. The economic energy brought by skilled entrepreneurs benefits the poor, while the rich's demand for housing space hurts the poor.

We now ask whether any finite set of cities has a unilateral incentive to unite in a single larger commuting zone, or even to concentrate in a single denser city at the cost of having a

reduced endowment of land. As opposed to our previous results, in this case, a key element in assessing consolidation or densification is whether these changes will attract more skilled entrepreneurs.

A natural benchmark is the case of where the cities in region r that are considering consolidation are identical to each other and to other cities in the region. As we will show below, this proves to be the setting that is least favorable to consolidation and most favorable to an urban network.

Proposition 7 *If a finite set of identical cities form a single wider commuting zone whose endowment is the sum of their endowments of labor, land and existing firms and amenities, their average level of amenities always increases. The number of skilled entrepreneurs in the set of cities rises if and only if*

$$\gamma_H + (1 - \zeta\tau + \phi\gamma_H) \frac{\mu + \eta - 1}{\sigma} > 0.$$

Consolidation creates more firms, produces more output, expands the housing stock, increases average wages, housing prices and land rents if and only if

$$\mu + \eta (1 + \gamma_H) > 1.$$

Consolidation increases the welfare of the less skilled if and only if

$$\gamma_L + (1 - \zeta\tau + \phi\gamma_L) \frac{\mu + \eta (1 + \gamma_H) - 1}{\sigma - \eta (1 - \zeta\tau + \phi\gamma_H)} > 0.$$

The first condition embeds the three major effects of consolidation on the fate of the skilled: higher amenities, higher housing prices and a change in the amount of spillovers, which can be positive or negative depending on $\mu + \eta - 1$. These three forces, and the first condition, also determine whether the skilled would prefer a region that produces a negligible amount of world output ($Y_{r,t}/Y_t = 0$) to be made up of larger commuting zones. The skilled are directly attracted by the greater amenities in the larger city (captured by the term γ_H). Consolidation can attract more skilled entrepreneurs, even if it reduces the profits to entrepreneurship, if the positive impact on amenities is sufficiently high. But the skilled will also be more attracted by a large city if local spillovers (η) and the persistence of local firms (μ) are sufficiently strong. Higher returns from new varieties and new innovation (lower σ) exacerbate the effect of increases or decreases in the level of spillovers.

The spatial equilibrium implies that if consolidation makes a set of cities more attractive to entrepreneurs, then sufficiently many entrepreneurs will move to the city so that their utility is equalized to the utility of the skilled elsewhere in the region. Since the set of cities merging is small, the overall utility of the skilled will not change.

The condition on whether larger commuting zones will increase wages and housing prices has weakened relative to the earlier condition ($\mu + \eta > 1$) found in Proposition 1, which applied to all region-wide changes. Larger commuting zones mean higher amenity levels, and this attracts more entrepreneurs to the city if the rest of the region remains fragmented. This flow of entrepreneurs into the larger city raises firm creation further by a factor $\eta\gamma_H$. This result emphasizes that consumer cities are not just about static welfare gains in consumption

(Glaeser, Kolko and Saiz 2001), but also about the dynamic gains that can result when skilled entrepreneurs locate in a particular area. The welfare consequences for the unskilled have a natural split. They benefit directly from amenities (γ_L) and indirectly from any inflow of the skilled, which raises incomes and amenities albeit also housing prices.

We now turn to the impact of consolidation of population into a single city with less land. As before, the creation of a single mega-city becomes less appealing when the overall density level rises dramatically.

Proposition 8 *If a finite set of identical cities shrink into a single denser commuting zone whose endowment is the sum of their endowments of labor and existing firms and amenities, but whose land area is that of a single city, then average amenities, housing prices and land rents always increase. The number of skilled entrepreneurs in the set of cities rises if and only if*

$$\gamma_H - \zeta\tau + \frac{\eta + \mu - 1}{\sigma}\left(1 - \zeta\tau + \phi\gamma_H\right) > 0.$$

Densification creates more firms, produces more output and increases average wages if and only if

$$\mu + \eta\left(1 - \zeta\tau + \gamma_H\right) > 1.$$

Densification increases the welfare of the less skilled if and only if

$$\gamma_L - \zeta\tau + \left(1 - \zeta\tau + \phi\gamma_L\right)\frac{\mu + \eta\left(1 - \zeta\tau + \gamma_H\right) - 1}{\sigma - \eta\left(1 - \zeta\tau + \phi\gamma_H\right)} > 0.$$

Densification expands the housing stock if and only if

$$\mu + \eta\left(1 - \zeta\tau + \gamma_H\right) - 1 > \zeta\left[\sigma - 1 + \mu + \eta\left(1 - \phi\right)\gamma_H\right].$$

The first three conditions in Proposition 8 are analogous to those in Proposition 7 immediately above, except that residents now suffer from greater scarcity of land and thus costlier housing, which is reflected in the additional terms $\zeta\tau$. Land loss makes consolidation into a single city less appealing, and the reduction in appeal is a function of the elasticity of housing supply and the importance of housing in the utility function. Again, this result is consistent with the observation that in places where housing supply is highly inelastic, such as much of Europe, urban networks are more common than mega-cities.

Higher density means higher housing costs and this also repels entrepreneurs, who also value cheap housing. Thus, an inelastic housing supply generates three costs of density to the unskilled. First, as in the case of wider commuting zones, a greater share of any income gains are absorbed by the housing sector. Second, land scarcity makes house prices rise even in the absence of income gains. Third, income gains are themselves muted because the skilled are less keen on moving to an expensive dense city. This combination of forces can be readily seen in the "affordable housing crises" which have now become the norm in successful cities, such as London and New York.

Just as abundant amenities can become an economic development strategy, because they attract skilled entrepreneurs, policies that make it easier to build cheap housing may also generate economic returns by attracting workers who create firms. If there are policies that both increase amenities and reduce housing supply, such as historic preservation districts or

greenbelts, then the impact on entrepreneurship and economic growth becomes a horse race between the impact on prices and the impact on the welfare of skilled.

The last result in Proposition 8 establishes that, somewhat paradoxically, less land can mean more housing supply as long as housing supply is sufficiently elastic (low ζ). This effect can occur only if density leads to an extremely sharp increase in housing demand, either because of higher amenities or greater productivity.

6.1 Urban Heterogeneity and the Case for Consolidation

We now asses the connection between heterogeneity and the impact of both densification and consolidation. We continue to characterize the rest of the region by a representative city. We also retain the assumption that all cities considering consolidation have identical endowments of land, so that a single denser city would simply have the same land endowment as any one city.

For tractability, we restrict and parametrize the types of heterogeneity within the set of cities that consider consolidation. We assume that for some fraction of these cities, all factors other than land move together in lockstep. For the rest of the cities, the three factors are independently distributed. These assumptions will allow us to characterize the joint empirical distribution of the endowments of labor, firms and amenities $(L_c, N_{c,t-1}, G_{c,t-1})$ across cities in the network with an index of heterogeneity in factor sizes and an index of correlation of factor sizes.

To capture the overall variance in factor sizes, we assume a non-degenerate empirical distribution $\hat{S}$ of shares: a set of K distinct values $\hat{s}_k \in (0,1)$ and associated empirical probabilities f_k such that $\sum_{k=1}^{K} f_k = 1$ and $\sum_{k=1}^{K} f_k \hat{s}_k = 1/m$. We define an empirical distribution S with the same empirical probabilities f_k and corresponding values

$$s_k = \theta \hat{s}_k + \frac{1-\theta}{m}. \tag{43}$$

This distribution is a linear combination of the non-degenerate $\hat{S}$ distribution, and the degenerate distribution that delivers the same amount to each city in the set. The index $\theta \in [0,1]$, which determines the weight on the non-degenerate distribution, becomes our measure of heterogeneity in endowment sizes within the set of cities.

We use $s_{F,c} \equiv F_c / \sum_{d=1}^{m} F_d$ to denote each city's share of the total factor endowment in the network for $F_c \in \{L_c, N_{c,t-1}, G_{c,t-1}\}$. We then assume that the sample of cities is split between a fraction ω of cities which have shares of each factor that are independent but identically distributed, with marginal distribution S; and a fraction $1-\omega$ of cities that have common factor proportions and whose shares $s_{L,c} = s_{N,c} = s_{G,c}$ have distribution S.[2] This distribution is a mixture of an independent distribution of the three endowments, and of a distribution with perfect correlation between the three. The index $\omega \in [0,1]$, which determines the weight on the perfectly correlated distribution, serves as our measure of correlation in cities' endowments across factors.

[2] We assume that the network comprises a sufficiently large number of cities for the empirical distribution of the three factor shares to be independent: $(1-\omega)\, m \geq 8$.

