# The Faster Growth of Larger, Less Crowded Locations

*By Jordan Rappaport*

Over the past few decades, the population and employment of small and large locations in the United States have been diverging. Most of the smallest locations in the United States—the approximate 1,200 counties and micropolitan areas with a population below 25,000—saw declining population and employment from 2000 to 2017 as their residents and jobs migrated to larger, more prosperous locations. Conversely, almost all medium and large metropolitan areas in the United States—those with a population of 500,000 or more—saw increasing population and employment from 2000 to 2017, many at well above the national rate.

An important question is whether this divergence between small and large locations has been driven by size itself. One possibility is that the benefits of size have become greater over time. For example, businesses may increasingly benefit from being near suppliers. Likewise, households may increasingly value access to services and amenities that are only available in larger locations. Alternatively, the divergence may be driven by characteristics that are correlated with size but not inherent to it. For example, the slower growth of smaller locations may simply reflect their disproportionate specialization in the manufacturing and agriculture sectors, which have seen declining employment.

In this article, I document the faster population and employment growth of medium and large metropolitan areas compared with smaller locations. Among these smaller locations—rural counties, micropolitan

*Jordan Rappaport is a senior economist at the Federal Reserve Bank of Kansas City. Anissa Khan, an intern at the bank, helped prepare the article. This article is on the bank's website at* ***www.KansasCityFed.org***

areas, and small metropolitan areas—growth is strongly positively correlated with population. Statistical analysis shows that most of this positive correlation is likely driven by size itself rather than location characteristics correlated with size. Among the medium and large metropolitan areas, growth is only weakly correlated with population but strongly negatively correlated with population density, a measure of crowdedness that moves closely with population. This negative correlation with density, too, is likely driven by density itself rather than correlated characteristics. Together, growth's positive correlation with population and negative correlation with density suggest that both the benefits and costs of size have increased over the past few decades.

Section I documents the relationship between population growth and size: population growth is positively correlated with size up to a population of about 500,000, uncorrelated with increases in size from 500,000 to 3 million, and negatively correlated with increases in size above 3 million. Section II lays out a framework for interpreting these correlations between growth and size: differences in locations' population and employment growth typically reflect relative changes in locations' productivity and amenities. Section III documents that the positive correlation of growth and size holds even after controlling for differences in local characteristics.

## I. The Positive Relationship between Population and Employment Growth and Size

To analyze the relationship between growth and size, I look at all locations within the continental United States. Specifically, I combine the 358 metropolitan and 554 micropolitan areas delineated after the 2000 decennial census with the 1,346 remaining counties that are not part of a metropolitan or micropolitan area. The metropolitan areas, which range in population from 52,000 to 18 million, are combinations of counties surrounding a dense core of at least 50,000 residents. Most are made up of two or more counties. For descriptive purposes, I divide the metropolitan areas into three groups: small (population up to 500,000), medium (population from 500,000 to 3 million), and large (population above 3 million). The micropolitan areas, which range in population from 13,000 to 182,000, are combinations of counties surrounding a dense core of 10,000 to 50,000 residents. Most are made up of a single county. The remaining counties (henceforth, "non-core") range in population from 67 to 97,000. I measure growth rates using a constant delineation of metropolitan and micropolitan land areas. Thus, any changes in metropolitan and micropolitan area populations attributable to changes in their land area are excluded from measured growth rates.[1]

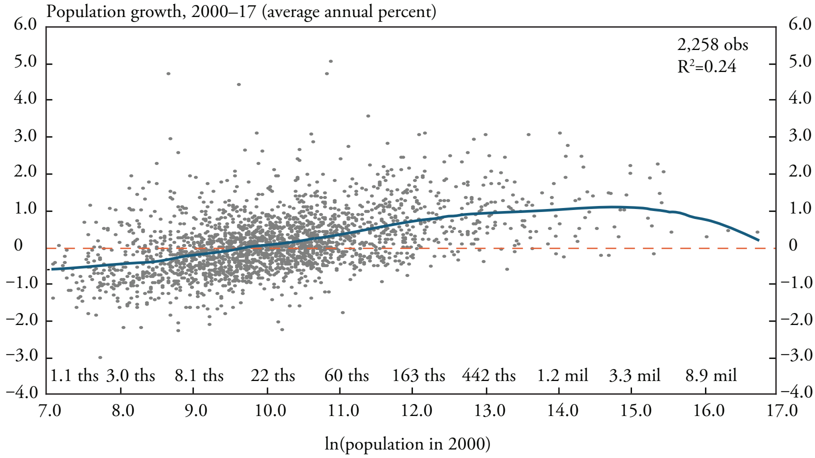
Chart 1 shows a scatter plot of the average annual growth rate of locations' population from 2000 to 2017 against the natural log of their population in 2000. I take the natural log of population to allow the horizontal axis to measure proportional rather than additive changes in population: each log point increase moving rightward along the horizontal axis represents a multiplicative increase in population by a factor of 2.7.[2]

The chart shows a clear, positive relationship between locations' population growth from 2000 to 2017 and their level of population in 2000. The blue line shows a smoothed average of this relationship, which can be interpreted as the predicted rate of growth based on initial size.[3] Predicted population growth ranges from modestly negative for locations with a low population in 2000 to moderately positive for most metropolitan areas. For example, a location with a population of 1,500 in 2000 (log population of 7.3) has a predicted population growth rate of –0.6 percent per year, leading to a 10 percent cumulative loss in population from 2000 to 2017. In contrast, a location with a population of 500,000 (log population of 13.1) has a predicted growth rate of just under 1 percent per year, leading to a 17 percent cumulative gain in population from 2000 to 2017. The predicted population growth rate is highest, 1.1 percent per year, for locations with a population close to 3 million (log population of 14.9).

Locations' predicted population growth falls off as their population in 2000 exceeds 3 million. Los Angeles and New York City, the two largest metropolitan areas with respective populations of 12.4 million and 18.3 million and log populations of 16.3 and 16.7, have predicted growth rates of 0.5 percent and 0.2 percent per year, meaningfully lower than the maximum 1.1 percent rate. But the decline from the maximum rate is slight for most large metropolitan areas. For example, Philadelphia, the fourth-largest metropolitan area in 2000 (with a population of 5.7 million and log population of 15.6) has a predicted

*Chart 1*

## Population Growth versus Initial Population, 2000–17




Notes: The blue line represents a prediction of locations' growth rates based on their population. The dashed orange line corresponds to a growth rate of zero. Replication code is available in an online data supplement.
Sources: U.S. Census Bureau and author's calculations.

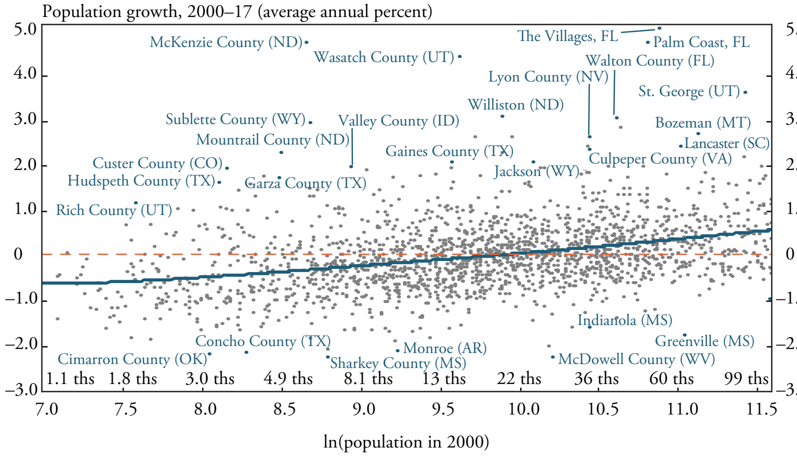
growth rate of 1.0 percent per year, only a tick below the maximum predicted rate.

Overall, locations with population above 500,000 on average have considerably higher predicted growth than medium-sized locations, which in turn have considerably higher predicted growth than small locations. Of course, the actual growth rates of many locations differed considerably from the predicted rate based on their size. The large vertical dispersion of the scatter above and below the average line reflects that characteristics other than initial size drove most of the variation in growth rates from 2000 to 2017.

Correspondingly, small size did not preclude rapid growth. Chart 2 zooms in on the left-hand side of Chart 1, showing the same scatter and average relationship for locations with a population in 2000 below 100,000 (log population of 11.5). Many of the locations that grew fastest relative to their predicted rates—that is, those furthest above the predicted growth line—are distinguished by natural amenities such as mountains (for example, Custer County, CO; Rich County, UT; Wasatch County, UT; St. George, UT; Valley County, ID; Jackson, WY; Bozeman, MT; and Lyon County, NV) or warm winter weather (for example, Palm Coast, FL; The Villages, FL; and Walton County, FL).[4] Others are adjacent to metropolitan areas (for example, Hudspeth

*Chart 2*

## Growth versus Initial Population, Smaller Locations




Notes: Metropolitan and micropolitan areas are labeled with the name of their largest city. The blue line represents a prediction of locations' growth rates based on their population and estimated using all 2,258 locations. The dashed orange line corresponds to a growth rate of 0. Marker for Issaquana County, MS (log population 7.7, growth rate –3.0 percent) is below the displayed area. Replication code is available in an online data supplement.
Sources: U.S. Census Bureau and author's calculations.

