THE
QUARTERLY JOURNAL
OF ECONOMICS

Vol. 127 May 2012 Issue 2

URBANIZATION AND STRUCTURAL TRANSFORMATION*

Guy Michaels
Ferdinand Rauch
Stephen J. Redding

We examine urbanization using new data that allow us to track the evolution of population in rural and urban areas in the United States from 1880 to 2000. We find a positive correlation between initial population density and subsequent population growth for intermediate densities, which increases the dispersion of the population density distribution over time. We use theory and empirical evidence to show this pattern of population growth is the result of differences in agriculture's initial share of employment across population densities, combined with structural transformation that shifts employment away from agriculture. *JEL* Codes: N10, O18, R11, R12.

I. Introduction

Urbanization—the concentration of population in cities and towns—is a key feature of economic development. The share of the world's population living in cities grew from less than one tenth in 1300 to around one sixth in 1900 and to more than one half today.[1] We examine urbanization using a new data set for

*Financial support for this paper was provided by the Globalization Programme of the Centre for Economic Performance (CEP), which is partly funded by the Economic and Social Research Council (ESRC). We thank Yu-Hsiang Lei, Ralph Ossa, Tomasz Swiecki, and Bart Thia for excellent research assistance. We are also grateful to the editor, four anonymous referees, Robin Burgess, Francesco Caselli, Tim Guinnane, Tom Holmes, Alan Manning, Rachel Ngai, Steve Pischke, Chris Pissarides, Richard Rogerson, Esteban Rossi-Hansberg, and seminar participants at AEA, CEPR, Helsinki, International Growth Center (IGC), LSE, Oxford, Tel Aviv University, Tinbergen Institute, University of Minnesota, University of Missouri, and Zurich for helpful comments. Thanks also to Gilles Duranton and Matt Turner for sharing the GIS roads and railroads data. Responsibility for any opinions, results, and errors lies with the authors alone.

1. The historical figures are from Bairoch (1988) and the modern ones are from United Nations (2008).

The Quarterly Journal of Economics (2012) 127, 535–586. doi:10.1093/qje/qjs003.
Advance Access publication on March 15, 2012.

the United States that tracks, for the first time, the evolution of population across both rural and urban areas from 1880 to 2000. We show that incorporating information on the full range of population densities considerably changes our understanding of the urbanization process. We provide theory and evidence that structural transformation away from agriculture explains our findings and demonstrate similar results for Brazil from 1970 to 2000. As other developing countries continue to undergo structural transformation away from agriculture, our findings suggest that it will exert a powerful influence on future patterns of urbanization.

Although most previous research on the population distribution has used data on cities, rural areas historically accounted for a large share of the population in developed countries, and they continue to do so in many developing countries today. We therefore develop a new data set that maps U.S. subcounty divisions—Minor Civil Divisions (MCDs)—to comparable spatial units over time. The small geographical area of MCDs allows us to clearly distinguish rural from urban areas over a long historical time period, but these data are only available for some states. We therefore complement our MCD data with a county-level data set, which covers almost all of the continental United States at 20-year time intervals.

To guide our empirical work, we develop a simple theoretical model, in which agents are mobile across locations and choose between agriculture and non-agriculture as sectors of employment.[2] A key feature of the model is that agriculture's share of initial employment varies with population density. At low population densities where agriculture dominates initial employment, mean reversion in agricultural employment growth generates a downward-sloping relationship between population growth and initial population density. At high population densities where nonagriculture dominates initial employment, a roughly constant rate of nonagricultural employment growth generates population growth that is largely uncorrelated with initial population density or size, as in existing empirical findings for cities and metropolitan areas. In between, the share of agriculture in initial employment is decreasing in population density, and structural transformation from agriculture to nonagriculture raises population growth at higher densities with lower shares of agriculture in employment.

2. "Nonagriculture" includes manufacturing and services.

Consistent with these theoretical predictions, our main empirical finding is that population growth from 1880 to 2000 is strongly increasing in initial population density for the range of intermediate densities at which the majority of the 1880 population lived. In particular, locations that started out with about 7 people per square kilometer did not grow on average, whereas those with 50 people per square kilometer more than tripled on average. This upward-sloping relationship between population growth and initial population density at intermediate densities increases the dispersion of population density from 1880 to 2000, with some locations experiencing rural depopulation as others experience urbanization.

We organize our empirical analysis around six stylized facts, which link this pattern of population growth to structural transformation away from agriculture. After developing these facts, we provide further evidence in support of this explanation using variation across U.S. subperiod and regions. We show that the increasing relationship between population growth and initial population density at intermediate densities is stronger from 1880 to 1960, during which substantial reallocation away from agriculture occurred, than from 1960 to 2000, during which such reallocation was much less important. We also show that this increasing relationship at intermediate densities is stronger for regions and 20-year periods characterized by greater structural transformation away from agriculture, such as the U.S. South from 1940 to 1960.

We provide evidence that our findings are not driven by a number of possible alternative explanations, including structural transformation within non-agriculture from manufacturing to services; suburbanization and changes in transport technology; locational fundamentals (including physical geography and institutions); and initial local differences in fertility, migration, and human capital. We find a similar pattern of results when we exclude metropolitan areas, using either modern or historical definitions, which suggests that our findings are not driven by suburbanization. Even when we focus solely on variation across MCDs within the same county with similar initial population densities and include controls for all of the foregoing potential alternative explanations, we continue to find statistically significant effects of structural transformation away from agriculture.

The remainder of the article is organized as follows. Section II discusses the related literature. Section III develops the model.

Section IV discusses our data. Section V presents our six key stylized facts about the evolution of population and employment and provides some baseline evidence on the role of structural transformation away from agriculture in explaining these stylized facts. Section VI presents further evidence in support of this explanation and against a variety of possible alternatives. Section VII concludes.

II. Related Literature

Our article is related to a large body of work in urban economics and economic geography. Recent research on the relationship between population growth and size for cities includes Gabaix (1999), Eeckhout (2004), Duranton (2007), and Rossi-Hansberg and Wright (2007). While the existing empirical literature on cities typically finds that population growth is largely uncorrelated with initial population size (Gibrat's Law), some evidence of departures from Gibrat's Law is found even for cities, as in Black and Henderson (2003), Soo (2007), González-Val, Lanaspa, and Sanz (2008), and Holmes and Lee (2010).[3] Recent theoretical research on cities emphasizes population mobility across locations and examines a number of different economic mechanisms underlying random city growth, including endogenous innovation and industry shocks. Both of these mechanisms can generate mean reversion in city population growth, so that the very largest cities grow somewhat more slowly, as in Duranton (2007) and Rossi-Hansberg and Wright (2007). As this existing literature on cities abstracts from rural areas, two recurring issues are the treatment of entry into the city-size distribution and the population threshold for being classified as a city.[4]

Our focus on the reallocation of economic activity from agriculture to non-agriculture also connects with theories of new economic geography, including Krugman (1991) and Fujita, Krugman, and Venables (1999). Although reductions in trade costs in these models can result in an increased dispersion of population across space, the new economic geography literature does not provide natural explanations for why Gibrat's Law is

3. Research on the empirical determinants of city growth includes among others Glaeser et al. (1992), Ioannides and Overman (2004), Glaeser and Gyourko (2005), and da Mata et al. (2007), and is surveyed in Gabaix and Ioannides (2004). The role of industrial specialization is emphasized in Henderson (1974).

4. For an analysis of the emergence of new cities as a source of growth in the urban population, see Henderson and Venables (2009).

a reasonable approximation for observed city population growth (see for example the discussion in Davis and Weinstein 2002) or for why Gibrat's Law is violated when both rural and urban areas are considered.

Though an empirical literature has examined the determinants of the distribution of economic activity across states and counties in the United States, including Kim (1995), Ellison and Glaeser (1999), Beeson, DeJong, and Troesken (2001), Rappaport and Sachs (2003), and Glaeser (2008), this literature has typically not emphasized structural transformation. Closest in spirit to our work is Caselli and Coleman (2001), which examines structural transformation and the convergence of incomes between Southern and Northern U.S. states. Also related is Desmet and Rossi-Hansberg (2009), which examines differences in patterns of employment growth between the manufacturing and service sectors using U.S. county data and relates these differences to technological diffusion and the age of sectors. Neither paper examines the relationship between structural transformation and urbanization—an analysis for which our subcounty data are especially well suited.

Our research is also related to the macroeconomics literature on structural transformation. The model developed herein captures the two key explanations for structural transformation proposed in the existing macroeconomics literature. The first is more rapid productivity growth in agriculture than in nonagriculture combined with inelastic demand across sectors, as in Baumol (1967), Ngai and Pissarides (2007), and Rogerson (2008). The second is nonhomothetic preferences in which the relative weight of agriculture in consumer preferences declines with real income, as in Echevarria (1997), Gollin, Parente, and Rogerson (2002), and Matsuyama (2002). Though both strands of this literature are concerned with the impact of structural transformation on macroeconomic aggregates, our focus is instead on its implications for the population distribution and the process of urbanization.

Finally, our article connects with a long line of research in the development and economic history literatures. Early work on structural change and economic development includes Lewis (1954) and is surveyed by Syrquin (1988), while more recent research on the interlinkages between industrial and agricultural development is reviewed in Foster and Rosenzweig (2008). Influential work on the history of urban development in the United

States includes Kim (2000) and Kim and Margo (2004), although for reasons of data availability this research has again largely concentrated on cities.

III. Model

To guide our empirical analysis, we develop a theoretical model that illustrates the mechanisms linking urbanization and structural transformation.[5] We consider an economy with many locations that are linked through goods trade and population mobility. The mechanism that drives an aggregate reallocation of employment from agriculture to nonagriculture is either more rapid productivity growth in agriculture combined with inelastic demand across the two goods or a change in relative demand for these two goods (e.g., as a result of nonhomothetic preferences). This aggregate reallocation affects the relationship between population growth and initial population density because agriculture's share of employment varies with population density. Agricultural specialization and population density are related because agriculture is land intensive, has weaker agglomeration forces, and exhibits greater mean reversion in productivity than nonagriculture, which implies that agriculture's share of employment declines at the highest population densities.

III.A. Preferences and Endowments

Time is discrete and indexed by t. The economy consists of a continuum of locations $i \in [0,1]$, which are grouped into larger statistical units called MCDs or counties. Each location is endowed with a measure H_i of land. The economy as a whole is endowed with a measure L_t of workers who are perfectly mobile across locations. Workers are infinitely lived and each is endowed with one unit of labor, which is supplied inelastically with zero disutility, so that employment equals population for each location.

Workers derive utility from consumption of goods, C_{it}, and residential land use, H_{Uit}, and for simplicity we assume that the utility function takes the Cobb-Douglas form:[6]

$$U\left(C_{it}, H_{Uit}\right) = C_{it}^{\alpha} H_{Uit}^{1-\alpha}, \qquad 0 < \alpha < 1. \tag{1}$$

5. The Online Appendix contains further details on the model and the technical derivations of the relationships reported in this section.

6. For empirical evidence using U.S. data in support of the constant housing expenditure share implied by the Cobb-Douglas functional form, see Davis and Ortalo-Magne (2011).

The goods consumption index, C_{it}, includes consumption of agriculture, c_{Ait}, and nonagriculture, c_{Nit}, and is assumed to take the constant elasticity of substitution (CES) form:

$$(2)\quad C_{it}=\left[a_t c_{Ait}^{\rho}+(1-a_t)\,c_{Nit}^{\rho}\right]^{1/\rho},\quad 0<\kappa=\frac{1}{1-\rho}<1,\quad 0<a_t<1,$$

where a_t captures the relative strength of demand for agriculture and, following a large literature in macroeconomics, we assume that the two goods are complements: $0<\kappa<1$.[7]

Expenditure on land in each location is redistributed lump sum to the workers residing in that location, as in Helpman (1998). Therefore total income in each location equals payments to labor and land used in production plus expenditure on residential land use.

III.B. Production Technology

Output in each sector is produced using labor and land according to a Cobb-Douglas production technology:

$$(3)\qquad Y_{jit}=L_{jit}^{\eta_j}\Gamma_{jt}\theta_{jit}L_{jit}^{\mu_j}H_{jit}^{1-\mu_j},\qquad 0<\mu_j<1,\ \eta_j\geq 0,$$

where Y_{jit}, L_{jit}, and H_{jit} denote output, employment, and commercial land use, respectively, for sector $j\in\{A,N\}$ in location i at time t; $L_{jit}^{\eta_j}$ captures external economies of scale in employment in the sector and location; Γ_{jt} is a component of sectoral productivity that is common across all locations (e.g., the aggregate state of technology); θ_{jit} is a component of sectoral productivity that is specific to each location (e.g., natural resources and weather).[8]

Within each sector, output is homogeneous and markets are perfectly competitive, with each good costlessly tradeable across locations.[9] Each firm is of measure zero and chooses its inputs of labor and land to maximize its profits taking as given productivity, goods and factor prices, and the location decisions of other firms and workers. Because economies of scale are external to the firm, they depend on aggregate (rather than firm) employment in a

7. See, for example, Ngai and Pissarides (2007).

8. While agglomeration forces are captured here through external economies of scale, see Duranton and Puga (2004) and Rosenthal and Strange (2004) for a discussion of other sources of agglomeration.

9. In a separate technical note (Michaels, Redding, and Rauch 2011), we develop a quantitative version of the model that features bilateral transport costs and yet remains tractable by introducing Eaton and Kortum (2002) heterogeneity and product differentiation within each sector.

sector and location. As such, each firm's production technology exhibits constant returns to scale in its own inputs of labor and land, which yields the standard result that payments to labor and land exactly exhaust the value of output. We assume that agriculture is more land intensive than nonagriculture ($0 < \mu_A < \mu_N < 1$) and that nonagriculture has stronger external economies of scale than agriculture ($0 \leq \eta_A < \eta_N$).