The overall joint empirical distribution is then described by

$$\Pr\left(s_{L,c} = s_i, s_{N,c} = s_j, s_{G,c} = s_k\right) = \omega f_i f_j f_k + \mathbf{1}_{i=j=k}\left(1-\omega\right) f_k. \tag{44}$$

With these assumptions, we can derive results on the ways in which heterogeneity influences the decision to create cities out of networks.

Proposition 9 *Consider a finite set of cities with identical land endowments. The distribution of their relative endowments of labor, firms and amenities is a mixture of a perfectly correlated distribution, with weight* $1-\omega$*, and three independent marginal distributions, with weight* ω*. Each of these distributions is a linear combination of equal shares, with weight* $1-\theta$*, and any unequal distribution of shares, with weight* $\theta \in [0,1]$*.*

If forming a single larger or denser city attracts more skilled entrepreneurs; or creates more firms, produces greater output and pays higher wages; or provides higher utility to unskilled workers; or entails a larger housing stock, then it also does so for any $\theta' \geq \theta$ *and any* $\omega' \geq \omega$*.*

Our baseline case of identical cities corresponds to $\theta = 0$. Proposition 9 establishes that heterogeneity in city size always makes it more likely that a single city should be preferred to a network of separate commuting zones. Heterogeneity in factor proportions within a city also makes forming a single city more attractive.

These results reflect the gains from trade that occur within larger commuting zone. Just as allowing trade between two identical countries does little in a standard Heckscher-Ohlin framework, merging two identical cities doesn't allow new combinations of past firms and unskilled labor. However, if some cities have many workers but few firms, and others have many firms but few workers, then there will be benefits from consolidation into a single city. When the communities are distinct, there is inefficiency because in the first set of cities it is difficult to create new firms but easy to grow them, while in the latter firms are easily created but then cannot reach their optimal size. Workers benefit from being able to commute to the places with higher firm density.

Higher values of ω imply a more frequent mismatch in endowments across cities. Higher values of θ exacerbate the size of any existing mismatch, between labor, firms and amenities (if $\omega > 0$) and between all of them and the endowment of land, which is distributed uniformly. Differences in density also make a larger commuting zone desirable: workers and firms from denser cities benefit from access to land in less dense parts of the network. As the size θ and the frequency ω of endowment mismatches grow, the network becomes less efficient, and the city becomes more advantageous.

When $\theta = 0$ all of the cities have perfectly equal shares of land and every other factor (then ω is irrelevant). This is exactly the case discussed above, which minimizes the gains from urban consolidation. When $\theta > \omega = 0$, then the cities differ in size, but all the factors except land move together. In this case, greater heterogeneity makes a network less likely to be preferred to a larger city, but it does not affect its ranking relative to a denser city.

When $\omega = 0$, the ratio of output in a network of m cities to output in a single denser city equals

$$m \sum_{k=1}^{1} f_k \left(\frac{1-\theta}{m} + \theta \hat{s}_k\right)^{\varepsilon_{Y,L}+\varepsilon_{Y,N}+\varepsilon_{Y,G}}, \tag{45}$$

where $\varepsilon_{Y,F}$ is the equilibrium elasticity of output to factor F, taking into account both its direct effect in production and its indirect effect through the induced influx of skilled entrepreneurs. For any distribution S and any heterogeneity θ, this ratio is less than one if and only if the sum of the elasticities is above one, namely if output displays increasing returns to $(L_c, N_{c,t-1}, G_{c,t-1})$. This condition determines whether the network is more productive than the denser city of not. The same condition with different elasticities applies to unskilled workers' welfare.

Increasing returns always make a single denser city better than the network. A rise in θ mitigates this preference by making the network relatively more appealing, because increasing returns also imply gains from inequality; but such a rise can never induce a preference reversal. Similarly, decreasing returns always make the network better than a single a denser city. A rise in θ again mitigates this preference by making the network relatively less appealing, because decreasing returns imply losses from inequality; but again such a rise cannot induce a preference reversal. Changes in θ affect the intensity of preferences for density, but not their sign.

However, a larger city is unambiguously preferred to a denser one because it enjoys the same gains from agglomeration but also benefits from a greater endowment of land. This explains why increases in θ can switch preferences from the network to a large city. Such a preference reversal takes place in a context of decreasing returns and losses from inequality. When θ is low the network is so much better than a denser city that it is also better than a larger one. As θ rises, the appeal of consolidation rises until a denser city is only a slightly worse than the network, and a larger one is better. The gain from undoing uneven density across the network more than offsets the losses from decreasing returns when average density does not have to increase.

The same reasoning apply to densification as soon as $\omega > 0$. In that case, increases in either variable always make it weakly more likely that the single city, whether denser or larger, is preferred to the network. This does not mean that increases in θ always increase the appeal of the city relative the network, but rather than when there is a range of parameters such that the network might be preferred to the city, increases in θ make the city more appealing and can trigger a change in the preference ordering. Just as in the case where $\omega = 0$, however, there are also parameter values such that the city is always preferred to the network, and in that case increases in θ reduce the city's advantage relative to the network

Proposition 9 gives us another means of understanding the relative appeal of networks in the Randstad and networks in Texas. European cities are generally more homogeneous, along many dimensions, than U.S. cities, which can differ dramatically in skill levels, capital investment and density. For example, Yonkers differs dramatically from Manhattan in many ways, which suggests that it is a good thing that Yonkers residents can commute to Manhattan. The homogeneity of many European urban networks suggests that the gains from consolidation are lower.

6.2 Urban Heterogeneity in the Long Run

We now extend this case to the long run, and we focus on the case where there is a steady state. We assume that there is a measure-zero set of cities that are considering merging. We must limit our analysis to the case in which their endowments of firms and amenities remain

finite. In that case, Lemma 1 follows.

Lemma 1 *If the economy consists of R regions and each region r comprises both measure C_r of identical cities and also measure zero of cities whose initial endowments $(L_c, T_c; N_{c,0}, G_{c,0})$ differ from those of the representative cities, then the economy converges to a steady state in which aggregate variables are independent of non-representative cities if and only if*

$$(1-\mu)\,\sigma > \eta\,(1-\zeta\tau+\gamma_H)\,.$$

Each non-representative city has a steady-state number of firms

$$\bar{N}_c = \Xi_r^N \left[L_c^{(\sigma-1)(1-\zeta\tau+\gamma_H)} T_c^{\sigma\zeta\tau}\right]^{\frac{\eta}{(1-\mu)\sigma-\eta(1-\zeta\tau+\gamma_H)}}$$

and steady-state amenities

$$\bar{G}_c = \Xi_r^G \left[L_c^{(1-\mu)(\sigma-1)} T_c^{\eta\zeta\tau}\right]^{\frac{1}{(1-\mu)\sigma-\eta(1-\zeta\tau+\gamma_H)}},$$

where the steady-state constants Ξ_r^N and Ξ_r^G are independent of conditions in the set of non-representative cities.

It is more difficult to guarantee convergence to a steady state when there are non-representative cities. When the cities are all intrinsically identical, ongoing migration to a single city is less likely than when cities differ. The migration that becomes more likely with heterogeneity means that in the long run all of the population might crowd into a particularly attractive city, and that reduces the possibility for a steady state in which a positive measure of cities continue to host skilled entrepreneurs, firms, and production. Consequently, the assumption in Lemma 1 is more restrictive than either Assumption 1 or $\mu < 1$, which guarantee convergence to a steady state when all cities in a region are identical.

Firms make entrepreneurs more productive and thus tend to persist. If $\mu < 1$, their persistence is insufficient to break stationarity at the aggregate level. However, at the city level firms also attract more entrepreneurs from other cities, which can mean that a city with greater initial endowments will eventually tend to attract the region's entire skilled population. Consequently, to have a steady state in which the full-measure set of representative cities do not disappear, the level of firm persistence must be lower, and housing prices must rise fast enough with the influx of high-skill agents.

When the condition in Lemma 1 is satisfied, we can extend Proposition 7 to consider steady-state changes.

Proposition 10 *If $(1-\mu)\,\sigma > \eta\,(1-\zeta\tau+\gamma_H)$, then if a finite set of identical cities consolidate into a single larger commuting zone their residents will enjoy greater average amenities in steady state. The zone will attract more skilled entrepreneurs than the network in steady state if and only if*

$$\gamma_H + (1-\zeta\tau+\gamma_H)\,\frac{\mu+\eta-1}{(1-\mu)\,\sigma} > 0.$$

Consolidation will increase the steady-state number of firms, output, housing stock, wages, house prices and land rents if and only if

$$\mu + \eta\,(1+\gamma_H) > 1.$$

Consolidation will increase the steady-state utility of the less killed if and only if

$$\gamma_L + (1 - \zeta\tau + \gamma_L) \frac{\mu + \eta(1 + \gamma_H) - 1}{(1 - \mu)\sigma - \eta(1 - \zeta\tau + \gamma_H)} > 0.$$

The steady state has two differences from the short run. First, changes in firm creation and income are magnified by a factor $1/(1-\mu)$. Second, amenities are fully endogenous and no longer depend on the initial persistence of existing amenities. This also magnifies the impact of income, because income ultimately determines the steady-state stock of amenities. Both changes mean that in the long run, the impact of consolidation depends more on whether $\mu + \eta > 1$.

