

Urban Growth Shadows

David Cuberes, Klaus Desmet, and Jordan Rappaport

November 2019; Revised July 2020

RWP 19-08

<http://doi.org/10.18651/RWP2019-08>

FEDERAL RESERVE BANK *of* KANSAS CITY



Urban Growth Shadows*

David Cuberes

Klaus Desmet

Jordan Rappaport

Clark University

SMU, NBER and CEPR

Kansas City Fed

July 2020

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 1840-2017, we document this tradeoff between urban shadows and urban spillovers. Proximity to large urban centers was negatively associated with growth between 1840 and 1920, and positively associated with growth after 1920. Using a two-city spatial equilibrium model with intra-city and inter-city commuting, we show that the secular evolution of commuting costs can account for this and other observed patterns in the data.

Keywords: urban shadows, agglomeration economies, spatial economics, urban systems, city growth, United States, 1840-2017

JEL Codes: R12, N93

“Cities were like stars or planets, with gravitational fields that attracted people and trade like miniature solar systems.”

— William Cronon, *Nature's Metropolis: Chicago and the Great West*

1 Introduction

In his account of the U.S. westward expansion during the nineteenth century, Cronon (1991) writes that land speculators on the frontier saw cities as having a gravitational pull akin to a law of nature that inexorably attracted migrants from the hinterland to the new urban centers.¹ This is

*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. We benefitted from presentations at the Urban Economic Association Meetings (New York and Amsterdam), Boston Federal Reserve Bank, Princeton University, and the University of North Dakota. 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. We thank McKenzie Humann, Anissa Khan, Da Li, John Limaldi and Isabel Steffens for outstanding research assistance.

¹This view of cities echoes that of central place theory (Christaller, 1933; Lösch, 1940).

consistent with a view that smaller places “close” to larger cities fall under the “urban shadow” of their neighbors, with increased competition for resources dampening their growth.² However, there is also an opposing view: the presence of nearby clusters of economic activity generates positive agglomeration spillovers, benefiting the growth of neighboring smaller places.³

Is the proximity of a large urban center beneficial or harmful to a location’s economic growth? This paper empirically and theoretically explores this question. Focusing on local population growth in the U.S. over almost two-hundred years, our empirical analysis identifies two distinct time periods: between 1840 and 1920, urban shadows dominated, and since then, between 1920 and today, urban spillovers have taken over. One key force that is likely to have driven the changing relative strength of urban shadows and urban spillovers is the evolution of intra- and inter-city commuting costs. After providing an overview of changes in commuting costs over the last two hundred years, we develop a two-city spatial model that incorporates both commuting and moving costs. We show that the long-run behavior of just one variable — commuting costs — can account for many of the observed patterns in the data, including the changing relative strength of urban shadows and urban spillovers over time and space.

Using U.S. county and metro population data from 1840 to 2017, our empirical analysis documents the changing correlation of local population growth with the presence of nearby large locations. In addition to establishing the important shift from urban shadows to urban spillovers around the year 1920, we identify three more stylized facts. First, since the turn of the twenty-first century, there has been a decrease in the positive correlation between proximity to a large urban center and growth, suggesting that urban spillovers have been weakening in the last decades. Second, there is evidence of the geographic reach of large urban centers expanding. Urban spillovers were very local between 1920 and 1940 and much more far-reaching between 2000 and 2017. Third, the greater the size of a nearby large location, the bigger the correlation with its hinterland’s growth. The evidence therefore suggests that larger locations exerted stronger urban shadows in the earlier time period, as well as stronger urban spillovers in the later time period.

We hypothesize that changes in commuting costs can account for these patterns. Before showing this with the help of a simple spatial model, the paper documents the evolution of commuting costs in the U.S. over the same period 1840-2017. Beginning in the mid-nineteenth century, a steady stream of transport innovations lowered the cost of commuting. The introduction of the streetcar facilitated longer-distance commutes, giving rise to the first “streetcar suburbs”. By the beginning of the twentieth century, streetcars had become widespread across cities in America. Drawing on cross-sectional data on streetcars for the time period 1900-1920, we identify a fifth

²See, e.g., Krugman (1993), Fujita et al. (1999), Black and Henderson (2003), and Bosker and Buringh (2017).

³See, amongst others, Davis and Weinstein (2002), Rosenthal and Strange (2003), Glaeser and Kahn (2004), Hanson (2005), and Redding and Sturm (2008).

stylized fact: streetcars strengthened large cities' urban shadows on their hinterland. The decline in commuting costs accelerated dramatically during the inter-war and post-war periods. The combination of the widespread adoption of the automobile and the building of the highway system connecting downtowns to hinterlands made it possible for people to live much further away from work. People residing in smaller nearby locations no longer had the need to permanently move to the larger cities to enjoy their productive benefits. Instead, they could reside in the hinterland, and commute to the urban centers for work. There is some indication that the continued drop in commuting costs has weakened in the last two decades. Several factors may have contributed to this: improvements in commuting technology have petered out; road and rail infrastructure investment has stalled; traffic congestion has worsened and the commuting speed has slowed down; and there has been a rise in the opportunity cost of time, due to longer work hours and the increase in double-income families.

To understand the role of commuting costs in shaping the relative strength of urban shadows and urban spillovers, we develop a simple spatial model 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, or she can commute for work to the other city without changing her residence. We then show how these choices change with the cost of commuting, the distance between cities and their relative sizes. We find that as the cost of commuting gradually drops, individuals switch from staying put in the smaller, least productive city, first to moving and later to commuting to the larger, more productive city. Hence, a gradual drop in transport costs first hurts growth in the smaller city, as it loses population to its larger neighbor, but a further drop eventually helps its growth, as its population commutes to the nearby larger city. That is, the smaller city goes from experiencing a negative urban shadow, to benefiting from positive urban spillovers.

The intuition for the non-monotonic relation between commuting costs and the growth of the smaller city is straightforward. The initial drop in transport costs lowers the cost of living in the larger city by more than in the smaller city, for the simple reason that intra-city commutes are on average longer in the larger city than in the smaller city. This makes it more attractive for residents of the smaller city to pay the one-time moving cost to relocate to the larger city. As in Cronon (1991), the large city uses its gravitational force to pull in migrants from the hinterland: an urban shadow. A further drop in transport costs continues to make the larger city more attractive than the smaller city, but it also facilitates inter-city commuting, which was hitherto too costly. This allows residents of the smaller city to work in the larger city without the need to move. The small city benefits from the proximity of the large city: an urban spillover.

When analyzing the observed long-run evolution of commuting costs through the lens of our model, we can account for the five main stylized facts uncovered in the data. Recall that commuting costs experienced three distinct regimes: slow decline between 1840 and 1920, rapid decline between

1920 and 2000, and stagnation since then. When interpreted by the model, these are consistent with urban shadows dominating in the early time period and urban spillovers dominating in the later time period, with some weakening of spillovers in the last decades. These effects are stronger in urban areas that experience more rapid improvements in their commuting infrastructure. The model also shows that as commuting costs decrease, the geographic reach of urban areas expands. In addition, the theory implies that an increase in the relative size of a large urban center strengthens the force it exerts on its hinterland. These theoretical predictions have their empirical counterparts in the stylized facts we highlight in our data analysis.

While the long-run evolution of just one variable – commuting costs – is able to capture the rise and decline of urban shadows, undoubtedly other forces might have been at play as well. One such force are technology or knowledge spillovers: in order for smaller locations to benefit from nearby larger locations, there may be no need to commute to that larger neighbor if technologies diffuse through space. We show that under certain assumptions our model is observationally equivalent to one without inter-city commuting but with technology spillovers. Another force that may drive urban spillovers is market access: the smaller location may benefit from proximity to its larger neighbor through trade. In Appendix B.2 we consider an alternative model with intra-city commuting costs and inter-city trade costs. Using the evidence in Glaeser and Kohlhase (2004) that points to trade costs having declined more rapidly than commuting costs, we show that this alternative model is also able to generate the rise and decline of urban shadows. Given these alternative explanations, we should not view the empirical evidence as proof of our conceptual framework, but rather as suggesting that the long-run decline in commuting costs is one plausible explanation for why locations close to large cities shifted from growing relatively less to growing relatively more.

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 leads to the emergence of a spatially organized hierarchy of locations of different sizes. A natural implication of central place theory 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).⁴

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 for 1990-2006, Partridge et al. (2009) uncover a positive impact of large urban clusters on nearby smaller places. Looking at the post-war period, Rappaport (2005) finds

⁴More 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 et al., 2018).

evidence of the populations of cities and suburbs moving together. Dobkins and Ioannides (2001) also study the interaction of cities of different sizes. 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 net effect of large neighbors was negative. Consistent with this, Rauch (2014) documents that historically larger European cities have been surrounded by larger hinterland areas. Most closely related to our work is Beltrán et al. (2017) who use data on Spanish municipalities for 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 from the Spanish experience for a variety of reasons: the U.S. was much less settled in the nineteenth century, modern-day mobility across cities and regions in the U.S. is greater, and the adoption of the automobile in the U.S. was swifter. In addition, our paper offers a theoretical framework that allows us to interpret the switch from urban shadows to urban spillovers by relating it to the secular decline in transport costs.

In this paper we do not consider the possible role of structural transformation in explaining the absorptive capacity of large locations. If the biggest cities disproportionately benefited from the rural exodus from nearby counties, this would generate an urban growth shadow. While plausible, a recent paper by Eckert and Peters (2018) on spatial structural change since 1880 finds little support for this alternative explanation. More specifically, it finds that the spatial reallocation of people from agricultural to non-agricultural labor markets accounted for almost none of the decline in agricultural employment since 1880. Instead, most of the structural transformation happened within counties. Hence, while our data do not allow us to take a strong stand on the possible role of structural transformation, this paper suggests it might not be a first-order driver of our stylized facts.

The rest of the paper is organized as follows. Section 2 presents the empirical findings on the changes in urban growth shadows and spillovers over the period 1840-2017. Section 3 documents the evolution of commuting costs over the same period 1840-2017. Section 4 proposes a conceptual framework that relates commuting costs to urban shadows and spillovers, and it shows that the theoretical predictions are consistent with the main patterns in the data. Section 5 concludes.

2 Urban Growth Shadows and Spillovers: 1840 to 2017

This section documents how the correlation of local population growth in the U.S. with the presence of nearby large locations has evolved over the period 1840-2017. In doing so, it aims to explore whether urban shadows or urban spillovers were more prominent in different time periods. It also elicits a number of additional stylized facts.

2.1 Data

We use county population data from the Census Bureau spanning the period 1840 to 2017. With the exception of the last period, we focus on successive twenty-year time frames: 1840-1860, 1860-1880, ...,1980-2000, 2000-2017. In constructing the dataset, we had to resolve two main issues: how to deal with changing county borders and how to delineate metro areas over time. In what follows we limit ourselves to a brief discussion, and point the interested reader to Desmet and Rappaport (2017) for more details. To get consistent county borders, we use a “county longitudinal template” 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 1850, we combine counties A_1 and A_2 to measure population growth of county A between 1840 and 1860. 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.⁵ This methodology gives us a separate dataset for each twenty-year period we study, as well as for 2000-2017.

When different counties form part of the same metropolitan area, we do not want to consider these counties as different locations. We therefore combine counties into metro areas, when and where we can delineate them. Our analysis is thus based on a hybrid of metropolitan areas and non-metropolitan counties. For 1940 and earlier, we merge counties to form metropolitan areas applying criteria promulgated by the Office of Management and Budget (OMB) in 1950 to population and economic conditions at the start of each twenty-year period (Gardner, 1999). For 1960 and later, we use the official delineations promulgated by OMB after each decennial census. As with the geographically-consistent counties, growth over any period is measured using the geographic borders of the initial year. The number of locations in our datasets increases rapidly from 862 for the period 1840-1860 to 2,370 for 1880-1900 and then more slowly to a maximum of 2,982 for 1940-1960, reflecting both the westward movement of the U.S. frontier and the splitting of geographically large counties as they became more densely settled into smaller ones. Thereafter, the number of locations steadily declines as more and more counties were absorbed into metropolitan areas. Our dataset for 2000-2017 has 2,369 locations.

The distribution of surrounding locations by size and distance systematically varies across different parts of the country, which in many periods had different average growth rates. For example, locations near the U.S. frontier during the nineteenth century tended to have few large neighbors and high average growth. To avoid an omitted variable bias, we extensively control for regional variation in order to isolate the correlation of growth with measures of surrounding

⁵This description applies to the most common case of counties splitting over time. If counties merge between, say, 1860 and 1880, then we would use geographic borders from 1880.

locations. A first set of 15 control variables are the terms from the third-order polynomial of latitude and longitude, $(1 + lat + lat^2 + lat^3)(1 + long + long^2 + long^3)$. A second set of control variables are indicators for eight of the nine U.S. census divisions. A third set of ten control variables are linear and quadratic terms of average low temperature in January, average high temperature in July, average daily humidity in July, average annual rainfall, and average number of days on which it rains (Rappaport, 2007). A fourth set of ten control variables are indicators of whether a location’s geographic centroid is within 80 kilometers of the coast and of a natural harbor along each of the north Atlantic, south Atlantic, Gulf of Mexico, Pacific, and Great Lakes (Rappaport and Sachs, 2003). A fifth set of two control variables are indicators of whether a location’s geographic centroid is within 40 kilometers of a river on which there was navigation in 1890 and whether it is in addition located within 80 kilometers of an ocean coast (Rappaport and Sachs, 2003). A final set of two variables is a quadratic specification of hilliness, measured as the standard deviation of altitude within a location normalized by the location’s land area. These six sets total 47 variables, which we include in all regressions beginning with the 1860 cross section. A handful of them are dropped for the 1840 cross section due to lack of variation (e.g., there were no locations in the Mountain and Pacific census regions).

2.2 Baseline Specification

Our main specification regresses population growth over successive twenty-year intervals on the presence of surrounding locations at specified distances with population above specified thresholds.

Let $d_{\ell k}$ denote the distance between locations ℓ and k , measured using a straight-line approximation between their geographic centroids. Let $\tilde{d} \in \{\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_D\}$ denote strictly increasing specified distances, e.g., $\{50\text{km}, 100\text{km}, \dots, 300\text{km}\}$. Finally, let L_k and \tilde{L} respectively denote the population of location k and a specified population threshold for considering a neighboring location to be large. For each ordered pair of locations, we construct an indicator variable, $\mathbb{I}_{\ell k}^{\tilde{L}, \tilde{d}}$, describing whether location k has population weakly above threshold \tilde{L} and distance from location ℓ weakly less than \tilde{d} :

$$\mathbb{I}_{\ell k}^{\tilde{L}, \tilde{d}} \equiv \mathbb{I}(L_k, d_{\ell k}; \tilde{L}, \tilde{d}) = \begin{cases} 1 & : L_k \geq \tilde{L} \ \& \ d_{\ell k} \leq \tilde{d} \\ 0 & : \text{otherwise} \end{cases}$$

For each location ℓ , we then construct a set of indicators, one for each specified distance, describing if there is at least one location, $k \neq \ell$, within that distance of location ℓ , that has population weakly

above \tilde{L} and no such location within a smaller distance of ℓ :

$$\mathbb{I}_{\ell}^{\tilde{L}, \tilde{d}} = \begin{cases} 1 & : \tilde{d} = \tilde{d}_1 \ \& \ \sum_{k \neq \ell} \mathbb{I}_{\ell k}^{\tilde{L}, \tilde{d}} \geq 1 \\ 1 & : \tilde{d} \in \{\tilde{d}_2, \dots, \tilde{d}_D\} \ \& \ \left(\prod_{d=\tilde{d}_1}^{\tilde{d}-1} (1 - \tilde{\mathbb{I}}_{\ell}^{\tilde{L}, d}) \right) \left(\sum_{k \neq \ell} \mathbb{I}_{\ell k}^{\tilde{L}, \tilde{d}} \right) \geq 1 \\ 0 & : \text{otherwise} \end{cases}$$

For each 20-year period from 1840 to 2000 and for the 17-year period from 2000 to 2017, we regress annual average population growth, g_{ℓ} , on the set of indicators, $\mathbf{I}_{\ell}^{\tilde{L}} = [\mathbb{I}_{\ell}^{\tilde{L}, \tilde{d}_1}, \mathbb{I}_{\ell}^{\tilde{L}, \tilde{d}_2}, \dots, \mathbb{I}_{\ell}^{\tilde{L}, \tilde{d}_D}]$, along with a fifth-order polynomial of a location’s initial population, $\mathbf{L}_{\ell} = [\log(L_{\ell}), (\log(L_{\ell}))^2, \dots, (\log(L_{\ell}))^5]$. The latter absorbs the non-monotonic relationship between growth and size throughout most of U.S. history (Michaels et al., 2012; Desmet and Rappaport, 2017). It is necessary to include these terms because the size distribution of neighbors closely depends on a location’s own size. For example, very small locations rarely have a very large neighbor. As described in the previous subsection, we also extensively control for geographic attributes with 47 variables, \mathbf{x}_{ℓ} . We thus specify a data generating process with reduced form

$$g_{\ell} = \mathbf{I}_{\ell}^{\tilde{L}} \beta + \mathbf{L}_{\ell} \gamma + \mathbf{x}_{\ell} \delta + \varepsilon_{\ell}. \quad (1)$$

The partial correlation between the growth of a location and the presence of larger neighbors unsurprisingly depends both on the threshold population above which we consider neighbors to be large, \tilde{L} , and the size of the location itself, L_{ℓ} . Because the size distribution of U.S. locations changed continually throughout U.S. history, we use relative measures of population both to set year-specific thresholds for considering a location large and to focus the analysis on the growth of locations that are not large. Specifically, we consider locations to be at least “moderately large” in a given year if their population is at or above the 95th percentile of the distribution across locations in that year. Analogously, we respectively consider locations to be “very large” in a given year if their population is at or above the 99th percentile in that year. Reciprocally, we exclude locations from our baseline regression analysis that have population above the 80th percentile. Our baseline regressions thus estimate the partial correlations between the growth rate of small and medium locations—those with population in the first through fourth quintiles—with the presence of nearby locations in the top portion of the fifth quintile. We also discuss how these correlations differ across sub-samples of locations by size.

