The Settlement of the U.S., 1790 to 2000: The Emergence of Gibrat's Law

Klaus Desmet
Universidad Carlos III

Jordan Rappaport
Federal Reserve Bank of Kansas City

September 2011
Preliminary and Incomplete

Abstract

This paper studies the long run spatial development of U.S. counties and metro areas between 1790 and 2000. The often-documented orthogonality between population growth and initial population size — Gibrat's law — only emerges recently. Between 1790 and 1940 population growth tends to be decreasing with county population, except among the largest counties, for which population growth is slightly increasing with county population. The negative correlation between size and growth among small locations is largely driven by the entry of new counties, and dissipates after 1940. Over time the relation between size and growth flattens, though it is only for the subgroup of counties that entered long ago that Gibrat's law cannot be rejected in recent years. A simple one-sector model successfully approximates this evolving distribution of local population. Locations differ by a permanent TFP parameter and a date of entry into the aggregate economy. Upon entry, an assumed friction slows the transition of locations' populations to their steady-state levels. Gibrat's law emerges as the system of locations approaches its steady state.

1 Introduction

The orthogonality of population growth and initial population — Gibrat's law — has long been considered a stylized fact of the growth of a system of localities (Eeckhout, 2004). Although recently doubts have been raised about its empirical validity (Holmes and Lee, 2008), even bigger doubts emerge when going back further in time (Michaels et al., 2010). Almost all the estimates establishing Gibrat's law are based on very recent experiences. But given the rapid expansion of settled land area and aggregate population of the United States over the course of the nineteenth century, it surely makes sense to examine whether the recent, arguably steady state characterization, applies equally to a system of locations in transition. Our paper therefore aims to shed light on the dynamics of local population growth as a country is gradually settled and then matures.

The paper draws on a panel dataset constructed from each Decennial Census from 1790 through 2000. In addition to this two century time span, the dataset has several unique characteristics. First, we construct county geographies to be unchanged over periods for which growth is being measured. Often, county geographies vary considerably over time due to mergers, splits, and other sorts of border changes. Failing to account for these changes biases results. Second, we track the years since a county is first included in a decennial census. This allows us to document the very different population levels and growth patterns of "young" and "old" counties. Third, we merge counties into metropolitan areas when the contemporary population and geographical characteristics meet predefined criteria. Doing so, we argue, makes our locations better correspond to an economically meaningful geography.

Breaking the dataset into 11 panels of 10 or 20 years each, we find two sets of empirical results. First, the orthogonality of population growth and initial population can be rejected for all 11 periods. During the nineteenth and early twentieth centuries, population growth was strongly negatively correlated with initial population for small and intermediate-sized locations. This negative correlation began to dissipate during the 1920s until by mid-century it had essentially disappeared. Over much of the same two-hundred year time span, there was a moderate positive correlation between growth and initial population for larger counties which weakened but remained present in recent time periods. The dissipation of the negative correlation and the weakening of the positive correlation have moved the U.S. system of locations closer to matching Gibrat's law. But as late as the 1980 to 2000 period, deviations from that match remained statistically significant and meaningful in size.

A second empirical finding is that counties in a similar age category exhibit similar growth patterns in different time periods. Newly delineated counties, independent of the time period in which such delineation occurred, experienced relatively high growth. This reflects that many newly created counties were far below their steady state. As the population levels of new counties rapidly increased, population growth rates began to taper. For more mature counties growth was more orthogonal to size in all time periods. Correspondingly, the changing growth patterns of counties over time is related to the changing composition of counties. As entry of new counties has slowed down, the overall growth pattern has become flatter. This is especially true for more mature counties. For that subgroup, Gibrat's law can no longer be rejected in recent years.

A relatively simple, one-sector general equilibrium model successfully approximates this evolving pattern of local population growth. A large number of locations across which there is high labor mobility differ with respect to when they enter the integrated system of locations and with respect to exogenous total factor productivity. Locations are assumed to have very low populations upon entry. A friction that dampens a location's productivity as an increasing function of its growth rate causes an extended transition to the system's steady state. During this transition, population growth is negatively correlated with size for small and intermediate-sized counties. [Additionally, we allow agglomeration economies to kick in when a location reaches a certain threshold size. This assumption relaxes the congestion that keeps down the populations of high-productivity locations. As a result, growth becomes positively correlated with size among intermediate to large-sized locations.] As entry ceases and the economy approaches its steady state, the relation between size and growth becomes orthogonal.

We are not the first to reject the applicability of Gibrat's law in the context of long-run development. Michaels et al. (2010) analyze the relation between size and growth of U.S. places between 1880 and 2000. They find a U-shaped relationship between population and growth, and propose a model of the structural transformation to explain this. However, by focusing on only two years, Michaels et al. (2010) do not allow for dynamics, and more importantly, they do not allow for entry of new places and for settlement of new land. In our model the entry of new locations and the transition to a steady state is key. This is consistent with the data, which suggest that the relation between size and growth changes dramatically over time when looking at *all* counties, but remains relatively stable over time when only looking at *young* counties or only at *old* counties. To understand the emergence of Gibrat's law, allowing for entry, and distinguishing between "young" and "old" counties turns out to be crucial. The theoretical model is then calibrated to match the empirical dynamics.

Our first conclusion is that historically Gibrat's law has not held. The relation between size and growth has been primarily characterized by convergence for smaller counties, and some degree of divergence between counties above a certain threshold. A second conclusion is that over time the relation between size and growth has become flatter, thus increasingly approximating Gibrat's law. A third conclusion is that this slow convergence is mainly due to the entry of new counties and to local frictions in population growth. The entry of new locations has two effects. It causes a strong force for convergence in young counties, and it slows down the transition of the entire system towards a situation of long-run steady state. This transition can be further slowed down by frictions in the local population growth rate. As entry loses importance, over time the negative correlation between size and growth for small counties disappears. However, the positive correlation between size and growth for mid-sized locations continues to be present, though it has also weakened, in more recent years. Desmet and Rossi-Hansberg (2009) explain this positive correlation in a life-cycle model of industries where young sectors benefit from agglomeration economies, whereas Michaels et al. (2010) emphasize the structural transformation and the increasing share of manufacturing in intermediate-sized locations. [While not the main focus of this paper, we incorporate these insights in a reduced-form manner in our theoretical model by allowing for agglomeration economies once locations reach a certain threshold size.]

The rest of the paper is organized as follows. Section 2 discusses the data, Section 3 reports the main empirical findings, Section 4 proposes and calibrates a theory of settlement, and Section 5 concludes.

2 Data

Our dataset is built using aggregate data for county and county-equivalents from each Decennial Census, from 1790 through 2000 (Haines, 2005). The nineteenth century was a period during which U.S. settled land area grew rapidly. As a result the number of reported county and county equivalents soared from under 300 in 1790 to almost 2900 in 1900. The number of counties continued to increase through the mid twentieth century, but at a more gradual pace.

To facilitate the analysis over time, we use a "county longitudinal template" (CLT) augmented by a map guide to decennial censuses to merge counties as necessary to keep geographies constant over time (Horan and Hargis, 1995; Thorndale and Dollarhide, 1987). By far the most common reason for doing so is the splitting of an existing county into two or more successor counties. For example, suppose county A splits into counties B and C in 1841. In this case we merge counties B and C together when measuring growth from

1840 or earlier to a year later than or equal to 1841. A similar procedure is used in the less frequent event of two or more counties joining into a single county. This is more common during the twentieth century than during the nineteenth century. For example, suppose counties A and B join together in 1901 to form county C. Addressing this requires merging together counties A and B from 1900 or earlier before comparing them to county C in 1901 or later. This adjustment for geographical changes implies that we construct a separate data set for each of the 11 multi-year periods we study (1790-1800, 1800-20, ... , 1980-2000). In other words, the 1790-1800 data set is based on geographies in 1790; the 1800-20 data set is based on geographies in 1800, etc.

The CLT also addresses two further issues. First, if there is a significant border change among two or more counties, it simply merges together all affected counties, both before and after the change. Second, in some areas there may be a small county equivalent completely surrounded by a much larger county. This circumstance is especially prevalent in Virginia, which labels many large municipalities as "independent cities" and considers them county equivalents. In this case, the smaller county equivalent and the larger surrounding county are simply merged for all years in which they are jointly present. Doing so removes an endogenous determinant of population and land area that could potentially bias results.

The various county merges implied by the CLT reduces the number of observed locations by about 30 percent during the first part of the nineteenth century and by a relatively modest amount during the twentieth century. An important limitation of the CLT is that it is formally constructed back only to 1840. For 1790 through 1830, we use 1840 merge codes with a handful of modifications. Hence if a county A were to split into B and C in 1791, the CLT would lack a code to merge B and C back together in 1800 and after to match county A in 1790. Moreover, either county B or county C would likely inherit the county A Census identification code for 1790. Reported population growth for county A will thus be strongly biased down due to the loss of significant land area. Results prior to 1840 should therefore be interpreted with some caution.

When metro areas exist, they better correspond to unified labor markets than do their constituent counties. We therefore combine counties to form metropolitan areas whenever groups of counties meet a set of criteria developed by the Census Bureau and the Office of Management and Budget. These criteria are applied to conditions at the start of any period over which growth is measured. In other words, if a certain group of adjacent counties do not meet the metro criteria in an initial year but do meet it 20 year later, they will *not* be combined. Doing otherwise would introduce a bias towards finding a positive correlation between growth and initial population.

