Urban growth shadows☆

David Cuberes [a,1], Klaus Desmet [b,d,e,1,*], Jordan Rappaport [c,1]

[a] Department of Economics, Clark University, United States
[b] Department of Economics and Cox School of Business, Southern Methodist University, United States
[c] Federal Reserve Bank of Kansas City, United States
[d] NBER, USA
[e] CEPR, UK

ARTICLE INFO

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 improve market access 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 access. 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 model, we show that the secular evolution of inter-city and intra-city commuting costs can account for this. Alternatively, the long-run decline in inter-city shipping costs relative to intra-city commuting costs is also consistent with these observed patterns.

"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.[2] This is 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.[3] However, there is also

an opposing view: the presence of nearby clusters of economic activity improves market access, benefiting the growth of neighboring smaller places.[4]

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 access has taken over. One key force that is likely to have driven the changing relative strength of urban shadows and urban access 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 intra- and inter-city commuting 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 access over time and space. Another key force we consider is the evolution of shipping costs. We show that the long-run decline in shipping costs relative to commuting costs provides an alternative explanation for the shift in importance from urban shadows to urban access.

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 access 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 access has been weakening in the last decades. Second, there is evidence of the geographic reach of large urban centers expanding. Urban access benefits were very local between 1920 and 1940 and much more far-reaching later on. 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 and provided improved urban access in the later period.

We hypothesize that changes in either commuting or shipping costs can account for these patterns. Beginning in the mid-nineteenth century, a steady stream of 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. Drawing on cross-sectional data on streetcars for the time period 1900–1920, we identify one more stylized fact: improvements in local commuting infrastructure strengthened large cities' urban shadows. 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. Rather than having to permanently move to enjoy the productive benefits of larger cities, people could now continue to reside in the hinterland and commute to the urban centers instead. There is some indication that the continued drop in commuting costs has weakened in the last two decades.[5] Over the last century and a half, shipping costs have also experienced a secular decline. A relevant observation is that this drop in shipping costs has been more dramatic than the decrease in commuting costs (Glaeser and Kahn, 2004).

To understand the role of commuting costs and shipping costs in shaping the relative strength of urban shadows and urban access, we develop a simple model of two cities. The model captures a basic trade-off: 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, either through commuting or through trade. In this framework, the evolution of different spatial frictions — inter-city commuting costs, intra-city commuting costs, and inter-city shipping costs — determine whether urban shadows or urban access dominate. To highlight the role of these different spatial frictions, we consider two special cases. The first ignores inter-city trade, whereas the second ignores inter-city commuting.

In the first special case without inter-city trade, 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 inter-city and intra-city commuting gradually drops, we first see individuals from the smaller, less productive city move to the larger, more productive city. As commuting costs continue to fall, those individuals prefer to commute, rather than to move, to the larger city. Hence, a gradual drop in commuting costs first hurts growth in the smaller city, as it loses population to its larger neighbor, but a further drop eventually helps its growth, by attracting residents who can commute to the nearby large city. That is, the smaller city goes from experiencing a negative urban shadow, to benefiting from improved urban access.

The intuition for the non-monotonic relation between commuting costs and the growth of the smaller city is straightforward. The initial drop in commuting 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. A further drop in commuting 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 the smaller city to attract residents who can work in the larger city. The small city benefits from the proximity of the large city.

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 access dominating in the later time period, with some weakening of urban access 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 model implies that an increase in the relative size of a large urban center strengthens the force it exerts on its hinterland. As such, the long-run evolution of just one variable – commuting costs – is able to capture the different stylized facts, in particular the rise and decline of urban shadows.

An alternative interpretation is that urban access has improved, not because of more inter-city commuting but because of more inter-city trade. The second special case of the model aims to capture this force by introducing the possibility of trade. As before, the rise of urban shadows is driven by the drop in intra-city commuting costs, which is beneficial for both cities, but more so for the larger one. However, as shipping costs also decline, market access for the smaller city improves, giving its residents less reason to move to the larger city. If, as suggested by Glaeser and Kahn (2004), shipping costs have dropped faster than commuting costs, we show that this alternative model is also able to generate the rise and decline of urban shadows.

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).[6]

---

[5] 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.

[6] More recently, there has been a growing interest in incorporating ordered space into economic geography models. This is particularly true of quantitative spatial models that

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 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. (forthcoming) 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 access 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 shadows and urban access over the period 1840–2017. Section 3 documents the evolution of commuting costs over the same period. Section 4 proposes a conceptual framework that relates the evolution of commuting and shipping costs to urban shadows and urban access, showing that the theoretical predictions are consistent with the main patterns in the data. Section 5 concludes.

2. Urban shadows and urban access: 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 access were more prominent in different time periods. It also reveals 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.[7] 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 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 km 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 km of a river on which there was navigation in 1890 and whether it is in addition located within 80 km 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

---

aim to bring the theory to the data in meaningful ways (Desmet and Rossi-Hansberg, 2014; Allen and Arkolakis, 2014; Desmet et al., 2018).

[7] This 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.

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 $\tilde{d} \in \{\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_D\}$ denote strictly increasing specified distances, e.g., $\{50\text{km}, 100\text{km}, \ldots, 300\text{km}\}$. 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:

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

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, $\mathbf{I}_\ell^{\tilde{L}} = [\mathbb{I}_\ell^{\tilde{L}, \tilde{d}_1}, \mathbb{I}_\ell^{\tilde{L}, \tilde{d}_2}, \ldots \mathbb{I}_\ell^{\tilde{L}, \tilde{d}_D}]$, along with a fifth-order polynomial of a location's initial population, $\mathbf{L}_\ell = [\log(L_\ell), (\log(L_\ell))^2, \ldots, (\log(L_\ell))^5]$. 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, $\mathbf{x}_\ell$. We thus specify a data generating process with reduced form

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

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 $\{\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_D\}$, 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 km. 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.[8] 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 relative strength of urban shadows and urban access 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 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 km. 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 points. The corresponding negative coefficients statistically differ from zero at the 0.05 or 0.10 levels. Estimated negative

---

4   [8] Using the same maximum distance for all time periods yields qualitatively similar results (Online Appendix Section A.1).

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

	Average Annual Population Growth of Small and Medium Locations (Quintiles 1–4)								
Distance to Nearest Neighbor with Pop ≥ 95th Pctile	(1) 1840-1860	(2) 1860-1880	(3) 1880-1900	(4) 1900-1920	(5) 1920-1940	(6) 1940-1960	(7) 1960-1980	(8) 1980-2000	(9) 2000-2017
1 to 50 km	-0.63*	-0.77***	-1.10***	-0.04	0.28**	1.11***	1.19***	0.50***	0.41**
	(0.35)	(0.28)	(0.21)	(0.14)	(0.12)	(0.15)	(0.19)	(0.14)	(0.20)
50 to 100 km	-0.67***	-0.41**	-0.86***	-0.28***	-0.07	0.16*	0.21*	0.26***	0.11
	(0.24)	(0.20)	(0.18)	(0.09)	(0.07)	(0.08)	(0.11)	(0.09)	(0.09)
100 to 150 km	-0.30***	-0.11	-0.63***						0.03
	(0.12)	(0.17)	(0.15)						(0.07)
Additional Controls	48	52	52	52	52	52	52	52	52
N	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
Adjusted R^2	0.846	0.778	0.718	0.521	0.393	0.352	0.376	0.412	0.298
Incremental R^2	0.002	0.002	0.008	0.002	0.003	0.028	0.033	0.010	0.004

Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above the 95th percentile is within the enumerated distance bin. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < .10$, ** $p < .05$, *** $p < .01$

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.[9]

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. The magnitude of the faster predicted growth is relatively modest from 1920 to 1940, when suburbanization was just getting underway. Then, both from 1940 to 1960 and from 1960 to 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 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 access, then we can summarize our findings in Table 1 as follows:

STYLIZED FACT 1: URBAN SHADOWS AND URBAN ACCESS. *Between 1840 and 1920 urban shadows dominated the U.S. economic geography,*

with locations in the vicinity of large places growing relatively slower, whereas between 1920 and 2017 urban access 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.[10]

2.4. Recent weakening of urban access

In this subsection we analyze whether there has effectively been a weakening in urban access 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 km 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 km 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.[11]

To assess the effect of changes in delineations, 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 starting circa 2000. This suggests that the transition

---

[9] In 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).

