Technical Appendix for Urbanization and Structural Transformation: Not for Publication

Guy Michaels
London School of Economics, CEP and CEPR
Ferdinand Rauch
London School of Economics and CEP
Stephen J. Redding
Princeton University, NBER and CEPR

November 9, 2011

I. Introduction

This web-based technical appendix contains additional supplementary material for the paper. Sections II.-VII. mirror the sections of the same name and number in the paper. In Section II., no additional material is required. In Section III., we provide further detail on the model, including the derivations of all expressions and results reported in the paper. In Section IV., we report detailed information on the construction of the U.S. MCD and county data. In Section V., we present additional results for the baseline empirical evidence section of the paper. In Section VI., we include additional results for the further evidence section of the paper. In Section VII., no additional material is required. In Section VIII., we discuss the construction of the Brazilian data and report the results of the robustness test using these data that is discussed in the paper.

II. Related Literature

No additional results required.

III. Theoretical Model

In this section, we develop in further detail the theoretical model outlined in the paper. We present the complete technical derivations for all the expressions and results reported in the paper. In the interests of clarity and to ensure that this section of the web appendix is self-contained, we reproduce some material from the paper, but also include the intermediate steps for the derivation of expressions.

We consider an economy in which workers are mobile across locations and choose between agriculture and non-agriculture as sectors of employment. The mechanism that drives an aggregate reallocation of employment from agriculture to non-agriculture is either more rapid productivity growth in agriculture combined with inelastic demand across the two goods or a change in relative demand for these two goods (e.g. as a result of non-homothetic preferences). This aggregate reallocation affects the relationship between population growth and initial population density because agriculture's share of employment varies with population density. Agricultural specialization and population density are related because agriculture is land intensive, exhibits weaker agglomeration forces, and is characterized by greater mean reversion in productivity than non-agriculture, which implies that agriculture's share of employment declines at the highest population densities.

III.A. Preferences and Endowments

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

Workers derive utility from consumption of goods, C_{it}, and residential land use, H_{Uit}, and for simplicity we assume that the utility function takes the Cobb-Douglas form:[1]

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

The goods consumption index, C_{it}, includes consumption of agriculture, c_{Ait}, and non-

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

agriculture, c_{Nit}, and is assumed to take the constant elasticity of substitution (CES) form:

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

where a_t captures the relative strength of demand for agriculture and, following a large literature in macroeconomics, we assume that the two goods are complements: $0 < \kappa < 1$.[2]

III.B. Production Technology

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

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

where Y_{jit}, L_{jit} and H_{jit} denote output, employment and commercial land use respectively in sector $j \in \{A, N\}$ in location i at time t; $L_{jit}^{\eta_j}$ captures external economies of scale in employment in a sector and location; Γ_{jt} is a component of sectoral productivity that is common across all locations (e.g. the aggregate state of technology); θ_{jit} is a component of sectoral productivity that is specific to each location (e.g. natural resources and weather).[3]

Within each sector, output is homogeneous and markets are perfectly competitive, with each good costlessly tradeable across locations.[4] Each firm is of measure zero and chooses its inputs of labor and land to maximize its profits taking as given productivity, goods and factor prices, and the location decisions of other firms and workers. Since economies of scale are external to the firm, they depend on aggregate (rather than firm) employment in a sector and location. As such, each firm's production technology exhibits constant returns to scale in its own inputs of labor and land, which yields the standard result that payments to labor and land exactly exhaust the value of output. We assume that agriculture is more land intensive than non-agriculture ($0 < \mu_A < \mu_N < 1$) and that non-agriculture has stronger external economies of scale than agriculture ($0 \leq \eta_A < \eta_N$).

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

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

[2] See, for example, Ngai and Pissarides (2007).

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

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

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

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

Land in each location can be allocated to residential or commercial use. When land is used commercially, we assume that it can be employed either in agriculture or non-agriculture but not in both sectors simultaneously, so that locations exhibit complete specialization in production. Since locations are grouped into larger statistical units (MCDs or counties) these larger statistical units exhibit incomplete specialization to the extent that they contain a mixture of agricultural and non-agricultural locations.[5]

III.C. Consumption Expenditure

Utility maximization implies that each worker allocates constant shares of income α and $(1-\alpha)$ to goods consumption and residential land use respectively:

$$C_{it} = \frac{\alpha \pi_{it}}{P_t}, \qquad H_{Uit} = \frac{(1-\alpha)\pi_{it}}{r_{it}}, \tag{5}$$

where π_{it} denotes income per worker; r_{it} denotes the rental rate on land; P_t is the price index for goods consumption. Since the goods consumption index (2) takes the CES form, the dual price index for goods consumption is:

$$P_t = \left[a_t^{\kappa} p_{At}^{1-\kappa} + (1-a_t)^{\kappa} p_{Nt}^{1-\kappa}\right]^{\frac{1}{1-\kappa}}, \tag{6}$$

where p_{At} and p_{Nt} are the prices of the agricultural and non-agricultural goods respectively. Costless trade ensures that these prices, and hence the price index for goods consumption, are the same across all locations.

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

Using utility maximization, consumer expenditure on the agricultural good is:

$$p_{At}c_{Ait} = a_t^{\kappa} p_{At}^{1-\kappa} E_{it} P_t^{\kappa-1},$$

where $E_{it} = P_t C_{it}$ denotes total goods consumption expenditure. Using the above expression and the price index for goods consumption, the share of agriculture in goods consumption expenditure can be written as follows:

$$e_{Ait} = e_{At} = \frac{p_{At}c_{Ait}}{P_t C_{it}} = \frac{1}{1 + \left(\frac{1-a_t}{a_t}\right)^{\kappa} \left(\frac{p_{Nt}}{p_{At}}\right)^{1-\kappa}}, \tag{7}$$

where, with inelastic demand between the two goods ($0 < \kappa < 1$), the share of agriculture in goods consumption expenditure is *increasing* in its relative price (p_{At}/p_{Nt}).

Expenditure on land in each location is redistributed lump sum to the workers residing in that location, as in Helpman (1998). Therefore total income in each location equals payments to labor and land used in production plus expenditure on residential land use. Using complete specialization in production, total income in a location producing good $j \in \{A, N\}$ is:

$$\pi_{it} L_{it} = p_{jt} Y_{jit} + (1-\alpha)\,\pi_{it} L_{it}.$$

Income per worker in each location is therefore proportional to the value of output per worker:

$$\pi_{it} = \frac{p_{jt} Y_{jit}}{\alpha L_{it}}. \tag{8}$$

III.D. Land Allocation

With competitive factor markets and complete specialization, commercial land in each location is employed in the sector with the higher value marginal product of land. As a result, for each agricultural productivity (θ_{Ait}) there is a cutoff non-agricultural productivity ($\theta^*_{Nit}(\theta_{Ait})$), such that a location produces the non-agricultural good for $\theta_{Nit} \geq \theta^*_{Nit}(\theta_{Ait})$. Similarly, for each non-agricultural productivity (θ_{Nit}), there is a cutoff agricultural productivity ($\theta^*_{Ait}(\theta_{Nit})$), such that a location produces the agricultural good for $\theta_{Ait} > \theta^*_{Ait}(\theta_{Nit})$. These cutoffs for non-agricultural and agricultural productivity are related to one another and defined by the equality of the value marginal products of land in the two sectors:

$$\begin{aligned} \theta^*_{Nit}(\theta_{Ait}) &= \frac{(1-\mu_A)\, p_{At} \Gamma_{At} \theta_{Ait} L_{it}^{\eta_A+\mu_A} H_{Ait}^{-\mu_A}}{(1-\mu_N)\, p_{Nt} \Gamma_{Nt} L_{it}^{\eta_N+\mu_N} H_{Nit}^{-\mu_N}}, \\ \theta^*_{Ait}(\theta_{Nit}) &= \frac{(1-\mu_N)\, p_{Nt} \Gamma_{Nt} \theta_{Nit} L_{it}^{\eta_N+\mu_N} H_{Nit}^{-\mu_N}}{(1-\mu_A)\, p_{At} \Gamma_{At} L_{it}^{\eta_A+\mu_A} H_{Ait}^{-\mu_A}}, \end{aligned} \tag{9}$$

where we assume that the non-agricultural good is produced in the case of indifference between the two patterns of complete specialization.

The equilibrium rental rate on land in each location is determined by the requirement that the total demand for land equals the total supply of land for that location. Since the total demand for land is the sum of residential and commercial land use, the land market clearing condition for a location producing good $j \in \{A, N\}$ is:

$$H_{Uit} + H_{jit} = H_i.$$

Using the Cobb-Douglas upper tier of utility and production technology, this land market clearing condition can be re-written in value terms as:

$$(1-\alpha)\,\pi_{it}L_{it} + \left(1-\mu_j\right)p_{jt}Y_{jit} = r_{it}H_i.$$

Noting from (8) that total income in each location is proportional to the value of production, the land market clearing condition becomes:

$$\left[(1-\alpha) + \left(1-\mu_j\right)\alpha\right]\pi_{it}L_{it} = r_{it}H_i. \tag{10}$$

Using this land market clearing condition and the fact from (5) that expenditure on residential land is a constant share of total income, the equilibrium allocation of land involves a constant fraction of land allocated to residential and commercial use:

$$\chi_{Uj} = \frac{1-\alpha}{(1-\alpha) + \left(1-\mu_j\right)\alpha}, \qquad \chi_{Yj} = \frac{\left(1-\mu_j\right)\alpha}{(1-\alpha) + \left(1-\mu_j\right)\alpha}, \tag{11}$$

where χ_{Uj} and χ_{Yj} denote the fractions of land used residentially and commercially, respectively, when good j is produced.

III.E. Population Mobility

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

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

where the dual price index for consumption goods, P_t, is the same for all locations; the rental rate on land, r_{it}, in general varies across locations.

Using land market clearing (10), the relationship between income per worker and the value of production (8), the production technology (3) and the equilibrium allocation of land (11), this population mobility condition can be re-expressed as:

$$\tilde{V}_{it} = \frac{p_{jt}\Gamma_{jt}\theta_{jit}\chi_{Yj}^{1-\mu_j} H_i^{1-\mu_j+\frac{1-\alpha}{\alpha}} L_{it}^{\eta_j-\left(1-\mu_j\right)-\frac{1-\alpha}{\alpha}}}{\beta_j} = V_t^{1/\alpha} P_t = \tilde{V}_t, \tag{12}$$

$$\beta_j \equiv \alpha \left[(1-\alpha) + \left(1-\mu_j\right)\alpha\right]^{\frac{1-\alpha}{\alpha}},$$

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

Labor market clearing requires that the population of all locations sums to the economy's labor endowment:

$$\int_0^1 L_{it} di = L_t. \tag{13}$$

A final equation comes from goods market clearing, which requires that the share of the agricultural good in aggregate revenue equals its share in aggregate expenditure. Using the consumer expenditure share (7), this goods market clearing condition is:

$$r_{At} = \frac{p_{At}Y_{At}}{p_{At}Y_{At} + p_{Nt}Y_{Nt}} = \frac{a_t^{\kappa} p_{At}^{1-\kappa}}{a_t^{\kappa} p_{At}^{1-\kappa} + (1-a_t)^{\kappa} p_{Nt}^{1-\kappa}} = e_{At}. \tag{14}$$

where Y_{At} and Y_{Nt} denote aggregate output of each good. Using the unit continuum of locations, the productivity cutoffs for agriculture and non-agriculture (9), and the land allocation (11), aggregate output of each good can be expressed as:

$$Y_{At} = \Gamma_{At} \int_{\underline{\theta}_N}^{\bar{\theta}_N} \left[\int_{\theta_A^*(\theta_{Nit})}^{\bar{\theta}_A} \theta_{Ait} L_{it}^{\eta_A+\mu_A} \left(\chi_{YA} H_i\right)^{1-\mu_A} z_A\left(\theta_{Ait}\right) d\theta_{Ait} \right] z_N\left(\theta_{Nit}\right) d\theta_{Nit}, \tag{15}$$

$$Y_{Nt} = \Gamma_{Nt} \int_{\underline{\theta}_A}^{\bar{\theta}_A} \left[\int_{\theta_N^*(\theta_{Akt})}^{\bar{\theta}_N} \theta_{Nkt} L_{kt}^{\eta_N+\mu_N} \left(\chi_{YN} H_k\right)^{1-\mu_N} z_N\left(\theta_{Nkt}\right) d\theta_{Nkt} \right] z_A\left(\theta_{Akt}\right) d\theta_{Akt}, \tag{16}$$

where $z_j\left(\theta_{jit}\right)$ is the limiting distribution of location-specific productivity in sector j , which depends on the law of motion for location-specific productivity (4) and the idiosyncratic productivity shocks $\left(g_j\left(\phi_{jit}\right)\right)$. We characterize these limiting distributions and their support $\left[\underline{\theta}_j, \bar{\theta}_j\right]$ in the proof of Proposition 1 below.

III.F. General Equilibrium

The general equilibrium of the model can be referenced by the limiting distributions of productivity in each sector $(z_A(\theta_{Ait}),\ z_N(\theta_{Nit}))$, the sets of locations producing each good $(\Omega_{At} \equiv \{\theta_{Ait} : \theta_{Ait} > \theta^*_{Ait}(\theta_{Nit})\}$ and $\Omega_{Nt} \equiv \{\theta_{Nit} : \theta_{Nit} \geq \theta^*_{Nit}(\theta_{Ait})\})$, the measure of workers in each location (L_{it}), and the relative price of the agricultural good (p_{At}), where we choose the non-agricultural good as the numeraire, so that $p_{Nt} = 1$. From this information, all other endogenous variables of the model can be determined, as shown in the proof of Proposition 1 below.

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

Proof. (a) Limiting productivity distributions: We begin by solving for the limiting productivity distributions in each sector. From (4), we have:

$$\ln\theta_{jit} = \nu_j \ln\theta_{jit-1} + \ln\phi_{jit}, \qquad t = \{1, \ldots, \infty\}, \qquad \ln\theta_{ji0} = \ln\phi_{ji0,} \tag{17}$$

where we assume $0 < \nu_A < \nu_N \leq 1$ and $\ln\phi_{jit}$ is independently and identically distributed with mean zero, constant variance $\sigma^2_{\phi j} > 0$ and bounded support.

(i) We begin with the case where $0 < \nu_j < 1$ for both sectors. Consider a given location i within a given sector j. Using (17), productivity in year t can be expressed as:

$$\ln\theta_{jit} = \sum_{k=0}^{t} \nu_j^k \ln\phi_{jit-k}. \tag{18}$$

From this expression, mean log productivity at time t for location i within sector j is:

$$\mathbb{E}\left[\ln\theta_{jit}\right] = 0,$$

while the variance of log productivity at time t for location i within sector j is:

$$\mathbb{V}\left[\ln\theta_{jit}\right] = \sum_{k=0}^{t} \left(\nu_j^2\right)^k \sigma^2_{\phi j}.$$

Taking the limit as $t \to \infty$, we have:

$$\lim_{t\to\infty} \mathbb{V}\left[\ln\theta_{jit}\right] = \frac{\sigma^2_{\phi j}}{1-\nu_j^2}. \tag{19}$$

Since $0 < \nu_j < 1$, it follows that the variance of log productivity is finite. Furthermore, the covariance of log productivity between times t and s depends only on the time difference $t-s$:

$$\mathbb{C}\left(\ln\theta_{jit}, \ln\theta_{jis}\right) = \frac{\sigma^2_{\phi j}\nu_j^{t-s}}{1-\nu_j^2}.$$

Since have shown that $\{\ln\theta_{jit}\}$ has a constant mean, finite second moments, and a covariance that depends only on the time difference, we have shown that $\{\ln\theta_{jit}\}$ is weakly or covariance stationary. From (18), we also have:

$$\mathbb{E}\left[\left(\ln\theta_{jit} - \sum_{k=0}^{t}\nu_j^k \ln\phi_{jit-k}\right)^2\right] = 0.$$

It follows that $\{\ln\theta_{jit}\}$ converges in mean square:

$$\lim_{t\to\infty}\mathbb{E}\left[\left(\ln\theta_{jit} - \sum_{k=0}^{t}\nu_j^k \ln\phi_{jit-k}\right)^2\right] = 0.$$

which in turn implies that $\{\ln\theta_{jit}\}$ converges in probability.

$$\Pr\left\{\lim_{t\to\infty}\ln\theta_{jit} = \lim_{t\to\infty}\sum_{k=0}^{t}\nu_j^k \ln\phi_{jit-k}\right\} = 1.$$

Therefore the limiting distribution for productivity for location i in sector j is determined by:

$$\lim_{t\to\infty}\ln\theta_{jit} = \lim_{t\to\infty}\sum_{k=0}^{t}\nu_j^k \ln\phi_{jit-k}, \qquad j\in\{A,N\} \text{ for } 0<\nu_j<1, \tag{20}$$

where since $\sum_{k=0}^{\infty}\nu_j^k < \infty$, the right-hand side is convergent, and there exists a strictly stationary limiting distribution for productivity.

Since the right-hand side of (20) is a linear combination of a sequence of independently and identically distributed continuous random variables $\ln\phi_{jit-k}$, it follows that the limiting distribution for productivity $(z_j(\theta_{jit}))$ is continuous in θ_{jit}.

Since the stochastic process for productivity (17) is the same for each location i within sector j, the limiting distribution determined by (20) is the same for all locations and corresponds to the limiting cross-section distribution of productivity across the unit continuum of locations $(z_j(\theta_{jit}))$.

To determine the support of this limiting distribution, note that θ_{jit}, is bounded from above since:

$$\sup\{\ln\theta_{jit}\} = \sum_{k=0}^{t}\nu_j^k \ln\overline{\phi}_j.$$

Taking the limit as $t \to \infty$ and using $0 < \nu_j < 1$, we have:

$$\ln \bar{\theta}_j = \lim_{t \to \infty} \sup \{\ln \theta_{jit}\} = \frac{\ln \overline{\phi}_j}{1 - \nu_j}, \tag{21}$$

Note that θ_{jit} is also bounded from below since:

$$\inf \{\ln \theta_{jit}\} = \sum_{k=0}^{t} \nu_j^k \ln \underline{\phi}_j.$$

Taking the limit as $t \to \infty$ and using $0 < \nu_j < 1$, we have:

$$\ln \underline{\theta}_j = \lim_{t \to \infty} \inf \{\ln \theta_{jit}\} = \frac{\ln \underline{\phi}_j}{1 - \nu_j}, \tag{22}$$

where, from our assumption that $g_j \left(\ln \phi_{jit}\right)$ has bounded support and a mean of zero, we have $-\infty < \ln \underline{\phi}_j < 0$ and hence $0 < \underline{\phi}_j < 1$.