The location-specific component of sectoral productivity (θ_{jit}) evolves stochastically over time as a result of idiosyncratic shocks to productivity (ϕ_{jit}):

$$\theta_{jit} = \phi_{jit}\theta_{jit-1}^{\nu_j}, \qquad t = \{1, \ldots, \infty\}, \quad \theta_{ji0} = \phi_{ji0}, \tag{4}$$

where $\ln \phi_{jit}$ is drawn from an independently and identically distributed continuous probability density function $g_j(\ln \phi_{jit})$, with mean zero, constant variance $\sigma_{\phi j}^2 > 0$, and bounded support $\left[\ln \underline{\phi}_j, \ln \overline{\phi}_j\right]$.

The parameter ν_j captures the degree of mean reversion in location-specific productivity. Since the relative productivity of locations in agriculture is heavily influenced by long-term fundamentals, such as soil and climate, we assume greater mean reversion in location-specific productivity in agriculture than in nonagriculture: $0 < \nu_A < \nu_N \leq 1$. From this law of motion for productivity (4) and the distribution of idiosyncratic productivity shocks $g_j(\ln \phi_{jit})$, we determine the limiting distribution of productivity in each sector j $(z_j(\theta_{jit}))$.

III.C. Land Allocation

Land in each location can be allocated to residential or commercial use. When land is used commercially, we assume that it can be employed either in agriculture or nonagriculture but not in both sectors simultaneously, so that locations exhibit complete specialization in production. Since locations are grouped into larger statistical units (MCDs or counties), these larger statistical units exhibit incomplete specialization to the extent that they contain a mix of agricultural and nonagricultural locations.[10]

Commercial land in each location is employed in the sector with the higher value marginal product for land. Given our

10. The assumption of complete specialization in production simplifies the allocation of land between residential and commercial use. While this assumption can be relaxed, this substantially complicates the characterization of general equilibrium without yielding much additional insight.

assumptions that production (3) and the upper tier of utility (1) are Cobb-Douglas and that specialization in production is complete, the equilibrium allocation of land takes a particularly tractable form with a constant fraction of land allocated to residential and commercial use:

$$(5) \qquad \chi_{Uj} = \frac{1-\alpha}{(1-\alpha)+(1-\mu_j)\,\alpha}, \qquad \chi_{Yj} = \frac{(1-\mu_j)\,\alpha}{(1-\alpha)+(1-\mu_j)\,\alpha},$$

where χ_{Uj} and χ_{Yj} denote the fractions of land used residentially and commercially, respectively, for locations in which commercial land is used in sector j.

III.D. Population Mobility

Workers are perfectly mobile across locations and can relocate instantaneously and at zero cost. After observing the vector of agricultural and non-agricultural productivity shocks across locations i in period t, ϕ_{Ait} and ϕ_{Nit}, each worker chooses their location to maximize their discounted stream of utility, taking as given goods and factor prices and the location decisions of other workers and firms. Because relocation is instantaneous and costless, this problem reduces to the static problem of maximizing their instantaneous flow of utility. Therefore population mobility implies the same real income across all populated locations:

$$\frac{\pi_{it}}{P_t^{\alpha} r_{it}^{1-\alpha}} = \frac{\pi_{kt}}{P_t^{\alpha} r_{kt}^{1-\alpha}} = V_t, \qquad \forall i, k,$$

where π_{it} denotes income per worker from labor and land; P_t is the dual price index for consumption goods derived from (2), which with costless trade is the same for all locations; r_{it} is the rental rate on land, which in general varies across locations.

Using land market clearing, the equality between income and payments to land and labor, and the equilibrium land allocation (5), the population mobility condition can be reexpressed as follows:

$$(6) \qquad \tilde{V}_{it} = \frac{p_{jt}\Gamma_{jt}\theta_{jit}\chi_{Yj}^{1-\mu_j} H_i^{1-\mu_j+\frac{1-\alpha}{\alpha}} L_{it}^{\eta_j-\left(1-\mu_j\right)-\frac{1-\alpha}{\alpha}}}{\alpha\left[(1-\alpha)+(1-\mu_j)\,\alpha\right]^{\frac{1-\alpha}{\alpha}}} = V_t^{1/\alpha} P_t = \tilde{V}_t,$$

where $\tilde{V}_t$ is a normalized common level of real income across all populated locations.

Both the good produced by a location and its population are determined by its productivities in the two sectors, since these determine the value marginal product of land in each of the alternative patterns of complete specialization and the value marginal product of labor, which together determine income per worker in each location.

III.E. General Equilibrium

The general equilibrium of the model can be referenced by the limiting distribution of productivities in each sector ($z_j\left(\theta_{jit}\right)$), the sets of locations specializing in agriculture and nonagriculture (Ω_{At}, Ω_{Nt}), the measure of workers in each location (L_{it}), and the price of the agricultural good (p_{At}), where we choose the nonagricultural good as the numeraire, so that $p_{Nt} = 1$. From this information, all other endogenous variables of the model can be determined, as shown in the proof of Proposition 1.

PROPOSITION 1. Assuming $\eta_j < (1 - \mu_j) + \frac{1-\alpha}{\alpha}$ for $j \in \{A, N\}$, there exists a unique stable equilibrium.

Proof. See the Online Appendix. ■

For parameter values satisfying the above inequality, the agglomeration forces from external economies of scale (η_j) are not too strong relative to the dispersion forces from an inelastic supply of land for residential and commercial use (α and μ_j). When this inequality holds, each location's real income (6) is monotonically decreasing in its population and there exists a unique stable equilibrium.[11]

PROPOSITION 2. With mean reversion in agricultural productivity ($0 < \nu_A < \nu_N \leq 1$) and approximately constant proportional growth in nonagricultural productivity ($\nu_N \to 1$): (*a*) the dispersion of population density across non-agricultural locations is greater than the dispersion of population density across agricultural locations, (*b*) the most densely populated locations only produce the nonagricultural good, (*c*) some less densely populated locations produce the agricultural good and

11. For parameter values for which the inequality is not satisfied, the model exhibits multiple equilibria, but each of the stable equilibria features a degenerate population distribution with all economic activity concentrated in a single location. Since we do not observe a degenerate population distribution in the data, we focus on parameter values for which the inequality is satisfied and there is a unique stable equilibrium.

there is a range of population densities at which the share of agriculture in employment strictly decreases with population density.

Proof. See the Online Appendix. ■

From the above proposition, one of the key predictions of the model is that the cross-section distribution of population densities differs between agricultural and nonagricultural locations. This difference in population density distributions reflects two sets of forces. First, agriculture is land intensive ($\mu_A < \mu_N$) and exhibits weaker agglomeration forces than nonagriculture ($\eta_A < \eta_N$), which reduces population density in agricultural locations relative to nonagricultural locations for given productivities. Second, in each sector, the variance and maximum value of productivity increase as the coefficient on lagged productivity ν_j rises toward 1. Therefore lower mean reversion in productivity in nonagriculture than in agriculture ($0 < \nu_A < \nu_N \leq 1$) implies a greater variance and maximum value for nonagricultural productivity for the same distribution of idiosyncratic shocks to productivity.[12] In the limit, as $\nu_N \to 1$ and $t \to \infty$, the variance and maximum value of nonagricultural productivity are unbounded. Noting that a location's population density is increasing in its productivity, it follows from these results that there is greater dispersion in population density across nonagricultural locations and that the most densely populated locations only produce the nonagricultural good. Because both goods are consumed and produced in equilibrium, some less densely populated locations produce the agricultural good, which in turn implies that there is a range of population densities where the share of agriculture in employment is strictly decreasing in population density.

PROPOSITION 3. A rise in aggregate productivity in agriculture (Γ_{At}) or a reduction in relative demand for agriculture (a_t) reallocate employment from agriculture to nonagriculture.

Proof. See the Online Appendix. ■

12. Using farm output per kilometer squared as a crude measure of agricultural productivity, we find evidence of substantial mean reversion in agricultural productivity in our county subperiods data set, as discussed in the Online Appendix. These findings are consistent with the historical literature on the development of U.S. agriculture, which discusses that some of the improvements in agricultural technology during our sample period favored areas that were initially less productive as a result of poorer climate and soil (see Cochrane 1979 and Olmstead and Rhode 2002).

Another key prediction of the model is that an increase in relative aggregate productivity in agriculture or a fall in the relative demand for agriculture reallocates employment away from agriculture, as formalized in the above proposition. With inelastic demand across goods, more rapid productivity growth in agriculture (a rise Γ_{At} in relative to Γ_{Nt}) leads to a more than proportionate decline in the relative price of the agricultural good.[13] This more than proportionate price reduction in turn reduces real income in agricultural locations relative to nonagricultural locations. Two mechanisms restore equilibrium in the model. First, population mobility from agricultural to nonagricultural locations arbitrages away real income differences, because real income per worker in each location is decreasing in its population. Second, land use in a given location can change endogenously from agriculture to nonagriculture. Both mechanisms increase employment in nonagriculture relative to agriculture until real income per worker is equalized across all agricultural and nonagricultural locations in the new equilibrium. Following similar reasoning, a fall in the relative demand for agriculture (a reduction in a_t) also reduces the relative price of the agricultural good and hence reallocates employment toward nonagriculture.[14]

PROPOSITION 4. (*a*) For locations that continue to produce the agricultural good, there is a decreasing relationship between population growth and initial population density. (*b*) For locations that continue to produce the nonagricultural good, population growth becomes uncorrelated with initial population density as $\nu_N \to 1$.

Proof. See the Online Appendix. ■

Even in the absence of changes in relative aggregate productivity or demand across sectors, idiosyncratic shocks to

13. Evidence of faster productivity growth in agriculture than nonagriculture comes from empirical growth accounting studies for the United States during our sample period, including Kuznets (1966) from 1870–1940 and Maddison (1980) from 1950–1976. These findings are also consistent with the economic history literature on the development of U.S. agriculture, which emphasizes the role of productivity-enhancing improvements in technology, such as mechanization and biological innovation (e.g., Rasmussen 1962).

14. One natural explanation for a fall in the relative demand for agriculture arises in the nonhomothetic CES preferences of Sato (1977), in which a_t falls as real income rises.

location-specific productivity induce changes in population across locations, as formalized in the above proposition. In locations that continue to produce the agricultural good, mean reversion in agricultural productivity induces a decreasing relationship between population growth and initial population density. In contrast, in locations that continue to produce the nonagricultural good, population growth is largely unrelated to initial population density for sufficiently small degrees of mean reversion in nonagricultural productivity ($\nu_N \to 1$).

From Propositions 2–4, the model yields a number of predictions that guide our empirical research.

Population Growth. The main prediction of the model is an initially decreasing, later increasing, and finally roughly constant relationship between population growth and initial population density. This prediction is driven by the following three forces. At high population densities where nonagriculture dominates initial employment (Proposition 2), constant proportional growth in nonagricultural productivity implies a roughly constant relationship between population growth and initial population density (Proposition 4). At low population densities where agriculture dominates initial employment (Proposition 2), mean version in agricultural productivity generates a decreasing relationship between population growth and initial population density (Proposition 4). In between, the share of agriculture in initial employment is decreasing in population density (Proposition 2), and structural transformation from agriculture to nonagriculture raises population growth at higher densities with lower shares of agriculture in employment (Proposition 3).

Employment Shares. A closely related prediction is that agriculture's share of employment varies with population density. From Proposition 2, mean reversion in agricultural productivity and constant proportional growth in nonagricultural productivity imply that the most densely populated locations only produce the nonagricultural good. Since some less densely populated locations produce the agricultural good, this in turn implies that there is a range of population densities at which agriculture's share of employment strictly decreases with population density.

Employment Density Dispersion. The model also predicts greater dispersion in employment density in nonagriculture than in agriculture. From Proposition 2, mean reversion in agricultural productivity and constant proportional growth in nonagricultural productivity imply greater variance in productivity and hence employment density in nonagriculture than in agriculture.

Population Density Dispersion. A related prediction stems from reallocation of employment from agriculture to non agriculture in Proposition 3. More rapid agricultural productivity growth or a shift in relative demand toward nonagriculture leads to a change in the composition of employment from agriculture to nonagriculture. Since nonagriculture has a greater dispersion of employment densities than agriculture (Proposition 2), this change in the composition of employment increases the dispersion of population densities.

Agricultural Mean Reversion. Another prediction is a direct implication of mean reversion in agricultural productivity growth. From Proposition 4, this implies mean reversion in employment growth in locations that continue to produce the agricultural good.

Nonagricultural Constant Growth. A final prediction is a direct implication of approximately constant proportional growth in nonagricultural productivity. From Proposition 4, this implies approximately constant proportional growth in employment in locations that continue to produce the nonagricultural good.

Guided by these theoretical predictions, we distinguish six stylized facts about population and employment in our empirical analysis. In later sections, we provide evidence in support of the model's mechanism of structural transformation away from agriculture as the explanation for these six facts.

IV. Data Description

To study the evolution of population and employment for both rural and urban areas, we require a data set that covers all the population and all the land—from the largest cities to the smallest farms. To distinguish rural areas from urban ones, we require that this data set uses geographic units that are stable over time and yet sufficiently spatially disaggregated.