[10] In Online Appendix A.2 we explore whether the westward expansion of the U.S. during the 19th and early 20th centuries might have affected the existence of urban shadows, and in Online Appendix A.3 we focus exclusively on the East Coast, where the size of counties has not changed much over time.

[11] In 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.

Table 2
Population Growth and the Presence of a Moderately Large Neighbor, 1960 Borders.

	Average Annual Population Growth of Small and Medium Locations (Quintiles 1–4)		
	(1)	(2)	(3)
Distance to Nearest Neighbor with Pop ≥ 95th Pctile	1960-1980	1980-2000	2000-2017
1 to 50 km	1.19***	1.43***	0.60***
	(0.19)	(0.30)	(0.12)
50 to 100 km	0.21*	0.59***	0.14**
	(0.11)	(0.23)	(0.06)
100 to 150 km		0.24	
		(0.20)	
150 to 200 km		0.13	
		(0.12)	
Additional Controls	52	52	52
N	2,283	2,282	2,283
Adjusted R^2	0.376	0.426	0.310
Incremental R^2	0.033	0.047	0.019

Metropolitan areas are delineated using the OMB standards following the 1960 decennial census. Non-metropolitan counties are delineated using their borders in 1960. Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above the 95th percentile is within the enumerated distance bin. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < .10$, ** $p < .05$, *** $p < .01$.

of metropolitan areas to a larger geographic footprint may be winding down.[12] We summarize these findings as follows:

STYLIZED FACT 2: RECENT WEAKENING OF URBAN ACCESS. *Urban access weakened circa 2000. In particular, during the period 2000–2017 benefits from urban access are less pronounced than during the periods from 1960 to 1980 and 1980–2000.*

2.5. Geographic reach

This subsection explores how the geographic reach of urban shadows and urban access 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 reach of urban shadows and urban access 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 250 km, 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 urban access, 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. For comparison, having a neighbor with population above the 95th percentile was associated with faster growth of only 1.19 percentage points.

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 REACH OF SHADOWS AND ACCESS. *Over the period 1920–2017 there is strong evidence of the geographic reach of urban access expanding, with the benefits from access 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 reach of urban shadows expanding between the late 19th century and early 20th century.*

2.6. Size of large neighbors

This subsection explores how the strength of urban shadows and urban access depends on the relative size of neighbors. In particular, we are interested in exploring whether growth's correlations with the presence of large neighbors increased with the size of neighbors.

Table 4 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. The corresponding coefficients measure the marginal boost to growth compared to having a largest neighbor at the threshold immediately below. 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 above the 95th percentile but not above the 99th percentile.

---

[12] Regressing 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.

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

Distance to Nearest Neighbor with Pop ≥ 99th Pctile	Average Annual Population Growth of Small and Medium Locations (Quintiles 1–4)								
	(1) 1840-1860	(2) 1860-1880	(3) 1880-1900	(4) 1900-1920	(5) 1920-1940	(6) 1940-1960	(7) 1960-1980	(8) 1980-2000	(9) 2000-2017
1 to 50 km	0.87	-1.01***	-0.79*	-0.28	0.73**	2.41***	3.20***	0.99*	-0.21
	(0.67)	(0.37)	(0.42)	(0.28)	(0.29)	(0.26)	(0.61)	(0.58)	(0.39)
50 to 100 km	-0.09	-0.67**	-0.87***	-0.64***	0.22	0.55***	0.97***	1.03***	0.69***
	(0.47)	(0.30)	(0.27)	(0.19)	(0.19)	(0.15)	(0.25)	(0.19)	(0.15)
100 to 150 km	0.28	-0.60***	-0.45*	-0.47***	0.08	0.14	0.30**	0.51***	0.39***
	(0.47)	(0.22)	(0.24)	(0.18)	(0.18)	(0.13)	(0.14)	(0.13)	(0.10)
150 to 200 km	0.49	-0.77***	-0.15	-0.52***	0.15	0.02	0.13	0.26***	0.28***
	(0.38)	(0.18)	(0.23)	(0.17)	(0.16)	(0.13)	(0.12)	(0.10)	(0.08)
200 to 250 km	0.00		-0.02	-0.19	0.18	-0.12	0.03	0.10	0.22***
	(0.24)		(0.17)	(0.16)	(0.15)	(0.13)	(0.11)	(0.08)	(0.07)
250 to 300 km	0.16		-0.02	-0.13	0.16				0.10
	(0.18)		(0.19)	(0.19)	(0.13)				(0.06)
Additional Controls	48	52	52	52	52	52	52	52	52
N	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
Adjusted R^2	0.844	0.781	0.712	0.521	0.392	0.362	0.393	0.436	0.322
Incremental R^2	0.001	0.004	0.003	0.004	0.003	0.039	0.050	0.035	0.028

Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above the 99th percentile is within the enumerated distance bin. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < .10$, ** $p < .05$, *** $p < .01$

During the negative regime, the increase in the magnitude of growth's relationship with the 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 points per year. If the largest such neighbor within 50km had population above the 90th percentile, predicted population growth is slower by an additional 0.37 percentage points per year. If the largest such neighbor had population above the 95th percentile, predicted population growth is slower by still an additional 0.74 percentage points 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 points 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 within 50km with a 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 points 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 size of the large neighbor affects the strength of urban shadows and urban access can be summarized as follows:

STYLIZED FACT 4: SIZE OF LARGE NEIGHBOR. *Urban shadows and urban access tend to strengthen in the size of the large neighbor. That is, the larger the neighbor, the stronger urban shadows and urban access.*

This result is robust to varying the relative size of neighbors: in the same way that urban shadows and urban access tend to be stronger the larger is the neighbor, they also tend to be stronger the smaller is the location itself (Online Appendix A.4).

3. Commuting costs

One important force that is bound to have shaped spatial growth dynamics in the hinterland of large population clusters are commuting costs. In this section we start by briefly documenting how local commuting costs in the U.S. have evolved since 1840. In doing so, we pay particular attention to changes in local transportation technology. We then explore cross-sectional evidence of the relation between local transportation infrastructure and urban shadows in the early twentieth century.

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 Kahn (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

Table 4
Population Growth and the Size of Large Neighbors.