(ii) We next turn to the case of $\nu_N = 1$. Consider a given location i within sector N. Using (17), productivity in year t can be expressed as:

$$\ln \theta_{Nit} = \sum_{k=0}^{t} \ln \phi_{Nit-k}. \tag{23}$$

From this expression, mean productivity at time t for location i within sector N is:

$$\mathbb{E} \left[\ln \theta_{Nit}\right] = 0,$$

while the variance of productivity at time t for location i within sector N is:

$$\mathbb{V} \left[\ln \theta_{Nit}\right] = \sum_{k=0}^{t} \sigma_{\phi N}^2.$$

Taking the limit as $t \to \infty$, we have:

$$\lim_{t \to \infty} \mathbb{V} \left[\ln \theta_{Nit}\right] = \infty. \tag{24}$$

Nonetheless a limiting distribution for productivity exists. Since $\ln \phi_{Nit}$ is independently and identically distributed with constant mean and variance, it follows from the Central Limit Theorem that $\ln \theta_{Nit}$ converges to a normal distribution as $t \to \infty$:

$$\lim_{t \to \infty} \ln \theta_{Nit} \sim \mathcal{N} \left(0, t\sigma_{\phi N}^2\right). \tag{25}$$

Therefore the limiting distribution for productivity in non-agriculture $(z_N(\theta_{Nit}))$ when $\nu_N = 1$ is log normal and continuous in θ_{Nit}.

Since the stochastic process for productivity (17) is the same for each location within sector N, it follows that the limiting cross-section distribution of productivity across the unit continuum of locations $(z_N(\theta_{Nit}))$ is also log normal.

Note that θ_{Nit} is unbounded from above since:

$$\sup\{\ln\theta_{Nit}\} = \sum_{k=0}^{t} \ln\overline{\phi}_N.$$

Taking the limit as $t \to \infty$, we have:

$$\ln\bar{\theta}_N = \lim_{t\to\infty} \sup\{\ln\theta_{Nit}\} = \lim_{t\to\infty} \sum_{k=0}^{t} \ln\overline{\phi}_N = \infty. \tag{26}$$

In contrast, note that θ_{Nit} is bounded from below by zero since:

$$\inf\{\ln\theta_{Nit}\} = \sum_{k=0}^{t} \ln\underline{\phi}_N.$$

$$\exp\{\inf\{\ln\theta_{Nit}\}\} = \prod_{k=0}^{t} \underline{\phi}_N.$$

Taking the limit as $t \to \infty$, we have:

$$\underline{\theta}_N = \lim_{t\to\infty} \exp\{\inf\{\ln\theta_{Nit}\}\} = \lim_{t\to\infty} \prod_{k=0}^{t} \underline{\phi}_N = 0, \tag{27}$$

since, from our assumption that $g_j\left(\ln\phi_{jit}\right)$ has bounded support and a mean of zero, we have $-\infty < \ln\underline{\phi}_j < 0$ and hence $0 < \underline{\phi}_j < 1$.

(b) Population and goods produced: As shown in part (a) of the proof above, the limiting productivity distribution in each sector $\{z_A(\theta_{Ait}),\ z_N(\theta_{Nit})\}$ can be determined independently of all other endogenous variables of the model. In part (b) of the proof, we now show how $\{L_{it}, \theta^*_{Ait}, \theta^*_{Nkt}, \Omega_{At}, \Omega_{Nt}\}$ can be determined given these limiting distributions and given the relative price of the agricultural good (p_{At}). In part (c) of the proof below, we endogenize the relative price of the agricultural good and jointly determine $\{L_{it}, \theta^*_{Ait}, \theta^*_{Nkt}, \Omega_{At}, \Omega_{Nt}, p_{At}\}$ in general equilibrium.

Using the population mobility (12) condition, the price index (6) and our choice of numeraire

($p_{Nt}=1$), we obtain the following closed-form expressions for the equilibrium population of each location when it produces the agricultural and non-agricultural good respectively:

$$\hat{L}_{Ait}=L_{Ait}\left(p_{At},V_t\right)=\left[\frac{\Gamma_{At}\chi_{YA}^{1-\mu_A}}{\left[a_t^{\kappa}+(1-a_t)^{\kappa}\,p_{At}^{\kappa-1}\right]^{\frac{1}{1-\kappa}}V_t^{1/\alpha}\beta_A}\right]^{\xi_A}\theta_{Ait}^{\xi_A}H_i^{\xi_A\eta_A+1}, \tag{28}$$

$$\hat{L}_{Nit}=L_{Nit}\left(p_{At},V_t\right)=\left[\frac{\Gamma_{Nt}\chi_{YN}^{1-\mu_N}}{\left[a_t^{\kappa}p_{At}^{1-\kappa}+(1-a_t)^{\kappa}\right]^{\frac{1}{1-\kappa}}V_t^{1/\alpha}\beta_N}\right]^{\xi_N}\theta_{Nit}^{\xi_N}H_i^{\xi_N\eta_N+1}, \tag{29}$$

$$\beta_j\equiv\alpha\left[(1-\alpha)+\left(1-\mu_j\right)\alpha\right]^{\frac{1-\alpha}{\alpha}}>0,\qquad j\in\{A,N\},$$

$$\xi_j\equiv\frac{1}{\frac{1-\alpha}{\alpha}+\left(1-\mu_j\right)-\eta_j}>0,\qquad j\in\{A,N\},$$

which determine equilibrium population as a function of the given relative price of the agricultural good (p_{At}), the common level of utility across all populated locations (V_t), productivity ($\{\theta_{Ait},\ \theta_{Nit}\}$ which has already been determined), and parameters (since $\{\chi_{YA},\ \chi_{YN}\}$ are determined as a function of parameters in (11) and H_i is an endowment).

Using the cutoff productivities for commercial land use (9), we also have:

$$\left.\begin{array}{r}
i\in\hat{\Omega}_{At}=\Omega_{At}\left(p_{At},V_t\right)=\left\{\theta_{Ait}:\theta_{Ait}>\hat{\theta}^*_{Ait}\right\},\\
k\in\hat{\Omega}_{Nt}=\Omega_{Nt}\left(p_{At},V_t\right)=\left\{\theta_{Nkt}:\theta_{Nkt}\geq\hat{\theta}^*_{Nkt}\right\},\\
\hat{\theta}^*_{Ait}=\theta^*_{Ait}\left(p_{At},V_t\right)=\frac{(1-\mu_N)\Gamma_{Nt}\theta_{Nit}\hat{L}_{Nit}^{\eta_N+\mu_N}(\chi_{YN}H_i)^{-\mu_N}}{(1-\mu_A)p_{At}\Gamma_{At}\hat{L}_{Ait}^{\eta_A+\mu_A}(\chi_{YA}H_i)^{-\mu_A}},\\
\hat{\theta}^*_{Nkt}=\theta^*_{Nkt}\left(p_{At},V_t\right)=\frac{(1-\mu_A)p_{At}\Gamma_{At}\theta_{Akt}\hat{L}_{Akt}^{\eta_A+\mu_A}(\chi_{YA}H_k)^{-\mu_A}}{(1-\mu_N)\Gamma_{Nt}\hat{L}_{Nkt}^{\eta_N+\mu_N}(\chi_{YN}H_k)^{-\mu_N}},
\end{array}\right\}, \tag{30}$$

which determines the sets of locations producing the agricultural and non-agricultural goods $\{\hat{\Omega}_{At}\ ,\ \hat{\Omega}_{Nt}\}$, as a function of the given relative price of the agricultural good (p_{At}), the common level of utility across all populated locations (V_t through $\hat{L}_{At}$ and $\hat{L}_{Nt}$), productivity ($\{\theta_{Ait},\ \theta_{Nit}\}$ which has already been determined), and parameters (since $\{\chi_{YA},\ \chi_{YN}\}$ are determined as a function of parameters in (11) and H_i is an endowment).

From the labor market clearing condition (13), we also have:

$$\int_0^1\hat{L}_{it}di=L_t, \tag{31}$$

$$\hat{L}_{it}=\left\{\begin{array}{ll}\hat{L}_{Ait} & \text{if } i\in\hat{\Omega}_{At},\\ \hat{L}_{Nit} & \text{if } i\in\hat{\Omega}_{Nt}\end{array}\right..$$

From (28) and (29), $\lim_{V_t \to 0} \hat{L}_{Ait} = \lim_{V_t \to 0} \hat{L}_{Nit} = \infty$, $\lim_{V_t \to \infty} \hat{L}_{Ait} = \lim_{V_t \to \infty} \hat{L}_{Nit} = 0$, and both $\hat{L}_{Ait}$ and $\hat{L}_{Nit}$ are monotonically decreasing in V_t. Using (28), (29) and (30) in the labor market clearing condition (31), and noting that population in each location is decreasing in V_t irrespective of which good is produced, the labor market clearing condition determines a unique value of V_t as a function of the given value of p_{At} and parameters. Having determined V_t, the equilibrium population and productivity cutoffs $\{\hat{L}_{Ait}, \hat{L}_{Nit}, \hat{\theta}^*_{Ait}, \hat{\theta}^*_{Nkt}\}$ in (28), (29) and (30) depend solely on the given value of p_{At} and parameters.

Thus the model's supply-side can be characterized in terms of relative goods prices, endowments and parameters. Using (28), (29) and (30), we can also establish the following comparative statics of $\{\hat{L}_{Ait}, \hat{L}_{Nit}, \hat{\theta}^*_{Ait}, \hat{\theta}^*_{Nkt}\}$ with respective to the given value of p_{At}:

$$\begin{array}{c} \frac{d\hat{L}_{Ait}}{dp_{At}} > 0, \\ \frac{d\hat{L}_{Nkt}}{dp_{At}} < 0, \\ \frac{d\hat{\theta}^*_{Ait}}{dp_{At}} < 0, \\ \frac{d\hat{\theta}^*_{Nkt}}{dp_{At}} > 0. \end{array} \tag{32}$$

Since there is a unit continuum of locations and we have established above that the limiting distributions of productivity in agriculture $(z_A(\theta_{Ait}))$ and non-agriculture $(z_N(\theta_{Nit}))$ are continuous, it follows from aggregate output of each good ((15) and (16)) and the comparative statics (32) that aggregate output of each good is a continuous function of the given value of p_{At}:

$$\begin{aligned} \frac{dY_{At}}{dp_{At}} &> 0, \\ \frac{dY_{Nt}}{dp_{At}} &< 0. \end{aligned} \tag{33}$$

(c) Relative prices: Having characterized the supply-side of the model as a function of p_{At}, we now use the goods market clearing condition (14) to combine supply and demand and jointly determine $\{L_{it}, \theta^*_{Ait}, \theta^*_{Nkt}, \Omega_{At}, \Omega_{Nt}, p_{At}\}$ in general equilibrium. The goods market clearing condition in value terms equates the share of the agricultural good in aggregate revenue (r_{At}) and expenditure (e_{At}). It proves convenient to rewrite this goods market clearing condition in terms of monotonic transformations of r_{At} and e_{At}:

$$\frac{r_{At}}{1 - r_{At}} = \frac{p_{At} Y_{At}}{Y_{Nt}} = \frac{a_t^{\kappa} p_{At}^{1-\kappa}}{(1 - a_t)^{\kappa}} = \frac{e_{At}}{1 - e_{At}},$$

where $0 < a_t < 1$ and we have used our choice of numeraire: $p_{Nt} = 1$. Taking logarithms of the left and right-hand sides, we obtain:

$$\ln\left(\frac{r_{At}}{1-r_{At}}\right) = \ln\left(\frac{Y_{At}}{Y_{Nt}}\right) + \ln p_{At}, \tag{34}$$

$$\ln\left(\frac{e_{At}}{1-e_{At}}\right) = \ln\left(\frac{a_t^{\kappa}}{(1-a_t)^{\kappa}}\right) + (1-\kappa)\ln p_{At}, \tag{35}$$

where $0 < \kappa < 1$. Differentiating these relationships yields:

$$\frac{d\ln\left(\frac{r_{At}}{1-r_{At}}\right)}{d\ln p_{At}} = \frac{d\ln\left(\frac{Y_{At}}{Y_{Nt}}\right)}{d\ln p_{At}} + 1, \tag{36}$$

$$\frac{d\ln\left(\frac{e_{At}}{1-e_{At}}\right)}{d\ln p_{At}} = (1-\kappa)\,. \tag{37}$$

From our analysis of the comparative statics of the model's supply-side in (33), we have:

$$\frac{d\ln\left(\frac{Y_{At}}{Y_{Nt}}\right)}{d\ln p_{At}} > 0, \tag{38}$$

which implies using (36) and (37):

$$\frac{d\ln\left(\frac{r_{At}}{1-r_{At}}\right)}{d\ln p_{At}} > \frac{d\ln\left(\frac{e_{At}}{1-e_{At}}\right)}{d\ln p_{At}}. \tag{39}$$

Furthermore, from aggregate output of each good (15) and (16), equilibrium population (28) and (29), commercial land use (30) and labor market clearing (31), we have $\lim_{p_{At}\to 0} Y_{At} = 0$ and $\lim_{p_{At}\to\infty} Y_{Nt} = 0$. It follows that:

$$\lim_{p_{At}\to 0} \ln\left(\frac{Y_{At}}{Y_{Nt}}\right) = -\infty,$$

$$\lim_{p_{At}\to\infty} \ln\left(\frac{Y_{At}}{Y_{Nt}}\right) = +\infty.$$

Combining these results with (34) and (35), and noting $0 < \kappa < 1$, we have:

$$\lim_{p_{At}\to 0}\left\{\ln\left(\frac{r_{At}}{1-r_{At}}\right) - \ln\left(\frac{e_{At}}{1-e_{At}}\right)\right\} < 0, \tag{40}$$

$$\lim_{p_{At}\to\infty}\left\{\ln\left(\frac{r_{At}}{1-r_{At}}\right) - \ln\left(\frac{e_{At}}{1-e_{At}}\right)\right\} > 0. \tag{41}$$

Using these two limiting results and the difference in elasticities in (39), it follows that there exists a single crossing-point for $\ln\left(\frac{r_{At}}{1-r_{At}}\right)$ and $\ln\left(\frac{e_{At}}{1-e_{At}}\right)$ at which $0 < p_{At} < \infty$ and

$0 < e_{At} = r_{At} < 1$.

We illustrate this single crossing-point in Figure A.17, which displays $\ln\left(\frac{r_{At}}{1-r_{At}}\right)$ and $\ln\left(\frac{e_{At}}{1-e_{At}}\right)$ on the y-axis against $\ln p_{At}$ on the x-axis. From (35), $\ln\left(\frac{e_{At}}{1-e_{At}}\right)$ takes the value $\ln\left(\frac{a_t^\kappa}{(1-a_t)^\kappa}\right)$ at the y-axis where $\ln p_{At} = 0$, and is linear in $\ln p_{At}$ with slope $0 < \kappa < 1$. From (34), (39), (40) and (41), $\ln\left(\frac{r_{At}}{1-r_{At}}\right)$ has a slope of strictly greater than one, and lies below $\ln\left(\frac{e_{At}}{1-e_{At}}\right)$ as $\ln p_{At} \to -\infty$ and above $\ln\left(\frac{e_{At}}{1-e_{At}}\right)$ as $\ln p_{At} \to +\infty$.

Given the unique equilibrium relative price of the agricultural good (p_{At}), the allocation of workers across locations and the good produced by each location are uniquely determined as shown above in part (b) of the proof above.

(d) Stability: To establish stability, we start from the unique equilibrium where $\tilde{V}_{it} = \tilde{V}_{kt}$ for all i, k in the population mobility condition (12) and reallocate a measure of workers from location k to location i such that the labor market clearing condition continues to hold:

$$\begin{aligned}\frac{d\tilde{V}_{it}}{dL_{it}}dL_{it} &= \left(\eta_j - \left(1-\mu_j\right) - \frac{1-\alpha}{\alpha}\right)\frac{\tilde{V}_{it}}{L_{it}}dL_{it} < 0, \\ \frac{d\tilde{V}_{kt}}{dL_{kt}}dL_{kt} &= \left(\eta_j - \left(1-\mu_j\right) - \frac{1-\alpha}{\alpha}\right)\frac{\tilde{V}_{kt}}{L_{kt}}dL_{kt} > 0,\end{aligned} \tag{42}$$

where $dL_{it} > 0$ and $dL_{kt} < 0$; $\eta_j < \left(1-\mu_j\right) + \frac{1-\alpha}{\alpha}$; and we have used the fact that each location is of measure zero relative to the economy as a whole and hence this reallocation leaves p_{At} unchanged. From (42), it follows that the unique equilibrium is stable.

