



WORKING PAPERS

RESEARCH DEPARTMENT

**WORKING PAPER NO. 13-1
REGIONAL RESILIENCE**

Jeffrey Lin
Federal Reserve Bank of Philadelphia

December 17, 2012

RESEARCH DEPARTMENT, FEDERAL RESERVE BANK OF PHILADELPHIA

Ten Independence Mall, Philadelphia, PA 19106-1574 • www.philadelphiafed.org/research-and-data/

Regional Resilience

Jeffrey Lin*

December 17, 2012

Abstract

In this paper, I study long-run population changes across U.S. metropolitan areas. First, I argue that changes over a long period of time in the geographic distribution of population can be informative about the so-called “resilience” of regions. Using the censuses of population from 1790 to 2010, I find that persistent declines, lasting two decades or more, are somewhat rare among metropolitan areas in U.S. history, though more common recently. Incorporating data on historical factors, I find that metropolitan areas that have experienced extended periods of weak population growth tend to be smaller in population, less industrially diverse, and less educated. These historical correlations inform the construction of a regional resilience index.

*Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106; jeff.lin@phil.frb.org. Thanks to Fuyuo Nagayama and Nick Reynolds for excellent research assistance, to Jeff Brinkman, Jerry Carlino, Paul Flora, Loretta Mester, Leonard Nakamura, Jason Novak, Elif Sen, and Mike Trebing for helpful comments, and to Jeremy Nowak for proposing this research. The views expressed here are my own and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. This paper is available free of charge at <http://www.philadelphiafed.org/research-and-data/publications/working-papers/>.

1 Introduction

There has been recent interest from policymakers and researchers in documenting and understanding regional “resilience” (e.g., Liu et al., 2011, and Weir et al., 2012). In this paper, my starting point is the idea that, in the long run, changes in regional populations may reflect changes in the underlying productive or amenity values of those locations. That is, whereas a decline in population over a short period of time may signal responses by households and firms to a temporary shock, a long-run decline strongly suggests a persistent reduction in the relative attractiveness of that region. Observed long-run changes in the distribution of population across regions, then, can be informative about the changing value of regions over time and a reasonable way to define the “resilience” of regions.

In Section 2, I discuss *spatial equilibrium*, a tool that urban economists use to understand the uneven geographic distribution of population and economic activity. In this framework, the distribution of population represents a balance between valuable amenities (e.g., rivers, houses, or product variety) and congestion costs (e.g., traffic, crime, or high land prices). Over the long run, *changes* in the relative population of regions may reflect changes in this balance between amenity benefits and costs from crowding. Thus, by examining long-term changes in the distribution of population across U.S. regions, we can better understand the changing relative values of these regions to households and firms. Put simply, if households and firms continue to choose to live in a region over a long period of time, that suggests that the value of regional amenities has remained resilient. In contrast, regions that are unable to successfully transition from older, obsolete amenities to newer, more valuable amenities will be more likely to suffer protracted, deep declines in population. This reasoning is the basis for my definition of regional resilience; I define resilient regions as those regions that manage to avoid episodes of persistent declines in population.

But first, in Section 3, I review recent annual employment data for U.S. metropolitan areas from 1991–2010. Conditional on using a population or employment data series to infer something about the resilience of regions, this exercise is useful for illustrating the weaknesses of an approach that relies on high-frequency changes versus a long-run approach. Notably, over

the most recent two decades, seemingly dissimilar regions such as the Detroit and San Francisco metropolitan areas exhibit similarly weak employment dynamics. I conclude that idiosyncratic shocks and short time horizons can make it difficult to draw inferences about any underlying changes in the relative value of locations. Instead, I suggest that a much longer time horizon and an emphasis on persistent declines would form a better approach.

I describe an approach that emphasizes long-run changes in population in section 4. First, I choose to examine U.S. metropolitan areas, which are groups of counties that share common labor, housing, and local product markets. Second, I collect county population data from every decennial census from 1790 to 2010. Third, because county boundaries change over time, I normalize these data to 2010 county boundaries. Fourth, I aggregate these data to regions based on the 2009 metropolitan and micropolitan area definitions used by the census. Finally, I calculate standardized population growth for each region and decade, relative to the average U.S. region’s growth in that period.

Using maps, I show that steep, persistent declines in population are relatively rare among regions in U.S. history, but these episodes have become more common since the mid-20th century. I define “persistent declines” as two consecutive decades of regional population growth that is (much) slower than the U.S. average. In section 5, using regressions, I show that these episodes of persistent relative decline are more likely to be experienced by regions that have (i) low population; (ii) low levels of industrial diversity; (iii) low levels of education; and (iv) high employment shares in traded industries. In addition, using census division indicators to control for fixed factors like geography and climate, I find that the Mountain and Pacific regions have been the most resilient in recent decades, while regions near the Great Lakes and the Middle Atlantic states have been the least resilient. Thus, historically, resilient regions—those that have avoided persistent relative declines in population—have been more populated, industrially diverse, highly educated, and service-based and have had access to natural amenities similar to those available in the western U.S. Finally, in section 6, I use these historical patterns to construct a resilience index based on contemporary regional characteristics.

2 Regional resilience and spatial equilibrium

A starting point for thinking about resilience is *spatial equilibrium*—a central way in which urban economists understand the geographic distribution of economic activity. In short, economists believe that, in the long run, people choose to live and work in places that have productive or amenity value. (This value may stem from a natural resource, such as a river; local institutions, such as well-defined property rights; or economies of density, such as those that make available the greater variety of goods and services in large cities. See Lin, 2012, for more discussion.) High-value places offer greater benefits; in exchange, people are willing to pay more for local goods, such as housing. Thus, economists tend to view the uneven geographic distribution of economic activity, at least in the long run, as a consequence of the balancing of two opposing forces: valuable amenities that attract firms and households to certain places versus the higher housing costs and other disamenities that result from increased crowding.

Similarly, we can understand *changes over time* in the geographic distribution of population. It follows that regions or cities that have grown in population relative to other cities have similarly increased in their relative productive or amenity value. In contrast, places that have experienced extended periods of relative decline in population seem likely to have also declined in the relative value that they offer residents and businesses. Over the long run, economists expect that changes in the relative population of regions reflect changes in the underlying balance of benefits from local amenities versus costs from increased crowding.

Of course, this is not to say that, for places that have maintained their relative level of economic activity, the *sources* of their productive or amenity value have remained unchanged. Over long stretches of history, given changes in technology and preferences and the depreciation of capital, it is unreasonable to expect that initial valuable factors, such as a harbor, continue to provide the same value. Rather, given a place with a constant or growing population, we expect that this place has successfully transitioned from its “first-nature” advantages (e.g., a harbor) to other, more modern advantages (e.g., a robust financial sector, highways and railroads, or skilled workers). (See Bleakley and Lin, 2012, for evidence of portage cities that

have successfully made this transition.) Put simply, if we observe that households and firms have continued to choose a region over a long period of time, that suggests that the value of regional amenities has remained resilient. In contrast, regions that are unable to successfully transition from older, obsolete amenities to newer, more valuable amenities will be more likely to suffer protracted, deep declines in population. This reasoning is the basis for my definition of regional resilience, which is explored further in Section 4.