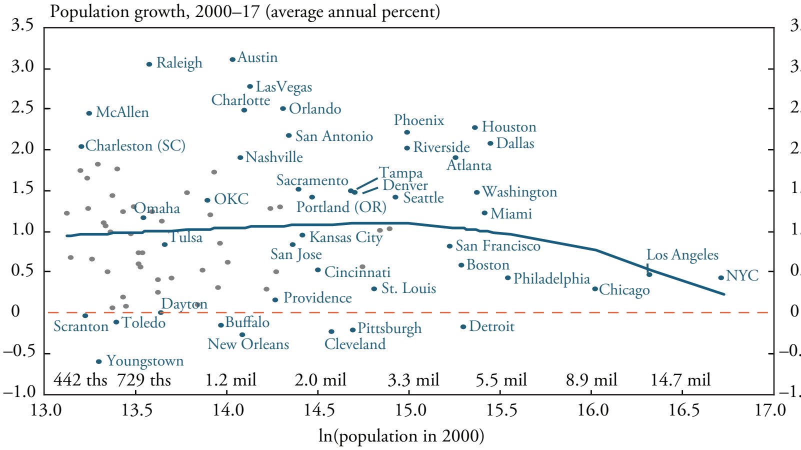
County, TX; Culpeper County, VA; and Lancaster, SC) or benefitted from oil and natural gas deposits that became accessible with the development of hydraulic fracturing technology (Williston, ND; Mountrail County, ND; McKenzie County, ND; Sublette County, WY; Garza County, TX; and Gaines County, TX).

Similarly, large size did not preclude population decline. Chart 3 zooms in on the right-hand side of Chart 1, showing the same scatter plot and average relationship for medium and large metropolitan areas, those with a population in 2000 of at least 500,000. In nine of these metros, population actually declined. Among these, eight are distinguished by an industrial composition skewed heavily toward manufacturing, a sector in which employment has been contracting for many decades. The disadvantages of this inherited industrial composition are likely to have offset any benefits from size.

Overall, however, declining population was relatively rare for medium and large metropolitan areas as well as for smaller metropolitan areas with a population in 2000 between 200,000 and 500,000 (Chart 4). In contrast, the majority of locations with a population

*Chart 3*

## Growth versus Initial Population, Medium and Large Metropolitan Areas




Notes: Metropolitan areas are labeled with the name of their largest city. The blue line represents a prediction of locations' growth rates based on their population and estimated using all 2,258 locations. The orange dashed line corresponds to a growth rate of 0. The Denver and Boulder metropolitan areas are combined. Replication code is available in an online data supplement.
Sources: U.S. Census Bureau and author's calculations.

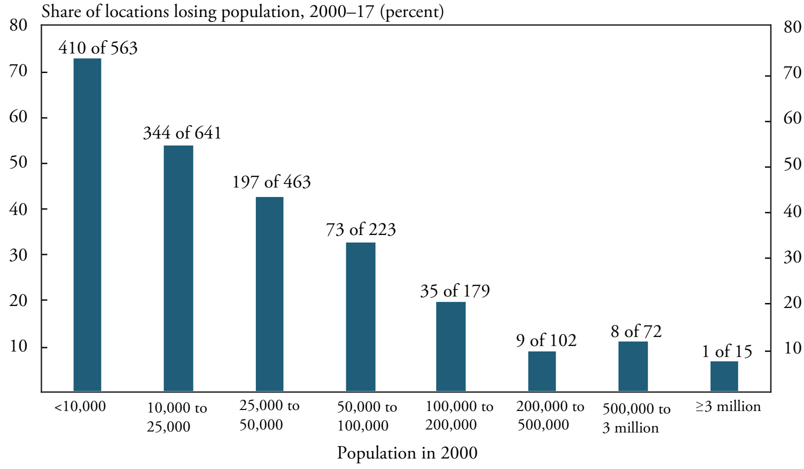
below 25,000 contracted, as did more than 40 percent of locations with a population between 25,000 and 50,000. Across all locations, population accounts for almost a quarter of the variation in growth rates from 2000 to 2017 (as estimated by the $R^2$ statistic). This is a high share attributable to a single characteristic.

The relationship between population growth and size from 2000 to 2017 continued a pattern that began in the mid-twentieth century. Chart 5 shows the predicted population growth rates from 1960 to 1980 (blue line) and from 1980 to 2000 (green line) based on the corresponding initial population levels. Both predicted relationships are characterized by a positive correlation between growth and size across most locations and a negative relationship across the largest locations.[5]

However, the relationship between growth and size evolved over these periods in four important ways. First, predicted growth shifted lower over time for locations with an initial population up to about 1 million, primarily reflecting slowing national population growth. As a result, the share of small locations with predicted population decline increased over time. Second, the downward slope in the relationship for

*Chart 4*

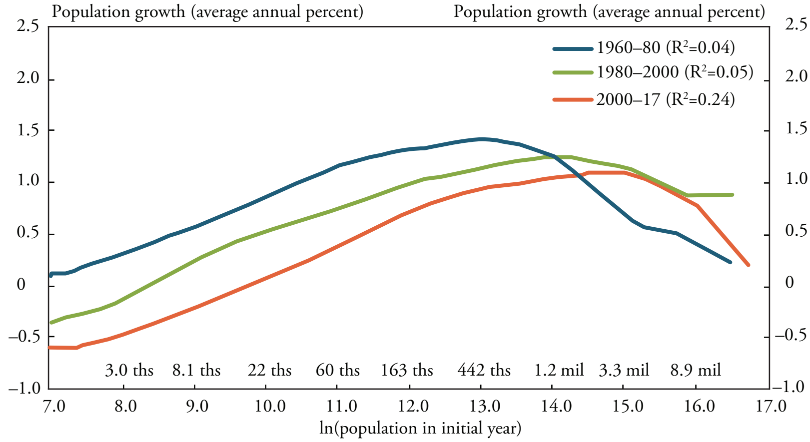
## Share of Locations Losing Population




Note: Replication code is available in an online data supplement.
Sources: U.S. Census Bureau and author's calculations.

*Chart 5*

## Historical Population Growth versus Initial Population



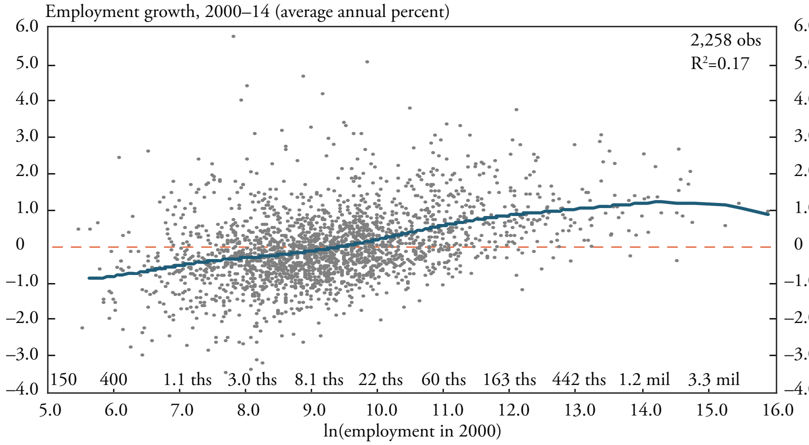

Sources: U.S. Census Bureau, Desmet and Rappaport (2017), and author's calculations.

the largest locations began at successively higher population levels over time. From 1960 to 1980, predicted growth began sloping down at a population of about 500,000. From 1980 to 2000, predicted growth began sloping down at a population of about 1.3 million. In the most recent period, it began sloping down at about 3.0 million. Third, the magnitude of the downturn for the largest locations varied over time. Predicted growth declined more in the 2000–17 period than in the 1980–2000 period but less than in the 1960–80 period. Fourth, the relationship between growth and size was strongest during the most recent period. Initial size accounted for 24 percent of the variation in growth from 2000 to 2017 compared with a maximum of 5 percent in the earlier periods. This increase in explanatory power may reflect size becoming a more important determinant of growth or other determinants of growth, such as suburbanization and the migration to the Sunbelt, becoming less important.

Growth's relationship with employment is similar to its relationship with population. Chart 6 shows a scatter of locations' employment growth from 2000 to 2014 plotted against their initial level of employment in 2000. The black line represents predicted growth.[6] The employment levels along the horizontal axis are lower than the population levels discussed previously, reflecting that the number of individuals with jobs, including both full time and part time, is less than one-half of the population in most locations. Employment growth from 2000 to 2014 was positively correlated with employment in 2000 up to a level of about 500,000 employed individuals (log employment of 13.1) and approximately uncorrelated with further increases in employment. Predicted average annual employment growth rose from about –0.8 percent for locations with fewer than 500 employed individuals in 2000 (log employment below 6.2) to about 1.0 percent for locations with more than 500,000 employed individuals.