Distance to Nearest Neighbor with Pop	Average Annual Population Growth of Small and Medium Locations (Quintiles 1–4)								
	(1) 1840-1860	(2) 1860-1880	(3) 1880-1900	(4) 1900-1920	(5) 1920-1940	(6) 1940-1960	(7) 1960-1980	(8) 1980-2000	(9) 2000-2017
≥ 80th Pctile									
1 to 50 km	-0.41*	-0.72***	-0.21*	-0.30***	0.10	-0.07	0.06	0.14**	0.06
	(0.21)	(0.19)	(0.12)	(0.09)	(0.07)	(0.08)	(0.07)	(0.06)	(0.05)
≥ 90th Pctile									
1 to 50 km	-0.37	0.25	-0.37**	0.13	-0.03	0.30***	0.15*	0.21**	0.00
	(0.34)	(0.27)	(0.19)	(0.14)	(0.11)	(0.11)	(0.08)	(0.09)	(0.08)
50 to 100 km	-0.31	-0.19	-0.29*						-0.01
	(0.25)	(0.22)	(0.16)						(0.06)
≥ 95th Pctile									
1 to 50 km	-0.17	-0.44	-0.74***	0.01	0.18	0.54***	0.72***	0.18	0.40*
	(0.43)	(0.30)	(0.23)	(0.20)	(0.15)	(0.17)	(0.15)	(0.13)	(0.21)
50 to 100 km	-0.46	-0.17	-0.54***	-0.22**	-0.09	0.06	0.04	0.17*	0.04
	(0.30)	(0.22)	(0.17)	(0.10)	(0.08)	(0.09)	(0.09)	(0.08)	(0.09)
100 to 150 km	-0.38***	0.05	-0.60***						0.00
	(0.13)	(0.20)	(0.15)						(0.06)
≥ 99th Pctile									
1 to 50 km	1.07	-0.48	-0.09	-0.31	0.55**	1.88***	2.43***	0.69	-0.56
	(0.82)	(0.36)	(0.46)	(0.29)	(0.26)	(0.28)	(0.59)	(0.52)	(0.38)
50 to 100 km	0.42	-0.40	-0.53**	-0.52**	0.28	0.59***	0.98***	0.97***	0.67***
	(0.45)	(0.30)	(0.26)	(0.20)	(0.18)	(0.16)	(0.24)	(0.17)	(0.15)
100 to 150 km	0.55	-0.59**	-0.14	-0.42**	0.09	0.16	0.28**	0.50***	0.39***
	(0.46)	(0.26)	(0.23)	(0.18)	(0.19)	(0.14)	(0.14)	(0.13)	(0.10)
150 to 200 km	0.65*	-0.75***	-0.04	-0.51***	0.17	0.06	0.11	0.24***	0.27***
	(0.35)	(0.19)	(0.21)	(0.17)	(0.16)	(0.13)	(0.11)	(0.09)	(0.07)
200 to 250 km	0.14		0.03	-0.21	0.19	-0.09	0.01	0.11	0.23***
	(0.23)		(0.16)	(0.15)	(0.15)	(0.13)	(0.11)	(0.08)	(0.07)
250 to 300 km	0.25		0.03	-0.13	0.16				0.12*
	(0.20)		(0.16)	(0.18)	(0.13)				(0.07)
Additional Controls	48	52	52	52	52	52	52	52	52
N	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
Adjusted R^2	0.847	0.784	0.719	0.523	0.393	0.374	0.411	0.447	0.323
Incremental R^2	0.005	0.008	0.011	0.006	0.006	0.052	0.069	0.046	0.032

Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above an enumerated threshold is within the enumerated distance bin. These thresholds are nested: a neighbor that is above the 99th percentile is also above the 90th and 95th percentiles. A coefficient on a given threshold thus estimates the marginal boost to predicted growth compared to having a neighbor with population only above the next highest of the enumerated thresholds. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < .10$, ** $p < .05$, *** $p < .01$.

affordable to the very rich (LeRoy and Sonstelie, 1983).[13] 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.[14] Commuter railroads appeared in the 1830s, although they were noisy and polluting, which led authorities to impose strict regulations, often limiting their use.[15]

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; Kopecky and 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.[16] 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 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

---

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

[14] LeRoy 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.

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

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

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 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).[17]

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 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.[18]

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, Duranton et al. (2020) compute the average speed of trips under 50 km from the National Household Travel Survey in 1995, 2001, 2009 and 2017, and find an almost monotone decrease in the speed of travel by car over the 1995–2017 period.[19]

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.[20]

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 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.[21]

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

---

[17] Suburbanization 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.

[18] The 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.

[19] It should be noted that comparing the NHTS data over time is a complicated exercise and so we should interpret this finding with caution. See Duranton et al. (2020) for more details.

[20] Of course, since this process of gentrification also displaces people, it is not clear whether this is associated with a decline or an increase in the center-city population.

[21] The 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.

Table 5

Population Growth and Large Neighbors, 1910–1920: Change in Streetcars.

	Average Annual Population Growth of Small and Medium Locations (Quintiles 1–4)			
	Neighbor with Pop ≥ 95th Pctile		Neighbor with Pop ≥ 90th Pctile	
	(1) 1910-1920 Change in Streetcar Presence	(2) 1910-1920 Change in Streetcar Log Miles	(3) 1910–1920 Change in Streetcar Presence	(4) 1910-1920 Change in Streetcar Log Miles
Distance to Nearest Neighbor with Pop ≥ 95th Pctile				
1 to 50 km	0.06	0.02	0.09	0.08
	(0.17)	(0.15)	(0.14)	(0.14)
Change in Streetcar 1–50 km	-1.02**	-0.28**	-0.18	-0.09
	(0.4)	(0.13)	(0.18)	(0.07)
50 to 100 km	-0.17	-0.12	-0.17	-0.15
	(0.12)	(0.11)	(0.11)	(0.11)
Change in Streetcar 50–100 km	-0.07	-0.25**	0.28	0.04
	(0.32)	(0.12)	(0.20)	(0.09)
Change in Own Streetcar	0.05	0.14**	0.02	0.13*
	(0.12)	(0.06)	(0.13)	(0.07)
Additional Controls	52	52	52	52
N	2,238	2,224	2,238	2,224
Adjusted R^2	0.154	0.156	0.155	0.154

Regressions are based on specification (2). Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile between 1910 and 1920 on categorical indicators if the nearest neighbor with a population at or above the 95th percentile (columns 1 and 2) or the 90th percentile (columns 3 and 4) is within the enumerated distance bin. The regressions include an interaction of these categorical indicators with either the change in streetcar presence (columns 1 and 3) or the change in streetcar miles (columns 2 and 4). All regressions include a constant, and control for either the change in streetcar presence or streetcar miles in the own location, as well as for initial population and additional geographic, weather, and topographical variables. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). * $p < .10$, ** $p < .05$, *** $p < .01$.

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 its streetcars expand, either along the extensive or the 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 main city where it operates and its length. Using a Python program, we match those cities to counties, and then manually correct any remaining errors.[22] In 1902, 521 out of 2,641 counties had streetcars, corresponding to a total mileage of almost 20,000 miles. By 1907, 693 out of 2,797 counties had 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.