The criteria themselves do change slightly over time. For the 1800 through 1940 censuses we rely on Gardner (1999), who applies Census Bureau criteria from 1950 retroactively using to IPUMS data from each of the censuses. The Census Bureau criteria, in turn, consider a range of characteristics.[1] For the 1960

[1]A metropolitan designation requires a contiguous group of counties that includes at least one urban center of at least 50,000 inhabitants; all counties must have no more than one-third of employed persons working in the agricultural sector; at least 10,000 nonagricultural workers; and at least 10 percent as many nonagricultural workers as in the primary county of the metropolitan area. If the number of nonagricultural workers is below one of these thresholds, the county could nevertheless be included in the metropolitan area if at least half of its population resides in a thickly settled area (at least 150 persons per

to 1980 period, we use the the "standard metropolitan statistical areas" delineations released by the Office of Management and Budget in 1963. As these were primarily based on 1960 census data, we are comfortable with their release several years into the period over which growth is measured. We additionally use the New England County Metropolitan Areas released in 1975. For the 1980 to 2000 period, we use the 1983 OMB metro definitions with the exception that we retain the 1963 delineation of of the New York City metro area. The 1983 Consolidate Metropolitan Area delineation for the New York City metro is far too large to be considered a single labor market. But its constituent Primary Metropolitan Statistical Area components tend to be too small. We also use the 1983 New England County Metropolitan area for New England.

In our subsequent analysis, county and metro "age" (taken as a measure of time since first reported in the census) plays an important role. We measure this based on the "entry date" of a raw county identification in the census data. In other words, if an observation with raw identification code 101 first appears in the 1790, that observation–and anything with which it may eventually be merged–is assigned an entry date of 1790. A drawback to this methodology is that following splits of raw counties, one or more of the "newly-identified" raw counties will be assigned an entry year equal to the time of the split. This time-of-secession value for the time of entry will then persist for all occurrences of the same raw county in future years. To the extent that such a split may be correlated with above-average population growth, dating entry in this way may introduce a bias towards finding that "young" observations grow more quickly than "old" ones. On the other hand, many historical secessions capture the opening of sparsely populated land to settlement. For example, the majority of New York State's land area in 1790 was split into just three counties: Montgomery, Ontario, and Indian Lands. By 1820, these three raw counties had been split into approximately 30 raw counties. Almost all of these will be assigned an entry year of 1800, 1810, or 1820, rather than 1790. To the extent that much of New York State was sparsely populated in 1790, such an assignment is probably appropriate.

3 Empirical Results

Because of a lack of historical data, most papers on the spatial distribution of population have focused on systems of locations that are probably close to their long-run steady state. Not much attention has been paid to growth patterns during probable transitions to long-run steady states. To the extent that Gibrat's law indeed exists presently, we know little about how it emerged over time. For the United States, this means we are largely ignorant about how local growth patters developed as the present-day land mass was gradually settled. By spanning the nineteenth and twentieth centuries, our analysis aims to address this shortcoming. Over this period, the continental U.S. population grew from less than 4 million to more than 275 million; the settled continental land area grew from less than 1 million square miles, primarily along the eastern seaboard, to over 7.5 million square miles, coast to coast.

square mile) contiguous to the central city.

Because growth patterns gradually changed over the course of these two centuries, we apply our analysis separately to each of the 20-year time spans from 1800 through 2000. In other words, we run separate regressions for the ten time spans starting with 1800 to 1820 and ending with 1980 to 2000. As described in the data section above, the set of locations for each of these ten time spans differs. This reflects the continual entry into the U.S. system of locations during the nineteenth century as well as numerous changes in county and metro boundaries through the twentieth century. But as also described above, we combine counties as necessary to maintain constant boundaries between the beginning and end of any 20 year time span.

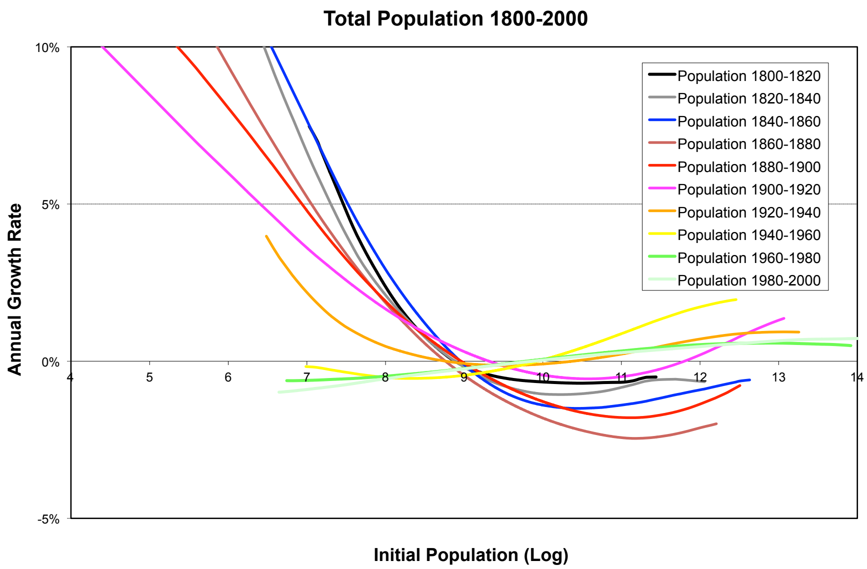

Figure 1: Growth and Size, 1800-2000

For the analysis itself, we run two main types of regressions. The first are non-linear kernels, which show a first-derivative continuous approximation of growth versus initial population level. These regressions allow for sharp visual comparisons of the growth versus initial level relationship as it develops over time. Second, we run comparable continuous, piecewise-linear spline regressions of growth on initial population. These allow for more quantitative measures of the dependence on growth on initial population including the ability to formally test for orthogonality. More specifically, the kernel regressions are of the form

$$(L_{i,t+20} - L_{i,t})/20 = \phi(L_{i,t}) + e_{i,t}$$

where $L_{i,t}$ is the log population for location i in year t. The left hand side is just the average rate of population growth over twenty years. The estimation uses an Epanechnikov kernel with optimal bandwidth (Desmet and Fafchamps, 2006). The continuous piecewise-linear regressions are of the form

$$(L_{i,t+20} - L_{i,t})/20 = \vec{\beta}(1 + \vec{L_{i,t}}) + e_{i,t}$$

where the vector $\vec{L_{i,t}}$ is 1 by k where k is the number of spline segments. The mapping of log population into its vector form is such that each spline segment measures the *marginal* affect of initial population size on growth while forcing the overall estimated spline to be piecewise continuous. Unless stated otherwise, the regressions take metro areas and rural counties to be the geographic unit of observation.

Both sets of regressions strongly reject the orthogonality of initial population and subsequent growth for virtually all time periods. The kernel regressions of the population growth rate as a function of the log of population at the beginning of each 20-year period in Figure 1 emphasize this visually. To facilitate the graphical representation, the population growth rate for each 20-year period has been demeaned.[2] If Gibrat's law were to hold, growth should be orthogonal to size, so that Figure 1 should only show horizontal lines coinciding with the horizontal axis. This is clearly not the case. A number of additional features stand out. First, until 1940 the relation between size and growth is very negative for the group of small counties, implying strong convergence. Second, for the intermediate-sized and large locations the relation between size and growth is positive, indicating divergence. Third, this negative correlation for the group of small counties dissipates over time, and is no longer present in the postwar period. Fourth, overall the relation between size and growth becomes flatter over time, and in that sense gets closer to Gibrat's law. In spite of that tendency, even in the most recent period, 1980-2000, Gibrat's law can still be rejected. To get a sense of the goodness of fit, Figures 2 and 3 provide scatter plots of the data, together with the estimated kernel regression and the 95% confidence intervals, for two time periods, 1840-1860 and 1980-2000.

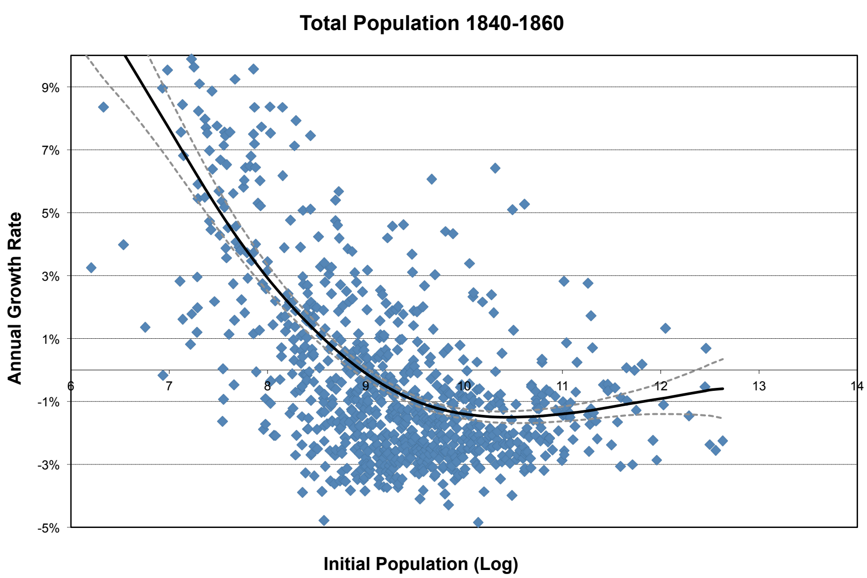

Figure 2: Growth and Size, 1840-1860

[2]That is, we have subtracted the average growth rate from the actual growth rate.