Partial correlations sensitively depend on the maximum distance, \tilde{d}_D , for which an indicator is included. To understand this, recognize that indicators of a large neighbor within distance intervals demarcated by $\{\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_D\}$, together with the excluded interval, $d_{\ell k} > d_D$, make up a disjoint set that fully partitions the observations in a regression. For a given maximum population

threshold, coefficients on each of the included indicators estimate the difference of predicted growth for observations with a corresponding positive value and the predicted growth of observations with a positive value of the excluded category. Estimated coefficients thus depend closely on the composition of the excluded category.

It is important to specify a maximum distance that is not too high. Failing to do so leaves few observations with a positive value for the excluded category. For example, in almost all years for which we run regressions, less than 15 percent of observations have no moderately large neighbor (one with population above the 95th percentile) within 300 kilometers. As these relatively isolated locations tended to grow slowly, a regression of growth on indicators for each of the distance intervals out to 300km must yield some positive coefficients. Hence it is important to choose a maximum distance that is not too large.

Conversely, it is also important to choose a maximum distance that is not too low. Many of the regressions estimate coefficients on the 50km-100km and 100km-150km indicators that are the same sign and similar in magnitude to their estimates on the 0km-50km indicator. For the 95th percentile and 99th percentile thresholds for large size, the number of observations with positive indicators for two further-away intervals far exceeds the number with positive indicators for the closest interval. In many cases, the majority of observations have positive values in the combined 50km-150km range. In consequence, regressions that include an indicator only for the 0km to 50km distance may not find much of a difference in predicted growth compared to locations in the excluded category.

To balance these two considerations, we specify our regressions to include indicators for 50km intervals out to the maximum distance that leaves at least 50 percent of observations in the excluded category.⁶ For example, 49 percent of the observations in the 1840 regression have a neighbor that is at least moderately large within 150km and 67 percent have one within 200km and so we use presence indicators for 0km to 50km, 50km to 100km, and 100km to 150km. Higher thresholds for considering a location to be large require a maximum distance that is further away. For the 1840 regression on the presence of neighbors that are very large (ones with population above 99th percentile), our rule implies including presence indicators out to a maximum distance of 300km.

2.3 Two Distinct Subperiods

This subsection explores the existence of urban growth shadows and urban growth spillovers between 1840 and 2017. When estimating the correlation of the population growth of small and medium locations with the presence of moderately large locations (population at or above the 95th

⁶Using the same maximum distance for all time periods yields qualitatively similar results (Appendix Tables A.1-A.3).

percentile), Table 1 shows two clearly distinct periods: a negative regime from 1840 and 1920, and a positive regime from 1920 through 2017.

The predicted population growth of small and medium locations was slower during each of the four 20-year periods from 1840 to 1920 if they had a moderately large neighbor. In 1840, the initial population of the small and medium locations ranged from 133 to 24,000 and the initial population of the 44 moderately large locations ranged from 62,000 to 425,000. Small and medium locations that had a moderately large neighbor within 50km had predicted annual population growth from 1840 to 1860 that was slower by 0.63 percentage points compared to the excluded locations, which did not have a moderately large neighbor within 150 kilometers. Locations whose nearest moderately large neighbor was between 50km to 100km away had predicted annual growth that was slower by 0.67 percentage points compared to excluded locations; and those whose nearest moderately large neighbor was 100km to 150km away had predicted lower annual growth that was slower by 0.30 percentage point. The corresponding negative coefficients statistically differ from zero at the 0.05 or 0.10 levels (respectively, dark and light blue typeface). Estimated negative coefficients are similar in magnitude for the 1860-1880 regression and a bit larger in magnitude for the 1880-1900 regression. Predicted growth from 1900 to 1920 was also slower for small and medium locations with a moderately large neighbor, although the magnitude of the difference compared to not having a moderately large neighbor was considerably less than during the earlier periods. Throughout the negative regime, the marginal share of the variation in growth accounted for by the indicators for a moderately large neighbors (the increase in R^2 compared to using only the control variables) is slight, ranging from 0.2 to 0.8 percentage points.⁷

The remaining columns of Table 1 describe the positive regime between population growth and the presence of a moderately large neighbor. For each of the five periods from 1920 to 2017, predicted population growth was faster for small and medium locations with a moderately large neighbor within 50km. For each of the three periods from 1940 to 2000, predicted growth was also slightly faster for locations whose nearest moderately large neighbor was located between 50km and 100km away. The corresponding positive coefficients statistically differ from zero at the 0.05 and 0.10 levels (respectively, dark and light red typeface). The magnitude of the faster predicted growth is relatively modest from 1920 to 1940, when suburbanization was just getting underway. Then, both from 1940-1960 and from 1960-1980, the presence of a moderately large neighbor within 50km predicted population growth that was higher by more than 1 percentage point (statistically significant at the 0.01 level). Smaller-magnitude coefficients seem to suggest that suburbanization waned from 1980 to 2000. But as we will describe in the next subsection, this is somewhat misleading, because it reflects many rapidly suburbanizing peripheral counties

⁷In Table 1 we refer to this as “ R^2 - R^2 controls”, i.e., the difference between the R^2 of our regression and the R^2 of a specification that only includes the controls (and hence leaves out the neighbor dummies).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
neighbor w/ pop ≥ 95th percentile @ distance:									
	1840- 1860	1860- 1880	1880- 1900	1900- 1920	1920- 1940	1940- 1960	1960- 1980	1980- 2000	2000- 2017
1 to 50 km	-0.63 (0.35)	-0.77 (0.28)	-1.10 (0.21)	-0.04 (0.14)	0.28 (0.12)	1.11 (0.15)	1.19 (0.19)	0.50 (0.14)	0.41 (0.20)
50 to 100 km	-0.67 (0.24)	-0.41 (0.20)	-0.86 (0.18)	-0.28 (0.09)	-0.07 (0.07)	0.16 (0.08)	0.21 (0.11)	0.26 (0.09)	0.11 0.09
100 to 150 km	-0.30 (0.12)	-0.11 (0.17)	-0.63 (0.15)						0.03 (0.07)
N (quints 1 to 4)	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
control vars	48	52	52	52	52	52	52	52	52
R²	0.857	0.787	0.726	0.533	0.407	0.366	0.391	0.427	0.318
Adj R²	0.846	0.778	0.718	0.521	0.393	0.352	0.376	0.412	0.298
R²-R² controls	0.002	0.002	0.008	0.002	0.003	0.028	0.033	0.010	0.004
pop≥95th pctl	62ths -425ths	49ths -1.4mn	51ths -2.5mn	65ths -4.9mn	80ths -8.5mn	100ths -11.7mn	172ths -14.2mn	250ths -14.5mn	377ths -18.3mn
N, ≥95th pctl	44	86	119	133	148	150	143	132	119
pop of obs	133 -24ths	103 -23ths	100 -26ths	104 -30ths	137 -32ths	285 -36ths	208 -42ths	408 -49ths	356 -65ths
share of pop	0.38	0.39	0.42	0.39	0.34	0.29	0.20	0.16	0.14

Table 1: **Population Growth and the Presence of a Moderately Large Neighbor.**

All regressions include a constant and control for initial population and up to 47 geographic covariates. Observations have population in the first four quintiles. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels. Dark and light red fonts respectively indicate a positive coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

having been reclassified as belonging to a metropolitan area following the 1970 and 1980 decennial censuses. The marginal share of the variation accounted for by the indicators of a moderately large neighbor is about 3 percentage points for the periods beginning in 1940 and 1960, but substantially lower for the other periods during the positive regime.

If we interpret the slower growth of locations with large neighbors as evidence of urban shadows, and the faster growth of those same locations as evidence of urban spillovers, then we can summarize our findings in Table 1 as follows:

Stylized Fact 1: Urban Shadows and Spillovers. *Between 1840 and 1920 urban growth shadows dominated the U.S. economic geography, with locations in the vicinity of large places growing relatively slower, whereas between 1920 and 2017 urban growth spillovers dominated, with locations in the vicinity of large places growing relatively faster.*

This division into a negative regime followed by a positive regime robustly holds for alternative threshold levels of largeness and widely varying specifications.

2.4 Recent Weakening of Urban Spillovers

In this subsection we analyze whether there has effectively been a weakening in urban spillovers since the 1980s, as suggested by some of the results reported above. To be precise, Table 1 showed that the expected growth boost from having a top-5 percent neighbor in the 1-to-50 kilometer range dropped by more than half, from 1.19 percentage points for the period 1960-1980 to 0.50 percentage points for the period 1980-2000. That fall may be partly explained by changing metro delineations: if a fast-growing location in one time period is also more likely to get absorbed into a metro area by the next time period, then this may cause a decline in growth of the locations in the 1-to-50 kilometer range. More generally, as re-delineated metro areas include more outlying counties, the continuing filling in of these counties is implicitly accounted for as migration within a location rather than between locations. This makes comparisons across periods more difficult.⁸

To assess the plausibility of this concern, Table 2 reports regressions for the three periods from 1960 to 2017 using metropolitan area borders established following the 1960 decennial census. Consistent with the possible bias we described, when keeping borders constant, the positive relationship between growth and the presence of large neighbors peaked 20 years later, during the period from 1980 to 2000, rather than during the period 1960 to 1980. In other words, we still see a weakening relation between growth and proximity to large locations, but only after the turn of the twenty-first century. This suggests that the transition of metropolitan areas to a larger geographic footprint may be winding down.⁹ We summarize these findings as follows:

Stylized Fact 2: Recent Weakening of Urban Spillovers. *Urban growth spillovers have been weakening since the turn of the 21st century. In particular, during the period 2000-2017 urban growth spillovers are less pronounced than during the periods from 1960-1980 and 1980-2000.*

⁸In addition, unobserved characteristics are likely to distinguish which surrounding counties at a given distance are absorbed into a metro, introducing a selection bias in making comparisons across periods. The re-delineations also leave fewer locations with nearby large neighbors, reflecting that metropolitan radiuses are becoming longer. The changing delineation of metros also affects metropolitan centroids, which are constructed as the population-weighted mean of constituent counties' centroids. Hence it also affects distances to large neighbors, which are measured between centroids.

⁹Regressing growth from 1960 to 1980 using metropolitan borders from 1940 modestly increases estimated coefficients on indicators of moderately large neighbors (compared to using 1960 borders) and modestly decreases estimate coefficients on indicators of very large neighbors. Regressing growth from 1940 to 1960 using metropolitan borders from 1920 modestly increases estimated coefficients on indicators of both moderately large and very large neighbors. Regardless of borders, the strength of suburbanization from 1940 to 1960 as estimated by the regressions was similar to the strength from 1960 to 1980.

	(1)	(2)	(3)
neighbor w/ pop ≥ 95th percentile @ distance:	1960- 1980	1980- 2000	2000- 2017
1 to 50 km	1.19 (0.19)	1.43 (0.30)	0.60 (0.12)
50 to 100 km	0.21 (0.11)	0.59 (0.23)	0.14 (0.06)
100 to 150 km		0.24 (0.20)	
150 to 200 km		0.13 (0.12)	
200 to 250 km			
N (quints 1 to 4)	2,283	2,282	2,283
control vars	52	52	52
R²	0.391	0.440	0.326
Adj R²	0.376	0.426	0.310
R²-R² controls	0.033	0.047	0.019
pop ≥ thresh	172ths -14.2mn	246ths -14.4mn	306ths- 15.7mn
N, ≥ thresh	143	143	143
pop of obs	208 -42ths	408 -54ths	67 -67ths

Table 2: **Population Growth and Large Neighbors, 1960 Metropolitan Borders.**

Metropolitan areas are delineated using the OMB standards following the 1960 decennial census. All regressions include a constant and control for initial population and 47 geographic covariates. Observations have population in the first four quintiles. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). 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 and light red fonts respectively indicate a positive coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

2.5 Geographic Span

This subsection explores how the geographic span of urban shadows and urban spillovers has changed over time. When focusing on moderately large neighbors (at or above the 95th percentile), as we have done so far, there are few observations with a positive value for the excluded category at far-away distances. This limits the maximum geographic distance we are able to consider, making it difficult to analyze how the geographic span of urban shadows and spillovers evolves over time. To get around this issue, Table 3 considers the presence of very large neighbors (at or above the 99th

percentile), allowing us to consider farther-away distances while maintaining enough observations with a positive value for the excluded category. As an example, for the period 1980-2000 we are able to include neighboring locations all the way to 250km, whereas for the same time period in Table 1 we only considered neighbors within a range of 100km.

Before discussing the spatial reach of urban shadows and spillovers, we show that increasing the size threshold from the 95th to the 99th percentile does not qualitatively change what we concluded before. There continues to be a negative regime and a positive regime, with the year 1920 separating the two (Table 3). The magnitudes of the coefficients of course differ, especially during the positive regime, when having a neighbor above the 99th percentile rather than above the 95th percentile was associated with a considerably greater boost in population growth. For example, predicted growth from 1960 to 1980 was 3.2 percentage points per year higher for small and medium locations that had a very large neighbor within 50km compared to the excluded locations, those whose nearest very large neighbor was at least 250km away.

We now analyze how the geographic reach of very large neighbors changes over time. During the negative regime, when comparing 1900-1920 to 1880-1900, the drop in growth from having a very large neighbor weakens at shorter distances below 50km but strengthens at farther-away distances above 150km. During the positive regime, the growth boost of having a very large neighbor starts off within a rather narrow 50km radius for the period 1920-1940, but then expands by 50km over each subsequent 20-year period, reaching 250km during 2000-2017. Of course, as is intuitive, growth's positive relationship with the presence of a very large neighbor weakens the more distant that neighbor is located. These findings constitute our third stylized fact:

Stylized Fact 3: Geographic Span of Shadows and Spillovers. *Over the period 1920-2017 there is strong evidence of the geographic span of urban growth spillovers expanding, with spillovers being very local between 1920-1940 and much more far-reaching in 2000-2017. Over the period 1840-1920 the evidence is mixed, though there is weak evidence of the geographic span of urban growth shadows expanding between the late 19th century and early 20th century.*

2.6 Relative Size of Locations and Neighbors

This subsection explores how the strength of urban shadows and spillovers depends on the relative size of locations and neighbors.

The Size of Locations. The qualitative relationship between a location's growth and the presence of a large neighbor is not too sensitive to the size of the locations, though the magnitude of the correlation shows some tendency to decline with size. In what follows, we make this point by focusing on two representative time periods, one for the positive regime, 1880 to 1900, and one for the negative regime, 1960 to 1980.

neighbor w/ pop ≥ 99th percentile @ distance:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1840- 1860	1860- 1880	1880- 1900	1900- 1920	1920- 1940	1940- 1960	1960- 1980	1980- 2000	2000- 2017
1 to 50 km	0.87 (0.67)	-1.01 (0.37)	-0.79 (0.42)	-0.28 (0.28)	0.73 (0.29)	2.41 (0.26)	3.20 (0.61)	0.99 (0.58)	-0.21 (0.39)
50 to 100 km	-0.09 (0.47)	-0.67 (0.30)	-0.87 (0.27)	-0.64 (0.19)	0.22 (0.19)	0.55 (0.15)	0.97 (0.25)	1.03 (0.19)	0.69 (0.15)
100 to 150 km	0.28 (0.47)	-0.60 (0.22)	-0.45 (0.24)	-0.47 (0.18)	0.08 (0.18)	0.14 (0.13)	0.30 (0.14)	0.51 (0.13)	0.39 (0.10)
150 to 200 km	0.49 (0.38)	-0.77 (0.18)	-0.15 (0.23)	-0.52 (0.17)	0.15 (0.16)	0.02 (0.13)	0.13 (0.12)	0.26 (0.10)	0.28 (0.08)
200 to 250 km	0.00 (0.24)		-0.02 (0.19)	-0.19 (0.16)	0.18 (0.15)	-0.12 (0.13)	0.03 (0.11)	0.10 (0.08)	0.22 (0.07)
250 to 300 km	0.16 (0.18)		-0.02 (0.19)	-0.13 (0.19)	0.16 (0.13)				0.10 (0.06)
N (quints 1 to 4)	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
control vars	48	52	52	52	52	52	52	52	52
R²	0.856	0.790	0.721	0.535	0.407	0.378	0.408	0.452	0.342
Adj R²	0.844	0.781	0.712	0.521	0.392	0.362	0.393	0.436	0.322
R²-R² controls	0.001	0.004	0.003	0.004	0.003	0.039	0.050	0.035	0.028
pop ≥ 99th pctl	171ths -425ths	139ths -1.4mn	139ths -2.5mn	197ths -4.9mn	321ths -8.5mn	442ths -11.7mn	810ths -14.2mn	1.3mn -14.5mn	2.0mn -18.3mn
N, ≥ 99th pctl	9	18	24	27	30	30	29	27	24
pop of obs	133 -24ths	103 -23ths	100 -26ths	104 -30ths	137 -32ths	285 -36ths	208 -42ths	408 -49ths	356 -65ths
share of pop	0.38	0.39	0.42	0.39	0.34	0.29	0.20	0.16	0.14

Table 3: **Population Growth and the Presence of a Very Large Neighbor.**

All regressions include a constant and control for initial population and up to 47 geographic covariates. Observations have population in the first four quintiles. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels. Dark and light red fonts respectively indicate a positive coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

For each of the first four quintiles of locations, Table 4 separately reports regressions of growth from 1880 to 1900 on the presence of a moderately large neighbor. All coefficients are estimated to be negative, with the magnitudes being lower for higher quintiles. When we consider growth of locations between the 80th and the 90th percentile, the magnitude of the correlation becomes even smaller, though it continues to be negative. Regressions for the other periods during the negative regime show similar results, both for the presence of moderately large and very large neighbors.¹⁰ Table 5 reports analogous regressions of growth from 1960 to 1980. All coefficients are

¹⁰The one exception concerns regressions of growth from 1880 to 1900 on the presence of very large neighbors. In

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<u>neighbor w/ population</u>	quints 1 to 4	quint 1	quint 2	quint 3	quint 4	decile 9	decile 10
≥ 95th pctile@							
1 to 50 km	-1.10 (0.21)	-4.78 (1.46)	-1.17 (0.26)	-0.35 (0.25)	-0.47 (0.11)	-0.25 (0.13)	0.07 (0.16)
50 to 100 km	-0.86 (0.18)	-1.13 (1.07)	-0.79 (0.31)	-0.40 (0.12)	-0.23 (0.10)		
100 to 150 km	-0.63 (0.15)	-1.84 (1.00)	-0.39 (0.19)				
150 to 200 km		-0.91 (0.97)					
200 to 250 km		-0.59 (0.96)					
250 to 300 km		-0.09 (0.96)					
300 to 350 km		-0.57 (1.01)					
N	1,844	422	476	472	474	237	237
control vars	52	52	50	51	52	48	50
R²	0.726	0.760	0.503	0.470	0.514	0.553	0.583
Adj R²	0.718	0.721	0.441	0.403	0.451	0.436	0.469
R²-R² controls	0.008	0.008	0.016	0.013	0.011	0.004	0.000
pop≥95th pctile	51ths -2.5mn	51ths -2.5mn	51ths -2.5mn	51ths -2.5mn	51ths -2.5mn	51ths -2.5mn	51ths -2.5mn
N, pop≥thresh	119	119	119	119	119	119	119
pop of obs	100	100	5.6ths	10.8ths	16.1ths	25.7ths	36.6ths
share of pop	-25.7ths 0.42	-5.6ths 0.02	-10.7ths 0.08	-16.1ths 0.13	-25.7ths 0.19	-36.6ths 0.14	-2.5mn 0.44

Table 4: **Population Growth by Quintile, 1880 to 1900.**

All regressions include a constant and control for initial population and up to 47 geographic covariates. Observations have population in the enumerated percentile range. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark blue and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

estimated to be positive, with in this case too some tendency, albeit weaker, for the magnitudes to decline with a location's size.

