

Urban Growth Shadows*

David Cuberes

Klaus Desmet

Jordan Rappaport

Clark University

SMU, NBER and CEPR

Kansas City Fed

September 2018

Abstract

Does a location's growth benefit or suffer from being geographically close to large economic centers? Spatial proximity may lead to competition and hurt growth, but it may also generate positive spillovers and enhance growth. Using data on U.S. counties and metro areas for the period 1840-2016, we document this tradeoff between urban shadows and urban spillovers. The effect of proximity to large urban centers on growth was negative between 1840 and 1940, became positive between 1940 and 2000, and turned slightly negative again between 2000 and 2016. Using a simple spatial model that incorporates commuting costs, moving costs and agglomeration economies, we account for the observed patterns.

Keywords: urban shadows, agglomeration economies, spatial economics, urban systems, cities, United States, 1840-2016

JEL Codes: R12, N93

1 Introduction

Is the proximity of a large urban center beneficial or harmful for a location's economic growth? The presence of nearby clusters of economic activity can generate positive agglomeration spillovers by improving productivity and providing greater market access (Davis and Weinstein, 2002; Rosenthal and Strange, 2003; Glaeser and Kahn, 2004; Hanson, 2005; Redding and Sturm, 2008). However, it can also dampen growth through increased competition for resources (Fujita et al., 1999; Black and Henderson, 2003; Bosker and Buringh, 2017).

This paper analyzes this tradeoff between urban spillovers and urban shadows in the context of population growth in U.S. counties and urban areas over the period 1840-2016. We observe a pronounced differential impact as time goes by: the presence of a nearby large population center

*Cuberes: Department of Economics, Clark University. E-mail: dcuberes@clarku.edu; Desmet: Department of Economics and Cox School of Business, Southern Methodist University. E-mail: kdesmet@smu.edu; Rappaport: Federal Reserve Bank of Kansas City. E-mail: jordan.rappaport@kc.frb.org. Preliminary and Incomplete. The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

has a negative effect on a location's population up to the 1920s. Starting in the 1940s, and for the rest of the twentieth century, this effect becomes positive, before experiencing a reversal after the turn of the twenty-first century. That is, before the 1920s urban growth shadows dominated, between the 1940s and 2000 positive urban spillovers took over, and in the most recent time period urban growth shadows reappeared.

At least two forces shape spatial growth dynamics in the hinterland of large population clusters. First, transportation technology is key, in particular in how it relates to commuting and accessing urban centers. The widespread adoption of cars after 1940 allowed a dramatic increase in the distance between residences and jobs (Glaeser and Kahn, 2004). The construction of highways connecting the central business districts of the large cities with the hinterland further contributed to this trend (Baum-Snow, 2010). This did not only facilitate commuting, it also improved overall market access to large urban centers. Coinciding with time periods of improved transportation and commuting technology, we would expect to see the hinterland gaining. Second, the productivity advantage of size determines to a large extent the incentives to live in (or move to) dense urban areas. The structural transformation from agriculture to manufacturing provided a first push to increased urbanization in the nineteenth and early twentieth century. In recent decades, the growing importance of information and communication technology, especially in the service sector, is generating a renewed incentive for economic activity to spatially cluster (Desmet and Rossi-Hansberg, 2009; Gaspar and Glaeser, 1998). Coinciding with time periods when the productivity advantages of density strengthen, we would expect a relative increase in the incentive to move from the hinterland to large cities.

To highlight the role of these forces, we provide a simple conceptual framework of two cities. An individual has three choices: she can work in the city where she initially resides, she can move to live and work in the other city, and she can commute for work to the other city without changing her residence. We then show how these choices differ with the cost of commuting, the strength of local agglomeration economies, and the distance between cities. We find that as the cost of commuting gradually drops, individuals switch from staying put in the smaller, least productive city to moving to the larger, more productive city, and later to commuting to the larger, more productive city. As the productivity advantage of large cities increases, this enhances the relative incentive of moving, rather than commuting, to the large city. The intuition is straightforward: commuting is essentially a time cost, whereas moving is not. Therefore, if the large city gains in productivity, the time cost of commuting increases, relative to the cost of moving.

Our empirical findings are consistent with how these forces have changed over time. Before the introduction of the car and the building of suburban highways and railroads, dynamics were driven by the growing importance of density, as the economy was going through the structural transformation from agriculture to manufacturing. In line with the model, we see large urban centers gaining population at the expense of nearby smaller locations – a negative sign on having

a large city closeby for the period 1840-1940. After the 1940s, the transport technology revolution became the dominant force. It made commuting relatively more attractive to moving residence, implying faster growth in places close to urban centers. Again, this is consistent with what we see in the data – a positive sign on having a large city closeby for the period 1840-1940. In more recent decades, the introduction of ICT has increased the benefit of urban density. In addition, the improvements in commuting technology have petered out for now, and there is evidence of a rise in the opportunity cost of time, due to longer work hours and the increase in double-income families. As this is making commuting more expensive, our conceptual framework would predict slower growth in the hinterland of large cities. The greater productivity benefits from density and the rising inter-city commuting costs make moving to large dense areas more attractive relative to commuting long distances. The most recent data are consistent with this – a negative sign on having a large city closeby for the period 2000-2016.

This paper is related to the literature that explicitly considers the spatial location of one place relative to another. Urban economics has until recently largely ignored the spatial distribution of cities (Fujita et al., 1999). An important early exception is central-place-theory (Christaller, 1933; Lösch, 1940). In that theory the tradeoff between scale economies and transportation costs lead to the emergence of a spatially organized hierarchy of locations of different sizes. A natural implication of this framework is that the presence of large urban centers may enhance population growth in nearby agglomerations through positive spillover effects, but it may also limit such growth through competition among cities (Krugman, 1993; Tabuchi and Thisse, 2011).¹

Most empirical studies that explore the effect of large agglomerations on other locations focus on the twentieth century. They tend to find positive growth effects from proximity to urban centers. Using U.S. data, Partridge et al. (2009) uncover a positive effect of large urban clusters on nearby smaller places.² Looking at the post-war period, Rappaport (2005) finds evidence of the populations of cities and suburbs moving together. Dobkins and Ioannides (2001) also conclude that there has been a positive effect of neighboring locations on growth since the 1950s. Liu et al. (2011) analyze the case of China, and likewise show that the impact of a high-tier city on its surrounding areas is positive.

A few papers have looked at earlier time periods, and find evidence of urban growth shadows. In pre-industrial Europe, Bosker and Buringh (2017) show that the negative effect of neighbors dominate the positive one. Consistent with this, Rauch (2014) shows that historically larger European cities have been surrounded by larger hinterland areas and larger countryside populations. Most closely related to our work is Beltrán et al. (2017) who use data on Spanish municipalities for

¹More recently, there has been a growing interest in incorporating ordered space into economic geography models. This is particularly true of quantitative spatial models that aim to bring the theory to the data in meaningful ways (Desmet and Rossi-Hansberg, 2014; Allen and Arkolakis, 2014; Desmet, Nagy and Rossi-Hansberg, 2018).

²In an earlier study for the time period 1950-2000, the same authors find negative effects from proximity to higher-tiered places.

the time period 1800-2000. They find that the influence of neighboring cities was negative between 1800 and 1950, to then become increasingly positive from 1950 onwards. Our work focuses on the U.S., a country where the urbanization process is likely to have differed for a variety of reasons: it was much less settled in the nineteenth century, modern-day mobility across cities and regions is greater, and the adoption of the automobile was swifter. In addition, by including the most recent time period, our analysis shows that the positive effect of proximity during the post-war period has reversed since the turn of the twenty-first century. Our theoretical framework allows us to interpret these two switches over the last two centuries, first from urban shadows to urban spillovers, and then back to urban shadows.

The rest of the paper is organized as follows. Section 2 presents the empirical findings on the changes in urban growth shadows over the period 1840-2000. Section 3 proposes a conceptual framework that relates commuting costs, moving costs and local agglomeration economies to urban shadows. Section 4 explores the model’s predictions for urban shadows in light of the major changes in transport costs and agglomeration economies in the U.S. over the last two-hundred years, and shows that these predictions are consistent with our empirical findings.

2 Empirical Results

2.1 Data

We use population data from the Census Bureau at the level of counties and metro areas spanning the period 1840 to 2016. With the exception of the last period, we focus on successive twenty-year time frames: 1840-1860, 1860-1880, ..., 1980-2000, 2000-2016. In constructing the dataset, we had to resolve two main issues: how to deal with changing county borders and how to define metro areas going back in time. In what follows we limit ourselves to a brief discussion, and point the interested reader to Desmet and Rappaport (2014) for more details. To get consistent county borders, we use a “county longitudinal template” (CLT) augmented by a map guide to decennial censuses and combine counties as necessary to create geographically-consistent county equivalents over successive twenty-year-periods (Horan and Hargis, 1995; Thorndale and Dollarhide, 1987). For example, if county A splits into counties A_1 and A_2 in 1870, we combine counties A_1 and A_2 to measure population growth of county A between 1860 and 1880. More generally, for growth between 1840 and 1860, we use geographic borders from 1840; for growth between 1860 and 1880, we use geographic borders from 1860; and so on. This methodology gives us a separate dataset for each twenty-year period we study, as well as for 2000-2016.

