

Economic Quarterly—Volume 93, Number 4—Fall 2007—Pages 341–360

The Evolution of City Population Density in the United States

Kevin A. Bryan, Brian D. Minton, and Pierre-Daniel G. Sarte

The answers to important questions in urban economics depend on the density of population, not the size of population. In particular, positive production or residential externalities, as well as negative externalities such as congestion, are typically modeled as a function of density (Chatterjee and Carlino 2001, Lucas and Rossi-Hansberg 2002). The speed with which new knowledge and production techniques propagate, the gain in property values from the construction of urban public works, and the level of labor productivity are all affected by density (Carlino, Chatterjee, and Hunt 2006, Ciccone and Hall 1996). Nonetheless, properties of the distribution of urban population size have been studied far more than properties of the urban density distribution.

Chatterjee and Carlino (2001) offer an insightful example as to why density can be more important than population size. They note that though Nebraska and San Francisco have the same population, urban interactions occur far less frequently in Nebraska because of its much larger area. Though the differences in the area of various cities are not quite so stark, there are meaningful heterogeneities in city densities. Given the importance of urban density, the stylized facts presented in the article ultimately require explanations such as those given for the evolution of city population.

This article makes two major contributions concerning urban density. First, we construct an electronic database containing land area, population, and urban density for every city with population greater than 25,000 in the

■ We wish to thank Kartik Athreya, Nashat Moin, Roy Webb, and especially Ned Prescott for their comments and suggestions. The views expressed in this article are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System. Data and replication files for this research can be found at http://www.richmondfed.org/research/research_economists/pierre-daniel_sarte.cfm. All errors are our own.

United States. Second, we document a number of stylized facts about the urban density distribution by constructing nonparametric estimates of the distribution of city densities over time and across regions.

We compile data for each decade from 1940 to 2000; by 2000, 1,507 cities meet the 25,000 threshold. In addition, we include those statistics for every “urbanized area” in the United States, decennially from 1950 to 2000. Though we also present data on Metropolitan Statistical Area (MSA) density evolution from 1950 to 1980, this definition of a city can be problematic for work with densities. A discussion of the inherent problems with using MSA data is found in Section 1. To the best of our knowledge, these data have not been previously collected in an electronic format.

Our findings document that the distribution of city densities in the United States has shifted leftward since 1940; that is, cities are becoming less dense. This shift is not confined to any particular decade. It is evident across regions, and it is driven both by new cities incorporating with lower densities, and by old cities adding land faster than they add population. The shift is seen among several different definitions of cities. A particularly surprising result is that “legal cities,” defined in this article as regions controlled by a local government, have greatly decreased in density during the period studied. That is, since 1940, local governments have been annexing territory fast enough to counteract the increase in urban population. Annexation is the only way that cities can simultaneously have increasing population, which is true of the vast majority of cities in our sample, and yet still have decreasing density.

This article is organized as follows. Section 1 describes how our database was constructed, and also discusses which definition of city is most appropriate in different contexts. Section 2 discusses our use of nonparametric techniques to estimate the distribution of urban density. Section 3 presents our results and discusses why cities might be decreasing in density. Section 4 concludes.

1. DATA

What is a city? There are at least three well-defined concepts of a city boundary in the United States that a researcher might use: the legal boundary of the city, the boundary of the built-up, urban region around a central city (an “urbanized area”), and the boundary of a census-defined Metropolitan Statistical Area (MSA). The legal boundary of a city is perhaps most relevant when investigating the area that state and local governments believe can be covered effectively with a single government. Legal boundaries also have the advantage of a consistent definition over the period studied; this is not completely true for urbanized areas, and even less true for MSAs. Urbanized areas parallel nicely with an economist’s mental image of an agglomeration, as they include the built-up suburban areas around a central city. MSAs, though commonly used in the population literature, offer a much vaguer in-

Table 1 Three Definitions of a City

Legal City	The region controlled by a local government or a similar unincorporated region (CDP). <i>Defined by local and state governments.</i>
Urbanized Area	A region incorporating a central city plus surrounding towns and cities meeting a density requirement. <i>Defined by the U.S. Census Bureau.</i>
MSA	A region incorporating a central city, the county containing that city, and surrounding counties meeting a requirement on the percentage of workers commuting to the center. <i>Defined by the U.S. Census Bureau.</i>

terpretation. Figure 1 displays the city, urbanized area, and MSA boundaries for Richmond, Virginia, and Las Vegas, Nevada, in the year 2000.