The Size of Neighbors. When comparing Table 1 and Table 3, we observe that growth's correlations with the presence of large neighbors increased with the size of neighbors during the positive

contrast to the combined regression, estimated coefficients using the first quintile of observations are positive.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<u>neighbor w/ population</u>	quints 1 to 4	quint 1	quint 2	quint 3	quint 4	decile 9	decile 10
≥ 95th pctile@							
1 to 50 km	1.19 (0.19)	1.69 (0.83)	1.34 (0.34)	1.03 (0.26)	1.16 (0.19)	0.27 (0.15)	0.12 (0.11)
50 to 100 km	0.21 (0.11)	1.12 (0.47)	0.15 (0.13)	0.18 (0.12)			
100 to 150 km		0.56 (0.31)					
150 to 200 km		0.60 (0.31)					
N	2,283	571	570	571	571	285	285
control vars	52	51	50	51	52	51	52
R²	0.391	0.521	0.452	0.459	0.410	0.427	0.647
Adj R²	0.376	0.470	0.397	0.404	0.349	0.298	0.566
R²-R² controls	0.033	0.022	0.032	0.034	0.086	0.005	0.001
pop≥95th pctile	172ths -14.2mn	172ths -14.2mn	172ths -14.2mn	172ths -14.2mn	172ths -14.2mn	172ths -14.2mn	172ths -14.2mn
N, pop≥thresh	143	143	143	143	143	143	143
pop of obs	208 -42.3ths	208 -7.8ths	7.8ths -13.5ths	13.5ths -21.3ths	21.3ths -42.3ths	42.3ths -79.5ths	79.5ths -14.2mn
share of pop	0.20	0.02	0.03	0.05	0.10	0.09	0.71

Table 5: **Population Growth by Quintile, 1960 to 1980.**

All regressions include a constant and control for initial population and up to 47 geographic covariates. Observations have population in the enumerated percentile range. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark and light red fonts respectively indicate a positive coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

regime but not the negative regime. A more general specification, however, establishes that magnitudes are increasing with the size of neighbors during both regimes. Table 6 shows results from regressing population growth on the presence of neighbors above four thresholds: the 80th, 90th, 95th, and 99th percentiles. These categories are nested in the sense that a neighbor that is above the 99th percentile is also above the 90th and 95th percentiles. Coefficients on these latter thresholds thus estimate the marginal boost to predicted growth compared to having a neighbor with population only above the next highest threshold. For example, a positive coefficient on the 99th percentile indicator estimates the additional predicted growth of locations that have a neighbor with population above the 99th percentile compared to locations with a largest neighbor with population between the 95th and 99th percentiles.

During the negative regime, the increase in the magnitude of growth's relationship with the

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<u>neighbor w/ population</u>	1840- 1860	1860- 1880	1880- 1900	1900- 1920	1920- 1940	1940- 1960	1960- 1980	1980- 2000	2000- 2017
≥ 80th pctile@									
1 to 50 km	-0.41 (0.21)	-0.72 (0.19)	-0.21 (0.12)	-0.30 (0.09)	0.10 (0.07)	-0.07 (0.08)	0.06 (0.07)	0.14 (0.06)	0.06 (0.05)
≥ 90th pctile@									
1 to 50 km	-0.37 (0.34)	0.25 (0.27)	-0.37 (0.19)	0.13 (0.14)	-0.03 (0.11)	0.30 (0.11)	0.15 (0.08)	0.21 (0.09)	0.00 (0.08)
50 to 100 km	-0.31 (0.25)	-0.19 (0.22)	-0.29 (0.16)						-0.01 (0.06)
≥ 95th pctile@									
1 to 50 km	-0.17 (0.43)	-0.44 (0.30)	-0.74 (0.23)	0.01 (0.20)	0.18 (0.15)	0.54 (0.17)	0.72 (0.15)	0.18 (0.13)	0.40 (0.21)
50 to 100 km	-0.46 (0.30)	-0.17 (0.22)	-0.54 (0.17)	-0.22 (0.10)	-0.09 (0.08)	0.06 (0.09)	0.04 (0.09)	0.17 (0.08)	0.04 (0.09)
100 to 150 km	-0.38 (0.13)	0.05 (0.20)	-0.60 (0.15)						0.00 (0.06)
≥ 99th pctile@									
1 to 50 km	1.07 (0.82)	-0.48 (0.36)	-0.09 (0.46)	-0.31 (0.29)	0.55 (0.26)	1.88 (0.28)	2.43 (0.59)	0.69 (0.52)	-0.56 (0.38)
50 to 100 km	0.42 (0.45)	-0.40 (0.30)	-0.53 (0.26)	-0.52 (0.20)	0.28 (0.18)	0.59 (0.16)	0.98 (0.24)	0.97 (0.17)	0.67 (0.15)
100 to 150 km	0.55 (0.46)	-0.59 (0.26)	-0.14 (0.23)	-0.42 (0.18)	0.09 (0.19)	0.16 (0.14)	0.28 (0.14)	0.50 (0.13)	0.39 (0.10)
150 to 200 km	0.65 (0.35)	-0.75 (0.19)	-0.04 (0.21)	-0.51 (0.17)	0.17 (0.16)	0.06 (0.13)	0.11 (0.11)	0.24 (0.09)	0.27 (0.07)
200 to 250 km	0.14 (0.23)		0.03 (0.16)	-0.21 (0.15)	0.19 (0.15)	-0.09 (0.13)	0.01 (0.11)	0.11 (0.08)	0.23 (0.07)
250 to 300 km	0.25 (0.20)		0.03 (0.16)	-0.13 (0.18)	0.16 (0.13)				0.12 (0.07)
N (quints 1 to 4)	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
control vars	48	52	52	52	52	52	52	52	52
R²	0.860	0.794	0.729	0.537	0.409	0.390	0.427	0.463	0.346
Adj R²	0.847	0.784	0.719	0.523	0.393	0.374	0.411	0.447	0.323
R²-R² controls	0.005	0.008	0.011	0.006	0.006	0.052	0.069	0.046	0.032
pop, pctile 80	24ths	23ths	26ths	30ths	32ths	36ths	42ths	49ths	65ths
pop, pctile 90	41ths	35ths	37ths	44ths	49ths	60ths	79ths	103ths	153ths
pop, pctile 95	62ths	49ths	51ths	65ths	80ths	100ths	172ths	150ths	377ths
pop, pctile 99	171ths	139ths	139ths	197ths	321ths	442ths	810ths	1.3mn	2.0mn

Table 6: Population Growth and the Size of Large Neighbors.

All regressions include a constant and control for initial population and up to 47 geographic covariates. Observations have population in the first four quintiles. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999).

population of its largest nearby locations is especially strong in the 1880-1900 regression. Having at least one neighbor within 50km that had population (weakly) above the 80th percentile is associated with slower predicted growth of 0.21 percentage point per year. If the largest such neighbor within 50km had population above the 90th percentile, predicted growth is slower by an additional 0.37 percentage point per year. If the largest such neighbor had population above the 95th percentile, predicted growth is slower by still an additional 0.74 percentage point per year. As an example, consider a location that has a neighbor with population at the 99th percentile between 50km and 100km away and no neighbor with population above the 90th percentile within 50km. During the period 1880-1900, such a location would have slower predicted population growth of 1.36 percentage point per year—the sum of the coefficients on the 50-to-100km indicators for the 90th, 95th, and 99th percentiles—compared to observations that do not have a neighbor in any of the categories included in the regression.

During the positive regime, the largest marginal increases in predicted growth are associated with having a neighbor with population at the 99th percentile rather than having one with population between the 95th and 99th percentiles. This is especially so during the 1960-1980 period, when the marginal increase was 2.43 percentage point per year. For neighbors located more than 50km away, only those with population at the 99th percentile are associated with a meaningful increase in predicted growth. For the period from 2000 to 2017, the statistically-significant boost from having a very large neighbor extends to those as much as 300km away. In contrast to the negative regime, the magnitude of the estimated differences in growth are modest for neighbors with population between the 80th and 90th percentiles.

Our findings of how the relative size of locations and neighbors affects the strength of urban shadows and spillovers can be summarized as follows:

Stylized Fact 4: Relative Size of Locations and Neighbors. *Urban shadows and urban spillovers tend to strengthen in the size difference between a location and its large neighbor. That is, the smaller a location and the larger its neighbor, the stronger urban shadows and urban spillovers.*

2.7 Regional Variation during the U.S. Westward Expansion

In this subsection we aim to understand to what extent the existence of urban shadows might have been related to the westward expansion of the U.S. during the 19th and early 20th centuries. If during that time period locations in the West, much of them isolated, grew fast because they were becoming settled, this would contribute to a negative correlation between local growth and proximity to a large neighbor. To assess this possibility, we rerun our baseline regression for three separate regions: the East, corresponding to the states covering the area of the original 13 colonies,

excluding Kentucky and Tennessee;¹¹ the Middle, corresponding to all other states east of the Mississippi River; and the West, corresponding to all states west of the Mississippi River.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<u>neighbor w/ pop ≥ 95th percentile @ distance:</u>	1860- 1880	1880- 1900	1900- 1920	1860- 1880	1880- 1900	1900- 1920	1860- 1880	1880- 1900	1900- 1920
	East			Middle			West		
1 to 50 km	-0.28 (0.31)	-0.31 (0.10)	0.50 (0.20)	-0.58 (0.24)	-0.62 (0.13)	0.06 (0.08)	-2.11 (1.88)	-1.78 (0.45)	-0.13 (0.27)
50 to 100 km [†]	-0.21 (0.26)			-0.46 (0.17)	-0.52 (0.14)		-2.08 (1.48)	-1.74 (0.31)	-0.38 (0.20)
100 to 150 km							-2.59 (1.29)	-1.41 (0.28)	-0.07 (0.18)
150 to 200 km							-2.30 (1.03)	-1.17 (0.26)	
200 to 250 km							-2.36 (0.90)	-0.94 (0.26)	
250 to 300 km							-1.77 (0.70)	-0.53 (0.23)	
300 to 350km							-1.41 (0.40)		
N (quints 1 to 4)	346	379	390	479	567	600	503	898	1,120
control vars	41	41	41	42	42	42	42	43	43
R²	0.527	0.445	0.363	0.742	0.690	0.389	0.828	0.761	0.589
Adj R²	0.459	0.375	0.286	0.716	0.664	0.342	0.810	0.748	0.572
R²-R² controls	0.004	0.004	0.017	0.004	0.009	0.000	0.007	0.008	0.001

Table 7: **Population Growth and Large Neighbors by Region, 1860 to 1920.**

All regressions include a constant and control for initial population and up to 47 geographic covariates. Observations have population in the first four quintiles of the national distribution. East region includes the states occupying the land area of the original 13 U.S. colonies excluding Kentucky and Tennessee. Middle region includes all other states east of the Mississippi River. West region includes all states west of the Mississippi River. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark blue and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

Table 7 reports our findings for the time period 1860-1920. Although, as expected, the magnitudes of the correlations are stronger in the more newly settled portions of the U.S. (the West) compared to portions that were settled earlier (the East), we continue to observe evidence of urban shadows in all regions of the country. Consistent with this variation across regions, the switch from the negative to the positive regime occurs 20 years earlier in the East. Thereafter, the

¹¹The land areas that would eventually constitute Kentucky and Tennessee were originally part of Virginia and North Carolina, respectively.

sign and magnitude of coefficients are similar to those for the nation (Appendix Tables A.4 and A.5).

3 Commuting Costs

One important force that shapes spatial growth dynamics in the hinterland of large population clusters relates to intra-city and inter-city commuting costs. As these will play a key role in the conceptual framework we present to interpret our empirical findings, in this section we briefly document how commuting costs in the U.S. have evolved since 1840. We start by focusing on changes in transportation technology, and then consider other factors that also affect commuting costs. In addition to documenting the long-term evolution of commuting costs, we also explore how the presence of local transportation infrastructure in the early twentieth century affects the strength of growth shadows in the cross-section.

3.1 Commuting Costs: 1840 to 2017

Transportation Technologies. Since the middle of the 19th century, there have been 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 (Fogel, 1964; Donaldson and Hornbeck, 2016), the building of canals (Shaw, 1990), the construction of the interstate highway system (Baum-Snow, 2007), and containerization (Bernhofen et al., 2016). To illustrate the magnitude of the decline in transport costs, 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 more central to improving short-distance transportation between neighboring or relatively close-by places. For the purpose of our paper, we are mostly interested in these latter improvements. In what follows we give a brief overview of the main innovations that have benefited short-distance transportation technology in the U.S. over the past two centuries.

Prior to the 1850s, many Americans worked near the central business district and walked to work. Other forms of transportation were expensive and slow. Horse-drawn carriages were available, but were only affordable to the very rich (LeRoy and Sonstelie, 1983).¹² The omnibus, a horse-drawn vehicle carrying twelve passengers, was first introduced in the 1820s and became more widely used in the 1840s. However, it was still a costly and not very fast way to travel.¹³ Commuter railroads appeared in the 1830s, although they were noisy and polluting, which led authorities to impose strict regulations, often limiting their use.¹⁴

¹²Regular steam ferry service began in the early 1810s but was limited to big coastal cities like New York.

¹³LeRoy 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.

¹⁴As in the case of the omnibus, commuter railroads were also quite expensive (LeRoy and Sonstelie, 1983).

Between 1850 and 1900 the U.S witnessed the arrival of the streetcar or trolley, which allowed for smoother travel and larger capacity than an omnibus. As with many other new modes of transport, initially only high-income individuals could pay the high fare of streetcars to commute to work on a regular basis. Nonetheless, the introduction of the streetcar allowed the larger cities to grow. Boston saw the first “streetcar suburbs”, well-off neighborhoods on the outskirts of the city (Warner, 1972; Mieszkowski and Mills, 1993; Kopeccky and Hon Suen, 2004). The streetcar was an important improvement over the omnibus in terms of capacity and speed: a two-horse streetcar could carry 40 passengers, and its speed was about one-third greater. Over time, animals were substituted by cleaner and more efficient motive powers. The first electric streetcar was operated in Montgomery, Alabama, in 1886, and by the end of 1903, 98 percent of the 30,000 miles of street railway had been electrified.¹⁵ By 1920, the streetcar had become an affordable mean to commute for almost every worker. However, by then the automobile had made its appearance, gradually replacing the streetcar as a way of commuting.

Several factors contributed to the streetcar facilitating longer-distance commutes, thus allowing large cities to grow bigger. One was an improvement in speed, another was the use of flat rates independent of distance, and a third was the construction of longer rail lines. In his study of Boston, Warner (1972) argues that the trolley triggered a substantial outward expansion of the city. In particular, he estimates this expansion to have been between 0.5 and 1.5 miles per decade. As Jackson (1985) explains, this translates into the outer limit of convenient commuting, defined as the distance that can be traversed in one hour or less, increasing from about 2 miles from Boston’s City Hall in 1850 to 6 miles in 1900, implying a considerably larger percentage increase in accessible land area.

While all these innovations significantly decreased transportation and commuting costs, it was not until the path-breaking invention of the automobile that these costs would experience radical change. The adoption of the car did not happen overnight: the affordability of automobiles for the middle class had to wait until the mass production of the Model-T in 1908. Other issues had to be resolved as well before cars could become wide-spread. Initially, regulations limited their use and speed to 4 miles per hour to avoid scaring horses. There was also a scarcity in gasoline stations and service facilities. More importantly, roads were still largely unpaved.

The growth in car ownership and use was tightly linked to the investment in roads and highways. New York opened the first part of its parkway system in 1908, which allowed drivers to increase their speed to 25 miles per hour. The Federal Highway Act of 1921 allowed the construction of similar highways across the country. In 1913, there was a motor vehicle to every eight people and, by the end of the 1920s, the car was used by 23 million people. The government effort was boosted years later with the Eisenhower Interstate Highway system, arguably the largest public

¹⁵Before the use of electricity, the use of steam engines was briefly tried, with limited success, in several U.S. cities.

works project in history and authorized by the Federal Highway Act of 1956. During this entire period, car ownership continued its upward ascent until the 1970s (Kopecky and Suen, 2004).

The combination of the mass use of the car and the expansion of the highway system translated into a huge wave of suburbanization, mostly in the post-WWII era. Many of these highways connected the downtown areas of large urban centers to the suburbs and the farther-off hinterland. According to Glaeser (2011), “the highway program was meant to connect the country, but subsidizing highways ended up encouraging people to commute by car”. 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 and Portland (Young, 2015).¹⁶

In addition to transport technology, other factors that determine the time cost of commuting are the spatial concentration of people and businesses, traffic congestion, and the opportunity cost of time. We turn to these factors next.