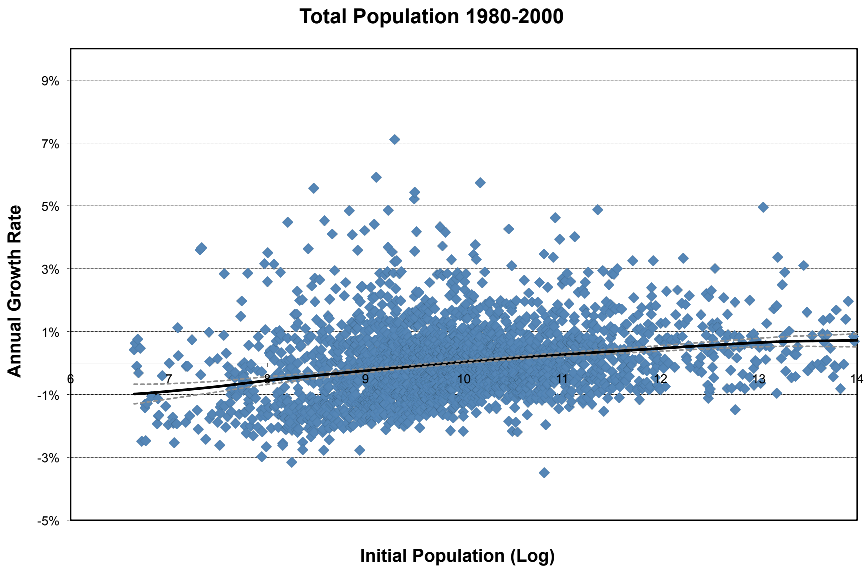

Figure 3: Growth and Size, 1980-2000

The corresponding continuous piecewise-linear regressions are reported in Table 1. The intervals go in increments of 1 (in terms of logs) but intervals are joined if there are not enough observations in a given interval. The results are consistent with those of the kernel regressions, but have the advantage of providing standard errors. One important observation is that for locations with a log population above 13 (which corresponds to a population level above 442,414) the relation between size and growth tends not to be statistically significant. In other words, for large counties Gibrat's law cannot be rejected. This suggests that Gibrat's law may hold for metropolitan areas — as suggested by the literature — but not for smaller places. Note that the R^2 is relatively low in 1790, rises throughout the 19th century, then starts declining, reaching values close to 0 in recent decades. This suggests that starting at 1790 something happened that moved the system away from steady state. Rapid expansion of settled land seems to be a good candidate.[3] As settlement progressed, the system gradually converged back to a new steady state.

Our interpretation for the strong convergence among small counties until the mid-twentieth century has to do with small counties being mostly below their steady state level of population, and therefore experiencing high growth rates. Over time, as those small counties that are below steady state move to the right of the distribution, the remaining small counties which did not move to the right are those with relatively low productivity and thus with relatively small steady state sizes. For intermediate-sized and large locations, there is a positive relation between size and growth, consistent with findings in Desmet and Rossi-Hansberg (2009) and Michaels et al. (2010). As time goes by, the relation becomes flatter as locations approach their respective steady states. According to this interpretation, the downward sloping part is driven

[3]In this context settled land should be understood as population density reaching a minimum threshold.

by “young” counties, and the relation between size and growth should be flatter for “older” counties than for “younger” counties. To explore the validity of this interpretation, for each 20-year time period, we split up our sample into “young” counties, “middle-aged” counties, and “old” counties. For a given 20-year time period, we define a county to be “young” if it entered in the 30 years prior to the starting year of our time period. Similarly, “middle aged” counties are between 40 and 70 years old, and “old” counties more than 80 years old.

All Locations (empirical)	**(1)** **1790 -1800**	**(2)** **1800 -1820**	**(3)** **1820-1840**	**(4)** **1840-1860**	**(5)** **1860-1880**	**(6)** **1880-1900**	**(7)** **1900-1920**	**(8)** **1920-1940**	**(9)** **1940-1960**	**(10)** **1960-1980**	**(11)** **1980-2000**
log(pop) bin:											
min to lowest lb	**-0.060** **(0.013)**	**-0.034** **(0.004)**	**-0.053** **(0.008)**	**-0.054** **(0.008)**	**-0.048** **(0.006)**	-0.008 (0.024)	**-0.023** **(0.008)**	**-0.039** **(0.006)**	0.004 (0.004)	-0.001 (0.004)	0.001 (0.002)
lpop.04to06						**-0.029** **(0.009)**					
lpop.05to07					**-0.045** **(0.006)**						
lpop.06to07						**-0.044** **(0.013)**					
lpop.07to08					**-0.032** **(0.007)**	**-0.025** **(0.007)**	**-0.023** **(0.007)**				
lpop.08to09			**-0.021** **(0.006)**	**-0.038** **(0.007)**	**-0.027** **(0.005)**	**-0.024** **(0.004)**	**-0.017** **(0.004)**	-0.002 (0.002)	**-0.007** **(0.002)**	0.003 (0.003)	**0.005** **(0.002)**
lpop.09to10	0.006 (0.008)	0.004 (0.004)	**-0.010** **(0.004)**	-0.005 (0.003)	**-0.009** **(0.002)**	**-0.011** **(0.002)**	**-0.009** **(0.002)**	0.001 (0.001)	**0.006** **(0.001)**	**0.003** **(0.001)**	**0.003** **(0.001)**
lpop.10to11		-0.001 (0.007)	**0.012** **(0.004)**	0.003 (0.003)	0.000 (0.002)	0.003 (0.002)	**0.010** **(0.002)**	**0.007** **(0.001)**	**0.011** **(0.001)**	**0.004** **(0.001)**	0.001 (0.001)
lpop.11to12					**0.013** **(0.004)**	**0.013** **(0.003)**	**0.010** **(0.003)**	**0.006** **(0.002)**	**0.012** **(0.003)**	0.001 (0.002)	**0.004** **(0.002)**
lpop.12to13								-0.004 (0.003)	**-0.006** **(0.003)**	0.000 (0.002)	0.000 (0.002)
lpop.13to14										0.002 (0.004)	0.002 (0.003)
highest ub to max	-0.002 (0.009)	0.002 (0.007)	-0.004 (0.007)	0.005 (0.004)	-0.004 (0.003)	0.002 (0.002)	-0.001 (0.002)	*0.003* *(0.002)*	0.000 (0.002)	**-0.007** **(0.002)**	-0.002 (0.002)
Bins	3	4	5	5	8	9	7	7	7	8	8
N	226	306	540	861	1,687	2,355	2,648	2,943	2,979	2,849	2,630
R^2	0.247	0.383	0.465	0.589	0.762	0.647	0.359	0.306	0.132	0.043	0.076

Table shows results from regressing average annual population growth rate for listed time period on a spline of initial population. The spline is constructed to be continuous with respect to population. Standard errors in parentheses are robust to a spatial correlation using the procedure discussed in the main text. Bold type signifies coefficients that statistically differ from zero at the 0.05 level. Italic type signifies coefficients that statistically differ from zero at the 0.10 level. Top results row is coefficient for the lowest bin, whatever it may be. Its range can be immediately inferred from the lower bound of the next highest bin for which a coefficient is reported. For example, in column 1 (1790-1800) the initial bin extends from zero up to 9. Similarly, lower bound of the final bin is upper bound of highest previously reported bin.

Table 1: Growth and Size, Piecewise-Linear Regressions, 1800-2000

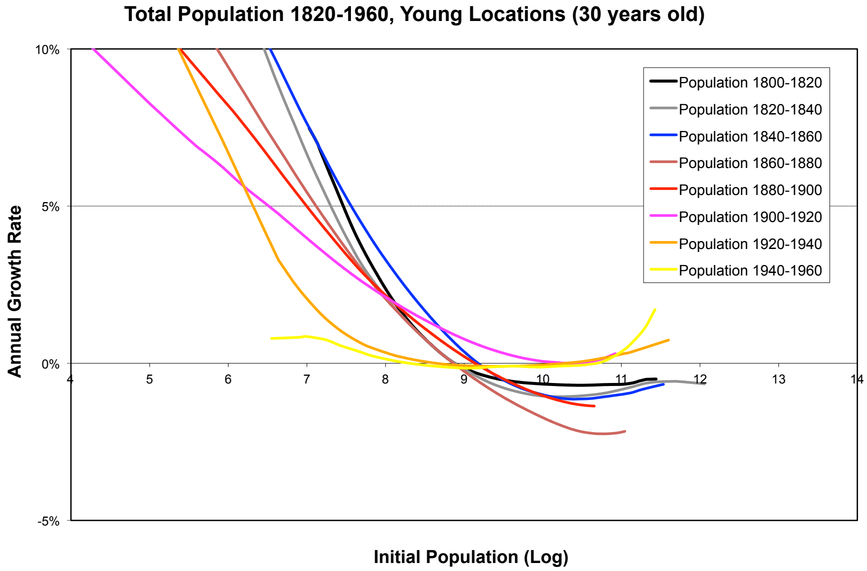

Figure 4: Young Locations

Figures 4, 5 and 6 show kernel regressions for young, middle-aged and old counties.[4] As can be seen, within each age category the growth patterns do not change dramatically over time. There is strong convergence for young counties across all time periods, except for 1940-1960. In that most recent time period, there are few young counties, and none of them are very small. We do not include young counties beyond 1940-1960 because of the lack of observations. For middle-aged counties, we see a U-shaped relation between size and growth. The smaller ones exhibit convergence and the larger ones divergence. Not surprisingly, compared to the young counties, their distribution shifts to the right. For old counties, we observe a slightly positive relation between size and growth, and their distribution shifts even more to the right. For more recent periods, the relation for old counties becomes flatter.

Another way of looking at the same evolution is by considering a given time period and comparing the behavior of locations in different age categories. Figure 7 focuses on 1900-1920 and Figure 8 on 1980-2000. To better see the change in patterns, for 1980-2000 we have split the "old" locations into three subgroups, "moderately old" (80-110 years), "very old" (120-150 years), and "extremely old" (more than 150 years). In Figure 7 young locations show strong convergence, middle-aged locations a U-shaped relation, and the old counties slight divergence. In Figure 8, we have dropped young locations because of a lack of observations. The pattern that stands out is that as locations age, the relation between size and growth flattens. In 1980-2000 for the "extremely old" category we can no longer reject Gibrat's law, as suggested by the fact that we can draw a horizontal line that is in between the confidence intervals (not shown in the figure).

[4]Note that the demeaning of the data is always done on the entire set of counties and metro areas, whereas the kernel regressions are run on a subset of those. This explains why in certain cases a kernel curve can be entirely above or below the horizontal axis which is centered at zero.