The similar relationships between size and growth of both population and employment reflect that employment and population growth are strongly positively correlated over the long term. In particular, increases in employment tend to be matched approximately one for one by inflows of workers (Rappaport 2012). I focus my subsequent analysis on population rather than employment, as it is the better measured of the two.

*Chart 6*
Employment Growth versus Initial Employment, 2000–14




Notes: The orange dashed line corresponds to a growth rate of 0. Several locations have employment growth rates outside the displayed range. Replication code is available in an online data supplement.
Sources: U.S. Census Bureau and author's calculations.

## II. Interpreting Correlations with Growth

Local population and employment growth are not inherently desirable or undesirable outcomes. Rather, they reflect changes in the underlying fundamentals determining where households and businesses choose to locate. In this section, I introduce a local growth framework that illustrates how these fundamentals are linked to population and employment. I then suggest some shifts in fundamentals that might be driving the observed relationship between growth and size.

### *A framework for understanding local population and employment growth*

The local growth framework has three key features. First, locations have different fundamental characteristics that affect the productivity of businesses or that serve as amenities for residents.[7] Some of these characteristics are exogenous in the sense that they do not depend on local outcomes such as income and population—for example, natural resources, a natural ocean harbor, natural recreational opportunities, and nice weather. Other characteristics are endogenous in the sense that they are themselves a local outcome—such as population, employment, and income—or partly depend on a local outcome. For example, a larger population may contribute to disamenities such as traffic congestion and pollution. Likewise, the age and income distribution of local residents may affect the variety of available goods and services (Glaeser, Kolko, and Saiz 2001; Diamond 2016).[8]

Second, an economy is in a spatial equilibrium if all households and firms prefer to remain where they are rather than move elsewhere (Rosen 1979; Roback 1982). In this equilibrium, businesses cannot increase their profits by moving somewhere else because higher productivity in other locations is offset by higher wages, land prices, and other costs. Similarly, households cannot benefit from moving somewhere with higher wages or more amenities because these advantages are offset by higher housing prices, more traffic congestion, and other costs. As is intuitive, locations with characteristics that contribute to high productivity or amenities have a larger equilibrium population (Rappaport 2008a, 2008b, 2016). The larger population pushes up land prices, housing prices, traffic congestion, and other costs to the level at which businesses and households are equally willing to live in locations with lower productivity and amenities but also lower costs.

Third, a location's transition from its current level of population to its equilibrium level, driven by net flows of households and businesses, takes considerable time (Rappaport 2004; Desmet and Rappaport 2017). Locations with a current population significantly below its equilibrium level will typically grow at an above-average rate for several decades as people gradually migrate there. Locations with an initial population significantly above its equilibrium level will typically grow at a below-average rate for several decades as people gradually migrate elsewhere. These transitions are gradual for several reasons, including physical moving costs, households' ties to family and friends in origin locations, the time it takes for housing and infrastructure to deteriorate in origin locations, and the time it takes to build new housing and infrastructure in destination locations (Glaeser and Gyuorko 2005; Kennan and Walker 2011; Davis and others 2013).

Although the local growth framework emphasizes location size as a consequence of productivity and amenities, size is also an important determinant of location productivity and amenities. For example, large size contributes negatively to productivity and amenities by increasing numerous types of congestion.[9] But large size also contributes positively to productivity and amenities in numerous ways. Such agglomerative benefits typically take the form of more sharing, better matching, and increased learning. (Duranton and Puga 2004). Agglomerative sharing captures large locations' ability to spread large fixed costs, such as building an airport or sports stadium, across a broad base of customers. It also captures large locations' ability to support a wider variety of business and consumer services, especially those that are more specialized. Agglomerative matching is exemplified by a larger pool of job candidates and firms that allows for a better fit of workers' skills to firms' needs. For example, research shows that a larger pool of employers has become more important over time as the share of couples with dual careers has risen (Costa and Kahn 2000). Agglomerative learning concerns the generation and diffusion of knowledge. For example, researchers tend to discover more, as measured by patents, when working near each other (Carlino and Kerr 2015; Buzard and others 2017). Likewise, when many workers in the same occupation are concentrated in one location, they learn from each other. As Alfred Marshall observed in 1890, "the mysteries of the trade become no mysteries; but are as it were in the air." Estimates based on wages suggest that these agglomeration effects together increase a location's productivity by between 2 and 6 percent for each log point increase in population. (Combes and Gobillon 2015). No comparable estimates exist for the effect of size on amenities.

An important implication of the local growth framework is that differences in growth rates typically reflect *changes* in underlying productivity and *changes* in underlying amenities rather than levels. At any point in time, the distribution of population across locations already captures many of the differences in local productivity and amenities, which tend to persist over very long periods. Metaphorically, differences in locations' current size result from how firms and people have "voted with their feet" up until that point in time (Tiebout 1956). Differences in growth rates, on the other hand, reflect firms and people changing their "votes." Because transitions are extended, this vote changing can persist for up to several decades following a change in productivity or amenities.[10]

### *Interpreting the relationship between growth and size*

The local growth framework suggests at least three possible interpretations of the empirical relationships between the levels of population and employment and their growth rates. One possible interpretation is that the agglomerative benefits from increases in population up to a level of 500,000 have become larger during recent decades.[11]

Following such an increase in agglomerative benefits, the equilibrium size of locations with a previous equilibrium population above 500,000 would have risen *relative* to the new equilibrium size of locations with a previous equilibrium population below 500,000.[12]

Under this first interpretation, the lack of correlation between growth and size across medium metropolitan areas suggests that the agglomerative benefits from increases in population from 500,000 to 3 million have remained approximately the same during recent decades. The negative correlation between growth and size across large metropolitan areas suggests that the agglomerative benefits from increases in population above 3 million have become smaller during recent decades.[13]

A second possible interpretation is that the agglomerative costs from increases in population up to a level of 500,000 have become smaller during recent decades. The spatial equilibrium of locations' size depends on the extent to which higher productivity and amenities are offset by higher housing prices, traffic congestion, and other agglomerative costs as more people compete for housing, road space, and other goods and services.[14] Thus, if agglomerative costs become less sensitive to population—for example, if cities relax zoning restrictions, expand public transit, or improve highway infrastructure—then larger locations will benefit proportionally more than smaller locations.[15] In essence, the lowered costs of size allow more households and businesses to crowd into locations with high productivity and amenities before rising house prices and congestion offset the gains from doing so.

Under this second interpretation, the flat and declining portions of the relationship between growth and population suggest that the agglomerative costs from increases in population from 500,000 to 3 million have remained approximately the same, while the agglomerative costs from increases in population above 3 million have become larger.[16]

A third possible interpretation is that the contributions of a location's exogenous characteristics to productivity and amenities have changed during recent decades. Such changes would likely induce correlations between growth and size because the same characteristics, through their previous contributions to productivity and amenities, helped determine the location's previous equilibrium population. For example, households during the nineteenth century were likely to have preferred, all else equal, to live where winters were less cold and summers were less hot and humid. As a result, locations with mild weather were likely to have, all else equal, a larger equilibrium population than locations that did not. During the early twentieth century, this preference for mild weather began strengthening (Rappaport 2007). The resulting faster population growth of locations with mild weather induced a positive correlation between growth and size, reflecting that locations with mild weather tended to have above-average size. However, the underlying impetus for this positive correlation was not intrinsically related to size. Similarly, the industry and occupation composition of many small locations is skewed toward agriculture and manufacturing, sectors for which employment has been declining during recent decades. The slower growth of smaller locations may partly reflect this industrial shift rather than size.

## III. The Increased Benefits and Costs of Size

The local growth framework suggests three possible interpretations of the empirical correlations between growth and size. To assess which interpretation is most likely, I first run regressions of population growth on initial population and several additional location characteristics to rule out that such characteristics, rather than changes in agglomerative benefits and costs, are driving the correlations. I then run regressions of population growth on initial population density to help distinguish whether changes in agglomerative benefits or agglomerative costs are driving the correlations.

### *The increased benefits of size for smaller locations*

Table 1 reports results from regressing average annual population growth from 2000 to 2017 on initial population in 2000 and additional characteristics such as geographic location and industry composition. I divide initial population into a "spline" of eight population ranges to allow the regression to approximate the smoothed relationship between predicted growth and population (the blue line in Chart 1). The regression coefficient on each of the population ranges estimates the slope of a linear segment corresponding to the curved line through that population range.

