# Urbanization and Structural Transformation*

Guy Michaels†
*London School of Economics, CEP and CEPR*
Ferdinand Rauch‡
*University of Vienna and CEP*
Stephen J. Redding§
*London School of Economics, CEP and CEPR*

14 April 2010

**Abstract**

We examine urbanization using new data that allows us to track the evolution of population across both rural and urban areas in the United States from 1880-2000. We find an upward-sloping relationship between initial population density and subsequent population growth for a range of intermediate densities, which induces polarization in the population distribution. We provide theory and evidence that these and our other findings about the population distribution are driven by structural transformation: an aggregate reallocation of employment away from agriculture and an agricultural employment share that decreases with initial population density.

Keywords: Urbanization, Structural Transformation, Population Growth
JEL: E00, N10, O18, R12

*This paper is produced as part of the Globalization Programme of the ESRC-funded Centre for Economic Performance (CEP). We thank Yu-Hsiang Lei, Ralph Ossa, and Bart Thia for excellent research assistance. We also thank Robin Burgess, Francesco Caselli, Tim Guinnane, Alan Manning, Rachel Ngai, Steve Pischke, Chris Pissarides and seminar participants at CEPR, Helsinki, LSE, Oxford, Tel-Aviv University, Tinbergen Institute, University of Minnesota, University of Missouri and Zurich for helpful comments. Responsibility for any opinions, results and errors lies with the authors alone.

†Department of Economics, Houghton Street, London, WC2A 2AE, United Kingdom, *tel:* (44) 20 7852 3518, *fax:* (44) 20 7955 7595, *email:* g.michaels@lse.ac.uk.

‡Department of Economics, Hohenstaufengasse 9, A-1010, Vienna, Austria, *tel:* (43) 1 4277 37455, *fax:* (43) 1 4277 9374, *email:* ferdinand.rauch@univie.ac.at.

§Department of Economics, Houghton Street, London, WC2A 2AE, United Kingdom, *tel:* (44) 20 7955 7483, *fax:* (44) 20 7955 7595, *email:* s.j.redding@lse.ac.uk.

# 1 Introduction

Urbanization – the concentration of population in cities and towns – is one of the most striking features of economic development.[1] 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.[2] In this paper, we examine urbanization using a new dataset that tracks, for the first time, the evolution of population across both rural and urban areas and over a long historical period of time. 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 across sectors is central to understanding the observed changes in the population distribution.

While most previous research on the population distribution has concentrated on cities, rural areas accounted for a large share of the population in developed countries historically, and they continue to account for large shares of the population in many developing countries today. To provide evidence on both rural and urban areas, we construct a new dataset for the United States from 1880 to 2000, which maps data on sub-county divisions – commonly referred to as Minor Civil Divisions (MCDs) – to comparable spatial units over time.

Our main finding is that population growth over this time period is strongly increasing in initial population density for the range of intermediate values of population density at which the majority of the 1880 population lived. At higher values for initial population density, population growth is largely uncorrelated with initial population density, which is consistent with existing empirical findings for cities and metropolitan areas.

The upward-sloping relationship between population growth and initial population density at intermediate densities is closely related to another feature of our data, which is a polarization of the distribution of U.S. population between 1880 and 2000. Despite the substantial increase in the overall U.S. population over this time period, there is an increase in the mass of both sparsely and densely-populated areas. While some regions experienced rapid urbanization, others experienced rural depopulation.

These substantial changes in the distribution of population have important implications for the organization of economic activity and public policy. Urbanization and rural depop-

[1] The U.S. Census Bureau defines an urban area as territory consisting of core census blocks with a population density of at least 1,000 people per square mile and surrounding census blocks with a population density of at least 500 people per square mile (Census 2000d).

[2] The historical figures are from Bairoch (1988) and the present-day figures from United Nations (2008).

ulation have starkly different effects on the incomes of immobile factors and the values of immobile durable goods such as housing. Understanding the reasons for and predicting the pattern of this population redistribution is central to the provision of public and private infrastructure and to the expenditure demands and revenue base facing state and local governments.

Incorporating information on rural areas into our analysis yields new insights for population growth at the high population densities observed in urban areas. Existing research on cities has had to take a stand on two related issues: entry into the city-size distribution and the population threshold for a city. The treatment of both issues affects the city-size distribution and the relationship between city population growth and size, against which existing theories of city growth are compared (see, for example, Black and Henderson 1999, Duranton 2007, Eeckhout 2004, Gabaix 1999, and Rossi-Hansberg and Wright 2007).[3] Both issues reflect the fact that cities, by construction, are a selected sample of locations that have already become densely populated. In contrast, our analysis makes use of the entire distribution of population densities to shed light on *both* the process through which locations become densely populated and the evolution of population among densely populated locations. We use our data to determine empirically the threshold at which population growth becomes largely uncorrelated with initial population density.

Our finding of an upward-sloping relationship between population growth and initial population density suggests that comparatively small initial differences can have large future effects, as growth rates compound over time, and can influence whether or not a location makes the transition from rural to urban. While MCDs with seven people per square kilometer barely grew between 1880 and 2000, those with fifty five people per square kilometer more than tripled in population over the same time period. Even within relatively densely-populated locations, we find variation in rates of population growth that is related to differences in initial patterns of specialization across sectors.

We develop a body of evidence, which shows that the upward-sloping relationship between population growth and initial population density is driven by structural transformation across sectors. We organize our empirical evidence around six stylized facts relating to the distribution of employment and population. We show that non-agricultural employment densities exhibit higher variance than agricultural employment densities, reflecting the fact

[3]For an analysis of the emergence of new cities as a source of growth in the urban population, see Henderson and Wang (2007) and Henderson and Venables (2008).

that employment is more spatially concentrated in non-agriculture than in agriculture. This difference in employment densities is associated with an initial share of non-agriculture in employment that is increasing in initial population density over a range of intermediate values for population density. We show that this range coincides exactly with the range of intermediate densities at which population growth between 1880 and 2000 is increasing in initial population density. Therefore reallocation of employment away from agriculture, combined with a non-agricultural employment share that increases with initial population density, generates the upward-sloping relationship between population growth and initial population density.

We demonstrate that these empirical findings are robust features of the U.S. data across a wide range of specifications. As an additional robustness check, we also replicate our entire analysis for Brazil for a period of substantial structural transformation from 1970-2000. Even though Brazil differs from the U.S. along a number of dimensions, including institutions and physical geography, and even though the data are collected at a different level of spatial aggregation and for a different time period, we find a strikingly similar pattern of results. This similarity of the results in a quite different context reassures us that our findings are not driven by idiosyncratic features of the data or the institutional environment for the U.S.

In Brazil, like the U.S., the range over which there is an upward sloping relationship between population growth and initial population density coincides exactly with the range over which non-agriculture accounts for an increasing share of employment. While this close relationship in both countries is strong evidence in support of our explanation based on structural transformation, we consider other potential explanations for our findings. For example, differences in locational fundamentals, such as physical geography or institutions, could be correlated with population growth, initial population density and the initial share of non-agriculture in employment. Or the upward-sloping relationship between population growth and initial population density could be driven by commuting or suburbanization around the fringes of densely-populated areas.

To rule out such alternative explanations, we include a wide range of additional controls and estimate a number of different specifications. To control for physical geography, we include measures of proximity to the coast, lakes and rivers and proximity to natural resource endowments. To control for physical geography and institutions, we include state fixed effects and even county fixed effects. In our cross-section specification, the state and

county fixed effects control for any determinant of population growth or levels that is common across MCDs within states and counties. Examples include state and county policies, but also climate, which in general varies little across MCDs within counties. To rule out explanations based on commuting and suburbanization, we construct alternative samples based on counties and aggregations of MCDs in the neighborhood of densely-populated areas. We also show that we observe a similar upward-sloping relationship even in a sample of MCDs of above-median distance from existing population centers (more than 170 kilometers away).

To interpret our empirical results, we develop a simple theoretical model, which shows how structural transformation can generate the changes in the population distribution observed in our data. Structural transformation arises from aggregate differences in productivity growth and inelastic demand across sectors. With inelastic demand, more rapid productivity growth in agriculture than in non-agriculture leads to a more than proportionate decline in the relative price of the agricultural good, and hence induces a reallocation of employment from agriculture to non-agriculture.[4]

In the model, non-agriculture is less land intensive than agriculture and exhibits stronger agglomeration forces, which generates an increasing relationship between the share of non-agriculture in employment and population density. In each sector, productivity is influenced by the aggregate trends noted above and by idiosyncratic shocks to productivity in each location. In our empirical results, we find that non-agricultural employment growth is uncorrelated with initial population density, while agricultural employment growth is sharply decreasing in initial population density. The model generates these different employment dynamics if non-agricultural productivity displays constant proportional growth and agricultural productivity exhibits mean reversion, which is consistent with non-agricultural productivity being less tied to persistent characteristics of locations, such as climate and soil. As the location-specific shocks to productivity have bounded support in each sector, these differences in productivity dynamics imply that the agricultural productivity distribution is bounded from above, whereas the non-agricultural productivity distribution is unbounded from above. This difference between the two productivity distributions further reinforces the

[4]In emphasizing differences in productivity growth across sectors and inelastic demand, we follow a long literature on structural transformation in macroeconomics, including Baumol (1967), Ngai and Pissarides (2007) and Rogerson (2008). The main competing explanation in the macroeconomics literature is non-homothetic preferences, as emphasized by Echevarria (1997), Gollin et al. (2002) and Matsuyama (2002), which can also be incorporated into the theoretical model developed below.

increasing relationship between the share of non-agriculture in employment and population density in the model.

As aggregate differences in productivity growth between the two sectors induce a decline in the relative price of agriculture, there is a reallocation of employment from agricultural to non-agricultural locations, and within locations there is an endogenous change in land use from agriculture to non-agriculture. Given the increasing relationship between the share of non-agriculture in employment and population density noted above, this reallocation of employment towards non-agriculture generates the upward-sloping relationship between population growth and density observed at intermediate densities. At low densities, agriculture dominates, and mean reversion in agricultural productivity generates a downward-sloping relationship between population growth and density. At high densities, non-agriculture dominates, and constant proportional growth in non-agricultural productivity generates population growth that is largely uncorrelated with population density.

To illustrate the quantitative relevance of structural transformation, we use relationships suggested by the model to predict rates of population growth between 1880 and 2000 based on economy-wide rates of employment growth in agriculture and non-agriculture combined with each location's initial share of employment in agriculture in 1880. Despite the parsimony of this explanation based on structural transformation, we show that it can account quantitatively for the upward-sloping relationship between population growth and density observed at intermediate densities. In contrast, we find that our controls for geography can account for little of the observed upward-sloping relationship.

Our paper is related to a large body of work in urban economics and economic geography. Recent research on the relationship between population growth and size in the literature on cities includes Duranton (2007), Eeckhout (2004), Gabaix (1999) and Rossi-Hansberg and Wright (2007). While population growth is typically found to be uncorrelated with population size in the cities literature (Gibrat's Law), Black and Henderson (2003), González-Val et al. (2008) and Soo (2007) find evidence of departures from Gibrat's Law even for cities.[5] Dividing the surface of the continental U.S. into a uniform grid of six-by-six mile squares, Holmes and Lee (2008) find that population growth from 1990-2000 is highest at intermediate values of initial population density.

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

Our focus on the reallocation of economic activity from agriculture to non-agriculture also connects with theories of new economic geography, including Fujita et al. (1999) and Krugman (1991). Although reductions in trade costs in these models can result in a polarization of population across space, they do not provide natural explanations for why Gibrat's Law is 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. While an empirical literature has examined the determinants of the distribution of economic activity across states and counties in the U.S., including Beeson et al. (2001), Ellison and Glaeser (1999), Glaeser (2008), Kim (1995) and Rappaport and Sachs (2003), this literature has typically not emphasized structural transformation. Closest in spirit to our work is Caselli and Coleman (2001), who examine structural transformation and the convergence of incomes between Southern and Northern U.S. states. Also related is Desmet and Rossi-Hansberg (2007), who examine differences in patterns of employment growth between the manufacturing and service sectors using U.S. county data, and relate 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 sub-county data are especially well suited.

Finally, our research is related to the development and economic history literatures. Early work on structural change and economic development is surveyed in 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 U.S. includes Kim (2000) and Kim and Margo (2004), although for reasons of data availability this research has again largely concentrated on cities.

The remainder of the paper is organized as follows. Section 2 discusses our main dataset for the U.S., outlines our empirical strategy, presents our main empirical findings, and reports the results of a number of robustness checks. Section 3 presents the results of an additional robustness check using Brazilian data. Section 4 outlines our theoretical model and Section 5 shows that structural transformation can account quantitatively for our empirical findings. Section 6 concludes.

# 2 U.S. Data and Stylized Facts

## 2.1 Data and Samples

This section begins by introducing the U.S. data that we use in this paper and the samples that we construct. We then document a set of stylized facts that shed light on the dynamics of urban and rural population growth from 1880-2000.[6]

In order to analyze these dynamics, we require data on land area, population, and sectoral employment for geographic units that are consistent over time. Since we are interested in both rural and urban areas, we also require that these geographic units partition the land area that we analyze. In other words, we want a dataset that covers the entire population and all the land - from the largest cities to the smallest farms. And since we are interested in examining rural and urban population dynamics, we prefer that our geographic units be fine enough to separate urban areas from rural ones.

While these criteria may seem natural, it is not easy to find an existing dataset that satisfies them all. The literature on urban growth in the U.S. has often analyzed counties or Metropolitan Statistical Areas (MSAs), which are groups of counties. And although counties satisfy most of our requirements, they often pool together urban centers with their surrounding countryside. So while we include counties in our analysis, we are also interested in data that provide finer spatial aggregation. One dataset that is less aggregated than the county dataset includes incorporated places - this is the dataset used by Eeckhout (2004). But while this data is useful for studying urban growth dynamics, it does not contain information on many rural areas, where the majority of the U.S. population lived before the 20th century.

Since existing datasets are not fully satisfactory for our purposes, we construct a new dataset using minor civil divisions (MCDs). MCDs have been used to report population in parts of the U.S., especially in the Northeast, since the first census in 1790 (see Census 2000c). But as we discuss below, we are interested not only in population but also in sectoral employment. And since the earliest available digitized employment data for MCDs comes from the 1880 Census, we chose 1880 as the starting year for our analysis. Over time, MCDs became a standard tool for partitioning counties throughout (almost) the entire U.S.[7] It is

[6] For further discussion of the U.S. data and the samples discussed below, see the web-based technical appendix.

[7] In many western states sub-county units were initially called MCDs but were reclassified as census county divisions (CCDs) in 1950, when the map of sub-county units in many of these states was redrawn. For simplicity, we refer to both MCDs and CCDs as MCDs (see chapter 8, Census 2000c) and discuss in

this feature of MCDs that makes them so suitable for our analysis: they provide the finest level of geographical disaggregation for which we can analyze urbanization and structural transformation over more than a century.[8]

The most common types of MCDs are towns and townships, but in some areas election precincts, magisterial districts, parishes, election districts, plantations, reservations, boroughs or other categories were used as MCDs. As some of these names suggest, in many states MCD boundaries coincide with those of local government bodies. In New England in particular, MCDs correspond to townships that are actively functioning units of local government, in many cases since the $17^{th}$ Century. But in other states MCDs are often statistical entities with few (or no) other functions (see Chapter 8 Census 2000c). Given the variation in their functions, it is not surprising that the size and shape of MCDs also vary from state to state. For example, in the Midwest MCDs often follow a chessboard patterns with squares of 6 miles per side; this design dates back to the Land Ordinance of 1785 and the Northwest Ordinance of 1787 (see Prescott 2003). As one travels West or South, the size of MCDs tends to grow, and they tend to become less regular and less stable over time. To address concerns that differences in the geographical and institutional organization of MCDs could affect population growth and employment structure, we report robustness checks where we consider states with similar geographical and institutional organizations of MCDs, and where we consider more aggregated spatial units such as counties.