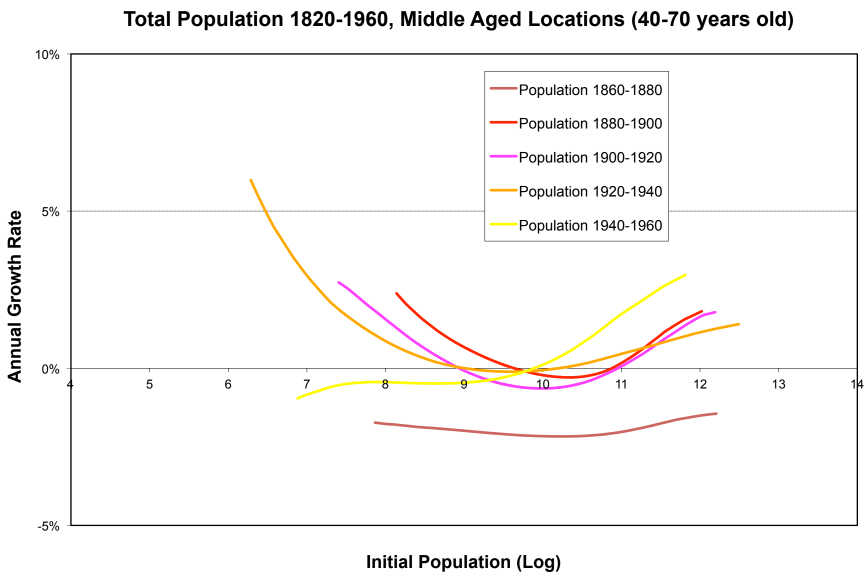

Figure 5: Middle Aged Locations

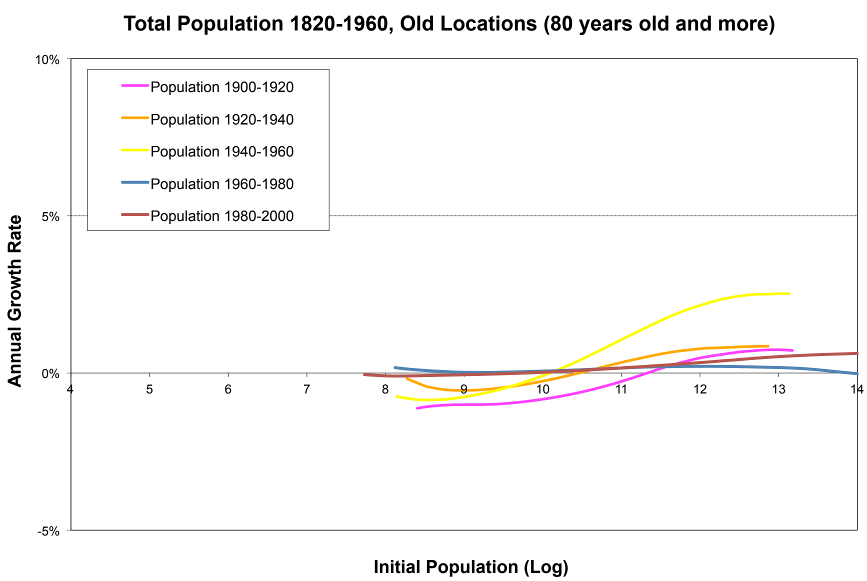

Figure 6: Old Locations

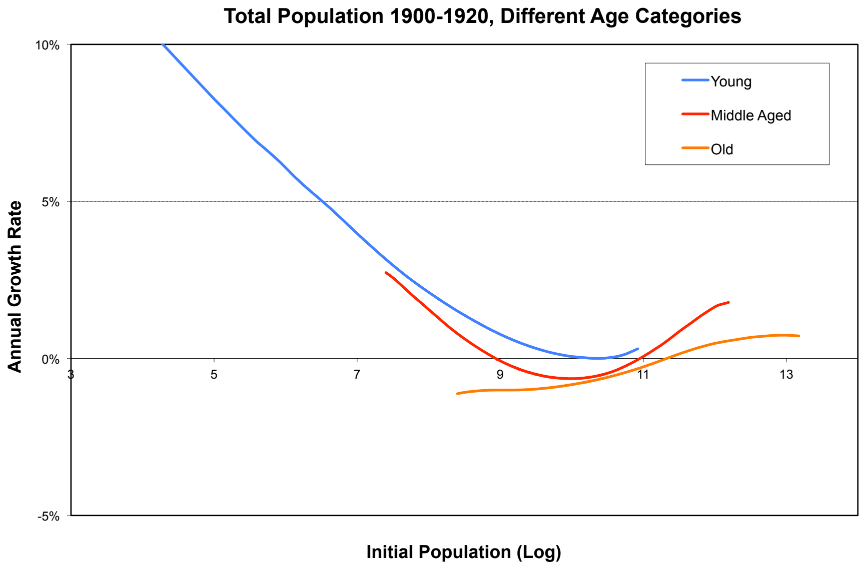

Figure 7: Different Age Categories, 1900-1920

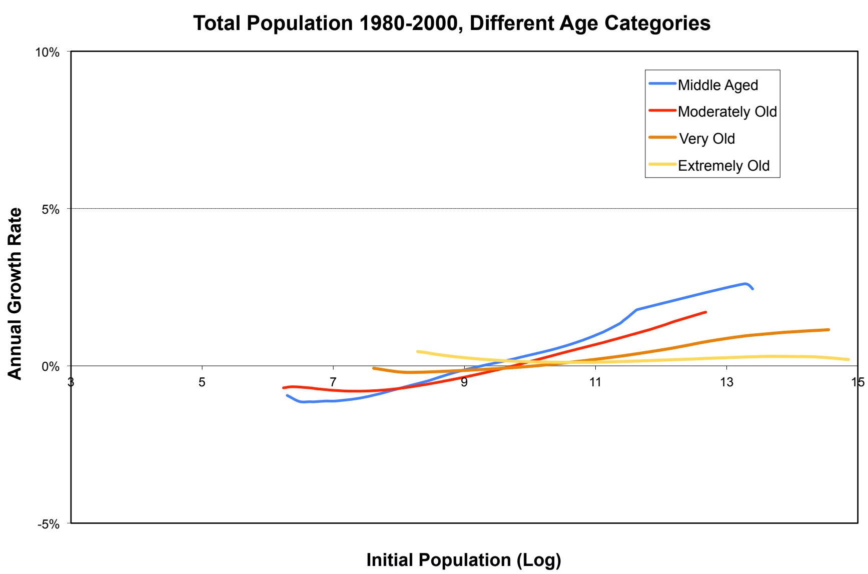

Figure 8: Different Age Categories, 1980-2000

The fact that it takes a long time for the relation between size and growth to become flatter is consistent with counties facing substantial frictions rapid population growth. If so, county growth rates should exhibit considerable persistence: those that grow faster than average one twenty-year period should grow faster than average the subsequent subsequent twenty-year period as well. Regressing county population growth on its lagged value establishes that persistence accounts for more than 20 percent of the variation in county growth rates (Table 2, top panel). This persistence derives about evenly from counties' growing at greater than the aggregate rate of existing locations and counties growing slower than this rate (Table3, bottom panel). In other words, location growing rapidly and locations growing slowly both exhibit persistence (Glaeser and Gyourko, 2005; Rappaport, 2004).

All the results in the empirical section take metropolitan areas and remaining counties as the units of observation. As a robustness check, we redid all exercises using all counties as the unit of observation. The results are very similar.

	(2) 1800-20 on 1790-1800	**(3)** 1820-40 on 1800-20	**(4)** 1840-60 on 1820-40	**(5)** 1860-80 on 1840-60	**(6)** 1880-1900 on 1860-80	**(7)** 1900-20 on 1880-1900	**(8)** 1920-40 on 1900-1920	**(9)** 1940-60 on 1920-1940	**(10)** 1960-80 on 1940-1960	**(11)** 1980-2000 on 1960-1980
Simple										
ρ		0.314	0.287	0.135	0.177	0.275	0.239	0.507	0.358	0.672
		(0.036)	(0.018)	(0.010)	(0.010)	(0.009)	(0.012)	(0.018)	(0.013)	(0.011)
N		309	544	865	1692	2369	2655	2950	2982	2852
R^2		0.201	0.325	0.174	0.165	0.265	0.136	0.210	0.197	0.581
By init grwth										
ρ_{below}		0.275	0.362	0.484	0.489	0.602	0.285	0.885	0.359	0.721
		(0.103)	(0.055)	(0.038)	(0.058)	(0.043)	(0.031)	(0.040)	(0.021)	(0.020)
N_{below}		201	327	518	950	1,419	1,879	2,221	2,399	1918
mrgnl R^2_{below}		0.019	0.055	0.143	0.035	0.059	0.027	0.125	0.077	0.189
ρ_{above}		0.327	0.263	0.058	0.147	0.230	0.221	0.329	0.357	0.629
		(0.047)	(0.024)	(0.012)	(0.011)	(0.011)	(0.016)	(0.025)	(0.029)	(0.018)
N_{above}		109	217	347	742	950	776	729	583	935
mrgnl R^2_{above}		0.124	0.147	0.019	0.086	0.134	0.060	0.046	0.040	0.172
R^2		0.201	0.327	0.254	0.180	0.283	0.137	0.239	0.197	0.582

Table reports results of regressing population growth for a given period on population growth from the preceeding period. Listed years are the start of the initial period. The second panel allows for different slope coefficients for counties that grew slower and faster than the aggregate rate. Standard errors are listed in parentheses. All coefficients differ from zero at the 0.01 level. N is the number of counties in each category. The marginal R-squared rows give the decrease in the a regressions R-squared value from constraining the coefficient on the lagged growth rates for the listed category to be zero.

Table 2: Persistence of Growth Rates

4 Model

We match the observed empirical behavior with a relatively simple, one-sector general equilibrium model of a system of locations transitioning towards a long run steady state. During the first part of the transition, new locations are rapidly entering the system. The transition is prolonged by a friction that dampens local total factor productivity as an increasing function of local population growth. More specifically, at an arbitrary starting year a small portion of potential locations is assumed to be active. Population is freely mobile so that wages equalize across locations. The endogenous partition of aggregate population across the *initially-active* locations is assumed to be frictionless. With an unchanged productivity distribution, no change in aggregate population and no entry of new locations, the initial partition would be a steady state. However, new locations are assumed to enter, aggregate population grows, and agglomeration economies change the productivity distribution. The entering of new locations, together with the friction, significantly lengthens the time it takes for the economy to reach its steady-state. In the long run the relative productivity of locations ceases to change, and the economy converges to Gibrat's law.