The results in column 1 show that regressing population growth on the spline without controlling for other characteristics approximately

*Table 1*

## Partial Correlation of Population Growth with Population

| Partial correlations of 2000–17 population growth | (1) Own size only | (2) Baseline | (3) Baseline and industry composition | (4) Baseline, industry composition, and occupation composition |
|---|---|---|---|---|
| Controls: | | | | |
| Metropolitan adjacency (2) | | x | x | x |
| Weather (10) | | x | x | x |
| Coast and river adjacency (7) | | x | x | x |
| Hilliness (2) | | x | x | x |
| Shale basin (6) | | x | x | x |
| Higher education (1) | | x | x | x |
| Industry composition (18) | | | x | x |
| Occupation composition (21) | | | | x |
| ln(pop) from 5.9 to 8 (3,000) | 0.06 | 0.16 | 0.36** | 0.41*** |
| *128 locations* | (0.12) | (0.10) | (0.15) | (0.16) |
| ln(pop) from 8 to 9 (8,100) | 0.27** | 0.35*** | 0.26*** | 0.28*** |
| *319 locations* | (0.12) | (0.07) | (0.07) | (0.07) |
| ln(pop) from 9 to 10 (22,000) | 0.30*** | 0.27*** | 0.18*** | 0.18*** |
| *627 locations* | (0.08) | (0.07) | (0.07) | (0.06) |
| ln(pop) from 10 to 11 (60,000) | 0.25*** | 0.19*** | 0.19*** | 0.20*** |
| *620 locations* | (0.09) | (0.06) | (0.06) | (0.05) |
| ln(pop) from 11 to 12 (163,000) | 0.46*** | 0.50*** | 0.39*** | 0.36*** |
| *292 locations* | (0.08) | (0.09) | (0.09) | (0.09) |
| ln(pop) from 12 to 13 (440,000) | 0.18 | 0.09 | 0.04 | 0.03 |
| *126 locations* | (0.13) | (0.10) | (0.09) | (0.09) |
| ln(pop) from 13 to 14 (1.2 million) | 0.12 | 0.17 | 0.00 | –0.06 |
| *60 locations* | (0.20) | (0.15) | (0.11) | (0.10) |
| ln(pop) from 14 to 16.7 (18.3 million) | –0.09 | –0.13 | –0.16 | –0.16 |
| *41 locations* | (0.13) | (0.12) | (0.10) | (0.10) |
| Observations | 2,258 | 2,258 | 2,258 | 2.258 |
| $R^2$ | 0.24 | 0.43 | 0.52 | 0.54 |
| Adjusted $R^2$ | 0.23 | 0.42 | 0.50 | 0.52 |
| Control variables | | 28 | 46 | 67 |
| $R^2$, control variables | | 0.30 | 0.48 | 0.51 |

** Significant at the 5 percent level
*** Significant at the 1 percent level

Notes: Dependent variable is average annual population growth (percent) from 2000 to 2017. Regressions also include a constant. The smallest location has a log population of 5.9 (population of 356). Standard errors are in parentheses and adjust for spatial correlation based on Conley (1999). Italicized text reports the number of locations with a population that lies within each spline segment. Coefficients on all variables included in the baseline regression are reported in appendix Table A-1. Replication code is available in an online data supplement.

matches the smoothed relationship between predicted growth and initial size. Each log point increase in population from 8 to 12 (corresponding to a population increase from about 3,000 to 163,000) is associated with between 0.25 and 0.46 percentage point faster predicted growth per year, implying large differences in cumulative growth. Three of the corresponding coefficients statistically differ from zero at the 1 percent level, and the fourth coefficient statistically differs from zero at the 5 percent level. Predicted growth also rises modestly as log population increases from 12 to 13 and from 13 to 14 and then falls modestly as log population rises above 14. But none of the coefficients on these segments statistically differs from zero. Overall, the initial population spline accounts for 24 percent of the variation in location growth rates, the same as the share accounted for by the smoothed relationship shown in Chart 1.

The statistically significant, positive relationship between growth and size for locations with log population from 8 to 12 continues to hold after accounting for numerous other characteristics, suggesting that the correlation is indeed driven by changes in agglomerative benefits and costs rather than by changes in the contributions of exogenous characteristics to productivity and amenities. Column 2 of Table 1 shows results from a regression that controls for 28 baseline characteristics likely to affect productivity and amenities and thereby drive both growth and size. On their own, these baseline characteristics—which describe adjacency to metropolitan areas, adjacency to coasts and rivers, weather, hilliness, energy deposits, and the presence of universities and colleges—account for a considerable portion of the variation in growth rates (30 percent, reported in the bottom row) and an even larger portion of the variation in the level of population in 2000 (40 percent, not shown).[17] But they leave the coefficients on the population spline mostly unchanged, ruling out that any of the baseline characteristics is driving the correlation between growth and size.

The positive correlation between growth and size similarly continues to hold after controlling for the industry and occupation composition of locations. Column 3 reports results from a regression that includes the baseline characteristics along with variables measuring the share of aggregate employment in each of 18 industries. Column 4 reports results from a regression that includes the baseline and industry characteristics along with variables measuring the share of aggregate employment in each of 21 occupations. In both regressions, the positive coefficients on the spline segments with log population between 8 and 12 are mostly unchanged. Both regressions also estimate a large, statistically significant positive coefficient on the lowest spline segment (log population from 5.9 to 8; population from 350 to 3,000), suggesting that even across the smallest locations, the net productivity and amenity benefits of size increased. In addition, including the industry and occupation controls boosts the magnitude of the negative coefficient on the uppermost spline segment and lowers its standard error. As a result, the coefficients on the uppermost segment in columns 3 and 4 statistically differ from 0 at only slightly above the 10 percent level, suggesting that the net productivity and amenity benefits of size may have decreased for large metropolitan areas.[18]

The estimated coefficients in Table 1 imply that differences in population among smaller locations predict large differences in growth rates from 2000 to 2017. To obtain the differences in growth rates between a location with log population in 2000 of 8 (a population of 3,000) and a location with log population of 12 (a population of 163,000), I sum the coefficients of the four spline segments from 8 to 12 for each of the specifications. This simple calculation shows that a location with log population in 2000 of 12 has from 1.0 to 1.3 percentage points faster predicted annual growth from 2000 to 2017, corresponding to a larger cumulative increase in population from 19 to 25 percentage points.

Although characteristics excluded from these regressions could account for the positive correlation between growth and size, such a possibility seems unlikely. On their own, the 67 characteristics included in the column 4 regression account for more than half of the variation in population growth, a high share given the many idiosyncratic circumstances affecting local growth. Moreover, many excluded characteristics are likely to endogenously depend on size. Including such endogenous variables in a regression may help identify channels through which size affects growth but might also mask the effect of size through all channels.

The positive, statistically significant coefficients reported in Table 1 suggest that the *net* benefits of larger size—the gross agglomerative benefits of higher productivity and amenities less the gross agglomerative costs of higher housing prices and more traffic congestion—increased for non-core counties, micropolitan areas, and the smallest metropolitan areas. But the coefficients do not distinguish whether this net change arose from an increase in agglomerative benefits or a decrease in agglomerative costs. The former seems more likely, as these groups of locations have historically been characterized by relatively low home prices, minimal commuting traffic, and few other congestion costs.

Conversely, the negative coefficients on the uppermost spline segment suggest that the net benefits of larger size may have decreased for metropolitan areas with population above 1.2 million. Again, the coefficients do not distinguish whether this arose from a decrease in agglomerative productivity and amenities or an increase in agglomerative costs, both of which seem plausible.

### *The increased costs of size for larger locations*

To distinguish whether changes in agglomerative benefits or costs are driving the relationship between growth and size, I look at the relationship between population growth and population density. While strongly positively correlated with the level of population, population density appears to be more closely related to home prices, a key agglomerative cost, than does population. In particular, population density accounts for more than twice the variation in median home prices across medium and large metropolitan areas.[19] In addition, population density is unlikely to affect businesses productivity. (Employment density, in contrast, is likely to affect businesses productivity by allowing more workers to interact with each other.)

Population density varies greatly within metropolitan areas, making "raw" density (total population divided by total land area) a poor summary measure of the density of the neighborhoods in which most residents live. For example, the raw density of the Las Vegas metropolitan area in 2000 was 174 persons per square mile. However, this measure is misleading, as 85 percent of Las Vegas residents lived in census tracts—small geographic units that typically include between 1,000 and 8,000 residents—with raw population of more than 2,100 persons per square mile.

To better reflect the density most residents actually experience, I measure mean population density, calculated as the population-weighted mean of each census tract's raw density (Glaeser and Kahn 2004; Rappaport 2008a).[20] Using this measure, the mean density of Las Vegas in 2000 was 6,500 persons per square mile. Across all locations, mean density ranged from less than 1 person per square mile for the 35 locations with the lowest value to more than 8,000 persons per square mile for the five locations with the highest value. Mean population density in the New York City metropolitan area, 32,600 persons per square mile, was almost three times that of the second most dense metropolitan area, Los Angeles.

Table 2 reports results from regressing population growth on a spline of mean population density in 2000. The results in column 1 show that before controlling for other characteristics, positive coefficients on three of the spline segments statistically differ from zero, implying that increases in mean density within each of the corresponding ranges predict faster population growth. However, controlling for the baseline characteristics and industry composition pushes down each of these positive coefficients to near zero (column 3).[21] In other words, any positive association between predicted growth and mean density may be driven by differences in the baseline characteristics and industry composition rather than by a change in agglomerative costs.