(e) Other endogenous variables: Having determined $\{z_A(\theta_{Ait}), z_N(\theta_{Nit}), \Omega_{At}, \Omega_{Nt}, L_{Ait}, L_{Nit}, p_{At}\}$, all other endogenous variables of the model can be determined. Income per worker in each location is given by:

$$\begin{aligned}\pi_{it} &= \frac{1}{\alpha}p_{At}\Gamma_{At}\theta_{Ait}\left(L_{it}\right)^{\eta_A+\mu_A-1}\left(\chi_{YA}H_i\right)^{1-\mu_A}, \qquad i \in \Omega_{At}, \\ \pi_{kt} &= \frac{1}{\alpha}\Gamma_{Nt}\theta_{Nkt}\left(L_{kt}\right)^{\eta_N+\mu_N-1}\left(\chi_{YN}H_k\right)^{1-\mu_N}, \qquad k \in \Omega_{Nt}.\end{aligned}$$

The share of expenditure on agriculture is uniquely determined by p_{At}:

$$e_{Ait} = e_{At} = \frac{1}{1 + \left(\frac{1-a_t}{a_t}\right)^\kappa \left(\frac{1}{p_{At}}\right)^{1-\kappa}}$$

Therefore equilibrium consumption of the agricultural and non-agricultural good are:

$$c_{Ait} = \frac{e_{At}\alpha\pi_{it}}{p_{At}}$$

$$c_{Nit} = (1 - e_{At})\, \alpha \pi_{it}.$$

The equilibrium rental rate on land is:

$$r_{it} = \frac{\left[(1-\alpha) + \left(1-\mu_j\right)\alpha\right] \pi_{it} L_{it}}{H_i}.$$

Therefore equilibrium residential land use is:

$$H_{Uit} = \frac{(1-\alpha)\,\pi_{it}}{r_{it}}.$$

■

Proposition 2 *With mean reversion in agricultural productivity ($0 < \nu_A < \nu_N \leq 1$) and approximately constant proportional growth in non-agricultural productivity ($\nu_N \to 1$):* ***(a)*** *the dispersion of population density across non-agricultural locations is greater than the dispersion of population density across agricultural locations,* ***(b)*** *the most-densely-populated locations only produce the non-agricultural good,* ***(c)*** *some less-densely-populated locations produce the agricultural good and there is a range of population densities at which the share of agriculture in employment strictly decreases with population density.*

Proof. (a) To establish that the dispersion of population across non-agricultural locations is greater than the dispersion of population across agricultural locations as $\nu_N \to 1$, we first use population mobility (12) to solve for population density in each location:

$$\frac{L_{it}}{H_i} = \Lambda_{jt}^{\xi_j} \theta_{jit}^{\xi_j} H_i^{\xi_j \eta_j}, \tag{43}$$

$$\Lambda_{jt}^{\xi_j} = \left[\frac{p_{jt} \Gamma_{jt} \chi_{Yj}^{1-\mu_j}}{P_t V_t^{1/\alpha} \beta_j}\right]^{\xi_j},$$

$$\xi_j \equiv \frac{1}{\frac{1-\alpha}{\alpha} + \left(1-\mu_j\right) - \eta_j} > 0,$$

$$\beta_j \equiv \alpha \left[(1-\alpha) + \left(1-\mu_j\right)\alpha\right]^{\frac{1-\alpha}{\alpha}} > 0.$$

Note that $\Lambda_{jt}^{\xi_j}$ takes the same value across all locations producing the same good. The variance of log population density across agricultural locations is:

$$\mathbb{V}_A\left(\ln\left(\frac{L_{it}}{H_i}\right)\right) = \xi_A^2 \mathbb{V}\left(\ln \theta_{Ait}\right) + \left(\xi_A \eta_A\right)^2 \mathbb{V}\left(\ln H_i\right), \tag{44}$$

for $i \in \Omega_{At}$; $\mathbb{V}(\cdot)$ denotes a variance; the covariance term is zero because the idiosyncratic shocks to productivity ($\ln \phi_{jit}$) are independently and identically distributed for all $t \in \{0, \ldots, \infty\}$. The variance of log population density across non-agricultural locations is:

$$\mathbb{V}_N\left(\ln\left(\frac{L_{kt}}{H_k}\right)\right) = \xi_N^2 \mathbb{V}(\ln \theta_{Nkt}) + (\xi_N \eta_N)^2 \mathbb{V}(\ln H_k), \tag{45}$$

for $k \in \Omega_{Nt}$; where the covariance term is again zero. From (19) and (24):

$$\lim_{t\to\infty} \mathbb{V}[\ln \theta_{Ait}] = \frac{\sigma_{\phi A}^2}{1-\nu_A^2}, \qquad 0 < \nu_A < \nu_N \leq 1.$$

$$\lim_{t\to\infty} \mathbb{V}[\ln \theta_{Nit}] = \frac{\sigma_{\phi N}^2}{1-\nu_N^2}, \qquad 0 < \nu_N < 1.$$

$$\lim_{t\to\infty} \mathbb{V}[\ln \theta_{Nit}] = \infty, \qquad \nu_N = 1.$$

Since $\mathbb{V}(\ln H_i) < \infty$, it follows from (44) and (45) that:

$$\lim_{\nu_N\to 1}\left\{\lim_{t\to\infty} \mathbb{V}_N\left(\ln\left(\frac{L_{kt}}{H_k}\right)\right)\right\} > \lim_{t\to\infty} \mathbb{V}_A\left(\ln\left(\frac{L_{it}}{H_i}\right)\right).$$

(b) We now establish that the most-densely-populated locations only produce the non-agricultural good. From equilibrium population (43), the highest log population density for an agricultural location $i \in \Omega_{At}$ is:

$$\sup\left\{\ln\left(\frac{L_{it}}{H_i}\right)\right\} = \xi_A \ln \Lambda_{At} + \sup\left\{\xi_A \ln \theta_{Ait} + \xi_A \eta_A \ln H_i\right\}, \tag{46}$$

while the highest population density for a non-agricultural location $k \in \Omega_{Nt}$ is:

$$\sup\left\{\ln\left(\frac{L_{kt}}{H_k}\right)\right\} = \xi_N \ln \Lambda_{Nt} + \sup\left\{\xi_N \ln \theta_{Nkt} + \xi_N \eta_N \ln H_k\right\}, \tag{47}$$

From (21) and (26):

$$\ln \bar{\theta}_A = \lim_{t\to\infty} \sup\left\{\ln \theta_{Ait}\right\} = \frac{\ln \overline{\phi}_A}{1-\nu_A}, \qquad 0 < \nu_A < \nu_N \leq 1,$$

$$\ln \bar{\theta}_N = \lim_{t\to\infty} \sup\left\{\ln \theta_{Nit}\right\} = \frac{\ln \overline{\phi}_N}{1-\nu_N}, \qquad 0 < \nu_N < 1,$$

$$\ln \bar{\theta}_N = \lim_{t\to\infty} \sup\left\{\ln \theta_{Nit}\right\} = \infty, \qquad \nu_N = 1.$$

Since $\ln H_i < \infty$ for all i, it follows from (46) and (47) that:

$$\lim_{\nu_N\to 1}\left\{\lim_{t\to\infty} \sup\left\{\ln\left(\frac{L_{kt}}{H_k}\right) : k \in \Omega_{Nt}\right\}\right\} > \lim_{t\to\infty} \sup\left\{\ln\left(\frac{L_{it}}{H_i}\right) : i \in \Omega_{At}\right\}.$$

(c) We now establish that there is a range of population densities at which the share of agriculture in employment strictly decreases with population density. Define the set of locations with log population densities in the interval $\left[\ln\left(\frac{L_b}{H_b}\right), \ln\left(\frac{L_B}{H_B}\right)\right)$ by Ω_b^B. The share of agriculture in employment for this set of locations is:

$$\varpi_A^{\Omega_b^B} = \frac{\int_{i\in\left\{\Omega_{At}\cap\Omega_b^B\right\}} L_{Ait}di}{\int_{i\in\left\{\Omega_{At}\cap\Omega_b^B\right\}} L_{Ait}di + \int_{k\in\left\{\Omega_{Nt}\cap\Omega_b^B\right\}} L_{Nkt}dk}.$$

From the proof of part **(b)** above, the most-densely-populated locations only produce the non-agricultural good. Therefore, for sufficiently high values of $\ln\left(\frac{L_b}{H_b}\right)$ and $\ln\left(\frac{L_B}{H_B}\right)$, there exists an interval Ω_b^B for which $\int_{i\in\left\{\Omega_{At}\cap\Omega_b^B\right\}} L_{Ait}di = 0$, $\int_{k\in\left\{\Omega_{Nt}\cap\Omega_b^B\right\}} L_{Nkt}dk > 0$, and hence $\varpi_A^{\Omega_b^B} = 0$.

Now, from the proof of Proposition 1, recall that we established the existence of a unique equilibrium relative price of the agricultural good $0 < p_{At} < \infty$ at which $0 < e_{At} = r_{At} < 1$. Since $r_{At} > 0$ at this unique equilibrium value of p_{At}, $Y_{At} > 0$ and some less-densely-populated locations produce the agricultural good.

Since some less-densely-populated locations produce the agricultural good, and there exists an interval of sufficiently high population densities at which no location produces the agricultural good, it follows that there exists an interval of population densities at which the share of agriculture in employment is strictly decreasing in population density. ■

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

Proof. From goods market clearing (14), the share of the agricultural good in aggregate revenue (r_{At}) is equal in equilibrium to its share in aggregate expenditure (e_{At}):

$$r_{At} = e_{At}, \tag{48}$$

$$r_{At} = \frac{p_{At}Y_{At}}{p_{At}Y_{At} + Y_{Nt}},$$

$$e_{At} = \frac{1}{1 + \left(\frac{1-a_t}{a_t}\right)^{\kappa} p_{At}^{\kappa-1}},$$

$$Y_{At} = \Gamma_{At}\int_{\underline{\theta}_N}^{\bar{\theta}_N}\left[\int_{\theta_A^*(\theta_{Nit})}^{\bar{\theta}_A} \theta_{Ait} L_{it}^{\eta_A+\mu_A}\left(\chi_{YA}H_i\right)^{1-\mu_A} z_A\left(\theta_{Ait}\right)d\theta_{Ait}\right] z_N\left(\theta_{Nit}\right)d\theta_{Nit},$$

$$Y_{Nt} = \Gamma_{Nt} \int_{\underline{\theta}_A}^{\bar{\theta}_A} \left[\int_{\theta_N^*(\theta_{Akt})}^{\bar{\theta}_N} \theta_{Nkt} L_{kt}^{\eta_N + \mu_N} \left(\chi_{YN} H_k\right)^{1-\mu_N} z_N\left(\theta_{Nkt}\right) d\theta_{Nkt} \right] z_A\left(\theta_{Akt}\right) d\theta_{Akt}.$$

To determine the general equilibrium effect of changes in $\{\Gamma_{At},\, a_t\}$, we first differentiate the goods market clearing condition (48) holding constant the allocation of factors of production at their initial equilibrium values and solve for the implied change in the relative price of the agricultural good (p_{At}). Using this implied change in p_{At}, we next determine how the allocation of factors of production must change in response to the change in $\{\Gamma_{At},\, a_t\}$.

Define the transformed variable $b_t = \frac{1-a_t}{a_t}$, which is monotonically decreasing in a_t.

Note the following partial derivatives of r_{At} and e_{At}, where these partial derivatives are evaluated holding constant the initial allocation of factors of production $\{L_{it},\, L_{kt},\, \theta_A^*\left(\theta_{Nit}\right),\, \theta_N^*\left(\theta_{Akt}\right) : i \in \Omega_{At}, k \in \Omega_{Nt}\}$:

$$\begin{aligned}
\tfrac{\partial r_{At}}{\partial b_t}\tfrac{b_t}{r_{At}} &= 0, \\
\tfrac{\partial e_{At}}{\partial b_t}\tfrac{b_t}{e_{At}} &= -\kappa\left(1-e_{At}\right), \\
\tfrac{\partial r_{At}}{\partial \Gamma_{At}}\tfrac{\Gamma_{At}}{r_{At}} &= \left(1-r_{At}\right), \\
\tfrac{\partial e_{At}}{\partial \Gamma_{At}}\tfrac{\Gamma_{At}}{e_{At}} &= 0, \\
\tfrac{\partial r_{At}}{\partial p_{At}}\tfrac{p_{At}}{r_{At}} &= \left(1-r_{At}\right), \\
\tfrac{\partial e_{At}}{\partial p_{At}}\tfrac{p_{At}}{e_{At}} &= \left(1-\kappa\right)\left(1-e_{At}\right), \\
\text{where } b_t &= \frac{1-a_t}{a_t}.
\end{aligned} \tag{49}$$

(i) Γ_{At} : Differentiating the revenue and expenditure shares at the initial equilibrium holding constant the initial allocation of factors of production $\{L_{it},\, L_{kt},\, \theta_A^*\left(\theta_{Nit}\right),\, \theta_N^*\left(\theta_{Akt}\right) : i \in \Omega_{At}, k \in \Omega_{Nt}\}$, we have:

$$dr_{At} = \frac{\partial r_{At}}{\partial \Gamma_{At}} d\Gamma_{At} + \frac{\partial r_{At}}{\partial p_{At}}\frac{\partial p_{At}}{\partial \Gamma_{At}} d\Gamma_{At},$$

$$de_{At} = \frac{\partial e_{At}}{\partial p_{At}}\frac{\partial p_{At}}{\partial \Gamma_{At}} d\Gamma_{At}.$$

These equations can be re-written as follows:

$$\frac{dr_{At}}{r_{At}} = \left[\frac{\partial r_{At}}{\partial \Gamma_{At}}\frac{\Gamma_{At}}{r_{At}}\right]\frac{d\Gamma_{At}}{\Gamma_{At}} + \left[\frac{\partial r_{At}}{\partial p_{At}}\frac{p_{At}}{r_{At}}\right]\left[\frac{\partial p_{At}}{\partial \Gamma_{At}}\frac{\Gamma_{At}}{p_{At}}\right]\frac{d\Gamma_{At}}{\Gamma_{At}},$$

$$\frac{de_{At}}{e_{At}} = \left[\frac{\partial e_{At}}{\partial p_{At}}\frac{p_{At}}{e_{At}}\right]\left[\frac{\partial p_{At}}{\partial \Gamma_{At}}\frac{\Gamma_{At}}{p_{At}}\right]\frac{d\Gamma_{At}}{\Gamma_{At}}.$$

Imposing the goods market clearing condition before and after the change in Γ_{At}, we have:

$$\frac{dr_{At}}{r_{At}} = \frac{de_{At}}{e_{At}}.$$

Substituting for dr_{At}/r_{At} and de_{At}/e_{At} in the above expression, and using the partial derivatives (49), we obtain:

$$(1 - r_{At}) + (1 - r_{At}) \left[\frac{\partial p_{At}}{\partial \Gamma_{At}} \frac{\Gamma_{At}}{p_{At}} \right] = (1 - \kappa)(1 - e_{At}) \left[\frac{\partial p_{At}}{\partial \Gamma_{At}} \frac{\Gamma_{At}}{p_{At}} \right].$$

Since these derivatives are taken from the initial equilibrium, where $r_{At} = e_{At}$, we obtain:

$$\frac{\partial p_{At}}{\partial \Gamma_{At}} \frac{\Gamma_{At}}{p_{At}} = -\frac{1}{\kappa}.$$

Therefore, holding constant the allocation of factors of production at their initial equilibrium values, a rise in Γ_{At} implies a reduction in p_{At}. Furthermore, since $0 < \kappa < 1$, the implied reduction in p_{At} is more than proportionate to the rise in Γ_{At}.

We now examine the implications of such a reduction in p_{At} for the equilibrium allocation of factors of production. From the population mobility condition (12), we have:

$$\tilde{V}_{it} = \tilde{V}_{kt}$$

$$\frac{p_{At}\Gamma_{At}\theta_{Ait}\chi_{YA}^{1-\mu_A} H_i^{1-\mu_A+\frac{1-\alpha}{\alpha}} L_{it}^{\eta_A-(1-\mu_A)-\frac{1-\alpha}{\alpha}}}{\beta_A} = \frac{\Gamma_{Nt}\theta_{Nkt}\chi_{YN}^{1-\mu_N} H_k^{1-\mu_N+\frac{1-\alpha}{\alpha}} L_{kt}^{\eta_N-(1-\mu_N)-\frac{1-\alpha}{\alpha}}}{\beta_N} \tag{50}$$

for $i \in \Omega_{At}$, $k \in \Omega_{Nt}$. The combination of a rise in Γ_{At} and a more than proportionate reduction in p_{At} reduces normalized real income in agricultural locations on the left-hand side of (50) relative to normalized real income in non-agricultural locations on the right-hand side of (50). As a result, population is reallocated from agricultural to non-agricultural locations. Additionally, the combination of a rise in Γ_{At} and a more than proportionate reduction in p_{At} reduces the value marginal product of land in agriculture relative to that in non-agriculture, which induces some locations to switch from agricultural to non-agricultural land use. It follows that a rise in Γ_{At} reallocates employment from agriculture to non-agriculture.

(ii) a_t: To examine the comparative statics of a reduction in a_t, we consider an increase in the transformed variable $b_t = \frac{1-a_t}{a_t}$. Differentiating the revenue and expenditure shares at the initial equilibrium holding constant the initial allocation of factors of production $\{L_{it}$, L_{kt}, $\theta_A^*(\theta_{Nit})$, $\theta_N^*(\theta_{Akt}) : i \in \Omega_{At}, k \in \Omega_{Nt}\}$, we have:

$$dr_{At} = \frac{\partial r_{At}}{dp_{At}} \frac{\partial p_{At}}{\partial b_t} db_t,$$

$$de_{At} = \frac{\partial e_{At}}{\partial b_t} db_t + \frac{\partial e_{At}}{\partial p_{At}} \frac{\partial p_{At}}{\partial b_t} db_t.$$

These equations can be re-written as follows:

$$\frac{dr_{At}}{r_{At}} = \left[\frac{\partial r_{At}}{dp_{At}}\frac{p_{At}}{r_{At}}\right]\left[\frac{\partial p_{At}}{\partial b_t}\frac{b_t}{p_{At}}\right]\frac{db_t}{b_t},$$

$$\frac{de_{At}}{e_{At}} = \left[\frac{\partial e_{At}}{\partial b_t}\frac{b_t}{e_{At}}\right]\frac{db_t}{b_t} + \left[\frac{\partial e_{At}}{\partial p_{At}}\frac{p_{At}}{e_{At}}\right]\left[\frac{\partial p_{At}}{\partial b_t}\frac{b_t}{p_{At}}\right]\frac{db_t}{b_t}.$$

Imposing the goods market clearing condition before and after the change in b_t, $dr_{At}/r_{At} = de_{At}/e_{At}$, and using the partial derivatives (49), we obtain:

$$-\kappa\left(1-e_{At}\right)+\left(1-\kappa\right)\left(1-e_{At}\right)\left[\frac{\partial p_{At}}{\partial b_t}\frac{b_t}{p_{At}}\right] = \left(1-r_{At}\right)\left[\frac{\partial p_{At}}{\partial b_t}\frac{b_t}{p_{At}}\right].$$

Since these derivatives are taken from the initial equilibrium, where $r_{At} = e_{At}$, we obtain:

$$\frac{\partial p_{At}}{\partial b_t}\frac{b_t}{p_{At}} = -1.$$

Therefore, holding constant the allocation of factors of production at their initial equilibrium values, a rise in relative demand for non-agriculture, b_t, implies a reduction in the relative price of agriculture, p_{At}. From the population mobility condition (50), this reduction in p_{At} reallocates population from agricultural to non-agricultural locations. Additionally, the reduction in p_{At} reduces the value marginal product of land in agriculture relative to that in non-agriculture, which induces some locations to switch from agricultural to non-agricultural land use. It follows that a reduction in a_t (a rise in $b_t = \frac{1-a_t}{a_t}$) also reallocates employment from agriculture to non-agriculture.

■

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

Proof. **(a)** For locations that continue to produce the agricultural good in time t and $t-1$, equilibrium population (43) and the stochastic process for productivity (4) imply:

$$\Delta \ln L_{Ait} = \delta_{Ait} - \left(1-\nu_A\right)\ln\left(L_{Ait-1}/H_i\right) + u_{Ait}, \tag{51}$$

$$\delta_{Ait} = \xi_A \ln\left(\frac{\Lambda_{At}}{\Lambda_{At-1}^{\nu_A}}\right) + \xi_A\eta_A\left(1-\nu_A\right)\ln H_i,$$

$$u_{Ait} = \xi_A \ln \phi_{Ait},$$

where Δ denotes the difference operator and u_{Ait} is independently and identically distributed. From (51) and $0 < \nu_A < 1$, it follows that there is a decreasing relationship between population growth and initial population density for locations that continue to produce the agricultural good.