A key consideration is the long run versus the short run. We expect that, in the short run, all regions are subject to idiosyncratic shocks, some of them negative. In observing these regions over a short time horizon, we will necessarily be uncertain about the sources of population movements. Is a yearly population decline related to changes in the underlying balance of local benefits and costs? Or is that decline because of a bad shock, say, storm damage or a business cycle trough, that will soon dissipate? Thus, while spatial equilibrium may be a useful tool in understanding the *long-run* geographic distribution of activities, it is less useful for describing differences across regions in the short run.

3 Using short-run employment changes to measure resilience

In this section, I review annual employment data for U.S. metropolitan areas from 1991–2010. Conditional on using a data series on population or employment to try to infer the underlying value of regions, this exercise is useful for illustrating the weaknesses of an approach that relies on high-frequency changes versus a long-run approach.

I collect employment and income data from the Quarterly Census of Employment and Wages for 87 of the largest metropolitan areas in the U.S. over 1991–2010, on an annual basis. Figure 1 shows the logarithm of annual changes in employment for eight of these metropolitan areas. Panel A groups four slow-growing metropolitan areas: the San Francisco Bay Area, Detroit, Philadelphia, and Tampa. On average, over this period these four metropolitan areas experienced annual employment growth of 0.2%, which lagged the U.S. metropolitan average of 1.1%. In contrast, Panel B shows

employment growth in some of the fastest-growing metropolitan areas in this period; annual employment growth among Houston, McAllen, Las Vegas, and Sacramento over this period averaged 2.6%. (Personal income growth was also slower in the four metropolitan areas in Panel A, at 4.3%, versus the U.S. average of 4.9% and the growth rate for the four metropolitan areas in Panel B, at 6.7%.)

These graphs demonstrate several peculiarities. A classification of metropolitan areas based on annual employment growth would tend to group together several seemingly dissimilar metropolitan areas into a “non-resilient” class. For example, San Francisco and Detroit are very different in terms of industrial structure, education, and amenities, yet display similar employment dynamics over the last twenty years. In contrast, the fastest-growing cities were those that experienced the housing boom most acutely, especially in the southwestern U.S.

A likely explanation for these peculiarities lies in the series of idiosyncratic shocks experienced over the past two decades: the housing boom in certain parts of the country, the dot-com boom and bust, and the continued stagnation of manufacturing employment meant that the slowest-growing metropolitan areas were high-technology coastal California metros and older, industrial cities in the Midwest and Northeast. Fast-growing metros tended to be in the South and interior West and often featured relatively low housing costs and low education levels.

However, note the difficulty in inferring from these correlations—between employment growth and characteristics such as industry mix, education, housing, and geography—an important causal link. First, patterns of employment growth over a short span of time may reflect the character of economic shocks, rather than the effect of regional factors on growth. Consider the following example. New Orleans experienced negative employment growth between 2005–2006. While that information might point to New Orleans’ poor “resilience,” a more cautious interpretation would point out that Hurricane Katrina and its effects were a very bad series of shocks. A better approach might take into account (i) regional shocks may be of varying intensity and (ii) the adjustment periods of years, if not decades, that households and firms need to fully respond to economic shocks.

A second weakness of a methodology that relies on high-frequency changes in population or employment is that many local factors—e.g., industry mix, education, housing, and geography—are very persistent, meaning that they change (if at all) very slowly. At any point in time, the geographic distribution of these factors likely reflects decisions that were made years, if not decades, before. In other words, annual variation in employment growth adds very little information because potential explanatory variables often do not change appreciably over two decades. Thus, even if there is some variation in the short-term employment dynamics of the metropolitan areas in Figure 1, Panel A, the fact that the average level of employment growth is similar across these regions means that regression analysis will assign common characteristics of these regions to low resilience.

4 Identifying episodes of persistent relative decline in U.S. history

In this section, I describe a strategy for identifying resilient regions that takes a longer-run perspective. Instead of annual data, as in the previous section, I use decennial data, drawn from U.S. censuses from 1790–2010.

By using a long panel, I hope to reduce uncertainty about the information contained in relative regional declines: over short time horizons, we cannot be certain if a negative result reflects the characteristics of a particular shock or characteristics of that region. A longer time horizon can reduce (though not eliminate) this uncertainty. Further, examining long-run changes in population can better account for long adjustment periods for households and firms to fully respond to economic shocks.

I define a region as “resilient” if it has avoided an episode of persistent, relative population decline over a long period of time. If we observe that a region has grown in population relative to other regions, that suggests that the value of regional amenities has remained resilient. In contrast, regions that are unable to successfully transition from older, obsolete amenities to newer, more valuable amenities will be more likely to suffer protracted, deep declines in population.

4.1 Data on historical population

I collect data on population for U.S. counties, for every decade from 1790 to 2010, from the National Historical Geographic Information System (NHGIS) database (Minnesota Population Center, 2011). These population data are in turn drawn from the decennial U.S. censuses of population. Thus, the choice of ten-year intervals is determined both by the availability of the underlying data, as well as a desire to focus on long-run changes in population. Population data are the best and most consistent information available in historical censuses; using population helps to maximize the sample size. Unfortunately, other types of data, such as employment, income, and industry mix, are only available sporadically, mostly in recent censuses.

A second main reason for using these census data is the availability of population information for small units of geography—in this case, counties—over a long period of time.

Then, because county boundaries change over time, I normalize these historical population counts to modern-day 2010 county boundaries. To do so, I compare maps of historical counties and modern counties, from NHGIS, to determine how boundaries have changed. Then, for historical counties that change their boundaries, I apportion their population across modern-day county boundaries according to land area. This results in panel data on population for normalized counties, over 23 decennial census years.

I then aggregate counties to present-day (2009) metropolitan and micropolitan area definitions. For counties that lay outside metropolitan and micropolitan areas, I group them by state remainders. Thus, for example, the Philadelphia metropolitan area contains 11 counties in Delaware, New Jersey, Maryland, and Pennsylvania, while Oklahoma counties that are not part of any metropolitan or micropolitan area are assigned to a “rural Oklahoma” remainder. I further limit my analysis to county groups and years that have at least a population of 50,000, in order to exclude extremely vast rural areas that do not necessarily represent common labor, housing, or product markets. Thus, I reduce the number of geographic units from 3,109 normalized counties to 468 county groups representing metropolitan areas, micropolitan areas, and more-populated rural regions. Note that these county groups cover a significant portion of the geographic extent of

the U.S., a much larger portion than is covered by standard metropolitan area definitions.

There is one more sample restriction based on date of county entry into the U.S., which I discuss later in this section.

For each region (group of counties) and year, I compute the logarithm of population growth. Because overall U.S. population growth varies over time, I do not necessarily want to compare, say, the population growth of Chicago in the 1820s to the population growth of Las Vegas in the 1990s. Further, our discussion of spatial equilibrium suggests that what we want is to compare contemporaneous population trends across cities and regions. For these reasons, I normalize population growth by year. Thus, for example, I compare the population growth of Chicago in the 1820s to the population growth of Philadelphia in the 1820s.