Though much of the existing literature on population growth in the United States uses data on cities, incorporated places, or metropolitan statistical areas (MSAs), each of these data sets selectively covers locations that became population concentrations. As a result, they largely exclude rural areas, where most of the U.S. population lived historically, and hence they are not well suited to analyze the process of urbanization and its implications for both rural and urban areas. Although counties cover all of the population and land area, they often pool together urban centers with their surrounding countryside, clouding the distinction between urban and rural areas.[15]

To address these limitations, we construct a new data set for the United States based on subcounty divisions. We follow common usage in referring to these sub-county divisions as Minor Civil Divisions (MCDs), but in certain parts of the United States they are now known as Census County Divisions (CCDs).[16] For 1880, data on employment by industry and population for each MCD can be obtained by aggregating the individual-level records of the 1880 U.S. population census. For 2000, data on employment by industry and population are available for each MCD from the American Factfinder of the U.S. Census Bureau. For the intermediate year of 1940, data on population alone are available for each MCD from the published volumes of the U.S. population census.[17]

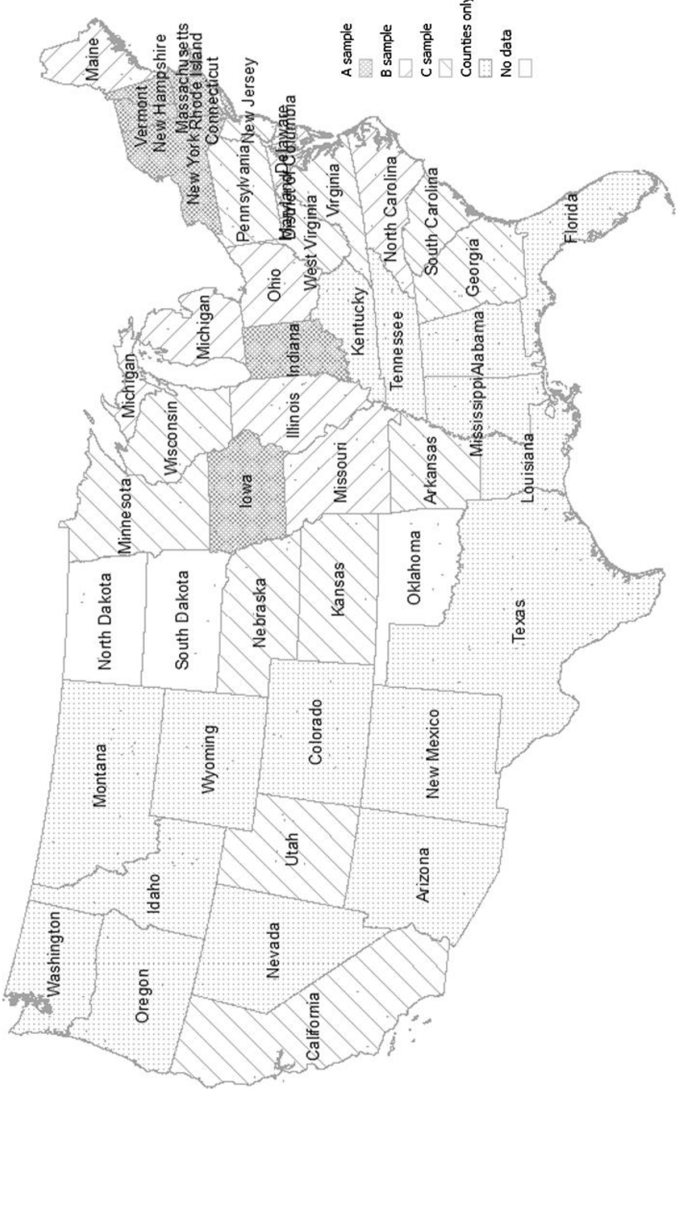
To construct consistent geographic units over time, we use Geographical Information Systems (GIS) software together with historical maps. We match the approximate centroid of each 1880 and 1940 MCD to the 2000 MCD in which it falls, aggregating any 2000 MCD that does not contain at least one 1880 MCD and one 1940 MCD to the nearest MCD within the same state that does.[18] The extent of aggregation required to construct time-consistent units varies across U.S. states and therefore we group the U.S. states into samples: little aggregation is required in A

15. As shown in Figure A.1 in the Online Appendix, using data on MCDs rather than counties enhances the density of observations for which we observe substantial variation in the share of agriculture in employment.

16. For expositional convenience, we refer to all subcounty divisions as MCDs. Therefore, our use of the term MCD includes CCDs, unorganized territories and all other historical terms used to refer to county sub-divisions, such as but not limited to boroughs and precincts.

17. For further discussion of the U.S. data sources and definitions, see the Online Appendix.

18. As in most cases the resulting geographic units consist of a single MCD, we refer to them for simplicity as MCDs, even though they sometimes consist of aggregations of MCDs.



MAP I
(continued)

states, more is needed in B states, and more still in C states. For some states, particularly those in the West and South, historical data on subcounty divisions are unavailable, which precludes the construction of consistent geographic units over time below the county level. The geographic distribution of states across the three samples is shown in Map I. We choose as our baseline sample the A and B states, where 1–1 matches between the 1880 and 2000 census that involve no aggregation exceed 70%. But we also examine the robustness of our results to the use of the alternative samples of A states, where very little aggregation is required, and A, B, and C states, which cover a wider geographical area.[19]

The key advantages of our MCD data set are that it covers both high and low population densities and is sufficiently spatially disaggregated to permit a sharp distinction between urban and rural areas. The average MCD in our baseline sample for the A and B states has an area of around 115 km^2 and a population of approximately 8,800 in 2000. As a point of comparison, the average county in the same sample has an area of around 1,500 km^2 and a population of approximately 115,000 in 2000.

⟵

MAP I

MCD Data by State

This map shows the states used for our various samples. Our baseline MCD sample consists of A and B states. The classification A, B, and C is based on the frequency with which MCDs in 1880 and 2000 can be matched 1-1 with no aggregation. In states classified as A (Connecticut, DC, Indiana, Iowa, Massachusetts, New Hampshire, New York, Rhode Island, Vermont), the 1-1 match rate between 1880 and 2000 MCDs is larger than 0.9. In states classified as B (Illinois, Maine, Maryland, Michigan, Missouri, North Carolina, Ohio), the 1-1 match rate is larger than 0.7. In states classified as C (Arkansas, California, Delaware, Georgia, Kansas, Minnesota, Nebraska, New Jersey, Pennsylvania, South Carolina, Utah, Virginia, West Virginia, Wisconsin), 1880 MCD data are available but the 1-1 match rate is lower than 0.7. For states in the counties sample (Alabama, Arizona, Colorado, Florida, Idaho, Kentucky, Louisiana, Mississippi, Montana, Nevada, New Mexico, Oregon, Tennessee, Texas, Washington, Wyoming), 1880 MCD data are not available. Our county subperiods data for 20-year intervals from 1880–2000 are available for the A, B, C, and counties only states except for Wyoming. We exclude Alaska, Hawaii, Oklahoma, North Dakota, and South Dakota, which had not attained statehood in 1880 and therefore are either not included in the 1880 census or did not have stable county boundaries at that time.

19. For further discussion of these samples and the construction of the U.S. MCD data, see the Online Appendix.

Two limitations of our MCD data set are its incomplete geographical coverage, with the sample of A and B states concentrated in the Midwest and Northeast of the United States, and the lack of data on industry employment for intervening years between 1880 and 2000. To address these limitations, we use complementary data on U.S. counties. Although counties are more spatially aggregated than MCDs, they again cover the full range of population densities. Furthermore, county data on population and industry employment can be constructed for almost all of the continental United States for 20-year periods from 1880 to 2000.[20] Using these county data, we can therefore examine variation in the timing of structural transformation across subperiods and regions.[21]

Finally, while we concentrate on results using our U.S. data, we also briefly discuss the results of a robustness test using Brazilian data from 1970 to 2000. A more detailed discussion of the Brazilian data and results is contained in the Online Appendix.

V. Baseline Empirical Results

In this section, we introduce our empirical specification, outline six stylized facts about population and employment that are closely related to the predictions of the model, and present our baseline evidence that structural transformation away from agriculture plays a central role in explaining these stylized facts and the relationship between them.

V.A. *Empirical Specification*

To analyze the evolution of population and employment over time, we adopt a nonparametric approach that imposes minimal structure on the data. To characterize the population density distribution, we divide the range of values for log population

20. We exclude Alaska, Hawaii, Oklahoma, North and South Dakota from our 1880 and 2000 data, because they had not attained statehood in 1880 and did not have stable county boundaries at that time. From our county subperiods data for 20-year time periods, we also exclude Wyoming because of missing information in the GIS shapefiles used to create these data. For further discussion of the U.S. counties data, see the Online Appendix.

21. One remaining concern is that MCDs and counties may not coincide with the economic boundaries between local markets. To address this concern, we undertake a number of robustness tests using modern and historical measures of metropolitan areas, as discussed further below.

density, x, into discrete bins of equal size δ. We index MCDs by m and bins by $b \in \{1, ..., B\}$. Denoting the set of MCDs with log population density in bin b by Φ_b and denoting the number of MCDs within this set by n_b, we estimate the population density distribution, $g(x_m)$, as follows:

$$(7) \qquad \hat{g}(x_m) = \frac{n_b}{n}, \qquad n = \sum_{b=1}^{B} n_b, \quad \text{for } x_m \in \Phi_b,$$

where a hat above a variable denotes an estimate. Thus the estimated probability of observing a population density within the range of values included in bin b equals the fraction of MCDs with population densities in this range. This corresponds to a simple histogram, which yields a consistent estimate of the true underlying probability density function (Scott 1979). We choose bin sizes of $\delta = 0.1$ log points, which provide a fine discretization of the range of log population densities, while generally preserving a relatively large number of MCDs within each bin. Although this approach provides a simple and flexible characterization of the population density distribution, which connects closely with the other components of our analysis, we also find similar results using related nonparametric approaches such as kernel density estimation.

To characterize the relationship between population growth and the initial population density distribution, we follow a similar approach. We approximate the continuous function relating population growth to initial population density using a discrete-step function consisting of mean population growth within each initial population density bin:

$$(8) \quad \hat{y}_{mt} = f(x_{mt-T}) = \sum_{b=1}^{B} I_{bt-T}\phi_{bt-T}, \qquad \phi_{bt-T} = \frac{1}{n_{bt-T}} \sum_{x_{mt-T} \in \Phi_{bt-T}} y_{mt},$$

where t indexes time. In this specification, bins are defined over initial population density, x_{mt-T}; y_{mt} is MCD population growth from $t-T$ to t; I_{bt-T} is an indicator variable equal to 1 if $x_{mt-T} \in \Phi_{bt-T}$ and 0 otherwise.

This specification corresponds to a regression of population growth on a full set of fixed effects for initial population density bins. We report both mean population growth and the 95% confidence intervals around mean population growth for each initial population density bin. The confidence intervals are based on heteroscedasticity-robust standard errors adjusted for clustering by county, which allows the errors to be correlated across MCDs within counties without imposing prior structure on the pattern of this correlation.[22] While this nonparametric specification allows for a flexible relationship between population growth and initial population density, we again find similar results using other related nonparametric approaches, such as locally weighted linear least squares regression and kernel regression.

V.B. Stylized Facts

We organize our baseline empirical analysis around six stylized facts about the evolution of population and employment. These six stylized facts are reported for our baseline sample of A and B states in column (1) of Table I and Panels A–F of Figure I.[23]

(1) Increased Dispersion of Population Density. Our first stylized fact is that the distribution of log population density across MCDs has become more dispersed from 1880 to 2000. As shown in Panel A of column (1) in Table I, there is an increase in the standard deviation of log population density over this period from 0.97 to 1.56, which is statistically significant and larger than the increase in mean log population density from 2.61 to 3.08. Panel A of Figure I confirms this increased dispersion in population densities by displaying the results from specification (7). Although the U.S. population increased substantially from 1880 to 2000, as reflected in Panel A of Figure I in an increased mass of densely populated areas, the figure also shows an increased mass of sparsely populated areas. The population density distribution therefore exhibits polarization, with some low-density areas depopulating while other higher-density areas experience rapid population growth.

(2) Correlation Between Population Growth and Initial Population Density at Low-Medium Densities. The second stylized fact is that Gibrat's Law of constant proportional growth is strongly rejected for population growth when both rural and urban areas are considered. In Panel B of Figure I, we display

22. Bertrand, Duflo, and Mullainathan (2004) examine several approaches to controlling for serial correlation and show that clustering the standard errors performs very well in settings with more than 50 clusters as in our application. In the Online Appendix, we report the results of a robustness check, in which we estimate standard errors using the alternative approach to allowing for spatial correlation of Bester, Conley, and Hansen (2011).

23. See Section V.B of the Online Appendix for further details on the construction of Panels A–F of Figure I.

TABLE I

U.S. STYLIZED FACTS AND THEIR ROBUSTNESS

		(1) MCDs Baseline: A and B states	(2) MCDs Only A states	(3) Counties, 45 states and DC[a]	(4) Counties, A and B sample[a]	(5) Hybrid MCD-county sample[b]	(6) MCDs British colonial claims	(7) MCDs pool 2000 MSAs	(8) MCDs exclude 2000 MSAs
Panel A	St dev of log pop dens in 1880 (σ_1)	0.967	1.025	1.757	0.963	1.273	1.222	0.930	1.268
	St dev of log pop dens in 2000 (σ_2)	1.556	1.631	1.450	1.303	1.687	1.680	1.234	1.352
	H_0: $\sigma_1 = \sigma_2$, vs. H_1: $\sigma_1 < \sigma_2$, p-value	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.008
	Stylized Fact 1: Distribution of log population density across geographic units became more dispersed from 1880–2000 (population became more concentrated)	Yes	Yes	No[c]	Yes	Yes	Yes	Yes	Yes
Panel B	Mean pop growth at log pop dens 0 ($\beta_g(0)$)	0.013	0.012	0.016	0.019	0.010	0.013	0.013	0.015
	Mean pop growth at log pop dens 2 ($\beta_g(2)$)	0.001	−0.001	0.007	0.007	0.002	0.005	0.000	0.000
	Mean pop growth at log pop dens 4 ($\beta_g(4)$)	0.009	0.010	0.014	0.014	0.011	0.012	0.005	0.010
	Mean pop growth at log pop dens larger 4 ($\beta_g(>4)$)	0.010	0.011	0.014	0.015	0.012	0.012	0.005	0.003
	H_0: $\beta_g(0) = \beta_g(2)$, H_1: $\beta_g(0) > \beta_g(2)$, p-value	0.000	0.000	0.000	0.000	0.000	0.036	0.000	0.000
	H_0: $\beta_g(2) = \beta_g(4)$, H_1: $\beta_g(2) < \beta_g(4)$, p-value	0.000	0.000	0.001	0.011	0.000	0.000	0.000	0.000
	H_0: $\beta_g(4) = \beta_g(>4)$, H_1: $\beta_g(>4) \neq \beta_g(4)$, p-value	0.489	0.913	0.937	0.606	0.645	0.930	0.740	0.317
	Stylized Fact 2: Increasing relationship between pop growth from 1880–2000 and log pop dens in 1880 at intermediate densities	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel C	Percent of agricultural in total employment at log pop dens 2 ($\beta_{sa}(2)$)	0.767	0.762	0.691	0.618	0.738	0.669	0.771	0.773
	Percent of agricultural in total employment at log pop dens 4 ($\beta_{sa}(4)$)	0.228	0.189	0.195	0.185	0.228	0.258	0.213	0.206