To overcome changes in MCD boundaries over time, we aggregate some MCDs to create geographic units that are stable over time. This aggregation process involved considerable work using historical maps and gazetteers, and it is described in further detail in the web-based technical appendix. To provide a brief idea of the aggregation process, we matched the approximate centroid of each 1880 and 1940 MCD to the 2000 MCD in which it fell. We then aggregated any 2000 MCD that did not contain at least one 1880 MCD and one 1940 MCD to the nearest 2000 MCD that did. This aggregation process enables us to track the evolution of population at a fine level of spatial detail over 60-year intervals.[9] One reason

further detail in the web-based technical appendix how we link MCDs over time.

[8] 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. Additionally, we use county data for some states where sub-county units are not comparable over time. We discuss in further detail below the robustness of our results across a wide range of samples and specifications.

[9] All MCDs in our baseline sample, which consists of the "A and B" states defined below, have non-zero population in all three years of our sample. But there are 7 MCDs in the "C" states, as defined below, that have zero population in 1880. These 7 MCDs are dropped when we construct population growth rates.

for restricting ourselves to these years is that adding more years would have forced us to aggregate further. But perhaps more importantly, we only know the employment structure of MCDs for 1880 (using the individual-level census records from the North Atlantic Population Project) and for the very recent censuses, such as 2000 (using data from the U.S. census American Factfinder tool, see Census 2000b). Since our analysis uses both population and employment data, adding more years for which we don't have employment data would have not contributed much. Finally, we used the 2000 census to calculate the land area in each geographic unit.

The extent of aggregation required to construct time-consistent units varies across states. In some states, especially in the Northeast and the Midwest, MCDs corresponded to local administrative units that were very stable over time, so little aggregation was required. We therefore divided states into samples: little aggregation was required in A states, more was needed in B states, and more still in C states. The geographic distribution of states across these three groups is shown in Map 1. In choosing our baseline sample, we sought to include as many states as possible while limiting the extent of aggregation, since the aggregation process might entail some imprecision. We therefore choose as our baseline sample the A and B states, for which $1-1$ matches between the 1880 and 2000 censuses involving no aggregation exceeded 70 percent.[10] But as we discuss below, we also construct alternative samples that either include more states (in some cases using county-level data) or restrict our sample to A states, where very little aggregation was required. In our baseline sample there are, on average, 13 units ("MCDs") per county. The average unit spans $115 km^2$, with a population of 2,400 in 1880 and 8,800 in 2000.

## 2.2 Empirical Strategy

We are interested in characterizing the population density distribution and the relationship between population growth and the initial population density distribution. In both cases, 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 density, $x$, into discrete bins of equal size $\delta$. 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 $\Phi_b$ and denoting the number of MCDs within this set by $n_b$, we estimate the population density

[10] Since in most cases our geographic units consist of a single MCD, we refer for simplicity to these units as "MCDs", even though they are sometimes aggregations of MCDs.

distribution, $\hat{g}(x_m)$, as follows:

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

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 space of values for log population density, while in general preserving a relatively large number of MCDs within bins. Although this approach provides a simple and flexible characterization of the population density distribution, which connects closely with the other components of our analysis below, we also find similar results using related non-parametric approaches such as kernel density estimation (Silverman 1986).

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:

$$y_{mt} = f(x_{mt-T}) = \sum_{b=1}^{B} I_b \phi_b, \qquad \phi_b = \frac{1}{n_b} \sum_{m \in \Phi_b} y_{mt}, \qquad \text{for } x_m \in \Phi_b, \tag{2}$$

where $t$ indexes time. In this specification, bins are defined over initial population density, $x_{mt-T}$; $y_{mt}$ is average population growth from $t-T$ to $t$; and $I_b$ is an indicator variable equal to one if $x_{mt-T} \in \Phi_b$ and zero 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 percent 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.[11] While this non-parametric specification allows for a flexible relationship between population growth and initial population density, we again find similar results using other related non-parametric approaches, such as locally weighted linear least squares regression

[11] When displaying the results of the specification (2) graphically, we remove the top and bottom one percent of the observations from the graphical representation, but not from the regressions. The bins at these extremes of the distribution contain few observations and have correspondingly large standard errors. Hence they tend to cloud rather than to illuminate the true picture.

(Cleveland 1979) and kernel regression (Härdle 1990). A key advantage of the specifications in (1) and (2) is that we can preserve the same discrete bins when analyzing the population density distribution, the relationship between population growth and the initial population density distribution, and analogous specifications for employment in the agricultural and non-agricultural sectors.

As our model yields predictions for the functional form of the relationship between population growth and initial population density, we also estimate parametric versions of specification (2) of the form:

$$y_{mt} = \rho x_{mt-T} + u_{mt}, \tag{3}$$

where $\rho$ is a parameter to be estimated, $u_{mt}$ is a stochastic error, and we again report standard errors clustered by county.

Finally, to examine the relationship between employment structure and the population distribution, we estimate specifications analogous to (1)-(3) for employment in the agricultural and non-agricultural sectors.

## 2.3 Stylized Facts

To better understand the process of urbanization and structural transformation in the U.S. from 1880-2000, we organize our empirical findings around 6 stylized facts. These facts highlight the *instability* of the spatial distribution of economic activity over this time period when urban areas are analyzed together with rural areas – a pattern of results that lies in stark contrast to the stability documented within the sample of cities in the literature on urban growth. These facts also suggest that this instability is closely related to structural transformation away from agriculture.

We begin by reporting a number of descriptive statistics for our baseline sample of "A and B" states in Column (1) of Table 1. Figures 1-6 then display the results of the non-parametric specifications (1) and (2) for population and for employment in the agricultural and non-agricultural sectors. Our first stylized fact is that the distribution of log population density across MCDs has become more dispersed from 1880-2000. As shown in Panel A of Column (1) in Table 1, the standard deviation of the distribution of log population density increased over this period from 0.97 to 1.56. This increase in the standard deviation is statistically significant and is larger than the increase in mean log population density. Figure 1 confirms this increase in dispersion by displaying the results from specification (1).

Although the U.S. population increased substantially from 1880-2000, as reflected in Figure 1 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 as other higher-density areas experience rapid population growth. This instability of the overall distribution of population stands in sharp contrast to the stability of the distribution of city sizes (e.g. Duranton 2007). Existing research for cities finds that the population size distribution is approximated by a (stable) Pareto distribution in the upper tail (e.g. Gabaix 1999) or a lognormal distribution for a wider range of city sizes (Eeckhout 2004).

Second, Gibrat's law, which states that population growth and population size are uncorrelated, is clearly violated when cities are considered together with rural areas. While the relationship between population growth and population size is typically estimated for shorter time horizons than available in our data, if Gibrat's Law holds for shorter time horizons, population growth should remain uncorrelated with population size over these longer time horizons. In Figure 2, we display the results from our population growth specification (2), where the dark solid line denotes mean population growth within each initial population density bin and the lighter dashed lines denote the 95 percent confidence intervals. As shown in the figure, log population density in 1880 is strongly predictive of population growth from 1880-2000. A similar relationship is found if we replace initial population density with initial population, as discussed further below.[12]

As Figure 2 shows, for low population densities, there is a negative correlation between population density in 1880 and subsequent population growth. But approximately between log population densities 2 and 4,[13] a range where more than half of the population in our sample resided in 1880, population density in 1880 is positively correlated with subsequent population growth.[14] The magnitudes of these departures from Gibrat's Law are substantial:

[12] While the existing literature on cities concentrates on the relationship between population growth and population size, we focus on the relationship between population growth and population density to control for differences in land area across sub-county units. Although our results are qualitatively the same if we instead use population size, the population density specification is more appropriate if land area varies across sub-county units and is derived directly from our theoretical model. In our data, there is a strong and approximately log linear relationship between population density and population size, which is consistent with the theoretical model developed below.

[13] Population densities in logs (levels) compare approximately as follows: 2 (7), 4 (55) and 6 (403), where these figures are expressed as the log number (number) of people per square kilometer.

[14] While classical measurement error in 1880 population could induce a negative correlation between population growth and 1880 population density, this does not 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

MCDs with log density of about 0 or 4 in 1880 experienced population growth at a rate of about 1 percent from 1880-2000. By contrast, MCDs with a log population density around 2 barely grew on average. As shown in Panel B of Column (1) in Table 1, these differences are statistically significant. We also note that at levels of log population density above 4 population density seems to be largely uncorrelated with population growth; this is the range that typically includes urbanized areas. Hence this finding is broadly consistent with the literature that finds Gibrat's law is a reasonable approximation for city population growth.

Third, the share of agriculture in employment drops steeply in the range where population density in 1880 and subsequent growth are positively correlated. Figure 3 presents the results from specification (2) using the share of agriculture in employment in 1880 as the left-hand side variable rather than population growth. As the figure shows, the agricultural employment share in 1880 drops 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 1 shows that this difference is statistically significant. For denser MCDs the share continues to decline, but at a much slower rate.[15]

Fourth, the distribution of employment per square kilometer across MCDs has a lower standard deviation in agriculture than in non-agriculture in both 1880 and 2000. As shown in Panel D of Column (1) in Table 1, this difference is statistically significant at conventional critical values. Figure 4 presents the results from specification (1) for employment in agriculture and non-agriculture in 1880 and 2000. As shown in the figure, the employment density distribution in agriculture has thinner tails than its non-agricultural counterpart.[16] Therefore, there are more observations with extreme low and high values of employment density for non-agriculture than for agriculture, reflecting the greater spatial concentration of non-agricultural employment. Furthermore, a comparison of Figures 1 and 4 suggests that the 1880 population was distributed in a similar way to the 1880 agricultural employment, while the 2000 population was more spatially concentrated and distributed in a similar way to the 2000 non-agricultural employment. This reflects the substantial decline in agricul-

Census data mitigates measurement error concerns.

[15] The share of employment in total population was about 0.33 in 1880 and 0.48 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 dynamics are a reasonable predictor of population dynamics.

[16] We also find that non-agricultural employment per square kilometer is more unequally distributed than agricultural employment 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.

ture's share of employment, which fell from 35 percent to 1 percent of total employment in our baseline sample of "A and B" states.

Fifth, agricultural employment growth appears to follow a mean-reverting process. To document this stylized fact, we consider the subsample of MCDs for which agriculture accounted for more than 80 percent of 1880 employment. Although the share of agricultural employment in this subsample was over 88 percent in 1880, it fell to below 10 percent in 2000, and hence this subsample does not entirely capture agricultural dynamics alone. Nevertheless, since this subsample was at least initially mostly agricultural, it is likely to capture the main features of agricultural growth.[17] Figure 5 displays the results from non-parametric specification (2) for this subsample using agricultural employment growth as the left-hand side variable. As apparent from the figure, densely-populated MCDs in this subsample exhibited much slower growth of agricultural employment from 1880-2000 than sparsely populated MCDs. Panel E of Column (1) in Table 1 reports the results from parametric specification (3) for this subsample, again using agricultural employment growth as the left-hand side variable. This confirms our finding of mean reversion: the coefficient on log population density in 1880 in the parametric specification is $-0.006$ and significant (p-value $< 0.001$). From the size of this coefficient, each additional log point of population density in 1880 is associated on average with just over half a percentage point lower rate of agricultural employment growth. We find very similar results if we instead relate agricultural employment growth to log agricultural employment density in 1880: the coefficient on initial log agricultural employment density is $-0.006$ and statistically significant.

Sixth, in contrast to the results for the agricultural sector, non-agricultural employment growth is uncorrelated with 1880 population density. To demonstrate this, we consider the subsample of MCDs for which agriculture accounted for less than 20 percent of 1880 employment. In this subsample the share of non-agricultural employment was higher than 90 percent in 1880 and higher than 98 percent in 2000. Figure 6 displays the results from non-parametric specification (2) using non-agricultural employment growth as the left-hand side variable, while Panel F of Column (1) in Table 1 reports the results from the analogous parametric specification (3). As apparent from the figure, non-agricultural employment grew at about 1.2 percent per year. This positive growth rate is very different from the

[17] We also find mean reverting processes when we consider population growth (rather than employment growth) for both 1880-2000 and 1880-1940 for the same agricultural subsample. During the 1880-1940 period, agriculture remained an important employer in much of the U.S. at both the beginning and end of the period.

(mostly) negative growth rates of agricultural employment shown in Figure 5. Moreover, in sharp contrast to the results for the agricultural sector, non-agricultural employment growth is uncorrelated with 1880 population density. As reported in Panel F of Column (1) in Table 1, the coefficient on log population density in 1880 in the parametric specification is $-0.0002$ and statistically insignificant (p-value = 0.515). We also find very similar results if we instead relate non-agricultural employment growth to log non-agricultural employment density in 1880. The coefficient on log non-agricultural employment density is $-0.00021$, which is more than an order of magnitude smaller than the corresponding coefficient for the agricultural sector, and statistically insignificant.

## 2.4 Robustness of the Stylized Facts

Having documented the 6 stylized facts for our preferred sample of MCDs, we now examine their robustness to different samples and specifications. The results of these robustness checks are summarized in Columns (2) to (8) of Table 1, while Figure 7 replicates the non-parametric population growth specification (2) displayed in Figure 2 for each of the robustness checks.

One potential concern about our preferred sample is that imperfect matching of MCDs across censuses could have affected our estimates. For example, some of the population and employment of MCDs with intermediate densities could have been assigned to MCDs with either higher or lower densities, which would affect relative population growth at different densities. To address this concern, the second column of Table 1 shows that all of our stylized facts remain intact when we restrict the sample to MCDs in the "A states" (to which we also refer as the restricted sample). In this restricted sample match rates are well over 90 percent, so imperfect matching is unlikely to be the cause of our finding. Figure 7 also shows non-parametrically that the U-shape we document in the second stylized fact is still strongly apparent in this sample.

Another possible concern is that we use a level of aggregation that is too fine. For example, people could live in one MCD and commute to work in another MCD, which could in turn influence the correlation between population growth and population density. As a first step to address this concern, we replicate our analysis using county-level data, since fewer people commute across county boundaries than across MCD boundaries. In the third column of Table 1, we report results using county-level data for 45 states and Washington DC.[18] As

[18] As noted in footnote 8, 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

this robustness check includes a more comprehensive set of states than our baseline "A and B" sample, it ensures that our findings are not being driven by the particular geographic distribution of states in the baseline "A and B" sample. To provide a comparison, the fourth column restricts the county sample to the baseline "A and B" states. And the fifth column reports results using a hybrid sample of MCDs for states where matching was possible and counties for other states.

Our results are robust across all three specifications with two exceptions. The first stylized fact does not hold in Column (3), where the standard deviation of log population in 1880 is higher than in 2000, though the difference is not statistically significant at conventional critical values. The sixth stylized fact does not hold in Columns (3) and (5), where we find some evidence of mean reversion in both agriculture and non-agriculture, though the estimated rate of mean reversion in non-agriculture is substantially lower than that in agriculture. These exceptions 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), are likely to be quite 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, accounting for the exception to stylized fact 3.[19] Relatedly, the future settlement of areas that were largely uninhabited in 1880 provides a natural explanation for mean reversion that is unrelated to employment structure, consistent with the exception to stylized fact 6. Despite these caveats, the remainder of the stylized facts hold in these specifications, and in areas that were well-settled by 1880 all of our results are robust to aggregating MCDs up to the county level.[20]

stable county boundaries at that time.