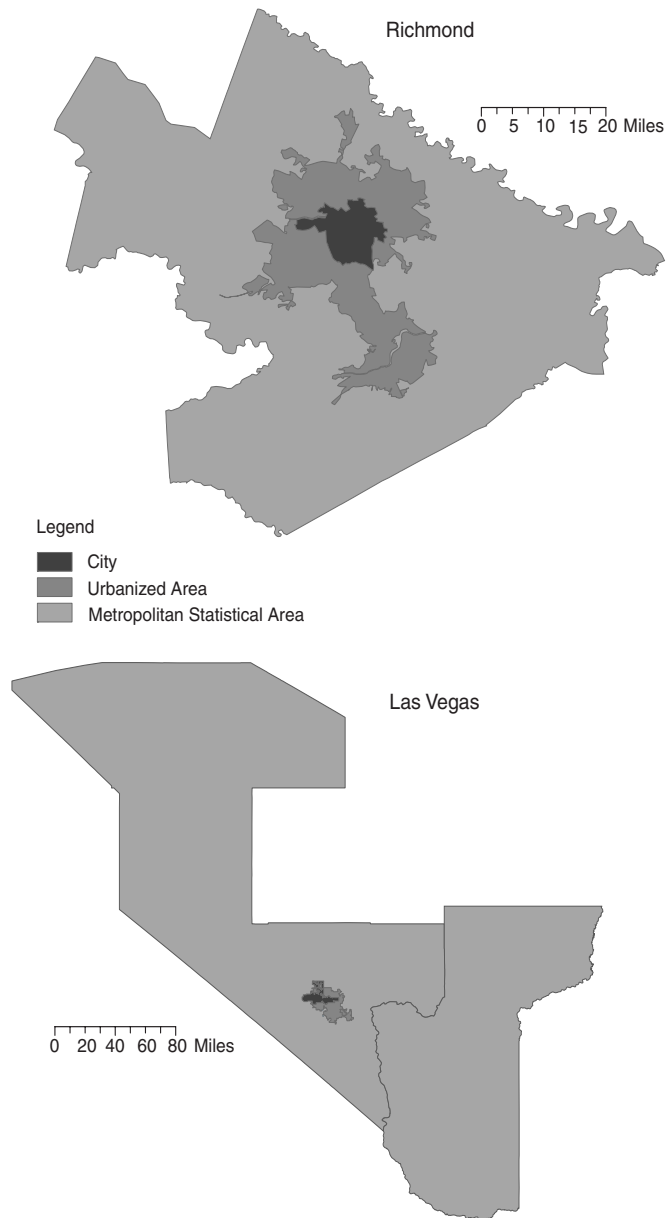
Our database of legal cities is constructed from the decennial U.S. Bureau of the Census *Number of Inhabitants*, which is published two to three years after each census is taken. Population and land area for every U.S. “place” with a population greater than 2,500 are listed. Places include cities, towns, villages, urban townships, and census-designated places (CDPs). Cities, towns, and townships are legally defined places containing some form of local government, while a census-designated place (called an “unincorporated place” before 1980) refers to unincorporated areas with a “settled concentration of population.” Some of these CDPs can be quite large; for instance, unincorporated Metairie, Louisiana has a population of nearly 150,000 in 2000. Though CDPs do not represent any legal entity, they are nonetheless defined in line with settlement patterns determined after census consultation with state and local officials, and are similar in size and density to incorporated cities.¹ Including CDPs in our database, and not simply incorporated cities, is particularly important as some states *only* have CDPs (such as Hawaii), and “towns” in eight states, including all of New England, are only counted as a place when they appear as a CDP.

From this list, we selected every place (including CDPs) with a population greater than 25,000 for each census from 1940 to 2000. There are 412 places in 1940 and 1,507 places in 2000 that meet this restriction. Each place was coded into one of nine geographical regions in line with the standard census region definition.² We also labeled each place as either “new” or “old.” An

¹ 1980 *Census of Population: Number of Inhabitants*. “Appendix A—Area Classification.” U.S. Department of Commerce, 1983. Note that CDPs did not appear in the 1940 Census.

² “Census Regions and Divisions of the United States.” Available online at http://www.census.gov/geo/www/us_regdiv.pdf.

Figure 1 A Graphic Representation of City Definitions



old place is a place that had a population greater than 25,000 in 1940 and still has a population greater than 25,000 in 2000. A new place is one that had a population less than 25,000 or did not exist at all in 1940, yet has a population greater than 25,000 in 2000. There are some places which had a population greater than 25,000 in 1940 but less than 25,000 in 2000 (for instance, a number of Rust Belt cities with declining populations); we considered these places neither new or old. Delineating places in this manner allows us to investigate whether the leftward shift of the distribution of U.S. cities was driven by newly founded cities having a larger area, or by old cities annexing area faster than their population increases.

In addition to legal cities, we also construct a series of urbanized areas from the *Number of Inhabitants* publication. Beginning in 1950, the U.S. Census defined urbanized areas as places with a population of 50,000 or more, meeting a minimum density requirement, plus an “urban fringe” consisting of places roughly contiguous with the central city meeting a small population requirement; as such, urbanized areas are defined in a similar way as agglomerations in many economic models. Aside from 1960, when the density requirement for central cities was lowered from approximately 2,000 people per square mile to 1,000 per square mile, changes in the definition of an urbanized area have been minor.³ Our database includes each urbanized area from 1950 to 2000; there were 157 such areas in 1950 and 452 in 2000.

Much of the literature on city population uses data on Metropolitan Statistical Areas (MSAs). An MSA is defined as a central urban city, the county containing that city, and outlying counties that meet certain requirements concerning population density and the number of residents who commute to the central city for work.⁴ We believe there are a number of reasons that this data can be problematic for investigating city density. First, it is difficult to get consistent data on metro areas. Before 1950, they were not defined at all, though Bogue (1953) constructed a series of MSA populations for 1900–1940 by adding up the population within the area of each MSA as defined in 1950. Because, by definition, Bogue holds MSA area constant for 1900–1950, this data set would not pick up any changes in density caused by the changing *area* of a city over time. Furthermore, there was a significant change in how MSAs are defined in 1983, with the addition of the “Consolidated Metropolitan Statistical Area” (CMSA). Because of this, MSAs between 1980 and 1990 are not comparable. Dobkins and Ioannides (2000) construct MSAs for 1990 using the 1980 definition, but no such series has been constructed for 2000.

Second, the delineation of MSAs is highly dependent on county definitions. Particularly in the West, counties are often much larger than in the

³ See the *Geographic Areas Reference Manual*, U.S. Bureau of the Census, chap. 12. Available online at: <http://www.census.gov/geo/www/garm.html>.

⁴ In New England, the town, rather than the county, is the relevant area.

Midwest and the East. For instance, in 1980, the Riverside-San Bernardino-Ontario, California MSA had an area of 27,279 square miles and a population density of 57 people per square mile.⁵ This MSA has an area three times the size of and a lower population density than Vermont.⁶ When looking solely at population, MSAs can still be useful because the population in outlying rural areas tends to be negligible; this is not the case with area, and therefore density.

Third, the number of MSAs is problematic in that it truncates the number of available cities such that only the far right-hand tail of the population distribution is included. For instance, Dobkins and Ioannides' (2000) MSA database includes only 162 cities in 1950, rising to 334 by 1990. For cities and census-designated places, three to four times as much data can be used. Eeckhout (2004) notes that the distribution of urban population size is completely different when using a full data set versus a truncated selection that includes only MSAs; it seems reasonable to believe that urban density might be similar in this regard. Further, nonparametric density estimation, as used in this article, requires a large data set. For completeness, we show in Section 3 that the distribution of densities in MSAs from 1950 to 1980, when the MSA definition was roughly consistent, follows a similar pattern to that of urbanized areas and legal cities.