Spatial Clustering. Commuting costs fall if it becomes easier to fit more people or businesses onto an acre of land, since this implies less people needing to commute long distances. One major factor facilitating density is the possibility of building vertically. Historically, this move upward was at first modest, as two-story buildings were gradually replaced by four- and six-story buildings (Glaeser, 2011). Heights were restricted by the cost of construction and the limits on people’s desire to climb stairs. As a result, the top floors of six-story buildings were typically occupied by the lowest-income tenants (Bernard, 2014). This all changed with the invention of the elevator. A first elevator engine was presented by Elisha Otis at the 1854 New York’s Crystal Palace Exposition, but its rudimentary technology was unsuitable to be used in tall buildings. In 1880, Werner von Siemens’ electric elevator made it possible to transport people to tall heights in a safe manner, hence enabling the construction of skyscrapers with functional uses.

As Glaeser (2011) points out, another challenge that had to be overcome to build skyscrapers was an architectural constraint: erecting tall buildings required thick walls, making skyscrapers unprofitable. The solution to this problem was the use of load-bearing steel skeletons, where the weight of the building rests on a skeleton frame. Building these type of structures became possible in large part thanks to the increasing affordability of steel in the late 19th century. The first

¹⁶Suburbanization was also facilitated by factors unrelated to transport technology: the home mortgage interest deduction, the introduction of government-guaranteed mortgages, the Federal Housing Administration loans that guaranteed up to 95 percent of mortgages for middle-income buyers, and the GI Bill that offered no down payment housing loans for veterans.

skyscraper is often attributed to William Le Baron Jenney's Home Insurance Building, a 138-foot structure built in Chicago in 1885. In the following decades, skyscrapers became a fixture in the skylines of American cities, especially in Manhattan, which witnessed a boom in the number of skyscrapers in the 1920s.¹⁷

Congestion. The speed of commuting is of course not only a function of available technology. As traffic congestion has become worse, the most recent decades have witnessed a slowdown or even a reversal in the trend of ever-faster commuting. As one indicator of this growing congestion, we use the travel time index (TTI) of the Texas A&M Transportation Institute. The TTI is defined as the ratio of travel time in the peak period to travel time at free-flow conditions. For example, a value of 1.10 indicates a 20-minute free-flow trip takes 22 minutes in the peak period. Between 1990 and 2010, the TTI increased from around 1.10 to 1.20. As another indicator of congestion, we compute the average speed of trips under 50 km from the National Household Travel Survey. Between 1983 and 2001, this speed was still increasing, from 23.3 miles per hour to 26.4 miles per hour. Since then, this speed has declined, and by 2017 it had fallen by nearly one quarter, to 20.3 miles per hour. A similar pattern can be observed for trips between 50 and 100 km. The average speed increased from 45.2 mph in 1983 to 49.0 mph in 1995, and has since then declined, reaching 39.7 mph in 2017.

Opportunity Cost of Time. Another factor contributing to the increasing time cost of commuting is the rising 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. Edlund and co-authors find that the increase in the number of couples where both partners work has contributed to gentrification and urban renewal in recent decades. 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.¹⁸

Summary. When focusing on 1840-2017, the above discussion suggests that we can distinguish three subperiods in the evolution of commuting costs. Between 1840 and 1920, there was a gradual decline in commuting costs, driven by the introduction of the omnibus and the streetcar, followed by the incipient adoption of the car. After 1920, there was a rapid decline in commuting costs, driven

¹⁷The growth in the number of skyscrapers diminished after 1933, as a result of stringent regulations based on the argument that these tall buildings severely reduced the amount of light available to pedestrians.

¹⁸Of 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.

by the mass adoption of the automobile, the construction of highways connecting urban areas with their hinterlands, and the expansion of suburban rail systems. By the turn of the 21st century, this continuous decline in commuting costs slowed down, because of the increase in congestion and the rising opportunity of time.¹⁹

3.2 Commuting Costs: Cross-Sectional Evidence from Streetcars

In this subsection we explore how the variation in local commuting infrastructure affects local population growth. For the later period, after 1920, we already know from the work by Baum-Snow (2007) that highways were key in promoting the growth of suburbs and exurbs. For the earlier period, before 1920, we know less. Arguably, in the pre-1920 period the most important urban transportation infrastructure were streetcars. We therefore collect data on streetcars in the early 20th century to see how they affect the strength of urban shadows. More specifically, we ask whether the urban shadow of a large neighbor is stronger when it has a streetcar (extensive margin) or when it has more streetcar mileage (intensive margin).

Data on Streetcars. The data on streetcars come from the *U.S. Census Special Report on Street and Electric Railways*. We use the two earliest editions of this special report: 1902 and 1907 (U.S. Bureau of the Census, 1905, 1910). They provide for each streetcar line, the city and the length. Using a Python program, we match those cities to counties, and then manually correct any remaining errors.²⁰ For 1902, there are 559 out of 2617 counties had streetcars, corresponding to a total mileage of almost 20,000 miles. The corresponding numbers for 1907 are 734 out of 2847 counties having streetcars, corresponding to a total mileage of more than 30,000 miles. Many streetcars pass through multiple counties. In those cases, we allocate the total mileage of the streetcar line according to the counties' populations.

Specifications. To investigate how the presence (or length) of streetcars in the neighboring large location might affect the strength of urban growth shadows, we take the baseline specification (1) and introduce an interaction term between the presence of a large neighbor and that large neighbor having a streetcar. We also control for whether the own county has a streetcar or not. This yields the following specification:

$$g_{\ell} = \mathbf{I}_{\ell}^{\tilde{}} \beta + \mathbf{I}_{\ell}^{\tilde{}} \cdot \mathbf{S}_{\ell}^{\tilde{}} \alpha + s_{\ell} \theta + \mathbf{L}_{\ell} \gamma + \mathbf{x}_{\ell} \delta + \varepsilon_{\ell}, \quad (2)$$

¹⁹The years that separate the different subperiods do not constitute precise breakpoints. For example, we can use either 1920 or 1940 to separate the first two subperiods, as the mass adoption of cars started after 1908, whereas the building of urban highways and suburban rail networks only started in earnest in the 1950s and the 1960s.

²⁰More specifically, for each streetcar line, we compute the bilateral distances between any two counties on that line. Whenever that distance was large relative to the rail length, we manually checked the matching generated by the Python code.

where $\mathbf{S}_\ell^{\tilde{L}}$ are indicator variables that take a value of one if the corresponding large locations in $\mathbf{I}_\ell^{\tilde{L}}$ have a streetcar and s_ℓ is an indicator variable that measures whether ℓ has a streetcar or not. In addition to the presence of streetcars, we are also interested in the length of streetcars. We therefore consider an alternative, where $\mathbf{S}_\ell^{\tilde{L}}$ measures the miles of streetcars of the corresponding large locations in $\mathbf{I}_\ell^{\tilde{L}}$ and s_ℓ measures the miles of streetcars in ℓ . When measuring miles, we take a log transformation, and to avoid ignoring the extensive margin, we consider the log of one plus the mileage. Since we have two years of data on streetcars, we will run (2) for 1900-1920 using the streetcar data of 1902 and for 1910-1920 using the streetcar data of 1907.

We can also exploit the two years of streetcar data to consider a second specification:

$$g_\ell = \mathbf{I}_\ell^{\tilde{L}} \beta + \mathbf{I}_\ell^{\tilde{L}} \cdot \Delta \mathbf{S}_\ell^{\tilde{L}} \alpha + \Delta s_\ell \theta + \mathbf{L}_\ell \gamma + \mathbf{x}_\ell \delta + \varepsilon_\ell, \quad (3)$$

where $\Delta \mathbf{S}_\ell^{\tilde{L}}$ and Δs_ℓ represent the difference in $\mathbf{S}_\ell^{\tilde{L}}$ and s_ℓ between two time periods (1902 and 1907). As before, these differences could be either in the indicator variables for streetcars or in the log of streetcars mileage. We will run (3) for 1910-1920 using the difference in streetcar presence or mileage between 1902 and 1907.

Results on streetcars. Table 8 reports our findings from running specification (2). Column (1) analyzes how the presence of streetcars in 1902 affects growth between 1900 and 1920. The predicted population growth of small and medium locations was slower if they had a moderately large neighbor with a streetcar, and it was larger if they themselves had a streetcar, although the former correlation is not statistically significant. Column (2) considers the length of streetcars, rather than its mere presence, and reports similar results. The lack of statistical significance of the coefficient on the presence of streetcars in the large neighbor might partly be due to the fact that almost all large locations already had streetcars in 1902, so that the differential effect of having a streetcar is hard to identify. (To be more specific, of the 221 observations that have a large location in a radius of 1 to 50 km, only 19 of them have large neighbors with no streetcars.)

To partly address this issue, columns (3) and (4) lower the threshold of what we consider a large neighbor to locations with a population at the 90th percentile or above. The results are similar, though the coefficient on the presence of streetcars in a large neighbor at a distance up to 50km is now statistically significant at the 10% level. Columns (5) through (8) are analogous to columns (1) through (4), with two differences: the dependent variable is growth between 1910 and 1920 and the streetcar data are for 1907. The results are similar to the ones for the period 1900-1920 in the sense that the coefficients on the presence or length of streetcars in the large neighbors are always negative. One difference is that more coefficients are now statistically significant at either the 5% or 10% level. Also as before, locations that themselves have a streetcar are predicted to grow faster.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	neighbor w/ pop \geq 95 th		neighbor w/ pop \geq 90 th		neighbor w/ pop \geq 95 th		neighbor w/ pop \geq 90 th	
	1900-1920 streetcar presence	1900-1920 streetcar log miles	1900-1920 streetcar presence	1900-1920 streetcar log miles	1910-1920 streetcar presence	1910-1920 streetcar log miles	1910-1920 streetcar presence	1910-1920 streetcar log miles
1 to 50 km	0.02 (0.44)	0.05 (0.37)	0.12 (0.29)	-0.14 (0.21)	0.15 (0.32)	-0.16 (0.28)	0.86 (0.45)	0.32 (0.19)
streetcar 1-50	-0.17 (0.43)	-0.05 (0.08)	-0.44 (0.26)	-0.05 (0.05)	-0.20 (0.31)	-0.01 (0.05)	-0.84 (0.43)	-0.07 (0.05)
50 to 100 km	-0.26 (0.21)	-0.14 (0.16)	-0.31 (0.21)	-0.23 (0.19)	0.20 (0.23)	0.26 (0.18)	0.43 (0.29)	0.30 (0.21)
streetcar 50-100	-0.02 (0.22)	-0.04 (0.04)	-0.12 (0.18)	-0.06 (0.05)	0.41 (0.23)	-0.11 (0.04)	-0.44 (0.41)	-0.12 (0.05)
own streetcar	0.67 (0.15)	0.33 (0.06)	0.69 (0.15)	0.34 (0.07)	0.23 (0.12)	0.12 (0.06)	0.20 (0.12)	0.12 (0.06)
N (quints 1 to 4)	2,110	2,110	2,110	2,110	2,238	2,238	2,110	2,110
control vars	54	54	54	54	54	54	54	54
R ²	0.538	0.535	0.541	0.54	0.176	0.177	0.541	0.54
Adj R ²	0.525	0.523	0.528	0.527	0.154	0.156	0.528	0.527

Table 8: **Population Growth and Large Neighbors, 1900 to 1920: Streetcars.**

All regressions include a constant and control for initial population and 47 geographic covariates. Observations have population in the first four quintiles of the national distribution. Columns (1)-(4) are based on streetcar data of 1902 and have annualized growth between 1900 and 1920 as dependent variable, Columns (5)-(8) are based on streetcar data of 1907 and have annualized growth between 1910 and 1920 as dependent variable. All columns are based on specification (2), with odd columns using the presence of streetcars and even columns using $\log(1+\text{mileage of streetcars})$. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark blue and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

Table 8 suggests that streetcars strengthen the urban shadows of large locations. We reach this conclusion by observing that in all but one case the coefficients on the presence (or length) of streetcars in neighboring large locations are negative. This conclusion is suggestive in the sense that the negative coefficients are often not statistically significant. Needless to say, our results cannot be interpreted as causal. While it could of course be that better commuting infrastructure in the neighboring large city attracts population from neighboring locations, it is also possible that the absorption of closeby population by the large city prompts it to build better commuting infrastructure.

Table 9 reports our findings from running specification (3). Column (1) analyzes how the change in the presence of streetcars between 1902 and 1907 correlates with growth between 1910 and 1920. Predicted growth of small and medium locations was lower if they had a large neighbor

	(1)	(2)	(3)	(4)
	neighbor w/ pop \geq 95 th percentile		neighbor w/ pop \geq 90 th percentile	
	1900-1920 change in streetcar presence	1900-1920 change in streetcar log miles	1900-1920 change in streetcar presence	1900-1920 change in streetcar log miles
1 to 50 km	0.06 (0.17)	0.02 (0.15)	0.09 (0.14)	0.08 (0.14)
change in streetcar 1-50	-1.02 (0.40)	-0.28 (0.13)	-0.18 (0.18)	-0.09 (0.07)
50 to 100 km	-0.17 0.12	-0.12 (0.11)	-0.17 (0.11)	-0.14 (0.11)
change in streetcar 50-100	-0.07 (0.32)	-0.25 (0.12)	0.28 (0.20)	0.04 (0.09)
change in own streetcar	0.05 0.13	0.14 (0.06)	0.02 (0.13)	0.13 (0.07)
N (quints 1 to 4)	2,238	2,224	2,238	2,224
control vars	54	54	54	54
R²	0.176	0.178	0.176	0.175
Adj R²	0.154	0.156	0.155	0.154

Table 9: **Population Growth and Large Neighbors, 1900-1910: Change in Streetcars.**

All regressions include a constant and control for initial population and 47 geographic covariates. Observations have population in the first four quintiles of the national distribution. Specification is (3) and dependent variable is annualized growth between 1910 and 1920. Odd columns use the difference in presence of streetcars between 1902 and 1907 and even columns use the difference in $\log(1+\text{mileage of streetcars})$ between 1902 and 1907. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark blue and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

that did not have a streetcar in 1902 but had one by 1907, and it was higher if they themselves got their first streetcar between 1902 and 1907. Column (2) is similar, but considers the log difference in streetcar length between 1902 and 1907. It shows that an expanding streetcar network in the nearby large location is predictive of lower growth, whereas expanding streetcar mileage in the own location is predictive of higher growth. In both columns changes in streetcar presence or length are mostly statistically significant, either at the 5% or 10% level. Columns (3) and (4) are analogous to columns (1) and (2), with the difference that the threshold for being considered a large county is now the 90th percentile. The results are similar, but the coefficients are mostly not statistically significant. From Table 9 we can conclude that large counties that either got their first streetcar or expanded their network of streetcars strengthened their urban shadow on nearby smaller locations. This is reflected in the negative coefficients on the change in the presence (or length) of streetcars in neighboring large locations.

Together, our findings in Tables 8 and 9 can be summarized as follows:

Stylized Fact 5: Urban Shadows and Local Commuting Infrastructure. *In the early 19th century, urban growth shadows were stronger when large locations disposed of better commuting infrastructure in the form of streetcars.*

4 Conceptual Framework

In this section we provide a two-city spatial model with commuting and moving costs that is able to account for the main stylized facts identified in the data.²¹ On the one hand, the smaller city may find it hard to survive in the shadow of the larger city, as its residents prefer to move to the more productive neighbor. On the other hand, the smaller city may thrive as its residents can access the neighbor's higher productivity through commuting. As commuting costs decline, we find that urban growth shadows dominate in the early stage, whereas urban growth spillovers dominate later on.

4.1 Setup and Equilibrium

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, denoted by ℓ and k . The set of individuals living closer to production point ℓ than to production point k comprises city ℓ . Whereas the total population, L , is exogenous, the populations of the two cities are endogenous. 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.²²

Technology. 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 . The price of the good is normalized to one. To produce, an individual needs to commute from his residence to one of the two production points.

The time cost of commuting per unit of distance is γ . An individual who resides in city ℓ at a distance d_ℓ from production point ℓ can choose between working in ℓ or k . If she works in her own city ℓ , she supplies one unit of labor net of the time lost in intra-city commuting $1 - \gamma d_\ell$, and earns an income of $A_\ell(1 - \gamma d_\ell)$. If she commutes to the other city k , we ignore differences in the

²¹Appendix B.2 provides an alternative model with inter-city trade costs and intra-city commuting costs that is also consistent with the shift from urban shadows to urban spillovers.

²²This ensures symmetry in a city's spatial structure on both sides of its production point.

residence location in the own city, and assume that she incurs an inter-city commuting distance $d_{\ell k}$.²³ In that case, she supplies $1 - \gamma d_{\ell k}$ units of labor, and earns an income $A_k(1 - \gamma d_{\ell k})$. Without loss of generalization, we assume that no one residing in k has an incentive to commute to ℓ . As a result, an individual residing in city k at a distance d_k from production point k supplies $1 - \gamma d_k$ units of labor, earning an income of $A_k(1 - \gamma d_k)$. Summarizing, depending on where an individual resides and works, there are three possible expressions for income y :

$$\begin{aligned} y_{\ell}(d_{\ell}) &= A_{\ell}(1 - \gamma d_{\ell}) \\ y_k(d_k) &= A_k(1 - \gamma d_k) \\ y_k(d_{\ell k}) &= A_k(1 - \gamma d_{\ell k}), \end{aligned} \tag{4}$$

where the subscript on y refers to the individual's workplace and the subscript on d refers to his commuting distance, and hence implicitly to his place of residence. For example, $y_k(d_{\ell k})$ refers to the income of an individual who works in k and covers a distance $d_{\ell k}$ to get to work, implying that she lives in ℓ .

The possibility of commuting from city ℓ to production point k is meant to capture the positive effect of urban 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 de facto discounted version of the neighboring city's productivity, $A_k(1 - \gamma d_{\ell k})$.²⁴ We could alternatively model this effect through direct technology spillovers, without the need of introducing inter-city commuting, as in Ahlfeldt et al. (2015): if technological spillovers decay at a rate of γ per unit of distance, then an agent who resides and works in city ℓ would have access to a discounted version of his neighbor's productivity, $A_k(1 - \gamma d_{\ell k})$, in exactly the same way as an agent who resides in ℓ and pays a time cost of γ per unit of distance to commute to k . In that sense, both interpretations are interchangeable in their effects on income. Hence, our results do not strictly hinge on the existence of inter-city commuting. We will return to this alternative interpretation when we discuss the model's results.²⁵

Preferences. Utility is equal to income, with two adjustments. First, agents who change residence from city ℓ to city k pay a consumption-equivalent utility moving cost $\mu d_{\ell k}$ that is increasing

²³This 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 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 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 .