In contrast, population growth is negatively correlated with increases in mean population density within the uppermost segment, a relationship that strengthens as additional controls are added to the regression. Controlling for the baseline characteristics and industry composition, the negative coefficient on the uppermost segment statistically differs from 0 at the 1 percent level (column 3). Additionally, controlling for occupation shares leaves this coefficient essentially unchanged (column 4). Holding the baseline characteristics, industry composition, and occupation composition constant, each log point increase in mean population density from 8 to 10.4 (that is, each 2.7 multiplicative increase in mean population density from 3,000 to 33,000 persons per square mile) is associated with 0.45 percentage point slower predicted population growth per year. This implies that the New York City metropolitan area would have had 1.1 percentage point per year higher predicted population growth if it had had St. Louis' mean population density in 2000 of 3,000 persons per square mile (rather than 32,600). Correspondingly, New York City's predicted increase in population from 2000 to 2017 would have been 20 percentage points higher.

*Table 2*

## Partial Correlation of Population Growth with Mean Population Density

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Partial correlations of 2000–17 population growth | Own size only | Baseline | Baseline and industry composition | Baseline, industry composition, and occupation composition |
| Controls: | | | | |
| Metropolitan adjacency (2) | | x | x | x |
| Weather (10) | | x | x | x |
| Coast and river adjacency (7) | | x | x | x |
| Hilliness (2) | | x | x | x |
| Shale basin (6) | | x | x | x |
| Higher education (1) | | x | x | x |
| Industry composition (18) | | | x | x |
| Occupation composition (21) | | | | x |
| ln(dens) up to 3 (20 persons/square mile) | 0.07 | 0.14*** | 0.05 | 0.08** |
| *398 locations* | (0.05) | (0.04) | (0.05) | (0.04) |
| ln(dens) from 3 to 4 (55) | 0.29*** | 0.12 | 0.07 | 0.02 |
| *277 locations* | (0.10) | (0.09) | (0.08) | (0.07) |
| ln(dens) from 4 to 5 (148) | –0.05 | 0.03 | –0.03 | –0.02 |
| *398 locations* | (0.10) | (0.08) | (0.07) | (0.06) |
| ln(dens) from 5 to 6 (403) | 0.21** | 0.12 | 0.06 | 0.06 |
| *383 locations* | (0.10) | (0.08) | (0.08) | (0.07) |
| ln(dens) from 6 to 7 (1,100) | 0.14 | 0.15 | 0.08 | 0.06 |
| *427 locations* | (0.11) | (0.10) | (0.08) | (0.08) |
| ln(dens) from 7 to 8 (3,000) | 0.60*** | 0.39*** | 0.06 | –0.02 |
| *299 locations* | (0.15) | (0.13) | (0.13) | (0.12) |
| ln(dens) from 8 to 10.4 (33,000) | –0.11 | –0.20 | –0.44*** | –0.45*** |
| *56 locations* | (0.14) | (0.15) | (0.15) | (0.16) |
| Observations | 2,258 | 2,258 | 2,258 | 2,258 |
| $R^2$ | 0.16 | 0.37 | 0.49 | 0.51 |
| Adjusted $R^2$ | 0.15 | 0.36 | 0.48 | 0.50 |
| Control variables | | 28 | 46 | 67 |
| $R^2$, control variables | | 0.30 | 0.48 | 0.51 |

** Significant at the 5 percent level
*** Significant at the 1 percent level

Notes: Dependent variable is average annual population growth (percent) from 2000 to 2017. Regressions also include a constant. Standard errors are in parentheses and adjust for spatial correlation based on Conley (1999). Italicized text reports the number of locations with mean population density that lies within each spline segment. Results for all variables included in the baseline regression are reported in appendix Table A-2. Replication code is available in an online data supplement.

Among medium and large metropolitan areas, the negative correlation between growth and density is especially strong when measured using each metro's 95th percentile density. At least 95 percent of a metro's residents live in a tract with raw density at or below its 95th percentile density and at least 5 percent live in a tract with raw density at or above its 95th percentile density. The negative relationship between growth and 95th percentile density stands out in a scatter plot of the former against the latter (Chart 7). The best-fit linear relationship between the two, shown by the black line, accounts for 12 percent of the variation in growth (as measured by an $R^2$ statistic; see column 1 of Table 3).[22]

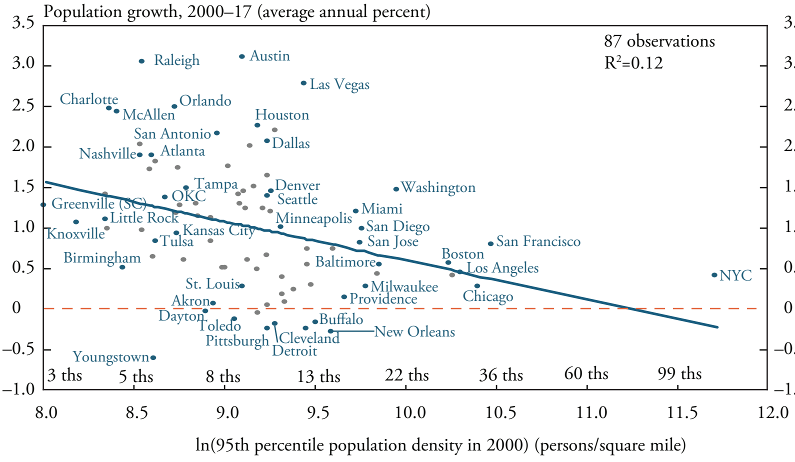
Although uncorrelated with population on its own, growth across medium and large metropolitan areas is strongly positively correlated with population after controlling for 95th percentile density, illustrating the partly offsetting benefits and costs of size (columns 2 and 3 of Table 3). More specifically, the positive coefficient on population in column 3 likely captures a gross increase in agglomerative benefits, while the negative coefficient on density—which is considerably larger in magnitude than when not controlling for population (column 3 versus column 1)—likely captures a gross increase in agglomerative costs. On net, the gross increase in costs dominates the gross increase in benefits, which is reflected in the negative coefficient on density when not controlling for size (column 1).

Population density, while unlikely to benefit businesses' productivity, may contribute positively to metros' amenities. For example, high population density helps support nearby urban amenities such as pedestrian access to varied restaurants, cafes, bars, retailers, and performance venues. Consistent with this possibility, growth is positively correlated with spikes in population density, measured by the increase in log density from a metro's 95th percentile tract to its 99th percentile tract (column 4 of Table 3). This positive correlation may be closely related to the increased tendency of young professionals to live near metropolitan central business districts (Couture and Handbury 2017; Baum-Snow and Hartley 2018)

Changes in agglomerative benefits and costs, as captured by partial correlations with population and population density, account for a considerable share of the variation in population growth across medium and large metropolitan areas. Together, population and 95th percentile

*Chart 7*

## Growth versus 95th Percentile Population Density, Medium and Large Metropolitan Areas




Notes: Metropolitan areas are labeled with the name of their largest city. The blue line represents the best fit based on a linear regression. The orange dashed line corresponds to a growth rate of 0. The Denver and Boulder metropolitan areas are combined. Replication code is available in an online data supplement.
Sources: U.S. Census Bureau and author's calculations.

*Table 3*

## Partial Correlation of Population Growth with Population and Population Density

| Partial correlations of 2000–17 population growth | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| ln(pop) | | 0.00<br>(0.09) | 0.38***<br>(0.11) | | 0.34***<br>(0.10) |
| ln(95th percentile density) | –0.48***<br>(0.10) | | –0.82***<br>(0.14) | | –0.79***<br>(0.13) |
| ln(99th percentile)–ln(95th percentile) | | | | 1.13**<br>(0.47) | 1.07***<br>(0.36) |
| Observations<br>$R^2$ | 87<br>0.12 | 87<br>0.00 | 87<br>0.20 | 87<br>0.09 | 87<br>0.28 |

** Significant at the 5 percent level
*** Significant at the 1 percent level

Notes: Dependent variable is average annual population growth (percent) from 2000 to 2017. Regressions also include a constant. Standard errors are in parentheses and adjust for spatial correlation based on Conley (1999). Replication code is available in an online data supplement.

population density account for 20 percent of the variation in growth (column 3 of Table 3). Including the difference in log density between the 99th and 95th percentile boosts the share of variation accounted for to 28 percent (column 5 of Table 3). Including the 25th and 75th percentile densities further boosts the share of variation accounted for to 38 percent (not shown).[23] This ability to account for more than one-third of the variation in population growth suggests that the shifting contributions of size, as measured by both population and density, to productivity and amenities have been among the most important determinants of recent metropolitan population growth.

## IV. Conclusions

The population and employment of small and large locations in the United States have been diverging for several decades. For locations with a population in 2000 up to about 500,000, population growth from 2000 to 2017 was positively correlated with initial population. For locations within this group with a population up to about 160,000, size itself is likely to have driven the positive correlation, reflecting a net increase in agglomerative productivity and amenities over the past few decades.

In contrast, population growth from 2000 to 2017 was negatively correlated with mean population density at high levels, likely reflecting a net increase in agglomerative costs such as housing prices and traffic congestion over the past few decades. Similarly, growth across medium and large metropolitan areas was strongly negatively correlated with population density measured at the 95th percentile.

This pattern of local population growth—positively correlated with population across smaller locations and negatively correlated with population density across larger locations—is likely to persist for a considerable time, as net flows of households and jobs gradually move locations toward a spatial equilibrium.