[19] 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. Indeed, 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.

[20] To further test whether our results are affected by the U.S.'s Westward expansion, we restricted our baseline "A and B" sample to states that were part of the original 13 colonies. All the stylized facts are robust to this restriction, except part of stylized fact 3 (the downward slope of the u-shape). We do not find that population growth for log density 0 is significantly larger than for log density 2. But this finding is not surprising, since only two MCDs fall in the category of log population density 0 in this restricted sample. When we further restrict our sample to A states within the 13 colonies (New York and New England, except Maine), the remaining stylized facts all hold, except that we find no significant mean reversion in the agricultural subsample (stylized fact 5). But this is probably again due to small sample size. There are only 78 observations (in 48 counties) in the agricultural subsample for A states that were part of the original colonies (out of 4439 observations for this sample), reflecting the relatively urban character of these states.

While results using county-level data are consistent with our previous results, a further concern is that the aggregation they provide is insufficient around large cities. Metropolitan Statistical Areas (MSAs) span multiple counties and may be characterized by commuting across county boundaries. Additionally, the suburbanization that took place during the second half of the twentieth century could have influenced population dynamics in the neighborhood of large cities even beyond county boundaries. To address these concerns, we undertake further aggregation. One possibility is to aggregate counties based on 20th-century definitions of MSAs, but MSA definitions are themselves endogenous to population growth during our sample period. Therefore we instead aggregate MCDs based on 1880 characteristics using a flexible approach that allows us to consider various levels of aggregation. Starting with our baseline sample, we identify as "cities" MCDs that had a log population per square kilometer 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 kilometers of each city.[21] We label the resulting sample a suburban sample, since it pools together large urban centers with their surrounding areas. As shown in Column (6) of Table 1 and Panel D of Figure 7, all of our stylized facts hold in this suburban sample. We also experimented with other ways of aggregating the areas surrounding cities, including defining "cities" as MCDs with 50,000 or 100,000 or more inhabitants in 1880 and using a distance threshold of 50 kilometers, and again found a similar pattern of results.

As a further robustness check, we examined whether the upward-sloping relationship between population growth and initial population density observed in Figure 2 for log densities in between 2 and 4 is robust to restricting the sample to MCDs with an above median distance to one of our "cities." Re-estimating our non-parametric specification (2) for this subsample, in which the distance to a "city" is greater than 170 kilometers, we continue to find a strong upward-sloping and highly statistically significant relationship between population growth and initial population density for log densities in between 2 and 4. Taking these results together, commuting and suburbanization in the neighborhood of large cities do not appear to be driving the upward-sloping relationship between population growth and initial population density observed in our data.[22]

Although we examine the relationship between population growth and initial population

[21] When two or more cities and their surrounding areas overlapped, we merged them together.

[22] While suburbanization is primarily associated with the use of the automobile as a form of mass transit, we also note that we find a very similar pattern of empirical results for the period 1880-1940, prior to the large-scale dissemination of the automobile after the end of the Second World War.

density to control for variation in land area across MCDs, existing research concentrates on the relationship between population growth and initial population size. Therefore, while initial population density and size are strongly and approximately log linearly related in our data, another concern is that the violation of Gibrat's Law is driven by the use of initial population density rather than initial population size. To address this concern, Column (7) of Table 1 and Panel E of Figure 7 display results using log initial population size. Given that log population is measured in different units from log population density, we do not expect the inflection point at which the population growth relationship switches from being downward-sloping to upward-sloping relationship to occur at the same numerical values, and therefore the statistical tests based on values of 0, 2 and 4 in Table 1 do not apply to this specification and are not reported. Nonetheless, we observe the same qualitative pattern, and each of our stylized facts holds if we use initial log population size instead of initial log population density.

Another alternative hypothesis is that the observed relationship between population growth and initial population density could be influenced by omitted institutions or natural endowments. While institutions and natural endowments are captured in the model developed below in so far as they influence location-specific productivities in the agricultural and non-agricultural sector, the empirical concern is that these variables have a direct effect on population growth and are correlated with initial population density. To explain our results, these omitted variables would need to have a non-linear relationship with population growth and initial population density, to have the same non-linear relationship with the share of agricultural employment and initial population density, and to have differential effects on the correlation between employment growth and initial employment in the agricultural and non-agricultural sectors.[23]

To provide evidence that such a direct effect of institutions or natural endowments is not driving our results, we first regress each of our left-hand side variables (population growth, the share of agriculture in employment, and employment growth in agriculture and non-agriculture) on state fixed effects (to control for state policies and institutions) and on measures of proximity to natural endowments (rivers, lakes and coastlines, and mineral endowments). We next take the residuals from these regressions and implement our tests

[23] As a first robustness check to address the concern about institutional differences, we also re-estimated our baseline specification for the subset of the A states that were part of the original 13 colonies. Within this subset of the A states, MCDs are towns and townships with similar administrative functions. Once again, we find a similar pattern of results, as discussed in footnote 20 above.

for Gibrat's Law (stylized fact 2), the share of agriculture in employment (stylized fact 3) and the relationship between employment growth and initial employment in agriculture and non-agriculture (stylized facts 5 and 6). As shown in Column (8) of Table 1 and Panel F of Figure 7, we continue to find a similar pattern of results after controlling for institutions and natural endowments.[24]

Finally, the population of urban locations can grow through a number of channels, including migration from rural areas, international migration or differences in fertility. While the model developed below abstracts from international migration and fertility, it could be extended to include them, and the assumption of population mobility implies that people are indifferent across locations. As a result of this indifference condition, the populations of all locations are linked together in the model. Although the U.S. has relatively high levels of population mobility, the presence of barriers to mobility could in principle break this link between locations' populations, with the result that local variation in international migration, fertility and mortality could directly affect local population. As a final robustness check, we therefore include a number of controls for initial demography, including international migration, fertility, education and race, using the same methodology as for Column (8) above. While these controls are likely to be themselves endogenous to employment structure, and are therefore not included in our baseline specification, we continue to find a similar pattern of results when they are included.[25]

Taken together, the evidence presented in this section shows that our stylized facts are robust characteristics of the U.S. growth experience in the 20th Century. But are they also relevant for more recent experiences of structural transformation in other countries? To shed more light on this issue, we next examine urbanization and structural transformation in Brazil.

[24] As the relationship between population and locational fundamentals can change over time, and as the relationship between employment and location fundamentals can differ between the agricultural and non-agricultural sectors, we do report standard deviations for log population and employment after controlling for locational fundamentals (stylized facts 1 and 4).

[25] All of the stylized facts are robust to the joint inclusion of the following four demographic control variables: the initial share of the population that is white, the share of the population aged 14-18 in education (as a measure of human capital), the share of the population that was born outside the U.S. (as a measure of international migration), and the share of the population aged less than six (as a measure of fertility).

# 3 Brazilian Data and Stylized Facts

## 3.1 Data and Samples

The most populous country in the Western Hemisphere after the U.S. is Brazil. Like the U.S., Brazil is divided into states, and just as U.S. states are divided into counties, Brazilian states are divided into municipalities. Since municipality boundaries have changed over time, the Instituto de Pesquisa Econômica Aplicada (IPEA) has created "áreas mínimas comparáveis" (AMCs), geographic units that are more stable over time. The 5,507 municipalities that existed in 1997 were pooled into 3,659 AMCs, which allow us to consistently analyze data from 1970-2000.[26] Although we could analyze Brazilian data before 1970, this would entail considerable further aggregation of municipalities, which would make it harder to distinguish urban from rural areas. Therefore we choose 1970 as the starting point for our analysis. It is worth noting that agriculture's share in employment in the average AMC declined from 71 percent to 43 percent from 1970-2000, and its share in overall employment fell from 46 percent to 20 percent. Therefore the period we analyze involved considerable structural transformation.

The average Brazilian AMC spans 2,323$km^2$, with a population of 25,817 in 1970 and 46,421 in 2000. While AMCs are on average larger than the units that we analyze in our U.S. sample, the difference is due in part to the fact that the interior regions of Brazil have larger and more sparsely populated AMCs. Therefore, while our baseline sample uses all of Brazil, we also demonstrate the robustness of our results to using a restricted sample that includes the Northeast, Southeast and South regions in Brazil only. In these areas, the average AMC spans 923$km^2$, and had a population of 26,013 in 1970 and 44,125 in 2000.

## 3.2 Stylized Facts

Having described Brazilian AMCs, we now examine whether their population dynamics are characterized (at least qualitatively) by the same stylized facts as for U.S. MCDs. Panel A in Figure 8 and Table 2 shows that the standard deviation of log population density across Brazilian AMCs increased from 1970-2000, confirming our first stylized fact. Additionally, Panel B in the same Figure and Table shows that low density areas and high density areas

[26] New municipalities were created after 2000, but the 1997 municipalities were used in the 2000 Census, the latest Census that we analyze in this paper. For further discussion of the Brazilian data and the samples discussed below, see the web-based technical appendix.

grew faster than areas of intermediate density. Therefore the U-shaped relationship between population growth and initial population density, characterized in stylized fact 2, also holds for Brazil. One quantitative difference between Brazil and the U.S. is, however, that the increasing segment of this U-shape is not 2-4 (as in the U.S.), but rather 4-6. This difference partly reflects differences in the relative distribution of agricultural and non-agricultural employment in Brazil and the U.S., as evident in Figures 4 and 8 (Panel D).

Furthermore, Panel C in Figure 8 and Table 2 shows that the increasing segment of the U-shaped population growth relationship is located in the same range of initial population densities where a sharp decline in agriculture's share of employment is observed, as in the U.S. (stylized fact 3). This provides further corroborating evidence that the U-shape is indeed related to employment structure. Panel D in Figure 8 and Table 2 also confirms that agricultural employment has a lower standard deviation than non-agricultural employment (stylized fact 4). Finally, the last two stylized facts - that agricultural employment is mean reverting and non-agricultural employment is uncorrelated with initial density, are also confirmed for Brazil, as shown most clearly in the final two panels of Table 2 and also in Figure 8.[27]

In summary, we find a striking similarity in the relationship between population growth and employment structure in Brazil and the U.S. This similarity of the results in two quite different contexts and time periods suggests that our results are unlikely to be driven by idiosyncratic features of the data or institutional environment for an individual country, but rather capture more systematic features of the relationship between urbanization and structural transformation.

# 4 The Model

To interpret our empirical results, this section develops a simple theoretical model that shows how structural transformation can account for the six stylized facts.[28] To isolate the role played by structural transformation, the model abstracts from a number of other potential determinants of population growth, such as physical geography and institutions. While the

[27] For Brazil, to ensure a sufficient sample size, we construct the non-agricultural subsample using AMCs that have an agricultural employment share in 1970 of less than less than 0.4 (instead of less than 0.2 for the U.S.). Nonetheless, if we also use a threshold of less than 0.2 for Brazil, we continue to find no statistically significant relationship between non-agricultural employment growth and initial population density.

[28] A more detailed exposition of the model is contained in a web-based technical appendix.

introduction of these additional elements would complicate the model, it would not negate the basic mechanism of structural transformation. In our empirical work, we carefully control for these other potential determinants of population growth, as discussed above.

The distribution of employment across locations in the model is determined by productivity in the agricultural and non-agricultural sectors. While agglomeration economies provide a force for the concentration of employment and hence population, an inelastic supply of land for commercial and residential use provides a force for the dispersion of employment and population.

As non-agriculture has stronger agglomeration economies and is less land-intensive than agriculture, the share of non-agriculture in employment is increasing in population density. Structural transformation occurs as a result of more rapid productivity growth in agriculture than in non-agriculture, which with inelastic demand reallocates employment towards non-agriculture. Structural transformation away from agriculture, combined with an increasing relationship between the share of non-agriculture in employment and population density, generates the upward-sloping relationship between population growth and density observed at intermediate densities.

## 4.1 Endowments, Preferences and Technology

Time is discrete and is indexed by $t$. The economy consists of a fixed number of locations $i \in \{1, ..., I\}$, which are grouped in our data into larger statistical units called MCDs. Each location is endowed with a quantity of land $H_i$, which can be used residentially or commercially. Land allocated to commercial use in each location can be employed in either agricultural or non-agricultural production, but cannot be simultaneously employed in both. Therefore each location specializes completely in either the agricultural or the non-agricultural good.[29] Furthermore, as the model abstracts from labor force participation, employment in a location's sector of specialization equals its population.[30] The economy as a whole is endowed with $S_t$ workers, who are mobile across locations, and are each endowed with one unit of labor that is supplied inelastically with zero disutility.

Each worker has the same Cobb-Douglas preferences and allocates a constant share of

[29] The assumption that locations are completely specialized in agriculture or non-agriculture simplifies the characterization of the model's dynamics. MCDs are in general incompletely specialized, as they can contain both agricultural and non-agricultural locations.

[30] The model's abstraction from labor force participation is motivated by the empirical finding noted above that labor force participation is not strongly related to population density in our data.

expenditure ($\alpha$) to a consumption index of tradeable goods and the remaining share ($1-\alpha$) to the consumption of residential land.[31] The tradeable goods consumption index ($C_{it}$) is defined over consumption of agriculture ($c_{Ait}$) and non-agriculture ($c_{Nit}$) and is assumed to take the constant elasticity of substitution (CES) form:

$$C_{it} = \left[\psi_{At} c_{Ait}^{\rho} + \psi_{Nt} c_{Nit}^{\rho}\right]^{1/\rho}, \qquad 0 < \kappa = \frac{1}{1-\rho} < 1, \quad \psi_{At}, \psi_{Nt} > 0, \tag{4}$$

where $\psi_{At}$ and $\psi_{Nt}$ are preference parameters that capture the relative strength of consumer preferences for the agricultural and non-agricultural goods. Consistent with empirical evidence and a large literature in macroeconomics, we assume that agricultural and non-agricultural consumption are complements, so that the elasticity of substitution between the two goods ($\kappa$) is strictly less than one.[32]

The non-agricultural and agricultural goods are produced under conditions of perfect competition and are costlessly tradeable across locations. Output in each sector ($Y_{jit}$) depends on labor input ($L_{jit}$), land input ($H_{jit}$), a productivity parameter ($\theta_{jit}$) and a local externality in the size of the sector ($S_{jt}^{\eta_j}$):

$$Y_{jit} = S_{it}^{\eta_j} \theta_{jit} L_{jit}^{\mu_j} H_{jit}^{1-\mu_j}, \qquad 0 < \mu_j < 1, \ 0 \leq \eta_j < 1. \tag{5}$$

where $j \in \{A, N\}$ indexes agriculture ($A$) and non-agriculture ($N$). While we allow for positive externalities in non-agriculture ($0 < \eta_N < 1$), we assume for simplicity that there are no externalities in agriculture ($\eta_A = 0$), although all we require is that externalities in agriculture are less strong than those in non-agriculture, which is consistent with the much greater spatial concentration of employment in non-agriculture discussed above.

Productivity in each sector is assumed to have an aggregate component ($\Gamma_{jt}$), which is common across locations but changes over time, and an idiosyncratic component ($\sigma_{jit}$), which varies across locations and over time:

$$\theta_{jit} = \Gamma_{jt} \left(1 + \sigma_{jit}\right) \theta_{jit-1}^{\nu_j}, \qquad 0 < \nu_j \leq 1, \tag{6}$$

where $\nu_j$ captures the degree of mean reversion in productivity over time, and the idiosyncratic component of productivity is assumed to be independently and identically distributed with mean zero, and bounded support satisfying $1 + \sigma_{jit} > 0$.