Specification. To investigate how the introduction of streetcars in the neighboring large location might affect the strength of urban shadows, we take the baseline specification (1) and introduce an interaction term between the presence of a large neighbor and that large neighbor introducing streetcars.[23] We also control for the possible introduction of streetcars in the own county. This yields the following specification:

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

where $\mathbf{S}_\ell^{\bar{L}}$ are indicator variables that take a value of one if the corresponding large locations in $\mathbf{I}_\ell^{\bar{L}}$ have a streetcar, s_ℓ is an indicator variable that measures whether ℓ has a streetcar or not, and $\Delta \mathbf{S}_\ell^{\bar{L}}$ and Δs_ℓ represent the difference in $\mathbf{S}_\ell^{\bar{L}}$ and s_ℓ between two time periods. We will run (2) for 1910–1920, where the difference in streetcar presence refers to the period between 1902 and 1907.

In addition to analyzing the effect of the introduction of streetcars (extensive margin), we are also interested in exploring the effect of the expansion in streetcar mileage (intensive margin). We therefore consider an alternative, where $\mathbf{S}_\ell^{\bar{L}}$ measures the miles of streetcars of the corresponding large locations in $\mathbf{I}_\ell^{\bar{L}}$ and s_ℓ measures the miles of streetcars in ℓ, with $\Delta \mathbf{S}_\ell^{\bar{L}}$ and Δs_ℓ denoting their corresponding differences between 1902 and 1907. When measuring miles, we take a log transformation, and to avoid ignoring the extensive margin, we consider the log of one plus the mileage.

Results on Streetcars. Table 5 reports our findings from running specification (2). 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 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 5 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 street-

---

[22] More specifically, for each streetcar line, we compute the bilateral distances between the centroids of 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.

[23] In Online Appendix A.5 we also consider an alternative specification that looks at the presence (rather than the introduction) of streetcars.

cars in neighboring large locations.[24] This finding can be summarized as follows:

STYLIZED FACT 5: URBAN SHADOWS AND LOCAL COMMUTING INFRASTRUCTURE. *In the early 19th century, urban shadows were stronger when large locations disposed of better commuting infrastructure in the form of streetcars.*

4. Conceptual framework

In this section we develop a two-city spatial model with commuting costs, moving costs and trade costs that is able to account for the main stylized facts identified in the data. The basic tradeoff the model captures is easy to understand: 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, either through commuting or through trade.

In this framework four types of spatial frictions – inter-city moving costs, intra-city commuting costs, inter-city commuting costs and inter-city trade costs – are key in determining the relative growth of the smaller city. To sharpen the analysis, we consider two special cases of the basic setup, each focusing on three of the four spatial frictions. The first special case ignores trade, and shows how the documented long-run decline in intra-city and inter-city commuting costs leads to urban shadows dominating in the early stage, and urban access dominating later on. The second special case allows for inter-city trade and intra-city commuting, but ignores the possibility of inter-city commuting. It shows that the secular decline in inter-city trade costs relative to intra-city commuting costs, documented by Glaeser and Kahn (2004), likewise generates urban shadows dominating early on, and urban access dominating later on.

In what follows, we start by describing the framework's general setup, and then analyze the two special cases and their relation to urban shadows and urban access.

4.1. Setup

Endowments. The economy consists of a continuum of points on a line. The density of land at all points of the line is one. There are L individuals, each residing on one unit of land. Each resident has one unit of time, which she divides between work and commuting. On the line there are two exogenously given production points, denoted by ℓ and k. The set of individuals living closer to production point ℓ than to production point k comprises city ℓ.

Of the L individuals, initially $L^{\ell 0}$ reside in city ℓ and L^{k0} reside in city k. Individuals from one city can choose to move and reside in the other city. Moving implies a utility cost $\mu d_{\ell k}$ that is increasing in inter-city distance $d_{\ell k}$, measured as the distance between production point ℓ and production point k. Examples of the utility cost of being a migrant include the psychological and social costs of having to leave friends and family behind. Consistent with this utility 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 introduces a possible difference between an individual's city of origin and the city of residence. In what follows, our notation uses superscripts for origin and subscripts for residence. For example, L_k^ℓ refers to the number of individuals from city ℓ who reside in city k.

The land rent in city ℓ at distance d_ℓ from production point ℓ is denoted by $r_\ell(d_\ell)$. Inter-city distance, $d_{\ell k}$, is big enough so that there

is at least some empty land between the two cities.[25] Land is owned by absentee landlords.

Technology. Each city produces a different good, and labor is the only factor of production. When a good of city ℓ is shipped to city k, a share γ_t is lost per unit of distance, so $1 - \gamma_t d_{\ell k}$ units arrive. Hence, the price of good ℓ in city k is $p_\ell / (1 - \gamma_t d_{\ell k})$, where p_ℓ is the free-on-board (f.o.b.) price of the good produced in ℓ. 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.[26]

To produce, an individual needs to commute from his residence to one of the two production points. The time cost of intra-city commuting per unit of distance is γ_c and the time cost of inter-city commuting per unit of distance is γ_i. 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 $1 - \gamma_c d_\ell$, and produces $A_\ell(1 - \gamma_c d_\ell)$ units of her own city's goods. Her income net of land rents is then $p_\ell A_\ell(1 - \gamma_c d_\ell) - r_\ell(d_\ell)$. If she commutes to the other city k, we assume that she incurs an inter-city commuting distance $d_{\ell k}$, independently of where she resides in city ℓ.[27] In that case, she supplies $1 - \gamma_i d_{\ell k}$ units of labor, and produces $A_k(1 - \gamma_i d_{\ell k})$ units of city k's goods. Her income net of land rents is then $p_k A_k(1 - \gamma_i d_{\ell k}) - r_\ell(d_\ell)$.

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

$$u_k^\ell = \left(c_{k\ell}^{\frac{\sigma-1}{\sigma}} + c_{kk}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} - \mu d_{\ell k} \tag{3}$$

whereas the utility of an individual originally from city ℓ who resides in her own city is

$$u_\ell^\ell = \left(c_{\ell\ell}^{\frac{\sigma-1}{\sigma}} + c_{\ell k}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{4}$$

where $\sigma > 1$ is the elasticity of substitution between both goods, and $c_{k\ell}$ denotes the consumption by a resident of city k of good ℓ (i.e., the first subscript refers to the residence of the consumer and the second subscript to the origin of the good).

4.2. Intra-City commuting and inter-City commuting

In this subsection we consider a special case of our general setup where the two goods are perfect substitutes. Since this amounts to having just one good in the economy, the model can be simplified in several ways. First, trade is no longer relevant, so we limit our focus to intra-city commuting costs, inter-city commuting costs and inter-city moving costs. To further simplify, we assume that intra-city and inter-city commuting costs per unit of distance are the same, so $\gamma_c = \gamma_i = \gamma$. Second, with only one good, we can normalize the good's price to one. Third, the indirect utility of an individual originally from city ℓ and residing in city k is now simply her income net of land rents and moving costs.