Figure 2 shows the raw data on relative population growth for 1820–1830 and 1930–1940, at the county (not regional) level. Each county is colored according to its relative population growth in that decade: Dark green counties were the fastest-growing counties; light green counties grew faster than the average county; orange counties grew slower than the average county; and red counties were the slowest-growing counties.

The most obvious feature in comparing these two maps is that of the expanding geographic footprint of the U.S. During the 1820s, large portions of the present-day U.S. were still politically unorganized or yet to be annexed. This can be seen in the large amount of missing data, shown as white space. In addition, the fastest-growing counties in the 1820s were those on the frontier, especially areas around the Great Lakes. In contrast, counties within the original 13 colonies experienced relatively balanced population growth, but at a rate slower than counties on the frontier.

By the 1930s, the entire present-day extent of the coterminous U.S. had been organized. The clearest pattern reflected in the map in Panel B of Figure 2 is the depopulation of the Great Plains associated with the Dust Bowl. This map also shows relatively slow growth in the Midwest compared with faster population growth in Florida, California, and Texas—the beginnings of larger migrations to the Sunbelt in subsequent decades.

Another feature evident in the maps is some degree of mean reversion

in population growth. Some counties experience fast population growth, only to experience at-pace or slow population growth in subsequent periods. An example of this is Chicago; large population increases in the mid-19th century, corresponding to its initial boom and the expansion of its railroad network, had dissipated by the early 20th century. One repeated pattern is that frontier areas tend to experience population booms in the initial decades after their entry into the U.S. (This pattern is consistent with the findings of Desmet and Rappaport, 2012, of population booms in “new” counties.) A concern that is relevant for the present study is that these booms often take a long time to dissipate after a county’s initial entry into the U.S. Because they are caused by entry into the system of regions, rather than adjustment among an existing group of regions, I exclude counties in the first 60 years of existence from my sample.

4.2 Identifying persistent relative decline using regional population growth

Next, I return to data on regional (that is, county group) population growth. Recall that there were at least two weaknesses in our initial analysis using annual employment: uncertainty about the size and nature of regional shocks, and uncertainty about the length of adjustment time required for households and firms to respond to these shocks. To address these weaknesses using the data on regional population growth, I adopt a very simple filter: I compare relative population growth for each region to population growth in the subsequent decade. For example, I compare relative population growth in Detroit from 1930–1940 to population growth in Detroit from 1940–1950. Only if regional growth is below average in both time periods do I consider the decline in the first period to be persistent.

More formally, define $\Delta P_{g,t} \equiv \ln P_{g,t+1} - \ln P_{g,t}$ as the log change in population for region g between census year t and $t + 1$. If $\Delta P_{g,t} < \mu_t + a\sigma_t$ and $\Delta P_{g,t+1} < \mu_{t+1} + a\sigma_{t+1}$, where μ_t is average regional population growth between t and $t + 1$, σ_t is the standard deviation of population growth, and a is some cutoff criteria, then I consider the decline in the first period t to be persistent. I try several values for a , including 0, -0.5 , and -1 . If $a = 0$, then the criterion is that growth is below average in two consecutive

decades. If $a = -1$, then the criterion is that growth is at least one standard deviation below average in consecutive decades. Note that this methodology requires two decades of data following the initial observation—so the last possible year for which I can identify persistent regional declines is 1990.

Conversely, I define *resilience* as regions that avoid these episodes of persistent decline. Thus, regions and years that do *not* experience two consecutive decades of below-average growth are assigned an indicator value of 1, and 0 otherwise. Table 1 shows the share of regions by year that pass these criteria, for three values of a : 0, -0.5 , and -1 . Thus, between 1930–1950, 54.5 percent of sample regions did not experience below-average growth in the 1930s and 1940s and were thus resilient. Similarly, using the lower value of a , 97.9 percent of regions between 1930–1950 did not experience growth that was slower than one standard deviation below average in the subsequent two decades. Note that, for $a = 1$, no region was non-resilient before 1920–1940.

4.3 Maps of persistent relative declines and increases in regional population

The series of maps of U.S. regions in Figures 3, 4, and 5 show these data on persistent regional growth. The maps show persistent relative increases in population, in light and dark green, in addition to persistent relative declines, in orange and red. (Note that these maps show data for all regions, including “new” regions and very rural regions that are later excluded from my analysis.)

In the maps, regions are colored yellow when they show moderate increases or decreases in population in two consecutive decades. More extreme increases or decreases (at least one-half of a standard deviation better or worse than average, in two consecutive decades) are shown in light green and orange, respectively.

These maps show that, for the first 140 years of U.S. history, persistent changes in relative population were rare and confined mostly to episodes of frontier booms. For example, Chicago from 1800–1830 experienced persistently faster population growth than the rest of the U.S., followed by Minneapolis from 1820–1840. Booms followed in Oklahoma in the late 19th

century and in Los Angeles and southern Florida starting around 1900. During this period, the U.S. overall experienced rapid population growth across nearly all regions, but most especially on the frontier. Even short, devastating losses across the U.S. South during the Civil War were reversed in the 1870s. Thus, these areas are largely colored yellow on the maps.

In fact, it is not until the Dust Bowl in rural 1930s Oklahoma that we see any large U.S. region experiencing persistently slow growth. This episode of persistently sluggish growth is followed by areas of the rural South in the mid-20th century, such as the Mississippi Delta, that experienced substantial African-American outmigration to northern cities.

In the late 20th century, the maps show continued relative declines in the rural Great Plains and the rural South. Starting in 1960, the maps show persistent declines in isolated, older industrial cities—such as Johnstown, Pennsylvania from 1960–1980; Youngstown, Ohio, starting around 1970; and western Pennsylvania and western New York state starting in 1990.

Table 2 lists regions, identified by their principal city and state, and years that have experienced two consecutive decades of growth slower than one standard deviation below average since 1930.

Importantly, while steep, persistent relative declines are rare in U.S. history, they seem to reflect actual changes in the productive or amenity value of certain regions and cities. During the Dust Bowl, there is evidence that rural areas of Oklahoma experienced real and persistent declines in the agricultural productivity of the soil. (Often, the productive layer of topsoil literally disappeared; for example, see Hornbeck, 2012, for evidence.) This correlation between persistent population declines and permanent declines in the amenity value of regions suggests that our method of identifying resilient regions is at least somewhat successful at differentiating the effects of idiosyncratic shocks on population from the effects of changes in the overall amenity value of regions.

In sum, persistent population changes among U.S. regions over 220 years are somewhat rare, but they have been more common in recent decades. An algorithm that is based on persistent changes in regional population seems to accurately classify regions in U.S. history that have experienced prolonged “busts”—that is, regions that have declined in amenity or productive value

in one decade and have been unable to return to average or above-average growth in the subsequent decade. The episodes of persistent decline identified in Figures 3, 4, and 5 seem to identify non-resilient regions.

5 Predicting persistent relative declines in U.S. history using historical factors

Next, I examine several historical factors that predict episodes of persistent decline in U.S. history. Although data availability varies by year, many historical censuses have information on factors that may predict regional resilience. In general, urban economic theory suggests that factors that are immobile, durable, or associated with strong positive spillovers may be associated with the resilience of regional amenities.