Other than the database used in this article, we know of no other complete panel data set of urban density for U.S. cities. For 1990 and 2000, a full listing of places with area and population is available online as part of the U.S. Census Gazetteer.⁷ The County and City Data Books, hosted by the University of Virginia, Geospatial and Statistical Data Center, hold population and area data for 1930, 1940, 1950, 1960, and 1975; these data were entered by hand during the 1970s from the same census books we used.⁸ However, crosschecking this data with the actual census publications revealed a number of minor errors, and further indicated that unincorporated places and urban towns were not included. For some states (for instance, Connecticut and Maryland), this means that very few places were included in the data set at all. Our data set rectifies these omissions.

⁵ The MSA was made up of two counties: Riverside County with an area of 7,214 square miles, and San Bernardino County with an area of 20,064 square miles.

⁶ In fact, the entire planet has a land area of around 58 million square miles and a population of 6.5 billion, giving a density of 112 people per square mile, or twice the density of the Riverside MSA.

⁷ The 1990 data can be found at <http://www.census.gov/tiger/tms/gazetteer/places.txt>. Data for 2000 are available at: <http://www.census.gov/tiger/tms/gazetteer/places2k.txt>.

⁸ County and City Data Books. University of Virginia, Geospatial and Statistical Data Center. Available online at: <http://fisher.lib.virginia.edu/collections/stats/ccdb/>.

2. NONPARAMETRIC ESTIMATION

With these density data, we estimate changes in the probability density function (*pdf*) over time for each definition of a city in order to examine, for instance, how the distribution of urban densities is changing over time. We use nonparametric techniques, rather than parametric estimation, because nonparametric estimators make no underlying assumption about the distribution of the data (for instance, the presence or lack of normality). Assuming, for instance, an underlying normal distribution might mask evidence of a true bimodal distribution, and given our lack of priors concerning the distribution of urban densities, nonparametric estimates offer more flexibility. Potential pitfalls in nonparametric estimation are the requirement of larger data sets, and the computational difficulty of calculating *pdf* estimates with more than two or three variables;⁹ however, our data sets are large and our estimated *pdfs* are univariate. Nonparametric estimates of a *pdf* are closely related to the histogram; a description of this link, and basic nonparametric concepts, is given in Appendix A.

One frequently used nonparametric *pdf* estimator is the Rosenblatt-Parzen estimator,

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K(\psi_i),$$

where n is the number of observations, h is a “smoothing factor” to be chosen below, $\psi_i = \frac{x-x_i}{h}$, and K is a nonparametric kernel. The smoothing factor determines the interval of points around x which are used to compute $\hat{f}(x)$, and the kernel determines the manner in which an estimator weighs those points. For instance, a uniform kernel would weigh all points in the interval equally.

In practice, the choice of kernel is relatively unimportant. In this article, we use one of the more common kernels, namely the Gaussian kernel,

$$K(\psi_i) = (2\pi)^{-0.5} e^{-\frac{\psi_i^2}{2}}.$$

This kernel uses a weighted average of all observations, with weights declining in the distance of each observation from x_i .

The choice of bandwidth h , on the other hand, can be important, and is often chosen so as to minimize an error function of bias and variance. Given a set of assumptions about the nature of $f(x)$, the Rosenblatt-Parzen estimator

⁹ Nonparametric estimates converge to their true values at a rate slower than \sqrt{n} .

$\hat{f}(x)$ is such that¹⁰

$$\text{Bias} = \frac{h^2}{2} \left[\int \psi^2 K(\psi) d\psi \right] f''(x) + O(h^2) \quad (1)$$

and

$$\text{Variance} = \frac{1}{nh} f(x) \int K^2(\psi) d\psi + O\left(\frac{1}{nh}\right). \quad (2)$$

A low bandwidth, h , gives low bias but high variance, whereas a high h will give high bias but low variance. That is, choosing too small of a value for h will cause the estimated density to lack smoothness since not enough sample points will be used to calculate each $\hat{f}(x_i)$, whereas too high a value for h will smooth out even relevant bumps such as the trough in a bimodal distribution. A description of the assumptions necessary for our bias and variance formulas can be found in Appendix B.

The integrated mean squared error is defined as

$$\int [\text{Bias}(\hat{f}(x))^2 + \text{V}(\hat{f}(x))] dx. \quad (3)$$

This function simultaneously accounts for bias and variance. It is analogous to the conventional mean squared error in a parametric estimation. When h is chosen to minimize (3) after substituting for the bias and variance using expressions (1) and (2) respectively, we obtain

$$h = cn^{-\frac{1}{5}} \text{ where } c = \left[\frac{\int K^2(\psi) d\psi}{\left[\int \psi^2 K(\psi) d\psi \right]^2 \int (f''(x))^2 dx} \right]^{\frac{1}{5}}.$$

Since $f(x)$ is unknown, and the formula for h involves knowing the true $f''(x)$, no more can be said about h without making some assumptions about the nature of $f(x)$. For example, if $f(x) \sim N(\mu, \sigma^2)$, then $c = 1.06\hat{\sigma}$, and therefore $h = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ exactly.¹¹ This formula is called Silverman's Rule of Thumb, and works very well for data that is approximately normally distributed (Silverman 1986). Silverman notes that this rule does not necessarily work well for bimodal or heavily skewed data, and some of the series in this article (for instance, city populations) are heavily skewed. In particular, outliers lead to large increases in the estimated standard deviation, $\hat{\sigma}$, and therefore a very large value for h . Consequently, this article instead uses Silverman's more general specification

$$h = .9Bn^{-\frac{1}{5}}$$

¹⁰ If $\frac{X_n}{n^k} \rightarrow$ some real number c as $n \rightarrow \infty$, then X_n is $O(n^k)$. $O(A)$ is the largest order of magnitude of a sequence of real numbers X_n .