The condition for whether consolidation will increase output and the other nominal variables is the same as in the short run. The magnitude of the impact expands in the long run, but the sign does not change. This result would change if we assumed that migration took time, and in that case, if $\mu + \eta(1 + \gamma_H) > 1 > \mu + \eta$, which means that output falls without migration but rises with migration, it would be to have a long run effect that has a different sign from the short run effect.

Moving from the short run to the long run also amplifies the role that income changes have on utility. For example, if $\mu + \eta(1 + \gamma_H) < 1$ and income falls with consolidation, then it will be harder for changes in amenities to make up for that shortfall in the long run than in the short run.

The next proposition deals with densification into a single city.

Proposition 11 *If $(1 - \mu)\sigma > \eta(1 - \zeta\tau + \gamma_H)$, then if a finite set of identical cities shrink into a single denser commuting zone with the land of only one city their residents will enjoy greater average amenities but pay higher house prices and land rents in steady state. The zone will attract more skilled entrepreneurs than the network in steady state if and only if*

$$\gamma_H - \zeta\tau + (1 - \zeta\tau + \gamma_H) \frac{\eta + \mu - 1}{(1 - \mu)\sigma} > 0.$$

Densification will increase the steady-state number of firms, output and wages if and only if

$$\mu + \eta(1 - \zeta\tau + \gamma_H) > 1.$$

Densification will increase the steady-state utility of the less killed if and only if

$$\gamma_L - \zeta\tau + (1 - \zeta\tau + \gamma_L) \frac{\mu + \eta(1 - \zeta\tau + \gamma_H) - 1}{(1 - \mu)\sigma - \eta(1 - \zeta\tau + \gamma_H)} > 0.$$

Densification will increase the steady-state housing stock if and only if

$$\mu + \eta(1 - \zeta\tau + \gamma_H) - 1 > \zeta(1 - \mu)(\sigma - 1).$$

The condition that determines whether densification increases income and other nominal variables is the same in the long run as in the short run. Just as with consolidation, in steady state the change in the stock of new firms is magnified. The sign of the impact of

densification on income becomes more important in determining whether welfare increases with densification because of the long-run endogeneity of amenities.

The magnification of long-run income changes also means that the housing stock is unambiguously more likely to rise with density in the steady state than in the short run. If the immediate increase in income compensates for the decline in the endowment of land, then there will be an even larger long-run increase in the housing stock. Naturally, this will be accompanied by large increases in housing costs.

Finally, we address the role of heterogeneity in factor endowments. Proposition 9 applies identically to the steady state, but the correlation (ω) between endowments of labor, past firms and historic amenities becomes irrelevant because neither firms nor amenities are persistent in the long run. As a consequence, from a long-run perspective a city's endowment is truly two-dimensional, and to consider correlation we must allow endowments of both land and labor to vary across cities.

When we extend the empirical distribution of Proposition 9 we obtain the following results, which are particularly relevant for the steady state but once again apply identically to the short run.

Corollary 3 *Consider a finite set of cities, the distribution of whose relative endowments of land, labor, firms and amenities is a mixture of a perfectly correlated distribution, with weight $1 - \omega$, and four independent marginal distributions, with weight ω. Each of these distributions is a linear combination of equal shares, with weight $1 - \theta$, and any unequal distribution of shares, with weight $\theta \in [0, 1]$.*

If forming a single larger city attracts more skilled entrepreneurs; or creates more firms, produces greater output and pays higher wages; or provides higher utility to unskilled workers; or entails a larger housing stock, then it also does so for any $\theta' \geq \theta$ and any $\omega' \geq \omega$.

If forming a single denser city attracts more skilled entrepreneurs; or creates more firms, produces greater output and pays higher wages; or provides higher utility to unskilled workers; or entails a larger housing stock, then it also does so for any $\omega' \geq \omega$.

Corollary 3 echoes the results on heterogeneity in Proposition 9 except for the relationship between densification and θ. Higher values of ω always make both densification and consolidation more appealing relative to the network. Higher values of σ also always make consolidation more appealing relative to the network. These results are the same when land is uniformly distributed and when the endowment of land is heterogeneous across cities.

With densification, however, the results on size heterogeneity become ambiguous once they apply to land endowments as well. Heterogeneity (θ) has an ambiguous effect on the trade-off between a dense mega-city and a network, and its sign depends on the correlation (ω) between endowments of different factors.

When $\omega = 0$, the ratio of output in a network of m cities to output in a single larger city equals

$$m \sum_{k=1}^{1} f_k \left(\frac{1-\theta}{m} + \theta \hat{s}_k \right)^{\varepsilon_{Y,T}+\varepsilon_{Y,L}+\varepsilon_{Y,N}+\varepsilon_{Y,G}}. \tag{46}$$

Consolidation into a larger commuting zone is desirable if and only if the sum of elasticities is above one, regardless of θ. In this case, output (or welfare, *mutatis mutandis*) displays increasing returns to scale with respect to all four endowments. In the long run, this means

that steady-state output displays increasing returns to the combination of land and labor, given that it is independent of initial endowments of firms and amenities.

Increasing returns also mean that greater size heterogeneity makes consolidation less attractive. Again, the preference ordering between a network and a single larger commuting zone cannot change, but the magnitude of preferences shrinks as θ increases. Densification, however, is desirable if and only if the gains from consolidation are so large as to offset the loss of land (formally, if the ratio is below $m^{-\varepsilon_{Y,L}}$). Therefore, if $\omega = 0$ and the endowments of all factors including land are perfectly correlated, densification is going to be attractive if and only if size θ is sufficiently low.

When instead $\omega > 0$, the appeal of consolidation behaves in Corollary 3 just as in Proposition 9. There must be decreasing returns for the network ever to be desirable. When θ is large enough, a larger city dominates an idea network because the gains from eliminating heterogeneity in factor proportions are large enough to offset losses from diminishing returns.

The relationship between size heterogeneity and densification is then ambiguous. If ω is small, differences in factor proportions are second-order. Then there must be increasing returns—or else the network would always be better than a denser city—and preferences can only switch from the city to the network as θ increases. If ω is large, densification may trade-off instead losses from decreasing returns and gains from eliminating inequality in factor proportions, as with consolidation and as in Proposition 9. The city dominates the network as θ increases.

7 Conclusion

This paper has examined the trade-off between urban networks and mega-cities in a model with recombinant growth and endogenous amenities. In our framework, larger cities lead to more amenity spillovers, which improves welfare and will attract the skilled. New firm creation may either rise or fall when networks merge into a single city depending on whether there are increasing or decreasing returns to local scale in innovation. Densification also carries the cost of lost land.

Large mega-cities dominate networks when there are increasing returns to scale in idea creation, but even when are decreasing returns to scale, the extra amenities added in a large city can tip the balance in favor of urban size. This conclusion would be muted if we added extra disamenities of size that are not present in our model.

The attraction of denser, not larger, mega-cities is determined also by the elasticity of housing supply. When it is easy to add extra homes on a narrow plot of land, as in Texas, then density becomes more attractive. European urban networks may well be the right answer because history and regulation makes it so hard to build in Europe's older cities. Even though China has usually been quite friendly towards skyscrapers, the sheer scale of the Chinese population may still make the case for urban networks.

A running theme of this paper is that the long run may be different from the short run. It could be, for example, that urban networks are good for Europe in the short run, but that in the long run, Europe will suffer from the lack of new firm creation that would come from larger mega-cities. A second theme is that skilled and unskilled workers may have different preferences, where the skilled typically have a keener taste for mega-cities, as opposed to

networks, because they have a stronger taste for urban amenities.

Yet, while this paper offers a framework for evaluating the trade-off between networks and mega-cities, it is only a model. Its implications depend on parameters that need to be estimated empirically. Three major empirical facts would need to be confirmed to make any recommendations: the connection between urban size and amenities, housing supply elasticities, and returns to scale in idea creation. Moreover, since these parameters can surely differ between, for example, China and the United States, we would need to have place-specific estimates.

The logic of the spatial equilibrium suggests that real wages, or wages corrected for housing costs, can provide a measure of local amenities. Glaeser, Kolko and Saiz (2001) found that real wages were higher in cities in 1970, suggesting that amenities were decreased by urban consolidation of densification. By 2000, there was, if anything, a slightly negative relationship between real wages and metropolitan area population, which provides one possible source of information about the connection between amenities and city size.

We have estimates of housing supply elasticity from Saiz (2010). His estimates document how different these elasticities are in different parts of the United States, which only reinforces the difficulty of providing a single answer about whether densification is better or worse than urban networks. In Section 2 we discussed some of the evidence on returns to scale in idea creation. This evidence is far from definitive, however. We hope that future work will add the empirical meat needed to turn the theory in this paper into a better guide for urban policy.