²⁴If we include the inter-city commuting time as part of the necessary time dedicated to work, we can interpret $A_k(1 - \gamma d_{\ell k})$ as the productivity of a commuter from the other city.

²⁵Yet another alternative would be to model this effect by introducing trade. We provide such a model in Appendix B.2.

in inter-city distance. As in the case of commuting, without loss of generality, no individual originally from k has an incentive to move to ℓ . Because of the moving cost, an agent's utility depends not just on where she resides and where she works, but also on where she is originally from. Second, occupying land is costly. In particular, an individual residing at a distance d_ℓ from the production point ℓ pays a land rent of $r_\ell(d_\ell)$, which reduces her net income and hence her utility.

Therefore, depending on an individual's place of origin, place of residence and place of work, there are four possible expressions for utility:

$$\begin{aligned}
u_\ell^\ell(d_\ell) &= A_\ell(1 - \gamma d_\ell) - r_\ell(d_\ell) \\
u_k^\ell(d_k) &= A_k(1 - \gamma d_k) - r_k(d_k) - \mu d_{\ell k} \\
u_k^\ell(d_\ell) &= A_k(1 - \gamma d_{\ell k}) - r_\ell(d_\ell) \\
u_k^k(d_\ell) &= A_k(1 - \gamma d_k) - r_k(d_k),
\end{aligned} \tag{5}$$

where the superscript on u refers to the individual's place of origin, the subscript on u refers to her workplace, and the subscript on d refers to her commuting distance, and hence implicitly to her place of residence. For example, $u_k^\ell(d_{\ell k})$ refers to the utility of an individual who is originally from ℓ , resides in ℓ and works in k , whereas $u_k^\ell(d_k)$ refers to the utility of an individual who is originally from ℓ , resides in k and works in k . To simplify notation, later in the paper we will sometimes refer to $u_\ell^\ell(d_\ell)$ as the staying utility U_S , to $u_k^\ell(d_k)$ as the moving utility U_M , and to $u_k^\ell(d_\ell)$ as the commuting utility U_C .

We now provide some more details about the moving cost $\mu d_{\ell k}$, and also relate it to the concept of urban shadows. Rather than interpreting the moving cost as a time cost, we think of it 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 behind. 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, and the larger city casts a growth shadow on the smaller city.

Residential Mobility within Cities. People can freely locate within cities. Where land is unoccupied, land rents are normalized to zero. Hence, at the city edge \bar{d}_ℓ land rents $r_\ell(\bar{d}_\ell) = 0$, whereas at other locations closer to the production center ℓ land rents are determined by the within-city residential free mobility condition. The same of course applies to city k .

To determine equilibrium land rents at different locations, note that in city ℓ there are potentially two types of agents: residents who work locally in ℓ , denoted by L_ℓ^ℓ , and residents who

commute to k , denoted by L_k^ℓ . The total cost of land rents and commuting costs incurred by a resident who lives at distance d_ℓ and works locally is $r_\ell(d_\ell) + A_\ell\gamma d_\ell$, whereas the analogous cost if she commutes to k is $r_\ell(d_\ell) + A_k\gamma d_{\ell k}$. Since all commuters to the other city k have to cover the same distance $d_{\ell k}$, independently of where they reside in city ℓ , they all prefer to live on the city edge and pay zero rent. As a result, there will be an area $L_k^\ell/2$ on both edges of the city where rents are zero. To be precise, for all $d_\ell \in [\bar{d}_\ell - L_k^\ell/2, \bar{d}_\ell]$ we have $r_\ell(d_\ell) = 0$. For all other locations closer to production point ℓ , occupied by residents who work locally, the sum of land rents plus commuting costs must equalize. Hence, for all $d_\ell \in [0, \bar{d}_\ell - L_k^\ell/2]$, we have $r_\ell(d_\ell) + A_\ell\gamma d_\ell = A_\ell\gamma(\bar{d}_\ell - L_k^\ell/2)$, so that $r_\ell(d_\ell) = A_\ell\gamma(\bar{d}_\ell - L_k^\ell/2 - d_\ell)$. Summarizing, equilibrium land rents in city ℓ are:

$$r_\ell(d_\ell) = \begin{cases} A_\ell\gamma(\bar{d}_\ell - \frac{L_k^\ell}{2} - d_\ell) & \text{if } d_\ell \in \left[0, \bar{d}_\ell - \frac{L_k^\ell}{2}\right] \\ 0 & \text{if } d_\ell \in \left[\bar{d}_\ell - \frac{L_k^\ell}{2}, \bar{d}_\ell\right]. \end{cases} \quad (6)$$

In city k , without loss of generality, no residents commute to ℓ , so land rents are simply:

$$r_k(d_k) = A_k\gamma(\bar{d}_k - d_k). \quad (7)$$

City Choice and Commuting Choice. The smaller city will lose population if its residents prefer to move to the larger neighbor, but it will gain population if its residents choose to commute to the larger city. To provide some intuition for when one situation is more likely than another, consider an individual who is originally from city ℓ and resides at a distance d_ℓ from the production point ℓ . She has three choices: she can stay in city ℓ and work at production point ℓ , earning a utility $U_S \equiv u_\ell^\ell(d_\ell)$; she can move to city k and reside at a distance d_k from her work at production point k , earning a utility $U_M \equiv u_k^\ell(d_k)$; or she can commute a distance $d_{\ell k}$ from city ℓ to production point k to work, earning a utility $U_C \equiv u_k^\ell(d_\ell)$. The expressions in (5) 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.

Building on this intuition, we can now characterize the equilibrium of the economy in terms of where individuals choose to reside and where they choose to work. We do so 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 which determine the size of both cities when populated by their original residents. 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 if all individuals work in the city they are originally from, the utility of a resident of ℓ is less than or equal to that of a resident of k .

Depending on the parameter values and on the initial sizes of both cities, the economy will be in one of four equilibria, represented by the four quadrants of Figure 1. First, if original residents

of 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 original residents of 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.²⁶ If, as illustrated in the top-right panel of Figure 1, the two utility levels equalize at a level above the utility from commuting, then we will say that we are in an *inter-city moving equilibrium*: some individuals of ℓ move to k , and the remainder lives and works in ℓ .²⁷

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 , until the utility from staying equalizes that of commuting. 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 from 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.²⁸ 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 of the economy for a given set of parameter values and for a given initial size of ℓ and k .

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:*

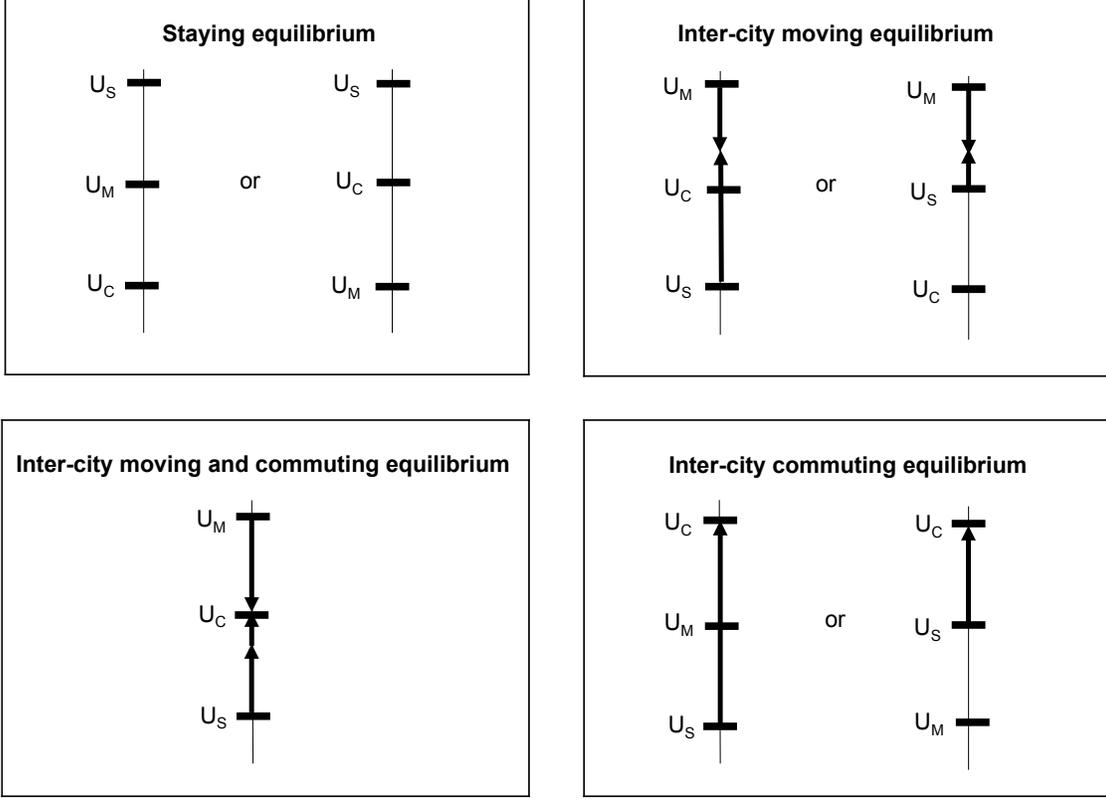
- i. *Staying equilibrium. If $A_\ell(1 - \gamma\bar{d}_\ell^0) \geq A_k(1 - \gamma\bar{d}_k^0) - \mu d_{\ell k}$ and $A_\ell(1 - \gamma\bar{d}_\ell^0) \geq A_k(1 - \gamma d_{\ell k})$, then no individual has an incentive to move from ℓ to k .*
- ii. *Inter-city moving equilibrium. If either $A_k(1 - \gamma\bar{d}_k^0) - \mu d_{\ell k} > A_\ell(1 - \gamma\bar{d}_\ell^0) \geq A_k(1 - \gamma d_{\ell k})$ 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 $A_k(1 - \gamma(\bar{d}_k^0 + m)) - \mu d_{\ell k} = A_\ell(1 - \gamma(\bar{d}_\ell^0 - m))$.*

²⁶The utility from commuting remains the same, since that utility depends on the distance between city ℓ and city k , which is unchanged.

²⁷It is also possible that all individuals move out of city ℓ before the two utility levels meet. This possibility is not shown in Figure 1.

²⁸Of course, if everyone commutes before that equality is reached, then we would have the entire city ℓ commuting to k .

Figure 1: Equilibrium Description



Given initial conditions, this figure graphically illustrates the four possible equilibrium configurations. Horizontal lines denote the initial utility levels for the different choices: U_S refers to the utility of an individual staying and working in her own city, U_M refers to the utility of an individual moving to the other city and working there, and U_C refers to the utility of an individual commuting to the other city. In the top-left corner individuals do not gain from either moving or commuting to the other city, so we have a *staying equilibrium*. In the top-right corner and bottom-left corner individuals get a higher utility from moving than from commuting or staying. If moving leads the utility to equalize to that of staying, we get a *moving equilibrium*, whereas if it leads the utility to equalize to that of commuting, we get a *moving and commuting equilibrium*. In the bottom-right corner individuals get a higher utility from commuting, so we have a *commuting equilibrium*.

- 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 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 $A_k(1 - \gamma d_{\ell k}) = A_\ell(1 - \gamma(\bar{d}_\ell^0 - c))$.

Now that we have formally defined the equilibrium, in the following subsection we analyze whether the smaller city gains or loses population when certain parameter values, such as commuting costs, change. This will provide us with valuable predictions on the incidence of urban shadows and urban spillovers, and it will allow us to link the theory's predictions to the stylized facts uncovered in the empirical part of the paper.

4.2 Urban Growth Shadows and Spillovers

In this subsection we explore how a drop in commuting costs affects urban shadows and urban spillovers. To gain some understanding of the role of commuting costs, consider an agent residing in the smaller, less productive city. 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 benefiting from low commuting costs. This intuition suggests that a drop in γ makes moving relatively more attractive than staying, and makes commuting relatively more attractive than moving.

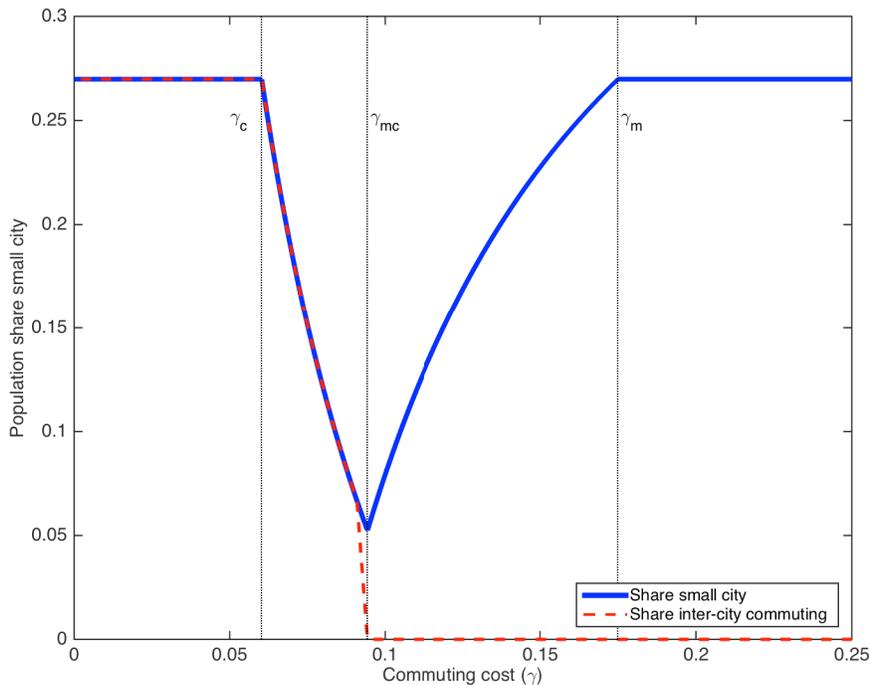
Starting off in a situation where all individuals have the same utility, we formalize this intuition and show that a gradual decrease in γ first shifts the economy from a staying equilibrium to an inter-city moving equilibrium, with some residents of the smaller low-productivity city moving to the larger high-productivity city. Later, as γ continues to drop, the economy shifts to an inter-city moving and commuting equilibrium, and then 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 sequentially from a staying equilibrium to an inter-city moving equilibrium, 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, whereas in the inter-city moving and commuting equilibrium the smaller city gains residents from the larger city.*

Proof. See Appendix B.1.

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 vs. Commuting Cost



This figure illustrates Result 1. It shows how the share of the population of the small city, as well as the share of commuters, changes with commuting costs. A drop in commuting costs leads the small city to first lose population, and then to gain population. For high commuting costs ($\gamma \geq \gamma_m$), individuals from the small, less productive city have no incentive to move or commute to the large, more productive city. As commuting costs drop ($\gamma_m > \gamma \geq \gamma_{mc}$), individuals from the small, less productive city move to the large, more productivity. The population of the small city declines: an urban growth shadow. For low levels of commuting costs ($\gamma_m < \gamma_{mc}$), individuals from the small, less productive city commute to the large, more productive city, and its original residents start to return: an urban spillovers. Eventually, the small city recovers its original population ($\gamma < \gamma_c$).

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 remains 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 residents from the small city start to commute to the large city. As the commuting cost further drops, the population of the small city slowly recovers, as some of the movers return and prefer to commute. Once the commuting cost falls to γ_c , the small city reaches its original population level, with all of its residents commuting to the large city.

Relation to Urban Shadows and Spillovers. What does Result 1 tell us about urban growth shadows and urban growth spillovers? As the commuting cost drops, residents of the smaller city move to the nearby larger city, and the smaller city loses population. The larger city casts an urban growth shadow: the nearby smaller city suffers in terms of population growth. Needless to say, if the initial drop in commuting is larger, the urban growth shadow is stronger. A further drop in the commuting cost reverses this trend, as residents of the smaller city find it more attractive to commute to the larger city than to move. The larger city no longer displays an urban growth shadow, but exhibits urban spillovers instead: the nearby smaller city gains in terms of population growth.

Relation to Empirical Stylized Facts. Our description of the evolution of commuting costs in the U.S. between 1840 and 2017 suggests a slow decline in γ between 1840 and 1920, a rapid fall in γ between 1920 and the turn of the 21st century, and a slowdown in the decrease in γ during the last two decades. In light of Result 1, this would be consistent with an early time period where growth shadows dominated and a later time period where growth spillovers dominated, with a weakening in those spillovers in more recent times. During the early time period, Result 1 also shows that growth shadows are stronger in locations that experience greater improvements in commuting infrastructure. This is consistent with our empirical findings for the U.S., as summarized in Stylized Fact 1, Stylized Fact 2 and Stylized Fact 5.

4.3 Geographic Span of Urban Shadows and Spillovers

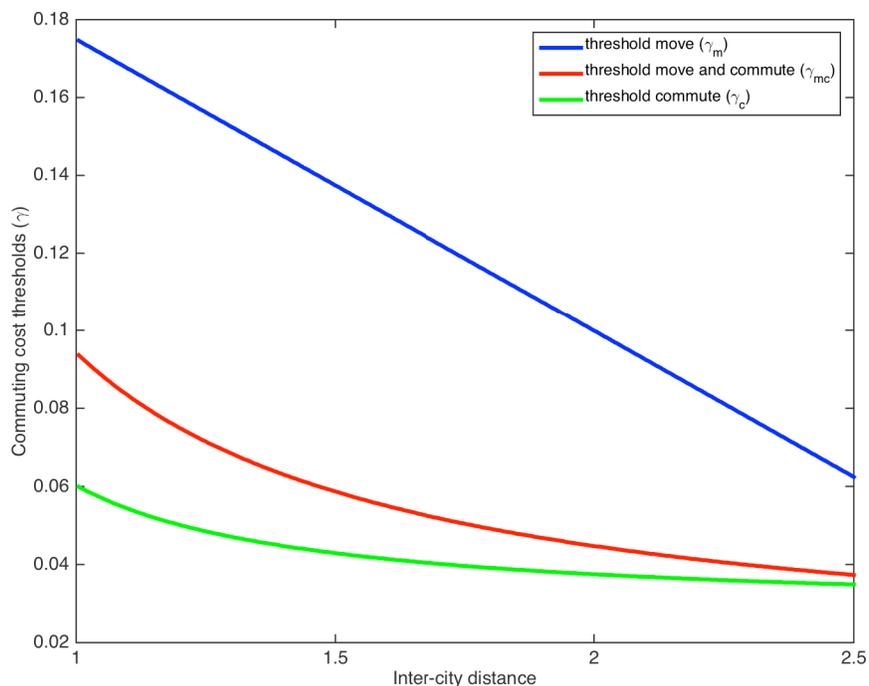
We now explore how urban shadows and urban spillovers depend on the distance to the larger city. The following result states that if inter-city distance increases, then all three threshold values of the commuting cost are lower.