¹¹ Note that this rule does not imply that the nonparametric estimate will look like a parametric normal distribution; it merely says that, given data that are roughly normal, $1.06\hat{\sigma}n^{-\frac{1}{5}}$ is the smoothing factor that minimizes both bias and variance.

given

$$B = \min(\hat{\sigma}, \frac{IQR}{1.34}),$$

where IQR is the interquartile range of sample data. This formula is much less sensitive to outliers than the Rule of Thumb. In practice, this has shown to be nearly optimal for somewhat skewed data.

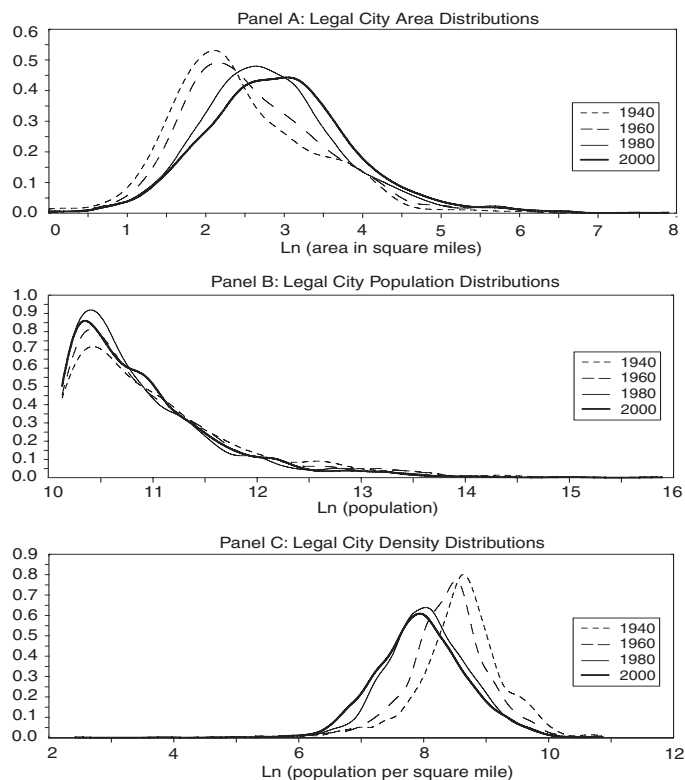
3. RESULTS

Using the kernel and smoothing parameter from the previous section, we can construct estimates of the *pdf* of the distribution of population, area, and urban density in each decade.

Figure 2 shows nonparametric estimations of the distributions of population size, area, and density for legal cities as defined in Section 1. Panel C shows a leftward shift of the distribution of city densities; that is, cities in 2000 are significantly less dense than in 1940. The mean population per square mile during that period fell from 6,742 to 3,802. This is being driven principally by an increase in the area of each city; mean area has increased from 19.2 square miles to 35.1 square miles between 1940 and 2000. The distribution of populations has remained relatively constant during this period.

One might imagine that this shift is being driven only by a subset of cities, such as rapidly-growing suburban and exurban cities, or cities in the West where land is less scarce. Hence, we divide cities into “new” and “old,” as defined in Section 1, as well as categorize each city into one of four regions: East, South, Midwest, and West. Figure 3 shows that the leftward shift in distribution is similar among both old and new cities; that is, city density is decreasing both because existing cities are annexing additional area, and because new cities have lower initial densities than in the past. The number of cities that change their legal boundaries in a given decade is surprising; for instance, between 1990 and 2000, nearly 36 percent of the cities in our data set added or lost at least one square mile. These changes vary enormously by state, however, in a state such as Massachusetts, where all of the land has been divided into towns for decades, there is very little opportunity for a city to add territory. Alternatively, in a state such as Oregon where the majority of land is unincorporated, annexation is much more common. Might it then be the case that the shift in city density is specific to the Midwest and West, where annexation is frequent?

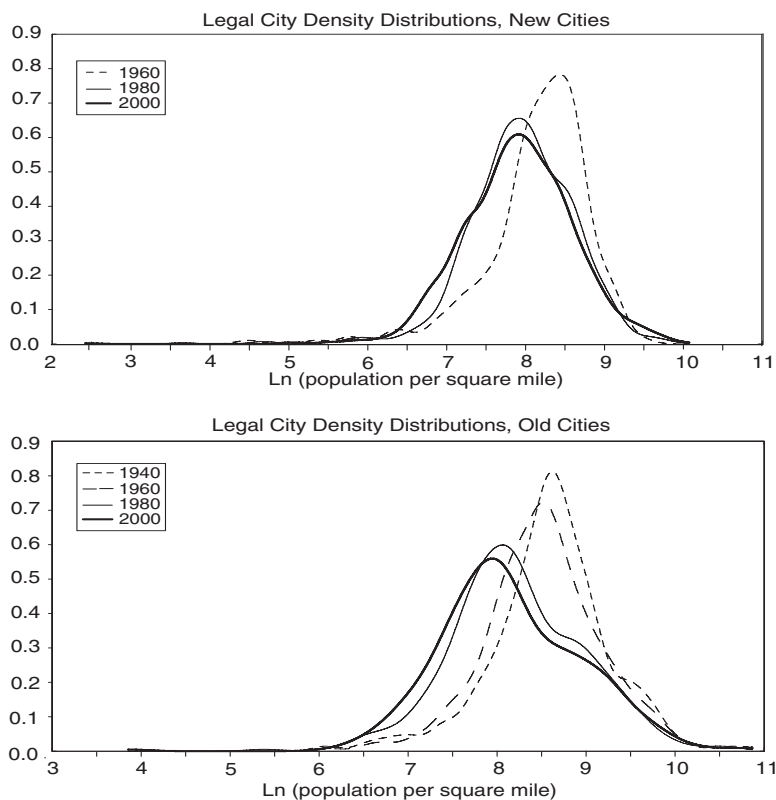
In fact, the leftward shift in city density does not appear to be a regional phenomenon. Figure 4 shows the distribution of densities in the East, South, Midwest, and West during the period 1940–2000. Each region showed a similar decline in density. The full distribution of log density from the Rosenblatt-Parzen estimator is particularly useful when examining the relatively small number of cities in each region when compared to a simple table of moments,

Figure 2 Legal City Area, Population, and Density

as extreme outliers in the data can result in high skewness. For instance, Juneau, Alaska, had an area of 2,716 square miles and a population of 30,711 in 2000, giving a density of approximately 11 people per square mile.