(b) For locations that continue to produce the non-agricultural locations at time t and $t-1$, equilibrium population (43) and the stochastic process for productivity (4) imply:

$$\Delta \ln L_{Nit} = \delta_{Nit} - (1-\nu_N) \ln (L_{Nit-1}/H_i) + u_{Nit}, \tag{52}$$

$$\delta_{Nit} = \xi_N \ln \left(\frac{\Lambda_{Nt}}{\Lambda_{Nt-1}^{\nu_N}} \right) + \xi_N \eta_N (1-\nu_N) \ln H_i,$$

$$u_{Nit} = \xi_N \ln \phi_{Nit},$$

where Δ again denotes the difference operator and u_{Nit} is independently and identically distributed. From (52), it follows that as $\nu_N \to 1$ population growth exhibits constant proportional growth for locations that continue to produce the non-agricultural good. ■

IV. Data Description

In Subsection IV.A., we report further details on the data sources and definitions for the U.S. MCD data for 1880, 1940 and 2000. In Subsection IV.B., we discuss the construction of comparable geographical units over time using the MCD data. In Subsection IV.C., we discuss the samples considered using the MCD data. In Subsection IV.D., we discuss the construction of the U.S. county sub-periods data for twenty-year intervals from 1880-2000.

IV.A. MCD Data Sources and Definitions

The main source of data on MCDs is the U.S. Census. We also use numerous other sources, including historical maps and gazetteers, as described below.

Data on MCD employment, population, land area and location in 2000 comes from the American Factfinder of the U.S. Census Bureau (Census 2000b). Agriculture, manufacturing and services are defined using the following sector classification: "agriculture" includes agriculture, forestry, fishing and hunting; "manufacturing" includes mining and construction as well as manufacturing; "services" include trade, transportation, warehousing, information, finance, insurance, real estate, professional, scientific, management, administrative, education, health, arts, entertainment, accommodation and food services; "non-agriculture" is defined as the sum of "manufacturing" and "services."

In addition to these 2000 data, we have data on population in 1940 and data on population and employment in 1880. The population information for 1940 comes from the 1940 Census Files (Census 1940), which also contain a full set of maps that allow us to identify the location of 1940 MCDs. However, our 1940 MCD data contain no employment information, which restricts their suitability for our analysis.

The 1880 population and employment data come from the North Atlantic Population Project (NAPP 2006). We use the 1950 occupation and industry classifications as provided by NAPP. We classify people for whom industry information is available into 3 categories: agriculture, manufacturing and services. Agricultural workers are those with 1950 industry codes $105-126$, which are mainly agriculture, forestry and fishing. Manufacturing workers are those with industry classifications $206-499$, and services include all other NAPP entries, except in the cases where the industry was illegible, missing, not reported, or not available.

Some people identified themselves as part of the labor force, but did not report their industry, or reported it in an illegible way, or were unclassifiable. In 1880 these amounted to about 15 percent of the workers classified above. In order to categorize these workers we use their self-reported occupations. If we classified most of the workers in a given occupation for which we did have industry information into, say, services, we also assigned all the workers in that occupation who did not report an industry to services. While this process may have introduced some error, for the vast majority of occupations one of the three sectors of agriculture, manufacturing and services accounts for a large majority of employment.

To determine the geographic location of MCDs in 1880 we used a variety of sources. For states for which 1880 maps of MCDs were available, we georeferenced those maps. For states for which 1880 maps of MCDs were not available, we started with the aforementioned 1940 maps and worked backwards through the microfilms for the 1930, 1920, 1910, 1900 and 1890 censuses, where changes to the names and organization of MCDs are documented in footnotes (NARA 2002, 2003). Finally, we supplemented this information with additional maps and gazetteers as reported in Appendix Table A.7 below.

The geographical control variables were created using maps from ESRI (1999). These geographical control variables are dummy variables equal to one if an MCD borders the ocean, if the distance between the centroid of an MCD and the closest river is less than 50 kilometers, if the distance between the centroid and the closest lake is less than 50 kilometers, and if the MCD contains coal.

IV.B. Linking 1880 and 1940 MCDs to 2000 MCDs

In matching MCDs from 1880, 1940, and 2000 we strove to cover all of the population and land area within each state in each of the three censuses, while consistently matching MCDs over time. This raised six challenges. First, some MCDs were renamed. Second, some MCDs merged over time. Third, in some areas county boundaries were redrawn, such that MCDs were reassigned to other counties. Fourth, in some areas the census did not provide sufficient geographical information. Fifth, some MCDs were split. Sixth, in some areas MCD boundaries were redefined.

In order to deal with these challenges, we aggregated some of the MCDs, and this process of aggregation required us to identify the geographic location of contemporary and historical MCDs. We started with a digital Geographic Information Systems (GIS) map from the Bureau of the Census of MCDs in 2000 (see Census 2000a,c,d). For the earlier censuses we assigned coordinates to the MCDs ourselves, using the 1940 MCD maps provided by the Bureau of the Census (Census 1940) and a variety of historical maps and gazetteers for 1880 (see Appendix Table A.7 below). Using these historical sources, we assigned geographic coordinates to MCDs in 1940 and 1880.[6]

To do so, we georeferenced the historical maps to the digital 2000 map using ArcGIS software. We then assigned the centroids manually in a point-shapefile. In total we assigned around 22,000 coordinates for 1880 and around 50,000 for 1940. In some states we were not able to assign coordinates, and these states are not divided into sub-county units in the final dataset. The geographical distribution of these states can be seen in Map 1 in the paper (these states are labeled as having "counties only" or "no data") or in Appendix Table A.6 below. For the other states we were able to determine the location of all MCDs in 1940 given the high quality of the maps provided by the census (Census 1940). For 1880 we were are able to determine the location of the vast majority of MCDs, with the main reason for unmatched 1880 MCDs being that the digital 1880 MCD data (NAPP 2006) contained

[6]When assigning the coordinates for 1940 we generally used the approximate geographic centroid of the MCD, except when the MCD was dominated by a single town. In this case we used the coordinates of the town. MCDs were categorized as being dominated by a single town if the census mentioned exactly one town within the MCD.

entries with missing names. Out of the approximately 22,000 MCDs listed by NAPP for 1880, only 150 MCDs remained unmatched (see below for further discussion).

We used the coordinates assigned to 1880 and 1940 MCDs to create geographic units that are stable over time and to which we could assign the data with reasonable confidence. To do so, we linked the 1880 and 1940 MCDs to the 2000 MCD in which their coordinates fell. In some cases multiple 1880 or 1940 MCDs fell into a single 2000 MCD, in which case we aggregated them into the single 2000 MCD.

We next proceeded in a number of steps. In a first step, we merged together 2000 MCDs in cases where they shared the same state, county and name. These were often cases where one MCD denoted a town and another denoted the surrounding area, and changes to the boundaries between the town and its surrounding area over time complicated the allocation of population to the two areas separately. This first step involved 1,163 aggregations.

In a second step, we aggregated some MCDs to the county level (using 1880 and 1940 county definitions) in counties for which we could not find all the 1880 MCDs on the map due to missing names. This second step involved the aggregation of 85 counties.

In a third step, we considered each 2000 MCD that had not been yet matched to at least one 1880 MCD and at least one 1940 MCD. These unmatched MCDs are referred to as "uncovered," while the other already-matched MCDs are referred to as "covered." Uncovered MCDs existed either if an older MCD was split (such that there were multiple MCDs in 2000 where there used to be one), or if boundaries were redrawn. In both cases we used proximity as a guide to solving the problem of uncovered MCDs. We determined the location of all 2000 centroids, and matched each uncovered MCDs to the closest 2000 MCD within the same county (using 1880 county definitions) that was covered.

Finally, we manually aggregated some additional units to deal with changes in municipal boundaries. We merged the MCDs of Bronx, Brooklyn, Manhattan, Queens, and Staten Island, since they are all parts of New York City. We also merged Saint Louis, Missouri, with its neighboring county, from which it split off at one point. Finally, we merged the MCDs Peoria and West Peoria in Illinois, which are part of the same urban area but not combined in the algorithm above.

IV.C. MCD Data Samples

This subsection discusses in further detail the samples used for our MCD data. Our baseline sample comprises "A and B" states (10,864 observations), and we also use a sample of "A" states (4,439 observations), a county sample (2,496 observations), and a hybrid sample (19,229 observations).

The "A and B" sample consists of states in which the ratios of the number of matched MCDs to the number of 1880 MCDs and 2000 MCDs are larger than 0.7. This restricts the extent to which we aggregate MCDs (a process that may involve imprecisions due to changes in boundaries), while maintaining a sizeable number of states. This sample consists of 15 states (plus Washington DC), most of which are found in the North-East and Mid-West of the U.S., as shown in Map 1 in the paper and listed in Appendix Table A.6 below.

The "A" sample is more restrictive: it only uses states for which the ratios of the number of matched MCDs to the number of 1880 MCDs and 2000 MCDs are larger than 0.9. These are the states in which there is a close correspondence between the 1880, 1940 and 2000 MCDs. This sample includes 8 states and Washington DC, and apart from Indiana and Iowa, all of these were part of the original 13 colonies.

To complement the MCD data, we use a sample of counties, which tackles the problem of representativeness by expanding the number of states that we use. The tradeoff is that in this sample we analyze data at a higher level of spatial aggregation. We exclude Alaska, Hawaii and Oklahoma, which were not included in the 1880 census. We also exclude North and South Dakota, which had not attained statehood in 1880, and did not have stable county boundaries at that time. For all other states, 1880 and 1940 counties are linked to 2000 counties using the centroids of the 1880 and 1940 counties.

The hybrid sample combines MCD and county data, and uses for each state the smallest unit for which we have data – MCDs for 30 states and counties for the remaining states in our sample.

A key advantage of the MCD data is their small spatial scale, which enables us to sharply draw the distinction between rural and urban areas that is central to our analysis. The average MCD in our baseline sample for the "A and B" states has an area of around 115 kilometers squared and a population of approximately 8,800 in 2000. In contrast, the average county in the "A and B" states has an area of around 1,500 kilometers squared and a population of approximately 115,000 in 2000. In Figure A.1 of this web appendix, we show that using MCDs rather than counties considerably enhances the density of observations for which we observe a range of initial shares of employment in agriculture in 1880.

IV.D. County Sub-Periods Data

We also construct a panel dataset on U.S. counties at twenty-year intervals from 1880-2000, which covers almost all of the area of the continental United States. We again exclude Alaska, Hawaii, Oklahoma, North Dakota and South Dakota, which had not attained statehood in 1880, and were either excluded from the 1880 Census or did not have stable county boundaries at that time. We also exclude Wyoming because of missing information in the GIS shapefiles used to create the county subperiods data. The data sources for this panel of counties are the Integrated Public Use Microdata Series (IPUMS), County Data Books from the Inter-University Consortium for Political and Social Research (ICPSR), National Historical Geographic Information System (NHGIS), and the American Factfinder of the U.S. Census Bureau.

For 1880, 1900 and 1920, we use IPUMS data, which are 1 percent samples from the U.S. Census microdata that are representative at the county level. For 2000, we use county-level data from American Factfinder. For 1940, 1960 and 1980, the IPUMS data do not include county identifiers, so we use instead county-level data from the County Data Books (ICPSR). While all three data sources (IPUMS, ICPSR, and American Factfinder) ultimately derive employment and population data from the U.S. Census, the aggregation of industries in the ICPSR data is coarser. When we decompose employment into sectors, we define "agriculture," "manufacturing" and "services" in the ICPSR and American Factfinder data in as consistent a way as possible with the definitions we use in the IPUMS data. Our county dataset also includes land area for 1880 (from ICPSR) and data from the Factfinder for 2000 on median house prices (variable H076001), and median wages (variable P068001). Finally, we include data from ICPSR on farm area and the nominal value of farm output from the agricultural census closest to the years included in our county sub-periods dataset, with the exception of 1920, for which no information on farm output is provided by ICPSR.

Over the 120-year period covered by our dataset, county boundaries change over time. In the 45 states (plus Washington DC) included in our sample, about 70 percent of 1880 counties have the same boundaries in 2000 (for the "A and B" states this figure is around 85 percent). Of the counties whose boundaries have changed, many were split, so the total number of counties across all states in our sample rises from 2,496 in 1880 to 2,984 in 2000. Apart from the splitting of counties, other challenges to matching counties over time are the renaming of some counties and some changes in county codes.

To construct geographic units that are stable, we follow a similar procedure as for our MCD data. Since we have GIS shapefiles for 1880 counties from NHGIS, and since most of the changes in boundaries involved splitting of counties, we use the 1880 counties as the base for matching counties over time. For each of the other years included in our dataset, we first construct the geographical centroid of each county in that year, again using GIS shapefiles from NHGIS. We next match each county centroid in each of the other years to the 1880 county in which it falls. Of the 1880 counties in our sample, 2,425 (about 97 percent) are matched to at least one county in each of the following years. We aggregate the remaining 71 unmatched 1880 counties with the nearest matched 1880 county within the same state based on the distance between the counties' centroids. From this procedure, we obtain a balanced panel of counties that is reasonably stable over time and covers all of the population and employment in each state in our sample in each year.

V. Baseline Empirical Results

V.A. Empirical Specification

No additional results required.

V.B. Stylized Facts

We begin by discussing in further detail the construction of the figures in the panels of Figure I in the paper. We next discuss a robustness test reported in the paper that uses initial log population size instead of initial log population density.

Panel A, Figure I in the paper: This figure shows the distribution of log population per square kilometer in 1880 and 2000 estimated using non-parametric specification (7) in the paper 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 greater than or equal to 0.1 and less than 0.2 are grouped together in bin 0.1.

Panel B, Figure I in the paper: The solid line shows the mean population growth rate from 1880-2000 within each population density bin based on estimating non-parametric specification (8) in the paper for the sample of "A and B" states. Population density bins are defined in the same way as in Panel A. 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.

Panel C, Figure I in the paper: The solid line shows the mean share of agriculture in 1880 employment within each population density bin based on estimating non-parametric specification (8) in the paper for the sample of "A and B" states. Population density bins and confidence intervals are constructed in the same way as in Panels A and B above. 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.

Panel D, Figure I in the paper: This figure shows the distribution of log agricultural employment and log non-agricultural employment (employment in manufacturing and services) per square kilometer in 1880 and 2000 estimated using non-parametric specification (7) in the paper for the sample of "A and B" states. Employment density bins are constructed in an analogous way as in Panels A and B above.

Panel E, Figure I in the paper: 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 (8) in the paper for the agricultural subsample (an agricultural share in 1880 employment of greater than 0.8) within "A and B" states. Population density bins and confidence intervals are constructed in the same way as in Panels A and B above. 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.

Panel F, Figure I in the paper: The solid line shows the mean growth rate of non-agricultural employment (employment in manufacturing and services) from 1880-2000 within each population density bin based on estimating non-parametric specification (8) in the paper for the non-agricultural subsample (an agricultural share in 1880 employment of less than 0.2) within "A and B" states. Population density bins and confidence intervals are constructed in the same way as in Panels A and B above. 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 A.2 in this web appendix: This figure displays the same specification as in Panel B in Figure I in the paper but uses initial log population size instead of initial log population density on the x-axis. We find a similar pattern of departures from Gibrat's Law of constant proportional growth: there is an initially decreasing, later increasing and ultimately roughly constant relationship between population growth and initial population size. This similarity of the results reflects the approximately log linear relationship between initial population size and initial population density in our data. Population density bins and confidence intervals are constructed in the same way as in Panels A and B of Figure I in the paper, as discussed above.

Table A.2: Using the nominal value of farm output per kilometer squared as a crude measure of agricultural productivity, we find evidence of mean reversion in agricultural productivity using our county sub-periods data. In Table A.2, we report the results of regressing log farm output per kilometer squared in a given year on log farm output per kilometer squared in a previous year.[7] Since the regression is run using the log level of productivity as the dependent variable (rather than productivity growth), a coefficient of less than one on lagged log productivity implies mean reversion in productivity. Each cell in the table corresponds to a separate regression across counties for a given pair of years. In general, we find estimated coefficients that are less than and statistically significantly different from one, with greater degrees of mean reversion occurring in earlier years. This pattern of results is consistent with the historical literature on the development of U.S. agriculture (see for example Cochrane 1979). As discussed by Olmstead and Rhode (2002) for the case of wheat, a number of the productivity-enhancing improvements in agricultural technology that occurred in the late nineteenth and early-twentieth centuries favored areas with poorer climate and soil. Since poor climate and soil are reflected in low initial levels of agricultural productivity, technological improvements that raise the relative productivity of areas with poorer climate and soil generate mean reversion in agricultural productivity.

V.C. Baseline Evidence on Structural Transformation

Results using the Regression prediction: Figure A.3 corresponds to Panel B of Figure II in the paper and displays results for population density, but uses the Regression prediction

[7] As discussed in Subsection IV.D. of this web appendix, data on farm output per kilometer squared are unavailable for 1920.

instead of the Employment Shares prediction. Again we find a close correspondence between the actual population distribution in 2000 and the predicted population distribution based on structural transformation away from agriculture. Figure A.4 corresponds to Panel C of Figure II in the paper and displays results for employment density in each industry, but uses the Regression prediction instead of the Employment Shares prediction. To generate the predicted employment density in 2000 for each sector using the Regression prediction, we estimate regression (10) in the paper for agricultural and non-agricultural employment growth separately. From the regression's fitted values, we obtain a prediction for employment growth in each sector j and MCD m from 1880-2000 ($\hat{g}_{Ejm}$). We next scale up observed 1880 MCD employment in each sector (E_{jm1880}) by this predicted employment growth rate $(1 + \hat{g}_{Ejm})$ to obtain predicted MCD 2000 employment for each sector ($\hat{E}_{jm2000} = (1 + \hat{g}_{Ejm})\, E_{jm1880}$) and hence predicted MCD 2000 employment density for each sector. As shown in Figure A.4, the predicted employment density distributions for each sector in 2000 are again close to the actual employment density distributions.

VI. Further Evidence

VI.A. Robustness

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

Employment Shares and Regression predictions disaggregating non-agriculture into manufacturing and services: Figure A.5 displays the shares of agriculture, manufacturing and services in aggregate U.S. employment for twenty-year intervals from 1880-2000 using the county sub-periods data discussed in Subsection IV.D. of this web appendix. While the left-hand panel shows results for our full sample of states (as discussed elsewhere, only Alaska, Hawaii, Oklahoma, North and South Dakota are excluded), the right-hand panel shows results for the "A and B" states. In both panels, the share of agriculture in aggregate U.S. employment declines rapidly until around 1960, after which it converges to less than two percent of aggregate employment. In contrast, the shares of manufacturing and services in aggregate employment increase rapidly alongside one another until around 1960, after which manufacturing's share declines and services's continues to rise.