References

[1] Alexander-Skipnes I (2007) *Cultural exchange between the Low Countries and Italy (1400-1600)*. Brepols, Turnhout

[2] Bettencourt LMA, Lobo J, Strumsky D (2007) Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. *Research Policy* 36 (1):107–120

[3] Christaller W (1933) *Die zentralen Orte in Süddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmäßigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen.* Gustav Fischer, Jena.

[4] Cohn D, Morin R (2008) American mobility: Who moves? Who stays put? Where's home? Pew Research Center social and demographics trends report

[5] Epifani P, Gancia G (2009) Openness, government size and the terms of trade. *Review of Economic Studies* 76: 629–668

[6] Ethier WJ (1979) Internationally decreasing costs and world trade. *Journal of International Economics* 9(1): 1–24

[7] Glaeser EL (2007) Entrepreneurship and the city. NBER working paper No. 13551

[8] Glaeser EL, Kahn M, Rappaport J (2008) Why do the poor live in cities? *Journal of Urban Economics* 63(1): 1–24

[9] Glaeser EL, Kerr WR (2009) Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain? *Journal of Economics and Management Strategy* 18(3): 623–663

[10] Glaeser EL, Kohlhase JE (2004) Cities, regions and the decline of transport costs. *Papers in Regional Science* 83(1): 197–228

[11] Glaeser EL, Kolko J, Saiz A (2001) Consumer city. *Journal of Economic Geography* 1: 27–50

[12] Glaeser EL, Tobio K (2008) The rise of the sunbelt. *Southern Economic Journal* 74: 610-643

[13] Howard MC (2012) *Transnationalism and society.* McFarland, Jefferson, NC

[14] Jacobs J (1969) *The economy of cities.* Random House, New York, NY

[15] Jefferson M (1939) The law of the primate city. *Geographical Review* 29: 226–232

[16] Mueller B (2000) From cities and towns to urban networks in Germany: Intra regional competition versus regional co-operation and inter regional competitiveness. *Urbani Izziv* 11(1): 134–138

[17] Overman HG, Bosquet C (2015) Why does birthplace matter so much? Sorting, learning and geography. London School of Economics working paper

[18] Saiz A (2010) The geographic determinants of housing supply. *Quarterly Journal of Economics* 125(3): 1253–1296

[19] Vermeule E (1960) The fall of the Mycenaean empire. *Archaeology* 13(1): 66–76

[20] Weitzman ML (1998) Recombinant growth. *Quarterly Journal of Economics* 113(2): 331–360

A Appendix

A.1. Proof of Proposition 1

By eq. (20),

$$\frac{\partial \ln N_t}{\partial \ln R} = 1 - \mu - \eta - \nu < 0 \text{ and } \frac{\partial \ln N_t}{\partial \ln C_r} = 1 - \mu - \eta. \tag{A1}$$

By eq. (21),

$$\frac{\partial \ln Y_t}{\partial \ln x} = \frac{1}{\sigma - 1} \frac{\partial \ln N_t}{\partial \ln x} \text{ for } x \in \{R, C_r\}. \tag{A2}$$

All other pecuniary variables increase if and only if the number of firms and output do: wages (eq. 3), profits per entrepreneur (eq. 8), the aggregate housing stock (eq. 13), house prices (eq. 14) and land rents (eq. 15).

By eqs. (22) and (23),

$$\frac{\partial \ln u_t^i}{\partial \ln x} = (1 - \zeta\tau + \phi\gamma_i) \frac{\partial \ln Y_t}{\partial \ln x} - \gamma_i \text{ for } i \in \{L, H\} \text{ and } x \in \{C_r, R\}, \tag{A3}$$

such that $\partial u_t^i / \partial R < 0$, while $\partial u_t^i / \partial C_r < 0$ if and only if the condition in Proposition 1 holds.

A.2. Proof of Proposition 2

Firm creation (eq. 20), output (eq. 21), wages (eq. 3) and profits per entrepreneur (eq. 8) do not depend on the endowment of land, so the condition for them to rise with fragmentation is the same as in Proposition 1.

By eqs. (22) and (23), $\partial \ln u_t^L / \partial \ln T = \partial \ln u_t^H / \partial \ln T = \zeta\tau$. Individual i's utility rises with concentration into denser commuting zones if and only if $\partial \ln u_t^i / \partial \ln C_r + \partial \ln u_t^i / \partial \ln T < 0$, and thus if and only if the condition in Proposition 2 holds.

Moreover, by eq. (13)

$$\frac{\partial \ln h_t}{\partial \ln C_r} + \frac{\partial \ln h_t}{\partial \ln T} = \zeta + (1 - \zeta) \frac{\partial \ln Y_t}{\partial \ln C_r} = \zeta + (1 - \zeta) \frac{1 - \mu - \eta}{\sigma - 1}, \tag{A4}$$

while by eqs. (14) and (15)

$$\frac{\partial \ln r_t}{\partial \ln C_r} + \frac{\partial \ln r_t}{\partial \ln T} = \zeta \left(\frac{\partial \ln \rho_t}{\partial \ln C_r} + \frac{\partial \ln \rho_t}{\partial \ln T} \right) = \zeta \left(\frac{\partial \ln Y_t}{\partial \ln C_r} - 1 \right) = \zeta \left(\frac{1 - \mu - \eta}{\sigma - 1} - 1 \right). \tag{A5}$$

A.3. Proof of Corollary 1

Firm creation (eq. 20), output (eq. 21), wages (eq. 3) and profits per entrepreneur (eq. 8) do not depend on the endowment of land, so they decline with R as in Proposition 1. A fortiori, house prices and land rents decline with R because

$$\frac{\partial \ln \rho_t}{\partial \ln R} + \frac{\partial \ln \rho_t}{\partial \ln T} = \frac{1}{\zeta} \left(\frac{\partial \ln r_t}{\partial \ln R} + \frac{\partial \ln r_t}{\partial \ln T} \right) = \frac{\partial \ln Y_t}{\partial \ln R} - 1 < \frac{\partial \ln Y_t}{\partial \ln R}. \tag{A6}$$

Individual i's utility rises with concentration into denser regions if and only if $\partial \ln u_t^i / \partial \ln R + \partial \ln u_t^i / \partial \ln T < 0$, and thus if and only if the condition in Corollary 1 holds.

Moreover, by eq. (13)

$$\frac{\partial \ln h_t}{\partial \ln R} + \frac{\partial \ln h_t}{\partial \ln T} = \zeta + (1-\zeta)\frac{\partial \ln Y_t}{\partial \ln R} = \zeta + (1-\zeta)\frac{1-\mu-\eta-\nu}{\sigma-1}, \tag{A7}$$

A.4. Proof of Proposition 3

The dynamics of the economy are characterized by a two-dimensional first-order vector autoregression:

$$\ln N_t = (1-\mu-\eta)\ln C_r + (1-\mu-\eta-\nu)\ln R + (\eta+\nu)\ln H + \mu \ln N_{t-1} + \ln \alpha_t \tag{A8}$$

and

$$\ln G_t = \phi\left[\ln \bar{\gamma} + \ln L + \frac{(1-\mu-\eta)\ln C_r + (1-\mu-\eta-\nu)\ln R + (\eta+\nu)\ln H}{\sigma-1}\right] + \frac{\mu\phi}{\sigma-1}\ln N_{t-1} + (1-\phi)\ln G_{t-1} + \frac{\phi}{\sigma-1}\ln \alpha_t + \phi \ln A_t + \ln \Gamma_t. \tag{A9}$$

By inspection, this process is stationary for $\mu < 1$, difference stationary for $\mu = 1$, and explosive for $\mu > 1$.

By eq. (24)

$$\frac{\partial \ln \bar{N}}{\partial \ln x} = \frac{1}{1-\mu}\frac{\partial \ln N_t}{\partial \ln x} \text{ for } x \in \{C_r, R\}, \tag{A10}$$

so the short-run effects of consolidation on all pecuniary variables are magnified in steady state.

By eqs. (26) and (27)

$$\frac{\partial \ln \bar{u}^i}{\partial \ln x} = (1-\zeta\tau+\gamma_i)\frac{\partial \ln Y_t}{\partial \ln x} - \gamma_i \text{ for } i \in \{L, H\} \text{ and } x \in \{C_r, R\}, \tag{A11}$$

so for an increase in R not only the short-run decline in output is magnified in steady state, but its welfare impact also rises for any $\phi < 1$.