My strategy in this section is to regress an indicator of regional resilience on several historical factors. As in the previous section, I define the resilience indicator to be 1 for regions and years that do not experience two consecutive decades of below-average growth. I experiment with several thresholds for “below-average growth,” using one and one-half standard deviation slower than the average county. However, for the initial presentation, I focus on all regions that experienced two consecutive decades of below-average growth as an indicator of non-resilience. As the dependent variable is a binary outcome, I use a probit model, which restricts predicted values to be between 0 and 1. Based on data availability, I include several historical factors as explanatory variables in the probit regression model.

In particular, one factor that may be associated with positive spillovers and is available in every decade is population density. Urban economists have long postulated the existence of economies of density, or increasing returns to the geographic concentration of economic activity. In the analysis that follows, I include the logarithm of population and the logarithm of land area.

Economists have also noted the role of human capital in economies of density, through knowledge spillovers or demand linkages. We include measures of metropolitan area educational attainment—the shares of the population over 25 years of age that have completed at least some college in

modern censuses, or the share of the adult population that is literate in certain historical 19th century censuses.

I also control for the nine census divisions, to control for persistent natural amenities such as climate and geography.

When available, I compute employment shares for sectors that may be associated with large fixed factors or associated with some economies of density—for brevity, I call these “traded sectors,” and they include manufacturing, information, finance, wholesale trade, and medical and educational services. Finally, Jacobs (1969) hypothesized that industrial diversity was related to the production and adaptation of new ideas. I also calculate a measure of industrial diversity—an inverted Herfindahl-Hirschman index—constructed by summing the squared industry employment shares for each region and year.

5.1 Regression results

Table 3 displays results from four probit regressions of the resilience indicator on historical factors. The dependent resilience variable takes a value of 1 if a region did not experience two decades of below-average growth. Each column shows estimated marginal effects from a separate probit regression. Thus, from column (1), all else equal, a 10% increase in population is associated with a 1.27% increase in resilience, for the average county.

Column (1) includes only population density, that is, the logarithm of population and land area, as regressors, and thus uses all of the available data. Column (2) includes a measure of the region’s educational attainment, and therefore narrows the sample to more modern censuses.

Greater population predicts resilience. This result is robust to the inclusion of census division and year indicators, as well as to other covariates, as seen in later columns. These marginal effect estimates suggest that, for the average county, a 10 percent increase in population, holding area fixed, predicts an 0.8–1.3 percent increase in resilience.

Regions with educated populations also tend to be more resilient. Following the estimates reported in columns (3) and (5), for the average county, a 5-percentage-point increase (approximately 1 standard deviation among 1990 regions) in the college graduate share is associated with a 3–5 percent

increase in resilience.

Industrial diversity, as measured by an inverted Herfindahl-Hirschman index using industry employment shares, is also related to resilience. According to the estimated marginal effects in columns (4) and (5), a one-standard-deviation increase in industrial diversity predicts a 3–4 percent increase in resilience for the average county.

Finally, greater employment shares in “traded” industries predict less resilience. While these industries tend to be associated with large fixed factors or strong economies of density, suggesting that they may aid resilience, these industry features can also predict reversals, where agglomerations disperse and re-form somewhere else.

The pattern of the results is mostly unchanged if we use alternative definitions of resilience based on the one-half or one-standard-deviation thresholds. Table 4 displays regression results using these alternative resilience measures as the dependent variable. (These estimates are also robust to alternative estimation strategies. Results using a linear probability model are qualitatively similar to the reported probit results.)

6 Building a resilience index using historical patterns

In this section, I use the historical correlations among regional factors and resilience shown in Table 3 to construct a ranking of 2010 regions. I calculate the predicted values for each U.S. region based on 2010 values of population density, industry mix, and education and the regression coefficients estimated in Table 3, column (5). In effect, I am assuming that historical correlations among these regional factors and resilience are informative about the future resilience of regions. By construction of the probit estimator, these predicted values are between 0 and 1. The result of this procedure is that I form an out-of-sample prediction for the resilience of each U.S. region based on 2010 regional data.

Figure 6 shows a map of U.S. regions. Each point represents a U.S. region according to the longitude and latitude of its principal city and is colored according to its predicted value of resilience. Several things are

notable about this map. First, nearly half of all U.S. regions have resilience scores above 0.75, and three-quarters of the regions have resilience scores above 0.50. This reflects a historical pattern in that very few regions in the past have experienced persistent relative declines in population.

Second, the largest regions by population are labeled on the map. In general, all of these cities have high predicted resilience. This pattern reflects the positive association between past resilience and population density. Notably, several large cities that have experienced recent episodes of persistent decline—e.g., Pittsburgh and Detroit—have predicted resilience values in 2010 much greater than those in 1980 and 1990. In fact, looking at only predicted values based on the probit regression, Pittsburgh’s predicted resilience has risen to 0.51 in 2010, from 0.24 in 1970.

Table 5 lists predicted resilience values for the largest 2010 regions by population, as well as confidence intervals based on the standard errors of the predicted values. All of these regions are large metropolitan areas with populations of at least 1 million in 2010. Note that nearly all of them have predicted resilience scores very close to 1, and almost all of them have confidence intervals for the predictions that include 1. Only a handful of the largest metropolitan areas have poor predicted resilience scores.

Regions with poor predicted resilience tend to be smaller and located near the Great Lakes or in the Great Plains. Again, refer to the map in Figure 6. This pattern corresponds to findings from the probit regression; in recent decades, regions in these areas have been the most likely to experience continued declines in population. While we cannot be certain as to the cause of these declines, one hypothesis is that it is probably related in part to changes in preferences and technology that have led to greater migration flows from the Northeast and Midwest to the West and South. Thus, our regression model predicts greater resilience for those regions in the West and South and less resilience for regions in the Midwest.

7 Conclusions

In this paper, I describe a method for constructing a “regional resilience index.” I define resilient regions as those that avoid persistent declines for

population—growth that lags the U.S. average in two consecutive decades. I use population data on U.S. regions—metropolitan, micropolitan, and rural areas—over 220 years. These episodes of persistent decline are relatively rare in U.S. history but are more common since the mid-20th century. Using probit regressions and data on historical regional factors, I find that U.S. regions that are densely populated, highly educated, industrially diverse, and service-based are more likely to be resilient. Then, using estimates from the probit regressions, I compute predicted values of resilience for U.S. regions using information on regional factors from the 2010 census.

So far, I have intentionally said very little about welfare and the policy implications of this work. In part this is because the welfare implications of persistent regional decline are not clear. If the productive or amenity value of a region declines, households and firms may leave this region and move to other regions. But to the extent that the absolute productive or amenity value of other regions improves, households and firms may not be worse off and, under certain conditions, may even be better off. If policymakers care about the welfare of households, rather than the welfare of places and landowners, these long-run population dynamics may not be very meaningful for policy.

My analysis has also isolated particular regional factors—population density, industry mix, and education—that explain resilience but are slow to evolve over time and are relatively insensitive to policy. In fact, variation in these factors across regions often has roots decades or even centuries in the past. In addition, the short-run elasticities appear relatively small, meaning that large differences in these factors are associated with small differences in resilience. There are also factors such as geography and climate that are outside the reach of local policymakers. Finally, the relationships between historical factors and resilience that I use to predict resilience scores in Section 6 are not *causal* estimates. Instead, they reflect correlations in the data that may change in the wake of changes in technology, preferences, or other shifts in the economy.