Since our focus is on the interaction between closeby locations and since we are interested in urban shadows, it combines counties into metro areas, when and where they exist. Hence, our analysis is based on a hybrid of metro areas and non-metro counties. For years prior to 1960, we apply the criteria promulgated by the Office of Management and Budget in 1950 to population and

economic conditions *at the start* of each twenty-year period to construct metro areas (Gardner, 1999). As with the geographically-consistent counties, growth over any period is measured using the geographic borders of the initial year.

2.2 Growth and Spatial Proximity

Throughout most of U.S. history, locations' population growth rates have been negatively correlated with the presence of larger surrounding locations and positively correlated with the population in smaller surrounding locations. In other words, places tended to grow more quickly when there were no nearby larger locations attracting people away from them and when there were more people in nearby smaller locations that they could attract. The 1940-2000 period is an exception to this: during the second half of the twentieth century, locations' population growth rates became positively correlated with proximity to large locations.

Nearby larger locations. Table 1 reports results from regressing average annual population growth on dummy variables indicating the presence of a location at least 3 log points larger than an observation. So for a location with population of 1,000 in an initial year, a neighboring location with population of at least 20,100 in that year would meet this criterion. The dummies additionally depend on the distance of the larger neighbor to the observation. If there is a larger neighbor whose centroid is within 50 kilometers of the centroid of an observation, then the dummy for the 1-to-50 kilometer distance band is set to 1.³ If there is no neighbor within 50 kilometers that meets the threshold and there is such a neighbor with centroid between 50 and 100 kilometers away, the dummy for the 50-to-100 kilometer distance band is set to 1. If no neighbor within 100 kilometers meets the threshold and there is one with centroid between 100 and 150 kilometers away, the dummy for the 100-to-150 km dummy is set to 1. And so on.

Locations' population growth throughout U.S. history closely depended on their own population (Desmet and Rappaport, 2017). Thus each of the reported regressions also includes a population spline to control for this possibly nonlinear relationship. For example, the 1840-to-1860 regression includes a five-segment spline for a location's own population in 1840, with breaks at log population equal to 8, 9, 10, and 11. Splines for cross-sections in later years include more segments, reflecting the increasing size range of U.S. locations. We suppress reporting the coefficients on the population spline in order to focus on the relationship between population growth and neighboring locations.⁴

The negative correlation of population growth and neighboring larger locations holds for all

³Centroids of counties are geographic, based on the latitude and longitude reported in decennial censuses. Centroids of metropolitan areas are constructed as the population-weighted mean of the county geographic centroids.

⁴Coefficient values are similar to those reported in Desmet and Rappaport (2017) (which did not control for neighboring population)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Average Annual Population Growth								
location w/ pop 3 log points higher at distance:	1840 to 1860	1860 to 1880	1880 to 1900	1900 to 1920	1920 to 1940	1940 to 1960	1960 to 1980	1980 to 2000	2000 to 2016
1 to 50 km	-0.037 (0.023)	-0.002 (0.008)	0.000 (0.007)	-0.004 (0.004)	0.004 (0.003)	0.019 (0.003)	0.018 (0.003)	0.009 (0.003)	0.002 (0.002)
50 to 100 km	-0.018 (0.007)	-0.019 (0.005)	-0.005 (0.004)	-0.009 (0.002)	-0.003 (0.002)	0.002 (0.002)	0.006 (0.002)	0.005 (0.002)	-0.003 (0.001)
100 to 150 km	-0.021 (0.007)	-0.014 (0.003)	-0.010 (0.002)	-0.007 (0.002)	-0.003 (0.002)	-0.002 (0.002)	0.003 (0.002)	0.001 (0.002)	-0.004 (0.001)
150 to 200 km	-0.019 (0.006)	-0.015 (0.003)	-0.009 (0.002)	-0.008 (0.001)	-0.003 (0.001)	-0.002 (0.002)	0.001 (0.002)	-0.001 (0.001)	-0.004 (0.001)
200 to 300 km	-0.010 (0.005)	-0.014 (0.003)	-0.007 (0.002)	-0.006 (0.002)	-0.002 (0.001)	-0.005 (0.002)	-0.001 (0.001)	-0.002 (0.001)	-0.004 (0.001)
init pop spline segments	5	7	7	7	7	7	8	8	9
RHS variables	10	12	12	12	12	12	13	13	14
N	860	1,686	2,365	2,652	2,947	2,980	2,851	2,632	2,368
R ²	0.655	0.781	0.715	0.417	0.317	0.172	0.111	0.120	0.249
Spline R ²	0.617	0.761	0.706	0.396	0.307	0.127	0.042	0.076	0.223
mean growth	0.030	0.033	0.024	0.012	0.006	0.002	0.008	0.005	0.001
std deviation	0.036	0.046	0.035	0.024	0.017	0.018	0.014	0.012	0.009
U.S. pop growth	0.031	0.024	0.021	0.017	0.011	0.015	0.012	0.011	0.009

Table 1: **Population Growth Near Larger Locations, 1840-2016.**

The variable indicating a location with population 3 log points above that of an observation at a given distance is contingent on there being no location with population 3 log points above at a closer distance. All regressions include a constant and a spline of an observation's initial population; the number of spline segments varies by year depending on the population range of counties and metropolitan areas. Standard errors, in parentheses, are robust to spatial correlation. Dark blue and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels. Red font indicates a positive coefficient that statistically differs from zero at the 0.05 level.

distance bands in each of the twenty-year cross-sections beginning in 1840 through 1900. It also holds for all but the nearest band in the 1920 cross-section.⁵ Most of the corresponding (negative) coefficients statistically differ from zero at the 0.05 level (dark blue font) and are economically meaningful in magnitude. In 1840, for example, the presence of at least one neighbor 50-to-100 kilometers away with population 3 log points higher than that of an observation, together with the absence of such neighbors within 50 kilometers away, was associated with lower average annual population growth from 1840 to 1860 of 1.8 percentage points, equivalent to half the standard deviation of average annual growth during these years. The negative correlation is nearly the same for nearest neighbors with population 3 log points higher that are located 100-to-150 kilometers

⁵The pattern is similar for the twenty-year cross-section beginning in 1800, although only one of the coefficients statistically differs from zero. For the 1820 cross-section, there is a small positive coefficient on the 1-to-50 kilometer dummy.

and 150-to-200 kilometers away. In other words, the presence of at least one location within 200 kilometers that had population at least 3 log points higher than an observation was associated with slower growth of about 2 percentage points per year. The five neighbor dummies make a moderate contribution to accounting for the variation in growth from 1840 to 1860, boosting the regression R-squared to 0.655 from 0.617 with only the population spline.

The strength of the negative correlation between population growth and neighboring larger locations mostly diminishes over time. Coefficient values for the regression starting in 1860 are approximately the same as those in the 1840 regression. The coefficient value for the presence of a large neighbor 50-to-100 kilometers away in 1880 is moderately lower, rebounds in 1900, and declines to relatively close to zero in 1920.

The relationship between growth and neighboring larger locations reversed during the second half of the twentieth century, a period of intense suburbanization. Growth for the cross-sections beginning in 1940, 1960, and 1980 was positively correlated with the presence of a location within 50 km that had a population at least 3 log points larger. The coefficient values correspond to faster annual average growth of almost 2 percentage points from 1940 to 1980 and of almost 1 percentage point from 1980 to 2000. The positive correlation from 1960 to 2000 also extended out as far as 100 kilometers, albeit with diminished magnitude. Neighboring larger locations make a significant contribution to accounting for variations in growth from 1940 to 2000, boosting R-squared values by about 5 percentage points compared to regressions on the population spline only.

The relationship between growth and neighboring larger locations reverses again in 2000, with population growth once again being negatively correlated with the presence of neighboring locations with population 3 log points higher. The coefficient magnitudes, however, are relatively small as is the increase in R-squared value compared to controlling for only the initial population spline. Another, possibly related, characteristic of growth from 2000 to 2016 is the large share of variation accounted for by the population. The R-squared value of the spline alone is 0.249, more than double its value from the previous two twenty-year periods.

Nearby smaller locations. Table 2 reports results from regressing average annual population growth on the log of the aggregate population in locations at an indicated distance with population at least 2 log points smaller than that of an observation. In other words, suppose an observation with population of 1,000 had three neighboring locations within 50 kilometers of it, two with population of 135 each (the highest integer below 2 log points smaller) and one with higher population than this. In this case In this case the variable for the 1-to-50 kilometer distance band would equal $\log(270)$. If the observation also had single location between 50 and 100 kilometers with population of 135 and no smaller locations in this distance band, the variable for the 50-to-100 kilometer band would equal the $\log(135)$.