Population and employment growth's dependence on size and density has some important public policy implications. First, small locations seeking to reverse declining population and employment face a formidable challenge, as they must offset the decreasing relative productivity and amenities attributable to their small size. Small locations that have succeeded in doing so have primarily relied on exogenous characteristics, such as natural resources, nice weather, natural recreational opportunities, the presence of a university, or adjacency to a large metropolitan area. For small locations that lack such offsetting characteristics, public policy may be more effective ameliorating the negative consequences of decline than reversing it.

Second, economic development strategies that attract new jobs may benefit existing local residents and businesses if the associated agglomerative benefits exceed the associated agglomerative costs. For example, the increase in local employment may sufficiently increase the productivity of existing businesses to allow them to pay higher wages. It may also attract households and workers who sufficiently increase the local tax base to offset any associated increases in public spending. In other cases, however, successfully attracting jobs may hurt existing residents and businesses. In particular, tax incentives narrowly targeted at one or a handful of businesses may lower government services and increase the tax burden for existing residents and businesses. Rather than narrowly targeting incentives, more effective public policy might focus on policies that broadly benefit local businesses and residents, both existing and new.

# Appendix

## Additional Tables

*Table A-1*

### Partial Correlation of Population Growth with Population and Baseline Controls (Regression Reported in Table 1, Column 2)

| Right-hand-side variable | Coefficient | Standard error | t-statistic | p-value |
|---|---|---|---|---|
| Initial population spline | | | | |
| ln(pop) from 5.9 to 8 (3,000) | 0.16 | 0.10 | 1.51 | 0.130 |
| ln(pop) from 8 to 9 (8,100) | 0.35*** | 0.07 | 5.28 | 0.000 |
| ln(pop) from 9 to 10 (22,000) | 0.27*** | 0.07 | 3.90 | 0.000 |
| ln(pop) from 10 to 11 (60,000) | 0.19*** | 0.06 | 2.90 | 0.004 |
| ln(pop) from 11 to 12 (163,000) | 0.50*** | 0.09 | 5.58 | 0.000 |
| ln(pop) from 12 to 13 (440,000) | 0.09 | 0.10 | 0.95 | 0.343 |
| ln(pop) from 13 to 14 (1.2 million) | 0.17 | 0.15 | 1.12 | 0.264 |
| ln(pop) from 14 to 16.7 (18.3 million) | –0.13 | 0.12 | –1.11 | 0.267 |
| Metropolitan adjacency (1/0 indicator) | | | | |
| Micro/non-core adjacent to metro with pop. > 1 million | 0.33*** | 0.07 | 4.59 | 0.000 |
| Micro/non-core adjacent to metro with pop. ≤ 1 million | 0.05 | 0.05 | 1.10 | 0.270 |
| Weather | | | | |
| Average max daily temp. in Jan. (linear) | 1.92E-02*** | 5.89E-03 | 3.25 | 0.001 |
| Average max daily temp. in Jan. (quadratic) | 7.16E-04*** | 2.16E-04 | 3.32 | 0.001 |
| Average max daily heat index in July (linear) | –3.81E-04 | 6.53E-03 | –0.06 | 0.954 |
| Average max daily heat index in July (quadratic) | –2.65E-04 | 3.63E-04 | –0.73 | 0.466 |
| Average mean daily relative humidity in July (linear) | –2.63E-02*** | 7.83E-03 | –3.36 | 0.001 |
| Average mean daily relative humidity in July (quadratic) | 6.26E-05 | 3.08E-04 | 0.20 | 0.839 |
| Average annual rainfall (linear) | –2.66E-03 | 8.19E-03 | –0.32 | 0.745 |
| Average annual rainfall (quadratic) | 3.15E-04*** | 1.16E-04 | 2.71 | 0.007 |
| Average annual number of days with rain (linear) | 6.60E-03* | 3.73E-03 | 1.77 | 0.077 |
| Average annual number of days with rain (quadratic) | –1.32E-04*** | 3.71E-05 | –3.57 | 0.000 |

*Table A-1 (continued)*

| Right-hand-side variable | Coefficient | Standard error | t-statistic | p-value |
|---|---|---|---|---|
| Coast and river adjacency (1/0 indicator) | | | | |
| Atlantic, Northeast census region | 0.19 | 0.15 | 1.29 | 0.196 |
| Atlantic, South Atlantic census division | 0.53*** | 0.15 | 3.58 | 0.000 |
| Gulf of Mexico | 0.12 | 0.17 | 0.71 | 0.478 |
| Pacific | -0.11 | 0.26 | –0.43 | 0.666 |
| Great Lakes | –0.32*** | 0.11 | –2.95 | 0.003 |
| Within 40 km of river on which nav. in 1968 | –0.19** | 0.08 | –2.44 | 0.015 |
| Within 40 km of major river | 0.02 | 0.05 | 0.31 | 0.755 |
| Hilliness | | | | |
| Ratio of std. dev. of altitude to land area (linear) | 1.43*** | 0.37 | 3.90 | 0.000 |
| Ratio of std. dev. of altitude to land area (quadratic) | –0.55*** | 0.13 | –4.36 | 0.000 |
| Shale oil basins (1/0 indicator) | | | | |
| Anadarko | 0.35*** | 0.13 | 2.76 | 0.006 |
| Bakken | 0.87*** | 0.20 | 4.32 | 0.000 |
| Eagle | 0.56*** | 0.19 | 3.00 | 0.003 |
| Haynesville | –0.13 | 0.11 | –1.13 | 0.259 |
| Niobrara | 0.24** | 0.11 | 2.27 | 0.023 |
| Permian | 0.08 | 0.15 | 0.50 | 0.617 |
| Higher education | | | | |
| Ratio of post-secondary students to pop. | 2.32*** | 0.37 | 6.30 | 0.000 |

* Significant at the 10 percent level
** Significant at the 5 percent level
*** Significant at the 1 percent level

Notes: Table reports estimation results for all variables included in the regression reported in column 2 of Table 1. Standard errors adjust for spatial correlation based on Conley (1999). The p-value is the probability that the absolute value of the t-statistic would exceed its regression value under the null hypothesis that population growth is uncorrelated with the corresponding right-hand-side variable. Linear weather coefficients estimate the partial derivative of growth with respect to each of the five weather measures for a location with the mean value of that measure. Replication code is available in an online data supplement.

*Table A-2*

## Partial Correlation of Population Growth with Mean Population Density and Baseline Controls (Regression Reported in Table 2, Column 2)

| Right-hand-side variable | Coefficient | Standard error | t-statistic | p-value |
|---|---|---|---|---|
| Initial population spline | | | | |
| ln(dens) up to 3 (20 persons/sq. mile) | 0.14*** | 0.04 | 4.01 | 0.000 |
| ln(dens) from 3 to 4 (55) | 0.12 | 0.09 | 1.44 | 0.149 |
| ln(dens) from 4 to 5 (148) | 0.03 | 0.08 | 0.33 | 0.740 |
| ln(dens) from 5 to 6 (403) | 0.12 | 0.08 | 1.42 | 0.155 |
| ln(dens) from 6 to 7 (1,100) | 0.15 | 0.10 | 1.60 | 0.110 |
| ln(dens) from 7 to 8 (3,000) | 0.39*** | 0.13 | 3.09 | 0.002 |
| ln(dens) from 8 to 10.4 (33,000) | −0.20 | 0.15 | −1.38 | 0.169 |
| Adjacency to metro area (1/0 indicator) | | | | |
| Micro/non-core adjacent to metro with pop. > 1 million | 0.30*** | 0.07 | 4.13 | 0.000 |
| Micro/non-core adjacent to metro with pop. ≤ 1 million | 0.01 | 0.04 | 0.11 | 0.911 |
| Weather | | | | |
| Average max daily temp in Jan. (linear) | 2.74e-02*** | 5.94e-03 | 4.62 | 0.000 |
| Average max daily temp in Jan. (quadratic) | 8.09e-04*** | 2.23e-04 | 3.63 | 0.000 |
| Average max daily heat index in July (linear) | −7.89e-03 | 6.81e-03 | −1.16 | 0.247 |
| Average max daily heat index in July (quadratic) | −4.52e-04 | 3.95e-04 | −1.15 | 0.252 |
| Average mean daily relative humidity in July (linear) | −2.59e-02*** | 8.45e-03 | −3.06 | 0.002 |
| Average mean daily relative humidity in July (quadratic) | 2.80e-04 | 3.19e-04 | 0.88 | 0.380 |
| Average annual rainfall (linear) | 7.32e-04 | 8.12e-03 | 0.09 | 0.928 |
| Average annual rainfall (quadratic) | 2.03e-04* | 1.21e-04 | 1.67 | 0.094 |
| Average annual number of days with rain (linear) | 7.92e-03** | 3.90e-03 | 2.03 | 0.043 |
| Average annual number of days with rain (quadratic) | −1.44e-04*** | 4.13e-05 | −3.49 | 0.000 |
| Coast and river adjacency (1/0 indicator) | | | | |
| Atlantic, Northeast census region | 0.28* | 0.15 | 1.90 | 0.058 |
| Atlantic, South Atlantic census division | 0.50*** | 0.19 | 2.63 | 0.009 |
| Gulf of Mexico | 0.17 | 0.20 | 0.85 | 0.394 |
| Pacific | 0.03 | 0.26 | 0.12 | 0.906 |
| Great Lakes | −0.26** | 0.10 | −2.46 | 0.014 |
| Within 40 km of river on which nav. in 1968 | −0.21** | 0.08 | −2.56 | 0.011 |
| Within 40 km of major river | 0.05 | 0.05 | 1.08 | 0.281 |