In Figure A.6 in the web appendix, we display the employment-weighted average of log population density for each sector from 1880-2000 using the county sub-periods data discussed in Subsection IV.D. of this web appendix. Employment-weighted average log population density for each sector j in year t (e_{jt}) is calculated as follows:

$$e_{jt} = \sum_{s=1}^{S} \frac{E_{jst}}{E_{st}} \ln \left(\frac{L_{st}}{H_s} \right), \tag{53}$$

where E_{jst} denotes employment in sector j in county s at time t; E_{st} and L_{st} are total employment and population; H_s is land area.

The sector average population density measure (53) reveals the extent to which employment in a sector is concentrated in high, medium or low population density counties. Consistent with other studies such as Desmet and Rossi-Hansberg (2009), Figure A.6 shows that in recent decades manufacturing has dispersed to lower densities, while services has continued to concentrate at higher densities. Over a longer time horizon from 1880 until around 1980, we find that employment in both manufacturing and services shifted towards higher densities, with services displaying the larger change.

Figures A.7 and A.8 correspond to Panel A of Figure II in the paper and display results using the Employment Shares and Regression predictions for our baseline sample of MCDs in the A and B states, but they disaggregate non-agriculture into manufacturing and services. Total employment growth in each MCD can be decomposed into employment growth in agriculture, manufacturing and services, weighted by the initial shares of employment in each MCD and sector. Following a similar approach as in the baseline specification in the paper, our Employment Shares specification predicts MCD population growth using *aggregate* employment growth in agriculture. manufacturing and services for the U.S. as a whole and each *MCD's* own initial employment in each sector. We first scale up observed 1880 employment in MCD m in sector j (E_{jm1880}) by the aggregate employment growth rate for the sector from 1880-2000 ($1 + g_{Ej}$) to obtain predicted MCD 2000 employment in each sector ($\hat{E}_{jm2000}$). Summing the predicted values for agriculture, manufacturing and services gives predicted total MCD 2000 employment ($\hat{E}_{m2000}$). We next scale up predicted total MCD 2000 employment by the observed aggregate ratio of population to employment for the U.S. as a whole in 2000 (k_{2000}) to obtain predicted MCD 2000 population ($\hat{L}_{m2000}$). From predicted MCD 2000 population and observed MCD 1880 population (L_{m1880}), we obtain predicted population growth from 1880-2000 ($\hat{g}_{Lm}$):

$$\begin{aligned} \hat{E}_{jm2000} &= E_{jm1880}\left(1+g_{Ej}\right), \\ \hat{E}_{m2000} &= \hat{E}_{Am2000}+\hat{E}_{Mm2000}+\hat{E}_{Sm2000}, \\ \hat{L}_{m2000} &= k_{2000}\hat{E}_{m2000}, \\ \hat{g}_{Lm} &= \ln\left(\frac{\hat{L}_{m2000}}{L_{m1880}}\right), \end{aligned} \tag{54}$$

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

Our Regression prediction for population growth is constructed in a similar way as in the paper. We regress actual employment growth on the share of agriculture in employment in 1880, the share of manufacturing in employment in 1880, log population density in 1880, and interactions between the 1880 shares of agriculture and manufacturing in employment and 1880 log population density:

$$\begin{aligned} \Delta\ln L_{mt} &= a_0+a_1\frac{E_{Amt-T}}{E_{mt-T}}+a_2\frac{E_{Mmt-T}}{E_{mt-T}}+a_3\ln\left(\frac{L_{mt-T}}{H_m}\right) \\ &\quad +a_4\left(\frac{E_{Amt-T}}{E_{mt-T}}\times\ln\left(\frac{L_{mt-T}}{H_m}\right)\right)+a_5\left(\frac{E_{Mmt-T}}{E_{mt-T}}\times\ln\left(\frac{L_{mt-T}}{H_m}\right)\right)+u_{mt}, \end{aligned} \tag{55}$$

where L_{mt}/H_m denotes population density; $\{a_0,\ a_1,\ a_2,\ a_3,\ a_4,\ a_5\}$ are parameters that we estimate; the main effect of initial population density (a_3) allows for the possibility of mean reversion in services; the coefficient on the agricultural interaction term (a_4) allows the degree of mean reversion to differ between agriculture and services; the coefficient on the manufacturing interaction term (a_5) allows the degree of mean reversion to differ between manufacturing and services; u_{mt} is a stochastic error.

From the regression's fitted values, we obtain a prediction for the total employment growth rate for each MCD from 1880-2000. Our Regression specification scales up observed MCD 1880 total employment (E_{m1880}) by this predicted total employment growth rate ($1+\hat{g}_{Em}$) to obtain predicted MCD 2000 total employment ($\hat{E}_{m2000}=\left(1+\hat{g}_{Em}\right)E_{m1880}$). Using predicted MCD 2000 total employment, we generate predictions for MCD 2000 population and MCD population growth from 1880-2000 following the same method as for the Employment Shares specification (54).

As shown in Figures A.7 and A.8 in this web appendix, we find that disaggregating non-agriculture contributes relatively little to the ability of structural transformation to explain population growth from 1880-2000. Further insight into this result can be gained by comparing the Employment Shares predictions with and without disaggregating non-agriculture. Our baseline Employment Shares prediction in the paper inflates each MCD's non-agricultural employment in 1880 (E_{Nit-T}) by the aggregate employment growth rate for non-agriculture $(1 + g_N)$:

$$E_{Nit} = E_{Nit-T} \left(1 + g_N\right). \tag{56}$$

In contrast, our Employment Shares prediction disaggregating non-agriculture into manufacturing and services inflates each MCD's employment in each of these two industries in 1880 (E_{Mit-T} and E_{Sit-T}) by their aggregate employment growth rates ($1 + g_M$ and $1 + g_S$):

$$E_{Nit} = E_{Mit-T} \left(1 + g_M\right) + E_{Sit-T} \left(1 + g_S\right),$$

which can be re-written as:

$$E_{Nit} = E_{Nit-T} \left[\frac{E_{Mit-T}}{E_{Nit-T}} \left(1 + g_M\right) + \left(1 - \frac{E_{Mit-T}}{E_{Nit-T}}\right) \left(1 + g_S\right) \right]. \tag{57}$$

Comparing (56) and (57), it is evident that the Employment Shares prediction disaggregating non-agriculture into manufacturing and services generates different predictions for population growth across initial population densities from those of our baseline specification to the extent that: (a) aggregate employment growth rates differ between manufacturing and services and (b) the share of manufacturing in non-agricultural employment varies with initial population density.

The similarity of the population growth predictions with and without disaggregating non-agriculture in Figures A.7 and A.8 reflects two features of the data. First, the difference in aggregate employment growth rates from 1880-2000 between agriculture and non-agriculture (-0.016 versus 0.017) is much larger than between manufacturing and services (0.013 versus 0.018). Second, the variation in the shares of agriculture and non-agriculture in 1880 total employment across 1880 population density bins (from around 0.8 to 0.1) is much larger than the variation in the share of manufacturing and services in 1880 non-agricultural employment (from around 0.2 to 0.4).

In Figure A.9 in this web appendix, we use our county sub-periods dataset to compare our baseline Employment Shares predictions with those disaggregating manufacturing and services for the three sub-periods of 1880-1920, 1920-1960 and 1960-2000. Panels A-C display actual population growth and the two Employment Shares predictions. Panels D-F display the difference between the two Employment Shares predictions. In Panels A and B, the two Employment Shares predictions are almost visually indistinguishable from one another for 1880-1920 and 1920-1960. In contrast, in Panel C, we find more of difference between the two Employment Shares predictions from 1960-2000. This pattern of results is consistent with our findings in Figures A.5 and A.6 that manufacturing and services exhibit greater differences in the evolution of their shares of aggregate employment and their average population densities in recent decades. However, even for 1960-2000, the difference in aggregate employment growth rates between agriculture and non-agriculture is larger than that between manufacturing and services, and the variation in the share of manufacturing in non-agricultural employment across initial population density bins is smaller than the variation in the share of non-agriculture in total employment. As a result, the differences between the two Employment Shares predictions for 1960-2000 remain relatively small, as shown in Panels C and Panel F (and as reflected in the scale of Panel F).

Predictions controlling for the characteristics of neighboring MCDs: Figure A.10 corresponds to Panel A of Figure II in the paper, but displays the results of augmenting the Regression prediction to include information on the characteristics of the county of which an MCD is part. We augment the regression (10) in the paper with the 1880 share of agriculture in employment in the MCD's county, the 1880 log population density in the MCD's county and the interaction between these two variables. As shown in Figure A.10, the predictions for population growth including county information (labeled "Reg prediction, MCD & county characs") lie close to those without county information (labeled "Reg prediction"). Therefore, while the county variables are statistically significant, including information on the characteristics of the county of which an MCD is part contributes relatively little to the explanatory power of structural transformation away from agriculture.

Robustness of the stylized facts: Table A.1 and Panels A-F of Figures A.11 and A.12 report the results of a number of tests of the robustness of the stylized facts discussed in the paper. Each panel of Figures A.11 and A.12 corresponds to Panel B of Figure I in the paper and shows the results of estimating the non-parametric specification (8) in the paper for different samples.

Table A.1: Column (1) of Table A.1 replicates Column (1) of Table I in the paper by reporting results for our baseline sample of "A and B" states. Column (2) of Table A.1 reports results based on aggregating MCDs in the "A and B" states that lie within the boundaries of each of our 1880 metropolitan areas. As discussed in the paper, 1880 metropolitan areas are constructed using a definition of a "city" as an MCD with a log population density in 1880 of greater than 6 and aggregating all MCDs within 25 kilometers of the city. When two or more cities and their surrounding areas overlap, we merge them together. Column (3) of Table A.1 presents results when we instead drop from the sample all MCDs in the "A and B" states that lie within the boundaries of an 1880 metropolitan area as defined above. Column (4) of Table A.1 reports results when we drop from the sample all MCDs in the "A and B" states with centroids within 100 kilometers of the centroid of a 2000 Metropolitan Statistical Area (MSA). Column (5) of Table A.1 presents results when we drop all MCDs in the "A and B" states with centroids within 100 kilometers of the centroid of an 1880 metropolitan area as defined above. Column (6) of Table A.1 displays results when we drop all MCDs in the "A and B" states that experience a decline in non-agricultural employment between 1880 and 2000. Across each of the columns in Table A.1, we find the same pattern of stylized facts as discussed in the paper.

Figure A.11: Panel A reports results for the sample of MCDs in the "A" states. Panel B presents results for the counties sample discussed in Subsection IV.C. of this web appendix. Panel C displays results for the subset of the counties sample in the "A and B" states. Panel D reports results for the hybrid sample of MCDs and counties discussed in Subsection IV.C. of this web appendix. Panel E presents results for the subset of MCDs in the "A and B" states that were part of British colonial claims in 1775. Panel F displays results aggregating MCDs in the "A and B" states that lie within the boundaries of each 2000 Metropolitan Statistical Area (MSA).

Figure A.12: Panel A reports results based on aggregating MCDs in the "A and B" states that lie within the boundaries of an 1880 metropolitan area. As discussed in the paper, 1880 metropolitan areas are constructed using a definition of a "city" as an MCD with a log population density in 1880 of greater than six and aggregating all MCDs within 25 kilometers of the city. When two or more cities and their surrounding areas overlap, we merge them together. Panel B presents results when we drop from the sample all MCDs in the "A and B" states that lie within the boundaries of a 2000 Metropolitan Statistical Area (MSA). Panel C displays results when we drop from the sample all MCDs in the "A and B" states that lie within the boundaries of an 1880 metropolitan area as defined above. Panel D reports results when we drop from the sample all MCDs in the "A and B" states with centroids within 100 kilometers of the centroid of a 2000 MSA. Panel E presents results when we drop from the sample all MCDs in the "A and B" states with centroids within 100 kilometers of the centroid of an 1880 metropolitan area as defined above. Panel F displays results when we drop all MCDs that experience a decline in non-agricultural employment between 1880 and 2000 from our baseline sample for the "A and B" states.

Across each of the specifications shown in Panels A-F of Figures A.11 and A.12, we find the same pattern of results as shown in Panel of B of Figure I in the paper. The pattern of an initially decreasing, later increasing and finally roughly constant relationship between population growth and initial population density is therefore robust to the consideration of each of these alternative samples and specifications.

VI.B. Timing of Structural Transformation

Figure A.13: This figure shows results using the county sub-periods data discussed in Subsection IV.D. of this web appendix for 1880-1920, 1920-1960 and 1960-2000. In Panels A-C of Figure A.13, we display mean population growth for each sub-period across initial log population density bins. In each panel, we show both mean actual population growth and mean predicted population growth based on the Employment Shares prediction. In Panels D-F of Figure A.13, we display the mean share of agriculture in employment for 1880, 1920 and 1960 respectively across initial log population density bins. Initial log population density bins are constructed in the same way as in Panels A and B of Figure I in the paper. To highlight the contrast between the sub-periods, we focus in Figure A.13 on initial population densities between 0 and 6 log points, but find a similar pattern of results across the full range of initial population densities.

In Panels A-B of Figure A.13, the increasing relationship between population growth and initial population density over a range of intermediate densities is strongly apparent for the first two sub-periods. Furthermore, this range of intermediate densities is the same range over which there is a sharp decline in the initial share of agriculture in employment in 1880 and 1920 in Panels D-E. As a result, the Employment Shares prediction captures the increasing relationship between population growth and initial population density observed at intermediate densities, confirming the explanatory power of structural transformation away from agriculture in predicting population growth.

In contrast, by 1960 agriculture is a small share of employment across all initial log population density bins in Panel F, so that the decline in the share of agriculture over a range intermediate densities is much more muted. This dominance of non-agriculture across the full range of initial population densities is accompanied by a largely constant rate of actual population growth in Panel C, which is well captured by the largely constant rate of predicted population growth based on structural transformation away from agriculture.

Figure A.14: This figure corresponds to Panel D of Figure III in the paper, but shows results using the Regression prediction instead of the Employment Shares prediction. Both figures are based on the county sub-periods data discussed in Subsection IV.D. of this web appendix. For each Census region and twenty-year period, Figure A.14 displays the difference in mean population growth for the ranges of 3-5 minus 1-3 log points of initial population density for both actual population growth (y-axis) and predicted population growth based on the Regression prediction (x-axis). Points are labeled according to Census region codes: MW (Mid-West), NE (North-East), S (South) and W (West). Washington D.C. is assigned to the South. Points are also labeled according to the final year of the interval over which population growth is computed, so that 1960 corresponds to the sub-period 1940-1960.

As for the Employment Shares prediction in the paper, we observe a strong positive relationship between the actual increase in population growth at intermediate densities and the predicted increase based on structural transformation away from agriculture. Regressing the actual increase in population growth between the two ranges (3-5 minus 1-3) on the predicted increase, we find a positive and statistically significant coefficient (standard error) of 0.619 (0.162), as shown in the regression line in Figure A.14. Augmenting the regression with region and sub-period fixed effects, we continue to find a positive and significant relationship, with a coefficient (standard error) of 1.167 (0.163).

VI.C. Alternative Potential Explanations

Table A.3: In Table IV in the paper, we report heteroscedasticity robust standard errors adjusted for clustering by county, which allows the standard errors to be correlated across MCDs within counties without imposing prior structure on the pattern of this correlation. In Table A.3, we report the results of a robustness test in which we instead use the alternative approach to allowing for spatial correlation in the regression errors of Bester et al. (2011). Both procedures result in similar standard errors, so that all statements about statistical significance are robust to the use of either approach.

VII. Conclusion

No additional results required.

VIII. Brazilian Evidence

VIII.A. Brazilian Data

Our Brazilian data mostly come from Instituto de Pesquisa Econômica Aplicada (IPEA), and Brazilian Census micro data compiled by Instituto Brasileiro de Geografia e Estatística (IBGE). 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.[8]

Data on AMC employment and population in both 1970 and 2000 comes from the Brazilian Census micro data (Brazil Census 1970, 2000). Data on AMC land area in 2000 comes from IPEA (2008). 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. Since 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, our sample period contains considerable structural transformation.

The average Brazilian AMC spans 2,323km^2 and had 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 subsample that includes the Northeast, Southeast and South regions in Brazil only. AMCs in these regions are relatively small, which permits a sharper distinction between rural and urban areas, and it is less likely that these regions were not fully settled in 1970. The average AMC in this subsample spans 923km^2 and had a population of 26,013 in 1970 and 44,125 in 2000. The three regions in

[8] New municipalities were created after 2000, but the 1997 municipalities were used in the 2000 Census, the latest Census that we analyze in this web appendix.

this subsample cover about 90 percent of Brazil's AMCs, 36 percent of its land area and 91 percent of its population in 1970.

While obtaining population data is straightforward, calculating sectoral employment involved some choices in the classification of workers into agriculture, manufacturing, and services. In particular, in the U.S. the logging sector is not considered part of agriculture, but in Brazil it proved more difficult to consistently separate logging from the rest of the agricultural sector for both 1970 and 2000. We therefore pooled the Brazilian logging industry with its agricultural sector. Otherwise the definitions of "agriculture," "manufacturing" and "services" for Brazil closely follow those used for the U.S.. Using 1970 sector definitions, we classified people employed in sectors 111-222 as agricultural workers, those in sectors 300-352 as industrial workers, and those in sectors 411-928 as service workers. In 2000, agricultural workers were those with sector classifications 01101-05002, manufacturing workers were those with sector classifications 10000-37000 or 45001-45999, and services workers were those with sector classifications 40010-41000 or 50010-93092.

In some of the robustness checks we also use a set of state fixed effects. To generate these fixed effects we use the 2000 classification of 27 Brazilian states, since some Brazilian state boundaries did change after 1970. In particular, Mato Grosso do Sul was separated from Mato Grosso in the 1970s; Guanabara and Rio de Janeiro merged in 1975 under the name of Rio de Janeiro; and Tocantins was formed in 1988 out of the northern part of Goiás.

In addition to state fixed effects, some of our specifications also use a range of geographic controls. These include indicators for mineral deposits of oil, nickel, manganese, iron, gold, copper, cobalt, and aluminum (bauxite). We also construct an indicator for whether an AMC borders on the ocean, or whether its centroid lies within 50 kilometers of a river. Finally, we construct a variable indicating whether an AMC's centroid is covered with tropical or subtropical moist broadleaf forest, or for if it is situated in the Amazonas area. The river shapefile is from ArcView Database Access (ESRI 1999). The broadleaf forest, minerals, and oil and gas data are from the Global GIS DVD (GIS 2003).