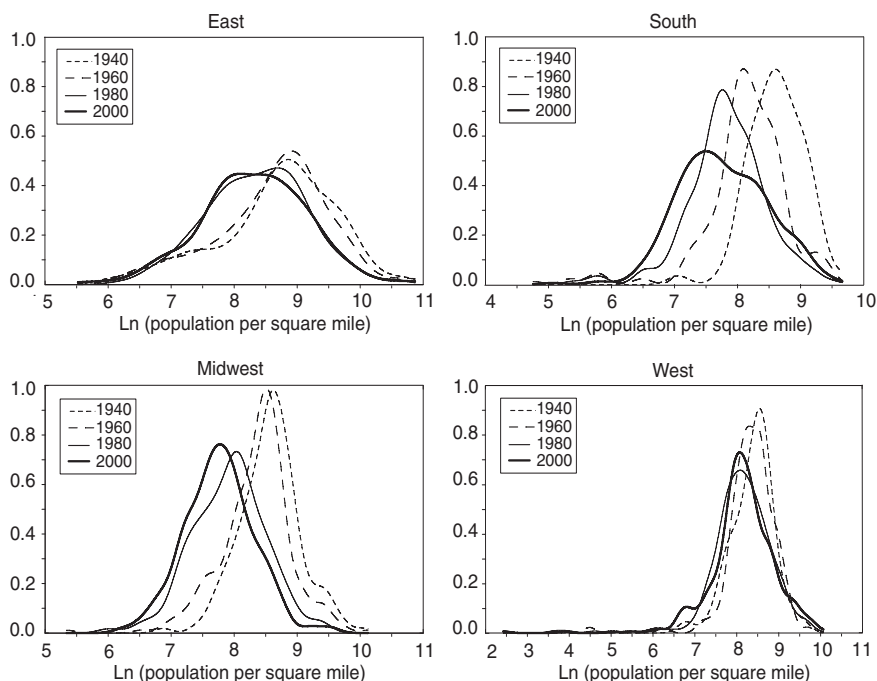
The trend in density is even clearer if we look at urbanized areas. Urbanized areas can be reasonably thought of as urban agglomerations; they represent the built-up area surrounding a central city. Figure 5 shows the estimated distribution of urbanized areas in 1960, 1980, and 2000. As in the case of legal cities, there has been a clear decrease in the density of urbanized areas during this period. Because the boundaries of urbanized areas and legal cities are quite different, it is rather striking that, under both definitions, the decrease in density has been so evident. That is, cities have not simply expanded into a mass of lower-density suburbs, but the individual cities and suburbs themselves have decreased in density, primarily by annexing land.

Finally, we consider the density of Metropolitan Statistical Areas. As noted in Section 1, there are only consistently defined MSA data available for the

Figure 3 Distributions of New and Old Cities

period 1950–1980. Furthermore, a decrease in the distribution of MSA density might simply reflect the increase in the number of MSAs in states with large counties, since each MSA by definition includes its own county. The urban economics literature concerning population size, however, often uses MSAs. Figure 6 shows that the distribution of MSA population density also appears to be shifting leftward in the same manner as legal cities and urbanized areas, but again, it is hazardous to give any interpretation to this shift. The definitional advantages and large data sample size for urbanized areas and legal cities potentially makes them preferable to MSAs for future work concerning urban density.

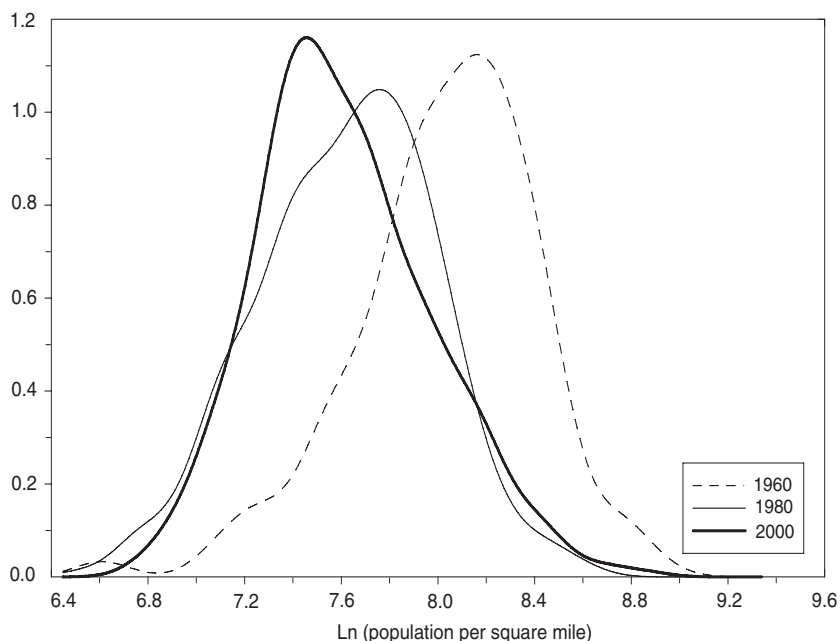
The importance of these shifts in urban density is underscored by the long-understood link between density and economic prosperity. Lucas (1988) cites approvingly Jane Jacobs' contention that dense cities, not simply cities, are the economic "nucleus of an atom," the central building block of development through their role in spurring human capital transfers. Ciccone and Hall

Figure 4 Distributions of Urban Density by Region

(1996), using county-level data, find that a doubling of employment density in a county increases labor productivity by 6 percent. In addition to knowledge transfer, agglomerations arise in order to facilitate effective matches between employer and employee and to take advantage of external economies of scale such as a common deepwater port.