This is not to say that the long-term effects of local amenities are always negligible. Bleakley and Lin (2012) show examples of small differences in initial value across regions that have large effects on population density and

income, a century or more after the sources of initial value became obsolete. Rather, the results that I have presented here can form a historical framework in which to better understand why regions that have superior factors are able to avoid persistent declines in population and value. For example, in Lin (2011), I show that regions that have educated populations are better able to adapt to new technologies. Carlino and Saiz (2008) similarly show that regions with greater consumption amenities are able to attract and retain educated workers. Taken together, these results suggest that local amenities may potentially have significant long-run effects.

In sum, the predicted resilience scores reflect historical relationships between regional factors and long-term, persistent relative declines in population. These regional factors do well in explaining past episodes of regional decline in U.S. history. The predictive ability of these scores therefore relies in part on stable relationships among these variables and may be more fragile in the event of significant changes to preferences or technologies. Finally, while persistent regional declines may induce costly adjustments for households and firms in the short run, the long-run welfare implications are less clear.

8 Works cited

Bleakley, Hoyt, and Jeffrey Lin (2012). “Portage and Path Dependence,” *Quarterly Journal of Economics* **127** (2); 587–644.

Carlino, Jerry, and Albert Saiz (2008). “City Beautiful,” Federal Reserve Bank of Philadelphia working paper 08-22.

Desmet, Klaus and Jordan Rappaport (2012). “The Settlement of the United States, 1800 to 2000: The Long Transition Towards Gibrat’s Law,” Federal Reserve Bank of Kansas City working paper.

Hornbeck, Richard (2012). “The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe,” *American Economic Review* **102** (4): 1477–1507.

Jacobs, Jane (1969). *The Economy of Cities*. New York: Random House.

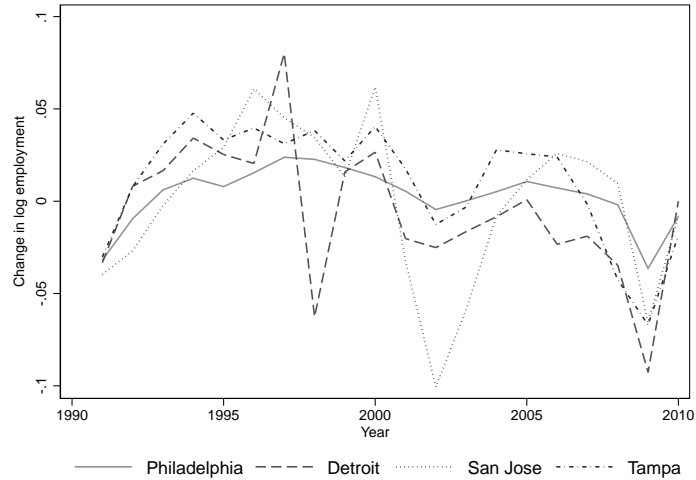
Lin, Jeffrey (2011). “Technological Adaptation, Cities, and New Work” (2011), *Review of Economics and Statistics* **93** (2): 554-574.

Lin, Jeffrey (2012). “Geography, History, Economies of Density, and the Location of Cities,” Federal Reserve Bank of Philadelphia *Business Review* **2012** (Q3): 18–24.

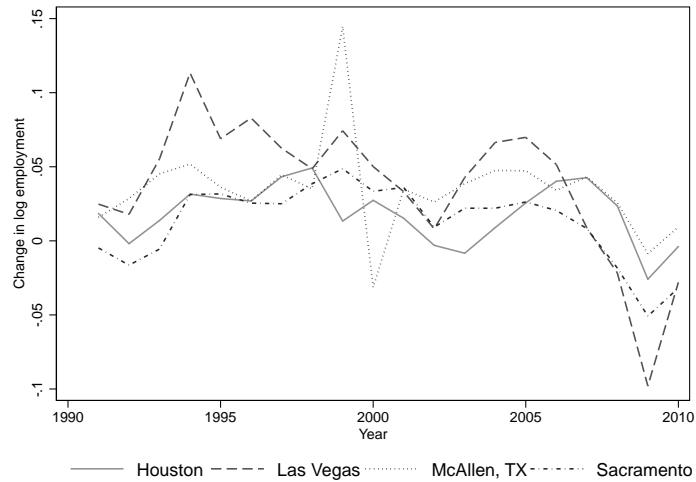
Liu, Amy, Roland V. Anglin, Richard M. Mizelle Jr., and Allison Plyer (eds.) (2011). *Resilience and Opportunity: Lessons from the U.S. Gulf Coast after Katrina and Rita*. Washington: The Brookings Institution.

Minnesota Population Center (2011). *National Historical Geographic Information System: Version 2.0*. Minneapolis: University of Minnesota.

Weir, Margaret, Nancy Pindus, Howard Wial, and Harold Wolman (eds.) (2012). *Urban and Regional Policy and Its Effects: Building Resilient Regions, Volume 4*. Washington: The Brookings Institution.



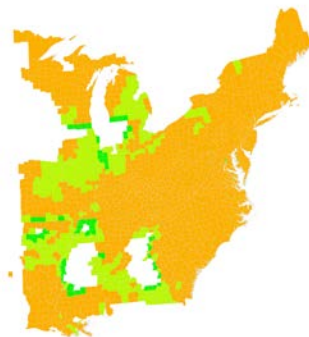
Panel A. Slow-growing metropolitan areas



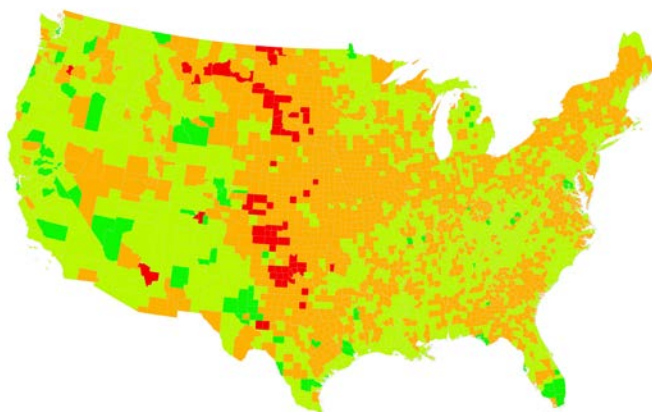
Panel B. Fast-growing metropolitan areas

Figure 1: Metropolitan area employment growth, 1991–2010

The four metropolitan areas in Panel A experienced annual employment growth that was slower (0.2%) than the U.S. metropolitan area average (1.1%) over the period 1991-2010. The four metropolitan areas in Panel B experienced annual employment growth of 2.6% over this period. Source data are from the Quarterly Census of Employment and Wages. Personal income growth was slower in the four metropolitan areas in Panel A (4.3%) versus the U.S. average (4.9%) and the four metropolitan areas in Panel B (6.7%).