Population growth during each of the six periods twenty-year periods from 1840 to 1960

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Average Annual Population Growth								
log total pop in locations w/ pop 2 log points lower at distance:	1840 to 1860	1860 to 1880	1880 to 1900	1900 to 1920	1920 to 1940	1940 to 1960	1960 to 1980	1980 to 2000	2000 to 2016
1 to 50 km	0.0011 (0.0006)	0.0008 (0.0004)	0.0003 (0.0004)	0.0004 (0.0003)	0.0002 (0.0003)	0.0000 (0.0002)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)
50 to 100 km	0.0006 (0.0002)	0.0004 (0.0002)	0.0005 (0.0002)	0.0003 (0.0003)	0.0004 (0.0002)	0.0005 (0.0002)	-0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)
100 to 150 km	0.0010 (0.0003)	0.0010 (0.0002)	0.0004 (0.0002)	0.0001 (0.0001)	0.0003 (0.0002)	0.0003 (0.0002)	-0.0001 (0.0001)	-0.0002 (0.0001)	0.0001 (0.0000)
150 to 200 km	0.0005 (0.0003)	0.0008 (0.0002)	0.0003 (0.0002)	0.0004 (0.0002)	0.0002 (0.0001)	0.0003 (0.0002)	0.0000 (0.0001)	-0.0002 (0.0001)	0.0000 (0.0001)
200 to 300 km	0.0014 (0.0004)	0.0013 (0.0003)	0.0007 (0.0002)	0.0004 (0.0002)	0.0001 (0.0002)	0.0001 (0.0002)	-0.0002 (0.0002)	0.0000 (0.0002)	0.0002 (0.0001)
init pop spline segments	5	7	7	7	7	7	8	8	9
RHS variables	10	12	12	12	12	12	13	13	14
N	860	1,686	2,365	2,652	2,947	2,980	2,851	2,632	2,368
R ²	0.675	0.798	0.717	0.408	0.320	0.139	0.045	0.080	0.235
Spline R ²	0.617	0.761	0.706	0.396	0.307	0.127	0.042	0.076	0.223
mean growth	0.030	0.033	0.024	0.012	0.006	0.002	0.008	0.005	0.001
std deviation	0.036	0.046	0.035	0.024	0.017	0.018	0.014	0.012	0.009
U.S. pop growth	0.031	0.024	0.021	0.017	0.011	0.015	0.012	0.011	0.009

Table 2: Population Growth Near Smaller Locations , 1840-2016.

RHS variables measure the aggregate population in locations at indicated distance that have population at least 2 log points below that of an observation. All regressions include a constant and a spline of an observation's initial population; the number of spline segments varies by year depending on the population range of counties and metropolitan areas. Standard errors, in parentheses, are robust to spatial correlation. Dark blue and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels. Dark red and light red fonts respectively indicate a positive coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

was positively correlated with the total population in the smaller locations in each of the distance bands. In the first four of these periods, from 1840 to 1920, most of these correlations statistically differed from zero at the 0.05 level. In the latter two periods, from 1920 to 1960, most of these correlations statistically differed from zero at the 0.10 level. The magnitudes of the corresponding coefficients are modest but economically meaningful. For example, a 1 log point increase in the population of such smaller locations between 50 and 100 kilometers away in 1840 is associated with faster average annual growth of 0.06 percentage points, which is small compared to the 3.6 percent standard deviation of growth during this period. But the log total population living in such smaller locations 50 to kilometers has a standard deviation of 3.7, implying that a 1 standard deviation increase is associated with faster average population growth from 1840 to 1860 of 0.2 percentage point per year.

This positive correlation can be seen as the flip side of the negative correlation of growth

and the presence of larger locations. To the extent that the slower growth of locations in the shadow of larger ones represents the loss of their own residents to the larger locations, larger locations tend to grow faster if they have a larger population in smaller locations from which to attract migrants.

3 Conceptual Framework

In this section we model the forces that shape the spatial interaction between cities. On the one hand, being close to larger cities may provide agglomeration economies, through technology spillovers and market access. On the other hand, smaller cities may find it hard to survive in the shadow of larger cities, as their residents may prefer to move to their high-productivity neighbors. Depending on whether agglomeration economies dominate agglomeration shadows, smaller locations in the hinterland of large cities may experience faster or slower growth.

3.1 Setup

Endowments. The economy consists of a continuum of points on a line. The density of land at all points of the line is one. There are L individuals, each residing on one unit of land. Each resident has one unit of time, which she divides between work and commuting. On the line there are two exogenously given production points, indexed by ℓ or k . The set of individuals living closer to production point ℓ than to the other production point comprises city ℓ . The land rent in city ℓ at distance d_ℓ from production point ℓ is denoted by $r_\ell(d_\ell)$. The distance between production points ℓ and k , denoted by $d_{\ell k}$, is big enough so that there is at least some empty land between the two cities.⁶

Technology and preferences. The economy produces one good, and labor is the only factor of production. Technology is linear, with one unit of labor producing A_ℓ units of the good at production point ℓ and A_k units of the good at production point k . To produce, an individual needs to commute from his residence to one of the two production points.

We illustrate the structure of commuting costs by focusing on an agent residing in city ℓ at a distance d_ℓ of production point ℓ . To commute to production point in the own city ℓ , she has to cover a distance d_ℓ . To commute to the production point in the other city k , we ignore differences in the residence location in the own city. To be precise, to commute to k , we assume she has to cover a distance $d_{\ell k}$. This simplifying assumption has the advantage of maintaining symmetry between agents who reside at a distance d_ℓ to the right of production point ℓ and those who reside at that same distance d_ℓ to the left of production point ℓ . It also implies that cities will be symmetric in shape: the number of residents living to the right and to the left of production point ℓ will be the

⁶This ensures symmetry in a city's spatial structure on both sides of the production point.

same. That is, the distance from production point ℓ to the edge of city ℓ , denoted by \bar{d}_ℓ , will be the same on both sides of ℓ .

The time cost of commuting per unit of distance is γ . Hence, an individual who resides in city ℓ at a distance d_ℓ from production point ℓ supplies $1 - \gamma d_\ell$ units of labor if she works at production point ℓ and $1 - \gamma d_{\ell k}$ units of labor if she works at production point k . The corresponding productivity levels of that individual are $A_\ell(1 - \gamma d_\ell)$ if she works at production point ℓ and $A_k(1 - \gamma d_{\ell k})$ if she works at production point k . We denote by $w_\ell(d_\ell)$ the income of an individual residing at a distance d_ℓ from production point ℓ . After normalizing the price of the good the economy produces to one, $w_\ell(d_\ell)$ is equal to $A_\ell(1 - \gamma d_\ell)$ if the individual works at ℓ and is equal to $A_k(1 - \gamma d_{\ell k})$ if she works at k .

The possibility of commuting from city ℓ to production point k is meant to capture the positive effect of productivity spillovers from the neighboring city. These spillovers decline with distance: by commuting from city ℓ to production point k , a resident loses working time at a rate of γ per unit of distance, giving him access to a discounted version of the neighboring city's productivity, $A_k(1 - \gamma d_{\ell k})$. We could alternatively model this effect through direct technology spillovers, without the need of introducing commuting, as in Ahlfeldt et al. (2015): if technological spillovers decay at a rate of γ per unit of distance, then an agent in city ℓ would have access to the same discounted version of his neighbor's productivity, $A_k(1 - \gamma d_{\ell k})$. Hence, both interpretations are interchangeable in their effects on productivity.

The utility of an individual who lives in a city ℓ at a distance d_ℓ from the production point ℓ is denoted by $u_\ell(d_\ell)$. Utility is linear, so that $u_\ell(d_\ell) = w_\ell(d_\ell) - r_\ell(d_\ell)$. People can freely locate within cities. As a result, the utility of all individuals who reside in city ℓ equalizes. Where land is unoccupied, land rents are normalized to zero. Because of free mobility, at the edge of city ℓ , land rents are zero, so $r_\ell(\bar{d}_\ell) = 0$.

To change residence between cities ℓ and k , agents need to pay a cost equal to $\mu d_{\ell k}$, so that the utility cost of inter-city moves is increasing in the distance between cities. Rather than as a time cost, we interpret the moving cost as the utility cost of being a migrant. For example, this could include the psychological and social costs of having to leave friends and family. Consistent with this interpretation, we assume that a return migrant does not pay a moving cost. That is, if an individual who moved from city ℓ to city k returns to her hometown, she does not pay a moving cost. The possibility of moving between cities is meant to capture urban shadows: an individual who resides in a low-productivity smaller city in the proximity of a high-productivity larger city may find it beneficial to move to its high-productivity neighbor. If so, the smaller city loses population, so the larger city casts a growth shadow on the smaller city.