[31] 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 (2008).

[32] The assumption of an elasticity of substitution between agriculture and non-agriculture of less than one is consistent with empirical findings of larger changes over time in nominal consumption shares than in real consumption shares (see for example Kravis et al. 1983).

## 4.2 Equilibrium Land Use and Population

After observing the vector of agricultural and non-agricultural productivity shocks in each period, each worker chooses location, consumption of the agricultural good, consumption of the non-agricultural good, and residential land use to maximize their utility, taking the population distribution as given. Since relocation is assumed to be costless, the worker's optimization problem reduces to choosing these variables to maximize their instantaneous flow of utility. The distribution of population across locations is therefore determined by the requirement that real wages are equalized across all the locations that are populated in equilibrium.

With perfectly competitive goods and factor markets, labor and land are paid their value marginal product. Equilibrium commercial land use in each location is determined by whichever sector offers the higher value marginal product for land. In general, the equilibrium rental rate for land varies across locations and is determined by the requirement that residential and commercial land use sum to the location's endowment of land. With a Cobb-Douglas production technology and upper tier of utility, firms expend a constant share of their revenue on payments to labor and commercial land use, and workers allocate a constant share of their income to goods consumption and residential land use. As a result, the equilibrium fraction of land allocated to residential and commercial use in each location depends solely on parameters of demand and technology. Since the factor intensity of production differs between agriculture and non-agriculture, the equilibrium fraction of land allocated to residential and commercial use varies across locations depending on which good is produced.

Combining real wage equalization and equilibrium land use, the equilibrium population density in each location can be determined as a function of its productivity in its sector of specialization and its land endowment:

$$\frac{S_{jit}}{H_i} = \Lambda_{jt}^{\xi_j} \theta_{jit}^{\xi_j} H_i^{\eta_j \xi_j}, \qquad \xi_j \equiv \frac{1}{\left(1-\mu_j\right) + \frac{1-\alpha}{\alpha} - \eta_j} > 0, \tag{7}$$

where $\Lambda_{jt}$ is constant across locations specialized in the same good $j$ at a given point in time $t$ and is defined in the web-based technical appendix.

Combining equilibrium population density (7) and productivity dynamics (6), we obtain the following relationship between population growth and initial population density for locations that remain specialized in the same sector over time:

$$\ln\left(\frac{S_{jit}}{S_{jit-1}}\right) = \vartheta_{jt} + \xi_j \ln\left(1+\sigma_{jit}\right) - \left(1-\nu_j\right)\ln\left(\frac{S_{jit-1}}{H_i}\right), \tag{8}$$

where $\vartheta_{jt}$ is constant across locations that specialize in the same good $j$ in both $t$ and $t-1$ and is defined in the web-based technical appendix.

Therefore, for locations that remain specialized in the same sector over time, the correlation between population growth and initial population density depends on the extent of mean reversion in productivity shocks over time. The model allows for differences in productivity dynamics between the two sectors and hence the extent of mean reversion in productivity shocks in each sector becomes an empirical question. As we find empirically that non-agricultural employment growth is largely uncorrelated with initial population density, we assume that $\nu_N = 1$, which implies constant proportional growth in non-agricultural productivity and the population of non-agricultural locations (Gibrat's Law). Similarly, as we find empirically that agricultural employment growth is negatively correlated with population density, we assume that $0 < \nu_A < 1$, which implies mean reversion in agricultural productivity and the population of agricultural locations.

These differences in productivity dynamics between the two sectors provide a further reason for why the share of non-agriculture in employment is ultimately increasing in population density. While mean reversion in agricultural productivity implies a productivity distribution that is bounded from above, constant proportional growth in non-agricultural productivity implies a productivity distribution that is unbounded from above. Therefore the very highest values of productivity that support the densest population concentrations are only observed in the non-agricultural sector. Mean reversion in agricultural productivity also explains the downward-sloping relationship between population growth and initial population density observed at low densities, since these locations are almost entirely specialized in the agricultural sector. Similarly, constant proportional growth in non-agricultural productivity also explains why population growth is largely uncorrelated with population density at high densities, since these locations are almost entirely specialized in the non-agricultural sector.[33]

With inelastic demand between the two tradeable consumption goods in (4), more rapid technological progress in the agricultural sector than in the non-agricultural sector leads to a more than proportionate fall in the relative price of the agricultural good and a reallocation of

[33]For empirical evidence of stronger mean reversion in agricultural productivity than in non-agricultural productivity, see for example Martin and Mitra (2001).

employment from agriculture to non-agriculture over time.[34] As this change in employment structure proceeds, population is reallocated away from locations with relatively high productivity in agriculture towards locations with relatively high productivity in non-agriculture. Additionally, the more than proportionate fall in the relative price of the agricultural good reduces the value marginal product of land in agriculture relative to that in non-agriculture, which results in endogenous switches in land use from agriculture to non-agriculture that are in general associated with violations of Gibrat's Law. Structural transformation away from agriculture, combined with the increasing relationship between the non-agricultural employment share and population density, generates the upward-sloping relationship between population growth and population density observed at intermediate densities, where MCDs are incompletely specialized in agriculture and non-agriculture.

# 5 Quantitative Predictions

In this section, we examine the quantitative relevance of structural transformation for accounting for observed patterns of population growth. We build on four key components of the model. First, as MCDs comprise multiple locations that specialize completely in either agriculture or non-agriculture, MCD population growth can be written as a weighted average of employment growth in agriculture and non-agriculture.[35] Second, the share of non-agriculture in employment is increasing in population density. Third, the relationship between employment growth and population density can differ between the agricultural and non-agricultural sectors. Fourth, the relationship between employment growth and initial population density depends on whether a location continues to specialize in the same sector in both time periods or whether it endogenously switches between sectors.

To illustrate the explanatory power of each of these components, we generate a sequence of predictions for MCD population growth, each of which uses progressively more components. We next compare the predicted relationship between population growth and initial

[34] See the web-based technical appendix for further discussion. While there is substantial empirical evidence of more rapid technological progress in agriculture than in agriculture (see again Martin and Mitra 2001) and inelastic demand between these broad categories of goods (see for example the discussion in Ngai and Pissarides 2007), structural transformation away from the agricultural sector could be also generated by labor-augmenting technological change and complementarity between labor and land in agriculture. Similarly, common technological progress in both sectors combined with non-homothetic preferences can also generate structural transformation, as discussed further in the web-based technical appendix.

[35] Consistent with the model's abstraction from labor force participation, we predict population growth using employment data, and compare the results to observed population growth.

population density to the actual relationship observed in the data. We undertake this comparison in two ways. First, we estimate our non-parametric specification (2) and display the results for predicted and actual population growth graphically in Figure 9. Second, to provide further evidence on the quantitative relevance of structural transformation, we regress actual population growth on predicted population growth and include a number of control variables.[36] We first undertake the analysis using our U.S. data before examining whether the model can also quantitatively account for our results using the Brazilian data. For brevity, we concentrate on results for the U.S. data with our baseline sample of "A and B" states. However, we find a qualitatively similar pattern with the other samples, as expected from the robustness checks above, and as discussed further below.

As a first step, Prediction 1 uses the property that MCD population growth is a weighted average of employment growth in agriculture and non-agriculture and makes the assumptions of (a) a common rate of employment growth within each sector across all MCDs and (b) no switching between agriculture and non-agriculture. In this prediction, the cross-section variation in population growth is predicted solely from the cross-section variation in the initial agricultural employment shares combined with common values of average employment growth within each sector for all MCDs. As evident from Figure 9, the employment share of an MCD in agriculture in 1880 goes a good way towards explaining its population growth from 1880-2000, providing strong evidence for the importance of structural transformation in shaping observed population dynamics.

Prediction 2 is the same as Prediction 1, except that it allows for mean reversion in agriculture. Whereas Prediction 1 regresses employment growth for each sector on a constant using the agricultural and non-agricultural samples from Table 1, Prediction 2 allows for agricultural mean reversion by augmenting the regression for agriculture with initial population density. The results of the regressions for agriculture and non-agriculture are reported in Columns (1) and (2) of Table 3. As shown in Figure 9, enriching the model in this way makes the downward-sloping relationship between population growth and initial population density observed at low densities more pronounced.[37]

[36] For each prediction, we also evaluate the implied population in 2000 and characterize the population distribution using the same methodology as used in Figure 2.

[37] As a robustness check, we also augmented the non-agricultural employment growth regression with initial population density, which although not shown in Figure 9 had no visible effect, since from Table 1 employment growth is largely uncorrelated with initial population density in non-agriculture. Finally, we experimented with allowing for richer forms of scale dependence within each sector by introducing polynomials in initial population density, which also had little effect on the relationship between predicted population growth and

In Predictions 1 and 2, we measured the common value of employment growth within each sector using employment growth in the most and least agricultural MCDs, which contain locations least likely to switch between sectors. In contrast, Prediction 3 takes into account the possibility of switching between sectors by allowing for a more flexible relationship between population growth and initial patterns of specialization in agriculture and non-agriculture.

Specifically, in Prediction 3, we regress total employment growth in each MCD on the 1880 agricultural employment share, the 1880 log population density, and the interaction term between these two variables. The inclusion of the initial agricultural employment share captures the role of structural transformation in shaping population growth, while the inclusion of initial log population density allows for the possibility of mean reversion in non-agriculture, and the inclusion of the interaction term between the two variables captures the extent to which mean reversion in agriculture differs from that in non-agriculture.

As column (3) of Table 4 shows, the agricultural employment share in 1880 is negatively correlated with subsequent population growth, reflecting structural transformation away from agriculture. Additionally, from the negative coefficient on the interaction term, the share of agriculture in 1880 employment has an even more negative effect on subsequent population growth in areas that were initially denser, reflecting mean reversion in agriculture. After controlling for these two terms, 1880 log population density is not significant, consistent with an absence of mean reversion in non-agriculture. We therefore use the coefficients from Column (4), which excludes initial log population density, to construct Prediction 3 shown in Figure 9.

As apparent from the figure, actual population growth rates are substantially more variable than predicted population growth rates and the actual data exhibit a sharper change in slope than the predicted values. Nonetheless, population growth in Prediction 3 closely replicates the observed pattern of violations of Gibrat's Law: the downward sloping relationship between population growth and initial population density at low densities, the upward sloping relationship at intermediate densities, and the largely flat relationship at high densities. The mean reversion in population growth rates at low initial population densities evident in Prediction 2 is further enhanced in Prediction 3, consistent with the idea that some of the mean reversion is the result of switches from agriculture to non-agriculture. Additionally,

initial population density.

mean population growth in Prediction 3 is closer to mean actual population growth, because some of the higher employment growth in non-agriculture is associated with these switches in land use, which are allowed for in Prediction 3.

To compare the quantitative relevance of structural transformation to that of other potential explanations for our findings, we also generate a Geographic Prediction, in which population growth is predicted solely from our set of geographic control variables. Except in the case where these geographic controls are orthogonal to population growth, the Geographic Prediction will typically capture some of the observed U-shaped relationship for population growth. However, the extent of the variation captured in the Geographic Prediction is substantially less than that explained by our structural transformation specifications. While each of Predictions 1-3 features a statistically significant difference in mean population growth between log population densities 2 and 4, the corresponding difference in mean population growth for the Geographic Prediction is statistically insignificant.

To provide further evidence on the quantitative relevance of structural transformation relative to alternative potential explanations, Table 4 reports regressions of actual against predicted population growth using our preferred specification of Prediction 3 and including a number of control variables. As a benchmark, we begin in Column (1) by regressing actual population growth rates on a constant. In Column (2), we augment that regression with the predicted population growth rates. Clearly there are many idiosyncratic factors affecting the population growth of individual MCDs that are not captured by our model, which results in a much larger variance of actual than of predicted population growth rates, as reflected in the regression $R^2$. Nonetheless, the coefficient on predicted population growth is positive, highly statistically significant and statistically indistinguishable from one.[38] Therefore, despite the much greater variance in the actual population growth rates, there is a close correspondence between actual and predicted population growth.

In Columns (3) to (5) of Table 4, we report a number of robustness checks for our baseline sample of "A and B" states, in which we show that the explanatory power of the model is robust to the inclusion of a number of control variables. After including measures of proximity to natural endowments, state fixed effects and county fixed effects, we continue to find a positive coefficient on predicted population growth that is large in magnitude and statistically significant. Columns (6) to (8) show that the same pattern of results holds for

[38] The standard errors in Table 4 are adjusted for predicted population growth being generated in a prior regression (Pagan 1984) and clustered on county.

the more restrictive sample of A states, the county sample and the suburban sample.

As the regressions in Columns (1) through (8) of Table 4 are estimated across MCDs, they exploit both variation across population density bins and variation across MCDs within population density bins. Similarly, the non-parametric estimates of specification (2) shown in Figure 9 exploited variation across population size bins. As a final step, we now examine whether structural transformation can explain variation in population growth within population density bins. In Column (9) of Table 4, we augment the baseline specification from Column (2) with a full set of fixed effects for initial population density bins. Even focusing solely on variation within population density bins, we continue to find a positive coefficient on predicted population growth that is large in magnitude and statistically significant. Column (10) of Table 4 shows that we continue to find the same pattern of results if we further augment this specification with our measures of proximity to natural endowments, and county fixed effects.

As an robustness check, the remainder of this section shows that we also find a very similar pattern of results for Brazil. Predictions 1-3 and the geography prediction are constructed in the same way for Brazil as for the U.S.[39] The employment growth regressions used in Predictions 2-3 for Brazil are reported in Table 5 (analogous to Table 3 for the U.S.). Having constructed these predictions, Figure 10 displays the results of estimating our non-parametric specification (2) for Brazil using actual and predicted population growth. As for the U.S., controlling simply for the initial agricultural employment share has considerable explanatory power for population growth (Prediction 1). Controlling for mean reversion in agriculture generates the downward-sloping relationship between population growth and initial population density at low densities (Prediction 2). A more flexible relationship between population growth and initial patterns of specialization to allow for switches from agriculture to non-agriculture again enhances the explanatory power of the model (Prediction 3). Finally, considering an alternative explanation, in which population growth is predicted based on our geographic control variables, fails to generate the upward–sloping relationship between population growth and initial population density observed at intermediate densities (Geographic Prediction).

Following the same structure as for the U.S., Table 6 reports the results of regressions of

[39] As noted above, to ensure a sufficient sample size, we construct the non-agricultural subsample for Brazil using AMCs that have an agricultural employment share in 1970 of less than less than 0.4 (instead of less than 0.2 for the U.S.).

actual against predicted population growth from the model using our preferred Prediction 3. While actual population growth again has a much higher variance than predicted population growth, the coefficient on predicted population growth is positive, highly statistically significant and statistically indistinguishable from one.[40] Therefore we again find a close correspondence between actual and predicted population growth. While Columns (1)-(4) include all AMCs, we find a similar pattern of results in Column (5), where we restrict attention to AMCs in the Northeast, Southeast and South of Brazil, which are smaller in geographic scope and are therefore likely to permit a finer discrimination between rural and urban areas. In Columns (6) and (7), we show that the model has explanatory power within as well as across population density bins by including a full set of fixed effects for population density bins.

Overall, there is considerable evidence that structural transformation can account for the quantitative as well as the qualitative patterns of observed population growth. Given the many differences between the U.S. and Brazil, and between the time periods considered, the similarity of the results in these two different contexts provides strong evidence in support of an explanation based on structural transformation.

# 6 Conclusion

While as recently as the nineteenth century less than one tenth 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 paper, we provide theory and evidence on urbanization using a new dataset that enables us to trace the transformation of the U.S. economy from a predominantly rural to largely urban society.