4.1 Endowments and Locations

At time t the economy is endowed with a total population of L_t which grows at an exogenous rate ℓ, so that

$$L_{t+1} = (1+\ell)L_t \tag{1}$$

Each agent supplies 1 unit of labor to the economy. The economy consists of N potential locations, indexed by i, each endowed with D units of land. All individuals, independently of their location, own an equal share of all land. At time t, a number $N_t \leq N$ of locations is *active*. The number of active locations increases exogenously over time until all N locations are active:

$$N_{t+1} = \begin{cases} N_t(1+\delta(\frac{N-N_t}{N})) & \text{if } N_t(1+\delta(\frac{N-N_t}{N})) < N \\ N & \text{else} \end{cases} \tag{2}$$

4.2 Technology

Locations exogenously only differ with respect to total factor productivity, $Z_{i,t}$. Local total factor productivity, in turn, derives from the multiplicative combination of a once-and-for-all TFP draw, Z_i^0, a location-specific friction that depends on a location's population growth, $f((L_{i,t} - L_{i,t-1})/L_{i,t-1})$ and on local agglomeration economies, $g(L_{i,t})$. Thus, at any point in time, local productivity of location i is given by

$$Z_{i,t} = Z_i^0 \cdot f((L_{i,t} - L_{i,t-1})/L_{i,t-1}) \cdot g(L_{i,t}) \tag{3}$$

The permanent component Z_i^0 is drawn upon activation of the location from a lognormal distribution. The friction factor, $f(\cdot)$, is equal to 1 below a specified population threshold, and then increasing in population growth:

$$f(\frac{L_{i,t} - L_{i,t-1}}{L_{i,t-1}}) = \begin{cases} 1 - \xi_1(\frac{L_{i,t} - \bar{L}_{i,t-1}}{\bar{L}_{i,t-1}})^{\xi_2} & \text{if } L_{i,t} > \bar{L} \\ 1 & \text{else} \end{cases} \tag{4}$$

Note that the frictions coming from population growth in a given location affect the productivity of the entire population of that location, and not just that of the new population. By modeling frictions this way, all agents in a given location will earn the same wage. This, together with free mobility, will imply that wages will equalize across all active locations in all time periods. The agglomeration factor, $g(L_{i,t})$, is assumed to be weakly increasing in a location's population, and takes the following functional form

$$g(L_{i,t}) = \min[\max(\varepsilon_1 L_{i,t}^{\varepsilon_2}, \underline{Z}), 1] \tag{5}$$

where $\underline{Z} < 1$. This implies that agglomeration economies kick in above a certain floor (when $\varepsilon_1 L_{i,t}^{\varepsilon_2} > \underline{Z}$) and that they are limited from above at 1. Taking into account frictions and agglomeration economies, Z_i^0 corresponds to the highest possible TFP level of location i.

Production is constant returns to scale. There is a measure one of firms in each location. All firms in all locations employ an identical production technology, except for the variation in location-specific productivity. Production of the stand-in firm in location i and time t is:

$$Y_{i,t} = Z_{i,t} \cdot L_{i,t}^{1-\alpha} \cdot D^{\alpha},$$

where $Y_{i,t}$, $L_{i,t}$, and D denote output, population, and land of the stand-in firm in location i and time t. The profit of the stand-in firm is

$$\Pi_{i,t} = Z_{i,t} \cdot L_{i,t}^{1-\alpha} \cdot D^{\alpha} - w_{i,t} L_{i,t}$$

where $w_{i,t}$ is the wage rate. The standard first-order condition of the profit maximization problem yields

$$w_{i,t} = (1-\alpha) Z_{i,t} \cdot (\frac{D}{L_{i,t}})^{\alpha} \tag{6}$$

As mentioned before, land is collectively owned by all individuals in the aggregate economy. Thus all factor payments to land are rebated lump-sum to individuals, irrespective of where they live. An assumption of absentee landlords would imply nearly identical results. The share of factor income accruing to land, α, is one of the model's critical parameters. With a high land factor share, the population tends to be more equally spread across locations, making the distribution more compressed. With a high land factor share, the low population density in low-productivity locations keeps those locations from becoming too small, whereas the high population density in high-productivity locations keeps those locations from becoming too large. Conversely, when land accounts for a smaller share of factor income, its relative abundance in low productivity locations is insufficiently valued to attract many residents. Similarly, land is a less effective congestion mechanism dampening the population of high productivity locations. The exclusion of reproducible capital from the production function is done to help keep the model simple. For a given land factor share, the inclusion of capital would not meaningfully change results.

Goods are freely tradable across locations. Given that population density differs across locations, the equal distribution of land rents across all individuals implies that there will be trade.

4.3 Preferences

In our simple framework preferences need not be explicitly defined. This reflects a combination of a single consumption good, the lack of a savings technology, homogeneous agents, and (implicitly) no variation in quality of life across locations. As a result, maximizing utility is the same as maximizing wages. In addition, since individuals are perfectly mobile across locations, wages will always be equalized across all locations, so that individuals only need to maximize their wage, period-by-period. The presence of frictions coming from population growth does not change this: since we model a location's frictions as affecting all individuals of that location through a reduction in TFP, free mobility indeed implies that all individuals in all locations will have the same wage in each time period.

4.4 Equilibrium

Although agents and firms are forward-looking, their individual decisions today do not affect their future decisions, so that the dynamic problem simplifies to a sequence of static problems. This means that there is no advantage to being in one place or another today. Hence, the dynamic equilibrium for the model economy is essentially a sequence of static equilibria that are linked through the laws of motion for the population and the number of locations. We now define the dynamic equilibrium

Definition of Equilibrium *An equilibrium is a sequence of variables L_t, N_t, $w_{i,t}$, $\Pi_{i,t}$, $Z_{i,t}$ for each location i and time period t such that*

1. *The number of active locations and the population grow according to (1) and (2).*
2. *Agents are indifferent between residing in one active location or another, so that wages equalize across all active locations.*
3. *Technology evolves according to (3), (4) and (5)*
4. *Firms maximize profits, so that (6) holds.*
5. *Goods market clear.*

This equilibrium converges to a steady state in which the population growth is equal across all locations.

Proposition *The economy converges to a steady state with constant growth in population across all locations.*

Proof: The proof is trivial. In the long-run the number of locations becomes constant. Following (3), (4) and (5) , productivity in all locations reaches a steady state level (net of the friction term). In then follows that population growth will be equal across locations. ∎

5 Numerical Results

In this section we calibrate the theoretical model, and show that it is able to account for the settlement patterns of the U.S. over the last 200 years.

symbol	interpretation	value	source	sensitivity of growth vs population
α	land share of factor income	0.15	arbitrary	? for fxd σ_{Z_i} †
$F(Z_i)$	distribution of tfp	lognormal	Eeckhout (2004)	?
$\mu_{\log(Z\_i)}$	asymptotic mean log(tfp)	0		none
$\sigma_{\log(Z\_i)}$	asymptotic std dev of log(tfp)	$\alpha \cdot \sigma_{\log(L\_i),1790}$	analytical	?
ξ^1	growth friction, level*	0.935	calibrated	moderate
ξ2	growth fricition, convexity	1	calibrated	moderate
$L_{entry,t}$	pre-entry population	$\min(L_{i,t})$	arbitrary	moderate
L_t	aggregate population	U.S. aggrgt, intrpltd	decenial census	
N_t	number of active locations	U.S. aggrgt, intrpltd	decenial census	

† Under the current calibration, what matters most for the growth versus population correlation is the quotient sigma $\sigma_{\log(Z\_i)}/\alpha$, which is assumed to always equal $\sigma_{\log(L\_i),1790}$. The specific value of either component is second order.

Table 3: Parameter Values

5.1 Calibration

The model depends on a few key parameter choices (Table 3). Following Eeckhout (2004), we assume total factor productivity is distributed log normally across locations. We set the land share of factor income, α, to be 0.15. This value is extremely high relative to estimates of land's share of factor income from manufacturing circa 2000. However, traded production, primarily agricultural, was obviously much more land intensive in 1790. Moreover, housing, which has a much higher land share, should be taken account to accurately gauge the congestion arising from a fixed land supply. For present purposes, α is largely irrelevant since it only matters in conjunction with the variance parameter governing the distribution of TFP across locations. For a shared wage level across locations, $\log(L_i)$ equals $\frac{w}{\alpha \log(Z_i)}$. So to match the empirical population variance, σ_D^2, $\mathrm{Var}(\log(Z_i))$ can be set to $\alpha^2 \mathrm{Var}(\log(L_i))$. Thus the higher the assumed land share, the higher the assumed TFP variance.

Two parameters that do have considerable influence on transitional dynamics are the level and convexity of the population growth friction, ξ_1 and ξ_2 in (4). For better intuition and without loss of generality, ξ_1 is mapped to $\widehat{\xi}_1$, which is defined as the multiplier on underlying productivity Z_i^0 at a growth rate of 4 percent. A grid search of combinations of $\widehat{\xi}_1$ and ξ_2 is done based on the criteria of minimizing a weighted root mean square error between the empirical distribution of simulated and empirical growth rates. The comparison included the growth distribution across all active locations during a given multi-year period as well as for each of three age groups, "young", "intermediate" and "old". Additionally a constraint was imposed that ξ_2 must be at least one. The combination that minimized the weighted root mean square error was $\widehat{\xi}_1 = 0.94$ and $\xi_2 = 1$. In other words, at a population growth rate of 4 percent, a location's TFP was "penalized" by 6 percent. This friction increased linearly so that at a growth rate of 8 percent, the TFP was penalized by 12 percent. The root mean square error function that this parameterization minimized was relatively similar across alternative weights. And transition dynamics was relatively similar across alternative choices for $\widehat{\xi}_1$ and ξ_2.