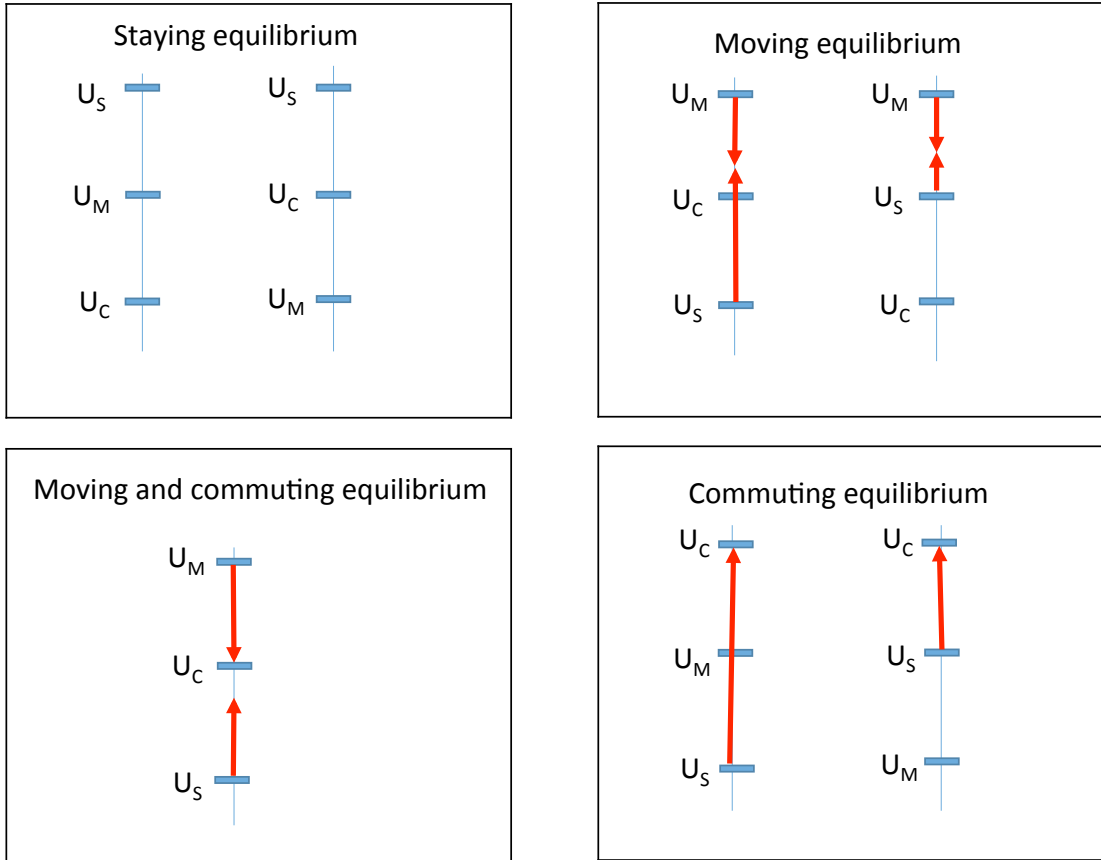
Residential and commuting choice. An individual who resides in city ℓ at a distance d_ℓ from production point ℓ has three choices: she can stay in city ℓ and work at production point ℓ ; she

can move to city k and reside at a distance $d(k)$ from her work at production point k ; or she can commute a distance $d_{\ell k}$ from city ℓ to production point k to work. The corresponding utility levels if she stays, $u_{\ell S}$, moves, $u_{\ell M}$, or commutes, $u_{\ell C}$, can be written as:

$$\begin{aligned}
 u_{\ell S} &= A_{\ell}(1 - \gamma d_{\ell}) - r_{\ell}(d_{\ell}) \\
 u_{\ell M} &= A_k(1 - \gamma d_k) - r_k(d_k) - \mu d_{\ell k} \\
 u_{\ell C} &= A_k(1 - \gamma d_{\ell k}) - r_{\ell}(d_{\ell}).
 \end{aligned} \tag{1}$$

These expressions suggest that staying is attractive if productivity differences are small, inter-city distances are large, commuting costs are high, and moving costs are big; moving is beneficial if commuting costs are not too high and moving costs are sufficiently low; and commuting is the preferred choice if commuting costs become sufficiently low.

Figure 1: Equilibrium Description



We now describe the equilibrium of the model for a given set of parameters A_{ℓ} , A_k , $d_{\ell k}$, μ and γ , and for given initial values \bar{d}_{ℓ}^0 and \bar{d}_k^0 . Without loss of generality, assume that $A_k(1 - \gamma \bar{d}_k^0) \geq A_{\ell}(1 - \gamma \bar{d}_{\ell}^0)$, implying that utility in city ℓ is less than or equal to utility in city k .

The equilibrium can be characterized by four different cases. First, if individuals in city ℓ

do not stand to gain from either moving to city k or commuting to city k , we will say that we are in a *staying equilibrium*: every individual stays and works in the city where she resided originally. This case is illustrated in the top-left panel of Figure 1. Second, if individuals in city ℓ get a higher utility from moving than from both commuting or staying, some individuals from city ℓ move to city k . As this happens, city ℓ becomes smaller and city k becomes larger, implying that the utility from moving goes down and the utility from staying goes up.⁷ If, as illustrated in the top-right panel of Figure 1, the two utility levels meet at a level above the utility from commuting or if everyone moves out of city ℓ before the two utility levels meet, then we will say that we are in an *inter-city moving equilibrium*: some (or all) individuals of ℓ move to k , and the remainder lives and works in ℓ . Third, starting in the same situation, with the utility from moving being higher than the utility from commuting or staying, it is possible that as people start moving, the utility from moving reaches the utility from commuting. At that point, some individuals in ℓ start commuting to k . In this case, shown in the bottom-left panel of Figure 1, the economy is in an *inter-city moving and commuting equilibrium*. Lastly, if the utility from commuting is higher than the utility of moving or staying, some individuals in ℓ commute to k . As this occurs, less people in ℓ work in production point ℓ . This weakens the competition for land in ℓ and lowers the land rent. As a result, the utility from staying increases, and the economy reaches an equilibrium when the utility from staying equalizes the utility from commuting.⁸ In this case, illustrated in the bottom-right panel of Figure 1, the economy is in an *inter-city commuting equilibrium*.

We are now ready to formally define the equilibrium for the four different cases that characterize it.

Equilibrium. Given A_ℓ , A_k , $d_{\ell k}$, μ and γ , and given initial values \bar{d}_ℓ^0 and \bar{d}_k^0 , with $A_k(1 - \gamma\bar{d}_k^0) > A_\ell(1 - \gamma\bar{d}_\ell^0)$, the economy will be in one of four equilibria:

- i. Staying equilibrium. If $A_\ell(1 - \gamma\bar{d}_\ell^0) \geq A_k(1 - \gamma\bar{d}_k^0) - \mu d_{\ell k}$ and $A_\ell(1 - \gamma\bar{d}_\ell^0) \geq A_k(1 - \gamma d_{\ell k})$, then no individual has an incentive to move from ℓ to k .
- ii. Inter-city moving equilibrium. If either $A_k(1 - \gamma\bar{d}_k^0) - \mu d_{\ell k} > A_\ell(1 - \gamma\bar{d}_\ell^0) \geq A_k(1 - \gamma d_{\ell k})$ or both $A_k(1 - \gamma\bar{d}_k^0) - \mu d_{\ell k} > A_k(1 - \gamma d_{\ell k}) > A_\ell(1 - \gamma\bar{d}_\ell^0)$ and $A_k(1 - \gamma(\bar{d}_k^0 + m)) - \mu d_{\ell k} \geq A_k(1 - \gamma d_{\ell k})$, then a share $\min(m, \bar{d}_\ell^0)$ moves from city ℓ to k , where m is the solution to $A_k(1 - \gamma(\bar{d}_k^0 + m)) - \mu d_{\ell k} = A_\ell(1 - \gamma(\bar{d}_\ell^0 - m))$.
- iii. Inter-city moving and commuting equilibrium. If $A_k(1 - \gamma\bar{d}_k^0) - \mu d_{\ell k} > A_k(1 - \gamma d_{\ell k}) > A_\ell(1 - \gamma\bar{d}_\ell^0)$ and $A_k(1 - \gamma(\bar{d}_k^0 + m)) - \mu d_{\ell k} < A_k(1 - \gamma d_{\ell k})$, then a share $\min(m', \bar{d}_\ell^0)$ people moves from

⁷The utility from commuting remains the same, since that utility depends on the distance between city ℓ and city k , which is unchanged.

⁸Of course, if everyone commutes before that equality is reached, then we would have the entire city ℓ commuting to k .

city ℓ to city k , where m' is the solution to $A_k(1 - \gamma(\bar{d}_k^0 + m')) - \mu d_{\ell k} = A_k(1 - \gamma d_{\ell k})$ and m is the solution to $A_k(1 - \gamma(\bar{d}_k^0 + m)) - \mu d_{\ell k} = A_\ell(1 - \gamma(\bar{d}_\ell^0 - m))$, and a share $\min(m'', \bar{d}_\ell^0 - \min(m', \bar{d}_\ell^0))$ people commute from city ℓ to city k , where m'' is the solution to $A_\ell(1 - \gamma(\bar{d}_\ell^0 - m' - m'')) = A_k(1 - \gamma d_{\ell k})$.

- iv. Inter-city commuting equilibrium. If $A_k(1 - \gamma d_{\ell k}) > A_\ell(1 - \gamma \bar{d}_\ell^0)$ and $A_k(1 - \gamma d_{\ell k}) > A_k(1 - \gamma \bar{d}_k^0) - \mu d_{\ell k}$, then $\min(c, \bar{d}_\ell^0)$ commutes from city ℓ to city k , where c is the solution to $A_k(1 - \gamma d_{\ell k}) = A_\ell(1 - \gamma(\bar{d}_\ell^0 - c))$.