For an increase in C_r, a preference reversal occurs if and only if eq. (28) holds, and therefore if and only if $\mu + \eta < 1$ and

$$\underline{\sigma}_s^i \equiv 1 + \left(\frac{1-\zeta\tau}{\gamma_i} + \phi\right)(1-\mu-\eta) < \sigma < \bar{\sigma}_s^i \equiv 1 + \left(\frac{1-\zeta\tau}{\gamma_i} + 1\right)\frac{1-\mu-\eta}{1-\mu}. \tag{A12}$$

The interval $(\underline{\sigma}_s^i, \bar{\sigma}_s^i)$ is a non-empty subset of $(1, \infty)$. By inspection,

$$\bar{\sigma}_s^i - \underline{\sigma}_s^i = (1-\mu-\eta)\left(\frac{1-\zeta\tau}{\gamma_i}\frac{\mu}{1-\mu} + \frac{1}{1-\mu} - \phi\right) \tag{A13}$$

is decreasing in η, ζ, τ, ϕ and γ_i.

A.5. Proof of Proposition 4

If $\mu + \eta > 1$ then $\partial \bar{Y}/\partial C_r < \partial Y_t/\partial C_r < 0$, so all pecuniary variables rise with city density in the short run, and more so in steady state. A preference reversal occurs if and only if eq. (29) holds, and therefore if and only if $\zeta\tau > \gamma_i$ and

$$\underline{\sigma}^i_{d^+} \equiv 1 + \frac{1 - \zeta\tau + \phi\gamma_i}{\zeta\tau - \gamma_i}(\mu + \eta - 1) < \sigma < \bar{\sigma}^i_{d^+} \equiv 1 + \frac{1 - \zeta\tau + \gamma_i}{\zeta\tau - \gamma_i}\frac{\mu + \eta - 1}{1 - \mu}. \tag{A14}$$

The interval $\left(\underline{\sigma}^i_{d^+}, \bar{\sigma}^i_{d^+}\right)$ is a non-empty subset of $(1, \infty)$. By inspection,

$$\bar{\sigma}^i_{d^+} - \underline{\sigma}^i_{d^+} = \left[\frac{\mu(1 - \zeta\tau)}{1 - \mu} + \left(\frac{1}{1 - \mu} - \phi\right)\gamma_i\right]\frac{\mu + \eta - 1}{\zeta\tau - \gamma_i} \tag{A15}$$

is increasing in μ, η and γ_i, and decreasing in ζ, τ and ϕ.

If $\mu + \eta < 1$ then $\partial \bar{Y}/\partial C_r > \partial Y_t/\partial C_r > 0$, so all pecuniary variables fall with city density in the short run, and more so in steady state. A preference reversal occurs if and only if neither inequality in eq. (29) holds, and therefore if and only if $\gamma_i > \zeta\tau$ and

$$\underline{\sigma}^i_{d^-} \equiv 1 + (1 - \zeta\tau + \phi\gamma_i)\frac{1 - \mu - \eta}{\gamma_i - \zeta\tau} < \sigma < \bar{\sigma}^i_{d^-} \equiv 1 + \frac{(1 - \zeta\tau + \gamma_i)}{\gamma_i - \zeta\tau}\frac{1 - \mu - \eta}{1 - \mu}. \tag{A16}$$

The interval $\left(\underline{\sigma}^i_{d^-}, \bar{\sigma}^i_{d^-}\right)$ is a non-empty subset of $(1, \infty)$. By inspection,

$$\bar{\sigma}^i_{d^-} - \underline{\sigma}^i_{d^-} = \left[\frac{\mu(1 - \zeta\tau)}{1 - \mu} + \left(\frac{1}{1 - \mu} - \phi\right)\gamma_i\right]\frac{1 - \mu - \eta}{\gamma_i - \zeta\tau} \tag{A17}$$

is decreasing in η, ϕ and γ_i, and increasing in ζ and τ.

A.6. Proof of Corollary 2

Since $\mu + \eta + \nu > 1$, all pecuniary variables rise with regional density in the short run, and more so in steady state: $\partial \bar{Y}/\partial R < \partial Y_t/\partial R < 0$. A preference reversal occurs if and only if

$$\begin{aligned}\frac{\partial \ln u^i_t}{\partial \ln R} + \frac{\partial \ln u^i_t}{\partial \ln T} &= \zeta\tau - \gamma_i - (1 - \zeta\tau + \phi\gamma_i)\frac{\mu + \eta + \nu - 1}{\sigma - 1} > 0 \\ &> \frac{\partial \ln \bar{u}^i}{\partial \ln R} + \frac{\partial \ln \bar{u}^i}{\partial \ln T} = \zeta\tau - \gamma_i - (1 - \zeta\tau + \gamma_i)\frac{\mu + \eta + \nu - 1}{(1 - \mu)(\sigma - 1)},\end{aligned} \tag{A18}$$

and therefore if and only if $\zeta\tau > \gamma_i$ and

$$\underline{\sigma}^i_D \equiv 1 + \frac{1 - \zeta\tau + \phi\gamma_i}{\zeta\tau - \gamma_i}(\mu + \eta + \nu - 1) < \sigma < \bar{\sigma}^i_D \equiv 1 + \frac{1 - \zeta\tau + \gamma_i}{\zeta\tau - \gamma_i}\frac{\mu + \eta + \nu - 1}{1 - \mu}. \tag{A19}$$

The interval $(\underline{\sigma}^i_D, \bar{\sigma}^i_D)$ is a non-empty subset of $(1, \infty)$. By inspection,

$$\bar{\sigma}^i_D - \underline{\sigma}^i_D = \left[\frac{\mu(1 - \zeta\tau)}{1 - \mu} + \left(\frac{1}{1 - \mu} - \phi\right)\gamma_i\right]\frac{\mu + \eta + \nu - 1}{\zeta\tau - \gamma_i} \tag{A20}$$

is increasing in μ, η, ν and γ_i, and decreasing in ζ, τ and ϕ.

A.7. Proof of Proposition 5

The evolution of firms (eq. A8) is independent of that of amenities. If and only if $\mu = 1$ it is a difference stationary process whose time-invariant mean growth (eq. 30) implies

$$\frac{\partial}{\partial \ln R}\mathbb{E}\left(\ln N_t - \ln N_{t-1}\right) = -\eta - \nu < \frac{\partial}{\partial \ln C_r}\mathbb{E}\left(\ln N_t - \ln N_{t-1}\right) = -\eta \text{ for all } t. \tag{A21}$$

A.8. Proof of Proposition 6

With asymmetric regions, spillovers are described by eq. (36) while for each region

$$\frac{\partial \ln Y_{r,t}}{\partial \ln N_{r,t}} = \frac{1}{\sigma}\left(1 + \frac{1}{\sigma - 1}\frac{Y_{r,t}}{Y_t}\right). \tag{A22}$$

If $\mu + \eta > 1 \Leftrightarrow \partial N_{r,t}/\partial C_r < 0$, denser commuting zones increase all incomes. A coordination failure occurs if and only if eq. (37) holds, and thus if and only if $\zeta\tau > \gamma_i$; moreover

$$(\mu + \eta - 1)\frac{1 - \zeta\tau + \gamma_i\phi}{\zeta\tau - \gamma_i} < \sigma < 1 + (\mu + \eta - 1)\frac{1 - \zeta\tau + \gamma_i\phi}{\zeta\tau - \gamma_i}, \tag{A23}$$

an interval whose intersection with $(1, \infty)$ is non-empty; and finally

$$\frac{Y_{r,t}}{Y_t} < \hat{y}^i_{d^+} \equiv (\sigma - 1)\left(\frac{\sigma}{\mu + \eta - 1}\frac{\zeta\tau - \gamma_i}{1 - \zeta\tau + \gamma_i\phi} - 1\right), \tag{A24}$$

a positive threshold that by inspection is increasing in σ, ζ and τ, and decreasing in μ, η, ϕ and γ_i.

If $\mu + \eta < 1 \Leftrightarrow \partial N_{r,t}/\partial C_r > 0$, denser or larger commuting zones increase all incomes. A coordination failure over density occurs if and only if both inequalities in eq. (37) fail, and thus if and only if $\gamma_i > \zeta\tau$; moreover

$$(1 - \mu - \eta)\frac{1 - \zeta\tau + \gamma_i\phi}{\gamma_i - \zeta\tau} < \sigma < 1 + (1 - \mu - \eta)\frac{1 - \zeta\tau + \gamma_i\phi}{\gamma_i - \zeta\tau}, \tag{A25}$$

an interval whose intersection with $(1, \infty)$ is non-empty; and finally

$$\frac{Y_{r,t}}{Y_t} < \hat{y}^i_{d^-} \equiv (\sigma - 1)\left(\frac{\sigma}{1 - \mu - \eta}\frac{\gamma_i - \zeta\tau}{1 - \zeta\tau + \gamma_i\phi} - 1\right), \tag{A26}$$

a positive threshold that by inspection is increasing in σ, μ, η and γ_i, and decreasing in ζ, τ and ϕ.