VIII.B. Brazilian Results

Table A.4 and Figure A.15: This table and figure report the results of our robustness test using the Brazilian data. In Column (1) of Table A.4 and in Figure A.15, we report results for all Brazilian AMCs. Panel A of the table and figure show that the standard deviation of log population density across Brazilian AMCs increased from 1970-2000, confirming our first stylized fact. Panel B of the table and figure show that low density areas and high density areas grew faster than areas of intermediate density. Therefore the pattern of an initially decreasing, later increasing and finally roughly constant relationship between population growth and initial population density, characterized in Stylized Fact 2, also holds for Brazil.

Panel C in Table A.4 and Figure A.15 show that the increasing segment of the 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). Panel D in the same table and figure also confirm 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 largely uncorrelated with initial density – are also confirmed for Brazil, as shown in the final two panels of the table and figure.[9]

While our baseline sample for Brazil includes all Brazilian AMCs, in Column (2) of Table A.4 we report results for the subsample of AMCs in the Northeast, Southeast and South regions of Brazil. As discussed in the previous subsection, AMCs in this subsample are somewhat smaller in geographical area, which facilitates a sharper distinction between rural and urban areas, and it is less likely that these regions were not fully settled in 1970. As shown in Column (2) of Table A.4, all of our stylized facts are confirmed for this subsample of AMCs.

Figure A.16: This figure corresponds to Panel A of Figure II in the paper, but displays actual and predicted population growth for Brazil instead of the U.S.. The Employment Shares and Regression predictions for Brazil are constructed in the same way as discussed in Subsection V.C of the paper. As for the U.S., we find that structural transformation away from agriculture has substantial explanatory power for population growth in Brazil. Regressing mean actual population growth on mean predicted population growth across the initial log population density bins shown in Figure A.16, we find a coefficient (standard error) of 0.498 (0.072) and a regression R^2 of 0.40 using the Employment Shares prediction. Results using the Regression prediction are similar: we find a coefficient (standard error) of 0.596 (0.061) and a regression R^2 of 0.63.

[9] 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.

Table A.5: This table corresponds to Table IV in the paper, but reports the results of regressions of actual on predicted population growth for Brazil rather than for the U.S.. Panel A uses the Employment Shares prediction, which is based on aggregate reallocation from agriculture to non-agriculture and each MCD's initial employment in each sector. Panel B uses the Regression prediction, which allows for different degrees of mean reversion in agriculture and non-agriculture.

In Column (1), we report the results of a regression of actual on predicted population growth with no controls. In Column (2), we augment the specification from Column (1) with controls for local differences in physical geography and natural endowments, including dummy variables for (a) the presence of oil, nickel, manganese, iron, gold, copper, cobalt, and aluminum (bauxite), (b) whether the AMC borders the ocean or lies within 50 kilometers of a river, (c) whether the AMC has a centroid covered with tropical or subtropical moist broadleaf forest or is located in the Amazonas area. In Column (3), we augment the specification from Column (1) with state fixed effects. In Column (4), we augment the specification from Column (1) with a full set of fixed effects for initial log population density bins. In Column (5), we re-estimate Column (1) for the sub-sample of AMCs in the North-East, South-east and South regions of Brazil. In Column (6), we simultaneously include all of the controls from Columns (2)-(4). In Column (7) we report results for the sub-sample of AMCs in the North-East, South-east and South regions of Brazil, simultaneously including all of the controls from Columns (2)-(4). Across each of these specifications and in both panels, we continue to find statistically significant effects of structural transformation away from agriculture.

In summary, we find a strikingly similar pattern of results for Brazil and the U.S.. While there are many differences between the U.S. from 1880-2000 and Brazil from 1970-2000, both are characterized by substantial structural transformation away from agriculture, and hence we would expect the stylized facts to apply. The striking similarity of the results in these two different contexts provides strong evidence that our findings are indeed capturing structural transformation away from agriculture and are not driven by idiosyncratic features of the data or institutional environment for the U.S..

References

Bester, C. Alan, Timothy G. Conley, and Christian B. Hansen (2011) "Inference with Dependent Data Using Cluster Covariance Estimators," *Journal of Econometrics*, 165(2), 137-151.

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

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.

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

Davis, Morris A. and François Ortalo-Magne (2011) "Household Expenditures, Wages, Rents," *Review of Economic Dynamics*, 14(2), 248-261.

Desmet, Klaus and Esteban Rossi-Hansberg (2009) "Spatial Growth and Industry Age," *Journal of Economic Theory*, 144(6), 2477-2502.

Duranton, Gilles and Diego Puga (2004) "Micro-foundations of Urban Agglomeration Economies," Chapter 48 in (eds.) J. V. Henderson and J.F. Thisse, *Handbook of Regional and Urban Economics*, Volume 4, Amsterdam: Elsevier North Holland.

Eaton, Jonathan and Samuel S. Kortum (2002) "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741-1779.

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

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.

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

Cambridge University Press, 33-54.

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

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

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]

NARA (2002) Descriptions of Census Enumeration Districts, 1830-1950, Microfilm Publication T1224, Washington, D.C.: National Archives and Records Administration, 2002.

NARA (2003) Enumeration District Maps for the Twelfth Through Sixteenth Censuses of the United States, 1900–1940, Microfilm Publication A3378, compiled by Claire Prechtel-Kluskens, Washington, DC : National Archives and Records Administration, 2003.

Ngai, Rachel and Chris Pissarides (2007) "Structural Change in a Multisector Model of Growth," *American Economic Review*, 97(1), 429-443.

Olmstead, Alan L. and Paul W. Rhode (2002) "The Red Queen and the Hard Reds: Productivity Growth in American Wheat, 1800-1940," *Journal of Economic History*, 62(4), 929-966.

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.

Table A.1: U.S. stylized facts, further robustness

		(1) MCDs Baseline: A and B states	(2) MCDs Pooling 1880 cities	(3) MCDs Excluding 1880 cities	(4) MCDs Excluding 2000 MSAs +100 km	(5) MCDs Excluding 1880 city +100 km	(6) MCDs No shrinking non-ag employment
Panel A	Standard deviation of log population density in 1880 (σ_1)	0.967	0.927	0.927	0.930	0.940	0.957
	Standard deviation of log population density in 2000 (σ_2)	1.556	1.457	1.457	1.234	1.402	1.526
	H_0: $\sigma_1 = \sigma_2$, vs. H_1: $\sigma_1 < \sigma_2$, p-value	0.000	0.000	0.000	0.000	0.000	0.000
	<u>Stylized Fact 1</u>: Distribution of log population density across geographic units became more dispersed from 1880-2000 (population became more concentrated)	Yes	Yes	Yes	Yes	Yes	Yes
Panel B	Mean population growth at log population density 0 ($\beta_g(0)$)	0.013	0.013	0.013	0.013	0.012	0.013
	Mean population growth at log population density 2 ($\beta_g(2)$)	0.001	0.001	0.001	0.000	0.001	0.002
	Mean population growth at log population density 4 ($\beta_g(4)$)	0.009	0.008	0.008	0.005	0.007	0.009
	Mean population growth at log population density larger 4 ($\beta_g(>4)$)	0.010	0.009	0.009	0.005	0.009	0.010
	H_0: $\beta_g(0) = \beta_g(2)$, H_1: $\beta_g(0) > \beta_g(2)$, p-value	0.000	0.000	0.000	0.000	0.000	0.000
	H_0: $\beta_g(2) = \beta_g(4)$, H_1: $\beta_g(2) < \beta_g(4)$, p-value	0.000	0.000	0.000	0.000	0.000	0.000
	H_0: $\beta_g(4) = \beta_g(>4)$, H_1: $\beta_g(>4) \neq \beta_g(4)$, p-value	0.489	0.470	0.470	0.739	0.248	0.666
	<u>Stylized Fact 2</u>: Increasing relationship between population growth from 1880-2000 and log population density in 1880 at intermediate densities	Yes	Yes	Yes	Yes	Yes	Yes
Panel C	Percent of agricultural in total employment at log population density 2 ($\beta_{sa}(2)$)	0.767	0.769	0.769	0.771	0.775	0.771
	Percent of agricultural in total employment at log population density 4 ($\beta_{sa}(4)$)	0.228	0.220	0.220	0.213	0.219	0.230
	H_0: $\beta_{sa}(2) = \beta_{sa}(4)$, H_1: $\beta_{sa}(2) > \beta_{sa}(4)$, p-value	0.000	0.000	0.000	0.000	0.000	0.000
	<u>Stylized Fact 3</u>: 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
Panel D	Standard deviation of agricultural employment in 1880 (σ_{1a})	0.820	0.820	0.820	0.896	0.901	0.802
	Standard deviation of non-agricultural employment in 1880 (σ_{1na})	1.520	1.437	1.437	1.431	1.457	1.518
	H_0: $\sigma_{1a} = \sigma_{1na}$, vs. H_1: $\sigma_{1a} < \sigma_{1na}$, p-value	0.000	0.000	0.000	0.000	0.000	0.000
	Standard deviation of agricultural employment in 2000 (σ_{2a})	0.858	0.851	0.851	0.868	0.881	0.845
	Standard deviation of non-agricultural employment in 2000 (σ_{2na})	1.623	1.509	1.509	1.293	1.468	1.591
	H_0: $\sigma_{2a} = \sigma_{2na}$, vs. H_1: $\sigma_{2a} < \sigma_{2na}$, p-value	0.000	0.000	0.000	0.000	0.000	0.000
	<u>Stylized Fact 4</u>: Standard deviation of non-agricultural employment is larger than standard deviation of agricultural employment in both years	Yes	Yes	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 1880, report slope coefficient (β_a)	-0.0060	-0.0061	-0.0061	-0.0060	-0.0061	-0.0060
	H_0: $\beta_a = 0$, H_1: $\beta_a \neq 0$, p-value	0.000	0.000	0.000	0.000	0.000	0.000
	<u>Stylized Fact 5</u>: Agricultural employment growth is negatively correlated with population density	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 (β_{na})	-0.0002	0.0001	0.0001	-0.0005	-0.0018	-0.0008
	H_0: $\beta_{na} = \beta_a$, H_1: $\beta_{na} > \beta_a$, p-value	0.000	0.000	0.000	0.000	0.000	0.000
	<u>Stylized Fact 6</u>: Stronger mean reversion in agricultural employment than in non-agricultural employment	Yes	Yes	Yes	Yes	Yes	Yes
	<u>Number of observations</u>	10,864	10,173	10,133	6,386	5,798	10,586

Note: This table reports robustness tests of our 6 stylized facts using US MCD data. Panels B, C, E, and F report tests based on regressions using robust standard errors clustered by county. See the text of the paper and this web appendix for further discussion of the construction of the data.

Table A.2: Correlation of farm output per kilometer squared over time

	(1) log 1900 farm output per km2	(2) log 1940 farm output per km2	(3) log 1960 farm output per km2	(4) log 1980 farm output per km2	(5) log 2000 farm output per km2
Panel A: Full sample of counties					
log 1880 farm output per km2	0.561*** (0.0300)	0.346*** (0.0231)	0.317*** (0.0241)	0.226*** (0.0299)	0.196*** (0.0262)
log 1900 farm. output per km2		0.617*** (0.0233)	0.555*** (0.0273)	0.438*** (0.0378)	0.419*** (0.0344)
log 1940 farm output per km2			0.905*** (0.0182)	0.823*** (0.0270)	0.846*** (0.0298)
log 1960 farm. output per km2				0.922*** (0.0171)	0.932*** (0.0228)
log 1980 farm output per km2					1.028*** (0.0156)
Panel B: Counties in the A and B states					
log 1880 farm output per km2	0.749*** (0.0441)	0.704*** (0.0482)	0.724*** (0.0504)	0.626*** (0.0672)	0.520*** (0.0593)
log 1900 farm output per km2		0.972*** (0.0368)	1.001*** (0.0472)	0.923*** (0.0742)	0.806*** (0.0560)
log 1940 farm output per km2			0.973*** (0.0335)	0.902*** (0.0473)	0.875*** (0.0438)
log 1960 farm output per km2				0.923*** (0.0338)	0.879*** (0.0421)
log 1980 farm output per km2					0.998*** (0.0319)

Note: Each cell reports the estimated coefficient from a regression of log nominal farm output per kilometer squared (measured in the year denoted in the column headings) on lagged log nominal farm output per kilometer squared (measured in the year denoted in the row headings). Each cell corresponds to a separate regression using observations on counties. Each specification includes state fixed effects, which control, for example, for state-specific changes in nominal prices for the relevant pair of years. The table uses our county sub-periods data. Robust standard errors are in parentheses.

Table A.3: Explanatory power of the Employment Shares and Regression predictions (alternative standard errors)

Actual population growth	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		As (1) with geographic controls	As (1) with demographic controls	As (1) with distance to 1947 highway	As (1) with distance to 1898 railroad	As (1) with distance to 1880 city	As (1) with manufacturing share in non-agriculture	As (1) with 1880 density bin controls	As (1) with county fixed effects	All controls used in (2) to (9)
Panel A : Employment Shares prediction										
Predicted Population Growth	0.449***	0.334***	0.262***	0.401***	0.400***	0.413***	0.418***	0.324***	0.271***	0.233***
	(0.031)	(0.0269)	(0.026)	(0.029)	(0.029)	(0.027)	(0.028)	(0.030)	(0.016)	(0.019)
Observations	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864
R-squared	0.09	0.18	0.14	0.14	0.14	0.09	0.10	0.15	0.62	0.67
Panel B : Regression prediction										
Predicted Population Growth	0.978***	0.718***	0.581***	0.905***	0.903***	0.892***	0.913***	0.645***	0.545***	0.362***
	(0.059)	(0.056)	(0.070)	(0.057)	(0.057)	(0.054)	(0.058)	(0.066)	(0.041)	(0.045)
Observations	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864	10,864
R-squared	0.10	0.18	0.15	0.15	0.16	0.20	0.10	0.15	0.62	0.66

Note: All regressions use MCD data for our baseline sample of A and B states. Observations are a cross-section of MCDs from 1880-2000. In all specifications, the dependent variable is actual population growth from 1880-2000. Panel A uses predicted population growth from the Employment Shares prediction. Panel B uses predicted population growth from the Regression prediction. In Column (2), the geographical controls are measures of proximity to rivers, lakes, coastlines and mineral resources. In Column (3), the demographic controls are the share of the population that is white, the share of the population born outside the state (as a measure of national and international migration), the share of the population aged less than six (as a measure of fertility), and the share of the population aged 14-18 in education (as a measure of educational attainment). Column (4) includes distance from the centroid of each MCD to the closest interstate highway in the 1947 plan. Column (5) includes distance from the centroid of each MCD to the closest railroad in the 1898 railroad network. Column (6) includes distance from the centroid of each MCD to the centroid of the closest 2000 MSA. In Column (7), we include the 1880 share of manufacturing in non-agricultural employment. Column (8) includes 1880 population density bin fixed effects. Column (9) includes county fixed effects. In Column (10), we include all controls from Columns (2)-(9). In Columns (4)-(6) and (10), we use log(1+distance to transportation system and/or city). Robust standard errors corrected for spatial correlation using the methodology of Bester, Conley and Hansen (2009). See the text of the paper and this web appendix for further discussion of the construction of the data.

Table A.4: Brazil stylized facts and their robustness

		(1) All of Brazil (AMCs)	(2) Brazil sub-sample (see table footnote)
Panel A	Standard deviation of log population density in 1970 (σ_1)	1.222	1.009
	Standard deviation of log population density in 2000 (σ_2)	1.323	1.197
	H_0: $\sigma_1 = \sigma_2$, vs. H_1: $\sigma_1 < \sigma_2$, p-value	<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
Panel B	Mean population growth at log population density 0 ($\beta_g(0)$)	0.024	0.015
	Mean population growth at log population density 4 ($\beta_g(4)$)	0.008	0.008
	Mean population growth at log population density 6 ($\beta_g(6)$)	0.021	0.021
	Mean population growth at log population density larger 6 ($\beta_g(>6)$)	0.032	0.032
	H_0: $\beta_g(0) = \beta_g(4)$, H_1: $\beta_g(0) > \beta_g(4)$, p-value	<0.001	<0.001
	H_0: $\beta_g(4) = \beta_g(6)$, H_1: $\beta_g(4) < \beta_g(6)$, p-value	<0.001	<0.001
	H_0: $\beta_g(6) = \beta_g(>6)$, H_1: $\beta_g(>6) \neq \beta_g(6)$, p-value	0.016	0.016
	Stylized Fact 2: Upward and downward sloping relationship between population growth from 1970-2000 and log population density in 1970	Yes	Yes
Panel C	Percent of agricultural in total employment in 1970 at log population density 4 ($\beta_{sa}(4)$)	0.671	0.671
	Percent of agricultural in total employment in 1970 at log population density 6 ($\beta_{sa}(6)$)	0.168	0.168
	H_0: $\beta_{sa}(4) = \beta_{sa}(6)$, H_1: $\beta_{sa}(4) > \beta_{sa}(6)$, p-value	<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
Panel D	Standard deviation of agricultural employment in 1970 (σ_{1a})	0.893	0.887
	Standard deviation of non-agricultural employment in 1970 (σ_{1na})	1.416	1.429
	H_0: $\sigma_{1a} = \sigma_{1na}$, vs. H_1: $\sigma_{1a} < \sigma_{1na}$, p-value	<0.001	<0.001
	Standard deviation of agricultural employment in 2000 (σ_{2a})	1.018	0.995
	Standard deviation of non-agricultural employment in 2000 (σ_{2na})	1.375	1.364
	H_0: $\sigma_{2a} = \sigma_{2na}$, vs. H_1: $\sigma_{2a} < \sigma_{2na}$, p-value	<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
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 (β_a)	-0.004	-0.004
	H_0: $\beta_a = 0$, H_1: $\beta_a \neq 0$, p-value	<0.001	<0.001
	Stylized Fact 5: Agricultural employment growth is negatively correlated with population density	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 (β_{na})	0.001	0.001
	H_0: $\beta_{na} = \beta_a$, H_1: $\beta_{na} > \beta_a$, p-value	0.000	0.000
	Stylized Fact 6: Stronger mean reversion in agricultural employment than in non-agricultural employment	Yes	Yes
	Observations	3,657	3,293

Note: This table reports robustness tests of our 6 stylized facts using data on Brazilian Áreas Mínimas Comparáveis (AMCs). Panels B, C, E, and F report tests based on regressions using robust standard errors. Column (2) uses only AMCs in the Northeast, Southeast, and South regions of Brazil. See the text of the paper and this web appendix for further discussion of the construction of the data.