Another important parameterization choice is the minimum population, $\bar{L}$, at which the friction kicks in. We set this equal to the lowest-population location in the period prior to their entry. Thus, if in 1790 the minimum population across counties was 1189, a county entering in 1791 would experience a loss of TFP due to the friction as its population rises above 1189. In other words, a county entering in 1791 would have its population adjust to the level at which its wage becomes equal to that of the rest of the economy, but any positive growth between the frictionless level of 1189 and its actual population would be subject to growth frictions. Allowing entry at higher population levels moderately dampens the negative correlation of population growth with initial population. The reason is that low population counties are made up of rapidly growing entrants and stagnant low TFP veterans.

5.2 Level Fit

A comparison of the simulated and empirical log population distributions for four different years shows the simulated and empirical to be a relatively tight match (Table 4). The simulated results are averages from 50 separate runs of the model, each with a different stochastic seed determining the simulated productivity draws. The empirical results are based on non-metro builds of the data, which is appropriate for level comparisons.[5]

The exact matching of aggregate and mean population (reported in logs) is by construction. More substantive is that at the 5th percentile and above, the simulated distribution for 1790 remains within 0.2 log points (22 percent) of the empirical distribution. From the 20th percentile to the 99th percentile, it remains within 0.1 log points of the empirical distribution. Not as good a fit is the full log point by which the minimum simulated population exceeds the minimum empirical population. But such a difference of this size at the far tails is relatively common in exercises of this sort, particularly when the simulation is averaged across a number of independent runs. The generally close matching of the simulated and empirical distributions reflects the ability of a log normal distribution to describe the empirical distribution. The link is that the 1790 simulated distribution is log normal by construction, being a multiplicative transformation of the assumed lognormal TFP distribution.

The simulated population distribution closely matches the empirical distribution through 1880. In that year, the simulated distribution remains within 0.2 log points of the empirical one from the 10th through the 95th percentiles. But during the twentieth century, the simulated and empirical distributions diverge. In particular, from the minimum population county up through the 80th percentile county, populations are at least moderately higher in the simulated ones. And at the very right of the population distribution, simulated populations are a fraction of their empirical counterparts.

[5] Alternatively, comparing the singleton simulated locations with metro combinations of what are typically already high-population counties leads to a significant mismatch between the upper portions of the two distributions.

	1790		1880		1920		2000	
	sim	**emp***	**sim**	**emp***	**sim**	**emp***	**sim**	**emp***
locations	232	232	2,397	2,397	3,014	3,014	3,067	3,067
log(aggregate pop)	15.2	15.2	17.7	17.7	18.5	18.4	19.4	19.4
log(mean pop)	9.7	9.7	9.9	9.9	10.5	10.4	11.4	11.4
log(s.d. pop)	9.7	9.7	10.0	10.7	10.5	11.1	11.4	12.2
log(median pop)	9.4	9.4	9.5	9.5	10.1	9.8	11.1	10.1
log(min pop)	7.0	6.0	7.1	1.1	7.4	3.6	8.2	4.2
log(pop.pctl.01)	7.5	7.2	7.5	3.2	7.9	7.0	9.2	7.0
log(pop.pctl.05)	8.0	8.2	7.8	6.8	8.5	8.1	9.7	8.1
log(pop.pctl.10)	8.3	8.5	8.1	7.9	8.9	8.6	10.0	8.6
log(pop.pctl.20)	8.7	8.7	8.6	8.6	9.3	9.1	10.4	9.1
log(pop.pctl.25)	8.8	8.8	8.8	8.8	9.5	9.2	10.5	9.3
log(pop.pctl.30)	9.0	9.0	9.0	9.0	9.6	9.4	10.6	9.5
log(pop.pctl.40)	9.2	9.1	9.3	9.3	9.9	9.6	10.9	9.8
log(pop.pctl.50)	9.4	9.4	9.5	9.5	10.1	9.8	11.1	10.1
log(pop.pctl.60)	9.6	9.6	9.8	9.7	10.3	10.0	11.3	10.5
log(pop.pctl.70)	9.8	9.8	10.0	9.9	10.6	10.2	11.5	10.8
log(pop.pctl.75)	10.0	9.9	10.2	10.0	10.7	10.3	11.6	11.0
log(pop.pctl.80)	10.1	10.1	10.3	10.2	10.8	10.4	11.8	11.3
log(pop.pctl.90)	10.4	10.5	10.7	10.5	11.2	10.9	12.1	12.1
log(pop.pctl.95)	10.7	10.8	11.0	10.9	11.5	11.4	12.4	12.8
log(pop.pctl.99)	11.3	11.1	11.6	11.9	12.1	12.8	13.0	14.0
log(max pop)	11.7	11.9	12.6	14.0	13.1	14.9	14.0	16.1

* non-metro build

Table 4: Level Fit

The main reason for this deterioration in match quality is the constant assumed land factor share in the simulation. Whatever the land factor share was in 1790, it almost surely declined significantly over the ensuing 210 year. A declining land share causes high productivity counties to grow quicker by relaxing the model's only source of long-run congestion. Conversely, a declining land share causes low productivity counties to lose population by reducing the input value of their abundant land. The model abstains from including a declining land share. In its one-sector framework, steady-state county populations are determined solely by underlying TFP. Hence a declining land share would be a significant force for divergence. In other words, essentially by assumption there would be a strong positive correlation between size and growth. In a richer model this need not be so. For example, a declining land share might correspond to the introduction of a new product or technology with a TFP draw across locations orthogonal to the TFP draw underlying current production. Consistent with this, the simulated log population distribution for 2000 fits its empirical counterpart relatively tightly when the assumed land share is approximately half its 1790 level.

5.3 Growth Fit

The match between simulated and empirical growth rates is also fairly good (Table 5). As with the levels above, the reported growth rates are the averages from 50 runs of the simulation. Through the middle percentiles of the 1820-40, 1860-80, and 1900-20 periods, simulated growth rates remained relatively close to empirical ones. For example, simulated growth rates for 1820-40 remained within 1 percentage point of their empirical counterparts from the 10th through the 70th percentiles; they remained within 1.3 percentage points from the 5th through the 95th percentiles.

All Locations*	**1820-40**			**1860-80**			**1900-20**			**1940-60**		
	sim	**emp**	**dif**	**sim**	**emp**	**dif**	**sim**	**emp**	**dif**	**sim**	**emp**	**dif**
Locations	545	544		1,707	1,692		2,696	2,655		3,064	2,982	
Agg (all)	2.9	2.9	(0.0)	2.1	2.1	(0.0)	1.7	1.7	(-0.0)	1.5	1.5	(-0.0)
Agg (existing)	2.1	2.0	(0.0)	1.6	2.2	(-0.6)	1.5	1.6	(-0.1)	1.5	1.5	(-0.0)
Mean	2.8	2.3	(0.5)	2.5	3.4	(-1.0)	1.7	1.3	(0.4)	1.5	0.2	(1.3)
Std.Dev.	3.0	3.2	(-0.2)	3.3	4.7	(-1.4)	1.7	2.5	(-0.8)	0.6	1.8	(-1.2)
Median	1.7	1.4	(0.3)	1.2	2.0	(-0.7)	1.0	0.7	(0.3)	1.3	0.0	(1.3)
Minimum	0.4	-3.8	(4.2)	-0.7	-3.6	(3.0)	0.8	-8.0	(8.8)	1.3	-4.8	(6.1)
Pctile 01	0.5	-2.8	(3.2)	-0.6	-0.6	(0.1)	0.8	-1.8	(2.6)	1.3	-3.1	(4.4)
Pctile 05	0.5	-0.9	(1.3)	-0.6	0.0	(-0.6)	0.8	-1.0	(1.9)	1.3	-2.2	(3.5)
Pctile 10	0.5	-0.3	(0.7)	-0.6	0.4	(-0.9)	0.8	-0.7	(1.5)	1.3	-1.7	(3.1)
Pctile 20	0.5	0.2	(0.2)	-0.5	0.8	(-1.3)	0.8	-0.3	(1.1)	1.3	-1.2	(2.5)
Pctile 25	0.5	0.4	(0.0)	-0.4	1.0	(-1.4)	0.8	-0.2	(1.0)	1.3	-1.0	(2.3)
Pctile 30	0.5	0.6	(-0.2)	-0.2	1.2	(-1.4)	0.8	0.0	(0.9)	1.3	-0.8	(2.1)
Pctile 40	0.5	1.0	(-0.6)	0.3	1.6	(-1.3)	0.9	0.3	(0.6)	1.3	-0.4	(1.7)
Pctile 50	1.7	1.4	(0.3)	1.2	2.0	(-0.7)	1.0	0.7	(0.3)	1.3	0.0	(1.3)
Pctile 60	2.6	2.0	(0.6)	2.4	2.4	(-0.0)	1.1	1.1	(0.0)	1.3	0.4	(0.9)
Pctile 70	3.8	2.9	(0.9)	3.7	3.2	(0.5)	1.5	1.6	(-0.1)	1.4	0.8	(0.6)
Pctile 75	4.5	3.4	(1.1)	4.5	3.7	(0.8)	1.8	2.0	(-0.2)	1.4	1.1	(0.3)
Pctile 80	5.3	4.0	(1.2)	5.4	4.5	(0.9)	2.2	2.4	(-0.3)	1.4	1.5	(-0.1)
Pctile 90	7.4	6.3	(1.1)	7.5	8.0	(-0.5)	3.5	4.0	(-0.5)	1.9	2.5	(-0.6)
Pctile 95	9.0	8.3	(0.7)	9.1	12.9	(-3.8)	5.2	5.8	(-0.6)	2.7	3.6	(-1.0)
Pctile 99	11.9	14.4	(-2.4)	12.0	24.9	(-12.9)	9.7	10.8	(-1.1)	4.5	6.2	(-1.7)
Maximum	14.9	21.4	(-6.4)	15.8	35.2	(-19.4)	15.6	28.2	(-12.6)	10.6	11.8	(-1.2)

* empirics are based on metro build

Table 5: Growth Fit

For these same three time periods, the fit at the tails is considerably worse. At the lower tail, for example, 1st percentile simulated growth in 1820-40 was 3.2 percentage points above its empirical value. The difference for minimum growth ranged from 3 percentage points in 1860-80 to almost 9 percentage points in 1900-20. This inability of the model to match what are typically negative empirical growth rates is structural. In the model, counties' stochastic TFP values are once and for all. Counties are also assigned a pre-entry population value, which in the current parameterization is the minimum population of already active counties in the year prior to entry. This pre-entry population serves as the denominator of an implicit growth rate with the actual entry population, which is determined endogenously. More specifically the same friction function, (4), applied to active locations' growth, is also applied to entering implicit growth. For almost all counties, local TFP will be sufficient to imply a steady-state population above the entry level. But for counties with extremely low TFP, steady-state population may be below the pre-entry assumption. Because there are no frictions on population decline, such counties will enter at what are their more-or-less steady-state populations. The absence of downward growth frictions and of location-specific post-entry shocks does away with most negative growth.