There are two main contributions of our analysis. First, we provide evidence of six stylized facts that are robust features of the urbanization process. 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 density observed at the intermediate densities where most of the population historically lived. In the same way that existing empirical regularities for

[40] Again the standard errors are adjusted for predicted population growth being generated in a prior regression (Pagan 1984).

densely-populated locations have proved fruitful in shaping theoretical models of cities, our new empirical regularities provide additional guidance for future theoretical research. A key challenge in modeling cities is arguably both explaining population growth among existing urban areas and accounting for the process through which urban areas are formed.

Second, we propose a simple theoretical explanation for our empirical findings based on structural transformation across sectors. Non-agriculture has stronger agglomeration forces and is less land-intensive than agriculture, which generates an increasing relationship between the share of non-agriculture in employment and population density. This increasing relationship is further reinforced in the model by mean reversion in agricultural productivity and constant proportional growth in non-agricultural productivity, which generates a non-agricultural productivity distribution that is unbounded from above. Structural transformation away from agriculture, combined with a non-agricultural employment share that increases with initial population density, generates the upward-sloping relationship between population growth and initial population density observed at intermediate densities.

While our explanation based on structural transformation abstracts from a number of other factors that are likely to influence population growth, the close connection between employment structure and population growth in our data, and the explanatory power of structural transformation, suggest that it is a key part of the urbanization process.

# References


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

Baumol, William J. (1967) "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crises," American Economic Review, 57(3), 415-26.

Beeson, Patricia E., David N. DeJong and Werner Troesken (2001) "Population growth in U.S. counties, 1840–1990," Regional Science and Urban Economics, 31(6), 669-699.

Black, Duncan and Vernon Henderson (1999) "A Theory of Urban Growth," Journal of Political Economy, 107(2), 252-284.

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

Brazil Census (1970, 2000), Integrated Public Use Microdata Series — International (IPUMS International)

Caselli, Francesco and Wilbor John Coleman II (2001) "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," Journal of Political Economy, 109(3), 584-616.

Census (1940), U.S. Census Bureau, Sixteenth U.S. Census of Population and Housing, www.census.gov/prod/www/abs/decennial/1940.htm

Census (2000a), U.S. Census Bureau, U.S. Census of Population and Housing, Shapefile downloaded from
www.census.gov/main/www/cen2000.html

Census (2000b), U.S. Census Bureau, American Factfinder, available at
factfinder.census.gov

Census (2000c), U.S. Census Bureau, Bureau's Geographic Areas Reference Manual, downloaded from
http://www.census.gov/geo/www/garm.html.

Census (2000d), U.S. Census Bureau, "Census 2000 Basics", Chapter 4: Geographic Areas, http://www.census.gov/mso/www/c2000basics/chapter4.htm.

Cleveland, William S. (1979) "Robust Locally Weighted Regression and Smoothing Scatterplots," Journal of the American Statistical Association, 74(368), 829-836.

Davis, Morris A. and François Ortalo-Magne (2008) "Household Expenditures, Wages, Rents," University of Wisconsin-Madison, mimeograph.

Davis, Donald R. and David Weinstein (2002) "Bones, Bombs, and Break Points: The Geography of Economic Activity ", American Economic Review, 92(5), 1269-1289.

Desmet, Klaus and Esteban Rossi-Hansberg (2007) " Spatial Growth and Industry Age," NBER Working Paper, 13302.

Duranton, Gilles (2007) "Urban Evolutions: The Fast, the Slow, and the Still," American Economic Review, 97(1), 197-221.

Echevarria, Cristina (1997) "Changes in Sectoral Composition Associated with Economic Growth," International Economic Review, 38(2), 431-452.

Eeckhout, Jan (2004) "Gibrat's Law for (All) Cities," American Economic Review, 94(5), 1429-1451.

Ellison, Glenn and Edward L. Glaeser (1999) "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" American Economic Review, 89(2), 311-316.

ESRI (1999), Arc View Database Access (Version 2.1a), Environmental Systems Research Institute, Inc.

Foster, Andrew D. and Mark R. Rosenzweig (2008) "Economic Development and the Decline of Agricultural Employment," Chapter 47 in (eds) T. Schultz and John Strauss, Handbook of Development Economics, Volume 4(5).

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

Gabaix, Xavier (1999) "Zipf's Law for Cities: An Explanation," Quarterly Journal of Economics, 114(3), 739-767.

Gabaix, Xavier and Yannis Ioannides (2004) "The Evolution of City Size Distributions," Handbook of Regional and Urban Economics, Volume 4, in (eds) Vernon Henderson and Jacques Thisse, North-Holland.

GIS (2003), Global GIS DVD-Global Coverage DVD 1st ed. U.S. Geological Data Series DDS-62 A-H Data Collections and Custom Programming the United States Geological Survey Published by the American Geological Institute, copyright 2003.

Glaeser, Edward L. (2008) "The Rise of the Sunbelt," Southern Economic Journal, 74(3), 610-643.

Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman and Andrei Shleifer (1992) "Growth in Cities," Journal of Political Economy, 100(6), 1126-1152.

Gollin, Douglas, Stephen Parente and Richard Rogerson (2002) "The Role of Agriculture in Development," American Economic Review, 92(2), 160-164.

González-Val, Rafael, Luis Lanaspa and Fernando Sanz (2008) "New Evidence on Gibrat's Law for Cities," MPRA Working Paper, 10411.

Härdle, Wolfgang (1990) "Applied Nonparametric Regression", Econometric Society Monographs, Cambridge University Press: Cambridge.

Henderson, J. Vernon (1974) "The Sizes and Types of Cities," American Economic Review, 64(4), 640-56.

Henderson, J. Vernon and Anthony J. Venables (2008) "The Dynamics of City Formation," NBER Working Paper, 13769.

Henderson, J. Vernon and Hyoung Gun Wang (2007) "Urbanization and City Growth: the Role of Institutions," Regional Science and Urban Economics, 37(3), 283-313.

Holmes, Thomas J. and Sanghoon Lee (2008) "Cities as Six-by-Six Mile Squares: Zipf's Law," University of Minnesota, mimeograph.

Ioannides, Yannis and Henry G. Overman (2004) "Spatial Evolution of the U.S. Urban System," Journal of Economic Geography, 4(2), 131–156.

IPEA (2008) - Instituto de Pesquisa Econômica Aplicada, Data Section, available at http://www.ipeadata.gov.br

Kim, Sukkoo (1995) "Expansion of Markets and the Geographic Distribution of Economic Activities: The trends in U.S.," Quarterly Journal of Economics, 110(4), 881-908.

Kim, Sukkoo (2000) "Urban Development in the United States, 1690-1990," Southern Economic Journal, 66(4), 855-880.

Kim, Sukkoo and Robert A. Margo (2004) "Historical Perspectives on U.S. Economic Geography" in (eds) Vernon Henderson and Jacques Thisse, Handbook of Regional and Urban Economics, Volume 4, North-Holland.

Kravis, Irving B., Alan W. Heston and Robert Summers (1983) "The Share of Services in Economic Growth," in F. Gerard Adams and Bert G. Hickman, eds., Global Econometrics: Essays in Honor of Lawrence R. Klein, Cambridge: MIT Press, 188-218.

Krugman, Paul (1991) "Increasing Returns and Economic Geography," Journal of Political Economy, 99(3), 483-499.

Kuznets, Simon (1955) "Economic Growth and Income Inequality," American Economic Review, 45, 1-28.

Martin, Will and Devashish Mitra (2001) "Productivity Growth and Convergence in Agriculture and Manufacturing," Economic Development and Cultural Change, 49(2), 403-422.
da Mata, Daniel, Uwe Deichmann, J. Vernon Henderson, Somik V. Lall and Hyoung G. Wang (2007) "Determinants of City Growth in Brazil," Journal of Urban Economics, 62(2), 252-272.
Matthews, Robert C., Charles H. Feinstein and John C. Odling-Smee (1982) "British Economic Growth 1856-1973", Stanford: Stanford University Press
Matsuyama, Kiminori (2002) "The Rise of Mass Consumption Societies," Journal of Political Economy, 110, 1035-1070.
Mckinsey (2008) "Preparing for China's Urban Billion", Mckinsey Global Institute: New York.
Ngai, Rachel and Chris Pissarides (2007) "Structural Change in a Multisector Model of Growth," American Economic Review, 97(1), 429-443.
NAPP 2006, North Atlantic Population Project and Minnesota Population Center. NAPP: Complete Count Microdata. NAPP Version 1.0 [computer files]. Minneapolis, MN: Minnesota Population Center [distributor], 2006. [www.nappdata.org]
Pagan, Adrian (1984) "Econometric Issues in the Analysis of Regressions with Generated Regressors," International Economic Review, 25(1), 221-247.
Prescott, Samuel T. (2003) "Federal Land Management - Current Issues and Background", Nova Publishing
Rappaport, Jordan and Jeffrey Sachs (2003) "The United States as a Coastal Nation," Journal of Economic Growth, 8(1), 5-46.
Rogerson, Richard (2008) "Structural Transformation and the Deterioration of European Labor Market Outcomes", Journal of Political Economy, 116(2), 235-259.
Rosen, Kenneth T. and Mitchel Resnick (1980) "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy", Journal of Urban Economics, 8(2), 165-186.
Rosenthal, Stuart S. and William C. Strange (2004) "Evidence on the Nature and Sources of Agglomeration Economies," Handbook of Urban and Regional Economics, (eds) J. Vernon Henderson and Jacques-Francois Thisse.
Rossi-Hansberg, Esteban and Mark L. Wright (2007) "Urban Structure and Growth," Review of Economic Studies, 74(2), 597-624.
Scott, David W. (1979) "On Optimal and Data-based Histograms," Biometrika, 66(3), 605-

610.

Silverman, B. W. (1986) "Density Estimation for Statistics and Data Analysis", Chapman and Hall: London.

Simon, Herbert A. (1955) "On a Class of Skew Distribution Functions," Biometrika, 42(3/4), 425–440.

Soo, Kwok Tong (2005) "Zipf's Law for Cities: A Cross-country Investigation", Regional Science and Urban Economics, 35(3), 239-263.

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

Syrquin, Moshe (1988) "Patterns of Structural Change," Chapter 7 in (eds) H. Chenery and T. N. Srinivasan, Handbook of Development Economics, 1, 205-273.

United Nations (2008) "World Urbanization Prospects - The 2007 Revision," published February 2008, accessible via www.un.org/esa/population/publications/wup2007/2007wup.htm.

# Table 1: US – Robustness of stylized facts

| | | (1) Baseline: A and B states | (2) Only A states | (3) Counties, 45 states and DC[1] | (4) Counties, A and B sample | (5) Hybrid Sample, 45 states and DC[2] | (6) Suburban A and B states[3] | (7) Log pop, not log density | (8) Baseline, geo controls[4] |
|---|---|---|---|---|---|---|---|---|---|
| Panel A | Standard deviation of log population density in 1880 ($\sigma_1$) | 0.967 | 1.025 | 1.757 | 0.963 | 1.272 | 0.904 | 0.833 | . |
| | Standard deviation of log population density in 2000 ($\sigma_2$) | 1.556 | 1.631 | 1.450 | 1.303 | 1.687 | 1.436 | 1.475 | . |
| | $H_0$: $\sigma_1 = \sigma_2$, vs. $H_1$: $\sigma_1 < \sigma_2$, p-value | <0.001 | <0.001 | 1.000 | <0.001 | <0.001 | <0.001 | <0.001 | . |
| | Stylized Fact 1: Distribution of log population density across geographic units became more dispersed from 1880-2000 (population became more concentrated) | Yes | Yes | No[5] | Yes | Yes | Yes | Yes | .[4] |
| Panel B | Mean population growth at log population density 0 ($\beta_g(0)$) | 0.013 | 0.012 | 0.016 | 0.019 | 0.010 | 0.013 | . | 0.013 |
| | Mean population growth at log population density 2 ($\beta_g(2)$) | 0.001 | -0.001 | 0.007 | 0.007 | 0.002 | 0.001 | . | 0.005 |
| | Mean population growth at log population density 4 ($\beta_g(4)$) | 0.009 | 0.010 | 0.014 | 0.014 | 0.011 | 0.008 | . | 0.011 |
| | $H_0$: $\beta_g(0) = \beta_g(2)$, $H_1$: $\beta_g(0) > \beta_g(2)$, p-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | . | <0.001 |
| | $H_0$: $\beta_g(2) = \beta_g(4)$, $H_1$: $\beta_g(2) < \beta_g(4)$, p-value | <0.001 | <0.001 | 0.001 | 0.011 | <0.001 | <0.001 | . | <0.001 |
| | Stylized Fact 2: U-shaped relationship between population growth from 1880-2000 and log population density in 1880 | Yes | Yes | Yes | Yes | Yes | Yes | .[6] | Yes |
| Panel C | Percent of agricultural in total employment at log population density 2 ($\beta_{sa}(2)$) | 0.767 | 0.762 | 0.691 | 0.618 | 0.738 | 0.769 | . | 0.743 |
| | Percent of agricultural in total employment at log population density 4 ($\beta_{sa}(4)$) | 0.227 | 0.189 | 0.195 | 0.185 | 0.228 | 0.221 | . | 0.235 |
| | $H_0$: $\beta_{sa}(2) = \beta_{sa}(4)$, $H_1$: $\beta_{sa}(2) > \beta_{sa}(4)$, p-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | . | <0.001 |
| | Stylized Fact 3: Share of agriculture in employment falls in the range where population density distribution in 1880 is positively correlated with population growth 1880-2000 | Yes | Yes | Yes | Yes | Yes | Yes | .[6] | Yes |
| Panel D | Standard deviation of agricultural employment in 1880 ($\sigma_{1a}$) | 0.820 | 0.722 | 1.677 | 0.810 | 1.084 | 0.820 | 0.820 | . |
| | Standard deviation of non-agricultural employment in 1880 ($\sigma_{1na}$) | 1.520 | 1.631 | 1.784 | 1.272 | 1.779 | 1.440 | 1.520 | . |
| | $H_0$: $\sigma_{1a} = \sigma_{1na}$, vs. $H_1$: $\sigma_{1a} < \sigma_{1na}$, p-value | <0.001 | <0.001 | 0.001 | <0.001 | <0.001 | <0.001 | <0.001 | . |
| | Standard deviation of agricultural employment in 2000 ($\sigma_{2a}$) | 0.858 | 0.853 | 0.806 | 0.617 | 0.936 | 0.851 | 0.858 | . |
| | Standard deviation of non-agricultural employment in 2000 ($\sigma_{2na}$) | 1.623 | 1.689 | 1.530 | 1.359 | 1.767 | 1.503 | 1.623 | . |
| | $H_0$: $\sigma_{2a} = \sigma_{2na}$, vs. $H_1$: $\sigma_{2a} < \sigma_{2na}$, p-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | . |
| | Stylized Fact 4: Standard deviation of non-agricultural employment is larger than standard deviation of agricultural employment in both years | Yes | Yes | Yes | Yes | Yes | Yes | Yes | .[4] |
| Panel E | Regress agricultural employment growth on log population density and intercept in subsample of units with agricultural employment share > 0.8 in 1880, report slope coefficient ($\beta_a$) | -0.0060 | -0.0077 | -0.0067 | -0.0054 | -0.0066 | -0.0060 | -0.0056 | -0.0055 |
| | $H_0$: $\beta_a = 0$, $H_1$: $\beta_a \neq 0$, p-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | 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 population density and intercept in subsample of units with non-agricultural employment share < 0.2 in 1880, report slope coefficient ($\beta_{na}$) | -0.0002 | -0.0006 | -0.0016 | -0.0006 | -0.0010 | -0.0001 | -0.0002 | -0.0005 |
| | $H_0$: $\beta_{na} = 0$, $H_1$: $\beta_{na} \neq 0$, p-value | 0.515 | 0.287 | <0.001 | 0.096 | <0.001 | 0.745 | 0.515 | 0.0991 |
| | Stylized Fact 6: Non-agricultural employment is uncorrelated with population density | Yes | Yes | No[7] | Yes | No[7] | Yes | Yes | Yes |
| | Number of observations | 10,864 | 4,439 | 2,496 | 819 | 19,229 | 10,159 | 10,864 | 10,864 |

Note: This table reports robustness tests of the 6 stylized facts using US data. All the regressions and tests reported in the table use robust standard errors clustered by county.