TABLE I
(CONTINUED)

		(1) MCDs Baseline: A and B states	(2) MCDs Only A states	(3) Counties, 45 states and DC[a]	(4) Counties, A and B sample[a]	(5) Hybrid MCD-county sample[b]	(6) MCDs British colonial claims	(7) MCDs pool 2000 MSAs	(8) MCDs exclude 2000 MSAs
	H$_0$: $\beta_{sa}(2) = \beta_{sa}(4)$, H$_1$: $\beta_{sa}(2) > \beta_{sa}(4)$, p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Stylized Fact 3: Share of agriculture in employment falls in the range where population density distribution in 1880 is positively correlated with pop growth 1880–2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel D	St dev of agricultural employment in 1880 (σ_{1a})	0.820	0.722	1.677	0.810	1.084	0.923	0.896	1.303
	St dev of nonagricultural employment in 1880 (σ_{1na})	1.520	1.631	1.784	1.272	1.778	1.722	1.431	1.601
	H$_0$: $\sigma_{1a} = \sigma_{1na}$, vs. H$_1$: $\sigma_{1a} < \sigma_{1na}$, p-value	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	St dev of agricultural employment in 2000 (σ_{2a})	0.858	0.853	0.806	0.617	0.936	1.061	0.868	0.983
	St dev of nonagricultural employment in 2000 (σ_{2na})	1.623	1.689	1.530	1.359	1.767	1.735	1.293	1.382
	H$_0$: $\sigma_{2a} = \sigma_{2na}$, vs. H$_1$: $\sigma_{2a} < \sigma_{2na}$, p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Stylized Fact 4: St dev of nonagricultural employment is larger than st dev of agricultural employment in both years	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel E	Regress agricultural employment growth on log pop dens and intercept in subsample of units with agricultural employment share > 0.8 in 1880, report slope coefficient (β_a)	−0.0060	−0.0077	−0.0067	−0.0054	−0.0066	−0.0049	−0.0060	−0.0060
	H$_0$: $\beta_a = 0, H_1$: $\beta_a \neq 0$, p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

TABLE I
(CONTINUED)

		(1) MCDs Baseline: A and B states	(2) MCDs Only A states	(3) Counties, 45 states and DC[a]	(4) Counties, A and B sample[a]	(5) Hybrid MCD-county sample[b]	(6) MCDs British colonial claims	(7) MCDs pool 2000 MSAs	(8) MCDs exclude 2000 MSAs
	Stylized Fact 5: Agricultural employment growth is negatively correlated with population density	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel F	Regress non agricultural employment growth on log pop dens and intercept in subsample of units with nonagricultural employment share < 0.2 in 1880, report slope coefficient (β_{na})	−0.0002	−0.0006	−0.0006	−0.0006	−0.0010	−0.0011	−0.0009	−0.0009
	H_0: $\beta_{na} = \beta_a$, H_1: $\beta_{na} > \beta_a$, p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Stylized Fact 6: Stronger mean reversion in agricultural employment than in non-agricultural employment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Number of observations	10,864	4,439	2,425	813	19,229	5,887	6,504	1,480

Notes. This table reports robustness tests of our 6 stylized facts using U.S. data. Panels B, C, E, and F report tests based on regressions using robust standard errors clustered by county. See the text and the Online Appendix for further discussion of the construction of the data.

[a] The county sample includes all U.S. states except Alaska, Hawaii, North Dakota, Oklahoma, and South Dakota, which had not attained statehood in 1880 and did not have stable county boundaries at that time.

[b] The hybrid sample uses the smallest geographical units available for each state. We use MCDs for the states in samples A, B, and C, and counties elsewhere. This sample excludes the five states described above.

[c] Since this sample includes states that were not fully settled in 1880, near-empty areas increase the standard deviation of the population density distribution in that year. When we restrict the analysis to counties in states A and B only, the stylized fact does hold (see column (4)). This is reassuring, since our model is concerned with an equilibrium, which is likely to be a better characterization of the longer-settled A and B states.

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

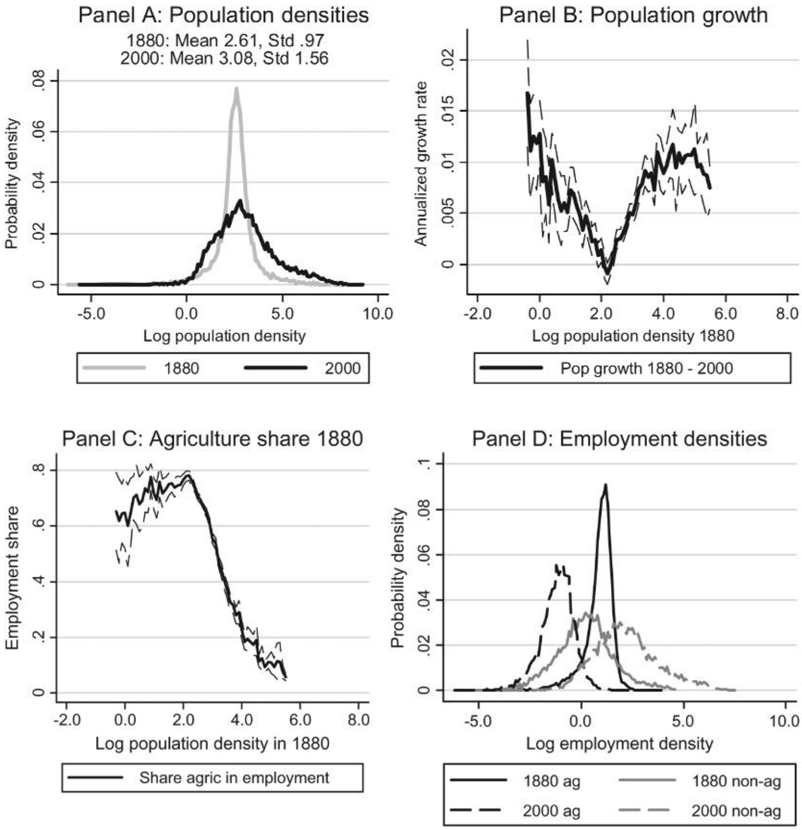


FIGURE I

Stylized Facts

This figure shows the six stylized facts for our baseline sample of MCDs from the A and B states. The x-axes are population density bins (or in Panel D, employment density bins), defined by rounding down log initial population density for each MCD to the nearest single digit after the decimal point. For example, all MCDs with log population density greater than or equal to 0.1 and less than 0.2 are grouped together in bin 0.1. The y-axes show means for each population density bin. In Panels B, C, E, and F, dashed lines show 95% confidence intervals, computed using robust standard errors clustered by county. Since population density bins at the extreme ends of the distribution typically contain few observations, the figures in these panels (but not the estimations) omit the 1% most and least dense MCDs in 1880. See the text and the Online Appendix for further discussion of the construction of the figures.

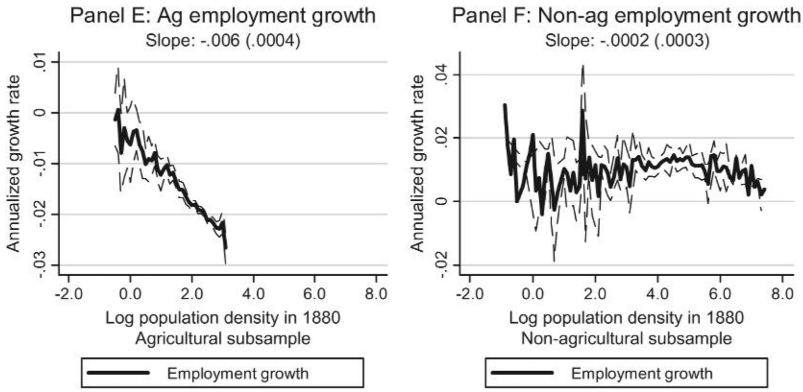


FIGURE I

(continued)

the results from specification (8), where the dark solid line shows mean population growth within each initial population density bin and the lighter dashed lines correspond to the 95% confidence intervals.[24]

For log population densities below 2 (7 people per km^2), there is a negative correlation between population density in 1880 and subsequent population growth. In contrast, for log population densities of between 2 and 4 (7–55 people per km^2), population density in 1880 is positively correlated with subsequent population growth.[25] Above log population densities of around 4 (e.g., from 4–6 log points or 55–403 people per km^2), population density in 1880 is largely uncorrelated with subsequent population growth.

We view the increasing relationship between population growth and initial population density at intermediate densities as our main empirical finding. More than half of the 1880 population in our sample lived at population densities between 2 and 4 log points, and the increase in population growth over this range

24. We do not display mean population growth for the top and bottom 1% of observations in terms of population density, although they are included in the estimation. The bins at these extremes of the distribution contain few observations and have correspondingly large standard errors. Hence they tend to cloud rather than illuminate the true picture.

25. While classical measurement error in 1880 population could induce a negative correlation between population growth and 1880 population density, it cannot account for the positive correlation between these variables observed above a log population density of around 2, and our use of individual-level records from census data mitigates measurement error concerns.

drives the increase in the dispersion of the population distribution in Stylized Fact 1. Furthermore, the magnitude of this departure from Gibrat's Law is substantial: MCDs with log density of about 4 in 1880 experienced population growth at an annualized rate of about 1% from 1880 to 2000, whereas those with a log population density around 2 barely grew on average. As shown in Panel B of column (1) in Table I, this difference in population growth rates is statistically significant.[26]

To place these ranges of population densities in context for our baseline sample from the A and B states, 41 MSAs as defined in 2000 and 129 counties had a 1880 population density of less than 2 log points. Examples include St. Cloud MSA, Minnesota (1.70), and Oxford County, Maine (1.80). In comparison, 170 MSAs and 647 counties had a 1880 population density from 2–4 log points, including Ann Arbor MSA, Michigan (3.13), and Fulton County, New York (3.18). By contrast, 15 MSAs and 33 counties had a 1880 population density of above 4 log points, including Philadelphia-Camden-Wilmington MSA, Pennsylvania–New Jersey–Delaware-Maryland (4.79), and Hartford County, Connecticut (4.19).

The range of population densities above 4 log points where population growth is roughly uncorrelated with initial population density includes the high densities observed in urban areas. Therefore, while our results are based on data on MCDs rather than metropolitan areas, cities, or incorporated places, our findings for densely populated locations are broadly consistent with the existing empirical literature's finding that Gibrat's Law of constant proportional growth is a reasonable approximation for cities.

Though we concentrate on the relationship between population growth and initial log population density to control for variation in land area across MCDs, we note that we find similar results using initial log population size instead of density. As shown in Figure A.2 in the Online Appendix, we observe the same pattern of an initially decreasing, later increasing, and finally roughly constant relationship between population growth and initial log population size. This pattern of results is consistent with the

26. We find a similar pattern of departures from constant growth if we use the 10th, 50th, or 90th percentile of population growth rather than mean population growth for each initial population density bin. In contrast, the variance of population growth is relatively constant across the range of initial population density bins from 0 to 5 log points, with a small increase at the lowest initial population densities, and a small decrease at the highest initial population densities.

approximately log linear relationship between population density and population size in our data.

(3) Agricultural Employment Share. Our third stylized fact is that the share of agriculture in 1880 employment declines sharply in the range where population density in 1880 and subsequent population growth are positively correlated. Panel C of Figure I presents the results from specification (8) using the share of agriculture in employment in 1880 as the left-hand variable rather than population growth. As shown in the figure, the agricultural employment share in 1880 declines from about 0.8 for MCDs with log density of 2 to about 0.2 for MCDs with log density of 4. Panel C of column (1) in Table I shows that this difference in specialization patterns is statistically significant. Although the agricultural employment share continues to decline for MCDs with a log density of greater than 4, it declines at a much slower rate.[27]

(4) Dispersion of Agricultural and Nonagricultural Employment. Our fourth stylized fact is that the employment density distribution is more dispersed in nonagriculture than in agriculture in both 1880 and 2000. As reported in Panel D of column (1) in Table I, the standard deviation of employment density is statistically significantly higher in nonagriculture in both years. We find a similar pattern of results using specification (7), as shown in Panel D of Figure I. This difference in employment density distributions reflects the greater spatial concentration of employment in nonagriculture, with more observations with extreme low and high values of employment density.[28]

From Panel D of Figure I, the employment density distributions for agriculture and nonagriculture shift to the left and right, respectively, over time. As a result, mean employment density decreases in agriculture and increases in non-agriculture, which results in less overlap between the two distributions in 2000 than in 1880. Furthermore, comparing Panels A and D in

27. The ratio of employment to total population was about 0.35 in 1880 and 0.47 in 2000. In both years, it was relatively stable across the population density distribution, suggesting that labor force participation is not strongly related to population density and hence that employment and population dynamics are similar.

28. We also find that employment per square kilometer is more unequally distributed in nonagriculture than in agriculture in both 1880 and 2000 using standard measures of inequality such as the Gini coefficient, the Theil index, the difference between the 90th and 10th percentiles, and the difference between the 99th and 1st percentiles.

Figure I, population in 1880 was distributed in a similar way to agricultural employment in 1880, whereas population in 2000 was more spatially concentrated and distributed in a similar way to nonagricultural employment in 2000. This reflects the substantial decline in agriculture's share of employment during our sample period: the average share of agriculture in MCD employment fell from 63% in 1880 to about 6% in 2000 in our baseline sample of "A and B" states.[29]

(5) Mean Reversion in Agriculture. Our fifth stylized fact is mean reversion in agricultural employment growth. In Panel E of Figure I, we estimate specification (8) for agricultural employment growth for a subsample of MCDs for which agriculture is more than 80% of 1880 employment.[30] As shown in the figure, sparsely populated MCDs in this subsample exhibited more rapid agricultural employment growth from 1880 to 2000 than densely populated MCDs. This mean reversion is confirmed using an OLS regression of agricultural employment growth on log 1880 population density. As reported in Panel E of column (1) in Table I, we find an estimated coefficient of -0.006 (p-value $< .001$), so that each additional log point of 1880 population density is associated on average with just over half a percentage point lower rate of agricultural employment growth.