Of course, in the real world county populations do decline, often over extended periods. It is these empirically declining counties —by definition located at the left tail— that the model can not match. Doing so can be accomplished in at least three ways. First would be to allow for a period-by-period county-specific stochastic shock. When a county received a large negative such shock, it would immediately jump down to its lower population steady state.[6] Unlike the case of entry, this decline would cause measured growth to be negative. But there would be no persistence. Second, as described in the previous subsection, would be if the land-factor share of income were gradually declining. In this case, low TFP counties would also gradually decline over time. Third would be to move to a two-sector model in which the TFP of a first sector was persistently declining relative to the TFP of a second sector. Depending on relative elasticities of supply and demand, counties that had high TFP in the first sector would persistently grow or decline relative to counties with high TFP in the second sector.

In the present model, the *only* way county populations decline is if counties on the upward transition to their steady state together "steal" enough population from counties already near their steady states such that the latter shrink. More precisely, the latter group of counties can be thought of as staying near location-specific steady-state populations that are themselves declining. Since this stealing is typically from a large number of counties with high aggregate population, the implied numerical declines are typically small. Moreover, aggregate population growth helps mitigate the need for any local population decline. For the 1860-80 time period, simulated population decline diminishes from -0.7 percent for the slowest county down to -0.2 percent for the simulated 30th percentile county.

Conversely, at the upper tail simulated growth rates fall considerably short of empirical growth rates. For example, in the 1860-80 regression the 99th percentile simulated growth rate trails the empirical 99th percentile growth by almost 13 percentage points. The maximum simulated growth rate does so by more than 19 percentile points. Given the smooth way in which the frictions dampen growth, this right-tail discord is inevitable. Fortunately, the discord really only applies above the 95th percentile.

The fit between simulated and empirical growth rates is less tight for the 1940-60 regressions. For example, the range over which simulated growth stays within 1 percentage point of empirical growth narrows to the 60th-90th percentiles. The difference between the simulated and empirical median growth rises to

[6] "Steady state" as used herein should be interpreted as varying over time, both due to the transition of other counties to their time-dependent steady states as well as due to entry. In contrast, the steady state of the system as a whole should be interpreted as the system's final distribution of population across counties if there were no future entry or parameter changes.

1.3 percentage points.[7] Also notice that simulated growth rates are relatively flat across all but the top 10th percentile of the 1940-60 distribution. The reason is that by 1940, most of the simulated counties have approximately attained their steady-state population.

	(2) 1800-20 on 1790-1800	**(3)** 1820-40 on 1800-20	**(4)** 1840-60 on 1820-40	**(5)** 1860-80 on 1840-60	**(6)** 1880-1900 on 1860-80	**(7)** 1900-20 on 1880-1900	**(8)** 1920-40 on 1900-1920	**(9)** 1940-60 on 1920-1940	**(10)** 1960-80 on 1940-1960	**(11)** 1980-2000 on 1960-80
Simple										
N	232	310	545	870	1707	2397	2696	3014	3064	3067
ρ	-0.447	0.233	0.221	0.245	0.201	0.225	0.219	0.219	0.203	0.197
	(0.060)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.198	0.994	0.991	0.997	0.971	0.991	0.989	0.992	0.995	0.996
By age										
N_{young}	all	78	313	560	1162	1527	989	617	368	to fill
ρ_{young}	locs	0.266	0.235	0.249	0.221	0.231	0.229	0.227	0.207	in
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	0.00	
mrgnl R^2_{young}	are old	0.237	0.588	0.712	0.749	0.773	0.691	0.679	0.719	
N_{int}				78	313	560	1162	1527	989	
ρ_{int}				0.222	0.112	0.179	0.138	0.096	0.149	
mrgnl R^2_{int}				0.014	0.015	0.020	0.009	0.001	0.005	
N_{old}		232	232	232	232	310	545	870	1707	
ρ_{old}		-0.121	0.162	0.221	0.103	0.179	0.144	0.107	0.149	
mrgnl R^2_{old}		0.001	0.032	0.044	0.012	0.014	0.009	0.002	0.007	
R^2		0.998	0.995	0.998	0.988	0.993	0.993	0.994	0.996	
By init grwth										
N_{below}	no	233	299	524	917	1,524	1,916	2,405	2,606	2656
ρ_{below}	lagged	-0.128	0.144	0.204	0.065	0.137	0.123	0.090	0.121	0.116
		(0.013)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)
mrgnl R^2_{below}	var.	0.001	0.019	0.027	0.005	0.010	0.007	0.001	0.003	0.003
N_{above}		77	246	346	790	873	780	609	458	411
ρ_{above}		0.267	0.240	0.254	0.240	0.240	0.232	0.228	0.209	0.200
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
mrgnl R^2_{above}		0.233	0.492	0.487	0.555	0.554	0.641	0.658	0.626	0.773
R^2		0.998	0.996	0.998	0.995	0.995	0.994	0.994	0.997	0.997

Table 6: Simulated Persistence in Growth Rates

[7] The increase in the difference between simulated and empirical *mean* growth is less informative. Its main source is the difference in the number of locations due to the combining of empirical counties to match metro-area formation. Indeed, starting in 1930 the number of empirical locations declines as metro areas subsume more and more counties. The main purpose of combining empirical counties into metro areas is to better match labor markets in an initial year with their subsequent growth. There is no obvious need or means to do this in the simulation.

5.4 Persistence

The simulated system of locations is characterized by extremely persistent growth (Table 6). Regressing population growth on lagged population growth beginning with 1820-40 (on 1800-20) through 1980-2000 always yields coefficient values of approximately 0.20. More importantly for present purposes, the regressions account for most of the variation in growth. R-squared values range from 0.971 to 0.997.

The ability of lagged growth to accurately reflect current growth can primarily be attributed to "young" locations. Allowing for different coefficients for each of three age groups and then constraining this regression to have a zero coefficient on young locations' lagged growth lowers R-squared from its non-constrained value by 24 percentage points for 1820-40 and by at least 59 percentage points for all other years. Correspondingly constraining the the intermediate and old coefficients causes marginal R-squared contributions of no more than 4 percentage points. Put differently, growth by the intermediate and old locations can mostly be captured by a constant. The fitted growth of young locations, in contrast, benefits greatly from taking account of lagged growth.

Unsurprisingly given the assumed positive growth frictions, the persistence is almost entirely driven by locations experiencing positive population growth. The bottom panel in Table 6 allows for different coefficients according to whether a location's lagged growth exceeded or fell short of aggregate population growth (of all locations active at the start of the lagged period). A comparison from constraining each of the below-average and above-average lagged growth coefficients with the unconstrained regression finds a marginal R-squared of 23 to 66 percentage points associated with the above-average locations but of no more than 3 percentage points from the below average locations. Similar to the division by age, the fitted growth of locations growing faster than average greatly improves when taking account of lagged growth.

5.5 The Simulated Emergence of Gibrat's Law: Growth versus Initial Population

Beginning with the 1800-20 cross-section, the simulated system of locations nicely matches the negative empirical correlation between growth and initial population size. It continues to do so through the 1920-40 cross-section (Table 7). This match is characterized by a quantitatively large (in absolute value) negative coefficient on initial population for the smaller population bins. More specifically, the negative correlation between simulated population and growth typically extends up to at least a log population of 10 (population of 22,000). Coefficients on the corresponding log population bins ranged from -0.004 to -0.059. In other words, a 1 percent increase in initial population is associated with slower population growth of 0.004 to 0.059 percentage points. The negative correlation of population growth with initial population size also generally holds for the population bins ranging above 10. But the magnitudes are smaller. The maximum coefficient (in absolute value) for these bins is -0.006.

The simulated coefficients are very similar to their 1800-20 to 1920-40 empirical counterparts. For log population up to 10, the statistically significant empirical coefficients range from -0.007 to -0.059, almost exactly the same as the simulated range. Above a log population of 10 all statistically-significant empirical coefficients are positive, ranging from 0.006 to 0.013. This empirical divergence of population at high levels cannot be matched by the model in its current form. To do so would require introducing allow TFP to increase with population size above some moderately high threshold. Doing so is a priority for future research.