[1] The county sample includes all US 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.

[2] 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 5 states mentioned in footnote 1.

[3] In the Suburban Sample we merge any MCD with more than 100,000 inhabitants in 1880 to all the MCDs whose centroids lie within 25 kilometers of its centroid.

[4] The geographic control variables are state fixed effects, an indicator for the presence of coal, and indicators for the unit bordering on the ocean and for its centroid being within 50 kilometers from a lake or a river. As these specifications include controls, we do not test stylized facts 1 and 4.

[5] Since this sample includes many states that were not fully settled in 1880, many 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).

[6] In this sample we do not expect the turning point of the U and the fall of the agriculture share at coefficient 2, and hence do not report these coefficients. The figures qualitatively show that there is a U-shape whose minimum coincides with the drop in agricultural employment.

[7] Since this sample includes many states that were not fully settled in 1880, many near-empty areas increase the standard deviation of the population density distribution in that year. The future settlement of areas that were near empty in 1880 is also likely to cause mean reversion that is unrelated to employment structure.

## Table 2: Brazil – Robustness of stylized facts

| | | (1) All of Brazil (AMCs) | (2) As (1) but with state fixed effects | (3) As (1) but with geo controls[1] | (4) As (2) but with geo controls | (5) Brazil sub-sample[2] | (6) As (5) but with state fixed effects | (7) As (5) but with geo controls | (8) As (6) but with geo controls |
|---|---|---|---|---|---|---|---|---|---|
| Panel A | Standard deviation of log population density in 1970 ($\sigma_1$) | 1.222 | .[3] | 1.222 | .[3] | 1.009 | .[3] | 1.009 | .[3] |
| | Standard deviation of log population density in 2000 ($\sigma_2$) | 1.323 | . | 1.323 | . | 1.197 | . | 1.197 | . |
| | $H_0$: $\sigma_1 = \sigma_2$, vs. $H_1$: $\sigma_1 < \sigma_2$, p-value | <0.001 | . | <0.001 | . | <0.001 | . | <0.001 | . |
| | Stylized Fact 1: Distribution of log population density across geographic units became more dispersed from 1970-2000 (population became more concentrated) | Yes | . | Yes | . | Yes | . | Yes | . |
| Panel B | Mean population growth at log population density 0 ($\beta_g(0)$) | 0.0239 | 0.0239 | 0.0239 | 0.0239 | 0.0146 | 0.0146 | 0.0146 | 0.0146 |
| | Mean population growth at log population density 4 ($\beta_g(4)$) | 0.0079 | 0.0134 | 0.0116 | 0.0146 | 0.0079 | 0.0053 | 0.0090 | 0.0100 |
| | Mean population growth at log population density 6 ($\beta_g(6)$) | 0.0214 | 0.0271 | 0.0265 | 0.0305 | 0.0214 | 0.0190 | 0.0240 | 0.0258 |
| | $H_0$: $\beta_g(0) = \beta_g(4)$, $H_1$: $\beta_g(0) > \beta_g(4)$, p-value | <0.001 | 0.015 | 0.002 | 0.016 | <0.001 | <0.001 | 0.001 | 0.006 |
| | $H_0$: $\beta_g(4) = \beta_g(6)$, $H_1$: $\beta_g(4) < \beta_g(6)$, p-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | Stylized Fact 2: U-shaped relationship between population growth from 1970-2000 and log population density in 1970 | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Panel C | Percent of agricultural in total employment in 1970 at log population density 4 ($\beta_{sa}(4)$) | 0.6710 | 0.6710 | 0.6710 | 0.6710 | 0.6710 | 0.6710 | 0.6710 | 0.6710 |
| | Percent of agricultural in total employment in 1970 at log population density 6 ($\beta_{sa}(6)$) | 0.1677 | 0.1933 | 0.1459 | 0.1689 | 0.1677 | 0.1933 | 0.1447 | 0.1686 |
| | $H_0$: $\beta_{sa}(4) = \beta_{sa}(6)$, $H_1$: $\beta_{sa}(4) > \beta_{sa}(6)$, p-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | Stylized Fact 3: Share of agriculture in employment falls in the range where population density distribution in 1970 is positively correlated with population growth 1970-2000 | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Panel D | Standard deviation of agricultural employment in 1970 ($\sigma_{1a}$) | 0.8933 | .[3] | 0.8933 | .[3] | 0.8869 | .[3] | 0.8869 | .[3] |
| | Standard deviation of non-agricultural employment in 1970 ($\sigma_{1na}$) | 1.4157 | . | 1.4157 | . | 1.4287 | . | 1.4287 | . |
| | $H_0$: $\sigma_{1a} = \sigma_{1na}$, vs. $H_1$: $\sigma_{1a} < \sigma_{1na}$, p-value | <0.001 | . | <0.001 | . | <0.001 | . | <0.001 | . |
| | Standard deviation of agricultural employment in 2000 ($\sigma_{2a}$) | 1.0176 | . | 1.0176 | . | 0.9954 | . | 0.9954 | . |
| | Standard deviation of non-agricultural employment in 2000 ($\sigma_{2na}$) | 1.3754 | . | 1.3754 | . | 1.3642 | . | 1.3642 | . |
| | $H_0$: $\sigma_{2a} = \sigma_{2na}$, vs. $H_1$: $\sigma_{2a} < \sigma_{2na}$, p-value | <0.001 | . | <0.001 | . | <0.001 | . | <0.001 | . |
| | Stylized Fact 4: Standard deviation of non-agricultural employment is larger than standard deviation of agricultural employment in both years | Yes | . | Yes | . | Yes | . | Yes | . |
| Panel E | Regress agricultural employment growth on log population density and intercept in subsample of units with agricultural employment share > 0.8 in 1970, report slope coefficient ($\beta_a$) | -0.0038 | -0.0036 | -0.0022 | -0.0037 | -0.0042 | -0.0031 | -0.0028 | -0.0031 |
| | $H_0$: $\beta_a = 0$, $H_1$: $\beta_a \neq 0$, p-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | 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 population density and intercept in subsample of units with agricultural employment share < 0.4 in 1970, report slope coefficient ($\beta_{na}$) | 0.00126 | 0.00176 | 0.00027 | 0.00090 | 0.0013 | 0.00156 | 0.00030 | 0.00059 |
| | $H_0$: $\beta_{na} = 0$, $H_1$: $\beta_{na} \neq 0$, p-value | 0.124 | 0.0503 | 0.758 | 0.342 | 0.108 | 0.074 | 0.729 | 0.521 |
| | Stylized Fact 6: Non-agricultural employment is uncorrelated with population density | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| | Number of observations | 3.659 | 3.659 | 3.659 | 3.659 | 3,293 | 3,293 | 3,293 | 3,293 |

Note: This table reports robustness tests of the 6 stylized facts using data on Brazilian municipalities (Áreas Mínimas Comparáveis (AMCs)). All the regressions and tests reported in the table use robust standard errors.

[1] The geographic controls are twelve dummy variables indicating the presence of oil, nickel, manganese, iron, gold, copper, cobalt, and aluminum, whether the AMC borders the ocean, lies within 50 kilometers of a river, has its centroid covered with tropical or subtropical moist broadleaf forest, or is contained in the Amazonas Area.

[2] This subsample uses only AMCs in the states of the Northeast, Southeast, and South official regions of Brazil, since AMCs in these regions are relatively small, allowing a clearer distinction between rural and urban areas. The three regions in this subsample cover about 90 percent of Brazil's AMCs, 36 percent of its land area and 91 percent of its population in 1970.

[3] As these specifications include controls, we do not test stylized facts 1 and 4, which involve measuring standard deviations.

## Table 3: US – Generate predictions

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | For prediction 2 | | For prediction 3 | |
| Employment growth rate, 1880-2000 | Non-agric. | Agric. | Total | Total |
| Constant | 0.011 | -0.005 | 0.014 | 0.014 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Log population density in 1880 | | -0.006 | -0.0002 | |
| | | (0.000) | (0.0003) | |
| Share of agriculture 1880 | | | -0.008 | -0.007 |
| | | | (0.002) | (0.001) |
| (Share of agriculture in 1880) x (log population density in 1880) | | | -0.0010 | -0.0013 |
| | | | (0.0005) | (0.0004) |
| Number of observations | 755 | 3,074 | 10,856 | 10,856 |
| $R^2$ | 0 | 0.31 | 0.063 | 0.063 |
| Sample: | A and B, non-agric | A and B, agric | A and B | A and B |

Note: This table reports the regressions used to generate predictions 2 and 3 for the US data. We construct prediction 2 using the predicted values of sectoral employment growth from the regressions reported in columns (1) and (2), as described in the text of the paper. We construct prediction 3 using the predicted values of employment growth from the regression reported in column (4), as described in the text. The non-agricultural subsample used in column (1) includes MCDs from our baseline A and B Sample for which agriculture's share of 1880 employment was less than 0.2. The agricultural subsample used in column (2) includes MCDs from our baseline A and B Sample for which agriculture's share of 1880 employment exceeded 0.8. Robust standard errors in parentheses are clustered by county.

## Table 4: US – Quantifying the explanatory power of prediction 3

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Inter-cept only | As (1) but with predicted growth | As (2) but with geo controls[1] | As (3) with state fixed effects | As (4) but with county fixed effects | As (2) with A Sample only | As (2) but with county sample | As (2) but with suburban sample | As (2) with log pop density bins[2] | As (5) with log pop density bins[2] |
| Actual population growth regression | | | | | | | | | | |
| Predicted population growth | | 1.041 | 0.798 | 0.629 | 0.633 | 1.221 | 1.011 | 1.057 | 0.648 | 0.674 |
| | | (0.06) | (0.055) | (0.067) | (0.047) | (0.078) | (0.057) | (0.062) | (0.079) | (0.047) |
| Intercept | 0.475 | -0.026 | | | | | | | | |
| | (0.034) | (0.045) | | | | | | | | |
| $R^2$ | 0 | 0.098 | 0.183 | 0.303 | 0.617 | 0.173 | 0.433 | 0.098 | 0.151 | 0.64 |
| Number of observations | 10,864 | 10,864 | 10,864 | 10,864 | 10,864 | 4,439 | 2,496 | 10,159 | 10,864 | 10,864 |
| Regression used to generate predicted population growth | | | | | | | | | | |
| Share of agriculture in 1880 | | -1.05 | -1.039 | -0.982 | -0.871 | -1.075 | -0.428 | -0.74 | -1.217 | -0.913 |
| | | (0.017) | (0.017) | (0.016) | (0.014) | (0.021) | (0.051) | (0.016) | (0.046) | (0.023) |
| (Share of agriculture in1880) x (log population density in 1880) | | -0.162 | -0.157 | -0.147 | -0.151 | -0.171 | -0.797 | -0.245 | -0.077 | -0.119 |
| | | (0.006) | (0.006) | (0.006) | (0.005) | (0.008) | (0.019) | (0.006) | (0.017) | (0.009) |
| F – statistic[3] | | 5,243 | 4,876 | 6,657 | 7,211 | 4,548 | 1,194 | 5,886 | 3,268 | 8,061 |

Note: This table shows the predictive power of prediction 3 for various specifications using US data. The upper panel of the table reports the regressions of actual population growth on predicted population growth. The lower panel of the table reports the regression whose fitted values are used for predicted population growth. The left-hand side variable in the lower panel of the table is total employment growth. Robust standard errors clustered by county are in parentheses. The standard errors in the upper panel of the table have been adjusted for the fact that predicted population growth is generated using a prior regression (Pagan 1984).

[1] The geographic control variables are an indicator for the presence of coal, and indicators for observations bordering on the ocean and for observations whose centroid lies within 50 kilometers of a lake or a river.

[2] The log population density bin fixed effects included in these regressions are a full set of dummy variables for MCDs having population densities within intervals of 0.1 log points. For example, all MCDs with log population density from 0.1 to 0.2 are grouped together in bin 0.1.

[3] The F-statistic reported is for an F-test that the coefficients on the share of agriculture and the interaction term are jointly equal to zero in the prior regression used to generate predicted population growth.

## Table 5: Brazil – Generating the predictions

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | For prediction 2 | | For prediction 3 | |
| Employment growth rate 1970-2000 | Non-agric. | Agric. | Total | Total |
| Constant | 0.039 | 0.00216 | 0.045 | 0.043 |
| | (0.001) | (0.00111) | (0.004) | (0.001) |
| Log population density in 1970 | | -0.0038 | -0.0005 | |
| | | (0.0004) | (0.0008) | |
| Share of agriculture 1970 | | | -0.0317 | -0.0291 |
| | | | (0.0044) | (0.0016) |
| (Share of agriculture in 1970) x (log population density in 1970) | | | -0.0037 | -0.0043 |
| | | | (0.0009) | (0.0004) |
| Number of observations | 384 | 1,651 | 3,659 | 3,659 |
| $R^2$ | 0 | 0.059 | 0.262 | 0.262 |
| Sample: | AMCs non-agric. | AMCs agric. | AMCs | AMCs |

Note: This table reports the regressions we used to construct predictions 2 and 3 for the Brazilian municipalities (Áreas Mínimas Comparáveis (AMCs)) data. We construct prediction 2 using the predicted values of sectoral employment growth from the regressions reported in columns (1) and (2), as described in the text of the paper. We construct prediction 3 using the predicted values of employment growth from the regression reported in column (4), as described in the text of the paper. The non-agricultural subsample used in column (1) includes AMCs for which agriculture's share of 1970 employment was less than 0.4 due to the small sample size using a threshold of 0.2 (but results are similar using a 0.2 threshold). The agricultural subsample used in column (2) includes AMCs for which agriculture's share of 1970 employment exceeded 0.8. Robust standard errors are in parentheses.

## Table 6: Brazil – Quantifying the explanatory power of prediction 3

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Intercept only | As (1) but with predicted growth | As (2) but with geo controls[1] | As (3) but with state fixed effects | As (4) but with subsample[4] only | As (2) but with log pop density bins[2] | As (4) but with log pop density bins |
| Actual population growth | | | | | | | |
| Predicted population growth | | 1.024 | 0.968 | 1.112 | 1.122 | 0.909 | 0.915 |
| | | (0.035) | (0.035) | (0.040) | (0.042) | (0.036) | (0.036) |
| Intercept | 0.269 | 0.010 | | | | | |
| | (0.009) | (0.010) | | | | | |
| $R^2$ | 0 | 0.196 | 0.315 | 0.378 | 0.350 | 0.287 | 0.385 |
| Number of observations | 3,659 | 3,659 | 3,659 | 3,659 | 3,659 | 3,659 | 3,659 |
| Regression used to generate predicted population growth | | | | | | | |
| Share of agriculture in 1970 | | -0.810 | -0.821 | -0.755 | -0.885 | -0.693 | -0.708 |
| | | (0.013) | (0.015) | (0.015) | (0.017) | (0.044) | (0.045) |
| (Share of agriculture in 1970) x (log population density in 1970) | | -0.122 | -0.125 | -0.129 | -0.073 | -0.158 | -0.159 |
| | | (0.003) | (0.003) | (0.004) | (0.005) | (0.012) | (0.012) |
| F – statistic[3] | | 5,460 | 5,651 | 4,728 | 4,088 | 4,042 | 4,212 |

Note: This table shows the predictive power of prediction 3 for various specifications using the Brazilian municipalities (Áreas Mínimas Comparáveis (AMCs)) data. The upper panel of the table reports the regression of actual population growth on predicted population growth. The lower panel of the table reports the regression whose fitted values are used for predicted population growth. The left-hand side variable in the lower panel of the table is total employment growth. Robust standard errors are in parentheses. The standard errors in the upper panel of the table have been adjusted for the fact that predicted population growth is generated using a prior regression (Pagan 1984).