A coordination failure over size occurs if and only if

$$\begin{aligned}\frac{\partial \ln u^i_{r,t}}{\partial \ln C_r} = (1 - \zeta\tau + \gamma_i\phi)\frac{1 - \mu - \eta}{(\sigma - 1)\sigma}\left(\sigma - 1 + \frac{Y_{r,t}}{Y_t}\right) - \gamma_i \\ < 0 < \sum_{s=1}^{R}\frac{\partial \ln u^i_{r,t}}{\partial \ln C_s} = (1 - \zeta\tau + \gamma_i\phi)\frac{1 - \mu - \eta}{\sigma - 1} - \gamma_i,\end{aligned} \tag{A27}$$

and thus if and only if

$$(1-\mu-\eta)\,\frac{1-\zeta\tau+\gamma_i\phi}{\gamma_i}<\sigma<1+(1-\mu-\eta)\,\frac{1-\zeta\tau+\gamma_i\phi}{\gamma_i}, \tag{A28}$$

an interval whose intersection with $(1,\infty)$ is non-empty; and moreover

$$\frac{Y_{r,t}}{Y_t}<\hat{y}_s\equiv(\sigma-1)\left(\frac{\sigma}{1-\mu-\eta}\frac{\gamma_i}{1-\zeta\tau+\gamma_i\phi}-1\right), \tag{A29}$$

a positive threshold that by inspection is increasing in σ, μ, η, ζ, τ and γ_i, and decreasing in ϕ.

A.9. Derivation of Equations (41) and (42)

Plugging eq. (40) into eq. (7) yields eq. (41) for

$$\Omega^N_{r,t}\equiv\alpha_{r,t}\frac{H_r^{\eta+\nu}}{C_r^{\eta}} \times\left\{\left[\left(\frac{N_{r,t-1}}{C_r}\right)^{\mu}\left(\frac{L_r}{C_r}\right)^{\sigma-1}\right]^{\frac{1-\zeta\tau+\phi\gamma_H}{\sigma}}\left(\frac{T_r}{C_r}\right)^{\zeta\tau}\left(\frac{G_{r,t-1}}{C_r}\right)^{(1-\phi)\gamma_H}\right\}^{-\frac{\sigma\eta}{\sigma-\eta(1-\zeta\tau+\phi\gamma_H)}}. \tag{A30}$$

These last two with eqs. (2), (11) and (16) yield eq. (42) for

$$\Omega^u_{r,t}\equiv\frac{\sigma-1}{\sigma}\left(\frac{\Psi_{r,t}}{\tau^{\zeta}}\right)^{\tau}\left(\bar{\gamma}^{\phi}\Gamma_{r,t}\right)^{\gamma_L}\left(Y_t\Omega^N_{r,t}A^{\sigma-1}_{r,t}\right)^{\frac{1-\zeta\tau+\phi\gamma_L}{\sigma}}. \tag{A31}$$

A.10. Proof of Propositions 7 and 8

For any outcome variable X that is isoelastic in the endowments $(L_c,T_c;N_{c,t-1},G_{c,t-1})$, the ratio of its sum in a finite set of $m>1$ cities to its value in a single commuting zone having the sum of their endowments can be written

$$\lambda_X=\sum_{c=1}^{m}\prod_{F\in\{L,T,G,N\}}s_{F,c}^{\varepsilon_{X,F}}, \tag{A32}$$

where $s_{F,c}$ denotes city c's share of the joint endowment of factor F, such that $\sum_{c=1}^{m}s_{F,c}=1$, while each outcome variable is characterized by its elasticities $\varepsilon_{X,F}$ with respect to each factor.

Forming a single wider commuting zone thus raises the outcome variable if and only if $\lambda_X<1$. If all m cities are identical, then the outcome rises if and only if

$$\ln\lambda_X=\left(1-\sum\nolimits_F\varepsilon_{X,F}\right)\ln M<0\Leftrightarrow\Sigma_X\equiv\sum\nolimits_F\varepsilon_{X,F}>1. \tag{A33}$$

If instead of the sum of the land endowments the single commuting zone has the average endowment of land, the ratio is $m^{\varepsilon_{X,T}}\lambda_X<\lambda_X$. Forming a single denser commuting zone then raises the outcome variable if and only if

$$\Sigma_X-\varepsilon_{X,T}>1. \tag{A34}$$

For the number of skilled entrepreneurs, the sum of elasticities is

$$\Sigma_H = 1 + \frac{\sigma\gamma_H + (\mu + \eta - 1)(1 - \zeta\tau + \phi\gamma_H)}{\sigma - \eta(1 - \zeta\tau + \phi\gamma_H)} \tag{A35}$$

while the elasticity with respect to the land endowment is

$$\varepsilon_{H,T} = \frac{\sigma\zeta\tau}{\sigma - \eta(1 - \zeta\tau + \phi\gamma_H)}. \tag{A36}$$

For the number of firms,

$$\Sigma_N = \mu + \eta\Sigma_H = 1 + \sigma\frac{\mu + \eta(1 + \gamma_H) - 1}{\sigma - \eta(1 - \zeta\tau + \phi\gamma_H)} \tag{A37}$$

and

$$\varepsilon_{N,T} = \eta\varepsilon_{H,T} \Rightarrow \Sigma_N - \varepsilon_{N,T} = 1 + \sigma\frac{\mu + \eta(1 - \zeta\tau + \gamma_H) - 1}{\sigma - \eta(1 - \zeta\tau + \phi\gamma_H)}. \tag{A38}$$

For output,

$$\Sigma_Y = 1 + \frac{\Sigma_N - 1}{\sigma} \tag{A39}$$

and

$$\varepsilon_{Y,T} = \frac{\varepsilon_{N,T}}{\sigma} \Rightarrow \Sigma_Y - \varepsilon_{Y,T} = 1 + \frac{\Sigma_N - \varepsilon_{N,T} - 1}{\sigma} \tag{A40}$$

so output rises if and only if the number of firm does.

The average wage rises if and only is so does the wage bill

$$w_{c,t}L_c = \frac{\sigma - 1}{\sigma}Y_{c,t}, \tag{A41}$$

and thus if and only if output and the number of firms rise.

For the housing stock,

$$\Sigma_h = \zeta + (1 - \zeta)\Sigma_Y \tag{A42}$$

so the housing stock rises in a wider commuting zone if and only if so do output and the number of firms, while

$$\varepsilon_{h,T} = \zeta + (1 - \zeta)\varepsilon_{Y,T} \tag{A43}$$

such that

$$\Sigma_h - \varepsilon_{h,T} = (1 - \zeta)(\Sigma_Y - \varepsilon_{Y,T}) = (1 - \zeta)\frac{\sigma - 1 + \mu + \eta(1 - \phi)\gamma_H}{\sigma - \eta(1 - \zeta\tau + \phi\gamma_H)}. \tag{A44}$$

Average house prices rise if and only if

$$\frac{\sum_c r_{c,t}h_{c,t}}{\sum_c h_{c,t}} = \frac{\tau^\zeta \sum_c Y_{c,t}}{\Psi_{r,t}\sum_c T_c^\zeta Y_{c,t}^{1-\zeta}} < r_{M,t} = \frac{1}{\Psi_{r,t}}\left(\tau\frac{Y_{M,t}}{T_M}\right)^\zeta. \tag{A45}$$

For a wider commuting zone this condition can be written

$$\sum_c \prod_F s_{F,c}^{\varepsilon_{Y,F}} < \sum_c s_{T,c}^\zeta \prod_F s_{F,c}^{(1-\zeta)\varepsilon_{Y,F}}, \tag{A46}$$

and for identical cities

$$\Sigma_Y > \zeta + (1-\zeta)\,\Sigma_Y \Leftrightarrow \Sigma_Y > 1. \tag{A47}$$

For a denser commuting zone it can be written

$$M^{\varepsilon_{Y,T}} \sum_c \prod_F s_{F,c}^{\varepsilon_{Y,F}} < M^{\zeta+(1-\zeta)\varepsilon_{Y,T}} \sum_c s_{T,c}^{\zeta} \prod_F s_{F,c}^{(1-\zeta)\varepsilon_{Y,F}}, \tag{A48}$$

and for identical cities

$$\zeta\,(\Sigma_Y - \varepsilon_{Y,T}) > 0, \tag{A49}$$

which is always satisfied.

Average land rents rise if and only if

$$\frac{\sum_c \rho_{c,t} T_c}{\sum_c T_c} = \frac{\zeta\tau \sum_c Y_{c,t}}{\sum_c T_c} < \rho_{M,t} = \zeta\tau \frac{Y_{c,t}}{T_M}. \tag{A50}$$

For a wider commuting zone they rise if and only if so does output. For a denser commuting zone they rise if and only if

$$\lambda_Y < M \Leftrightarrow \Sigma_Y > 0, \tag{A51}$$

which is always satisfied.