Panel A. 1820–1830



Panel B. 1930–1940

Figure 2: Population growth in U.S. counties, 1820–1830 and 1930–1940

This map shows population growth in U.S. counties (normalized to 2010 boundaries) in two decades, 1820–1830 (Panel A) and 1930–1940 (Panel B). Each county is colored according to its relative growth in that decade. Dark green counties were the fastest-growing counties (growth was greater than one standard deviation above average); light green counties grew faster than the average county; orange counties grew slower than the average county; and red counties were the slowest-growing counties (growth was slower than one standard deviation below average). Counties in Panel A that are not colored were not yet organized in 1820.



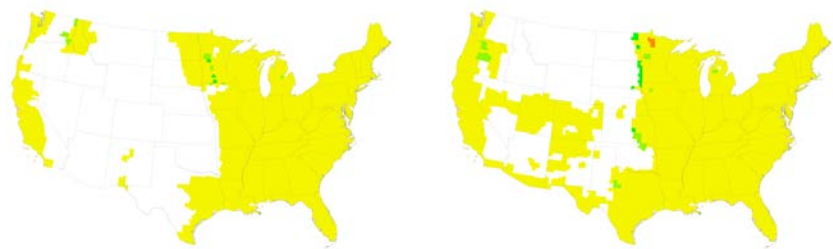
1790-1810; 1800-1820



1810-1830; 1820-1840

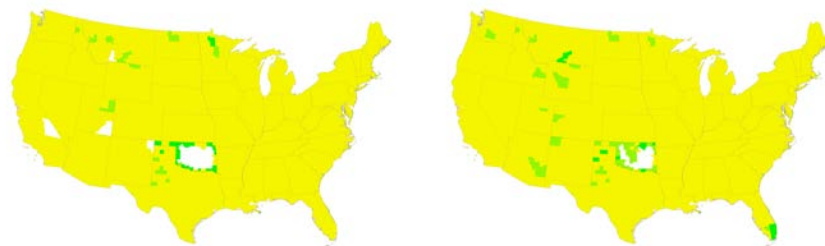


1830-1850; 1840-1860

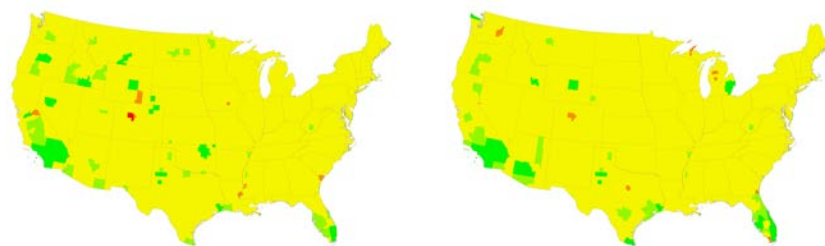


1850-1870; 1860-1880

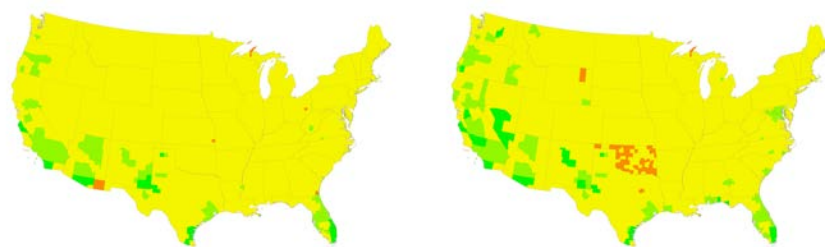
Figure 3: Filtered relative population growth of U.S. regions, 1790-1880



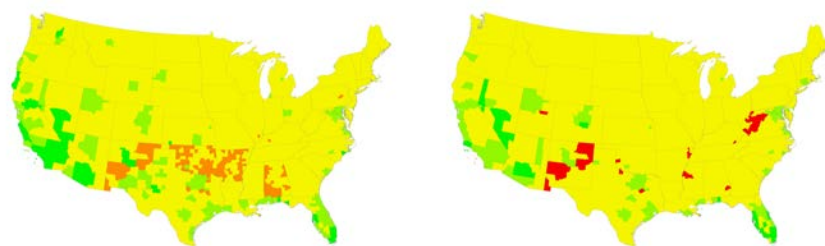
1880–1900; 1890–1910



1910–1920; 1910–1930

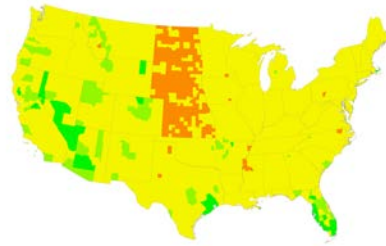


1920–1940; 1930–1950

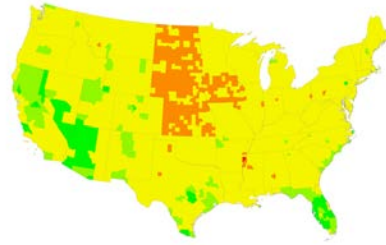


1940–1960; 1950–1970

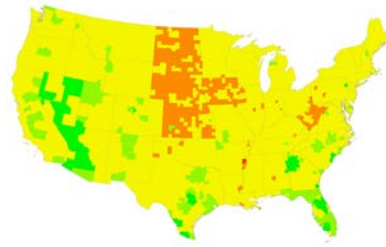
Figure 4: Filtered relative population growth of U.S. regions, 1880–1970



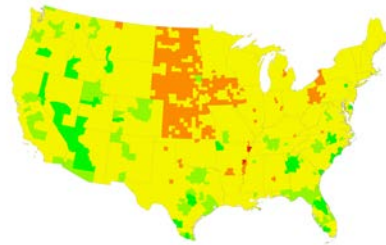
1960–1980



1970–1990



1980–2000



1990–2010

Figure 5: Filtered relative population growth of U.S. regions, 1960–2010

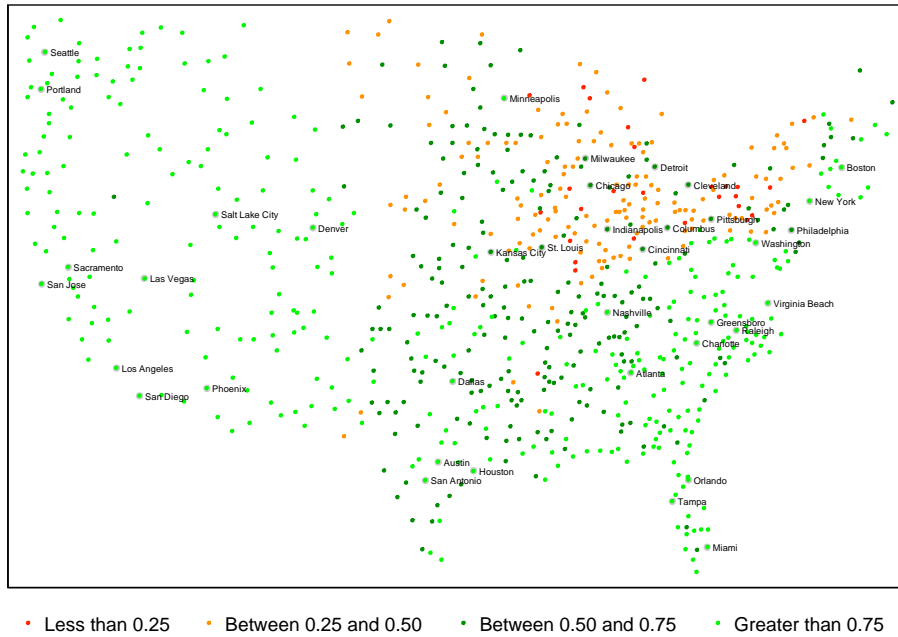


Figure 6: Predicted resilience for 2010 regions

Notes: This figure shows resilience scores for U.S. regions based on 2010 regional data on population density, industry mix, and education. Scores are predicted values based on estimates from regression (4) in Table 3. By construction of the probit model, scores are between 0 and 1.