Indirect Utility. Without loss of generality, we assume that no one from city k has an incentive to move or work in city ℓ. Depending on an

individual's city of origin, city of residence, city of work and distance of residence to city center, there are four possible expressions for utility:

$$u_\ell^\ell(\ell, d(\ell)) = A_\ell(1 - \gamma d_\ell) - r_\ell(d_\ell)$$

$$u_k^\ell(k, d(k)) = A_k(1 - \gamma d_k) - r_k(d_k) - \mu d_{\ell k}$$

$$u_\ell^\ell(k, d(\ell)) = A_k(1 - \gamma d_{\ell k}) - r_\ell(d_\ell)$$

$$u_k^k(k, d(k)) = A_k(1 - \gamma d_k) - r_k(d_k), \tag{5}$$

where the superscript on u refers to the individual's place of origin, the subscript on u refers to her place of residence, the first element in the brackets refers to the place of work, and the second element in the brackets refers to the distance of the residence to the city center. For example, $u_\ell^\ell(k, d(\ell))$ refers to the utility of an individual who is originally from ℓ, resides in ℓ at distance $d(\ell)$ from the city center and works in k, whereas $u_k^\ell(k, d(k))$ refers to the utility of an individual who is originally from ℓ, resides in k at distance $d(k)$ from the city center and works in k. Since no one from k has an incentive to move or work in ℓ, the increase or decrease in population of ℓ is solely driven by the decisions of the original residents of ℓ, who can either stay in ℓ, move to k, or commute to k. To simplify notation, we will sometimes refer to $u_\ell^\ell(\ell, d(\ell))$ as the staying utility U_S, to $u_k^\ell(k, d(k))$ as the moving utility U_M, and to $u_\ell^\ell(k, d(\ell))$ as the commuting utility U_C.

Residential Mobility within Cities. Individuals can freely locate within cities. Where land is unoccupied, land rents are normalized to zero. Hence, at the city edge $\bar{d}_\ell$ 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.

To determine equilibrium land rents at different locations, note that in city ℓ there are potentially two types of residents: those who work locally in ℓ, denoted by $L_\ell^\ell(\ell)$, and those who commute to k, denoted by $L_\ell^\ell(k)$. 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 $L_\ell^\ell(k)/2$ on both edges of the city where rents are zero. To be precise, for all $d_\ell \in [\bar{d}_\ell - L_\ell^\ell(k)/2, \bar{d}_\ell]$ 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 $d_\ell \in [0, \bar{d}_\ell - L_\ell^\ell(k)/2]$, we have $r_\ell(d_\ell) + A_\ell \gamma d_\ell = A_\ell \gamma (\bar{d}_\ell - L_\ell^\ell(k)/2)$, so that $r_\ell(d_\ell) = A_\ell \gamma (\bar{d}_\ell - L_\ell^\ell(k)/2 - d_\ell)$. Summarizing, equilibrium land rents in city ℓ are:

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

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

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

City Choice and Commuting Choice. A city loses population if its residents prefer to move to the other city, but it gains population if its residents move back and instead commute to the other 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 $U_S \equiv u_\ell^\ell(\ell, d(\ell))$; she can move to city k at a distance $d(k)$ from her work at production point k, earning a utility $U_M \equiv u_k^\ell(k, d(k))$; or she can commute a distance $d_{\ell k}$ from city ℓ to production point k to work, earning a utility $U_C \equiv u_\ell^\ell(k, d(\ell))$. 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 $\bar{d}_\ell^0$ 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 Fig. 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 Fig. 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.[28] If, as illustrated in the top-right panel of Fig. 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.[29]

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 Fig. 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 at 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 where the utility from staying equalizes the utility from commuting.[30] In this case, illustrated in the bottom-right panel of Fig. 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 $\bar{d}_\ell^0$ 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 $A_k(1 - \gamma(\bar{d}_k^0 + m)) - \mu d_{\ell k} = A_\ell(1 - \gamma(\bar{d}_\ell^0 - m))$.

---

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

[29] It 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.

[30] Of course, if everyone commutes before that equality is reached, then we would have the entire city ℓ commuting to k.
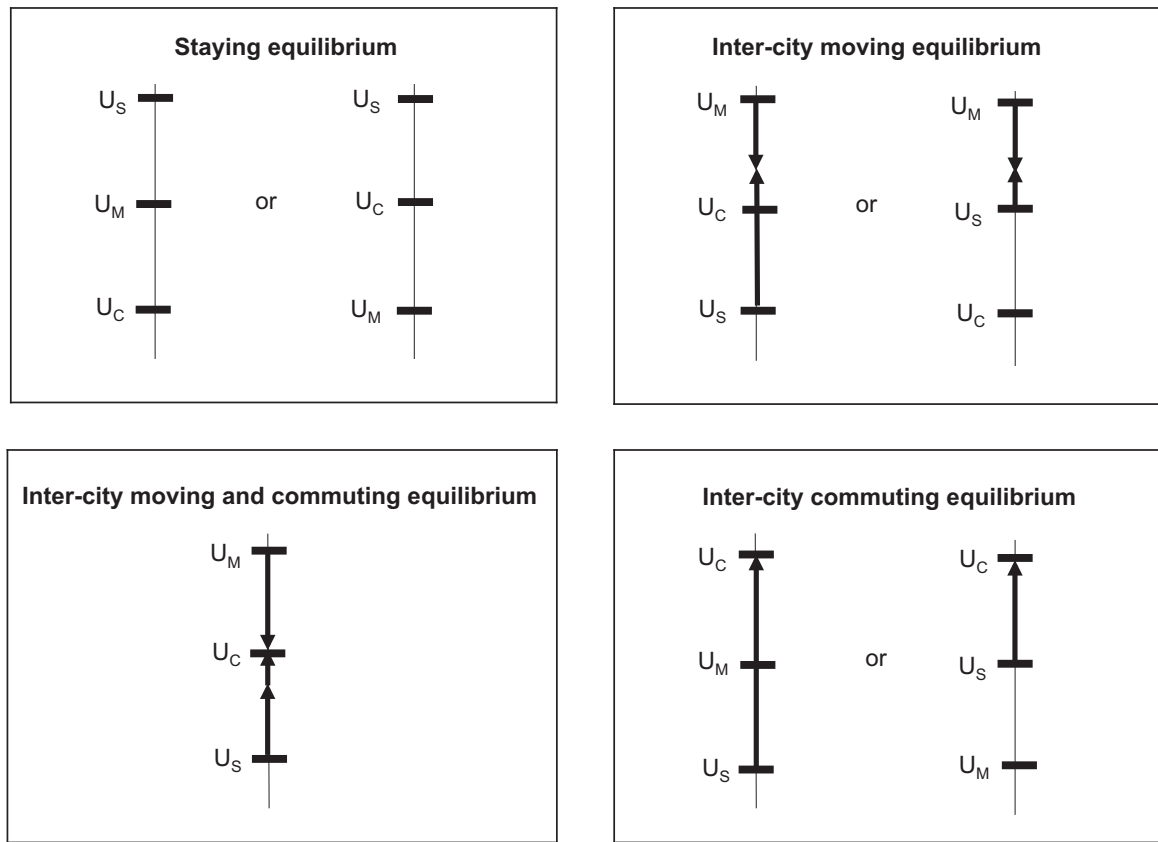
Fig. 1. Equilibrium Description. Given initial conditions, this figure graphically illustrates the four possible equilibrium configurations. Horizontal lines denote the initial utility levels for the different choices: U_S refers to the utility of an individual staying and working in her own city, U_M 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 $\min(m'', \bar{d}_\ell^0 - \min(m', d_\ell^0))$ 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 $A_k(1 - \gamma d_{\ell k}) = A_\ell(1 - \gamma(\bar{d}_\ell^0 - c))$.*

Urban Shadows and Urban Access. In Section 3.2 we documented the secular decline in commuting costs in the U.S. over the past 150 years. Our goal here is to explore how this drop affects urban shadows and urban access. To fix ideas, start off in a situation where all individuals reside in their city of origin and all have the same utility. This implies that the more productive city is also the larger one. Now consider a gradual drop in commuting costs γ. As long as commuting costs continue to be relatively high, no one has an incentive to move or commute to the larger city, because moving requires paying a fixed cost and inter-city commuting is still too expensive. However, if commuting costs drop to an intermediate level, the relative cost of intra-city commuting falls enough in the larger city for moving to become attractive. If commuting costs fall still further, commuting to the larger city becomes the better choice, as it saves on the fixed cost of moving.