Average amenities rise in a wider commuting zone if and only if

$$\lambda_G < M \Leftrightarrow \Sigma_G > 0, \tag{A52}$$

and in a denser commuting zone if and only if

$$\lambda_G < M^{1-\varepsilon_{G,T}} \Leftrightarrow \Sigma_G > \varepsilon_{G,T}, \tag{A53}$$

both of which are always satisfied.

The average utility of unskilled workers rises if and only if so does the sum of their utilities

$$U_{c,t}^L \equiv w_{c,t} L_c = \frac{\sigma - 1}{\sigma} \left[\Psi_{r,t} \left(\frac{T_c}{\tau} \right)^{\zeta} \right]^{\tau} \left(\bar{\gamma}^{\phi} \Gamma_{r,t} G_{c,t-1}^{1-\phi} \right)^{\gamma_L} Y_{c,t}^{1-\zeta\tau+\phi\gamma_L}, \tag{A54}$$

which has

$$\begin{aligned} \Sigma_U &= 1 + \gamma_L + (1 - \zeta\tau + \phi\gamma_L)\,(\Sigma_Y - 1) \\ &= 1 + \gamma_L + (1 - \zeta\tau + \phi\gamma_L)\,\frac{\mu + \eta\,(1+\gamma_H) - 1}{\sigma - \eta\,(1 - \zeta\tau + \phi\gamma_H)} \end{aligned} \tag{A55}$$

and

$$\varepsilon_{U,T} = \zeta\tau + (1 - \zeta\tau + \phi\gamma_L)\,\varepsilon_{Y,T} = \frac{\sigma - \eta\phi\,(\gamma_H - \gamma_L)}{\sigma - \eta\,(1 - \zeta\tau + \phi\gamma_H)} \zeta\tau. \tag{A56}$$

A.11. Proof of Proposition 9 and Corollary 3

If $s_{T,c} = 1/m$ for all c while the empirical distribution of $(s_{L,c}, s_{G,c}, s_{N,c})$ is described by eq. (44) then

$$\lambda_X = m^{1-\varepsilon_{X,T}} \times \left[\omega \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,L}} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,G}} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,N}} + (1-\omega) \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,L}+\varepsilon_{X,G}+\varepsilon_{X,N}}\right], \quad \text{(A57)}$$

such that

$$\frac{\partial \lambda_X}{\partial \omega} = m^{1-\varepsilon_{X,T}} \left[\sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,L}} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,G}} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,N}} - \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,L}+\varepsilon_{X,G}+\varepsilon_{X,N}}\right] \leq 0 \quad \text{(A58)}$$

because for any distribution S and any positive elasticities

$$\mathbb{E}s^{\varepsilon_{X,L}+\varepsilon_{X,G}+\varepsilon_{X,N}} \geq \mathbb{E}s^{\varepsilon_{X,L}} \mathbb{E}s^{\varepsilon_{X,G}+\varepsilon_{X,N}} \geq \mathbb{E}s^{\varepsilon_{X,L}} \mathbb{E}s^{\varepsilon_{X,G}} \mathbb{E}s^{\varepsilon_{X,N}} \quad \text{(A59)}$$

given that the covariance of a random variable and its monotone increasing transformation cannot be negative. It follows that

$$\lambda_X \leq m^{1-\varepsilon_{X,T}} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,L}+\varepsilon_{X,G}+\varepsilon_{X,N}} < m^{-\varepsilon_{X,T}} \text{ for all } \varepsilon_{X,L} + \varepsilon_{X,G} + \varepsilon_{X,N} > 1. \quad \text{(A60)}$$

Eq. (43) implies that

$$\frac{\partial \lambda_X}{\partial \theta} = \frac{m^{1-\varepsilon_{X,T}}}{\theta} \times \left[\begin{array}{c} \omega\varepsilon_{X,L} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,G}} \sum_{k=1}^{K} f_k s_k^{\varepsilon,N} \sum_{k=1}^{K} f_k \left(s_k - \frac{1}{m}\right) s_k^{\varepsilon_{X,L}-1} \\ +\omega\varepsilon_{X,G} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,L}} \sum_{k=1}^{K} f_k s_k^{\varepsilon,N} \sum_{k=1}^{K} f_k \left(s_k - \frac{1}{m}\right) s_k^{\varepsilon_{X,G}-1} \\ +\omega\varepsilon_{X,N} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,L}} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,G}} \sum_{k=1}^{K} f_k \left(s_k - \frac{1}{m}\right) s_k^{\varepsilon_{X,N}-1} \\ +(1-\omega)\left(\varepsilon_{X,L} + \varepsilon_{X,G} + \varepsilon_{X,N}\right) \sum_{k=1}^{K} f_k \left(s_k - \frac{1}{m}\right) s_k^{\varepsilon_{X,L}+\varepsilon_{X,G}+\varepsilon_{X,N}-1} \end{array}\right] \leq 0 \quad \text{(A61)}$$

for all $\varepsilon_{X,L} + \varepsilon_{X,G} + \varepsilon_{X,N} \leq 1$. Thus, either $\lambda_X < m^{1-\varepsilon_{X,T}}$ for all θ and ω—so the outcome X always rises in a denser commuting zone and a fortiori in a larger one—or λ_X is decreasing in θ as well as in ω.

If $s_{T,c}$ is also heterogeneous and the joint empirical distribution of $(s_{T,c}, s_{L,c}, s_{G,c}, s_{N,c})$ is described by

$$\Pr\left(s_{T,c} = s_i, s_{L,c} = s_j, s_{N,c} = s_k, s_{G,c} = s_l\right) = \omega f_i f_j f_k f_l + \mathbf{1}_{i=j=k=l}\left(1-\omega\right) f_i, \quad \text{(A62)}$$

then

$$\lambda_X = m\left[\omega \prod_{F} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,F}} + (1-\omega) \sum_{k=1}^{K} f_k s_k^{\Sigma_X}\right], \quad \text{(A63)}$$

such that, as above, $\partial\lambda_X/\partial\omega \leq 0$ and

$$\lambda_X \leq m \sum_{k=1}^{K} f_k s_k^{\Sigma_X} < 1 \text{ for all } \Sigma_X > 1. \tag{A64}$$

As above,

$$\frac{\partial \lambda_X}{\partial \theta} = \frac{m}{\theta} \left[\begin{array}{r} \omega \sum_F \varepsilon_{X,F} \sum_{k=1}^{K} f_k \left(s_k - \frac{1}{m}\right) s_k^{\varepsilon_{X,F}-1} \prod_{J \neq F} \sum_{k=1}^{K} f_k s_k^{\varepsilon_{X,J}} \\ + (1-\omega) \Sigma_X \sum_{k=1}^{K} f_k \left(s_k - \frac{1}{m}\right) s_k^{\Sigma_X - 1} \end{array} \right] \leq 0 \tag{A65}$$

for all $\Sigma_X \leq 1$. Thus, either $\lambda_X < 1$ for all θ and ω—so the outcome X always rises in a larger commuting zone—or λ_X is decreasing in θ as well as in ω.

A.12. Proof of Lemma 1

Firms in city c evolve according to

$$\begin{aligned} \ln N_{c,t} = \eta \frac{(\sigma-1)\left(1-\zeta\tau+\phi\gamma_H\right)\ln L_c + \sigma\zeta\tau \ln T_c}{\sigma - \eta\left(1-\zeta\tau+\phi\gamma_H\right)} \\ + \sigma \frac{\mu \ln N_{c,t-1} + \eta\left(1-\phi\right)\gamma_H \ln G_{c,t-1}}{\sigma - \eta\left(1-\zeta\tau+\phi\gamma_H\right)} + \ln \Omega_{r,t}^N \end{aligned} \tag{A66}$$

and amenities according to

$$\begin{aligned} \ln G_{c,t} = \phi \frac{(\sigma-1)\ln L_c + \eta\zeta\tau \ln T_c}{\sigma - \eta\left(1-\zeta\tau+\phi\gamma_H\right)} \\ + \frac{\mu\phi \ln N_{c,t-1} + (1-\phi)\left[\sigma - \eta\left(1-\zeta\tau\right)\right]\ln G_{c,t-1}}{\sigma - \eta\left(1-\zeta\tau+\phi\gamma_H\right)} + \ln \Omega_{r,t}^G. \end{aligned} \tag{A67}$$

The idiosyncratic component of these dynamics is deterministic and converges to a steady state if and only if the transition matrix of the difference equation,

$$\frac{1}{\sigma - \eta\left(1-\zeta\tau+\phi\gamma_H\right)} \begin{bmatrix} \mu\sigma & \eta\sigma\left(1-\phi\right)\gamma_H \\ \mu\phi & \left[\sigma - \eta\left(1-\zeta\tau\right)\right]\left(1-\phi\right) \end{bmatrix}, \tag{A68}$$

has two eigenvalues within the unit circle. The characteristic polynomial is

$$x^2 - \frac{\mu\sigma + \left[\sigma - \eta\left(1-\zeta\tau\right)\right]\left(1-\phi\right)}{\sigma - \eta\left(1-\zeta\tau+\phi\gamma_H\right)} x + \frac{\mu\sigma\left(1-\phi\right)}{\sigma - \eta\left(1-\zeta\tau+\phi\gamma_H\right)} = 0. \tag{A69}$$