*Table A-2 (continued)*

| Right-hand-side variable | Coefficient | Standard error | t-statistic | p-value |
|---|---|---|---|---|
| Hilliness | | | | |
| Ratio of std. dev. of altitude to land area (linear) | 1.06*** | 0.35 | 2.98 | 0.003 |
| Ratio of std. dev. of altitude to land area (quadratic) | −0.43*** | 0.13 | −3.41 | 0.001 |
| Shale oil basins (1/0 indicator) | | | | |
| Anadarko | 0.35*** | 0.10 | 3.32 | 0.001 |
| Bakken | 0.85*** | 0.20 | 4.20 | 0.000 |
| Eagle | 0.70*** | 0.21 | 3.37 | 0.001 |
| Haynesville | −0.06 | 0.13 | −0.46 | 0.647 |
| Niobrara | 0.24** | 0.10 | 2.46 | 0.014 |
| Permian | −0.07 | 0.17 | −0.41 | 0.683 |
| Higher education | | | | |
| Ratio of post-secondary students to population | 2.70*** | 0.41 | 6.66 | 0.000 |

* Significant at the 10 percent level
** Significant at the 5 percent level
*** Significant at the 1 percent level

Notes: Table reports estimation results for all variables included in the regression reported in column 2 of Table 2. Standard errors adjust for spatial correlation based on Conley (1999). The p-value is the probability that the absolute value of the t-statistic would exceed its regression value under the null hypothesis that population growth is uncorrelated with the corresponding right-hand-side variable. Linear weather coefficients estimate the partial derivative of growth with respect to each of the five weather measures for a location with the mean value of that measure. Replication code is available in an online data supplement.

## Endnotes

[1]Holding delineations constant causes the calculated population growth rates of many fast-growing metropolitan areas to be lower than those calculated using population numbers published by the U.S. Census Bureau. The difference reflects that the official borders of metropolitan areas, delineated by the Office of Management and Budget, were redrawn after the 2010 decennial census. These border changes led to the inclusion of additional counties in many fast-growing metropolitan areas, reflecting the spread of suburbs into previously undeveloped land.

[2]Using a standard additive scale, the horizontal distance between the 1 log point increase from 3.3 million to 8.9 million (a 2.7 multiplicative increase) would be 3,000 times larger than the 1 log point increase from 1,100 to 3,000 (also a 2.7 multiplicative increase).

[3]I measure the average relationship by a linear Epanechnikov kernel with a bandwidth of 1.5 log points. Replication code is available in an online data supplement.

[4]The Palm Coast metropolitan area was merged into the Delton-Daytona Beach metropolitan area following the 2010 decennial census.

[5]Throughout the nineteenth and early twentieth centuries, population growth was negatively correlated with size across small locations. Beginning in the late nineteenth century, population growth was strongly positively correlated with size across medium and large locations. This steep positive relationship flattened in about 1960 (Desmet and Rappaport 2017).

[6]I measure employment in 2014 by values reported in the 2016 American Community Survey five-year summary file, which is based on households' responses to surveys from 2012 through 2016. I measure employment in 2000 by the number of individuals reporting they were employed the week prior to filling out their census questionnaire as disseminated in the 2000 decennial census summary files. Alternatively, measuring employment based on administrative data collected from firms, the positive relationship between growth and size is considerably weaker. This difference in the relationship between growth and the level of employment may reflect that a larger share of workers in small locations are self-employed or hold other positions for which administrative data, which are based on firms' payment of unemployment insurance taxes, are not collected.

[7]Productivity measures the efficiency with which firms transform labor and other inputs into a final output good or service. A location characteristic can be interpreted as increasing productivity if it allows businesses to pay higher wages for labor and higher prices for other inputs without hurting their profits. Low rates of taxes that fall on businesses can thus be interpreted as positively contributing to a location's productivity, reflecting that businesses care about after-tax profits. But tax incentives to lure a single business to a location typically leave the after-tax productivity of most existing businesses unaffected. A location characteristic can

be interpreted as increasing amenities if it makes households willing to pay higher house prices and accept lower wages. Low rates of taxes that fall on individuals can be interpreted as positive amenities, as they make individuals willing to accept lower pre-tax wages and pay higher housing prices compared to living elsewhere. However, low taxes may result in lower amenities in the form of public services.

[8]Many characteristics have both exogenous and endogenous components. For example, many seaports are protected by a constructed breakwater or require periodic dredging. For these ports, location along an ocean coast is clearly exogenous, while the breakwater and dredging are likely to be endogenous, based on judgments about the economic potential of the location. The industrial composition of locations similarly combines exogenous and endogenous components. In part, industrial composition depends on economic considerations from the distant past, which may no longer be relevant today. But industrial composition can also evolve over time in response to changing location productivity, amenities, and other economic circumstances.

[9]Congestion exemplifies a nonpecuniary cost, meaning that it does not take the form of an explicit monetary price. In contrast, higher land and house prices are pecuniary costs, which do not directly affect productivity and amenities.

[10]Of course, workers' skills differ as do their tastes for different consumption amenities. This is a second reason, in addition to variation in house prices, that population in a spatial equilibrium is distributed across many locations rather than clustered in a handful of locations with the highest productivity and amenities. Over the past few decades, workers have increasingly sorted into different metropolitan areas based on their skill type (Moretti 2012).

[11]This increase in agglomerative benefits can equivalently be thought of as the disadvantages of small size worsening during recent decades. Consistent with this, a measure of business dynamism has been declining in small locations relative to large locations (Brown 2018).

[12]The increase in the equilibrium population of larger locations relative to the equilibrium population of smaller locations (for locations with an initial population below 500,000) is consistent with the level of equilibrium population increasing for both types of locations. In this case, all that is required is that the proportional increase in the equilibrium population of the large locations exceed the proportional increase of the population of the small locations.

[13]This interpretation corresponds with a rise in the elasticity of productivity and amenities with respect to size for increases in population up to 500,000; an unchanged elasticity with respect to size for increases in population between 500,000 and 3 million; and a decline in the elasticity for increases in population above 3 million.

[14]Numerical results from a model of metropolitan size suggest that higher housing prices and traffic congestion contribute about equally to agglomerative costs (Rappaport 2016). If commuting speeds were to remain at their free-flow

level, metropolitan areas with the highest productivity would be an order-of-magnitude larger than they actually are.

[15]Research finds that building more highways significantly increases the number of commuters, leaving travel times mostly unchanged (Duranton and Turner 2011). Building highways can thus increase a location's equilibrium population, allowing migration from elsewhere until traffic congestion returns to its previous level.

[16]Consistent with the second interpretation, recent research suggests that the elasticity of agglomerative costs with respect to population is increasing (Combes, Duranton, and Gobillon 2015; Rappaport 2016). In other words, proportional increases in agglomerative costs due to an increase in location size are higher for larger locations. Thus, if the population of all locations proportionally increases by the same amount— for example, due to national population growth—then agglomerative costs will rise more for larger locations.

[17]The baseline controls include two indicator variables for metropolitan adjacency: the first variable takes a value of 1 for micropolitan areas and non-core counties adjacent to a metropolitan area with a population below 1 million and 0 otherwise, while the second takes a value of 1 for micropolitan and non-core counties adjacent to a metropolitan area with a population above 1 million and 0 otherwise. The 10 weather variables are linear and quadratic measures of winter temperature, the summer heat index, summer humidity, annual rainfall, and annual rainy days. The five coast variables are indicators taking the value of 1 if a location borders a coast along the Great Lakes, the Pacific Ocean, the Gulf of Mexico, the North Atlantic (Maryland north to Maine), and the South Atlantic (Virginia south to Florida). The two river variables are indicators that take a value 1 for locations that touch a major river and for locations that touch a river on which there was commercial navigation in 1968. The two hilliness variables are the linear and quadratic ratio of the standard deviation of altitude within a location, measured across 1.25-arc-minute grid cells, to the location's total land area. The six shale basin variables are indicators taking the value of 1 for locations in each of the Anadarko, Bakken, Eagle, Haynesville, Niobrara, and Permian basins. The presence of colleges and universities is measured by the share of a location's population enrolled in post-secondary classes. The appendix reports the results of the column 2 regression for all of these control variables. The variables themselves are included in the online data supplement.

[18]The respective p-values on the uppermost spline segment for the regressions reported in columns 3 and 4 are 0.116 and 0.103.

[19]The variation in log population accounts for 71 percent of the variation in the log of mean population density across all locations and 47 percent of the variation across medium and large metropolitan areas (as measured by $R^2$ statistics). Across medium and large metropolitan areas, log mean population density and log population account for 36 percent and 17 percent, respectively, of the variation in log median home price. Across all locations, however, log mean population

density accounts for a smaller share of the variation in log median home prices than does log population (25 percent versus 37 percent).