[1] The geographic controls are twelve dummy variables indicating the presence of oil, nickel, manganese, iron, gold, copper, cobalt, and aluminum, whether the AMC borders the ocean, lies within 50 kilometers of a river, has its centroid covered with tropical or subtropical moist broadleaf forest, or is contained in the Amazonas Area.

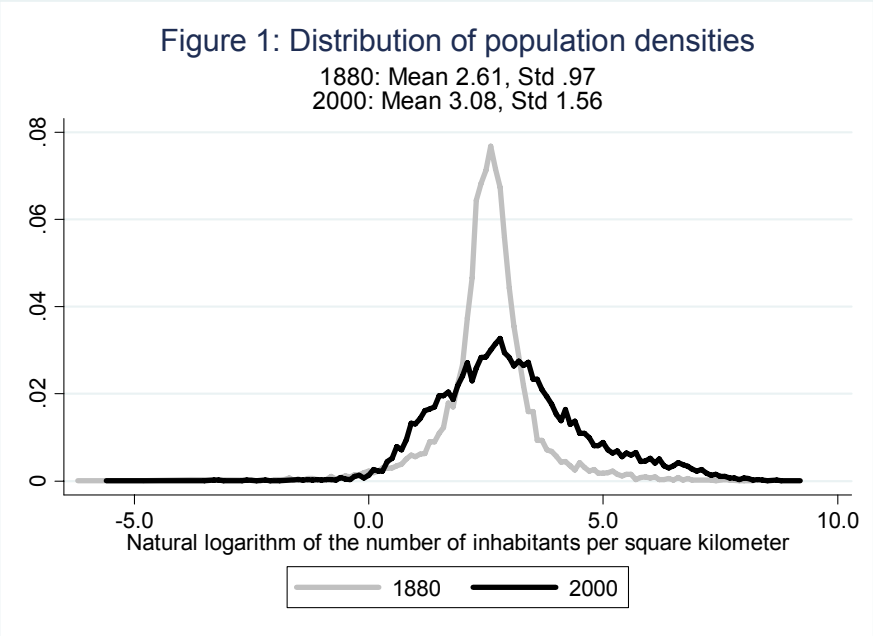
[2] The log population density bin fixed effects included in these regressions are a full set of dummy variables for MCDs having population densities within intervals of 0.1 log points. For example, all AMCs with log population density from 0.1 to 0.2 are grouped together in bin 0.1.

[3] The F-statistic reported is for an F-test that the coefficients on the share of agriculture and the interaction term are jointly equal to zero in the prior regression used to generate predicted population growth.

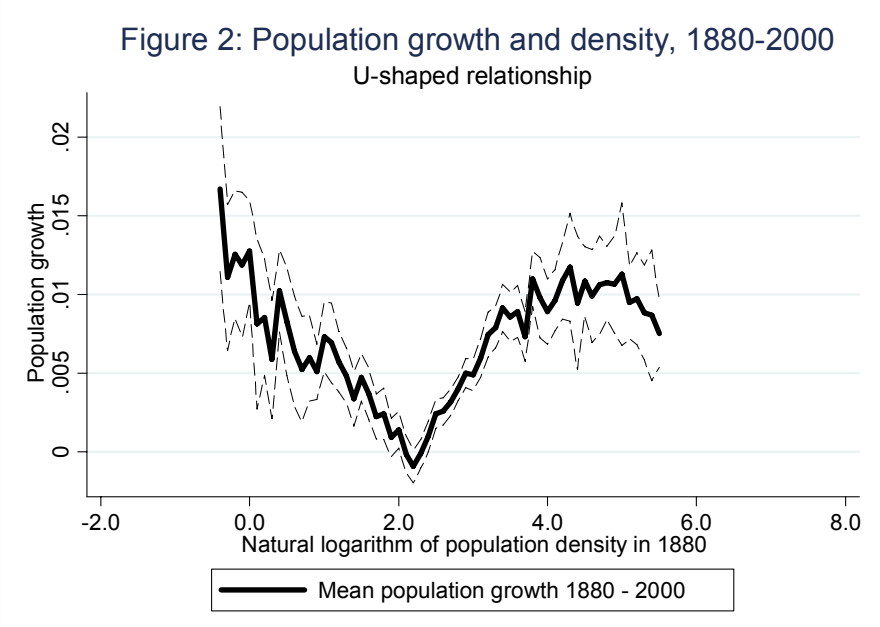
[4] This subsample uses only AMCs in the states of Northeast, Southeast, and South macro regions of Brazil, since AMCs in these regions are relatively small, allowing a clearer distinction between rural and urban areas. These three macro regions in this subsample cover about 90 percent of Brazil's AMCs, 36 percent of its land area and 91 percent of its population in 1970.

## Map 1: US MCD data by state and county

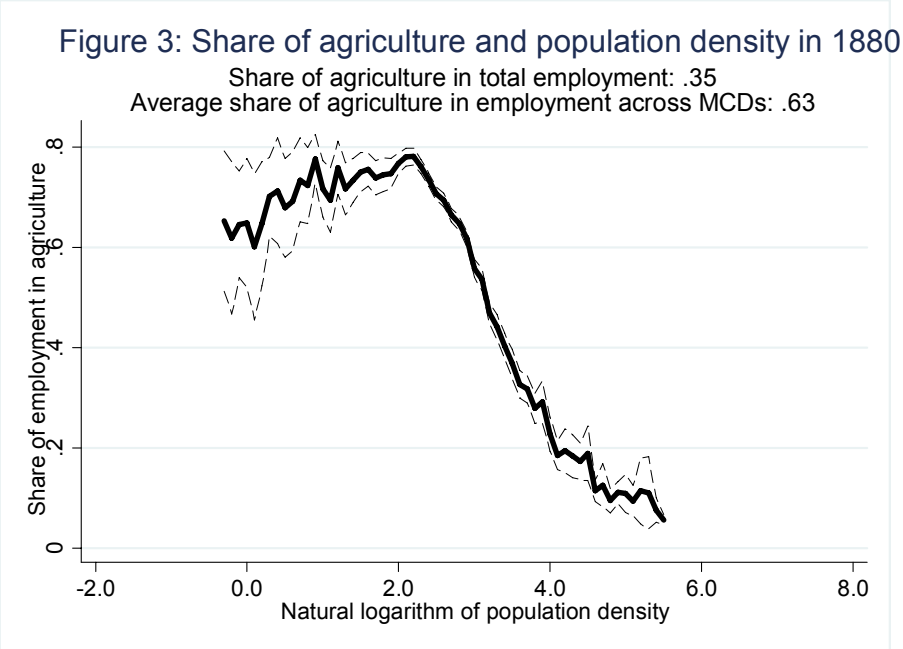

Note: This map shows the states used for our various samples. Our baseline sample consists of A and B states. The classification A, B and C corresponds to the quality of the match rate between 1880 and 2000 MCDs. In states classified as A (Connecticut, DC, Indiana, Iowa, Massachusetts, New Hampshire, 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 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 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. 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.



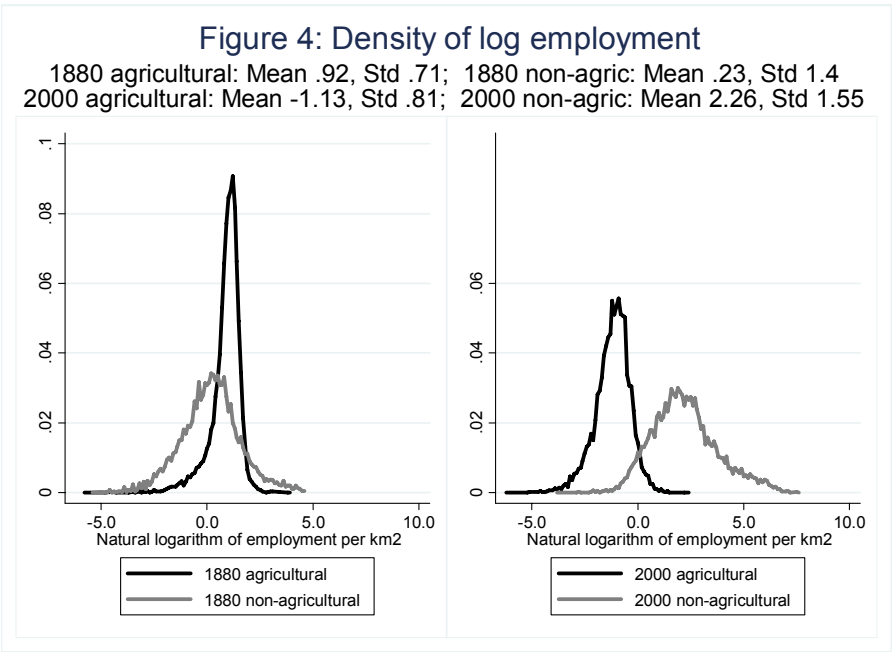

Note: This figure shows the distribution of log population per square kilometer in 1880 and 2000 estimated using non-parametric specification (1) for the sample of "A and B" states. Population density bins are defined by rounding down log population density for each MCD to the nearest single digit after the decimal point. For example, all MCDs with log population density ≥ 0.1 and < 0.2 are grouped together in bin 0.1. See the web-based technical appendix for further details on data.



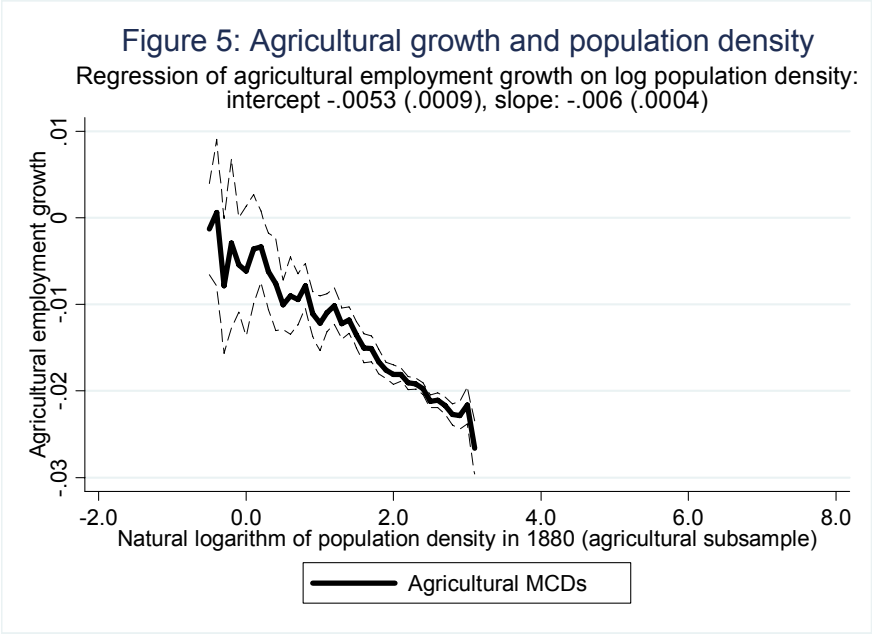

Note: The solid line shows mean population growth rate from 1880-2000 within each population density bin based on estimating non-parametric specification (2) for the sample of "A and B" states. Population density bins are defined by rounding down log population density for each MCD to the nearest single digit after the decimal point. The dashed lines show 95 percent confidence intervals based on robust standard errors clustered by county. Since population density bins at the extreme ends of the distribution typically contain at most one observation, the figure (but not the estimation) omits the 1 percent most and least dense MCDs in 1880. See the web-based technical appendix for further details on data.



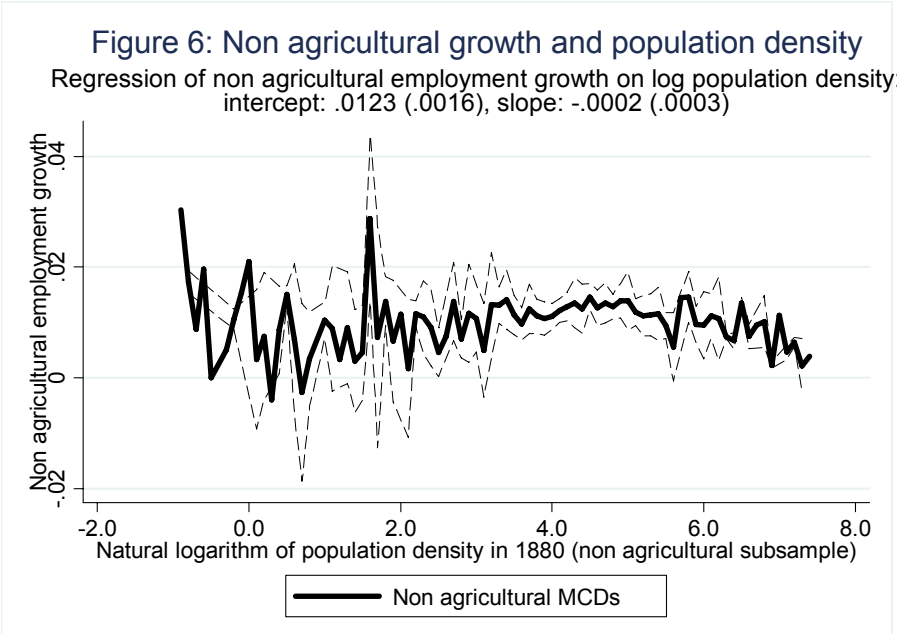

Note: The solid line shows the mean share of agriculture in 1880 employment within each population density bin based on estimating non-parametric specification (2) for the sample of "A and B" states. Population density bins are defined by rounding down log population density for each MCD to the nearest single digit after the decimal point. The dashed lines show 95 percent confidence intervals based on robust standard errors clustered by county. Since population density bins at the extreme ends of the distribution typically contain at most one observation, the figure (but not the estimation) omits the 1 percent most and least dense MCDs in 1880. See the web-based technical appendix for further details on data.




Note: This figure shows the distribution of log agricultural employment and log non-agricultural employment (employment in industry and services) per square kilometer in 1880 and 2000 estimated using non-parametric specification (1) for the sample of "A and B" states. Employment density bins are defined by rounding down log employment density for each MCD to the nearest single digit after the decimal point. Since population density bins at the extreme ends of the distribution typically contain at most one observation, the figure (but not the estimation) omits the 1 percent most and least dense MCDs in 1880. See the web-based technical appendix for further details on data.