Table 1: Resilient regions by year and resilience measure

Period	Total Regions	Resilience measure		
		$a = 0$	$a = -0.5$	$a = -1$
1920–1940	302	0.513	0.954	0.993
1930–1950	330	0.545	0.906	0.979
1940–1960	348	0.552	0.805	0.948
1950–1970	372	0.642	0.844	0.965
1960–1980	413	0.646	0.884	0.973
1970–1990	424	0.535	0.797	0.972
1980–2000	466	0.564	0.783	0.961
1990–2010	468	0.568	0.780	0.953

This table shows the number and share of regions I classify as “resilient,” based on population growth over the two decades indicated in the first column. The second column shows the number of regions with valid data in each period. The sample is an unbalanced panel of consistent-boundary U.S. regions (groups of 2010-boundary counties representing metropolitan, micropolitan, or rural areas), with a population of at least 50,000 in the initial year, whose date of entry is no less than sixty years before the initial year. The sample expands as more regions become more populated over time. Each cell shows the proportion of regions that satisfy resilience criteria as noted in the column headings. “Not below average” ($a = 0$) indicates that a region has not experienced two consecutive decades of below-average population growth, beginning in the initial year. “Not 0.5 s.d. below average” ($a = -0.5$) indicates that a region has not experienced two consecutive decades of population growth slower than 0.5 standard deviation below average, beginning in the year noted in the row heading. “Not 1 s.d. below average” ($a = -1$) indicates that a region has not experienced two consecutive decades of population growth slower than 1 standard deviation below average, beginning in the year noted in the row heading.

Table 2: Non-resilient regions with two decades of slow population growth

	Twenty-year period ending in:									
	1930	1940	1950	1960	1970	1980	1990	2000	2010	
Beckley, WV					X					
Bluefield, WV								X		
Blytheville, AR					X	X			X	
Buffalo, NY									X	
Burlington, IA							X			
Butte, MT			X							
Cleveland, MS					X	X				
Clinton, IA								X		
Corsicana, TX			X	X						
Danville, IL							X		X	
Decatur, IL								X	X	
Elmira, NY									X	
Fort Madison, IA								X		
Galesburg, IL								X	X	
Greenville, MS									X	
Greenwood, MS				X	X	X	X			
Hope, AR				X						
Houghton, MI	X	X	X	X						
Indianola, MS					X					
Jamestown, NY									X	
Johnstown, PA						X	X	X	X	
Marion, IN							X	X	X	
Mason City, IA									X	
Morgan City, LA								X		
Natchez, MS								X	X	
Oil City, PA								X	X	
Paris, TX				X						
Pittsburg, KS		X								
Pittsburgh, PA									X	
Pittsfield, MA									X	
Pottsville, PA				X	X	X				
Richmond, IN									X	
Roanoke Rapids, NC						X				
Rural AL				X						
Rural AR				X						
Rural GA				X						
Rural IA						X	X			
Rural KS			X	X	X	X	X	X	X	
Rural KY					X					
Rural MI				X						
Rural MO				X						
Rural ND			X	X	X	X	X	X	X	
Rural NE			X	X	X	X	X	X	X	
Rural NM				X	X					
Rural OK			X	X						
Rural SD						X	X			
Rural TX				X	X					
Rural WV					X			X		
Saginaw, MI									X	
Scranton, PA				X						
Selma, AL							X	X		
Steubenville, OH							X	X	X	
Wheeling, WV								X	X	
Youngstown, OH									X	

This table shows all regions in my sample that have experienced episodes of slow growth in U.S. history. The sample is an unbalanced panel of consistent-boundary U.S. regions (groups of 2010-boundary counties representing metropolitan, micropolitan, or rural areas), with a population of at least 50,000 in the initial year, whose date of entry is no less than sixty years before the initial year. These regions at some point have experienced two consecutive decades of population growth slower than 1 standard deviation below average. These periods are marked by X, indicating the twenty-year period ending in the year noted in each column heading. There were no regions that satisfied these criteria prior to 1910–1930. Regions are named for their principal city and state.

Table 3: Probit regression of regional resilience on historical factors

	(0)	(1)	(2)	(3)	(4)
Log population	12.2 (1.12)	0.127 (0.014)**	0.136 (0.015)**	0.082 (0.021)**	0.088 (0.023)**
Log land area	22.5 (1.12)	-0.111 (0.017)**	-0.124 (0.018)**	-0.066 (0.024)**	-0.075 (0.027)**
Share population educated	0.14 (0.05)		0.598 (0.180)**		1.089 (0.368)**
Industrial diversity	7.92 (1.14)			0.033 (0.014)*	0.023 (0.014)†
Employment share in traded industries	0.74 (0.07)				-0.602 (0.305)*
Census division indicators		X	X	X	X
Census year indicators		X	X	X	X
Number of observations		4,183	3,351	1,358	1,358
Number of regions		489	488	476	476
Pseudo- R^2		0.222	0.180	0.221	0.231

Each column in this table reports estimated marginal effects of the row variables from a separate probit regression. Dependent variable is an indicator for regional resilience, with a value of zero if the region has two consecutive decades of below-average population growth and one otherwise. Standard errors, clustered on region, are in parentheses. **—Significant at the 99% level of confidence; *—95%; †—90%. The sample is an unbalanced panel of consistent-boundary U.S. regions (groups of 2010-boundary counties representing metropolitan, micropolitan, or rural areas), with a population of at least 50,000 in the initial year, whose date of entry is no less than sixty years before the initial year. Sample sizes are smaller in some regressions due to the limited availability of data on the additional regressors. Each regression includes indicators for census divisions and census years. Population share educated is the share of the adult population with a college degree. In some historical censuses, this variable is the share of the adult population that is literate. Industrial diversity is measured as the inverse of a Herfindahl-Hirschman index of industry employment. Traded industries are defined as manufacturing, information, finance, trade, and educational and medical services. Column (0) reports means and standard deviations of variables for 1990 regions.