Its discriminant is

$$\Delta^2 = \frac{\left\{\mu\sigma - \left[\sigma - \eta\left(1-\zeta\tau\right)\right]\left(1-\phi\right)\right\}^2 + 4\eta\mu\sigma\phi\left(1-\phi\right)\gamma_H}{\left[\sigma - \eta\left(1-\zeta\tau+\phi\gamma_H\right)\right]^2} > 0, \tag{A70}$$

so the transition matrix has two real eigenvalues. They are both positive by Assumption 1 and less than one if and only if moreover

$$\left(1-\mu\right)\sigma > \eta\left(1-\zeta\tau+\gamma_H\right). \tag{A71}$$

The steady-state firms and amenities in each non-representative city c in region r are jointly defined by

$$\bar{N}_c = \bar{\Omega}_r^N \left\{ \bar{N}_c^{\mu} \left[L_c^{\frac{\sigma-1}{\sigma}(1-\zeta\tau+\phi\gamma_H)} T_c^{\zeta\tau} \bar{G}_c^{(1-\phi)\gamma_H} \right]^{\eta} \right\}^{\frac{\sigma}{\sigma-\eta(1-\zeta\tau+\phi\gamma_H)}} \tag{A72}$$

and

$$\bar{G}_c = \bar{\Omega}_r^G \left\{ \left(\bar{N}_c^{\mu} L_c^{\sigma-1} T_c^{\eta\zeta\tau} \right)^{\phi} \bar{G}_c^{(1-\phi)[\sigma-\eta(1-\zeta\tau)]} \right\}^{\frac{1}{\sigma-\eta(1-\zeta\tau+\phi\gamma_H)}}, \tag{A73}$$

whose solution is provided in the statement of the Lemma for

$$\Xi_r^N \equiv \left[\left(\bar{\Omega}_r^N \right)^{\frac{\sigma-\eta(1-\zeta\tau+\gamma_H)}{\eta}} \left(\bar{\Omega}_r^G \right)^{\frac{\sigma(1-\phi)\gamma_H}{\phi}} \right]^{\frac{\eta}{(1-\mu)\sigma-\eta(1-\zeta\tau+\gamma_H)}} \tag{A74}$$

and

$$\Xi_r^G \equiv \left[\left(\bar{\Omega}_r^N \right)^{\mu} \left(\bar{\Omega}_r^G \right)^{\frac{(1-\mu)\sigma-\eta(1-\zeta\tau+\phi\gamma_H)}{\phi}} \right]^{\frac{1}{(1-\mu)\sigma-\eta(1-\zeta\tau+\gamma_H)}}. \tag{A75}$$

A.13. Proof of Propositions 10 and 11

By eqs. (40) the steady-state number of skilled entrepreneurs is proportional to

$$\bar{H}_c \propto \left[\left(\bar{N}_c^{\mu} L_c^{\sigma-1} \right)^{\frac{1-\zeta\tau+\phi\gamma_H}{\sigma}} T_c^{\zeta\tau} \bar{G}_c^{(1-\phi)\gamma_H} \right]^{\frac{\sigma}{\sigma-\eta(1-\zeta\tau+\phi\gamma_H)}}. \tag{A76}$$

By Lemma 1, the sum of its elasticities to the endowments of labor and land (L_c, T_c) is then

$$\Sigma_{\bar{H}} = 1 + \frac{(1-\mu)\sigma\gamma_H + (\eta+\mu-1)(1-\zeta\tau+\gamma_H)}{(1-\mu)\sigma - \eta(1-\zeta\tau+\gamma_H)}, \tag{A77}$$

while the elasticity to the land endowment alone is

$$\varepsilon_{\bar{H},T} = \frac{(1-\mu)\sigma\zeta\tau}{(1-\mu)\sigma - \eta(1-\zeta\tau+\gamma_H)}. \tag{A78}$$

For the steady-state number of firms from Lemma 1,

$$\Sigma_{\bar{N}} = 1 + \sigma \frac{\mu + \eta(1+\gamma_H) - 1}{(1-\mu)\sigma - \eta(1-\zeta\tau+\gamma_H)} \tag{A79}$$

and

$$\varepsilon_{\bar{N},T} = \frac{\eta\sigma\zeta\tau}{(1-\mu)\sigma - \eta(1-\zeta\tau+\gamma_H)}. \tag{A80}$$

For output, average wages, average house prices and average land rents the results are then unchanged from Propositions 7 and 8.

As in Proposition 7, the housing stock rises in a larger city if and only if the number of firms does. It rises in a denser city if and only if

$$\Sigma_{\bar{h}} - \varepsilon_{\bar{h},T} = (1-\zeta)\left(1 + \frac{\bar{\Sigma}_N - \bar{\varepsilon}_{N,T} - 1}{\sigma}\right) = \frac{(1-\zeta)(1-\mu)(\sigma-1)}{(1-\mu)\sigma - \eta(1-\zeta\tau+\gamma_H)} > 1. \tag{A81}$$

For steady-state amenities from Lemma 1,

$$\Sigma_{\bar{G}} = \frac{(1-\mu)(\sigma-1)+\eta\zeta\tau}{(1-\mu)\sigma-\eta(1-\zeta\tau+\gamma_H)} > \varepsilon_{\bar{G},T} = \frac{\eta\zeta\tau}{(1-\mu)\sigma-\eta(1-\zeta\tau+\gamma_H)} > 0, \tag{A82}$$

so as in Propositions 7 and 8 average amenities rise both in a larger and in a denser city.

The sum of unskilled workers' utilities is proportional to

$$\bar{U}_c^L \propto T_c^{\zeta\tau}\bar{G}_c^{(1-\phi)\gamma_L}\bar{Y}_c^{1-\zeta\tau+\phi\gamma_L}, \tag{A83}$$

which has

$$\begin{aligned}\Sigma_{\bar{U}} &= \zeta\tau + (1-\phi)\gamma_L\Sigma_{\bar{G}} + (1-\zeta\tau+\phi\gamma_L)\Sigma_{\bar{Y}} \\ &= \zeta\tau + (1-\phi)\gamma_L\Sigma_{\bar{G}} + (1-\zeta\tau+\phi\gamma_L)\left(1+\frac{\Sigma_{\bar{N}}-1}{\sigma}\right) \\ &= 1+\gamma_L + (1-\zeta\tau+\gamma_L)\frac{\mu+\eta(1+\gamma_H)-1}{(1-\mu)\sigma-\eta(1-\zeta\tau+\gamma_H)}\end{aligned} \tag{A84}$$

and

$$\begin{aligned}\varepsilon_{\bar{U},T} &= \zeta\tau + (1-\phi)\gamma_L\varepsilon_{\bar{G},T} + (1-\zeta\tau+\phi\gamma_L)\varepsilon_{\bar{Y},T} \\ &= \zeta\tau + (1-\phi)\gamma_L\varepsilon_{\bar{G},T} + \frac{1-\zeta\tau+\phi\gamma_L}{\sigma}\varepsilon_{\bar{N},T} = \frac{(1-\mu)\sigma-\eta(\gamma_H-\gamma_L)}{(1-\mu)\sigma-\eta(1-\zeta\tau+\gamma_H)}\zeta\tau.\end{aligned} \tag{A85}$$

Table 1 – Backward Transportation Linkages Across U.S. Metropolitan Areas

Destination Metro Area	Origin Metro Area	Origin's Share of Shipments to Destination	Destination's Share of Shipments to Origin	Industry of Largest Shipment from Origin to Destination
Tampa-St. Petersburg, FL	Seattle, WA	.507	<.001	Transportation Equipment (336)
Portland,OR	Seattle, WA	.463	.084	Transportation Equipment (336)
Memphis, TN	Seattle, WA	.430	.021	Transportation Equipment (336)
Baton Rouge, LA	New Orleans, LA	.420	.009	Petroleum (and Products) Wholesale (4247)
Virginia Beach, VA	Seattle, WA	.387	<.001	Transportation Equipment (336)
Jacksonville, FL	New York, NY	.316	.001	Chemical Manufacturing (325)
Charleston, SC	Greenville, SC	.315	.012	Transportation Equipment (336)
Savannah, GA	New Orleans, LA	.273	<.001	Raw Farm Products Wholesale (4245)
San Antonio, TX	Corpus Christi, TX	.235	.008	Petroleum Products Manufacturing (324)
Columbus, OH	Philadelphia, PA	.232	.002	Electronic Shopping and Mail Order (4541)

Notes: Value of physical shipments across the United States (Commodity Flow Survey 2012)