Measuring the nature of local knowledge transfer, and in particular whether the relevant area has expanded as transportation and communication technologies have fallen, is difficult. Jaffe, Trajtenberg, and Henderson (1993) find evidence that, given the existing distribution of industries and research activity, new patents tend to cite existing patents from the same state and MSA at an unexpectedly high level. Using data on the urbanized portion of a metropolitan area, Carlino, Chatterjee, and Hunt (2006) find that patents per capita rise 20 percent as the employment density of a city doubles. They also find that the benefits of density are diminishing over density, so that cities with employment densities similar to Philadelphia and Baltimore, around 2,100 jobs per square mile, are optimal.

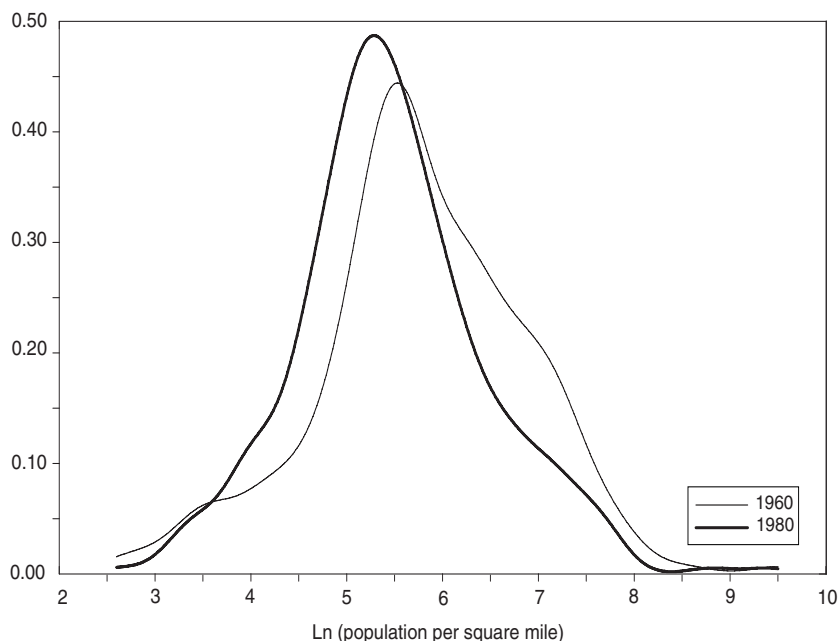
Given the economic benefits of density, the changes in the urban density distribution presented in this article suggest two questions. First, why

Figure 5 Distribution of Urbanized Areas

have agglomeration densities decreased? Second, why have the areas of legal jurisdictions increased?

Decreased densities in urban areas have been explained by a number of processes in the literature, including federal mortgage insurance, the Interstate Highway System, racial tension, and schooling considerations. Mieszkowski and Mills (1993) counter that these explanations tend to be both unique to the United States and are phenomena of the postwar period, whereas a decrease in urban density began as early as 1900 and has occurred across the developed world. Two theories remain.

First, the decreased transportation costs brought about by the automobile and the streetcar has allowed congestion in central cities to be avoided by firms and consumers. Glaeser and Kahn (2003) point out that the automobile also has a supply-side effect in that it allows factories and other places of work to decentralize by eliminating the economies of scale seen with barges and railroads; the rail industry was three times larger than trucking in 1947, but trucks now carry 86 percent of all commodities in the United States. Whereas the wealthy in the nineteenth century might have preferred to live in the center of a city while the poor were forced to walk from the outskirts, the modern

Figure 6 Distribution of MSAs

well-to-do are less constrained by transport times and, therefore, occupy land in less-dense suburban and exurban cities.

Rossi-Hansberg, Sarte, and Owens (2005) present a model in which firms set up non-integrated operations such that managers work in cities in order to take advantage of knowledge transfer externalities but production workers tend to work at the periphery of a city where land costs are lower. They then show that, as city population grows, the internal structure of cities changes along a number of dimensions that are consistent with the data.

A second theory, not entirely independent from the first, posits that cities have become less dense because of a desire for homogenization. When a large group with relatively homogenous preferences for tax rates and school quality is able to occupy its own jurisdiction, it can use land-use controls to segregate itself from potential residents with a different set of preferences. Mieszkowski and Mills (1993) argue that land-use restrictions have become more stringent in the postwar era, and that segregation into income-homogenous areas may be contributing to decreased densities.

There are fewer existent theories about why legal jurisdictions, at a given population level, have increased in area. Glaeser and Kahn (2003) note that effective land use requires larger jurisdictions as transportation costs fall. That

is, if a city wished to limit sprawl in an era with high transportation costs, it could enact effective land-use regulations within small city boundaries. In an era with low transportation costs, however, such a regulation would simply push residents into another bedroom community and have no effect on sprawl or traffic. The growing number of regional land-use planning commissions, such as Portland's Metropolitan Service District and Atlanta's Regional Commission, speak to this trend (Song and Knaap 2004).

Austin (1999) discusses reasons why cities may want to annex territory, including controlling development on the urban fringe, increasing the tax base, lowering the cost of municipal services, lowering municipal service costs by exploiting returns to scale, or altering the characteristics of the city, such as decreasing the minority proportion of population. External areas may wish to be annexed because of urban economies of scale, and because urban areas offer benefits such as cheaper bond issuance than suburban and unincorporated areas. Austin finds evidence that cities annex for both political and economic reasons, but that increasing the tax base does not appear to be a relevant factor, perhaps because of the growing ability of high-wealth areas to avoid annexation by poorer cities.