(6) Substantially Less Mean Reversion in Nonagriculture. Our sixth stylized fact is substantially less mean reversion in nonagricultural employment. In Panel F of Figure I, we estimate specification (8) for nonagricultural employment growth for a subsample of MCDs for which agriculture accounted for less than 20% of 1880 employment. As shown in the figure, nonagricultural employment growth in this subsample is largely uncorrelated with 1880 population density. This pattern of results is confirmed using an OLS regression of nonagricultural employment growth on log 1880 population density. As reported in Panel F of

29. The MCD population distribution is skewed, with a long lower tail of MCDs with low population but high shares of agriculture in employment. The share of agriculture in aggregate employment for our baseline sample of MCDs in the A and B states fell from 42% in 1880 to less than 2% in 2000.

30. Although the agricultural employment share for this subsample was over 88% in 1880, it fell to below 10% in 2000. Hence this subsample does not entirely capture agricultural dynamics alone. Nevertheless, since this subsample was initially mostly agricultural, it is likely to capture the main features of agricultural growth. In our counties data, we find mean reversion in agricultural employment from 1880 to 1900 for counties with more than 80% of their employment in agriculture in both 1880 and 1900.

column (1) in Table I, we find an estimated coefficient of −0.0002 (p-value = .515), which while negative is statistically insignificant and more than an order of magnitude smaller than for agricultural employment. As shown in the table, the null hypothesis that mean reversion in nonagriculture is the same as in agriculture is easily rejected at conventional levels of statistical significance.

In the next section, we present our baseline evidence on the role of structural transformation away from agriculture in explaining these six stylized facts, before returning in Section VI.A to demonstrate the robustness of the stylized facts across a wide range of samples and specifications.

V.C. Baseline Evidence on Structural Transformation

Our first approach to examining the explanatory power of structural transformation away from agriculture builds on standard accounting or decomposition methods. Total employment growth in each MCD can be decomposed into employment growth in agriculture and nonagriculture weighted by the initial shares of each sector in employment. In our Employment Shares specification, we therefore predict MCD population growth using *aggregate* employment growth in agriculture and nonagriculture for the United States as a whole and each *MCD's* own initial employment in each sector.

We first scale up observed 1880 employment in MCD m in sector j (E_{jm1880}) by the aggregate employment growth rate for the sector from 1880 to 2000 ($1 + g_{Ej}$) to obtain predicted MCD 2000 employment in each sector ($\hat{E}_{jm2000}$). Summing the predicted values for agriculture and nonagriculture gives predicted total MCD 2000 employment ($\hat{E}_{m2000}$). We next scale up predicted total MCD 2000 employment by the observed aggregate ratio of population to employment for the United States as a whole in 2000 (k_{2000}) to obtain predicted MCD 2000 population ($\hat{L}_{m2000}$). From predicted MCD 2000 population and observed MCD 1880 population (L_{m1880}), we obtain predicted population growth from 1880 to 2000 ($\hat{g}_{Lm}$):

$$\begin{aligned} \hat{E}_{jm2000} &= E_{jm1880}\left(1+g_{Ej}\right), \\ \hat{E}_{m2000} &= \hat{E}_{Am2000} + \hat{E}_{Nm2000}, \\ \hat{L}_{m2000} &= k_{2000}\hat{E}_{m2000}, \\ \hat{g}_{Lm} &= \ln\left(\frac{\hat{L}_{m2000}}{L_{m1880}}\right). \end{aligned} \tag{9}$$

where a hat above a variable denotes a prediction. Note that this measure of predicted population growth only varies across MCDs because of differences in the 1880 shares of agriculture and non-agriculture in MCD employment.

In Panel A of Figure II, we display the mean of actual and predicted population growth for each 1880 log population density bin. The predicted population growth rate (labeled Emp prediction) captures the relationship between population growth and initial population density in Stylized Fact 2. Regressing mean actual population growth on mean predicted population growth across the initial log population density bins shown in Panel A of

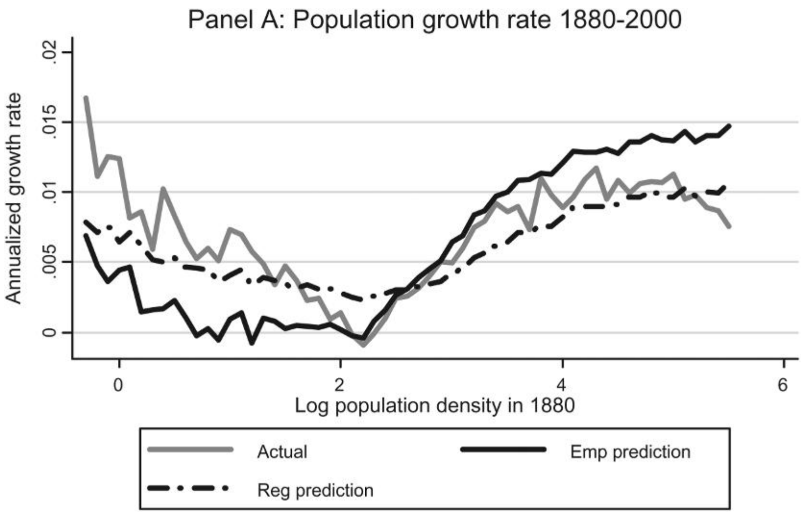


FIGURE II

Employment Share and Regression Predictions

This figure displays actual population growth, population densities, and employment densities as well as the predictions for each of these variables from the Employment Shares prediction. The figures are based on our baseline sample of MCDs from the A and B states. The x-axes are population density bins, defined by rounding down log initial population density for each MCD to the nearest single digit after the decimal point. For example, all MCDs with log population density greater than or equal to 0.1 and less than 0.2 are grouped together in bin 0.1. The y-axes show means for each population density bin. Since population density bins at the extreme ends of the distribution typically contain few observations, the figure in Panel A (but not the estimation) omits the 1% most and least dense MCDs in 1880. See the text and the Online Appendix for further discussion of the construction of the figures.

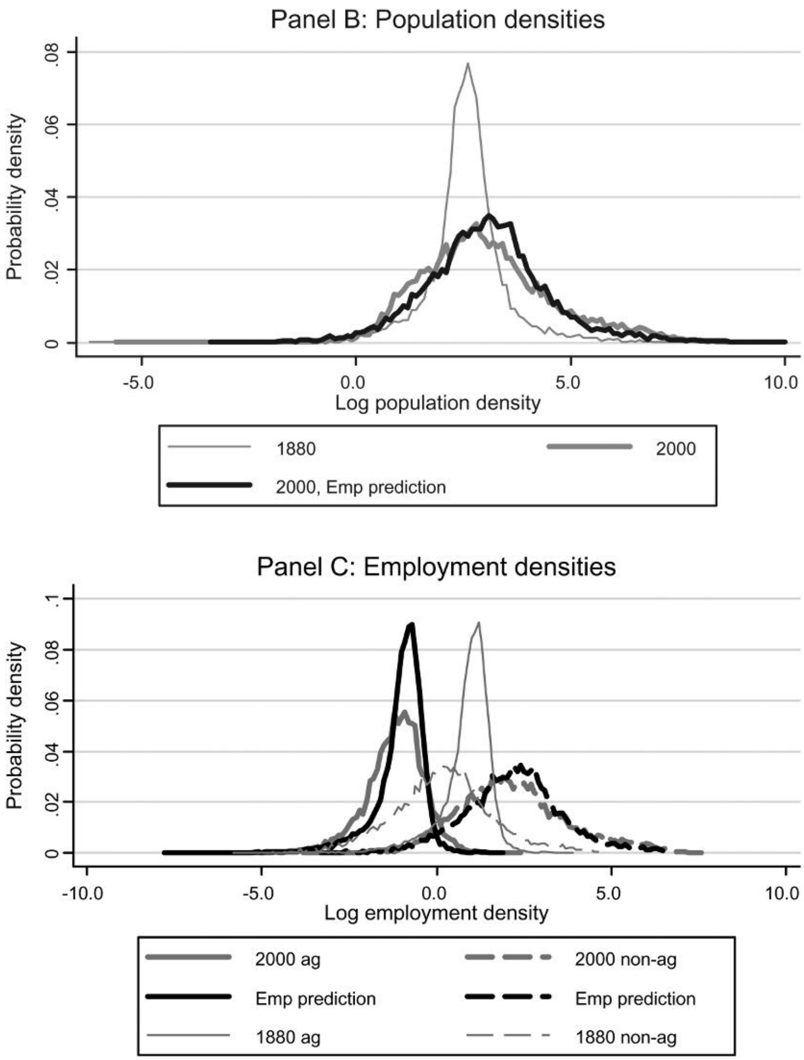


FIGURE II

(continued)

Figure II, we find a highly statistically significant coefficient of 0.438 (standard error 0.053) and a regression R^2 of 0.41.

Although the Employment Shares prediction captures the increasing relationship between population growth and initial population densities at intermediate densities and the roughly constant relationship at high densities, the decreasing relationship at low densities is less apparent than for actual population growth. One potential reason is that the Employment Shares prediction assumes a constant growth rate of employment in each sector equal to the value for the United States as a whole, and hence does not allow for mean reversion in agricultural employment growth (Stylized Fact 5). To allow the relationship between employment growth and initial population density to vary with the initial share of agriculture in employment, we consider a second Regression specification, which is based on regressing actual employment growth on log population density in 1880, the share of agriculture in employment in 1880, and the interaction between these two variables:

$$\begin{aligned}\Delta \ln E_{mt} = b_0 + b_1 \frac{E_{Amt-T}}{E_{mt-T}} + b_2 \ln \frac{L_{mt-T}}{H_m} \\ + b_3 \left(\frac{E_{Amt-T}}{E_{mt-T}} \times \ln \frac{L_{mt-T}}{H_m} \right) + u_{mt},\end{aligned} \tag{10}$$

where $\frac{L_{mt}}{H_m}$ denotes population density; $\{b_0, b_1, b_2, b_3\}$ are parameters that we estimate; the main effect of initial population density (b_2) allows for the possibility of mean reversion in nonagriculture; the coefficient on the interaction term (b_3) allows the degree of mean reversion to vary with the share of agriculture in employment; u_{mt} is a stochastic error.

From the regression's fitted values, we obtain a prediction for the total employment growth rate for each MCD from 1880 to 2000. Our Regression specification scales up observed MCD 1880 total employment (E_{m1880}) by this predicted total employment growth rate ($1 + \hat{g}_{Em}$) to obtain predicted MCD 2000 total employment ($\hat{E}_{m2000} = (1 + \hat{g}_{Em}) E_{m1880}$). Using predicted MCD 2000 total employment, we generate predictions for MCD 2000 population and MCD population growth from 1880 to 2000 following the same method as for the Employment Shares specification (9)

Table II reports the results of estimating the regression (10). In column (1), we find a negative and statistically significant relationship between population growth and the initial share of agriculture in employment, which is robust to controlling for initial log population density in column (2). When we include the interaction term in column (3), we find a negative and statistically

TABLE II
REGRESSION PREDICTION FOR MCDs

Employment growth 1880–2000	(1)	(2)	(3)
Agricultural employment share in 1880	−0.0096***	−0.011***	−0.0083***
	(0.0007)	(0.00092)	(0.0015)
Log population density in 1880		0.00067***	−0.00022
		(0.00022)	(0.00025)
Interaction term			−0.00099**
			(0.0005)
Observations	10,856	10,856	10,856
R-squared	0.06	0.06	0.06

Notes. This table reports the regressions used to generate the Regression prediction for our U.S. MCD data. Observations are a cross-section of MCDs from 1880–2000 for our baseline sample of A and B states. The agricultural employment share is employment in agriculture as a fraction of total employment. The interaction term is the product of the other two variables. Robust standard errors in parentheses are clustered by county. See the text and the Online Appendix for further discussion of the construction of the data.
* Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

significant coefficient on this variable, which is consistent with greater mean reversion in employment growth in agriculture than in nonagriculture. We use the fitted values from column (3) in our Regression prediction.

In Panel A of Figure II, we display mean predicted population growth (labeled Reg prediction) across initial log population density bins. We again capture the initially decreasing, later increasing, and finally roughly constant relationship between population growth and initial population density. Regressing mean actual population growth on mean predicted population growth across the initial log population density bins shown in Panel A of Figure II, we find a positive and statistically significant coefficient of 1.162 (standard error 0.118) and a regression R^2 of 0.65.

In Panels B and C of Figure II, we show that the Employment Shares prediction can account not only for population growth (Stylized Fact 2) but also for changes in the population density distribution (Stylized Fact 1) and the employment density distributions for each sector (Stylized Fact 4). We find a similar pattern of results using the Regression prediction, as shown in Figures A.3 and A.4 in the Online Appendix. In both cases, there is a close correspondence between the actual and predicted distributions for population density and employment density in each sector.

VI. Further Evidence

Having shown that structural transformation away from agriculture provides a parsimonious explanation for the stylized facts, we now provide further evidence in support of this explanation and against potential alternative explanations.

VI.A. Robustness

In this section, we first examine the robustness of the Employment Shares and Regression predictions in the previous section. We next demonstrate the robustness of the stylized facts in Table I across a number of different samples and specifications.

A first concern about the Employment Shares and Regression predictions is the role of structural transformation within nonagriculture from manufacturing to services. In Figure A.5 in the Online Appendix, we use our county subperiods data to show the shares of agriculture, manufacturing, and services in aggregate U.S. employment over time. In Figure A.6 in the Online Appendix, we use the same data to examine the population densities at which these sectors are concentrated. For each sector, we calculate its employment-weighted average of county log population densities. Consistent with other studies such as Desmet and Rossi-Hansberg (2009), we find that in recent decades manufacturing has dispersed to lower densities, while services has continued to concentrate at higher densities. Over a longer time horizon from 1880 until around 1980, we find that employment in both manufacturing and services shifted toward higher densities, with services displaying the larger change.