This intuition suggests 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 A. □

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.

What does Result 1 tell us about urban shadows and urban access? 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 shadow: the nearby smaller city suffers in terms of population growth. Needless to say, if the initial drop in commuting is larger, the urban 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 shadow, but provides urban access instead: the nearby smaller city gains in terms of population growth.

How does this relate to our 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 urban shadows dominated and a later time period where urban access dominated, with a weakening in the benefits of urban access in more recent times. During the early time period, Result 1 also suggests 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.

Geographic Reach of Urban Shadows and Urban Access. We now explore how urban shadows and urban access depend on the distance to the larger city. The following result states that if inter-city distance increases, 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 A. □

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.

How does this relate to our empirical stylized facts? Result 2 allows us to trace the changing geographic reach of urban shadows and urban access as commuting costs fall. Initially, it predicts urban shadows at relatively short distances, that gradually expand as transport costs drop. Eventually, shadows are dominated by access, again first at relatively short distances, but later at farther away distances as the spatial reach of urban access expands. This is consistent with our empirical findings for the U.S., as summarized in STYLIZED FACT 3.

Relative Size of Large City. How do urban shadows and urban access 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 A. □

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. How does this relate to our empirical stylized facts? Result 3 implies that as commuting costs decline, it is the largest cities that first cast their urban shadow on their smaller neighbors, and likewise, it is the largest cities that first improve urban

access for their smaller neighbors. This is consistent with our empirical findings for the U.S., as summarized in STYLIZED FACT 4.

Model Assessment. By focusing on the documented long-run decline of commuting costs, our model is able to account for the main stylized facts we identified empirically when studying the relative importance of urban shadows and urban access 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 improved urban access through inter-city commuting 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 access 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. More generally, the influence of large metro areas may stretch further than standard forces would suggest, because of overlapping regions of interaction (Kerr and Kominers, 2015).

An alternative is that inter-city commuting costs are not the driving force behind what we observe in the data. One possibility is to reinterpret the benefits from inter-city commuting as technological spillovers. When commuting from city ℓ to city k, a resident of ℓ 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, $A_k(1 - \gamma d_{\ell k})$. We could alternatively model this effect through 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 the same discounted version of the neighboring city's productivity. In that sense, both interpretations are interchangeable in their effects on income.

Another possibility is that the attractiveness of the larger, more productive city comes from market access through inter-city trade, rather than through inter-city commuting. We turn to this possibility next.
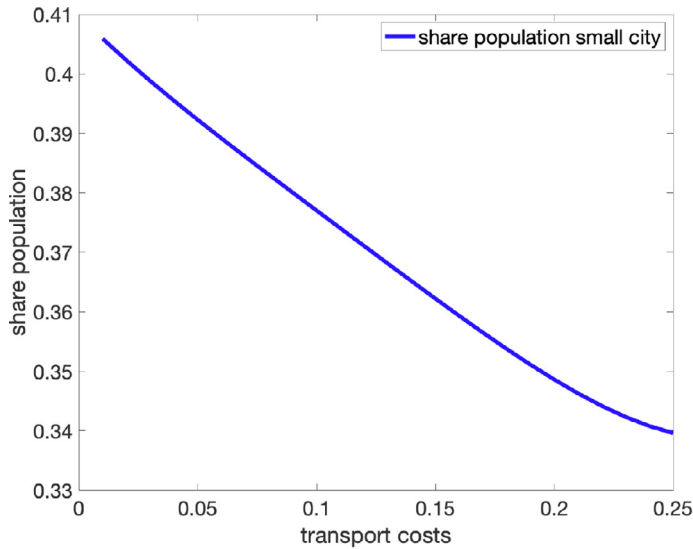
4.3. Intra-City commuting and inter-City trade

In this subsection we consider a second special case of our general setup by assuming that the inter-city commuting cost parameter, γ_i, is too high for individuals to commute between cities. That is, we focus on a model with intra-city commuting, inter-city trade and inter-city moving, but no inter-city commuting. To have a role for inter-city trade, we no longer assume the two goods to be perfect substitutes. As a result, we cannot normalize all goods prices to one. Instead, goods prices will have to be derived from standard goods markets clearing conditions.

Residential Mobility within Cities and Income. People can freely choose where to reside in their city. This implies that for all residents of ℓ, income net of land rents, y_ℓ, equalizes across all locations within a city. Since at the edge of ℓ land rents are zero,

$$y_\ell = p_\ell A_\ell (1 - \gamma_c \bar{d}_\ell). \tag{8}$$
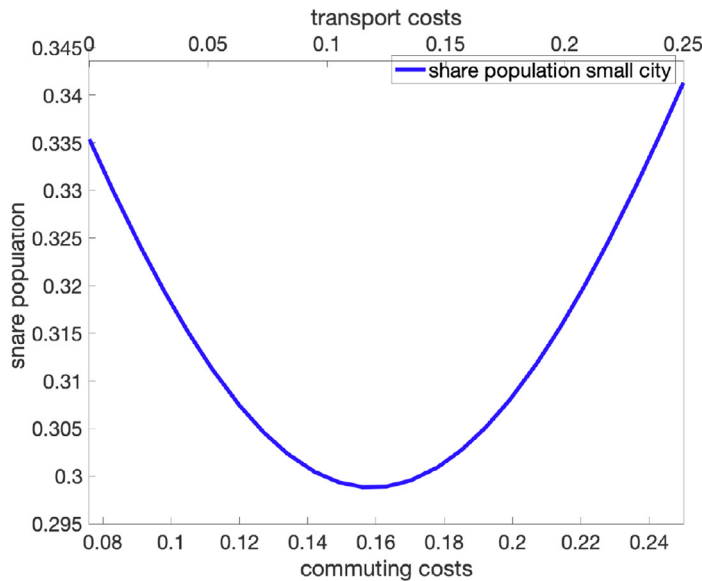
Since there is no inter-city commuting, expression (8) does not distinguish between residence and workplace.

(a) Drop in inter-city trade costs
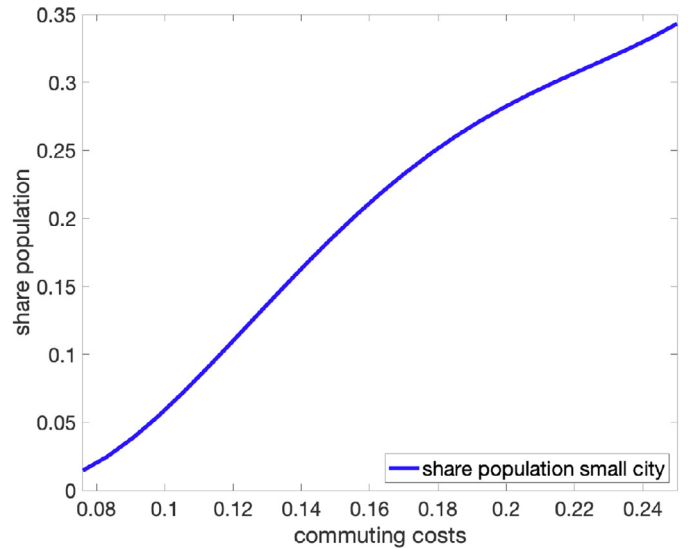
(b) Drop in intra-city commuting costs

(c) Relative drop in inter-city trade costs

Fig. 2. Model with inter-city trade costs and intra-city commuting costs.