3.2 Results

Drop in commuting costs. To understand the intuition for how commuting costs affect the different choices, focus on an agent residing in the smaller, less productive city.⁹ If commuting costs are large, an agent has less incentive to move or to commute to the larger, more productive city, because in either case she would be facing longer commuting distances. If commuting costs are at an intermediate level, moving becomes more attractive: although the commuting distance increases, it does so by less than if the agent were to commute from the smaller to the larger city. If commuting costs drop far enough, commuting to the larger city becomes the better choice: she saves the fixed cost of moving, while not suffering an important increase in the cost of commuting. This intuition suggests that a drop in γ makes moving relatively more attractive than staying, and makes commuting relatively more attractive than moving.

Starting off in a situation where all individuals have the same utility, we formalize this intuition and show that a gradual decrease in γ first moves the economy from a staying equilibrium to an inter-city moving equilibrium, with some residents of the smaller low-productivity city moving to the larger high-productivity city. Later the economy moves to an inter-city moving and commuting equilibrium, and to an inter-city commuting equilibrium, with some original residents of the smaller low-productivity city commuting to the larger high-productivity city. This is stated in the following result.

Result 1. *Start off in an equilibrium where $A_k > A_\ell$ and where the utility of all individuals is identical. For a value of μ that is sufficiently small, a gradual drop in commuting costs, γ , moves the economy from a staying equilibrium to an inter-city moving equilibrium and eventually to an inter-city moving and commuting equilibrium and an inter-city commuting equilibrium. In the inter-city moving equilibrium the smaller city loses residents to the larger city.*

Proof. Initially $A_\ell(1 - \gamma \bar{d}_\ell^0) = A_k(1 - \gamma \bar{d}_k^0)$. In this case, $A_\ell(1 - \gamma \bar{d}_\ell^0) \geq A_k(1 - \gamma \bar{d}_k^0) - \mu d_{\ell k}$

⁹There are of course also situations where an agent residing in the larger, more productive might want to move to the smaller, less productive city. For example, this could happen if commuting costs are large, and the bigger city is too large. We will return to this possibility later in the discussion.

and $A_\ell(1 - \gamma\bar{d}_\ell^0) \geq A_k(1 - \gamma d_{\ell k})$ because $d_{\ell k} \geq \bar{d}_k^0$ by construction. Without loss of generality, assume $A_k > A_\ell$, so that $\bar{d}_k^0 > \bar{d}_\ell^0$. As a result, $-\partial A_\ell(1 - \gamma\bar{d}_\ell^0)/\partial\gamma < -\partial A_k(1 - \gamma\bar{d}_k^0)/\partial\gamma$, so that a drop in γ leads to $A_k(1 - \gamma\bar{d}_k^0) > A_\ell(1 - \gamma\bar{d}_\ell^0)$. If γ continues to drop, at some point $A_k(1 - \gamma\bar{d}_k^0) - \mu d_{\ell k} = A_\ell(1 - \gamma\bar{d}_\ell^0)$. We refer to this threshold as γ_m . Suppose μ is small enough so that at this point $A_k(1 - \gamma\bar{d}_{\ell k}^0) < A_k(1 - \gamma\bar{d}_k^0) - \mu d_{\ell k}$. If γ continues to fall, then some of original residents of ℓ will want to move to k . To be precise, $\min[m, \bar{d}_\ell^0]$ people who originally lived in ℓ will move to k , where $A_k(1 - \gamma(\bar{d}_k^0 + m)) - \mu d_{\ell k} = A_\ell(1 - \gamma(\bar{d}_\ell^0 - m))$. As γ continues to gradually drop, m will gradually increase. At some point, the drop in γ reaches $A_k(1 - \gamma(\bar{d}_k^0 + m)) - \mu d_{\ell k} = A_k(1 - \gamma d_{\ell k})$. We refer to this threshold as γ_{mc} . Any further drop in γ will now imply that some of the original residents of ℓ will prefer to commute to k . If γ continues to drop, an increasing share of the original residents of ℓ commute. There is a threshold γ_c below which all original residents of ℓ commute to k . \square

The above result implies three threshold values of γ . A high threshold, γ_m , a middle threshold, γ_{mc} , and a low threshold, γ_c , such that for $\gamma \geq \gamma_m$, we are in a staying equilibrium, for $\gamma_m > \gamma \geq \gamma_{mc}$, we are in an inter-city moving equilibrium, for $\gamma_{mc} > \gamma \geq \gamma_c$, we are in an inter-city moving and commuting equilibrium, and for $\gamma < \gamma_c$, we are in an inter-city commuting equilibrium.

Figure 2: Inter-city Moving and Commuting

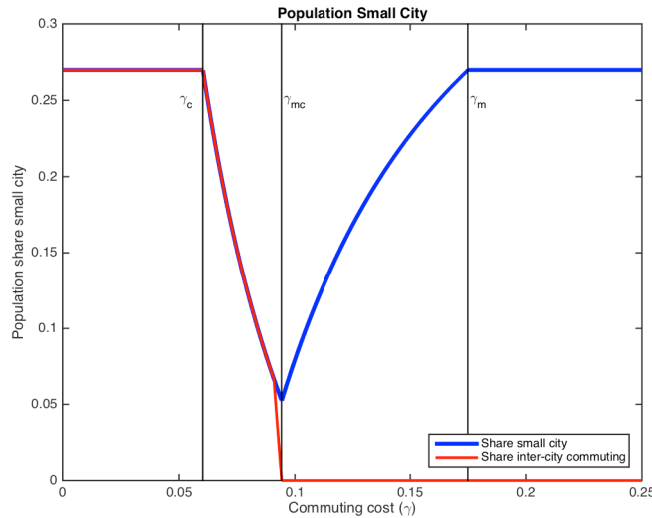


Figure 2 illustrates Result 1 with a simple numerical example. The productivity values are set such that the large city has a TFP that is 10% higher than the small city: $A_\ell = 1$ and $A_k = 1.1$. The inter-city distance is set to 1, and the overall population to 1.5: with cities being symmetric around their production points, this implies 75% of the land between the two cities is occupied. We set the moving cost parameter μ to 0.03. When taking the initial income in the small city as reference, this amounts to a cost of slightly more than 3% in income-equivalence terms. We choose the initial value of γ to be 0.25, implying that people would lose 25% of their income if they were

to commute to the other city.

Starting off with a geographic distribution of population such that utility levels are the same in both cities, we analyze what happens to the population share of the small city and to the population share commuting to the large city as we lower γ from 0.25 all the way to zero. As can be seen in Figure 2, when $\gamma > \gamma_m$, there is no inter-city moving or commuting. All residents of the small city stay put, so that its population is constant. Once γ drops below γ_m , some residents of the small city move to the large city, and the population of the small city gradually declines as the commuting cost continues to decrease. When γ falls below γ_{mc} , some of the residents who had moved to the large city now return and prefer to commute. The population of the small city slowly recovers. Once the commuting cost drops to γ_c , the small city has reached its original population level, but now the entire community commutes to the large city.

We now explore how these thresholds depend on the degree of geographic isolation of both cities. We measure the degree of geographic isolation by $\bar{d}_{\ell k}$. As stated in the following result, if the cities are geographically more isolated, then both thresholds are lower.

Result 2. *Thresholds γ_m and γ_c are declining in $\bar{d}_{\ell k}$. That is, for cities that are more isolated geographically, the shift from a staying equilibrium to a moving equilibrium and from an inter-city moving equilibrium to an inter-city commuting equilibrium occurs for lower values of the commuting cost γ .*

Proof. From the proof of Result 1, we can rewrite $\gamma_m = (A_k - A_\ell - \mu d_{\ell k}) / (A_k \bar{d}_k^0 - A_\ell \bar{d}_\ell^0)$. It is clear that $d\gamma_m / dd_{\ell k} < 0$. From the same proof of Result 1, we can write $\gamma_c = \mu d_{\ell k} / (A_k (d_{\ell k} - \bar{d}_k^0 - m))$, where m can be written as $(A_k - A_\ell - A_k \gamma_c \bar{d}_k^0 + A_\ell \gamma_c \bar{d}_\ell^0 - \mu d_{\ell k}) / (\gamma_c (A_\ell + A_k))$. Together, this implies that $\gamma_c = (A_k (A_k - A_\ell) + A_\ell d_{\ell k} \mu) / (A_k (A_k + A_\ell) d_{\ell k} - A_k A_\ell (\bar{d}_k^0 + \bar{d}_\ell^0))$. Here as well, it is immediate that $d\gamma_c / dd_{\ell k} < 0$. \square

The above result says that when cities are geographically farther apart, commuting costs need to drop more before individuals from the smaller city want to move to the bigger city, and they also need to drop more before they find it profitable to commute to the bigger city. Recall from our earlier discussion that γ can also be interpreted as the spatial decay rate of technological spillovers, with a lower value of γ implying a greater geographic span of technological spillovers. Under this alternative interpretation, the above result says that, when cities are geographically more distant from each other, technological spillovers need to have a wider spatial reach for individuals to move to the high-productivity city, and they need to have an even greater reach for individuals to commute to the high-productivity city.