To show that our predictions for MCD population growth in the previous section capture reallocation from agriculture to nonagriculture as a whole, rather than reallocation within nonagriculture, we replicate the Employment Shares and Regression predictions treating manufacturing and services as separate sectors. As shown in Figures A.7 and A.8 in the Online Appendix, we find that disaggregating nonagriculture contributes relatively little to the ability of structural transformation to explain patterns of population growth from 1880–2000. This finding reflects the following two features of the data, as discussed further in the Online Appendix. First, the difference in aggregate employment growth rates from 1880 to 2000 between agriculture and nonagriculture (–0.016 versus 0.017) is much larger than that between manufacturing and services (0.013 versus 0.018). Second, the variation in the shares of agriculture and nonagriculture in 1880 total employment across 1880 population density bins (from around 0.8 to 0.1) is much larger than the variation in the share of manufacturing and services in 1880 nonagricultural employment (from around 0.2 to 0.4). We present further evidence on the robustness of our findings to structural transformation within nonagriculture when we consider subperiods (Section VI.B) and when we augment our population growth regressions with additional controls (Section VI.C.).[31]

A second concern about the Employment Shares and Regression predictions is the extent to which population growth in each MCD is affected by the characteristics of neighboring MCDs. To address this concern, we augment the Regression prediction (10) with the initial agricultural employment share and initial population density for the county of which the MCD is part as well as their interaction. Although the county variables are statistically significant, their inclusion again adds little to the ability of structural transformation to explain patterns of population growth, as shown in Figure A.10 in the Online Appendix. In the remainder of our empirical analysis, we control for spatial autocorrelation by clustering the standard errors on county and including county fixed effects in some of our empirical specifications.

We now turn to the robustness of the stylized facts in Table I. A first potential concern is their sensitivity to imperfections in the matching of MCDs across censuses. For example, in cases where MCDs are aggregated to construct geographic units that are consistent over time, some of the population and employment of MCDs with intermediate densities could be assigned to MCDs with either higher or lower densities, which could in turn influence relative population growth at different densities. To address this concern, the second column of Table I shows that all our stylized facts remain intact when we restrict the sample to MCDs in the "A states," where match rates with no aggregation are over 90% for every state. As further confirmation of the robustness of our results, Panel A of Figure A.11

31. When we use our county subperiods data and focus on the subperiod 1960–2000, we find that disaggregating nonagriculture into manufacturing and services has more of an impact on predicted population growth across initial population densities, as shown in Figure A.9 and discussed further in the Online Appendix. But even for this more recent subperiod the differences between manufacturing and services are smaller than those between agriculture and nonagriculture as a whole.

in the Online Appendix shows that the pattern of an initially decreasing, later increasing, and finally roughly constant relationship between population growth and initial population density (Stylized Fact 2) is still strongly apparent in this sample.[32]

Two other concerns are the extent to which the stylized facts are affected by the incomplete geographical coverage of our baseline sample (A and B states) and the particular level of spatial aggregation used in the analysis (MCDs rather than counties). To address both points, the third column of Table I reports results using our county data set, which covers almost all of the continental United States, as discussed in Section IV. In the fourth column, we compare results using MCD and county data for a common geographical sample by restricting the counties sample to the A and B states. In the fifth column, we provide additional evidence that our findings are not driven by incomplete geographical coverage by reporting results for a hybrid sample, which includes MCDs for all states for which we have subcounty data (A, B, and C states) and counties otherwise.

Across columns (3)–(5), we find a similar pattern of results with two caveats. In column (3), the standard deviation of log population in 1880 is higher than in 2000, although the difference is not statistically significant at conventional critical values. In columns (3) and (5), there is some evidence of mean reversion in both agriculture and nonagriculture, although in both cases there is substantially less mean reversion in nonagriculture than in agriculture. These findings are perhaps not surprising because the samples in columns (3) and (5) include western states that were not yet fully settled in 1880. Early settlement dynamics in these states, around the time of the "closing of the frontier" (identified in the 1890 census), could be different from those elsewhere. As the western states include areas that were largely uninhabited in 1880, they have correspondingly high standard deviations of log population in 1880.[33] The subsequent settlement of these largely uninhabited areas provides a natural explanation for some mean reversion in nonagriculture. Despite these caveats,

32. See the Online Appendix for further discussion of the samples and specifications in Table I and Figure A.11.

33. Consistent with this, we find that the higher standard deviation of log population in 1880 than in 2000 is driven by a tail of very sparsely populated counties in 1880. The interquartile range of the population distribution is greater in 2000 than in 1880, so that Stylized Fact 3 is confirmed using measures of dispersion that are less sensitive to the tails of the distribution.

Panels B–D of Figure A.11 show that the pattern of departures from constant population growth (Stylized Fact 2) is strongly apparent in all three samples.

While columns (3)–(5) address the concern of imperfect geographical coverage, a related but separate issue is the sensitivity of our results to the westwards expansion of the frontier of U.S. settlement. In particular, the higher average growth in low-density locations could be driven by the dramatic growth of a small number of places in the West that were largely unpopulated in 1880 and grew rapidly in the succeeding decades. To address this concern, we restrict the MCD sample to the subset of the A and B states that were part of British colonial claims in 1775.[34] In this subsample, which includes only states along the eastern seaboard of the United States, we find a very similar pattern of results, as shown in column (6) of Table I and Panel E of Figure A.11.

Although MCDs provide a fine level of spatial disaggregation, there remains the concern that they may not correspond to economic units if their boundaries do not coincide with local labor and product markets, especially around cities. To address this concern, we aggregate MCDs within the boundaries of each 2000 MSA. As shown in column (7) of Table I and Panel F of Figure A.11, we observe the same pattern of stylized facts in this sample. One potential concern about this specification is the endogeneity of 2000 MSA boundaries to population growth during our sample period. To address this concern, we also construct historical metropolitan areas based on classifying cities as MCDs where log population per square kilometer in 1880 was larger than 6. To each of these cities, we add the land area, population, and employment of any MCD whose geographic centroid lies within 25 km of that city.[35] As shown in column (2) of Table A.1 and Panel A of Figure A.12 in the Online Appendix, we again find the same pattern of results.

A related but somewhat different concern is the role of suburbanization or a more general shift of population to lower-density

34. The states included in this sample are Connecticut, Delaware, Georgia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, North Carolina, Pennsylvania, Rhode Island, South Carolina, Vermont, Virginia, and West Virginia.

35. When two or more cities and their surrounding areas overlap, we merge them together. We experimented with other historical definitions of metropolitan areas, including defining "cities" as MCDs with 50,000 or 100,000 or more inhabitants in 1880 and using a distance threshold of 50 km. Using these alternative definitions yields a similar pattern of results.

areas. As a first step toward addressing this concern, column (8) of Table I and Panel B of Figure A.12 report results in which we drop from our sample all MCDs within the boundaries of a 2000 MSA. In column (3) of Table A.1 and Panel C of Figure A.12, we report analogous results in which we drop from our sample all MCDs within the boundaries of a historical metropolitan area as defined above. For both these nonmetropolitan samples, the stylized facts are confirmed.

As a further robustness check, we exclude MCDs proximate to urban areas by restricting the sample to MCDs with centroids more than 100 km from the centroid of a metropolitan area, as defined using either 2000 MSAs or the historical metropolitan areas discussed above. As shown in columns (4) and (5) of Table A.1 and Panels D and E of Figure A.12, we find a similar pattern of results in both subsamples. Although these findings already provide strong evidence against an explanation based on suburbanization, we present further evidence below when we consider subperiods (Section VI.B) and when we augment our population growth regressions with additional controls (Section VI.C).

Another potential concern is that converting land from agricultural to non-agricultural use could be easier than converting land from nonagricultural to agricultural use, which could be responsible for differences in the degree of mean reversion in employment growth across sectors. Because our sample period is characterized by a large-scale reallocation of employment and land from agricultural to nonagricultural use, difficulties in the conversion of land in the reverse direction are unlikely to be the dominant influence on employment and population growth. But they could contribute towards differences in the degree of mean reversion across sectors. To address this concern, column (6) of Table A.1 and Panel F of Figure A.12 report results excluding the 278 MCDs in our baseline sample that experienced a decline in nonagricultural employment between 1880 and 2000. Again the stylized facts are confirmed with greater mean reversion in agriculture than in nonagriculture.

Finally, although the results in this section confirm that our stylized facts are robust features of the evolution of population and employment in the United States from 1880 to 2000, we have also replicated our entire analysis for Brazil from 1970 to 2000. Although there are many differences between the two countries and time periods, both are characterized by substantial structural transformation away from agriculture, and hence we would expect the stylized facts to apply. As discussed in the Online Appendix, we find a strikingly similar pattern of results for Brazil, which reassures us that our findings are not driven by idiosyncratic features of the data or institutional environment for the United States and confirms the relevance of our results for a developing country in recent decades.

VI.B. Timing of Structural Transformation

According to our explanation for the stylized facts, the increase in population growth over the range of intermediate initial population densities is driven by an aggregate reallocation away from agriculture combined with a sharp decline in agriculture's share of employment over this range of intermediate densities. A key implication of this explanation is that the increase in population growth over the range of intermediate densities should be stronger in subperiods and regions characterized by greater structural transformation away from agriculture.

To examine this implication, we require data on both population growth and sectoral employment for years in between 1880 and 2000, which are available for counties but not MCDs. Since structural transformation away from agriculture was largely complete in the United States by 1960, we begin by splitting our sample into the three subperiods of 1880–1920, 1920–1960 and, 1960–2000.[36] For each of these subperiods, we compute actual and predicted population growth using the same approach as in Section V.C. While we focus on the Employment Shares predictions, which use aggregate sectoral employment growth for each subperiod and initial shares of agriculture in MCD employment at the beginning of each subperiod, we find similar results with the Regression predictions.

As shown in Panels A–C of Figure III, the increase in population growth over the range of intermediate densities is strongly apparent from 1880–1920 and 1920–1960 for both actual and predicted population growth, which is consistent with the sharp decline in agriculture's share of aggregate employment during these periods.[37] In contrast, actual and predicted population growth are

36. As shown in Figure A.5 in the Online Appendix, the share of agriculture in aggregate U.S. employment declines rapidly until around 1960, after which it converges to less than 2%.

37. While Panels A–C of Figure III focus on initial population densities between 0 and 6 log points to highlight the contrasts between the earlier and later subperiods, a similar pattern of results is observed across the full range of initial population densities.

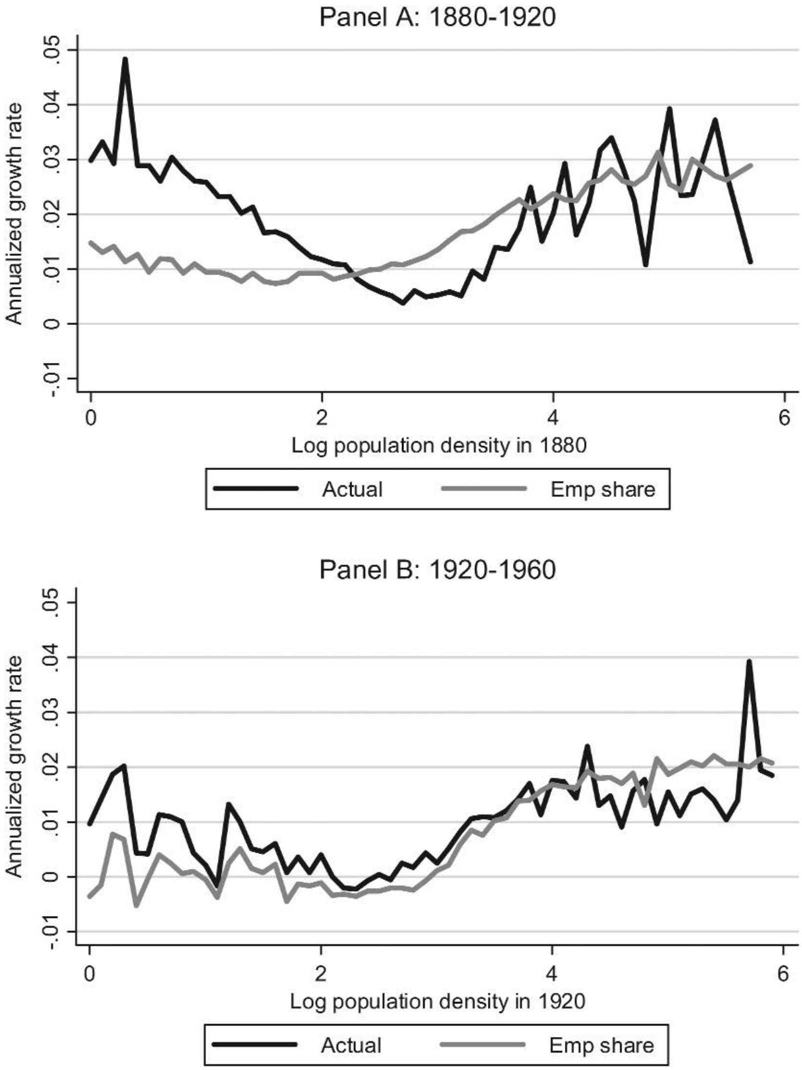


FIGURE III

County Subperiod Results

Panels A, B, and C show actual population growth and predicted population growth based on the Employment Shares prediction for the full sample of states from our county subperiods data set. In these panels, the x-axes are population density bins, defined by rounding down log initial population density for each MCD to the nearest single digit after the decimal point. The y-axes show means for each population density bin. Since population density bins at the extreme ends of the