Aggregate Supply and Demand. Net of the payments to land owners, each individual in ℓ produces $A_\ell(1 - \gamma_c \bar{d}_\ell)$. As a result, aggregate supply net of what is lost to land owners is

$$Q_\ell = 2A_\ell \bar{d}_\ell (1 - \gamma_c \bar{d}_\ell) = A_\ell L_\ell (1 - \gamma_c \bar{d}_\ell). \tag{9}$$

An agent faces two decisions: in which city to reside and how much to consume of each good. Both decisions can be separated. We start by describing the consumption decision. An agent who resides in ℓ maximizes[31]

$$\left(c_{\ell\ell}^{\frac{\sigma-1}{\sigma}} + c_{\ell k}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

---

subject to $p_\ell A_\ell (1 - \gamma_c \bar{d}_\ell) = p_\ell c_{\ell\ell} + p_k / (1 - \gamma_t d_{\ell k}) c_{\ell k}$. The first order conditions yield the following demand for each one of the two goods:

$$c_{\ell\ell} = \frac{y_\ell (p_\ell)^{-\sigma}}{p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma_t d_{\ell k}}\right)^{1-\sigma}}$$

$$c_{\ell k} = \frac{y_\ell \left(\frac{p_k}{1-\gamma_t d_{\ell k}}\right)^{-\sigma}}{p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma_t d_{\ell k}}\right)^{1-\sigma}}. \tag{10}$$

Aggregate demand for goods produced in location ℓ is then:

$$C_\ell = \frac{y_\ell L_\ell (p_\ell)^{-\sigma}}{p_\ell^{1-\sigma} + \left(\frac{p_k}{1-\gamma_t d_{\ell k}}\right)^{1-\sigma}} + \frac{y_k L_k \left(\frac{p_\ell}{1-\gamma_t d_{\ell k}}\right)^{-\sigma}}{\left(\frac{p_\ell}{1-\gamma_t d_{\ell k}}\right)^{1-\sigma} + p_k^{1-\sigma}}. \tag{11}$$

Analogous expressions to (9)-(11) can be written down for city k.

City Choice. An individual originally from 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

$$u_\ell^\ell = \frac{p_\ell A_\ell (1 - \gamma_c \bar{d}_\ell)}{\left(p_\ell^{1-\sigma} + (\frac{p_k}{1-\gamma_t d_{\ell k}})^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} \qquad (12)$$

whereas his indirect utility if he moves to k is

$$u_k^\ell = \frac{p_k A_k (1 - \gamma_c \bar{d}_k)}{\left((\frac{p_\ell}{1-\gamma_t d_{\ell k}})^{1-\sigma} + p_k^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} - \mu d_{\ell k}. \qquad (13)$$

In equilibrium, the population of ℓ should be such that an individual from city ℓ is indifferent between staying in his own city ℓ and moving to k:

$$\frac{p_\ell A_\ell (1 - \gamma_c \bar{d}_\ell)}{\left(p_\ell^{1-\sigma} + (\frac{p_k}{1-\gamma_t d_{\ell k}})^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} = \frac{p_k A_k (1 - \gamma_c (\frac{L}{2} - \bar{d}_\ell))}{\left((\frac{p_\ell}{1-\gamma_t d_{\ell k}})^{1-\sigma} + p_k^{1-\sigma}\right)^{\frac{1}{1-\sigma}}} - \mu d_{\ell k} \qquad (14)$$

where $\bar{d}_\ell = L/2 - \bar{d}_k$ and $\tilde{d}_\ell$ denotes the value of $\bar{d}_\ell$ that equalizes u_ℓ^ℓ and u_k^ℓ. An analogous expression to (14) for city k implicitly defines $\tilde{d}_k$. 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 $2(\bar{d}_k^0 - \tilde{d}_k)$, no one moves and everyone lives in their original location of residence. Hence, city choice implies:

$$\bar{d}_\ell = \begin{cases} \tilde{d}_\ell & \text{if } \tilde{d}_\ell < \bar{d}_\ell^0 \\ \bar{d}_\ell + (\bar{d}_k - \tilde{d}_k) & \text{if } \tilde{d}_k < \bar{d}_k^0 \\ \bar{d}_\ell^0 & \text{otherwise} \end{cases} \qquad (15)$$

Equilibrium. For given L, A_ℓ, A_k, $d_{\ell k}$, μ, γ_c, γ_t and σ, and for a given initial distribution of individuals across cities, $\bar{d}_\ell^0$ 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 (8), (14), (15), goods market clearing, labor market clearing $L = L_\ell + L_k$, land market clearing $L_\ell = 2\bar{d}_\ell$, as well as equivalent conditions for city k.

Urban Shadows and Urban Access. In their study of the long-run decline in transport costs, (Glaeser and Kahn, 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".

We use a simple numerical example to illustrate how this drop in trade costs relative to commuting costs affects urban shadows and urban access. 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, σ, is set 3. The total population L is set to 6, and inter-city distance $d_{\ell k}$ is set to 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 around 0.3% in terms of utility. For the initial commuting cost and trade cost parameters, we choose $\gamma_c = 0.25$ and $\gamma_t = 0.25$. Using these initial parameters, we distribute population between the two cities to equalize utility.

Before analyzing the effect of a drop in trade costs relative to commuting costs on the distribution of population across the two cities, it is useful to discuss both effects separately. 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. Moving from right to left in Panel (a) of Fig. 2 shows the gain in the population share of the smaller city as inter-city trade costs fall. 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. Moving from right to left in Panel (b) of Fig. 2 shows the gain in the population share of the larger city as intra-city commuting costs fall.

Fig. 2 Panel (c) combines both effects, with trade costs declining faster than commuting costs, as in Glaeser and Kahn (2004). When moving from left to right, inter-city trade costs decline from 0.25 to 0,

whereas intra-city commuting costs decline from 0.25 to 0.075. This first benefits the larger city (urban shadows dominate), and eventually the small city recovers its population (urban access dominate). Where does this non-monotonicity come from? From Panels (a) and (b), we would be tempted to conclude that urban shadows should dominate throughout. Indeed, for the same decreases in trade and commuting costs as in Panel (c), the gain to the large city from lowering commuting costs in Panel (b) is bigger than the gain to the small city from lowering trade costs in Panel (a). However, this ignores the interaction between both types of spatial frictions. As trade costs drop, residents from the small city have better access to the good produced in the large city, reducing the marginal gain from moving. As a result, the gain to the large city from lowering commuting costs is smaller when trade costs are lower. This leads to a weakening of urban shadows, and to the emergence of urban access as the dominant force.

Model Assessment. By focusing on the greater decline in trade than in commuting costs, our framework is able to account for STYLIZED FACT 1, our main empirical finding when studying the relative importance of urban shadows and urban access in the U.S. since the mid-19th century.