Now compare two situations, one where the cities are closer to each other, and one where the cities are further apart. To distinguish between both situations, we use prime superscripts to denote variables that relate to the situation of geographically more isolated cities. Hence, $\bar{d}'_{\ell k} > \bar{d}_{\ell k}$,

$\gamma'_m < \gamma_m$ and $\gamma'_c < \gamma_c$. For now assume that $\gamma_c < \gamma'_m$, so that $\gamma'_c < \gamma_c < \gamma'_m < \gamma_m$. We can then write down the following corollary.

Corollary 1. *Compare two situations, one where the two cities are closer to each other and one where the two cities are geographically further apart (denoted by “prime” superscript), where $\gamma'_c < \gamma_c < \gamma'_m < \gamma_m$. When γ drops below threshold γ_m but is still above threshold γ'_m , the less isolated small city loses population, whereas the more isolated small city does not. When γ drops below threshold γ_c , the less isolated city no longer loses population, whereas the more isolated city does.*

The above result implies that (i) for intermediate levels of commuting costs (or for intermediate levels of spatial technology spillovers), growth will be relatively lower if there is a large city closeby, and (ii) for low levels of commuting costs (or high levels of spatial technological spillovers), growth will be relatively higher if there is a large city closeby.

3.3 Agglomeration Economies

It is straightforward to extend the model and allow for local productivity advantages of density. Assume that $A(\ell)$ increases proportionately with the labor employed in city ℓ . This makes moving to the more productive city more attractive compared to commuting to the more productive city. The opportunity cost of time lost by commuting goes up, increasing the incentive to move rather than to commute. This intuition suggests that a strengthening of local productivity advantages reduces γ_c . We summarize this in the next result.

Result 3. *Start off in an equilibrium where $A_k > A_\ell$ where the utility of all individuals is identical. A strengthening of local productivity advantages increases γ_m and lowers γ_c . Moving to the more productive city becomes more attractive relative to residing and working in the least productive city and relative to commuting to the more productive city.*

Proof. First, consider the economy is at threshold γ_m . At that point, individuals are indifferent between staying and residing in the least productive city and moving to the more productive city, so $A_\ell(1 - \gamma_m d_\ell) = A_k(1 - \gamma_m \bar{d}_k) - \mu d_{\ell k}$. Hence, $A_\ell(1 - \gamma_m d_\ell) < A_k(1 - \gamma_m \bar{d}_k)$. Now increase the local productivity of density, so that A_k increases proportionately more than A_ℓ . This implies that the positive difference between $A_k(1 - \gamma_m \bar{d}_k)$ and $A_\ell(1 - \gamma_m d_\ell)$ increases. As a result, at the original threshold γ_m individuals prefer moving to the more productive city. This implies that γ_m increases.

Second, consider the economy is at threshold γ_c . At that point, individuals are indifferent between moving to the more productive city and commuting to the more productive city, so $A_k(1 - \gamma_c d_{\ell k}) = A_k(1 - \gamma_c \bar{d}_k) - \mu d_{\ell k}$. Now increase the local productivity advantage of density. Because

$d_{\ell k}$ is greater than \bar{d}_k , this makes moving more attractive than commuting, implying that γ_c has to go down. \square

Hence, if we are in a situation where people are commuting from the smaller, less productive city to the larger, more productive city, then any strengthening of local agglomeration economies will incentivize people to move to the larger, more productive city. Likewise, if we are in a situation where people are residing and working in the least productive city, then any strengthening of local agglomeration economies will incentivize people to move to the larger, more productive city.

4 Evidence on Commuting Costs and Agglomeration Economies

In this section we discuss some of the main forces that have shaped the spatial interaction between locations in the U.S. over the last two-hundred years. We focus on two such forces: changes in commuting costs and changes in the strength of the economies of density. In light of these changes, we explore the theoretical predictions of our conceptual framework, and show that they are consistent with our empirical findings.

4.1 Commuting Costs

The long-run evolution of commuting costs depends on changes in transportation technologies and on other changes in the cost of commuting.

Transportation technologies. The last two centuries have seen enormous improvements in transportation technologies. Some of those have greatly enhanced long-distance trade and market integration. Examples that come to mind include the railroad network (Donaldson and Hornbeck, 2016; Fogel, 1964), the building of canals (Shaw, 1990), the construction of the inter-state highway system, and containerization. To give an idea of the magnitude of decline, Glaeser and Kohlhase (2004) document that the real cost per ton-mile of railroad transportation dropped by nearly 90% between 1890 and 2000. Other changes have been key in improving short-distance transportation between neighboring or relatively close-by places. For the purpose of our paper, we are mostly interested in the latter improvements.

We now give a brief overview of the main innovations that occurred in short-distance transportation technology in the U.S. over the past two centuries. Prior to the 1830s, almost all Americans worked near the central business district and walked to work. Horse-drawn carriages were available, but were only affordable to the very rich (LeRoy and Sonstelie, 1983).¹⁰ Beginning in the 1830s, several important innovations in mass transportation helped cities like New York,

¹⁰Regular steam ferry service began in the early 1810s but was limited to big coastal cities like New York.

Boston, and Philadelphia expand and spread to many other locations. The omnibus, a horse-drawn vehicle carrying twelve passengers that first appeared in the 1820s became widely used in the 1840s (Kopecky and Hon Suen, 2004). However, it was still an expensive and not very fast way to travel.¹¹ Commuter railroads appeared in the 1830s, although they were noisy and polluting, which led authorities to impose strict regulations, often limiting their use.¹²

Between 1850 and 1900 the U.S witnessed the emergence of horse-drawn and electric streetcars, with the latter generating the so-called ‘streetcars suburbs’ (Warner, 1962; Mieszkowski and Mills, 1993; Kopecky and Hon Suen, 2004). Initially only high-income individuals could afford using streetcars to commute, making the first streetcar suburbs in cities like Boston areas of well-off families.¹³ By 1920 it had become an affordable means of commuting for almost any worker. In addition to streetcars, the first subways appeared in 1900, although only in very large cities.

While all these innovations significantly decreased transportation and commuting costs, it was not until the path-breaking invention of the internal combustion engine that these costs decreased dramatically. The internal combustion engine led to the introduction of the automobile in the day-to-day life of Americans. Cars became increasingly affordable to the middle class starting with the mass production of the Model-T in 1908. Automobile ownership started to increase rapidly in the 1920s and continued its upward trend until the 1970s (Kopecky and Hon Suen, 2004). As explained in Glaeser and Kahn (2004), cars had two effects on population decentralization. First, they dramatically reduced transport costs, thus increasing the possible distance between residences and jobs. This was especially important for commuting between cities and between centers and increasingly far-off suburbs. Second, they eliminated the scale economies involved in older transportation technologies, which implied that cities no longer needed to develop around ports or railroad hubs, allowing employment to decentralize throughout metropolitan areas and throughout the country.

The impact of the car on commuting costs accelerated after 1944, with the construction of the national highway system. Many of these highways connected the downtown areas of large urban centers to the suburbs and the hinterland. Baum-Snow (2007) argues that cars and highways were a fundamental determinant of the suburbanization of American cities. His estimations show that, between 1950 and 1990, the construction of one new highway passing through a central city reduced its population by about 18 percent. Another major transportation change starting around 1950 was the construction of suburban rail terminals. In cities like San Francisco and Washington, D.C. heavy-rail systems were established, while light-rail systems followed in cities like San Diego,

¹¹LeRoy and Sonstelie (1983) document that an omnibus fare ranged from 12 cents to 50 cents at a time when a laborer might earn \$1.00 a day. Its average speed was slow – about 6 miles per hour.

¹²As in the case of the omnibus, commuter railroads were also quite expensive (Leroy and Sonstelie, 1983).

¹³A two-horse streetcar could carry 40 passengers, while an omnibus could only transport 12 passengers. The streetcar speed was about one-third greater than that of the omnibus and its costs was 5 cents per trip.

Portland in later decades (Oxford Research Encyclopedia of American History).

Other commuting costs. Other factors that affect the cost of commuting are gasoline prices and the opportunity cost of time. With the exception of the oil crises of the 1970s, real gas prices remained relatively low between the 1940s and the end of the twentieth century, but have experienced an important increase since then. Consistent with this rise in the cost of commuting by car, the share of Americans who commute to work by car, after peaking at 87.9 percent in 2000, has slightly decreased to 85.3% in 2016 (American Community Survey Report (2013, 2016)).

The cost of commuting is also partly a time cost. Several papers argue that the hours worked by households and individuals have increased in recent decades, hence increasing the opportunity cost of time. Edlund et al. (2016) focus on the increase in double-income high-skilled households between 1980 and 2010. Dual-earner couples have less time, making commuting more costly, giving them an incentive to live closer to work. Su (2018) makes a similar point, but focuses on individuals between 1990 and 2010. The percentage of those working long hours has increased for all skill classes, though the effect is larger for the college educated. To economize on the commuting time, the high-skilled are disproportionately moving to the city centers. Su further argues that this improves amenities in city center, further attracting other individuals, not just the high-skilled. Of course, since this process of gentrification also displaces people, it is not clear whether this is associated with a decline or an increase in the center-city population. However, Glaeser et al. (2011) find that high-amenity locations tend to grow faster, so the improvements in amenities are likely to attract a net inflow of population.