4. CONCLUDING REMARKS

This article provides two novel contributions. First, it constructs an electronic data set of urban densities in the United States during the previous seven decades for three different definitions of a city. Second, it applies non-parametric techniques to estimate the distribution of those densities, and finds that there has been a stark decrease in density during the period studied. This deconcentration has been occurring continuously since at least 1940, in every area of the United States, and among both new and old cities. This result is striking; increasing population and increasing area across cities do not, by themselves, tell us what will happen to density.

Falling urban densities suggest that, over the past seven decades, the productivity benefits of dense cities have been weakening. Decreasing costs of transportation and communication have allowed firms to move production workers out of high-rent areas, and have allowed residents to move away from downtowns. It is unclear what effect these changes in the urban landscape will have on knowledge accumulation and growth in the future. For instance, it is conceivable that the productivity loss from ever-decreasing spatial density might be counteracted by decreased long-range communication costs. Understanding the broad properties of urban density in modern economies is merely a necessary first step in understanding how these changing properties of cities will affect the broader economy.

APPENDIX A: NONPARAMETRIC ESTIMATORS

Classical density estimation assumes a parametric form for a data set and uses sample data to estimate those parameters. For instance, if an underlying process is assumed to generate normal data, the estimated density is

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where σ and μ are the sample standard deviation and mean.

Nonparametric density estimation, on the other hand, allows a researcher to estimate a complete density function from sample data, and therefore estimate each moment of that data, without assuming any underlying functional form. For instance, if a given distribution is bimodal, estimating moments under the assumption of normally distributed data will be misleading. Knowing the full distribution of data also makes clear what stylized facts need to be explained in theory; if the data were skewed heavily to the right and suffered from leptokurtosis, a theory explaining that data should be able to replicate these properties. Nonparametric estimation generally requires a larger data set than parametric estimation to achieve consistency, but is becoming more common in the literature. Given that our city data set is large, we use nonparametric techniques in this article. A brief introduction to these techniques can be found in Greene (2003), while a more complete treatment is found in Pagan and Ullah (1999).

At its core, a nonparametric density estimate is simply a smoothed histogram. Therefore, the nonparametric estimator can be motivated by beginning with a histogram. In a histogram, the full range of n sample values is partitioned into non-overlapping bins of equal width h . Each bin has a height equal to the number of sample observations within the range of that bin divided by the total number of observations. Given an indicator function $\mathbf{I}(A)$, defined as equal to 1 if the statement A is true, and 0 if the statement A is false, the height of a bin centered at some point x_0 , with width h , is

$$H(x_0) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(x_0 - \frac{h}{2} < x_i \leq x_0 + \frac{h}{2}).$$

That is, we are simply counting the number of sample observations in each bin of width h , and dividing that frequency by the sample size; the resulting height of each bin is the relative frequency. If there are 40 observations, of which 10 are in the bin $(1,2]$, with $h = 1$, then the histogram has height $H(1.5) = .25$ for all x in $(1,2]$.

This concept can be extended by computing a “local” histogram for each point x in the range $(x_{\min} - \frac{h}{2}, x_{\max} + \frac{h}{2}]$, where x_{\min} and x_{\max} are the minimum

and maximum values in the sample data.¹² In the histogram above, we computed $H(x_0)$ for only h points in the range; x_0 was required to be the midpoint of a bin. The local histogram will instead calculate $\hat{f}(x)$ for every x in $(x_{\min} - \frac{h}{2}, x_{\max} + \frac{h}{2})$, where $\hat{f}(x)$ evaluated at a given point x_0 is equal to the number of sample observations within $(x_0 - \frac{h}{2}, x_0 + \frac{h}{2})$, divided by n to give a frequency.¹³ That is,

$$\begin{aligned} \hat{f}(x) &= \frac{1}{n} \sum_{i=1}^n \mathbf{I}(x - \frac{h}{2} < x_i < x + \frac{h}{2}) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{I}(|\frac{x - x_i}{h}| < \frac{1}{2}) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{I}(|\psi(x_i)| < \frac{1}{2}), \end{aligned}$$

where $\psi(x_i) = \frac{x-x_i}{h}$. $\hat{f}(x)$ is a proper density function if, first, it is greater than or equal to zero for all x , which is guaranteed since the indicator function is always either 0 or 1, and second, if $\int_{-\infty}^{\infty} \hat{f}(x)dx = 1$. Dividing $\hat{f}(x)$ by h ensures that the function integrates to one. To see this, observe first that

$$\int_{-\infty}^{\infty} \mathbf{I}(|\psi(x_i)| < \frac{1}{2})d\psi = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{I}(|\psi(x_i)| < \frac{1}{2})d\psi = \int_{-\frac{1}{2}}^{\frac{1}{2}} d\psi = 1.$$

In addition, since $\psi(x_i) = \frac{x-x_i}{h}$,

$$\begin{aligned} \frac{1}{h} \int_{-\infty}^{\infty} \hat{f}(x)dx &= \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbf{I}(|\frac{x - x_i}{h}| < \frac{1}{2})dx \\ &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbf{I}(|\frac{x - x_i}{h}| < \frac{1}{2})d\psi \\ &= 1. \end{aligned}$$

While local histograms certainly provide a nonparametric estimate of density, and are smoother than proper histograms, they are still discontinuous. It seems sensible, then, to attempt to smooth the histogram. This is done by

¹²The local histogram $\hat{f}(x)$ must be computed for $(x_{\min} - \frac{h}{2}, x_{\max} + \frac{h}{2}]$ and not simply for $(x_{\min}, x_{\max}]$, because $\hat{f}(x) > 0$ for points outside of $(x_{\min}, x_{\max}]$. For instance, if $h = 1$ and $(x_{\min}, x_{\max}] = (0, 10]$, $\hat{f}(10.4)$ will be greater than zero because it will count the sample observation $x_0 = 10$.