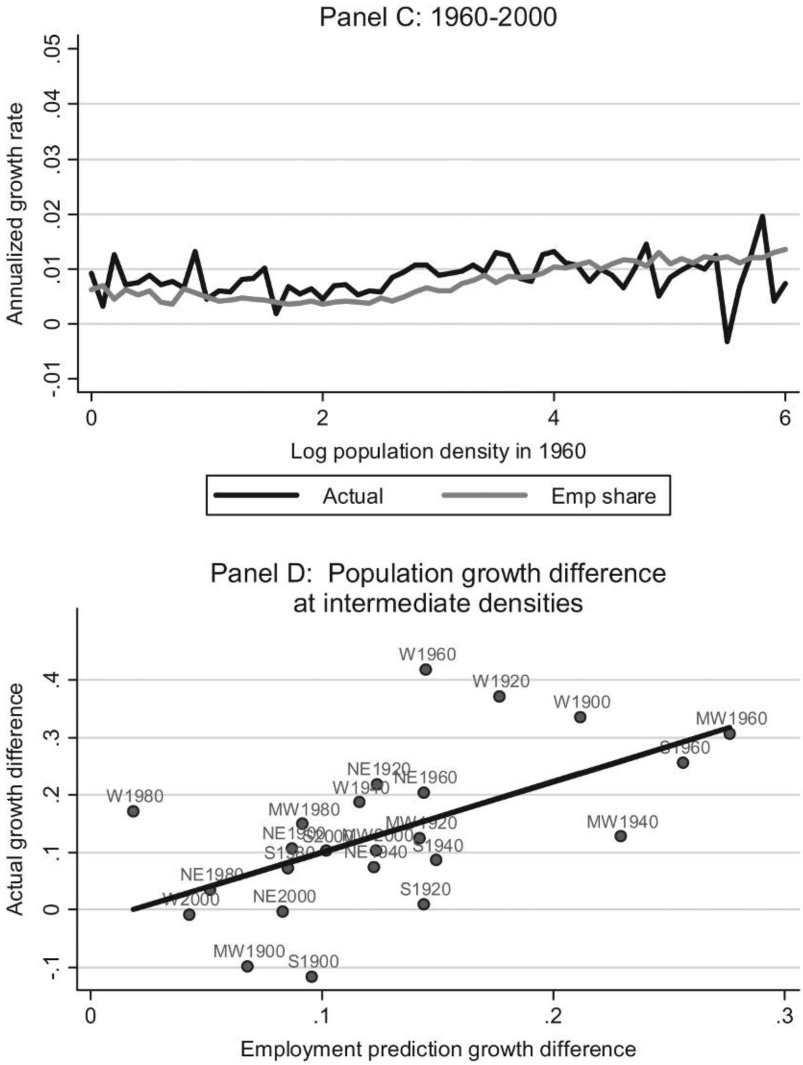


FIGURE III

(continued)

distribution typically contain few observations, the figures in panels A–C (but not the estimations) focus on the density range 0–6. Panel D shows the difference in mean population growth for 3–5 minus 1–3 log points of initial population density. The y-axis and x-axis show this difference for actual population growth and predicted population growth based on the Employment Shares prediction, respectively. The figure uses our county subperiods data. See the text and the Online Appendix for further discussion of the construction of the figures.

largely uncorrelated with initial population density from 1960 to 2000, which is consistent with agriculture's small share of aggregate employment during this period and the lack of correlation between nonagricultural employment growth and initial population density. In Figure A.13 of the Online Appendix, we show that the rise in population growth at intermediate densities in 1880–1920 and 1920–1960 coincides with a sharp decline in the share of agriculture in employment at these intermediate densities.[38]

The contrast between Panels A–B and Panel C of Figure III provides further evidence against an alternative explanation based on reallocation from manufacturing to services. Such reallocation was more important from 1960–2000 than from 1880–1920 and 1920–1960, yet the increasing relationship between population growth and initial population density at intermediate densities is stronger in the two earlier subperiods than in the later subperiod.[39] These results also provide further evidence against an alternative explanation based on suburbanization, which was more important from 1960–2000.[40]

To tighten the link between the timing of structural transformation and the increase in population growth over the range of intermediate initial population densities, we now use variation across each 20-year interval and Census region. For both MCDs (Panel B of Figure I) and counties (Panels A–B of Figure III), the increasing relationship between population growth and initial population density is observed for initial population densities in between roughly 2 and 4 log points. Because there are far fewer counties than MCDs and we now consider Census regions and subperiods separately, some initial population density bins have few counties for some Census regions. For example, the West and North-East Census regions each have around 200–250 counties compared to more than 10,000 MCDs in

38. For later 20-year periods, we find some evidence of a decline in the population growth of counties with the highest initial population densities, which is consistent with the results of Holmes and Lee (2010) using census tract data for 1990–2000. However, this feature appears toward the end of our sample period and is less precisely estimated using county rather than census tract data, which makes it harder to distinguish from constant population growth.

39. As shown in Figure A.5 in the Online Appendix, the shares of manufacturing and services in aggregate employment track one another relatively closely until around 1960, after which manufacturing's share declines and services's continues to rise.

40. For empirical evidence on suburbanization following the construction of the interstate highway system from the late 1950s onward, see Baum-Snow (2007).

our baseline sample for the United States as a whole. Therefore, we examine mean population growth rates over ranges of initial population density bins centered on 2 and 4 log points. In particular, we consider the ranges of 1–3 versus 3–5 log points, which capture the increase in population growth at intermediate initial population densities in Stylized Fact 2.

In Panel D of Figure III, we display the difference in mean population growth between these two ranges (3–5 minus 1–3) for both actual population growth (y-axis) and predicted population growth based on the Employment Shares prediction (x-axis). Points are labeled according to Census region codes and the final year of the interval over which population growth is computed, so that 1960 corresponds to the subperiod 1940–1960.[41]

Across subperiods and regions, we observe a strong positive relationship between the actual increase in population growth at intermediate densities and the predicted increase based on structural transformation away from agriculture. Regressing the actual increase in population growth between the two ranges (3–5 minus 1–3) on the predicted increase, we find a positive and statistically significant coefficient (standard error) of 1.230 (0.322), as shown in the regression line in Panel D of Figure III. Augmenting this regression with region and subperiod fixed effects, we continue to find a positive and significant relationship, with a coefficient (standard error) of 1.480 (0.295).

Looking across 20-year intervals in Panel D, the size of the actual and predicted increase in population growth at intermediate densities is larger for 1940–1960 (labeled 1960) than for 1920–1940 (labeled 1940) or for later subperiods. This pattern of variation is consistent with the historical literature on the development of U.S. agriculture, which emphasizes the role of technological change as a driver for structural transformation, including mechanization (U.S. Department of Agriculture 1947) and biological innovation (Griliches 1957). This historical literature highlights a deceleration of technological change from 1920 to 1940, which includes the agricultural depression of the 1920s and the Great Depression of the 1930s, and an acceleration of technological change from 1940 to 1960, as increased demand for U.S. agricultural products in the years surrounding World War II stimulated the adoption of productivity-enhancing technologies (see for example Rasmussen 1962).

41. The region labels are: MW (Midwest), NE (Northeast), S (South), and W (West). Washington D.C. is assigned to the South.

Looking across regions in Panel D, the size of the actual and predicted increase in population growth at intermediate densities from 1940 to 1960 (labeled 1960) is larger for the Midwest, South, and West than the Northeast. This pattern accords with the literature on regional development in the United States, which emphasizes the later timing of structural transformation in the South than in the North (Wright 1986; Caselli and Coleman 2001), and the role of technological change in the years surrounding World War II in increasing agricultural productivity in the Midwest and West (Cochrane 1979).

While Panel D is based on the Employment Shares prediction, we find a similar pattern of results using the Regression prediction, as shown in Figure A.14 in the Online Appendix.[42] For both sets of predictions, there are a small number of cases of zero or small negative increases in population growth between the ranges of 1–3 and 3–5 log points, which reflects mean reversion that raises population growth rates toward the bottom of the 1–3 range.

As a final piece of evidence on the timing of structural transformation, we use our full panel of data on counties and 20-year subperiods. Columns (1) and (2) of Table III show that actual and predicted population growth rates (using both the Employment Shares and Regression predictions) are strongly correlated even after controlling for county and subperiod fixed effects. This specification has a "differences-in-differences" interpretation: in counties and subperiods where predicted population growth is higher than implied by the means for the county and subperiod, actual population growth is also higher than implied by the means for the county and subperiod. In columns (3)–(4) of Table III, we show that these findings are robust to replacing the subperiod fixed effects with state×subperiod fixed effects, which allow for heterogeneous growth rates across states. In both specifications, we find a close connection between actual population growth and predicted population growth based on structural transformation away from agriculture.

42. Regressing the actual increase in population growth between the ranges of 1–3 and 3–5 log points of initial population density on the predicted increase from the Regression prediction, we find a positive and statistically significant coefficient (standard error) of 0.619 (0.162). Augmenting this regression with region and subperiod fixed effects, we continue to find a positive and statistically significant relationship, with a coefficient (standard error) of 1.167 (0.163).

TABLE III

EXPLANATORY POWER OF THE EMPLOYMENT SHARE AND REGRESSION PREDICTIONS FOR COUNTIES

Population growth	(1)	(2)	(3)	(4)
Population growth predicted by Employment Shares	0.503*** (0.0354)		0.382*** (0.0377)	
Population growth predicted by Regression		0.641*** (0.0183)		0.651*** (0.0201)
Year fixed effects	Yes	Yes		
County fixed effects	Yes	Yes	Yes	Yes
State-year fixed effects			Yes	Yes
Observations	14,496	14,696	14,496	14,496
R-squared	0.42	0.62	0.52	0.68

Notes. The table uses our county subperiods data. Observations are counties across 20-year time intervals from 1880–2000. Population growth and predicted population growth rates are measured by annualized log differences. Robust standard errors clustered by county are shown in parentheses.The sample includes all U.S. states except Alaska, Hawaii, North Dakota, Oklahoma, South Dakota, and Wyoming. See the text and the Online Appendix for further discussion of the construction of the data.
* Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

VI.C. *Potential Alternative Explanations*

In this section, we return to our more spatially disaggregated MCD data to provide additional evidence against a variety of possible alternative explanations. Table IV reports the results of regressing actual population growth on predicted population growth based on structural transformation away from agriculture for our baseline MCD sample for the A and B states from 1880 to 2000. Panels A and B use the Employment Shares and Regression predictions, respectively.[43]

In column (1) of Table IV, we regress actual on predicted population growth without controls and find a positive and statistically significant coefficient. For the Employment Shares prediction, the coefficient on predicted population growth is less than 1, which reflects the fact that actual population growth responds more sharply to initial population density at low and intermediate densities than predicted population growth (see Panel A of Figure II). While many idiosyncratic factors can affect actual population

43. While Table IV reports standard errors clustered by county, Table A.3 in the Online Appendix reports standard errors based on the alternative approach to allowing for spatial correlation of Bester, Conley, and Hansen (2011). Both procedures result in similar standard errors, so that all statements about statistical significance are robust to the use of either approach.

TABLE IV

EXPLANATORY POWER OF THE EMPLOYMENT SHARE AND REGRESSION PREDICTIONS

Actual population growth	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		As (1) with geographic controls	As (1) with demographic controls	As (1) with distance to 1947 highway	As (1) with distance to 1898 railroad	As (1) with distance to 2000 MSA	As (1) with manufacture share in non-agriculture	As (1) with 1880 density bin controls	As (1) with county fixed effects	All controls used in (2) to (9)
Panel A : Employment Shares prediction										
Predicted Population	0.449^{***}	0.334^{***}	0.262^{***}	0.401^{***}	0.400^{***}	0.413^{***}	0.418^{***}	0.324^{***}	0.271^{***}	0.223^{***}
Growth	(0.028)	(0.0244)	(0.025)	(0.026)	(0.026)	(0.025)	(0.026)	(0.029)	(0.016)	(0.020)
Observations	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864
R-squared	0.09	0.18	0.14	0.14	0.14	0.09	0.10	0.15	0.62	0.67
Panel B : Regression prediction										
Predicted Population	0.978^{***}	0.718^{***}	0.581^{***}	0.905^{***}	0.903^{***}	0.892^{***}	0.913^{***}	0.645^{***}	0.545^{***}	0.362^{***}
Growth	(0.054)	(0.052)	(0.067)	(0.052)	(0.052)	(0.050)	(0.053)	(0.065)	(0.040)	(0.047)
Observations	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864
R-squared	0.10	0.18	0.15	0.15	0.16	0.20	0.10	0.15	0.62	0.66

Notes. All regressions use MCD data for our baseline sample of A and B states. Observations are a cross-section of MCDs from 1880–2000. In all specifications, the dependent variable is actual population growth from 1880–2000. Panel A uses predicted population growth from the Employment Shares prediction. Panel B uses predicted population growth from the Regression prediction. In column (2), the geographical controls are measures of proximity to rivers, lakes, coastlines, and mineral resources. In column (3), the demographic controls are the share of the population that is white, the share of the population born outside the state (as a measure of national and international migration), the share of the population aged less than 6 (as a measure of fertility), and the share of the population aged 14–18 in education (as a measure of educational attainment). Column (4) includes distance from the centroid of each MCD to the closest interstate highway in the 1947 plan. Column (5) includes distance from the centroid of each MCD to the closest railroad in the 1898 railroad network. Column (6) includes distance from the centroid of each MCD to the centroid of the closest 2000 MSA. In column (7), we include the 1880 share of manufacturing in nonagricultural employment. Column (8) includes 1880 population density bin fixed effects. Column (9) includes county fixed effects. In Column (10), we include all controls from columns (2) through (9). In columns (4) through (6) and (10), we use log(1+distance to transportation system and/or city). See the text and the Online Appendix for further discussion of the construction of the data. Robust standard errors clustered by county are shown in parentheses.
* Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

growth in individual MCDs, predicted population growth alone accounts for around 10% of the variation in actual population growth in both panels of Table IV.