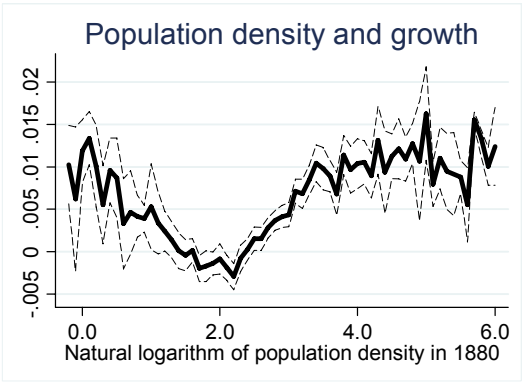



Note: The solid line shows the mean growth rate of agricultural employment from 1880-2000 within each population density bin based on estimating non-parametric specification (2) for the agricultural subsample (an agricultural share in 1880 employment of greater than 0.8) within "A and B" states. Population density bins are defined by rounding down log population density for each MCD to the nearest single digit after the decimal point. The dashed lines show 95 percent confidence intervals based on robust standard errors clustered by county. Since population density bins at the extreme ends of the distribution typically contain at most one observation, the figure (but not the estimation) omits the 1 percent most and least dense MCDs in 1880. See the web-based technical appendix for further details on data.

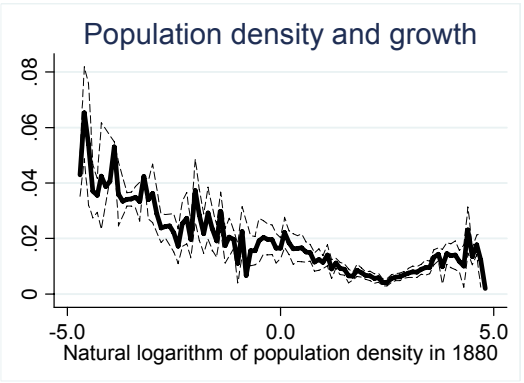



Note: The solid line shows the mean growth rate of non-agricultural employment (employment in industry and services) from 1880-2000 within each population density bin based on estimating non-parametric specification (2) for the non-agricultural subsample (an agricultural share in 1880 employment of less than 0.2) within "A and B" states. Population density bins are defined by rounding down log population density for each MCD to the nearest single digit after the decimal point. The dashed lines show the 95 percent confidence intervals based on robust standard errors clustered by county. Since population density bins at the extreme ends of the distribution typically contain at most one observation, the figure (but not the estimation) omits the 1 percent most and least dense MCDs in 1880. See the web-based technical appendix for further details on data.

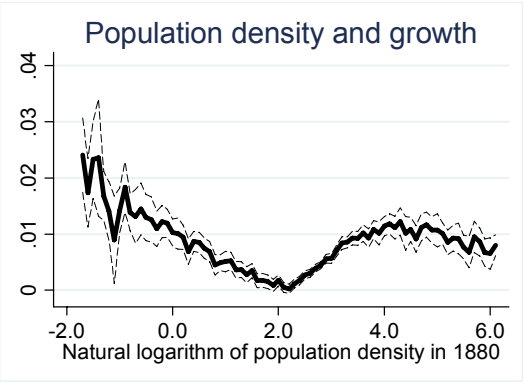
Figure 7: Robustness of U-shaped population growth relationship



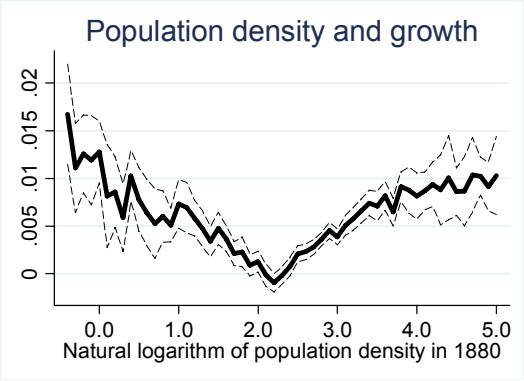

Panel A: A states sample



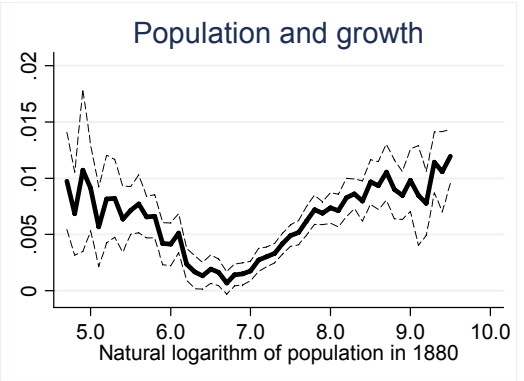

Panel B: Counties sample



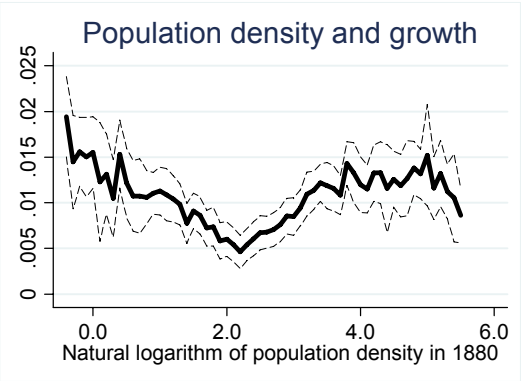

Panel C: Hybrid sample




Panel D: Suburban sample



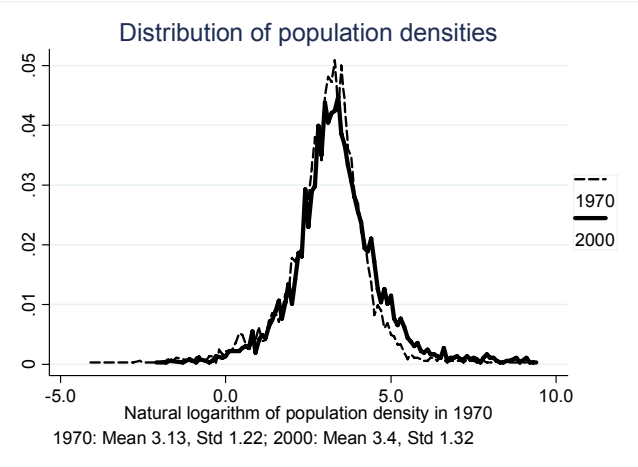

Panel E: A and B, population not densities



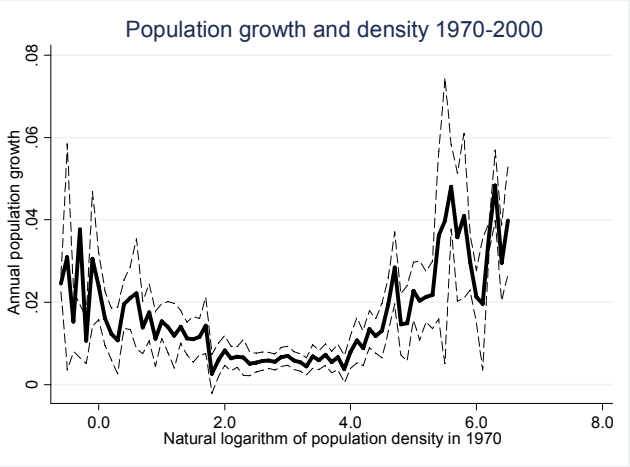

Panel F: A and B sample, state controls

Note: This figure shows the robustness of the U-shaped relationship for population growth (Figure 2) by reproducing it for other samples. The various samples used here are described in the web-based technical appendix. Since population density bins at the extreme ends of the distribution typically contain at most one observation, the figure (but not the estimation) omits the 1 percent most and least dense MCDs in 1880.

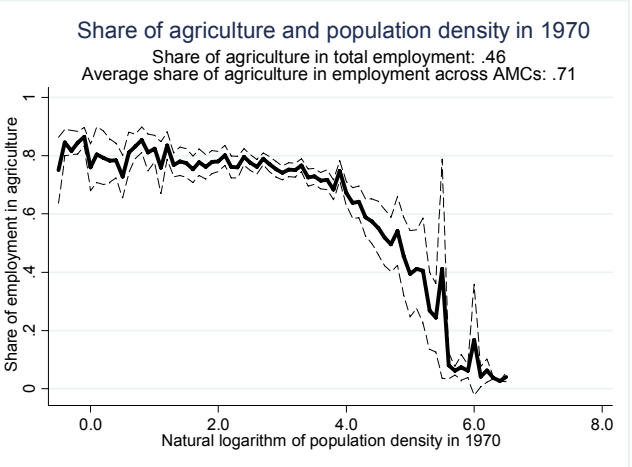
Figure 8: Brazilian results



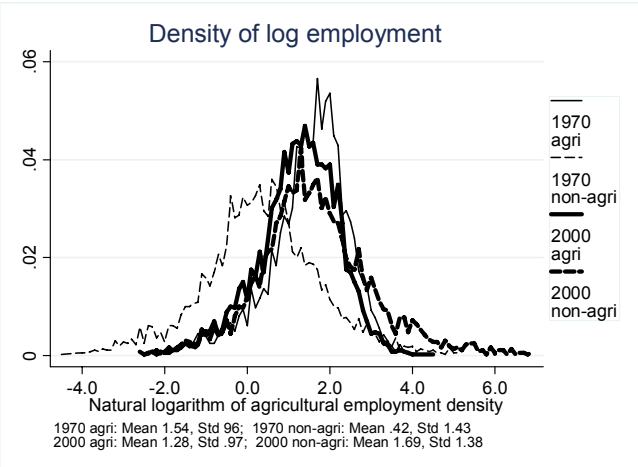

Panel A



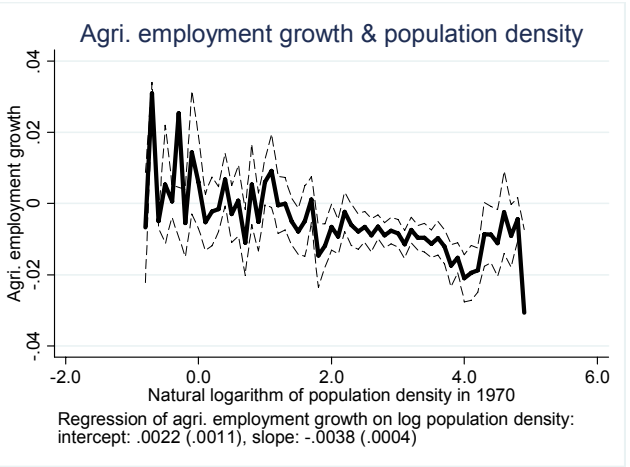

Panel B



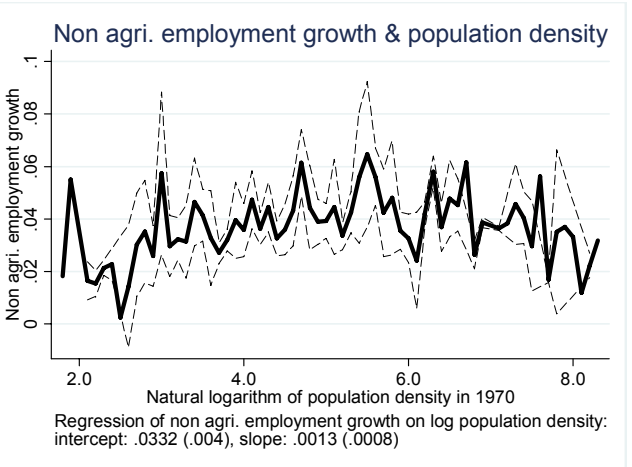

Panel C




Panel D



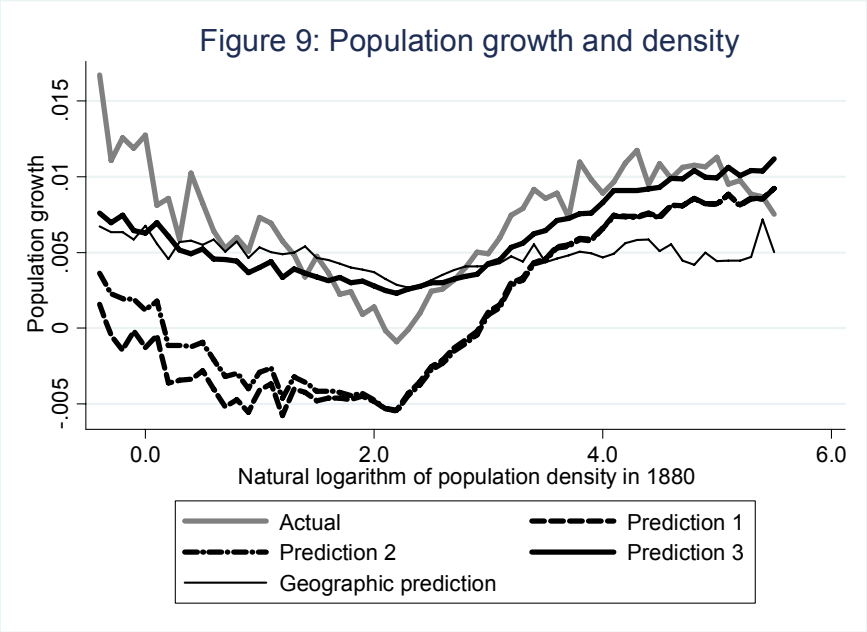

Panel E



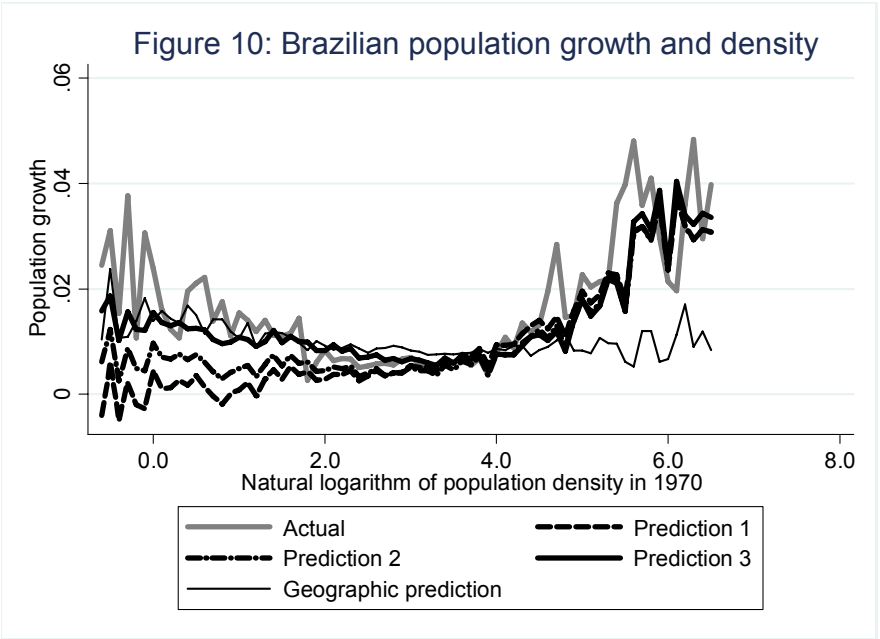

Panel F

Note: This figure reproduces Figures 1 to 6 but uses Brazilian instead of US data.

.




Note: Mean actual and predicted population growth from 1880-2000 within population density bins based on estimating non-parametric specification (2) for the sample of "A and B" states. Population density bins are defined by rounding down log population density for each MCD to the nearest single digit after the decimal point. The figure (but not the estimation) omits the 1 percent densest and sparsest MCDs in 1880. Predictions 1-3 use progressively more components of the model to generate predicted population growth: Prediction 1 allows initial employment shares to vary by MCD. Prediction 2 allows for mean reversion in agriculture and Prediction 3 allows for switching of sectors within MCD. The geographical prediction predicts population growth using dummies indicating proximity to lakes, rivers, the sea and coal as discussed in the paper.




Note: This figure reproduces Figure 9 but uses Brazilian instead of US data. Predictions 1-3 and the Geographic Prediction are constructed in a similar way as for the U.S. as discussed in the paper.