¹³In practice, $\hat{f}(x)$ can only be computed for a finite number of points. The distributions we display in Section 5 have been computed at 1,000 points evenly divided on the range (x_{\min}, x_{\max}) .

replacing the indicator function in

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \mathbf{I}\left(\left|\frac{x - x_i}{h}\right| < \frac{1}{2}\right)$$

with another function called a kernel, $K(\psi)$, such that $\hat{f}(x) \geq 0$, integrates to one and is smooth. An estimator of the form

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K(\psi_i), \text{ where } \psi_i = \frac{x - x_i}{h},$$

is a Rosenblatt-Parzen kernel estimator, and the resulting function $\hat{f}(x)$ depends on the choice of h , called a bandwidth or smoothing parameter, and the choice of kernel. A “good” density estimate will have low bias (that is, $E(\hat{f}(x)) - f(x)$, where $f(x)$ is the true density of the data) and low variance.

APPENDIX B: ROSENBLATT-PARZEN BIAS AND VARIANCE

Bias and variance of a nonparametric estimator can be calculated given the following four assumptions:

- 1) The sample observations are i.i.d.
- 2) The kernel is symmetric around zero and satisfies $\int_{-\infty}^{\infty} K(\psi)d\psi = 1$, $\int_{-\infty}^{\infty} \psi^2 K(\psi)d\psi \neq 0$, and $\int_{-\infty}^{\infty} K^2(\psi)d\psi < \infty$.
- 3) The second-order derivatives of \hat{f} are continuous and bounded around x , and
- 4) $h \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$.

It can be shown that the Rosenblatt-Parzen estimator $\hat{f}(x)$ has

$$\text{Bias} = \frac{h^2}{2} \left[\int \psi^2 K(\psi)d\psi \right] f''(x) + O(h^2)$$

and

$$\text{Variance} = \frac{1}{nh} f(x) \int K^2(\psi)d\psi + O\left(\frac{1}{nh}\right).$$

The integrated mean squared error (MISE) is defined as

$$\int [\text{Bias}(\hat{f}(x))^2 + \text{V}(\hat{f}(x))] dx.$$

Substituting the formulas for bias and variance, and ignoring the higher order terms, $O(h^2)$ and $O(\frac{1}{nh})$, respectively, gives the asymptotic integrated

mean squared error (AMISE):

$$\begin{aligned} & \frac{h^4}{4} \left[\int \psi^2 K(\psi) d\psi \right]^2 \int (f''(x))^2 dx + \frac{1}{nh} \int f(x) dx \int K^2(\psi) d\psi \\ &= \frac{h^4}{4} \left[\int \psi^2 K(\psi) d\psi \right]^2 \int (f''(x))^2 dx + \frac{1}{nh} \int K^2(\psi) d\psi. \end{aligned}$$

Differentiating with respect to h and setting the result equal to zero, we have

$$h^3 \left[\int \psi^2 K(\psi) d\psi \right]^2 \int (f''(x))^2 dx - \frac{1}{nh^2} \int K^2(\psi) d\psi = 0$$

or

$$h = cn^{-\frac{1}{5}}, \text{ where } c = \left[\frac{\int K^2(\psi) d\psi}{\left[\int \psi^2 K(\psi) d\psi \right]^2 \int (f''(x))^2 dx} \right]^{\frac{1}{5}}.$$

REFERENCES

- Austin, D. Andrew. 1999. "Politics vs. Economics: Evidence from Municipal Annexation." *Journal of Urban Economics* 45 (3): 501–32.
- Bogue, Donald J. 1953. *Population Growth in Standard Metropolitan Areas 1900–1950*. Oxford, Ohio: Scripps Foundation in Research in Population Problems.
- Carlino, Gerald, Satyajit Chatterjee, and Robert M. Hunt. 2006. "Urban Density and the Rate of Invention." Federal Reserve Bank of Philadelphia Working Paper No. 06-14.
- Chatterjee, Satyajit, and Gerald A. Carlino. 2001. "Aggregate Metropolitan Employment Growth and the Deconcentration of Metropolitan Employment." *Journal of Monetary Economics* 48 (3): 549–83.
- Ciccone, Antonio, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86 (1): 54–70.
- Dobkins, Linda, and Yannis Ioannides. 2000. "Dynamic Evolution of the Size Distribution of U.S. Cities." In *The Economics of Cities*, eds. J. Huriot and J. Thisse. New York, NY: Cambridge University Press.
- Eeckhout, Jan. 2004. "Gibrat's Law for (All) Cities." *American Economic Review* 94 (5): 1,429–51.
- Glaeser, Edward L., and Matthew E. Kahn. 2003. "Sprawl and Urban Growth." In *Handbook of Regional and Urban Economics*, eds. J. V.

- Henderson and J. F. Thisse, 1st ed., vol. 4, chap. 56. North Holland: Elsevier.
- Greene, William. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.
- Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108 (3): 577–98.
- Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (1): 3–42.
- Lucas, Robert E., Jr., and Esteban Rossi-Hansberg. 2002. "On the Internal Structure of Cities." *Econometrica* 70 (4): 1,445–76.
- Marshall, Alfred. 1920. *Principles of Economics*. 8th ed. London: Macmillan and Co., Ltd.
- Mieszkowski, Peter, and Edwin S. Mills. 1993. "The Causes of Metropolitan Suburbanization." *The Journal of Economic Perspectives* 7 (3): 135–47.
- Pagan, Adrian, and Aman Ullah. 1999. *Nonparametric Econometrics*. Cambridge, UK: Cambridge University Press.
- Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Raymond Owens III. 2005. "Firm Fragmentation and Urban Patterns." Federal Reserve Bank of Richmond Working Paper No. 05-03.
- Silverman, B. W. 1986. *Density Estimation*. London: Chapman and Hall.
- Song, Yan, and Gerritt-Jan Knaap. 2004. "Measuring Urban Form: Is Portland Winning the War on Sprawl?" *Journal of the American Planning Association* 70 (2): 210–25.
- U.S. Bureau of the Census. "Number of Inhabitants: United States Summary." Washington, DC: U.S. Government Printing Office 1941, 1952, 1961, 1971, and 1981.
- U.S. Bureau of the Census. 1994. *Geographic Areas Reference Manual*. Available online at <http://www.census.gov/geo/www/garm.html> (accessed September 4, 2007).