Result 2. *Thresholds γ_m , γ_{mc} and γ_c are declining in $d_{\ell k}$. That is, if the distance to the larger city increases, the shift from a staying equilibrium to an inter-city moving equilibrium, from an inter-city moving equilibrium to an inter-city moving and commuting equilibrium, and from an inter-city moving and commuting equilibrium to an inter-city commuting equilibrium, occurs for lower values of the commuting cost γ .*

Proof. See Appendix B.1.

The above result says that when the larger city is geographically farther away, 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.

Figure 3: Inter-city Moving and Commuting Thresholds vs. Inter-City Distance



This figure illustrates Result 2 by showing how the three commuting cost thresholds (γ_m , γ_{mc} and γ_c) decrease as inter-city distance increases. This result implies that as commuting costs decline, urban shadows first expand in space as the moving equilibrium applies to increasingly farther-away locations, and then urban spillovers expand in space as the moving & commuting equilibrium applies to increasingly farther-away locations.

Figure 3 illustrates Result 2 with a simple numerical example. We use the same parameter values as before, with the exception of inter-city distance $d_{\ell k}$, which we now vary from 1.0 to 2.5.²⁹ For each level of inter-city distance, we plot the three threshold values. As can be seen, the larger the inter-city distance, the more commuting costs need to drop before people start moving to the larger city, and before they start commuting to the larger city.

Relation to Geographic Span of Shadows and Spillovers. What do these findings tell us about the geographic span of urban shadows and urban spillovers? Using Figure 3, consider a commuting cost that is relatively high, say, $\gamma = 0.14$. In that case, population growth in the

²⁹To make the results comparable to the other figures, for each value of $d_{\ell k}$, population is allocated across cities such that utility equalizes for $\gamma = 0.25$.

smaller city is relatively lower if the larger city is close-by than if it is farther away. We see urban growth shadows at close distances, and no effect at further distances. As the commuting cost falls from this relatively high level, the larger city’s urban growth shadow increases its geographic reach. Once commuting costs become low enough, we see the emergence of urban spillovers. Consider, for example, a commuting cost of $\gamma = 0.07$. In that case, population growth is relatively higher if the larger city is in the vicinity, and it is relatively lower if the larger city is at a greater distance. That is, urban spillovers dominate at short distances, and urban shadows dominate at farther distances. As the commuting cost falls further, the spatial reach of urban spillovers increases.

Relation to Empirical Stylized Facts. Result 2 allows us to trace the changing geographic reach of urban shadows and spillovers as commuting costs fall. Initially, it predicts urban shadows at relatively short distances, that gradually expand as transport costs drop. Eventually, these shadows are replaced by spillovers, again first at relatively short distances, but later at farther away distances as the spatial reach of spillovers expands. This is consistent with our empirical findings for the U.S., as summarized in Stylized Fact 3.

4.4 Relative Size of Large City

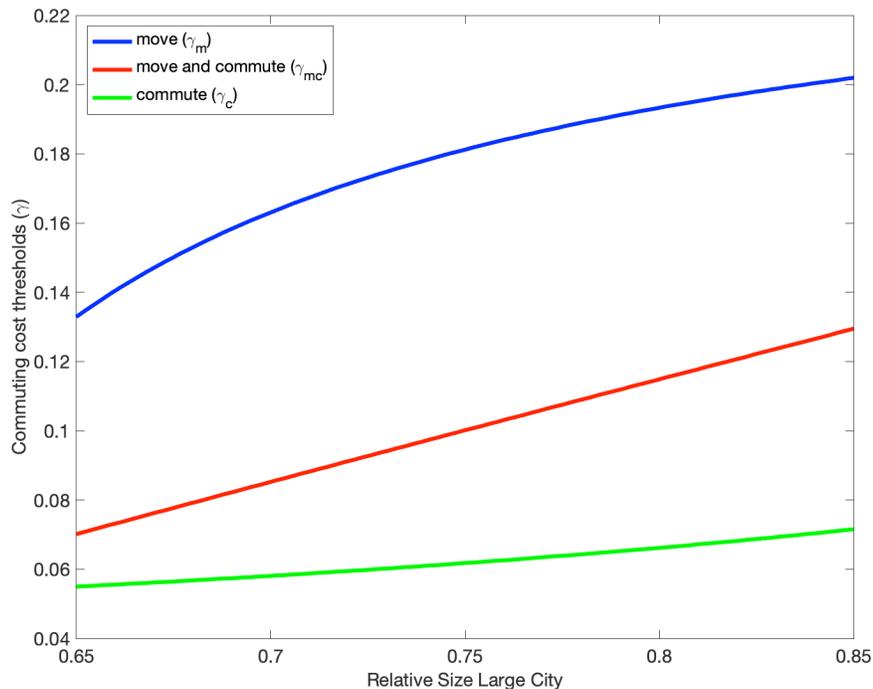
How do urban shadows and urban spillovers depend on the relative size of the large city? The following result shows that the moving threshold is increasing in the relative size of the large city. That is, commuting costs have to fall by less before a large city starts attracting the population of its hinterland.

Result 3. *Keeping population-weighted productivity unchanged, the threshold γ_m is increasing in the relative size of the larger city. That is, the shift from a staying equilibrium to an inter-city moving equilibrium occurs for a higher value of the commuting cost if the larger city has a bigger relative size.*

Proof. See Appendix B.1.

The above result shows that larger cities exert a stronger gravitational pull on their hinterland, as they start casting their urban shadows at higher levels of commuting costs. Figure 4 illustrates this result, and further shows that the same applies to urban spillovers. We use the same parameter values as before, with the exception of the productivity A_k which we now vary in order for the city size to change. As the relative size of k increases from 0.65 to 0.85, all three threshold values increase.

Figure 4: Inter-city Moving and Commuting Thresholds vs. Inter-City Distance



This figure illustrates Result 3 by showing how the three commuting cost thresholds (γ_m , γ_{mc} and γ_c) increase as the relative size of the large city increases. This result implies that as commuting costs decline, it is the largest cities that first cast their growth shadow on their smaller neighbors, and likewise, it is the largest cities that first spread their growth spillovers to their smaller neighbors.

Relation to Empirical Stylized Facts. Result 3 implies that as commuting costs decline, it is the largest cities that first cast their urban growth shadow on their smaller neighbors, and likewise, it is the largest cities that first spread their urban growth spillovers to their smaller neighbors. The stronger urban shadows and urban spillovers of larger cities is consistent with our empirical findings for the U.S., as summarized in Stylized Fact 4.

4.5 Alternative Interpretations

Commuting costs are central to our conceptual framework. Their evolution affects whether urban growth shadows or urban growth spillovers dominate, and they are also key in determining the geographic reach of these shadows and spillovers. Indeed, by focusing on the evolution of just this variable, our conceptual framework is able to account for the main stylized facts we identified when empirically studying growth shadows and spillovers in the U.S. over the period 1840 to 2017.

One potential issue with our interpretation is that in some of the later time periods, after 1980, the geographic span of urban spillovers reached 200km. At face value this seems well beyond standard inter-city commuting distances, so one could doubt whether in this most recent time period the conceptual framework captures the essence of what we observe in the data. There are at least

three reasons why our interpretation may still hold. First, although between 1980 and 2017 we find evidence of urban spillovers having a large geographic reach, those effects dissipate with distance. For example, correlations between 150km and 200km are one-half to one-quarter their magnitudes between 1km and 100km. Second, the empirical correlations should always be interpreted relative to the excluded category (e.g., locations that have no large neighbors within 300km). If in recent time periods geographically isolated locations have been experiencing particularly low growth, this pushes up the relative growth rate of all other locations, including those that are, say, 200km away from large neighbors. Third, although a distance of 200km between the centroid of a rural county and the centroid of a large metro area may be beyond standard commuting distances, the distance between that same county and the edge of a large metro area may very well still be within reasonable commuting time.

That being said, an alternative is that inter-city commuting costs are not the driving force behind what we observe in the data. As mentioned before, our conceptual framework is equivalent to one with intra-city (but no inter-city) commuting costs, and with technological spillovers that decay with distance. If in this alternative model technological spillovers decrease by a share γ per unit of distance in exactly the same way as the hours worked of a commuter decrease by a share γ per unit of distance, then both models are observationally equivalent. Workers of the smaller city have access to the same discounted productivity of their larger neighbor: either by benefitting from technological spillovers from the larger city or by commuting to the larger city.

In addition to inter-city commuting costs and spatial technological spillovers, a third force that may contribute to the attractiveness of the larger city is trade and market access. In Appendix B.2 we consider an alternative model that incorporates trade between both cities. Once again, we switch off the possibility of inter-city commuting and only allow for intra-city commuting. Because intra-city commuting is more costly in the large than in the small city, falling transport costs make it relatively more attractive to move to the large city: an urban growth shadow. At the same time, falling transport costs make it easier to trade with the large city, thus increasing the incentive to reside and produce in the nearby small city: an urban growth spillover. In Appendix B.2 we use simulations to show that such an alternative model can generate similar dynamics to the ones we observe in the U.S. data between 1840 and 2017.

5 Concluding Remarks

In this paper we have analyzed whether a location's growth benefits or suffers from being geographically close to a large urban center. To do so, we have focused on U.S. counties and metro areas over the time period 1840-2017. We have found evidence of urban shadows between 1840 and 1920 and of urban spillovers between 1920 and 2017. Proximity to large urban clusters was negatively correlated with a location's growth in the early time period, and positively correlated in the later

time period, albeit with some weakening of this positive correlation in the last decades.

The conceptual framework we have developed suggests that as the cost of commuting drops, individuals first have an incentive to move from smaller closeby cities to larger urban centers. Later, if commuting costs continue to fall, individuals prefer to commute, rather than to move, from the smaller to the larger cities. This implies that falling transport costs first hurt, and then help, the growth of smaller locations in the vicinity of large urban centers. After documenting the long-run evolution of commuting costs, we have shown that our framework is consistent with the empirical evidence, both in the time series and the cross-section.

As such, a single variable — commuting costs — is able to capture the growth patterns of small cities in the hinterland of large urban clusters over the time period stretching from 1840 to 2017. Other factors are of course likely to have contributed to these spatial growth patterns. In this context, we have discussed the role of the spatial diffusion of technology, as well as the possible importance of market access and trade.

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, 2127-2189.
- [2] Allen, T., and C. Arkolakis (2014). “Trade and the Topography of the Spatial Economy,” *Quarterly Journal of Economics*, 129, 1085-1140.
- [3] Baum-Snow, N. (2007). “Did Highways Cause Suburbanization?,” *Quarterly Journal of Economics*, 122, 775-805.
- [4] 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.
- [5] Bernard, A. (2014). *Lifted: A Cultural History of the Elevator*, New York: NYU Press.
- [6] Bernhofen, D. M., El-Sahli, Z., and R. Kneller (2016). “Estimating the Effects of the Container Revolution on World Trade,” *Journal of International Economics*, 98, 36-50.
- [7] Black, D., and V. Henderson (2003). “Urban Evolution in the USA,” *Journal of Economic Geography*, 3, 343-372.
- [8] Bosker, M., and E. Buringh (2017). “City Seeds: Geography and the Origins of European Cities,” *Journal of Urban Economics*, 98, 139-157.

- [9] Christaller, W. (1933). *Central Places in Southern Germany*. Jena, Germany: Fischer (English translation by C. W. Baskin, London: Prentice Hall, 1966).
- [10] Conley, T. (1999). “GMM Estimation with Cross Sectional Dependence,” *Journal of Econometrics*, 92, 1-45.
- [11] Cronon, W. (1991). *Nature’s Metropolis: Chicago and the Great West*, New York: W.W. Norton.
- [12] Davis, D. R., and D. E. Weinstein (2002). “Bones, Bombs, and Break Points: The Geography of Economic Activity,” *American Economic Review*, 92, 1269-1289.
- [13] Desmet, K., D. Nagy, and E. Rossi-Hansberg (2018). “The Geography of Development,” *Journal of Political Economy*, 126, 903-983.
- [14] Desmet, K., and J. Rappaport (2017). “The Settlement of the United States, 1800-2000: The Long Transition towards Gibrat’s Law,” *Journal of Urban Economics*, 98, 50-68.
- [15] Desmet, K., and E. Rossi-Hansberg (2014). “Spatial Development,” *American Economic Review*, 104, 1211-1243.
- [16] Dobkins, L. H., and Y. Ioannides (2001). “Spatial Interactions among U.S. Cities,” *Regional Science and Urban Economics*, 31, 701-731.
- [17] Donaldson, D., and R. Hornbeck (2016). “Railroads and American Economic Growth: A Market Access Approach,” *Quarterly Journal of Economics*, 131, 799-858.
- [18] Eckert, F., and M. Peters (2018). “Spatial Structural Change,” unpublished manuscript, Yale University.
- [19] Edlund, L., C. Machado, and M. Sviatschi (2016). “Bright Minds, Big Rent: Gentrification and the Rising Returns to Skill,” NBER Working Paper # 21729.
- [20] Fogel, R. W. (1964). *Railroads and American Economic Growth: Essays in Econometric History*, Baltimore, MD: Johns Hopkins University Press.
- [21] Fujita, M., P. Krugman, and A. J. Venables (1999). *The Spatial Economy*, Cambridge, MA: MIT Press.
- [22] Gardner, T. (1999). “Metropolitan Classification for Census Years before World War II,” *Historical Methods*, 32, 139-150.
- [23] Glaeser, E. L. (2011), *Triumph of the City*, London: MacMillan.

- [24] 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.
- [25] Glaeser, E. L., and M. Kohlhase (2004). "Cities, Regions and the Decline of Transport Costs," *Papers in Regional Science*, 83, 197-228.
- [26] Hanson, G. (2005). "Market Potential, Increasing Returns and Geographic Concentration," *Journal of International Economics*, 67, 1-24.
- [27] Horan, P. M., and P. G. Hargis (1995). "County Longitudinal Template, 1840-1990." [computer file]. ICPSR Study 6576. Inter-university Consortium for Political and Social Research [distributor]. Corrected and amended by Patricia E. Beeson and David N. DeJong, Department of Economics, University of Pittsburgh, 2001. Corrected and amended by Jordan Rappaport, Federal Reserve Bank of Kansas City, 2010.
- [28] Jackson, K. T. (1985). *Crapgrass Frontier. The Suburbanization of the United States*. Oxford University Press.
- [29] Kopecky, K., and M. H. Suen (2004). "Economie d'Avant Garde: Suburbanization and the Automobile." *Research Report No. 6*.
- [30] Krugman, P. (1993). "On the Number and Location of Cities," *European Economic Review*, 37, 293-298.
- [31] LeRoy, S. F., and J. Sonstelie (1983), "Paradise Lost and Regained: Transportation Innovation, Income, and Residential Location," *Journal of Urban Economics*, 13, 67-89.
- [32] 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.
- [33] Lösch, A. (1940). *The Economics of Location*, Jena, Germany: Fischer (English translation, New Haven, CT: Yale University Press, 1954).
- [34] Michaels, G., F. Rauch, and S. J. Redding (2012). "Urbanization and Structural Transformation," *Quarterly Journal of Economics*, 127, 535-586.
- [35] Mieszkowski, P., and E. S. Mills (1993). "The Causes of Metropolitan Suburbanization," *Journal of Economic Perspectives*, 7, 135-147.
- [36] 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, 445-466.

- [37] Rappaport, J. (2005). “The Shared Fortunes of Cities and Suburbs,” *Federal Reserve Bank of Kansas City Economic Review*, Third Quarter, 33-59.
- [38] Rappaport, J. (2007). “Moving to Nice Weather,” *Regional Science and Urban Economics*, 37, 375-398.
- [39] Rappaport, J., and J. D. Sachs (2003). “The United States as a Coastal Nation,” *Journal of Economic Growth*, 8, 5-46.
- [40] Rauch, F. (2014). “Cities as Spatial Clusters,” *Journal of Economic Geography*, 14, 759-773.
- [41] Redding, S., and D. Sturm (2008). “The Costs of Remoteness: Evidence from German Division and Reunification,” *American Economic Review*, 98, 1766-1797.
- [42] Rosenthal, S. S, and W. C. Strange (2003). “Geography, Industrial Organization, and Agglomeration,” *Review of Economics and Statistics* 85, 377-393.
- [43] Shaw, R. E. (1990). *Canals for a Nation. The Canal Era in the United States, 1790-1860*, The University Press of Kentucky.
- [44] Su, Y. (2018). “The Rising Value of Time and the Origin of Urban Gentrification,” unpublished manuscript.
- [45] Tabuchi, T., and J-F. Thisse (2011). “A New Economic Geography Model of Central Places,” *Journal of Urban Economics*, 69, 240-252.
- [46] Thorndale, W., and W. Dollarhide (1987). *Map Guide to the Federal Censuses, 1790-1920*, Genealogical Publishing Company, Baltimore.
- [47] U.S. Bureau of the Census (1905). *Street and Electric Railways 1902*, Government Printing Office, Washington.
- [48] U.S. Bureau of the Census (1910). *Street and Electric Railways 1907*, Government Printing Office, Washington.
- [49] Warner, S. B. Jr. (1972). *Streetcar Suburbs. The Process of Growth in Boston, 1870-1900*, Harvard University Press and the MIT Press.
- [50] Young, J. (2015). “Infrastructure: Mass Transit in 19th- and 20th-Century Urban America.” *Oxford Research Encyclopedia of American History*. Online Publication Date: March 2015.