All Locations	(1) 1790 -1800	(2) 1800 -1820	(3) 1820-1840	(4) 1840-1860	(5) 1860-1880	(6) 1880-1900	(7) 1900-1920	(8) 1920-1940	(9) 1940-1960	(10) 1960-1980	(11) 1980-2000
log(pop) bin:											
min to lowest lb	0.000	-0.029	-0.032	-0.032	-0.037	-0.035	-0.059	-0.040	-0.001	0.000	0.000
lpop.08to09		-0.046	-0.032	-0.036	-0.033	-0.041	-0.023	-0.040			
lpop.09to10	0.000	-0.005	-0.014	-0.013	-0.020	-0.010	-0.004	-0.010	0.000	0.000	0.000
lpop.10to11		0.001	-0.006	-0.004	-0.005	-0.001	-0.003	0.000	-0.001	0.000	0.000
lpop.11to12							-0.001	0.000	-0.001	0.000	0.000
lpop.12to13											0.000
highest ub to max	0.000	-0.001	-0.003	-0.002	-0.002	-0.003	0.000	0.000	0.000	0.000	0.000
Seeds	50	50	50	50	50	50	50	50	50	50	50
N/Seed	232	310	545	870	1,707	2,397	2,696	3,014	3,064	3,067	3,067
Bins	3	5	5	5	5	5	6	6	5	5	6
R^2	0.001	0.521	0.355	0.348	0.393	0.422	0.199	0.322	0.007	0.001	0.005
R^2 across 50 Seeds											
mean		0.513	0.342	0.337	0.383	0.413	0.196	0.320	0.009	0.002	0.006
min		0.303	0.152	0.153	0.173	0.230	0.085	0.219	0.003	0.000	0.003
max		0.728	0.519	0.564	0.572	0.613	0.328	0.481	0.023	0.004	0.013

Table shows results from regressing each locations' average annual population growth rate for listed time period on a spline of initial population. The spline is constructed to be continuous with respect to population. Top results row is coefficient for the lowest bin, whatever it may be. It can be immediately inferred from the lower bound of the next highest bin for which a coefficient is reported. Results are from 50 independent runs of the simulation, each using a different stochastic seed. Each location's initial population and growth rate are stacked on top of each other. Coefficients can thus be interpreted as an approximate average of coefficients from each

Table 7: Simulated Growth on Initial Population

As is the case empirically, an initial population spline accounts for a large share of the variation in location growth rates. For the 1800-1820 through 1920-1940 cross-sections, the mean R-squared value across the 50 stochastic seeds ranged from a low of 0.196 for 1900-20 to a high of 0.513 for 1800 to 1820 (third panel of Table 8). For some seeds in some years, R-squared values ran as high as 0.728. In contrast, initial population accounts for almost no variation in simulated growth from 1790 to 1800. This reflects that for this cross section, there is no variation to account for. The system of locations in 1790 is essentially at its steady state because, by assumption, these initial locations enter the system with no frictions limiting the distribution of aggregate population. Changes in these locations' population over the first decade come strictly from aggregate population growth, partly offset from the "pull" of population away from them from entering locations. As a result, growth is identical across locations regardless of size. Similarly, simulated R-squared values are close to zero for the 1940-60 through 1980-2000 regressions. This reflects that by about

1940, the simulated system of locations is rapidly approaching its steady state.[8]

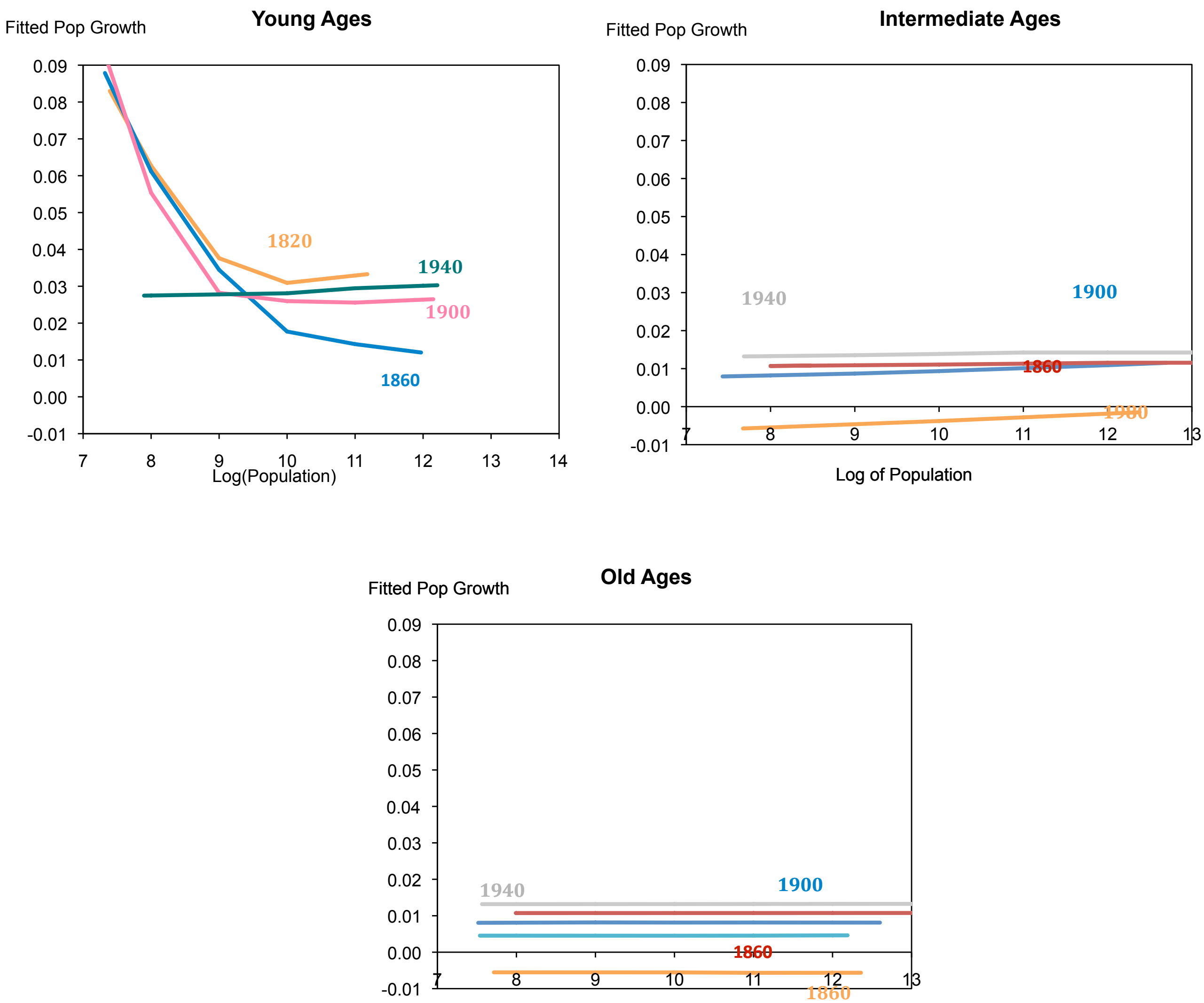

Figure 9: Simulated Growth of Young, Middle Aged and Old Counties

The decline of population growth as initial population size increases is driven primarily by young locations (Figure 9). For the 1820-40 through the 1920-40 cross sections, the coefficients on log population up to 10 ranged from -0.025 to -0.055. In other words, a 1 percent increase in a young location's population was associated with a 0.025 to 0.055 percentage point of faster growth. As above, these are relatively large values considering that a 1 percent increase at log population equal to 10 is just 220. R-squared values are similar to those for the all locations simulated regressions.

[8]As shown in Table 5, about 10 percent of locations during the 1940-60 cross section are growing significantly faster than remaining observations. The fastest growing location over this period, grew by 9 percentage points faster than average. By 1980-2000, only the top 1 percent of locations are growing faster than remaining, the fastest by about a percentage point.

Similar regressions for intermediate and old locations show growth to be essentially orthogonal to initial population. For intermediate aged locations (active for 40 to 79 years), regressions for the 1860-80 through 1980-2000 cross-sections admit coefficients no larger than 0.001 in absolute value. None of these coefficients statistically differs from zero at the 0.05 level. For old locations (active 80 or more years), the range of coefficient values is even closer to zero.

6 Concluding Remarks

This paper studies the long run spatial development of U.S. counties and metro areas between 1800 and 2000. It shows that the often-documented orthogonality between population growth and initial population size — Gibrat's law — can be rejected across the entire two centuries. However, the economy gradually approaches Gibrat's law over time. This is especially true for counties that date back to the early 19th century. For that subset of counties we can no longer reject Gibrat's law in recent decades. A simple one-sector model with entry and frictions can be shown to capture these dynamics. Our findings on Gibrat's law have clear implications for how the cross-sectional distribution of population evolved over time. We study the dynamics of the emergence of Zipf's law and the lognormal distribution of population in a companion paper.

References

[1] Desmet, K. and Fafchamps, M., 2006. “Employment concentration across U.S. counties,” *Regional Science and Urban Economics*, 36, 482-509.

[2] Desmet, K. and Rossi-Hansberg, E., 2009. “Spatial growth and industry age,” *Journal of Economic Theory*, 144, 2477-2502.

[3] Eeckhout, J., 2004. “Gibrat’s Law for (All) Cities,” *American Economic Review*, 94, 1429-1451.

[4] Gardner, T., 1999. “Metropolitan Classification for Census Years before World War II”, *Historical Methods*, 32, 3, 139–150.

[5] Glaeser, E.L. and Gyourko, J. 2005. “Urban Decline and Durable Housing,” *Journal of Political Economy*, 113, 345-375.

[6] Haines, M.R., 2005. “Historical, Demographic, Economic, and Social Data: The United States 1790-2002” [Computer File]. ICPSR Study 2896. Inter-university Consortium for Political and Social Research [distributor].

[7] Horan, P.M. and Hargis, P.G., 1995. “County Longitudinal Template,1840–1990.” [computer file]. ICPSR Study 6576. Inter-university Consortium for Political and Social Research [distributor]. Corrected and amended by Patricia E. Beeson and David N. DeJong, Department of Economics, University of Pittsburgh, 2001. Corrected and amended by Jordan Rappaport, Federal Reserve Bank of Kansas City, 2010.

[8] Holmes, T.J. and Lee, S., 2010. “Cities as Six-by-Six-Mile Squares: Zipf’s Law?,” NBER Chapters, in: *Agglomeration Economics*, 105-131, National Bureau of Economic Research.

[9] Michaels, G., Rauch, F. and Redding, S., 2010. “Urbanization and Structural Transformation,” unpublished manuscript, Princeton University.

[10] Rappaport, J. (2004). “Why Are Population Flows So Persistent?” *Journal of Urban Economics*, 56, 554-580.