[20]The raw density of a metropolitan area is arithmetically equal to the land-weighted mean of the raw population density of each tract.

[21]The appendix reports the results of the column 2 regression for all control variables.

[22]The negative correlation between growth and density across medium and large metropolitan areas is much weaker for more standard benchmarks of density: mean population density can account for only 4 percent of the variation in growth, and median density cannot account for any of it. Instead, the negative correlation becomes meaningful at 75th percentile density, which accounts for 4 percent of the variation in growth. One possible explanation is that density measured at high percentiles reflects opportunities for apartment construction at sites near metropolitan centers (Rappaport 2017). The negative correlation between growth and density is also considerably weaker at the highest percentiles: measured at the 99th percentile, density accounts for only 5 percent of the variation in growth; measured at the maximum density within each metro, density accounts for only 2 percent of the variation in growth.

[23]A regression on the 25th, 75th, 95th, and 99th percentile densities in 2000 accounts for 37 percent of the variation in population growth from 2000 to 2017. An arithmetically equivalent regression has right-hand-side variables for the 25th percentile density as well as the increase in density from the 25th to 75th percentile, the increase from the 75th to the 95th percentile, and the increase from the 95th to the 99th percentile. The corresponding coefficients, each of which statistically differs from 0 at the 1 percent level, are negative on the first three variables and positive on the last variable. The negative coefficients are increasing in magnitude, implying that a 1 log point increase in 25th percentile density is associated with a smaller decrease in predicted growth than a 1 log point larger increase in density from the 25th to 75th percentiles, which in turn is associated with a smaller decrease in predicted growth than a 1 log point larger increase in density from the 75th to the 95th percentiles. Additionally, including log population boosts the share of variation accounted for to 38 percent, but the associated coefficient on log population does not statistically differ from 0. The online data supplement includes metropolitan density measured at numerous percentile benchmarks ranging from the 1st to the 99th.

## References


Baum-Snow, Nathaniel, and Daniel Hartley. 2018. "Accounting for Central Neighborhood Change, 1980–2010." Working paper, July.

Blomquist, Glenn C., Mark C. Berger, and John P. Hoehn. 1988. "New Estimates of Quality of Life in Urban Areas." *American Economic Review*, vol. 96, pp. 89–107.

Brown, Jason P. 2018. "The Widening Divide in Business Turnover between Large and Small Urban Areas." *Economic Review.* Federal Reserve Bank of Kansas City, Third Quarter, pp. 5–25.

Buzard, Kristy, Gerald A. Carlino, Robert M. Hunt, Jake K. Carr, and Tony E. Smith. 2017. "Localized Knowledge Spillovers: Evidence from the Spatial Clustering of R&D Labs and Patent Citations." Federal Reserve Bank of Philadelphia, Working Paper no. 17-32. Available at https://www.philadelphiafed.org/-/media/research-and-data/publications/working-papers/2017/wp17-32.pdf

Carlino, Gerald, and William R. Kerr. 2015. "Agglomeration and Innovation." In Gilles Duranton, J. Vernon Henderson, and William Strange, eds., *Handbook of Regional and Urban Economics,* vol. 5A, pp. 349–404. Amsterdam: Elsevier. Available at https://doi.org/10.1016/B978-0-444-59517-1.00006-4

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2018. "The Costs of Agglomeration: House and Land Prices in French Cities." Working Paper. Available at http://real.wharton.upenn.edu/~duranton/research.html

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In Gilles Duranton, J. Vernon Henderson, and William Strange, eds., *Handbook of Regional and Urban Economics,* vol. 5A, pp. 247–348. Amsterdam: Elsevier. Available at https://doi.org/10.1016/B978-0-444-59517-1.00005-2

Conley, Timothy G. 1999. "GMM Estimation with Cross-Sectional Dependence." *Journal of Econometrics*, vol. 91, no. 1, pp. 1–45. Available at https://doi.org/10.1016/S0304-4076(98)00084-0

Costa, Dora L., and Matthew E. Kahn. 2000. "Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990. *The Quarterly Journal of Economics*, vol. 115, no. 4, pp. 1287–1315. Available at https://doi.org/10.1162/003355300555079

Couture, Victor, and Jessie Handbury. 2017. "Urban Revival in America, 2000 to 2010." NBER Working Paper no. 24084, November. Available at https://doi.org/10.3386/w24084

Davis, Morris A., Jonas D.M. Fisher, and Marcelo Veracierto. 2013. "Gross Migration, Housing and Urban Population Dynamics." Federal Reserve Bank of Chicago, Working Paper no. 2013-19. Available at https://www.chicagofed.org/publications/working-papers/2013/wp-19

Desmet, Klaus, and Jordan Rappaport. 2017. "The Settlement of the United States, 1800–2000: The Long Transition towards Girbrat's Law." *Journal of Urban Economics*, vol. 98, March, pp. 50–68. Available at https://doi.org/10.1016/j.jue.2015.03.004

Diamond, Rebecca. 2016. “The Determinants and Welfare Implications of U.S. Workers’ Diverging Location Choices by Skill: 1980-2000.” *American Economic Review*, vol. 106, no. 3, pp. 479–524. Available at https://doi.org/10.1257/aer.20131706

Duranton, Gilles, and Diego Puga. 2004. “Micro-Foundations of Urban Agglomeration Economies.” In J. Vernon Henderson and Jacques-François Thisse, eds., *Handbook of Regional and Urban Economics,* vol. 4, pp. 2063–2117. Amsterdam: Elsevier. Available at https://doi.org/10.1016/s1574-0080(04)80005-1

Duranton, Gilles, and Matthew A. Turner. 2011. “The Fundamental Law of Road Congestion: Evidence from American Cities.” *American Economic Review*, vol. 101, no. 6, pp. 2616–2652. Available at https://doi.org/10.1257/aer.101.6.2616

Glaeser, Edward L., and Joseph Gyourko. 2005. “Urban Decline and Durable Housing.” *Journal of Political Economy*, vol. 113, no. 2, pp. 345–375. Available at https://doi.org/10.1086/427465

Glaeser, Edward L., and Matthew E. Kahn. 2004. “Sprawl and Urban Growth.” In J. Vernon Henderson and Jacques-François Thisse, eds., *Handbook of Regional and Urban Economics*, vol. 4, pp. 2481–2527. Amsterdam: Elsevier. Available at https://doi.org/10.1016/S1574-0080(04)80013-0

Glaeser, Edward L., Jed Kolko, and Albert Saiz. 2001. “Consumer City.” *Journal of Economic Geography*, vol. 1, no. 1, pp. 27–50. Available at https://doi.org/10.1093/jeg/1.1.27

Kennan, John, and James R. Walker. 2011. “The Effect of Predicted Income on Individual Migration Decisions.” *Econometrica*, vol. 79, no. 1, pp. 211–251. Available at https://doi.org/10.3982/ECTA4657

Marshall, Alfred. 1890. *Principles of Economics.* London: Macmillan.

Moretti, Enrico. 2012. *The New Geography of Jobs.* Boston and New York: Houghton Mifflin Harcourt.

Rappaport, Jordan. 2017. “Crowdedness, Centralized Employment, and Multifamily Home Construction.” *Economic Review.* Federal Reserve Bank of Kansas City, vol. 102, no. 1, pp. 41–83.

______. 2016. “Productivity, Congested Commuting, and Metro Size.” Federal Reserve Bank of Kansas City, Research Working Paper 16-03. Available at https://doi.org/10.18651/RWP2016-03

______. 2012. “Why Does Unemployment Differ Persistently across Metro Areas?” *Economic Review.* Federal Reserve Bank of Kansas City, Second Quarter, pp. 1–31.

______. 2008a. “A Productivity Model of City Crowdedness.” *Journal of Urban Economics*, vol. 63, no. 2, pp. 715–722. Available at https://doi.org/10.1016/j.jue.2007.04.008

______. 2008b. “Consumption Amenities and City Population Density.” *Regional Science and Urban Economics,* vol. 38, no. 6, pp. 533–552. Available at https://doi.org/10.1016/j.regsciurbeco.2008.02.001

______. 2007. “Moving to Nice Weather.” *Regional Science and Urban Economics*, vol. 37, no. 3, pp. 375–398. Available at https://doi.org/10.1016/j.regsciurbeco.2006.11.004

______. 2004. "Why Are Population Flows So Persistent?" *Journal of Urban Economics*, vol. 56, no. 3, pp. 554–580. Available at https://doi.org/10.1016/j.jue.2004.07.002

Rappaport, Jordan, and Jeffrey D. Sachs. 2003. "The United States as a Coastal Nation." *Journal of Economic Growth*, vol. 8, no. 1, pp. 5–46. Available at https://doi.org/10.1023/A:1022870216673

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy,* vol. 90, no. 6, pp. 1257–1278. Available at https://doi.org/10.1086/261120

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." In Peter M. Miezkowski and Mahlon R. Straszheim, eds., *Current Issues in Urban Economics*. Baltimore: Johns Hopkins University Press.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy*, vol. 64, no. 5, pp. 416–424. Available at https://doi.org/10.1086/257839