A Additional Tables

A.1 Constant Distances across Time Periods

In the empirical analysis in the main text of the paper we specify our regressions to include indicators for 50km intervals out to a maximum distance that leaves at least 50 percent of the observations in the excluded category. As a robustness check, in what follows we report versions of the main tables, keeping the maximum distance constant across time periods.

neighbor w/ pop \geq 95th percentile @ distance:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1840- 1860	1860- 1880	1880- 1900	1900- 1920	1920- 1940	1940- 1960	1960- 1980	1980- 2000	2000- 2017
1 to 50 km	-0.63 (0.35)	-0.77 (0.28)	-1.10 (0.21)	-0.19 (0.15)	0.15 (0.14)	0.99 (0.17)	1.29 (0.25)	0.59 (0.18)	0.41 (0.20)
50 to 100 km	-0.67 (0.24)	-0.41 (0.20)	-0.86 (0.18)	-0.42 (0.12)	-0.20 (0.10)	0.04 (0.12)	0.31 (0.20)	0.35 (0.15)	0.11 0.09
100 to 150 km	-0.30 (0.12)	-0.11 (0.17)	-0.63 (0.15)	-0.27 (0.13)	-0.24 (0.09)	-0.22 (0.10)	0.17 (0.18)	0.14 (0.13)	0.03 (0.07)
N (quints 1 to 4)	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
control vars	48	52	52	52	52	52	52	52	52
R²	0.857	0.787	0.726	0.534	0.409	0.368	0.392	0.428	0.318
Adj R²	0.846	0.778	0.718	0.522	0.395	0.353	0.377	0.413	0.298
R²-R² controls	0.002	0.002	0.008	0.003	0.005	0.030	0.034	0.011	0.004
pop\geq95th pctile	62ths -425ths	49ths -1.4mn	51ths -2.5mn	65ths -4.9mn	80ths -8.5mn	100ths -11.7mn	172ths -14.2mn	250ths -14.5mn	377ths -18.3mn
N, \geq95th pctile excluded share	44 0.51	86 0.54	119 0.53	133 0.43	148 0.41	150 0.38	143 0.39	132 0.41	119 0.50
pop of obs	133 -24ths	103 -23ths	100 -26ths	104 -30ths	137 -32ths	285 -36ths	208 -42ths	408 -49ths	356 -65ths
share of pop	0.38	0.39	0.42	0.39	0.34	0.29	0.20	0.16	0.14

Table A.1: Population Growth and the Presence of a Moderately Large Neighbor.

All regressions include a constant and control for initial population and up to 47 geographic covariates. Observations have population in the first four quintiles. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels. Dark and light red fonts respectively indicate a positive coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

As can be seen when comparing Table A.1 to Table 1, the results are qualitatively similar. One minor difference is worth discussing: in Table A.1 the positive coefficient on the presence of a large neighbor less than 50km away for the period 1920-1940 is smaller and no longer statistically significant, whereas expanding the maximum distance to 150km now yields a negative and statistically significant coefficient on the 100 to 150km distance interval for both the 1920-1940 and

<u>neighbor w/ pop ≥ 99th percentile @ distance:</u>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1840- 1860	1860- 1880	1880- 1900	1900- 1920	1920- 1940	1940- 1960	1960- 1980	1980- 2000	2000- 2017
1 to 50 km	0.77 (0.63)	-1.51 (0.46)	-0.78 (0.42)	-0.22 (0.28)	0.66 (0.28)	2.41 (0.26)	3.20 (0.61)	0.99 (0.58)	-0.26 (0.39)
50 to 100 km [†]	-0.20 (0.43)	-1.18 (0.40)	-0.86 (0.26)	-0.58 (0.19)	0.15 (0.17)	0.55 (0.15)	0.97 (0.25)	1.03 (0.19)	0.65 (0.14)
100 to 150 km	0.18 (0.43)	-1.11 (0.33)	-0.45 (0.22)	-0.41 (0.18)	0.01 (0.17)	0.14 (0.13)	0.30 (0.14)	0.51 (0.13)	0.35 (0.10)
150 to 200 km	0.39 (0.33)	-1.28 (0.31)	-0.14 (0.21)	-0.47 (0.16)	0.08 (0.14)	0.02 (0.13)	0.13 (0.12)	0.26 (0.10)	0.24 (0.07)
200 to 250 km	-0.11 (0.17)	-0.96 (0.27)	-0.02 (0.15)	-0.14 (0.13)	0.12 (0.12)	-0.12 (0.13)	0.03 (0.11)	0.10 (0.08)	0.18 (0.06)
N (quints 1 to 4)	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
control vars	48	52	52	52	52	52	52	52	52
R²	0.856	0.794	0.721	0.534	0.407	0.378	0.408	0.452	0.341
Adj R²	0.844	0.785	0.712	0.521	0.392	0.362	0.393	0.436	0.320
R²-R² controls	0.001	0.008	0.003	0.003	0.003	0.039	0.050	0.035	0.027
pop ≥ 99th pctile	171ths -425ths	139ths -1.4mn	139ths -2.5mn	197ths -4.9mn	321ths -8.5mn	442ths -11.7mn	810ths -14.2mn	1.3mn -14.5mn	2.0mn -18.3mn
N, ≥99th pctile	9	18	24	27	30	30	29	27	24
excluded share	0.68	0.47	0.64	0.62	0.64	0.54	0.54	0.56	0.64
pop of obs	133 -24ths	103 -23ths	100 -26ths	104 -30ths	137 -32ths	285 -36ths	208 -42ths	408 -49ths	356 -65ths
share of pop	0.38	0.39	0.42	0.39	0.34	0.29	0.20	0.16	0.14

Table A.2: **Population Growth and the Presence of a Very Large Neighbor.**

All regressions include a constant and control for initial population and up to 47 geographic covariates. Observations have population in the first four quintiles. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels. Dark and light red fonts respectively indicate a positive coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

1940-1960 periods. These findings are related. With the economy transitioning from the negative regime to the positive regime during the period 1920-1940, we should not be surprised to see a negative coefficient on the presence of a large neighbor at more far-off distances. This negative coefficient on the 100 to 150km shifts down the coefficients at shorter distances, explaining the smaller coefficient on the presence of a large neighbor within 50km. We also generate versions of Tables 3 and 6 that keep the maximum distance constant across time periods. These are reported in Tables A.2 and A.3. The results are largely unchanged.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<u>neighbor w/ population</u>	1840- 1860	1860- 1880	1880- 1900	1900- 1920	1920- 1940	1940- 1960	1960- 1980	1980- 2000	2000- 2017
≥ 80th pctile@									
1 to 50 km	-0.40 (0.21)	-0.66 (0.16)	-0.21 (0.12)	-0.26 (0.09)	0.12 (0.08)	-0.05 (0.09)	0.06 (0.07)	0.14 (0.06)	0.05 (0.05)
≥ 90th pctile@									
1 to 50 km	-0.35 (0.34)	0.21 (0.24)	-0.37 (0.19)	-0.12 (0.19)	-0.11 (0.16)	0.39 (0.15)	0.11 (0.10)	0.28 (0.11)	0.00 (0.08)
50 to 100 km	-0.29 (0.24)	-0.15 (0.19)	-0.29 (0.16)	-0.35 (0.13)	-0.09 (0.09)	0.13 (0.12)	-0.07 (0.08)	0.10 (0.07)	0.00 (0.06)
≥ 95th pctile@									
1 to 50 km	-0.19 (0.44)	-0.37 (0.31)	-0.74 (0.23)	0.03 (0.23)	0.09 (0.17)	0.34 (0.20)	0.80 (0.20)	0.18 (0.17)	0.38 (0.21)
50 to 100 km	-0.47 (0.30)	-0.21 (0.25)	-0.54 (0.18)	-0.10 (0.15)	-0.16 (0.12)	-0.17 (0.15)	0.14 (0.19)	0.13 (0.15)	0.03 (0.09)
100 to 150 km	-0.35 (0.13)	0.00 (0.20)	-0.60 (0.16)	-0.21 (0.12)	-0.24 (0.10)	-0.27 (0.10)	0.10 (0.18)	0.04 (0.12)	-0.01 (0.06)
≥ 99th pctile@									
1 to 50 km	0.91 (0.78)	-1.00 (0.45)	-0.10 (0.47)	-0.23 (0.29)	0.49 (0.26)	1.88 (0.28)	2.42 (0.58)	0.67 (0.50)	-0.61 (0.39)
50 to 100 km [†]	0.24 (0.43)	-0.86 (0.38)	-0.54 (0.26)	-0.44 (0.19)	0.23 (0.16)	0.59 (0.15)	0.98 (0.24)	0.96 (0.17)	0.63 (0.15)
100 to 150 km	0.37 (0.43)	-1.03 (0.34)	-0.16 (0.23)	-0.29 (0.17)	0.08 (0.18)	0.22 (0.13)	0.26 (0.13)	0.49 (0.11)	0.35 (0.09)
150 to 200 km	0.49 (0.31)	-1.22 (0.31)	-0.05 (0.20)	-0.42 (0.15)	0.11 (0.14)	0.03 (0.13)	0.12 (0.11)	0.25 (0.09)	0.23 (0.07)
200 to 250 km	-0.02 (0.16)	-0.87 (0.26)	0.02 (0.15)	-0.15 (0.13)	0.13 (0.12)	-0.09 (0.13)	0.01 (0.10)	0.11 (0.07)	0.19 (0.06)
N (quints 1 to 4)	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
control vars	48	52	52	52	52	52	52	52	52
R²	0.860	0.797	0.729	0.540	0.411	0.392	0.427	0.464	0.344
Adj R²	0.847	0.787	0.719	0.526	0.395	0.376	0.411	0.447	0.322
R²-R² controls	0.005	0.012	0.011	0.009	0.008	0.054	0.070	0.047	0.030
pop, pctile 80	24ths	23ths	26ths	30ths	32ths	36ths	42ths	49ths	65ths
pop, pctile 90	41ths	35ths	37ths	44ths	49ths	60ths	79ths	103ths	153ths
pop, pctile 95	62ths	49ths	51ths	65ths	80ths	100ths	172ths	150ths	377ths
pop, pctile 99	171ths	139ths	139ths	197ths	321ths	442ths	810ths	1.3mn	2.0mn
pop of obs	133 -24ths	103 -23ths	100 -26ths	104 -30ths	137 -32ths	285 -36ths	208 -42ths	408 -49ths	356 -65ths

Table A.3: Population Growth and the Size of Large Neighbors.

All regressions include a constant and control for initial population and up to 47 geographic covariates. Observations have population in the first four quintiles. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999).

A.2 Further Exploration of East Coast

In this subsection we further explore the local growth dynamics in the East. Focusing on the East has the advantage that the size of counties has not changed much over time. Tables A.4 and A.5 report how predictive having a large neighbor (95th percentile and above) is for local growth. The first table sets the maximum distance as in the paper, requiring that at least 50 percent of the observations are in the excluded category, whereas the second table keeps the maximum distance constant across time periods. When comparing with Table 1, the results for the East are similar to the ones for the U.S. as a whole. However, there are two differences worth pointing out. First, the growth shadows in the early time period are weaker, and second, the growth spillovers in the later time period start two decades earlier.

neighbor w/ pop \geq 95th percentile @ distance:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1840- 1860	1860- 1880	1880- 1900	1900- 1920	1920- 1940	1940- 1960	1960- 1980	1980- 2000	2000- 2017
1 to 50 km	0.29 (0.21)	-0.28 (0.31)	-0.31 (0.10)	0.50 (0.20)	0.24 (0.11)	0.78 (0.23)	1.15 (0.18)	0.34 (0.14)	0.31 (0.25)
50 to 100 km	-0.29 (0.16)	-0.21 (0.26)							0.02 (0.16)
N (quints 1 to 4)	235	346	379	390	423	431	407	355	307
control vars	41	41	41	41	41	41	41	41	41
R²	0.563	0.527	0.445	0.363	0.405	0.422	0.381	0.506	0.372
Adj R²	0.465	0.459	0.375	0.286	0.339	0.360	0.310	0.439	0.269
R²-R² controls	0.014	0.004	0.004	0.017	0.005	0.035	0.093	0.005	0.002

Table A.4: **Population Growth and the Size of Large Neighbors: East.**

All regressions include a constant and control for initial population and up to 37 geographic covariates. Observations have population in the first four quintiles of the national distribution. East region includes the states occupying the land area of the original 13 U.S. colonies excluding Kentucky and Tennessee. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark blue and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

What might explain these differences? The statistically weaker urban shadows could partly be due to the smaller number of observations. They might also be related to East Coast cities being closer to their long-run steady state. In our framework, declining commuting costs make large cities more attractive, prompting them to absorb population from their hinterlands. The further those cities are from their steady-state population, the greater that absorptive capacity might be. While this paper does not focus on transitions, locations in the West were further away from their steady than locations in the East, a point emphasized by Desmet and Rappaport (2017). The earlier emergence of urban spillovers may be due to East Coast counties being on average smaller than

neighbor w/ pop \geq 95th percentile @ distance:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1840- 1860	1860- 1880	1880- 1900	1900- 1920	1920- 1940	1940- 1960	1960- 1980	1980- 2000	2000- 2017
1 to 50 km	0.29 (0.21)	-0.28 (0.31)	-0.18 (0.23)	0.67 (0.28)	0.21 (0.12)	0.84 (0.24)	1.14 (0.18)	0.39 (0.15)	0.31 (0.25)
50 to 100 km	-0.29 (0.16)	-0.21 (0.26)	0.16 (0.25)	0.22 (0.14)	-0.03 (0.13)	0.08 (0.09)	-0.02 (0.21)	0.06 (0.10)	0.02 (0.16)
N (quints 1 to 4)	235	346	379	390	423	431	407	355	307
control vars	41	41	41	41	41	41	41	41	41
R²	0.563	0.527	0.446	0.366	0.405	0.423	0.381	0.506	0.372
Adj R²	0.465	0.459	0.375	0.288	0.337	0.358	0.308	0.438	0.269
R²-R² controls	0.014	0.004	0.006	0.020	0.005	0.035	0.093	0.005	0.002

Table A.5: **Population Growth and the Size of Large Neighbors: East.**

All regressions include a constant and control for initial population and 37 geographic covariates. Observations have population in the first four quintiles of the national distribution. East region includes the states occupying the land area of the original 13 U.S. colonies excluding Kentucky and Tennessee. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). Dark blue and light blue fonts respectively indicate a negative coefficient that statistically differs from zero at the 0.05 and 0.10 levels.

counties in the rest of the country. Our empirical results have already shown that the switch from the negative regime (lower growth in closeby places) to the positive regime (higher growth in closeby places) happens earlier for places “very close” to a large location than for places slightly further away. Because counties in the East Coast tend to be smaller, there are more locations that are “very close”, so that this shift becomes visible earlier.

B Theoretical Appendix

B.1 Proofs of Results

Proof of Result 1. Initially $A_\ell(1 - \gamma\bar{d}_\ell^0) = A_k(1 - \gamma\bar{d}_k^0)$ and $A_k > A_\ell$, so that $\bar{d}_k^0 > \bar{d}_\ell^0$. In this case, $A_\ell(1 - \gamma\bar{d}_\ell^0) \geq A_k(1 - \gamma\bar{d}_k^0) - \mu d_{\ell k}$ and $A_\ell(1 - \gamma\bar{d}_\ell^0) \geq A_k(1 - \gamma d_{\ell k})$ because $d_{\ell k} \geq \bar{d}_\ell^0$ by construction. Because $A_k\bar{d}_k^0 > A_\ell\bar{d}_\ell^0$, $-\partial A_\ell(1 - \gamma\bar{d}_\ell^0)/\partial\gamma < -\partial A_k(1 - \gamma\bar{d}_k^0)/\partial\gamma$, 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 and $\mu < (A_k - A_\ell)/d_{\ell k}$, at some point $A_k(1 - \gamma\bar{d}_k^0) - \mu d_{\ell k} = A_\ell(1 - \gamma\bar{d}_\ell^0)$. This occurs when γ reaches the threshold $\gamma_m = (A_k - A_\ell - \mu d_{\ell k})/(A_k\bar{d}_k^0 - A_\ell\bar{d}_\ell^0)$. If $A_k(1 - \gamma_m\bar{d}_k^0) < A_k(1 - \gamma_m\bar{d}_k^0) - \mu d_{\ell k}$, which requires $\mu < (A_k(d_{\ell k} - \bar{d}_k^0)(A_k - A_\ell))/(d_{\ell k}(d_{\ell k} + A_k\bar{d}_k^0 - A_\ell\bar{d}_\ell^0))$, then as soon as γ falls below γ_m , some of the 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 drop, m will increase. At some point, the drop in γ reaches $A_k(1 - \gamma(\bar{d}_k^0 + \min[m, \bar{d}_\ell^0])) - \mu d_{\ell k} = A_k(1 - \gamma d_{\ell k})$.