Table A.5: Explanatory power of the Employment Shares and Regression predictions for Brazil

Actual population growth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		As (1) with geographic controls[1]	As (1) with state fixed effects	As (1) with log pop density bin fixed effects	As (1) in the AMC sub sample	As (1) with all controls from (2)-(4)	As (5) with all controls from (2)-(4)
Panel A : Employment Shares prediction							
Predicted Population Growth	0.717***	0.849***	0.929***	0.674***	0.735***	0.693***	0.661***
	(0.128)	(0.099)	(0.131)	(0.138)	(0.150)	(0.089)	(0.095)
Observations	3,657	3,657	3,657	3,657	3,291	3,657	3,291
R-squared	0.15	0.38	0.34	0.29	0.18	0.50	0.48
Panel B : Regression prediction							
Predicted Population Growth	0.813***	0.871***	0.945***	0.678***	0.735***	0.691***	0.657***
	(0.122)	(0.093)	(0.124)	(0.126)	(0.145)	(0.079)	(0.084)
Observations	3,657	3,657	3,657	3,657	3,291	3,657	3,291
R-squared	0.21	0.41	0.36	0.30	0.20	0.51	0.48

Note: Observations are a cross-section of Brazilian AMCs from 1970-2000. In all specifications, the dependent variable is actual population growth. Panel A uses predicted population growth from the Employment Shares prediction. Panel B uses predicted population growth from the Regression prediction. In Columns (2), (6) and (7), the geographical controls are indicators for mineral deposits of oil, nickel, manganese, iron, gold, copper, cobalt, and aluminum (bauxite); an indicator for whether an AMC borders on the ocean or whether its centroid lies within 50 kilometers of a river; and an indicator for whether an AMC's centroid is covered with tropical or subtropical moist broadleaf forest or it is situated in the Amazonas area. The AMC sub-sample in Columns (5) and (7) includes AMCs in the Northeast, Southeast and South regions of Brazil only. Robust standard errors are shown in parentheses. See the text of the paper and this web appendix for further discussion of the construction of the data.

Appendix Table A.6: U.S. MCD match quality by state

	1880 MCDs	1940 MCDs	2000 MCDs	Final (pooled) MCDs	Final/ 1880	Final/ 2000	Class-ification	Aggreg-ation Step 1	Aggreg-ation Step 2	Aggreg-ation Step 3
Arkansas	887	1482	1330	806	0.91	0.61	C	0	0	524
California	419	504	387	133	0.32	0.34	C	0	69	185
Connecticut	167	169	169	167	1.00	0.99	A	0	0	2
Delaware	33	31	27	21	0.64	0.78	C	0	0	6
Washington DC	1	1	1	1	1.00	1.00	A	0	0	0
Georgia	1232	1648	577	505	0.41	0.88	C	0	0	72
Illinois	1583	1638	1708	1446	0.91	0.85	B	0	104	158
Indiana	1011	1015	1009	997	0.99	0.99	A	0	0	12
Iowa	1545	1676	1654	1509	0.98	0.91	A	7	0	145
Kansas	1066	1686	1492	982	0.92	0.66	C	43	135	375
Maine	574	712	530	487	0.85	0.92	B	1	0	43
Maryland	236	302	293	212	0.90	0.72	B	0	15	66
Massachusetts	344	349	351	337	0.98	0.96	A	0	0	14
Michigan	1110	1428	1425	1044	0.94	0.73	B	104	0	381
Minnesota	1220	2911	2506	897	0.74	0.36	C	269	485	1124
Missouri	1134	1303	1379	1099	0.97	0.80	B	0	28	252
Nebraska	654	1506	1198	526	0.80	0.44	C	36	328	344
New Hampshire	245	249	258	236	0.96	0.91	A	1	0	22
New Jersey	265	563	549	254	0.96	0.46	C	17	0	295
New York	969	1006	986	913	0.94	0.93	A	27	21	48
North Carolina	871	1027	1053	834	0.96	0.79	B	2	0	219
Ohio	1373	1445	1548	1320	0.96	0.85	B	31	44	184
Pennsylvania	1995	2567	2469	1671	0.84	0.68	C	110	167	631
Rhode Island	36	39	39	36	1.00	0.92	A	0	0	3
South Carolina	411	574	296	246	0.60	0.83	C	0	0	50
Utah	219	422	90	55	0.25	0.61	C	0	20	15
Vermont	248	252	251	243	0.98	0.97	A	4	0	8
Virginia	434	473	544	362	0.83	0.67	C	0	17	165
West Virginia	326	352	240	208	0.64	0.87	C	0	0	32
Wisconsin	952	1808	1646	902	0.95	0.55	C	255	0	744

Note: This table shows the number of MCDs in each state in each of the three census years (1880, 1940, and 2000). It also reports the number of observations in the final dataset, and ratios of this number to 1880 MCDs and to 2000 MCDs. The table reports the classification of the match quality in each state (the classification is A if both ratios are ≥ 0.9, B if both ratios are ≥ 0.7 but one or more of them is less than 0.9, and C otherwise). The baseline sample we use consists of the A and B states. This sample and the other samples used in robustness checks are described in the paper and this web appendix. The table also reports the number of MCDs aggregated in each step of the data creation process. In the first step we merged together 2000 MCDs with identical state, county and names; in the second step we pooled together MCDs in 1880 counties when we could not identify the location of some of the MCDs; and in the third step we pooled together 2000 MCDs that were not matched to data from 1880 or 1940 to the nearest 2000 MCD that was matched and lay within the same 1880 county. For a detailed discussion of the matching and aggregation process, see this web appendix. Five states are excluded from our analysis (Alaska, Hawaii, Oklahoma, North Dakota and South Dakota), because they had not attained statehood in 1880 and are either not included in the 1880 census or did not have stable county boundaries at that time.

Appendix Table A.7: U.S. geographical sources by state

State	Map sources
Alabama:	No sufficient 1880 map found. Only county data available.
Alaska:	Not included in the 1880 census and so excluded from the dataset.
Arizona:	In 1880 precincts were not separately returned by the enumerators. Only county data available.
Arkansas:	United States Library of Congress map collection, Rand McNally and Co: "Arkansas Administrative Railroad and Township Map", 1898.
California:	National Archives and Records Administration (NARA), Washington DC, Microfilm publication A3378: "Enumeration District Maps of the Twelfth trough Sixteenth Censuses of the United States, 1900 - 1940", County Maps on Microfilm Roll numbers 4 to 6. Additionally (for the counties of Colusa, Napa, Solano and Ventura): Blum, George W., California Book Map, Compiled and Published by Geo. W. Blum, 330 Pine St., S.F. Edward Denny and Co., Agents. Copyrighted 1895 By Geo. W. Blum, San Francisco, Cal., available at David Rumsey Map Collection (www.davidrumsey.com)
Colorado:	In 1880 precincts were not separately returned by the enumerators. Only county data available.
Connecticut:	Mitchell, Samuel A. "Township map of the States of Massachusetts, Connecticut and Rhode Island, Drawn and engraved by W.H. Gamble, Philadelphia. Copyright 1887 by Wm. M. Bradley and Bro. (1890)", Publisher: John Y. Huber and Co, available at David Rumsey Map Collection (www.davidrumsey.com).
Delaware:	Library of Congress, "The Township Map of Delaware", Mc Connell School supply company, copyright McConnell (Philadelphia), 1990.
Florida:	No sufficient 1880 map found. Only county data available.
Georgia:	National Archives and Records Administration (NARA), Washington DC, Microfilm publication A3378: "Enumeration District Maps of the Twelfth trough Sixteenth Censuses of the United States, 1900 - 1940", County Maps on Microfilm Roll numbers 11 and 12.
Hawaii:	Not included in the 1880 census and so excluded from the dataset.
Idaho:	In 1880 precincts were not separately returned by the enumerators. Only county data.
Illinois:	Library of Congress, Rufus Blanchard (cartographer), "Blanchard's township map Illinois", 1867. Additionally Mitchell, Samuel Augustus: "County and Township map of the State of Illinois", (1880) available at David Rumsey Map Collection (www.davidrumsey.com).
Indiana:	Representative Districts Indiana. Published by Baskin, Forster and Co. Lakeside Building Chicago, Ills. 1876. Engraved and Printed by Chas. Shober and Co. Props. of Chicago Lithographing Co.), Andreas, A. T., 1839-1900. Additionally: Gazetteer from United States Geological Survey (geonames.usgs.gov).
Iowa:	Sectional map of Iowa showing civil and congressional townships, all towns, post offices, railroads, streams. Compiled by D.W. Ensign, published by A.T. Andreas, Chicago, Ills., 1875. (Lakeside Building, Chicago, Ills. Engraved and printed by Chas. Shober and Co., Props. of Chicago Lithographing Co.), available at David Rumsey Map Collection (www.davidrumsey.com).
Kansas:	The official state atlas of Kansas compiled from government surveys, county records and personal investigations. Philadelphia. L.H. Everts and Co. 1887. Copyright, 1887, L.H. Everts and Co. (with view:), Additionally: Gazetteer from United States Geological Survey (geonames.usgs.gov).
Kentucky:	No sufficient map found. Only county data available.
Louisiana:	No sufficient map found. Only county data available.
Maine:	Mitchell, Samuel A. "Township map of the State of Maine", Drawn and engraved by W.H. Gamble, Philadelphia. Copyright 1887 by Wm. M. Bradley and Bro. (1890)", Publisher: John Y. Huber and Co, available at David Rumsey Map Collection (www.davidrumsey.com).
Maryland:	Gazetteer from United States Geological Survey (geonames.usgs.gov).
Massachusetts:	Mitchell, Samuel A. "Township map of the States of Massachusetts, Connecticut and Rhode Island", Drawn and engraved by W.H. Gamble, Philadelphia. Copyright 1887 by Wm. M. Bradley and Bro. (1890)", Publisher: John Y. Huber and Co, available at David Rumsey Map Collection (www.davidrumsey.com).
Michigan:	Mitchell, Samuel A. "Township map of the States of Michigan and Wisconsin", Drawn and engraved by W.H. Gamble, Philadelphia. Copyright 1887 by Wm. M. Bradley and Bro. (1890)", Publisher: John Y. Huber and Co, available at David Rumsey Map Collection (www.davidrumsey.com).
Minnesota:	United States Library of Congress, "Map of the state of Minnesota", The Anderson Publishing Company, Arthur Gibson.
Mississippi:	No sufficient map found. Only county data available.
Missouri:	Gazetteer from United States Geological Survey (geonames.usgs.gov).
Montana:	In 1880 precincts were not separately returned by the enumerators. Only county data available.
Nebraska:	The official state Atlas of Nebraska. Compiled from government surveys, county records and personal investigations. Philadelphia, Everts and Kirk, 1885. Copyright, 1885, Everts and Kirk. Additionally gazetteer from United States Geological Survey (geonames.usgs.gov).
Nevada:	In 1880 precincts were not separately returned by the enumerators. Only county data available.
New Hampshire:	County and township map of Vermont and New Hampshire. Copyright 1887 by William M. Bradley and Brother, John Y. Huber Company, Publishers, Philadelphia and St. Louis. (1890), available at David Rumsey Map Collection (www.davidrumsey.com).
New Jersey:	Johnson's new Illustrated (Steel Plate) Family Atlas, With Descriptions, Geographical, Statistical, And Historical. Compiled, Drawn, and Engraved Under The Supervision Of J.H. Colton And A.J. Johnson. New York: Johnson And Browning, Formerly (Successors To J.H. Colton And Company) No. 133 Nassau Street. 1860, available at David Rumsey Map Collection (www.davidrumsey.com).
New Mexico:	In 1880 precincts were not separately returned by the enumerators. Only county data are available.
New York:	Johnson's new Illustrated (Steel Plate) Family Atlas, With Descriptions, Geographical, Statistical, And Historical. Compiled, Drawn, and Engraved Under The Supervision Of J.H. Colton And A.J. Johnson. New York: Johnson And Browning, Formerly (Successors To J.H. Colton And Company,) No. 133 Nassau Street. 1860, available at David Rumsey Map Collection (www.davidrumsey.com).
North Carolina:	National Archives and Records Administration (NARA), Washington DC, Microfilm publication A3378: "Enumeration District Maps of the Twelfth trough Sixteenth Censuses of the United States, 1900 - 1940", County Maps on Microfilm Roll numbers 44 and 45.
North Dakota:	Had not attained statehood in 1880 and did not have stable county boundaries at that time. Excluded from the dataset.
Ohio:	United States Library of Congress, "Colton's Ohio", published by J.H. Colton (New York), 1898.
Oklahoma:	Not included in the 1880 census and so excluded from the dataset.
Oregon:	No sufficient map found. Only county data available.

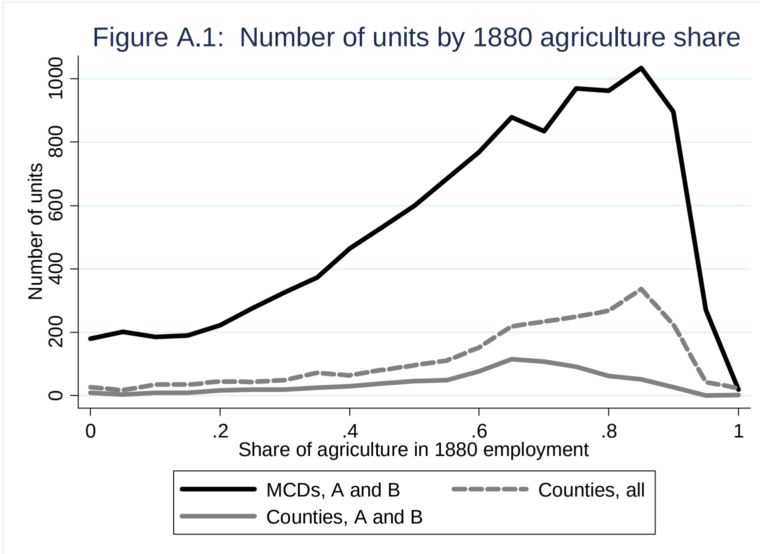
Pennsylvania:	United States Library of Congress, "Pennsylvania administrative Township map", Rand McNally and Co. (Publishers), 1898.
Rhode Island:	Mitchell, Samuel A. "Township map of the States of Massachusetts, Connecticut and Rhode Island", Drawn and engraved by W.H. Gamble, Philadelphia. Copyright 1887 by Wm. M. Bradley and Bro. (1890)", Publisher: John Y. Huber and Co, available at David Rumsey Map Collection (www.davidrumsey.com).
South Carolina:	National Archives and Records Administration (NARA), Washington DC, Microfilm publication A3378: "Enumeration District Maps of the Twelfth trough Sixteenth Censuses of the United States, 1900 - 1940", County Maps on Microfilm Roll number 58.
South Dakota:	Had not attained statehood in 1880 and did not have stable county boundaries at that time. Excluded from the dataset.
Tennessee:	No sufficient map found. Only county data available.
Texas:	No sufficient map found. Only county data available.
Utah:	No additional map source was found, but the 1880 MCDs could be identified using the 1940 MCD maps.
Vermont:	County and township map of Vermont and New Hampshire. Copyright 1887 by Wm. M. Bradley and Bro., John Y. Huber Company, Publishers, Philadelphia and St. Louis. (1890), available at David Rumsey Map Collection (www.davidrumsey.com).
Virginia:	Gazetteer from United States Geological Survey (geonames.usgs.gov).
Washington:	In 1880 precincts were not separately returned by the enumerators. Only county data available.
West Virginia:	White's political map of West Virginia. Drawn and engraved by W.H. Gamble, Philadelphia. Entered according to Act of Congress in the year 1873 by M. Wood White in the Office of the Librarian of Congress at Washington, 1873, available at David Rumsey Map Collection (www.davidrumsey.com).
Wisconsin:	Mitchell, Samuel A. "Township map of the States of Michigan and Wisconsin", Drawn and engraved by W.H. Gamble, Philadelphia. Copyright 1887 by Wm. M. Bradley and Bro. (1890)", Publisher: John Y. Huber and Co. Additionally county maps on Lincoln county and Marathon county (Map of Wisconsin showing congressional and judicial districts. Copyright 1877 by Snyder, Van Vechten and Co. (Compiled and published by Snyder, Van Vechten and Co., Milwaukee. 1878)). Both available at David Rumsey Map Collection (www.davidrumsey.com).
Wyoming:	In 1880 precincts were not separately returned by the enumerators. Only county data available.

The states excluded from our analysis are: Alaska, Hawaii and Oklahoma (which are not included in the 1880 census) and North and South Dakota (which had not attained statehood in 1880 and did not have stable county boundaries at that time).

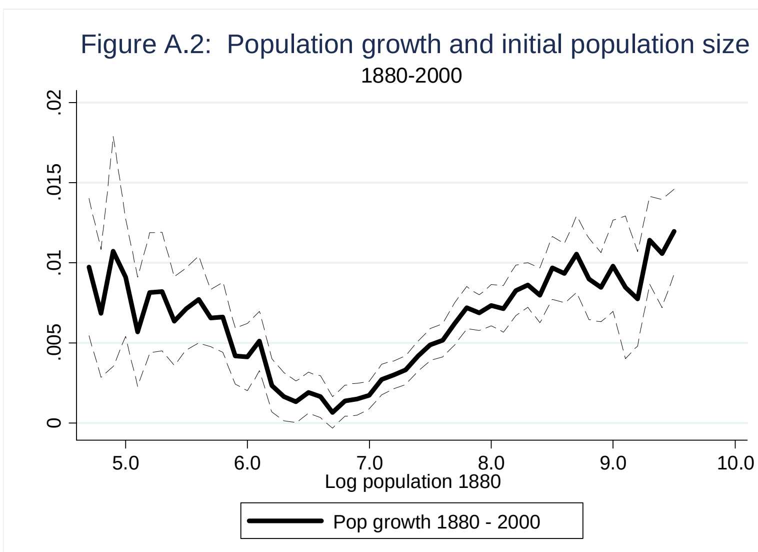
The included states for which we could not create data at the sub-county level are: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Washington and Wyoming (for these 10 states the 1880 census contains a note saying that: "As the precincts in the different counties were not separately returned by the enumerators, the counties cannot be published in detail", which precludes obtaining information on sub-county divisions); the states of Alabama, Florida, Kentucky, Louisiana, Mississippi, Oregon, Tennessee and Texas (in these states we were unable to find sufficient maps to determine the location of the 1880 MCDs). The main problem with these states was that many of the MCD entries contained numbers instead of names, such as "Beat 1" or "Precinct 5". These entries are much harder to find on maps than names, since there are many competing numbering schemes applied to maps of this period, and number schemes are changed more frequently than names.