In column (2), we consider local differences in physical geography and natural endowments, as highlighted by for example Rappaport and Sachs (2003). Controlling for measures for proximity to rivers, lakes, coastlines, and mineral resources, we continue to find a positive and statistically significant effect of structural transformation. In column (3), we examine local variation in demography, fertility, and education, as discussed for example in Beeson, DeJong, and Troesken (2001). Including the share of each MCD's population that is white, the share born outside of the MCD's state (as a measure of national and international migration), the share aged less than 6 (as a measure of fertility), and the share aged 14–18 in education (as a measure of educational investments), the effect of structural transformation again remains robust.

In columns (4)–(6), we examine changes in transport technology and suburbanization, as considered by Baum-Snow (2007), Michaels (2008), and Duranton and Turner (2010), which could have contributed to a more general shift in population toward lower densities. In column (4), we include the distance from the centroid of each MCD to the closest interstate highway based on the 1947 plan. In column (5), we introduce distance from the centroid of each MCD to the closest railroad based on the 1898 railroad network. In column (6), we control for distance between the centroid of each MCD and the centroid of the closest 2000 MSA. Across all three columns, we continue to find a statistically significant effect of structural transformation away from agriculture. While column (6) is based on 2000 MSAs, we find a similar pattern of results using the historical definition of metropolitan areas discussed in Section VI.A.[44]

In column (7), we examine structural transformation within nonagriculture from manufacturing to services, as considered in Desmet and Rossi-Hansberg (2009). To control for such reallocation, we include the initial share of manufacturing in nonagricultural employment in each MCD as an additional regressor. Again we find similar results, consistent with our findings in Section VI.A that disaggregating nonagriculture into

44. Reestimating the two specifications in column (6) using the historical definition of metropolitan areas, we find estimated coefficients (standard errors) in Panels A and B of 0.427 and 0.906 (0.025 and 0.054), respectively.

manufacturing and services contributes relatively little to the explanatory power of structural transformation over our long historical time period.

In column (8), we include a full set of fixed effects for initial log population density bins. While our earlier analysis of population growth in Panel A of Figure II used variation across initial log population density bins, here we only exploit variation within initial log population density bins. Again we find a positive and statistically significant coefficient. Therefore, even among MCDs with similar initial log population densities, structural transformation away from agriculture has predictive power for population growth.

In column (9), we include a full set of county fixed effects, which control for any observed or unobserved characteristics of counties that affect rates of population growth. These characteristics include state and county policies and institutions as well as physical geography, including climate as emphasized by Glaeser (2008), since climate typically varies relatively little within counties. Again we find a positive and statistically significant effect of structural transformation away from agriculture.

In column (10), we demonstrate that the statistical significance of structural transformation is robust to simultaneously including the full set of controls considered in columns (2)–(9). Some of these controls, such as demography, fertility, and education, are likely to be endogenously influenced by structural transformation away from agriculture, which implies that some of structural transformation's effect is likely to be attributed to the controls. Furthermore, this specification includes both county fixed effects and initial population density bins, which implies that the effect of structural transformation is identified from variation across MCDs within counties with similar initial population densities but different initial patterns of specialization between agriculture and nonagriculture. Remarkably, even using this limited variation and including all of our controls, we continue to find statistically significant effects of structural transformation away from agriculture.

VII. Conclusion

While as recently as the nineteenth century around one sixth of the world's population lived in cities, urban residents now account for a growing majority of the world's population.

Arguably few other economic changes have involved as dramatic a transformation in the organization of society. In this article, we provide theory and evidence on patterns of urbanization using a new data set that enables us to trace the transformation of the U.S. economy from a predominantly rural to a largely urban society.

Our analysis has two main contributions. First, we provide evidence of six stylized facts that are robust features of data sets that cover both rural and urban areas over time periods characterized by substantial structural change. These stylized facts encompass empirical regularities from existing research for densely populated locations, but also introduce hitherto neglected features of the data, such as an increasing relationship between population growth and initial population density observed at the intermediate densities where most of the population lived historically.

Second, we provide a simple explanation for the stylized facts, which emphasizes an aggregate reallocation of employment away from agriculture combined with cross-section differences in agriculture's share of initial employment. At low population densities where agriculture dominates initial employment, mean reversion in agricultural productivity generates a decreasing relationship between population growth and initial population density. At high population densities where nonagriculture dominates initial employment, a largely constant rate of nonagricultural productivity growth generates population growth that is largely uncorrelated with initial population density. In between, the share of agriculture in initial employment is decreasing in population density, and structural transformation from agriculture to nonagriculture raises population growth at higher densities with lower shares of agriculture in employment.

According to our explanation for the stylized facts, the increasing relationship between population growth and initial population density at intermediate densities should be more pronounced in periods and regions characterized by greater structural transformation away from agriculture. We find strong confirmation that this is indeed the case. Although there are many factors that can potentially influence population growth, the close relationship between employment structure and population growth in both the United States and Brazil, the tight connection between the timing of structural transformation and our findings, and the predictive power and robustness of our results suggest that structural transformation is a key part of the urbanization process.

London School of Economics
London School of Economics
Princeton University

Supplementary Material

An Online Appendix for this article can be found at QJE online (qje.oxfordjournals.org).

References

Bairoch, Paul, *Cities and Economic Development: From the Dawn of History to the Present* (Chicago, University of Chicago Press, 1988).

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics*, 122 (2007), 775–805.

Baumol, William J., "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crises," *American Economic Review*, 57 (1967), 415–426.

Beeson, Patricia E., David N. DeJong, and Werner Troesken, "Population growth in U.S. counties, 1840–1990," *Regional Science and Urban Economics*, 31 (2001), 669–699.

Bertrand, Marianne., Esther Duflo, and Sendhil Mullainathan, "How Much Should we Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119 (2004), 249–275.

Bester, C. Alan., Timothy G. Conley, and Christian B. Hansen, "Inference with Dependent Data Using Cluster Covariance Estimators," *Journal of Econometrics*, 165 (2011), 137–151.

Black, Duncan, and Vernon Henderson, "Urban Evolution in the U.S.A.," *Journal of Economic Geography*, 3 (2003), 343–372.

Caselli, Francesco, and Wilbor John Coleman II., "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, 109 (2001), 584–616.

Cochrane, Willard W., *The Development of American Agriculture: A Historical Analysis* (Minneapolis, University of Minnesota Press, 1979).

da Mata, Daniel, Uwe Deichmann, J. Vernon Henderson, Somik V. Lall, and Hyoung G. Wang, "Determinants of City Growth in Brazil," *Journal of Urban Economics*, 62, (2007), 252–272.

Davis, Donald R., and David Weinstein, "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, 92 (2002), 1269–1289.

Davis, Morris A., and François Ortalo-Magne, "Household Expenditures, Wages, Rents.," *Review of Economic Dynamics*, 14 (2011), 248–261.

Desmet, Klaus, and Esteban Rossi-Hansberg, "Spatial Growth and Industry Age," *Journal of Economic Theory*, 144 (2009), 2477–2502.

Duranton, Gilles, "Urban Evolutions: The Fast, the Slow, and the Still," *American Economic Review*, 97 (2007), 197–221.

Duranton, Gilles, and Diego Puga, "Micro-foundations of Urban Agglomeration Economies," in *Handbook of Regional and Urban Economics*, J. V. Henderson and J.F. Thisse, eds., vol. 4 (Amsterdam, Elsevier North Holland, 2004).

Duranton, Gilles, and Matthew A. Turner, "Urban Growth and Transportation," University of Toronto, Mimeo, 2010.

Eaton Jonathan, and Samuel S. Kortum, "Technology, Geography, and Trade," *Econometrica*, 70 (2002), 1741–1779.

Echevarria, Cristina, "Changes in Sectoral Composition Associated with Economic Growth," *International Economic Review*, 38 (1997), 431–452.

Eeckhout, Jan, "Gibrat's Law for (All) Cities," *American Economic Review*, 94 (2004), 1429–1451.

Ellison, Glenn, and Edward L. Glaeser, "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review*, 89 (1999), 311–316.

Foster, Andrew D., and Mark R. Rosenzweig, "Economic Development and the Decline of Agricultural Employment," in *Handbook of Development Economics*, T. Schultz and John Strauss, eds., vol. 4 (Amsterdam: Elsevier North-Holland, 2008).

Fujita, Masahisa., Paul Krugman, and Anthony J. Venables, *The Spatial Economy: Cities, Regions and International Trade* (Cambridge MA: MIT Press, 1999).

Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, 114 (1999), 739–767.

Gabaix, Xavier, and Yannis Ioannides, "The Evolution of City Size Distributions," in *Handbook of Regional and Urban Economics*, J. V. Henderson and J.F. Thisse, eds., vol. 4 (Amsterdam: Elsevier North-Holland, 2004).

Glaeser, Edward L., "The Rise of the Sunbelt," *Southern Economic Journal*, 74 (2008), 610–643.

Glaeser, Edward L., and Joseph Gyourko, "Urban Decline and Durable Housing," *Journal of Political Economy*, 113 (2005), 345–375.

Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer, "Growth in Cities," *Journal of Political Economy*, 100 (1992), 1126–1152.

Gollin, Douglas., Stephen Parente, and Richard Rogerson, "The Role of Agriculture in Development," *American Economic Review*, 92 (2002), 160–164.

González-Val, Rafael, Luis Lanaspa, and Fernando Sanz, "New Evidence on Gibrat's Law for Cities," MPRA Working Paper No. 10411, 2008.

Griliches, Zvi, "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 25 (1957), 501–522.

Helpman, Elhanan, "The Size of Regions," in *Topics in Public Economics: Theoretical and Applied Analysis*, David Pines, Efraim Sadka, and Itzhak Zilcha, eds. (Cambridge: Cambridge University Press, 1998).

Henderson, J. Vernon, "The Sizes and Types of Cities," *American Economic Review*, 64 (1974), 640–656.

Henderson, J. Vernon, and Anthony J. Venables, "The Dynamics of City Formation," *Review of Economic Dynamics*, 12 (2009), 233–254.

Holmes, Thomas J., and Sanghoon Lee, "Cities as Six-by-Six Mile Squares: Zipf's Law," in *Agglomeration Economics*, Edward L. Glaeser, ed. (Chicago: University of Chicago Press, 2010).

Ioannides, Yannis, and Henry G. Overman, "Spatial Evolution of the U.S. Urban System," *Journal of Economic Geography*, 4 (2004), 131–156.

Kim, Sukkoo, "Expansion of Markets and the Geographic Distribution of Economic Activities: The trends in U.S.," *Quarterly Journal of Economics*, 110 (1995), 881–908.

———, "Urban Development in the United States, 1690–1990." *Southern Economic Journal*, 66 (2000), 855–880.

Kim, Sukkoo, and Robert A. Margo, "Historical Perspectives on U.S. Economic Geography," in *Handbook of Regional and Urban Economics*, J. V. Henderson and J.F. Thisse, eds., vol 4 (Amsterdam: Elsevier North-Holland, 2004).

Krugman, Paul, "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99 (1991), 483–499.

Kuznets, Simon, *Modern Economic Growth: Rate, Structure and Spread* (New Haven, CT:Yale University Press, 1966).

Lewis, W. Arthur, "Economic Development with Unlimited Supplies of Labor," *Manchester School of Economic and Social Studies*, 22 (1954), 139–191.

Maddison, Angus, "Economic Growth and Structural Change in the Advanced Economies," in *Western Economies in Transition*, Irving Leveson and Jimmy W. Wheeler, eds. (London: Croom Helm, 1980).

Matsuyama, Kiminori, "The Rise of Mass Consumption Societies," *Journal of Political Economy*, 110 (2002), 1035–1070.

Michaels, Guy, "The Effect of Trade on the Demand for Skill—Evidence from the Interstate Highway System," *Review of Economics and Statistics*, 90 (2008), 683–701.

Michaels, Guy, Stephen J. Redding, and Ferdinand Rauch, "Technical Note: An Eaton and Kortum (2002) Model of Urbanization and Structural Transformation." Mimeo, London School of Economics and Princeton University, 2011.

Ngai, Rachel, and Chris Pissarides, "Structural Change in a Multisector Model of Growth," *American Economic Review*, 97 (2007), 429–443.

Olmstead, Alan L., and Paul W. Rhode, "The Red Queen and the Hard Reds: Productivity Growth in American Wheat, 1800–1940," *Journal of Economic History*, 62 (2002), 929–966.

Rappaport, Jordan, and Jeffrey Sachs, "The United States as a Coastal Nation," *Journal of Economic Growth*, 8 (2003), 5–46.

Rasmussen, Wayne D., "The Impact of Technological Change on American Agriculture, 1862–1962," *Journal of Economic History*, 22 (1962), 578–591.

Rogerson, Richard, "Structural Transformation and the Deterioration of European Labor Market Outcomes,"*Journal of Political Economy*, 116 (2008), 235–259.

Rosenthal, Stuart S., and William C. Strange, "Evidence on the Nature and Sources of Agglomeration Economies," in *Handbook of Urban and Regional Economics*, J. V. Henderson and J.F. Thisse, eds., vol. 4 (Amsterdam: Elsevier North-Holland, 2004).

Rossi-Hansberg, Esteban, and Mark L. Wright, "Urban Structure and Growth," *Review of Economic Studies*, 74 (2007), 597–624.

Sato, Ryuzo, "Homothetic and Non-Homothetic CES Production Functions," *American Economic Review*, 67 (1977), 559–569.

Scott, David W, "On Optimal and Data-based Histograms," *Biometrika*, 66 (1979), 605–610.

Soo, Kwok Tong, "Zipf's Law and Urban Growth in Malaysia," *Urban Studies*, 44 (2007), 1–14.

Syrquin, Moshe, "Patterns of Structural Change," in *Handbook of Development Economics*, H. Chenery and T. N. Srinivasan, eds., vol. 1 (Amsterdam: Elsevier North-Holland, 1988).

United Nations, "World Urbanization Prospects—The 2007 Revision," 2008, www.un.org/esa/population/publications/wup2007/2007wup.htm.

U.S. Department of Agriculture, *Progress of Mechanization*, Miscellaneous Publication No. 630 (Washington, DC: Department of Agriculture, 1947).

Wright, Gavin, *Old South New South: Revolutions in the Southern Economy since the Civil War* (New York: Basic Books, 1986).