We refer to this threshold as γ_{mc} , where $\gamma_{mc} = \mu d_{\ell k} / (A_k(d_{\ell k} - \bar{d}_k^0 - \min[m, \bar{d}_\ell^0]))$. 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 $\gamma_c = \mu d_{\ell k} / (A_k(d_{\ell k} - \bar{d}_k^0))$, below which all original residents of ℓ commute to k . \square

Proof of Result 2. From the proof of Result 1, we can write $\gamma_m = (A_k - A_\ell - \mu d_{\ell k}) / (A_k \bar{d}_k^0 - A_\ell \bar{d}_\ell^0)$. It is clear that $d\gamma_m / dd_{\ell k} < 0$. From the same proof of Result 1, we can write $\gamma_{mc} = \mu d_{\ell k} / (A_k(d_{\ell k} - \bar{d}_k^0 - \min[m, \bar{d}_\ell^0]))$, where m can be written as $(A_k - A_\ell - A_k \gamma_{mc} \bar{d}_k^0 + A_\ell \gamma_{mc} \bar{d}_\ell^0 - \mu d_{\ell k}) / (\gamma_{mc}(A_\ell + A_k))$. Together, this implies that $\gamma_{mc} = \max[\mu d_{\ell k} / (A_k(d_{\ell k} - \bar{d}_k^0 - \bar{d}_\ell^0)), (A_k(A_k - A_\ell) + A_\ell d_{\ell k} \mu) / (A_k(A_k + A_\ell)d_{\ell k} - A_k A_\ell(\bar{d}_k^0 + \bar{d}_\ell^0))]$. Here as well, it is immediate that $d\gamma_{mc} / dd_{\ell k} < 0$. Threshold γ_c is reached when m in the above expression is equal to zero, so $\gamma_c = \mu d_{\ell k} / (A_k(d_{\ell k} - \bar{d}_k^0))$. It is immediate that $d\gamma_c / dd_{\ell k} < 0$. \square

Proof of Result 3. From the proof of Result 1, we can write $\gamma_m = \frac{A_k - A_\ell}{A_k \bar{d}_k^0 - A_\ell \bar{d}_\ell^0} - \frac{\mu d_{\ell k}}{A_k \bar{d}_k^0 - A_\ell \bar{d}_\ell^0}$. Our aim is to show that γ_m is increasing in $\bar{d}_k^0 / \bar{d}_\ell^0$. To do so, we consider the two terms in the γ_m expression separately. Because we start off in an equilibrium where $A_\ell(1 - \gamma^0 \bar{d}_\ell^0) = A_k(1 - \gamma^0 \bar{d}_k^0)$, where γ^0 is the initial value of γ , it follows that $\frac{A_k - A_\ell}{A_k \bar{d}_k^0 - A_\ell \bar{d}_\ell^0} = \gamma^0$. Hence, the first term of the γ_m expression above does not depend on the relative size $\bar{d}_k^0 / \bar{d}_\ell^0$. This leaves us with the second term, $-\frac{\mu d_{\ell k}}{A_k \bar{d}_k^0 - A_\ell \bar{d}_\ell^0}$. Because $A_\ell(1 - \gamma^0 \bar{d}_\ell^0) = A_k(1 - \gamma^0 \bar{d}_k^0)$, it follows that $\frac{A_k}{A_\ell} = \frac{1 - \gamma^0 \bar{d}_\ell^0}{1 - \gamma^0 \bar{d}_k^0}$. If $\bar{d}_k^0 / \bar{d}_\ell^0$ increases, we know that \bar{d}_k^0 increases and \bar{d}_ℓ^0 decreases, since $\bar{d}_\ell^0 + \bar{d}_k^0$ is a constant. As a result, if $\bar{d}_k^0 / \bar{d}_\ell^0$ increases, it follows that A_k / A_ℓ increases. Recall that we are keeping population-weighted productivity the same, so $A_k \bar{d}_k^0 + A_\ell \bar{d}_\ell^0$ is a constant we denote by λ . Hence, $A_k \bar{d}_k^0 - A_\ell \bar{d}_\ell^0 = \lambda - 2A_\ell \bar{d}_\ell^0$. If the larger city becomes larger and its relative productivity increases and the overall productivity is unchanged, it must be that the productivity of the small city decreases. It hence follows that $A_k \bar{d}_k^0 - A_\ell \bar{d}_\ell^0$ increases. This implies that the second term, $-\frac{\mu d_{\ell k}}{A_k \bar{d}_k^0 - A_\ell \bar{d}_\ell^0}$ is increasing in A_k / A_ℓ , so that γ_m is increasing in A_k / A_ℓ . \square

B.2 An Alternative Model with Trade and Market Access

In this subsection we propose an alternative model without inter-city commuting but with inter-city trade, and show the existence of the same fundamental tradeoff between urban shadows and urban spillovers.

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 . Whereas the total

population, L , is exogenous, the populations of the two cities, L_ℓ and L_k , are endogenous. 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. Land is owned by absentee landlords.

Technology and preferences. Each city produces a different good, firms are competitive, 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 to the production point of the city where she resides. In contrast to the model in the main paper, there is *no* inter-city commuting. The time cost of intra-city 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, and produces $A_\ell(1 - \gamma d_\ell)$ units of her city's good. Her wage income, $w_\ell(d_\ell)$, is therefore $p_\ell A_\ell(1 - \gamma d_\ell)$, where p_ℓ is the free-on-board (f.o.b.) price of the good produced in city ℓ . Her income net of land rents paid, $y_\ell(d_\ell)$, is $w_\ell(d_\ell) - r_\ell(d_\ell)$. Land rents are paid in terms of the local good to the absentee landlords, and then disappear from the economy. When a good of city ℓ is shipped to city k , a share γ' is lost per unit of distance, so $1 - \gamma' d_{\ell k}$ units arrive. Hence, the price of good ℓ in city k is $p_\ell / (1 - \gamma' d_{\ell k})$.

People can freely choose where to reside in their city. This implies that y_ℓ equalizes across all locations within a city. At the edge of city ℓ , land rents are zero, so $r_\ell(\bar{d}_\ell) = 0$, where \bar{d}_ℓ refers to the distance between the city center and the city edge. Hence, for all residents of city ℓ , income net of land rents is

$$y_\ell = p_\ell A_\ell (1 - \gamma \bar{d}_\ell). \quad (\text{A.1})$$

To move to another city, an individual has to pay a utility cost $\mu d_{\ell k}$.³⁰ 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.

Agents have CES preferences over the two different goods. The utility of an individual originally from city k and residing in city ℓ can then be defined as

$$u_\ell^k = \left(c_{\ell\ell}^{\frac{\sigma-1}{\sigma}} + c_{\ell k}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} - I_\ell^k \mu d_{\ell k} \quad (\text{A.2})$$

with

$$I_\ell^k = \begin{cases} 1 & \text{if } k \neq \ell \\ 0 & \text{otherwise} \end{cases}$$

³⁰This introduces a utility difference between the original residents of a city and the immigrant residents from the other city. However, it does not lead to a difference in their income net of land rents, so that (A.1) applies to both the original residents and the immigrant residents.

where $c_{\ell\ell}$ and $c_{\ell k}$ denote the consumption of good ℓ and k by a resident of city ℓ , $\sigma > 1$ is the elasticity of substitution between both goods, and I_ℓ^k is an indicator value equal to zero if the individual is an original resident of ℓ and equal to one if she is an immigrant from the other city.

Aggregate production. In city ℓ , total production is $\int_0^{\bar{d}_\ell} 2A_\ell(1 - \gamma d_\ell) dd_\ell = 2A_\ell \bar{d}_\ell(1 - \frac{1}{2}\gamma \bar{d}_\ell)$, where $\bar{d}_\ell = \frac{L_\ell}{2}$. A part of total production is paid out to the absentee land owners, and disappears from the economy. Net of the payments to land owners, each individual in ℓ produces $A_\ell(1 - \gamma \bar{d}_\ell)$. As a result, total production net of the payouts to land owners is

$$Q_\ell = 2A_\ell \bar{d}_\ell(1 - \gamma \bar{d}_\ell) = A_\ell L_\ell(1 - \gamma \bar{d}_\ell). \quad (\text{A.3})$$

Aggregate consumption. When solving the utility maximization problem, we can separate the consumption decision and the residential decision. We start by describing the consumption decision. An agent who resides in ℓ maximizes (A.2) subject to

$$p_\ell A_\ell(1 - \gamma \bar{d}_\ell) = p_\ell c_{\ell\ell} + \frac{p_k}{1 - \gamma' d_{\ell k}} c_{\ell k}. \quad (\text{A.4})$$

The first order conditions yield the following demand for each one of the two goods:

$$\begin{aligned} c_{\ell\ell} &= \frac{y_\ell(p_\ell)^{-\sigma}}{p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma' d_{\ell k}}\right)^{1-\sigma}} \\ c_{\ell k} &= \frac{y_\ell \left(\frac{p_k}{1-\gamma' d_{\ell k}}\right)^{-\sigma}}{p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma' d_{\ell k}}\right)^{1-\sigma}}. \end{aligned} \quad (\text{A.5})$$

Aggregate demand for goods produced in location ℓ is:

$$C_\ell = \frac{y_\ell L_\ell (p_\ell)^{-\sigma}}{p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma' d_{\ell k}}\right)^{1-\sigma}} + \frac{y_k L_k \left(\frac{p_\ell}{1-\gamma' d_{\ell k}}\right)^{-\sigma}}{\left(\frac{p_\ell}{1-\gamma' d_{\ell k}}\right)^{1-\sigma} + p_k^{1-\sigma}}. \quad (\text{A.6})$$

Residential choice. An individual who originally resides in city ℓ has a choice to stay in city ℓ or to move to city k . His indirect utility if he stays in city ℓ is:

$$u_\ell^\ell = \frac{y_\ell}{\left(p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma' d_{\ell k}}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} = \frac{p_\ell A_\ell(1 - \gamma \bar{d}_\ell)}{\left(p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma' d_{\ell k}}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} \quad (\text{A.7})$$

whereas his indirect utility if he moves to k is

$$u_k^\ell = \frac{y_k}{\left(\left(\frac{p_\ell}{1-\gamma' d_{\ell k}}\right)^{1-\sigma} + p_k^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} - \mu d_{\ell k} = \frac{p_k A_k(1 - \gamma \bar{d}_k)}{\left(\left(\frac{p_\ell}{1-\gamma' d_{\ell k}}\right)^{1-\sigma} + p_k^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} - \mu d_{\ell k}. \quad (\text{A.8})$$

Note that $\bar{d}_\ell = L/2 - \bar{d}_k$. Denote by \tilde{d}_ℓ the value of \bar{d}_ℓ that equalizes u_ℓ^ℓ and u_ℓ^k . That is,

$$\frac{p_\ell A_\ell (1 - \gamma \tilde{d}_\ell)}{\left(p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma' d_{\ell k}}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} = \frac{p_k A_k (1 - \gamma(\frac{L}{2} - \tilde{d}_\ell))}{\left(\left(\frac{p_\ell}{1-\gamma' d_{\ell k}}\right)^{1-\sigma} + p_k^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} - \mu d_{\ell k} \quad (\text{A.9})$$

By analogy, denote by \tilde{d}_k the value of \bar{d}_k that equalizes u_k^k and u_k^ℓ . That is,

$$\frac{p_k A_k (1 - \gamma \tilde{d}_k)}{\left(\left(\frac{p_\ell}{1-\gamma' d_{\ell k}}\right)^{1-\sigma} + p_k^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} = \frac{p_\ell A_\ell (1 - \gamma(\frac{L}{2} - \tilde{d}_k))}{\left(p_\ell^{1-\sigma} + \frac{p_k}{1-\gamma' d_{\ell k}}\right)^{\frac{1}{1-\sigma}}} - \mu d_{\ell k}. \quad (\text{A.10})$$

The original distribution of population has $2\bar{d}_\ell^0$ individuals living in ℓ and $2\bar{d}_k^0$ individuals living in k , where $2\bar{d}_\ell^0 + 2\bar{d}_k^0 = L$. If $\tilde{d}_\ell < \bar{d}_\ell^0$, then $2(\bar{d}_\ell^0 - \tilde{d}_\ell)$ people move from ℓ to k . If $\tilde{d}_k < \bar{d}_k^0$, then $2(\bar{d}_k^0 - \tilde{d}_k)$ people move from k to ℓ . If neither $\tilde{d}_\ell < \bar{d}_\ell^0$ nor $\tilde{d}_k < \bar{d}_k^0$, then $2(\bar{d}_\ell^0 - \tilde{d}_\ell)$, no one moves and everyone lives in their original location of residence.

Equilibrium. For given parameter values $L, A_\ell, A_k, d_{\ell k}, \mu, \gamma, \gamma'$ and σ , and for a given initial distribution of individuals across cities, \bar{d}_ℓ^0 and \bar{d}_k^0 , an equilibrium is a collection of variables $\{p_\ell, p_k, L_\ell, L_k, \bar{d}_\ell, \bar{d}_k, \tilde{d}_\ell, \tilde{d}_k\}$ that satisfy conditions (A.1), (A.9), (A.10) and:

1. Goods market clearing:

$$\begin{aligned} L_\ell p_\ell A_\ell (1 - \gamma \bar{d}_\ell) &= \frac{y_\ell L_\ell (p_\ell)^{1-\sigma}}{p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma' d_{\ell k}}\right)^{1-\sigma}} + \frac{y_k L_k \left(\frac{p_\ell}{1-\gamma' d_{\ell k}}\right)^{1-\sigma}}{\left(\frac{p_\ell}{1-\gamma' d_{\ell k}}\right)^{1-\sigma} + p_k^{1-\sigma}} \\ L_k p_k A_k (1 - \gamma \bar{d}_k) &= \frac{y_\ell L_\ell \left(\frac{p_k}{1-\gamma' d_{\ell k}}\right)^{1-\sigma}}{p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma' d_{\ell k}}\right)^{1-\sigma}} + \frac{y_k L_k (p_k)^{1-\sigma}}{\left(\frac{p_\ell}{1-\gamma' d_{\ell k}}\right)^{1-\sigma} + p_k^{1-\sigma}} \end{aligned} \quad (\text{A.11})$$

2. Labor market clearing:

$$L = L_\ell + L_k \quad (\text{A.12})$$

3. Land market clearing:

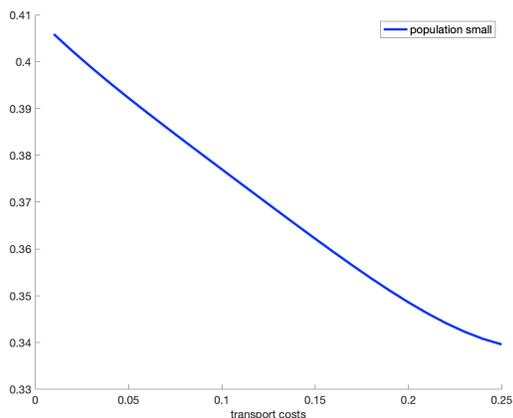
$$\begin{aligned} L_\ell &= 2\bar{d}_\ell \\ L_k &= 2\bar{d}_k \end{aligned} \quad (\text{A.13})$$

4. Labor mobility:

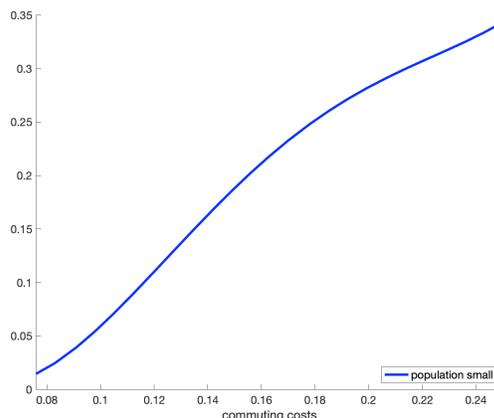
$$\bar{d}_\ell = \begin{cases} \tilde{d}_\ell & \text{if } \tilde{d}_\ell < \bar{d}_\ell^0 \\ \bar{d}_\ell + (\tilde{d}_k - \bar{d}_k) & \text{if } \tilde{d}_k < \bar{d}_k^0 \\ \bar{d}_\ell^0 & \text{otherwise} \end{cases} \quad (\text{A.14})$$

Quantitative properties. We use a simple numerical example to explore the quantitative properties of the model. We make the large city 50% more productive than the small city: $A_\ell = 1.0$ and $A_k = 1.5$. The elasticity of substitution between both goods, $\sigma = 3$. The total population $L = 6$, and inter-city distance $d_{\ell k} = 3$. The moving cost parameter is set to $\mu = 0.001$. Given the inter-city distance and the initial utility in both cities, this amounts to a little more than 0.3% in terms of utility. For the initial commuting cost and trade cost parameters, we choose $\gamma = 0.25$ and $\gamma' = 0.25$. Using these initial parameters, we distribute population between the two cities to equalize utility.

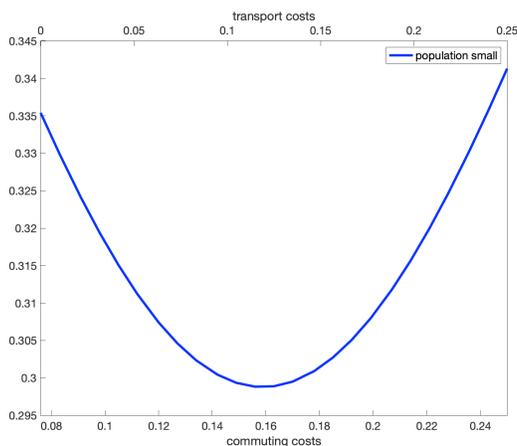
Figure B.1: Model with inter-city trade costs and intra-city commuting costs



(a) Drop in inter-city trade costs



(b) Drop in intra-city commuting costs



(c) Relative drop in inter-city trade costs

Before turning to the comparative statics exercises that illustrate the workings of the model, it is useful to discuss the differential effects of trade costs and commuting costs on the distribution of population across the two cities. A drop in trade costs improves the price index for both locations, but more so for the smaller, less productive city, because it has worse access to the cheaper good. A drop in commuting costs is also beneficial for both locations, but more so for the larger, more

productive city, because it suffers from greater congestion. As a result, the population share of the smaller city depend on which effect dominates. The following exercises show these different effects.

- (i) *Drop in trade costs, keeping commuting costs constant.* A decline in inter-city trade costs, keeping intra-city commuting costs unchanged, benefits the smaller, less productive city. Its price index goes down more than that of the larger city. Moving from right to left in B.1 Panel (a) shows the gain of the smaller city as inter-city trade costs fall.
- (ii) *Drop in commuting costs, keeping trade costs constant.* A decline in intra-city commuting costs, keeping inter-city trade costs unchanged, benefits the larger city. This is easy to understand: intra-city commuting is more costly in the larger than in the smaller city, so any decline in intra-city commuting will be more beneficial to the bigger city. Moving from right to left in B.1 Panel (b) shows the gain of the larger city as intra-city commuting costs fall.
- (iii) *Drop in both trade costs and commuting costs, with trade cost falling faster.* Figure B.1 Panel (c) depicts a situation where, when moving from left to right, intra-city commuting costs (represented on the bottom horizontal axis) decline at a slower pace than inter-city transport costs (represented on the top horizontal axis). To be more precise, intra-city commuting costs decline from 0.25 to 0.075, whereas inter-city trade costs decline from 0.25 to 0. We see that this decline first benefits the larger city (urban shadows dominate), and that eventually the small city recovers its population (urban spillovers dominate).

Which one of these exercises might be a better description of the long-run evolution of the U.S. economy? In their study of the long-run decline in transport costs, Glaeser and Kohlhase (2004) argued that the decline in trade costs has been more rapid than in commuting costs, saying that in today's world "it is essentially free to move goods, but expensive to move people". From that point of view, Figure B.1 Panel (c) is a reasonable description of reality. Moreover, it generates the same qualitative patterns as the model in the main text: urban growth shadows initially, and urban spillovers eventually. We therefore conclude that a model with intra-city commuting costs and inter-city trade costs is also able to explain what we see in the data.