Interpretation of empirical results. When focusing on 1840-2016, the above discussion suggests that we can distinguish three subperiods in the evolution of commuting costs. Between 1840 and 1940, there was a gradual decline in commuting costs, driven by the introduction of the omnibus and the streetcar. Between 1940 and 2000, there was a rapid decline in commuting costs, driven by the mass adoption of the automobile, the construction of highways connecting urban areas with their hinterlands, and the expansion of suburban rail systems. Between 2000 and 2016, this decline in commuting costs came to a halt, and even reversed, because of the increasing opportunity cost of time, and rising gas prices.¹⁴

Using the notation of the model, this implies a slow decline in γ between 1840 and 1940, a rapid decline in γ between 1940 and 2000, and a slow increase in γ between 2000 and 2016. To interpret what this implies about the effect of nearby large cities on population growth of smaller

¹⁴The years that separate the three subperiods, 1940 and 2000, do not constitute precise breakpoints. The mass adoption of cars started in the 1920s, whereas the building of urban highways and suburban rail networks only started in earnest in the 1950s and the 1960s. This justifies using 1940 to separate the first two subperiods. The increase in the opportunity cost of time started between 1980 and 1990, whereas the rise in gas prices occurred post-2000. Given that we focus on twenty-year subperiods, this could justify using either 1980 or 2000 as the year separating the last two subperiods.

cities in the context of our model, we first turn to Result 1. Starting in a situation where no one has an incentive to move or commute to another city, Result 1 says that if γ declines, initially individuals will want to move from the small city to the nearby large city, and if γ continues to decline, individuals will prefer to commute from the small city to the large city. Of course, going the other way, if γ increases, people may choose to move to the large city rather than commute from the small to the large city. Given the evolution of γ in the data, Result 1 therefore implies an inverted U-shaped effect of having a large city nearby: from negative, to no effect, to negative again.

Result 1 analyzes the effect of a large city on a nearby smaller city, but it does not explore how this effect differs at different distances. For that purpose, Corollary 1 is more relevant. If the large city is closer, γ needs to drop by less before people from the small city switch from moving to the large city to commuting to the large city. As a result, in a world where transport costs are gradually declining, people will start commuting from closeby small cities before they start commuting from more far-away small cities. Hence, in relative terms, small cities at shorter distances to the large cities will fare better than small cities at more far-off distances to the large city. When focusing on the time period 1940-2000, this is consistent with the positive coefficient on having a large city very closeby relative to the negative coefficient on having a large city at a slightly farther away distance.

4.2 Local Agglomeration Economies

Although providing accurate estimates of agglomeration economies for a substantial number of years is a daunting task,¹⁵ there is some evidence that agglomeration economies have become stronger over time (Davis and Weinstein, 2002). Focusing on the period 1880-1987, Kim (1999) shows that the importance of agglomeration economies relative to first-nature forces in shaping the distribution of economic activity in the U.S. has increased. This may have been driven by technological progress and structural change, crucial features of the U.S. economy during this period (Duranton 1999; Michaels et al., 2012; Desmet and Henderson, 2015). In most countries industrialization and urbanization have evolved hand-in-hand.

Recent time period. Several studies suggest that agglomeration economies have increased in recent decades. First, the spread in the use of information and communication technologies (ICT) started in the early 1980s. As argued by Desmet and Rossi-Hansberg (2009), agglomeration economies tend to increase in times when new general purpose technologies are introduced. Large agglomerations provide laboratories to discover the many applications of ICT or any other disruptive technologies, thus increasing the attractiveness of cities. Consistent with this, Desmet

¹⁵See Rosenthal and Strange (2004) and De La Roca and Puga (2017) for a discussion of the challenges involved in estimating agglomeration economies.

and Rappaport (2014) find an increase in the divergence in population growth rates after 1980, benefitting larger locations relative to smaller locations. This trend has continued after the turn of the century. Second, Baum-Snow et al. (2018) find that agglomeration economies have been increasing in each decade since 1980 for the skilled. Although the opposite is true for the unskilled, the positive effect for the skilled is larger than the negative effect for the unskilled. This, together with the workforce becoming more educated, points towards an average strengthening in agglomeration economies in recent decades. Finally, the increase in consumption amenities triggered by the gentrification process discussed above has increased the appeal of large cities.

Interpretation of empirical results. The secular increase in agglomeration economies has made moving to large cities more attractive. In the context of our model, Result 3 says that this would imply a negative sign on the coefficient of nearby population. If this were the only force at play, we would expect urban growth shadows to dominate throughout the entire period from 1840 to 2016. The exception to this is the time period spanning 1940-2000, which witnessed the largest improvements in commuting between closeby locations.

5 Concluding Remarks

In this paper we have analyzed whether a location's growth benefits or suffers from being geographically close to large urban centers. To do so, we have focused on U.S. counties and metro areas over the time period 1840-2016. We have documented that, with the exception of 1940-2000, proximity to urban clusters has had a negative effect on a location's growth. This negative impact includes the most recent time period 2000-2016.

References

- [1] Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015). “The Economics of Density: Evidence from the Berlin Wall,” *Econometrica*, 83 (6): pp. 2127-2189.
- [2] Allen, T. and C. Arkolakis, C. (2014). “Trade and the Topography of the Spatial Economy,” *Quarterly Journal of Economics*, 129(3): 1085-1140.
- [3] Atack, J., F. Bateman, M. Haines, and R. Margo (2010). “Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850-1860,” *Social Science History*, 34(2): 171-197.
- [4] Baum-Snow, N. (2007). “Did Highways Cause Suburbanization?,” *Quarterly Journal of Economics*, 122(2): 775-805.
- [5] Baum-Snow, N. (2010). “Changes in Transportation Infrastructure and Commuting Patterns in U.S. Metropolitan Areas, 1960-2000,” *American Economic Review Papers and Proceedings*, 100(2): 378-382.
- [6] Baum-Snow, N., M. Freedman, and R. Pavan (2018). “Why Has Urrban Inequality Increased?,” *American Economic Journal: Applied Economics*, forthcoming.
- [7] Beeson, P. E., D. N. DeJong, and W. Troesken (2001). “Population Growth in U.S. Counties, 1840-1990,” *Regional Science and Urban Economics*, 31(6): 669-699.
- [8] Bellis, M. (2018). “American Farm Machinery and Technology Changes from 1776-1990,” ThoughtCo., <https://www.thoughtco.com/american-farm-tech-development-4083328>.
- [9] Beltrán, F. J., A. Díez-Minguela, and J. Martínez-Galarraga (2017). “The Shadow of Cities: Size, Location, and the Spatial Distribution of Population in Spain,” *Cambridge Working Paper Economics* 1749.
- [10] Beltrán, F. J., A. Díez-Minguela, and J. Martínez-Galarraga (2018). “Tracing the Evolution of Agglomeration Economies: Spain, 1860-1991,” *Journal of Economic History* 78(1): 81-117.
- [11] Black, D., and V. Henderson (2003). “Urban Evolution in the USA,” *Journal of Economic Geography*, 3(4): 343-372.
- [12] Bleakley, H., and J. Lin (2012). “Portage and Path Dependence,” *Quarterly Journal of Economics*, 127(2): 587-644.
- [13] Bodenhorn, H., and D. Cuberes (2018). “Banks and City Growth in Nineteenth-Century New York,” *Journal of Urban Economics*, 104: 47-58.

- [14] Borchert, J. (1963). *The Urbanization of the Upper Midwest*, Upper Midwest Economic Study. Retrieved from the University of Minnesota Digital Conservancy, <http://hdl.handle.net/11299/160786>.
- [15] Borchert, J. (1972). "America's Changing Metropolitan Regions," *Annals of the Association of American Geographers*, 62: 352-373.
- [16] Bosker, M., and E. Buringh (2017). "City Seeds: Geography and the Origins of European Cities," *Journal of Urban Economics*, 98: 139-157.
- [17] Bosker, M., S. Brakman, H. Garretsen, H. de Jong, and M. Schramm (2008). "Ports, Plagues and Politics: Explaining Italian City Growth 1300-1861," *European Review of Economic History*, 12, 97-131.
- [18] Cheshire, P. C., and S. Magrini (2002). "The Distinctive Determinants of European Urban Growth: Does One Size Fit All?," Research Papers in Environmental and Spatial Analysis, Department of Geography and Environment, London School of Economics and Political Science, London, UK.
- [19] Christaller, W. (1933). *Central Places in Southern Germany*. Jena, Germany: Fischer (English translation by C. W. Baskin, London: Prentice Hall, 1966).
- [20] Cuberes, D. (2011). "Sequential City Growth: Empirical Evidence" *Journal of Urban Economics*, 69: 229-239.
- [21] Da Mata, D., U. Deichmann, J. V. Henderson, S. V. Lall, and H.G. Wang (2007). "Determinants of City Growth in Brazil," *Journal of Urban Economics*, 62: 252-272.
- [22] Davis, D. R., and D. E. Weinstein (2002). "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, 92(5): 1269-1289.
- [23] De la Roca, J. and D. Puga (2017). "Learning by Working in Big Cities," *Review of Economic Studies*, 84(1): 106-142.
- [24] Desmet, K., and M. Fafchamps (2005). "Changes in the Spatial Concentration of Employment across US Counties: a Sectoral Analysis 1972-2000," *Journal of Economic Geography*, 5: 261-284.
- [25] Desmet, K., and V. Henderson (2015). "The Geography of Development within Countries," in: Duranton, G., Henderson, J.V. and Strange, W. (Eds.), *Handbook of Regional and Urban Economics*, vol. 5., Elsevier: Amsterdam.
- [26] Desmet, K., D. Nagy, and E. Rossi-Hansberg (2018). "The Geography of Development," *Journal of Political Economy*, 126(3): 903-983.