Table 4: Probit regressions, alternative resilience measures

	Resilience measure		
	$a = 0$	$a = -0.5$	$a = -1$
	(0)	(1)	(2)
Log population	0.088 (0.023)**	0.037 (0.022)†	0.010 (0.012)
Log land area	-0.075 (0.027)**	-0.006 (0.024)	0.003 (0.012)
Population share educated	1.089 (0.368)**	1.882 (0.369)**	0.445 (0.228)*
Industrial diversity	0.023 (0.014)†	0.029 (0.013)*	0.010 (0.007)
Employment share in traded industries	-0.602 (0.305)*	-0.036 (0.281)	0.399 (0.136)**
Pseudo- R^2	0.231	0.183	0.126

Notes: Each column reports marginal effects of the row variables from a separate probit regression. Dependent variable is an indicator for regional resilience, as noted in the column headings. Standard errors, clustered on region, are in parentheses. **—Significant at the 99% level of confidence; *—95%; †—90%. Column (0) repeats estimates from Table 3. “Not below average” ($a = 0$) indicates that a region has not experienced two consecutive decades of below-average population growth, beginning in the initial year. “Not 0.5 s.d. below average” ($a = -0.5$) indicates that a region has not experienced two consecutive decades of population growth slower than 0.5 standard deviation below average, beginning in the year noted in the row heading. “Not 1 s.d. below average” ($a = -1$) indicates that a region has not experienced two consecutive decades of population growth slower than 1 standard deviation below average, beginning in the year noted in the row heading. See notes from Table 3 for more details.

Table 5: Predicted resilience index for largest 2010 regions

Metropolitan area name	Resilience	C.I.	
San Jose-San Francisco-Oakland, CA Combined Statistical Area	0.999	0.608	1.000
San Diego-Carlsbad-San Marcos, CA Metropolitan Statistical Area	0.998	0.632	1.000
Denver-Aurora-Boulder, CO Combined Statistical Area	0.998	0.637	1.000
Seattle-Tacoma-Olympia, WA Combined Statistical Area	0.995	0.665	1.000
Los Angeles-Long Beach-Riverside, CA Combined Statistical Area	0.995	0.663	1.000
Portland-Vancouver-Hillsboro, OR-WA Metropolitan Statistical Area	0.994	0.681	1.000
Sacramento-Arden-Arcade-Yuba City, CA-NV Combined Statistical Area	0.993	0.677	1.000
Washington-Baltimore-Northern Virginia, DC-MD-VA-WV Combined Statistical Area	0.991	0.633	1.000
Phoenix-Mesa-Glendale, AZ Metropolitan Statistical Area	0.991	0.659	1.000
Salt Lake City-Ogden-Clearfield, UT Combined Statistical Area	0.988	0.702	1.000
Miami-Fort Lauderdale-Pompano Beach, FL Metropolitan Statistical Area	0.984	0.669	1.000
Fresno-Madera, CA Combined Statistical Area	0.975	0.667	1.000
Virginia Beach-Norfolk-Newport News, VA-NC Metropolitan Statistical Area	0.973	0.680	1.000
Tampa-St. Petersburg-Clearwater, FL Metropolitan Statistical Area	0.971	0.678	1.000
Raleigh-Durham-Cary, NC Combined Statistical Area	0.970	0.690	1.000
Las Vegas-Paradise-Pahrump, NV Combined Statistical Area	0.964	0.645	1.000
Orlando-Deltona-Daytona Beach, FL Combined Statistical Area	0.959	0.684	1.000
Jacksonville, FL Metropolitan Statistical Area	0.949	0.695	1.000
Atlanta-Sandy Springs-Gainesville, GA-AL Combined Statistical Area	0.947	0.696	1.000
Boston-Worcester-Manchester, MA-RI-NH Combined Statistical Area	0.941	0.580	1.000
Charlotte-Gastonia-Salisbury, NC-SC Combined Statistical Area	0.925	0.691	1.000
Austin-Round Rock-Marble Falls, TX Combined Statistical Area	0.923	0.630	1.000
Richmond, VA Metropolitan Statistical Area	0.914	0.682	1.000
Hartford-West Hartford-Willimantic, CT Combined Statistical Area	0.912	0.573	1.000
Houston-Baytown-Huntsville, TX Combined Statistical Area	0.911	0.613	1.000
Greensboro-Winston-Salem-High Point, NC Combined Statistical Area	0.911	0.683	1.000
Dallas-Fort Worth, TX Combined Statistical Area	0.906	0.628	1.000
Greenville-Spartanburg-Anderson, SC Combined Statistical Area	0.897	0.678	1.000
San Antonio-New Braunfels, TX Metropolitan Statistical Area	0.893	0.618	1.000
New Orleans-Metairie-Bogalusa, LA Combined Statistical Area	0.875	0.598	1.000
Oklahoma City-Shawnee, OK Combined Statistical Area	0.871	0.600	1.000
Nashville-Davidson-Murfreesboro-Columbia, TN Combined Statistical Area	0.832	0.552	1.000
Minneapolis-St. Paul-St. Cloud, MN-WI Combined Statistical Area	0.832	0.535	1.000
New York-Newark-Bridgeport, NY-NJ-CT-PA Combined Statistical Area	0.821	0.427	1.000
Memphis, TN-MS-AR Metropolitan Statistical Area	0.812	0.530	1.000
Louisville/Jefferson County-Elizabethtown-Scottsburg, KY-IN Combined Statistical Area	0.810	0.531	1.000
Birmingham-Hoover-Cullman, AL Combined Statistical Area	0.801	0.525	1.000
Knoxville-Sevierville-La Follette, TN Combined Statistical Area	0.792	0.508	1.000
Milwaukee-Racine-Waukesha, WI Combined Statistical Area	0.745	0.412	1.000
Kansas City-Overland Park-Kansas City, MO-KS Combined Statistical Area	0.744	0.471	1.000
St. Louis-St. Charles-Farmington, MO-IL Combined Statistical Area	0.742	0.455	1.000
Philadelphia-Camden-Vineland, PA-NJ-DE-MD Combined Statistical Area	0.715	0.363	1.000
Chicago-Naperville-Michigan City, IL-IN-WI Combined Statistical Area	0.711	0.379	1.000
Detroit-Warren-Flint, MI Combined Statistical Area	0.670	0.355	0.985
Cleveland-Akron-Elyria, OH Combined Statistical Area	0.659	0.346	0.972
Indianapolis-Anderson-Columbus, IN Combined Statistical Area	0.553	0.264	0.841
Columbus-Marion-Chillicothe, OH Combined Statistical Area	0.522	0.239	0.806
Cincinnati-Middletown-Wilmington, OH-KY-IN Combined Statistical Area	0.517	0.229	0.805
Pittsburgh-New Castle, PA Combined Statistical Area	0.509	0.207	0.810
Buffalo-Niagara-Cattaraugus, NY Combined Statistical Area	0.506	0.204	0.808
Albany-Schenectady-Amsterdam, NY Combined Statistical Area	0.496	0.189	0.803
Dayton-Springfield-Greenville, OH Combined Statistical Area	0.485	0.199	0.770
Grand Rapids-Muskegon-Holland, MI Combined Statistical Area	0.483	0.199	0.768
Rochester-Batavia-Seneca Falls, NY Combined Statistical Area	0.480	0.188	0.772

Notes: This table shows predicted resilience indexes for the largest U.S. regions by population, which are all metropolitan areas, based on 2010 regional data on population density, industry mix, and education. Resilience indexes for smaller U.S. regions are not shown in this table. Predicted values based on estimates from regression (4) in Table 3. By construction of the probit model, scores are between 0 and 1. The second and third columns show the lower and upper 67% confidence interval based on the standard errors of the predicted values.