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 access 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 commuting costs first hurt, and then help, the growth of smaller locations in the vicinity of large urban centers. 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 particular, the growth of smaller locations might have benefited from improved access to large urban clusters through trade, rather than through commuting. Using an alternative conceptual framework that introduces trade between cities, we show that the rise and decline of urban shadows is also consistent with the observed faster drop in shipping costs than in commuting costs.

Appendix A. 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 $A_k (1 - \gamma_m \bar{d}_\ell^0) < A_k (1 - \gamma_m \bar{d}_k^0) - \mu d_{\ell k}$, which requires $\mu < (A_k (d_{\ell k} - \bar{d}_\ell^0)(A_k - A_\ell))/(d_{\ell k}(d_{\ell k} + A_k \bar{d}_k^0 - A_\ell \bar{d}_\ell^0))$, 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 to 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 $\bar{d}_\ell^0$ 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. $\square$

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jue.2021.103334.

References

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

Allen, T., Arkolakis, C., 2014. Trade and the topography of the spatial economy. Quarterly Journal of Economics 129, 1085–1140.

Baum-Snow, N., 2007. Did highways cause suburbanization? Quarterly Journal of Economics 122, 775–805.

Beltrán, F. J., Díez-Minguela, A., Martínez-Galarraga, J., forthcoming. The shadow of cities: Size, location, and the spatial distribution of population in Spain. Annals of Regional Science.

Bernard, A., 2014. Lifted: A cultural history of the elevator,. NYU Press, New York.

Bernhofen, D.M., El-Sahli, Z., Kneller, R., 2016. Estimating the effects of the container revolution on world trade. J. Int. Econ. 98, 36–50.

Black, D., Henderson, V., 2003. Urban evolution in the USA. Journal of Economic Geography 3, 343–372.

Bosker, M., Buringh, E., 2017. City seeds: geography and the origins of European cities. J. Urban Econ. 98, 139–157.

Christaller, W., 1933. Central places in southern Germany. Jena,, Germany. Fischer (English translation by C. W. Baskin, London: Prentice Hall, 1966

Conley, T., 1999. GMM Estimation with cross sectional dependence. J. Econom. 92, 1–45.

Cronon, W., 1991. Nature's metropolis: Chicago and the Great West. W. W. Norton, New York.

Davis, D.R., Weinstein, D.E., 2002. Bones, bombs, and break points: the geography of economic activity. American Economic Review 92, 1269–1289.

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

Desmet, K., Rappaport, J., 2017. The settlement of the United States, 1800–2000: the long transition towards Gibrat's law. J. Urban Econ. 98, 50–68.

Desmet, K., Rossi-Hansberg, E., 2014. Spatial development. American Economic Review 104, 1211–1243.

Dobkins, L.H., Ioannides, Y., 2001. Spatial interactions among U.S. cities. Reg. Sci. Urban Econ. 31, 701–731.

Donaldson, D., Hornbeck, R., 2016. Railroads and American economic growth: a market access approach. Quarterly Journal of Economics 131, 799–858.

Duranton, G., Nagpal, G., Turner, M., 2020. Transportation infrastructure in the U.S. NBER Working Paper # 27254.

Eckert, F., Peters, M., 2018. Spatial structural change. Yale University. Unpublished manuscript

Edlund, L., Machado, C., Sviatschi, M., 2016. Bright minds, big rent: Gentrification and the rising returns to skill. NBER Working Paper # 21729.

Fogel, R.W., 1964. Railroads and american economic growth: Essays in econometric history,. Johns Hopkins University Press, Baltimore, MD.

Fujita, M., Krugman, P., Venables, A.J., 1999. The spatial economy. MIT Press, Cambridge, MA.

Gardner, T., 1999. Metropolitan classification for census years before World War II. Hist. Methods 32, 139–150.

Glaeser, E.L., 2011. Triumph of the city,. MacMillan, London.

Glaeser, E.L., Kahn, M., 2004. Sprawl and urban growth. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, Vol.4.. Elsevier.

Hanson, G., 2005. Market potential, increasing returns and geographic concentration. J. Int. Econ. 67, 1–24.

Horan, P.M., Hargis, P.G., 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.

Jackson, K.T., 1985. Crapgrass frontier. the suburbanization of the United States. Oxford University Press.

Kerr, W.R., Kominers, S.D., 2015. Agglomerative forces and cluster shapes. Review of Economics and Statistics 97, 877–899.

Kopecky, K., Suen, M.H., 2004. Economie d'avant garde: Suburbanization and the Automobile. Research Report No. 6.

Krugman, P., 1993. On the number and location of cities. Eur. Econ. Rev. 37, 293–298.

Krugman, P., Venables, A.J., 1995. Globalization and the inequality of nations. Quarterly Journal of Economics 110, 857–880.

Lösch, A., 1940. The economics of location,. Jena,, Germany. Fischer (English translation, New Haven, CT: Yale University Press, 1954).

LeRoy, S.F., Sonstelie, J., 1983. Paradise lost and regained: transportation innovation, income, and residential location. J. Urban Econ. 13, 67–89.

Liu, Y., Wang, X., Wu, J., 2011. Do bigger cities contribute to economic growth in surrounding areas? Evidence from County-Level Data in China," unpublished manuscript.

Michaels, G., Rauch, F., Redding, S.J., 2012. Urbanization and structural transformation. Quarterly Journal of Economics 127, 535–586.

Mieszkowski, P., Mills, E.S., 1993. The causes of metropolitan suburbanization. Journal of Economic Perspectives 7, 135–147.

Partridge, M.D., Rickman, D.S., Ali, K., Olfert, M.R., 2009. Do new economic geography agglomeration shadows underlie current population dynamics across the urban hierarchy? Papers in Regional Science 88, 445–466.

Rappaport, J., 2005. The shared fortunes of cities and suburbs. Federal Reserve Bank of Kansas City Economic Review. Third Quarter, 33–59

Rappaport, J., 2007. Moving to nice weather. Reg. Sci. Urban Econ. 37, 375–398.

Rappaport, J., Sachs, J.D., 2003. The United States as a coastal nation. Journal of Economic Growth 8, 5–46.

Rauch, F., 2014. Cities as spatial clusters. Journal of Economic Geography 14, 759–773.

Redding, S., Sturm, D., 2008. The costs of remoteness: evidence from German division and reunification. American Economic Review 98, 1766–1797.

Rosenthal, S.S., Strange, W.C., 2003. Geography, industrial organization, and agglomeration. Review of Economics and Statistics 85, 377–393.

Shaw, R.E., 1990. Canals for a nation. the canal era in the United States, 1790–1860,. The University Press of Kentucky.

Su, Y., 2018. The rising value of time and the origin of urban gentrification. Unpublished manuscript.

Tabuchi, T., Thisse, J.F., 2011. A new economic geography model of central places. J. Urban Econ. 69, 240–252.

Thorndale, W., Dollarhide, W., 1987. Map guide to the Federal Censuses, 1790–1920,. Genealogical Publishing Company, Baltimore.

U. S. Bureau of the Census, 1905. Street and electric railways 1902. Government Printing Office, Washington.

U. S. Bureau of the Census, 1910. Street and electric railways 1907. Government Printing Office, Washington.

Warner Jr., S.B., 1972. Streetcar suburbs. the process of growth in Boston, 1870–1900,. Harvard University Press and the MIT Press.

Young, M., 2015. Infrastructure: Mass transit in 19th and 20th century urban America. Oxford Research Encyclopedia of American History. Online Publication Date: March 2015