- [27] Desmet, K., and J. Rappaport (2015). “The Settlement of the United States, 1800-2000: The Long Transition towards Gibrat’s Law,” *Journal of Urban Economics*, 98: 50-68.
- [28] Desmet, K., and E. Rossi-Hansberg (2014). “Spatial Development,” *American Economic Review*, 104(4): 1211-1243.
- [29] Dobkins, L. H., and Y. Ioannides (2001). “Spatial Interactions among U.S. Cities,” *Regional Science and Urban Economics*, 31(6): 701-731.
- [30] Donaldson, D., and R. Hornbeck (2016). “Railroads and American Economic Growth: A Market Access Approach,” *Quarterly Journal of Economics*, 131(2): 799-858.
- [31] Duranton, G. (1999). “Distance, Land, and Proximity: Economic Analysis and the Evolution of Cities,” *Environment and Planning A*, 31(12): 2169-2188.
- [32] Duranton, G., and D. Puga (2014). “The Growth of Cities,” *Handbook of Economic Growth*, Elsevier, volume 2, 781-853.
- [33] Edlund, L., C. Machado, and M. Sviatschi (2016). “Bright Minds, Big Rent: Gentrification and the Rising Returns to Skill,” NBER Working Paper # 21729.
- [34] Elliott, H.M. (1981). “Macrocalifornia and the Urban Gradient,” *The California Geographer*, 21: 1-17.
- [35] Faber, B. (2014). “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System” *Review of Economic Studies*, 81(3): 1046-1070.
- [36] Fang, C. L., and H. Y. Liu (2007). “The Spatial Deprivation and the Corresponding Controlling Paths in China’s Urbanization Process” *Acta Geographica Sinica (in Chinese)*, 62(8): 849-860.
- [37] Fogel, R. W. (1964). *Railroads and American Economic Growth: Essays in Econometric History*, Baltimore, MD: Johns Hopkins University Press.
- [38] Fujita, M., P. Krugman, and A. J. Venables (1999). *The Spatial Economy*, MIT Press.
- [39] Fujita, M., and T. Mori (1996). “The Role of Ports in the Making of Major Cities: Self-Agglomeration and the Hub-Effect,” *Journal of Development Economics*, 49(1): 93-120.
- [40] Gardner, T. (1999). “Metropolitan Classification for Census Years before World War II,” *Historical Methods*, 32(3): 139-150.
- [41] Gaspar, J., and E. L. Glaeser (1998). “Information Technology and the Future of Cities,” *Journal of Urban Economics*, 43: 136-156.

- [42] Glaeser, E.L., and M. Kahn (2004). “Sprawl and Urban Growth,” In: Henderson J.V, Thisse J.F (ed.), *Handbook of Regional and Urban Economics*, Vol.4. Elsevier.
- [43] Glaeser, E. L., J. Kolko, and A. Saiz (2001). “Consumer City,” *Journal of Economic Geography*, 1: 27-50.
- [44] González-Val, R. (2017). “When Zipf Meets Space,” unpublished manuscript.
- [45] Greenstone, M., R. Hornbeck, and E. Moretti (2010). “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,” *Journal of Political Economy*, 118(3): 536-598.
- [46] Griffith, D. A. (1976). “Spatial Structure and Spatial Interaction: A Review,” *Environment and Planning A: Economy and Space*.
- [47] Guinness, P. G., and M. Bradshaw (1985). *North America: a Human Geography*, London: Hodder and Stoughton.
- [48] Hanson, G. (2005). “Market Potential, Increasing Returns and Geographic Concentration,” *Journal of International Economics*, 67: 1-24.
- [49] Henderson, V., A. Storeygard, T. Squires, and D. N. Weil (2018). “The Global Spatial Distribution of Economic Activity: Nature, History, and the Role of Trade,” *Quarterly Journal of Economics*, 133(1): 357-406.
- [50] Hsu, W-T., T. J. Holmes, and F. Morgan (2014). “Optimal City Hierarchy: A Dynamic Programming Approach to Central Place Theory,” *Journal of Economic Theory*, 154: 245-273.
- [51] Huff, D. L.(1973). “The Delineation of a National System of Planning Regions on the Basis of Urban Spheres of Influence,” *Regional Studies*, 7: 323-329.
- [52] Ioannides, Y., and J. Zhang (2017). “Walled Cities in Late Imperial China,” *Journal of Urban Economics*, 97: 71-88.
- [53] Kim, S. (1999). “Regions, Resources and Economic Geography: The Sources of U.S. Regional Comparative Advantage, 1880-1987,” *Regional Science and Urban Economics*, 29 (1): 1-32.
- [54] Krugman, P. (1991). “Increasing Returns and Economic Geography,” *Journal of Political Economy*, 99: 483-499.
- [55] Krugman, P. (1993). “On the Number and Location of Cities,” *European Economic Review*, 37(2-3): 293-298.

- [56] Liu, Y., X. Wang, and J. Wu (2011). “Do Bigger Cities Contribute to Economic Growth in Surrounding Areas? Evidence from County-Level Data in China,” unpublished manuscript.
- [57] Lösch, A. (1940). *The Economics of Location*, Jena, Germany: Fischer (English translation, New Haven, CT: Yale University Press, 1954).
- [58] Michaels, G., F. Rauch, and S. J. Redding (2012). “Urbanization and Structural Transformation,” *Quarterly Journal of Economics*, 127(2): 535-586.
- [59] North, D. (1955). “Location Theory and Regional Economic Growth,” *Journal of Political Economy*, 63: 243-258.
- [60] Partridge M. D., D. S. Rickman, K. Ali, and M. R. Olfert (2008). “Lost in Space: Population Dynamics in the American Hinterlands and Small Cities,” *Journal of Economic Geography* 8: 727-757.
- [61] Partridge M. D., D. S. Rickman, K. Ali, and M. R. Olfert (2009). “Do New Economic Geography Agglomeration Shadows Underlie Current Population Dynamics across the Urban Hierarchy?,” *Papers in Regional Science*, 88(2): 445-466.
- [62] Partridge, M. D., K. Ali, and M. R. Olfert (2010). “Rural-to-Urban Commuting: Three Degrees of Integration,” *Growth and Change*, 41(2): 303-335.
- [63] Poor, H. V. (1860). *History of Railroads and Canals of the United States*, J.H Schultz and Company.
- [64] Rappaport, J. (2005). “The Shared Fortunes of Cities and Suburbs,” *Federal Reserve Bank of Kansas City Economic Review*, Third Quarter, 33-59.
- [65] Rauch, F. (2014). “Cities as Spatial Clusters,” *Journal of Economic Geography*, 14: 759-773.
- [66] Redding, S., and D. Sturm (2008). “The Costs of Remoteness: Evidence from German Division and Reunification,” *American Economic Review*, 98(5): 1766-1797.
- [67] Redding, S. J., and M. A. Turner (2015). “Transportation Costs and the Spatial Organization of Economic Activity,” in: *Handbook of Regional and Urban Economics*. Edited by G. Duranton, J. V. Henderson and W. C. Strange.
- [68] Rosenthal, S. S, and W. C. Strange (2003). “Geography, Industrial Organization, and Agglomeration,” *Review of Economics and Statistics* 85(2): 377-393.
- [69] Rosenthal, S. S, and W. C. Strange (2004). “Evidence on the Nature and Sources of Agglomeration Economies,” in: *Handbook of Regional and Urban Economics*, Edited by J. V. Henderson and J. F. Thisse.

- [70] Shaw, R. E. (1990). *Canals for a Nation. The Canal Era in the United States, 1790-1860*, The University Press of Kentucky.
- [71] Su, Y. (2018). “The Rising Value of Time and the Origin of Urban Gentrification,” unpublished manuscript.
- [72] Tabuchi, T., and J-F. Thisse (2011). “A New Economic Geography Model of Central Places,” *Journal of Urban Economics*, 69: 240-252.
- [73] Taylor, G. R. (1966). *The Transportation Revolution, 1815-1860*, New York: Holt, Rinehart and Winston.
- [74] Thorndale, W., and W. Dollarhide (1987). *Map Guide to the Federal Censuses, 1790-1920*, Genealogical Publishing Company, Baltimore.
- [75] Warner, S. B. Jr. (1972). *Streetcar Suburbs. The Process of Growth in Boston, 1870-1900*, Harvard University Press and the MIT Press.