All MCDs with coordinates are present in one of the sources listed above apart from the following MCDs. These were found in the footnotes of the censuses 1890 – 1930 (the number indicates the county):

FERGUSON, 430, Arkansas; BLACKWOOD, 610, Arkansas; TREMONT, 1030, Arkansas; BRIDGE BEND, 1170, Arkansas; SPRING CREEK, 1490, Arkansas; HAPPY CAMP, 150, California; MOUNTAIN, 150, California; HOT SPRING, 490, California; SOUTH FORK, 490, California; HOT SPRINGS, 550, California; OSO FLACO, 790, California; SALINAS, 790, California; SAN JOSE, 790, California; SALMON, 930, California; SOUTH, 930, California; CASCADE, 1030, California; LASSEN, 1030, California; BUCKEYE, 1130, California; FAIRVIEW, 1130, California; MERRITT, 1130, California; ALLEN, 650, Illinois; ALLIN, 1130, Illinois; " ", 1770, Indiana; ELWOOD, 110, Kansas; KIOWA, 110, Kansas; LAKE CITY, 110, Kansas; MEDICINE LODGE, 110, Kansas; SUN CITY, 110, Kansas; MILLROOK, 670, Kansas; MILLROOK, 710, Kansas; Anthony, 750, Kansas; NOBLE, 1310, Kansas; VALLEY, 1310, Kansas; TWIN MOUND, 1370, Kansas; LUDWICK, 1510, Kansas; WEST WATERVILLE, 110, Maine; MUSCLE RIDGE, 130, Maine; MUSCONGUS ISLAND, 150, Maine; PINE, 170, Maine; BURBANK T, 210, Maine; MOUNT KINEO, 210, Maine; PERKINS, 230, Maine; HOLDEN, 250, Maine; PLEASANT VALLEY, 410, Maryland; DISTRICT 6, CLOBOURNES, 450, Maryland; CENTER, 1010, Minnesota; TOWNSHIP 105, RANGE 42, 1010, Minnesota; TOWNSHIP 103, RANGE 42, 1050, Minnesota; TOWNSHIP 103, RANGE 43, 1050, Minnesota; EAST BATTLE LAKE, 1110, Minnesota; TOWNSHIP 132, 1110, Minnesota; TOWNSHIP 133, RANGE 49, 1110, Minnesota; TOWNSHIP 135, RANGE 42, 1110, Minnesota; TOWNSHIP 136, RANGE 36, 1110, Minnesota; TOWNSHIP 136, RANGE 37, 1110, Minnesota; TOWNSHIP 137, RANGE 36, 1110, Minnesota; TOWNSHIP 137, RANGE 37, 1110, Minnesota; TOWNSHIP 137, RANGE 38, 1110, Minnesota; RESERVE, 1230, Minnesota; TOWN 111 RANGE 38, 1270, Minnesota; DULUTH (I), 1370, Minnesota; ONEOTA, 1370, Minnesota; SAHLMARK, 1490, Minnesota; TOWNSHIP 124 RANGE 44, 1490, Minnesota; TOWNSHIP 125 RANGE 44, 1490, Minnesota; TOWNSHIP 126 RANGE 44, 1490, Minnesota; MORITZIUS (I), 1710, Minnesota; OTIS, 1730, Minnesota; TOWNSHIP 114 RANGE 46, 1730, Minnesota; GERMAN, 170, Missouri; BENTON, 430, Missouri; GALLOWAY, 430, Missouri; MARION, 430, Missouri; WESTPORT, 950, Missouri; GERMAN, 1230, Missouri; EAST, 1430, Missouri; LYNN, 1490, Missouri; OAK GROVE, 1490, Missouri; MARION, 1530, Missouri; FOURCHEE, 1810, Missouri; CARONDELET, 1890, Missouri; CENTRAL, 1890, Missouri; JEFFERSON, 1950, Missouri; COURT-HOUSE ROCK, 330, Nebraska; SCOTT, 350, Nebraska; CEDAR VALLEY, 810, Nebraska; PLATTE, 810, Nebraska; SPRING CREEK, 830, Nebraska; SPRINGBROOK, 830, Nebraska; CAPITAL, 1090, Nebraska; MIDLAND, 1090, Nebraska; BOHNART, 1290, Nebraska; SPRING VALLEY, 1290, Nebraska; JOHNSON CREEK, 1510, Nebraska; GRANT, 1690, Nebraska; LISBON, 450, New York; GRAMPION 10 Utah; TERRACE 30 Utah; HILLSDALE, 210, Utah; LITTLE PINTO, 210, Utah; TINTIC, 230, Utah; BELLEVUE, 250, Utah; DUNCANS RETREAT, 250, Utah; GRAFTON, 250, Utah; JOHNSON, 250, Utah; PAH REAH, 250, Utah; SHUNESBURG, 250, Utah; KANYON, 290, Utah; MEADOWVILLE, 330, Utah; FREEDOM, 390, Utah; PETTY, 390, Utah; VERMILLION, 410, Utah; WILLOW BEND, 410, Utah; HAYTSVILLE, 430, Utah; BATESVILLE, 450, Utah; JACOB CITY, 450, Utah; MILL, 450, Utah; HEBRON, 530, Utah; PINTO, 530, Utah; PRICE CITY, 530, Utah; SILVER REEF, 530, Utah; LYNNE, 570, Utah; WALKER, 1950, Virginia; LEE, 130, West Virginia; SULLIVAN, 530, Wisconsin; CARPENTER, 670, Wisconsin; BRANNAN, 990, Wisconsin.

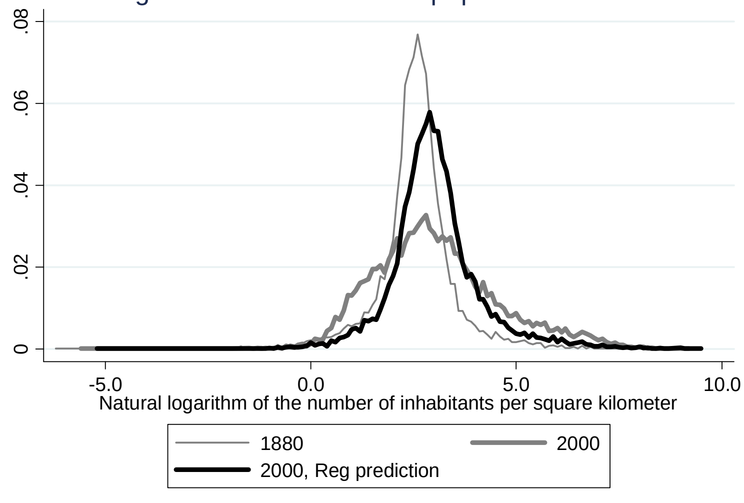


Note: The number of observations (MCDs and counties) by share of agriculture in 1880 employment. The x-axis displays bins of size 0.05, defined by rounding down the agricultural employment share for each unit. For example, all units with a share of agriculture greater than or equal to 0.1 and less than 0.15 are grouped together in bin 0.1. See the text of the paper and this web appendix for further discussion of the construction of the data.



Note: Mean population growth from 1880-2000 within each population bin is based on estimating equation (2) in the paper for MCDs in the A and B states. 95 percent confidence intervals are computed using robust standard errors clustered by county. Since population density bins at the extreme ends of the distribution typically contain few observations, the figure (but not the estimation) omits the 1 percent most and least dense MCDs in 1880. See the text of the paper and this web appendix for further discussion of the construction of the data.

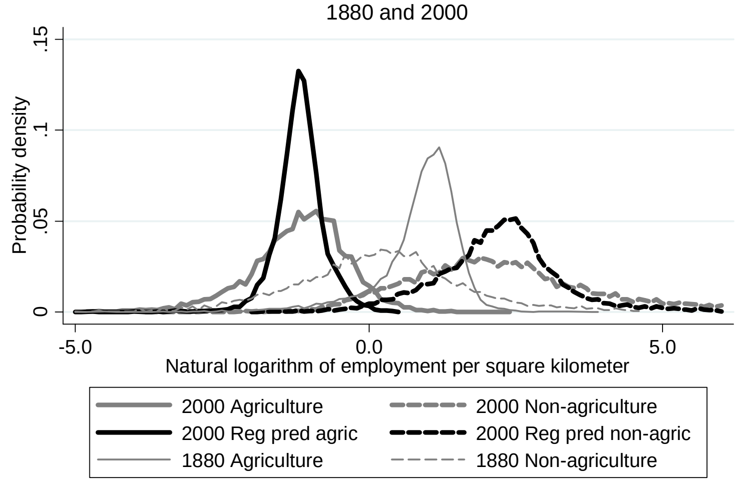
Figure A.3: Distribution of population densities



Note: This figure reports the distribution of population density for MCDs in the A and B states in 1880 and 2000 and predicted population density for 2000 using the Regression prediction. The x-axis uses population density bins, 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 greater than or equal 0.1 and less than 0.2 are grouped together in bin 0.1. See the text of the paper and this web appendix for further discussion of the construction of the data.

Figure A.4: Distribution of employment densities

1880 and 2000



Note: The figure uses MCD data for the sample of A and B states. The x-axis uses population density bins, 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 greater than or equal to 0.1 and less than 0.2 are grouped together in bin 0.1. See the text of the paper and this web appendix for further discussion of the construction of the data.

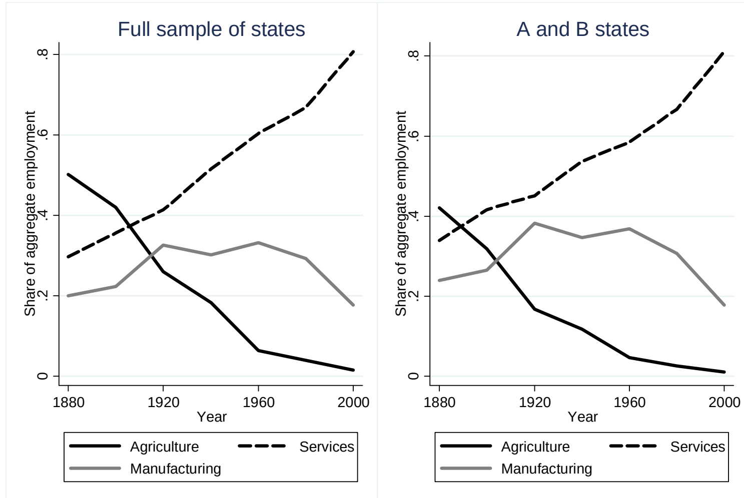


Figure A.5: Aggregate employment shares over time

Note: Shares of agriculture, manufacturing and services in total employment over time in our county sub-periods data, using all states in these data (left panel) and A and B states (right panel). See this web appendix for further details on the sector definitions and data construction.

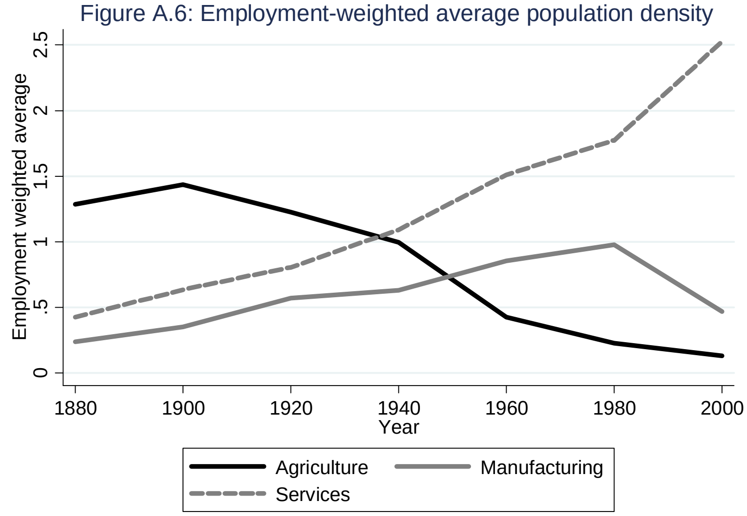
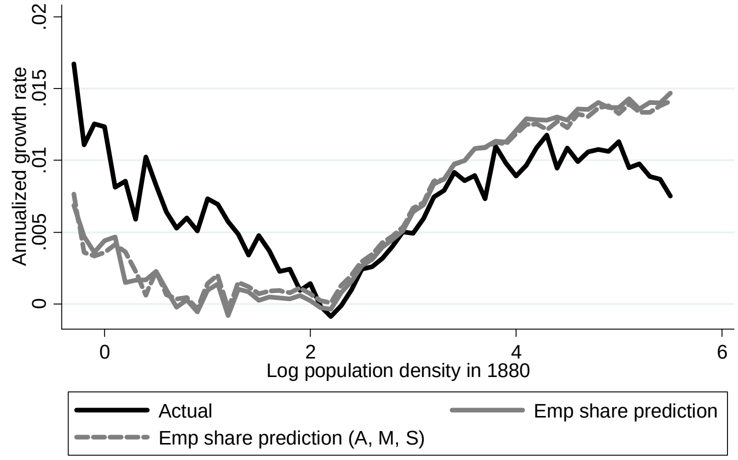


Figure A.6: Employment-weighted average population density

Note: Each line shows average log population density for an industry, calculated as the industry-employment-weighted average of log population density in all counties. The figure uses our county sub-periods dataset and includes all states in these data. See the text of the paper and this web appendix for further details on the sector definitions and data construction.

Figure A.7: Actual and predicted population growth

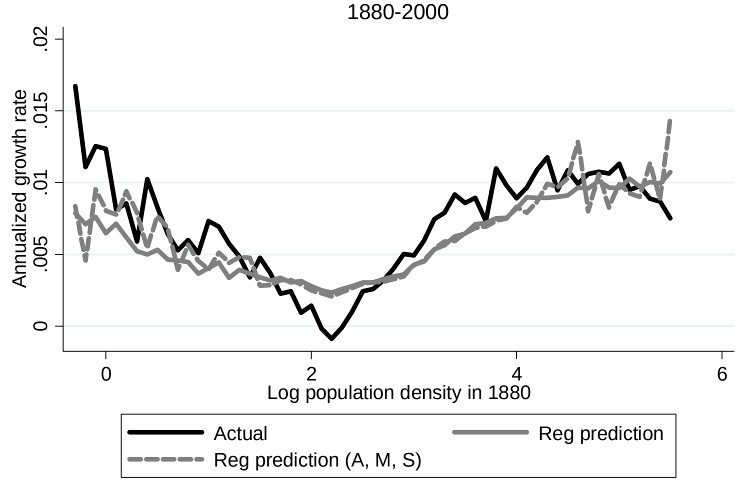
1880-2000



Note: The black line shows mean population growth from 1880-2000 within each population density bin based on estimating equation (2) in the paper for MCDs in the A and B states. "Emp share prediction (A, M, S)" is the same as the Employment Shares prediction, except that it separates non-agriculture into manufacturing and services. Population density bins are defined by rounding down log population 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 few observations, the figure (but not the estimation) omits the 1 percent most and least dense MCDs in 1880. See the text of the paper and this web appendix for further discussion of the construction of the data.

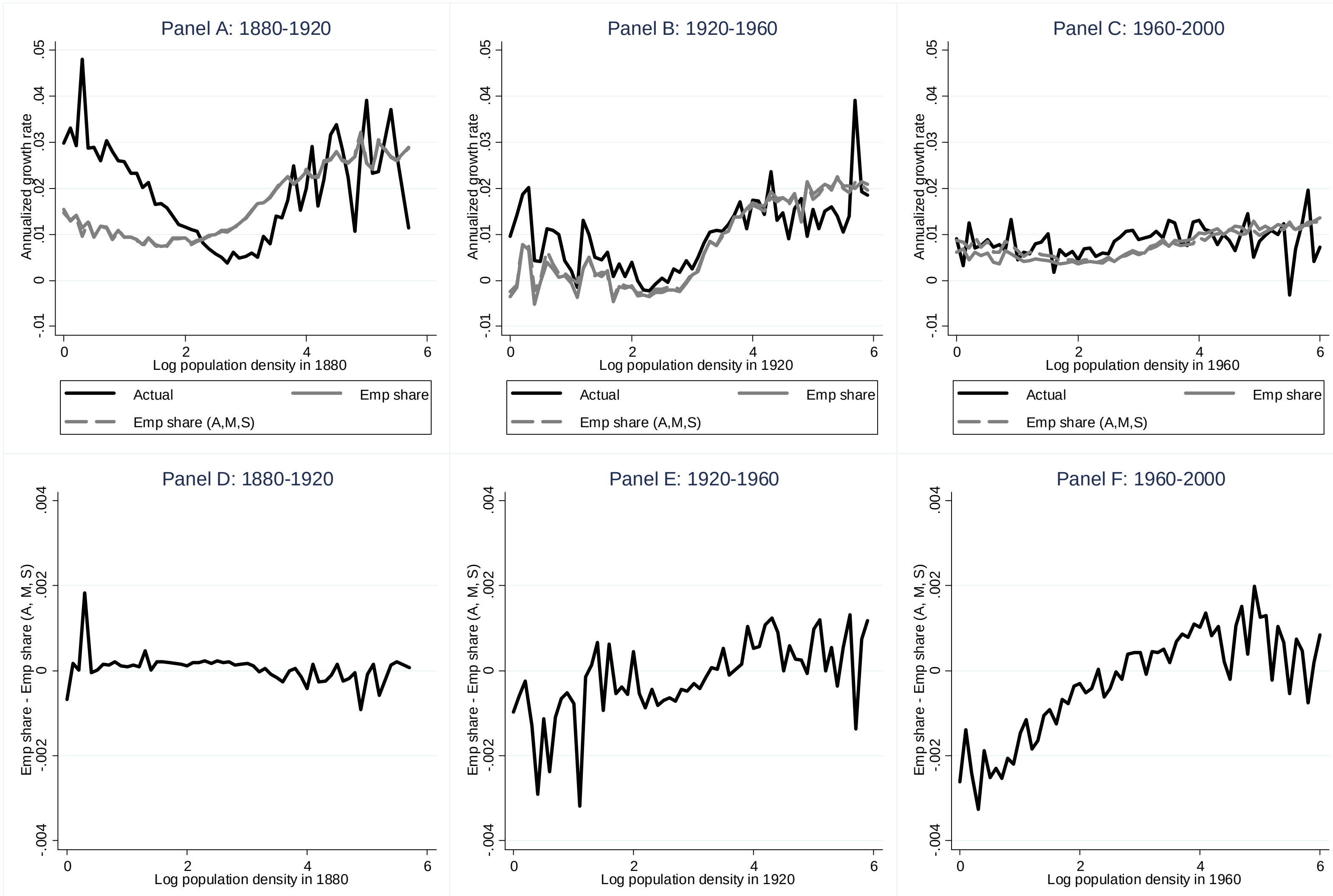
Figure A.8: Actual and predicted population growth

1880-2000

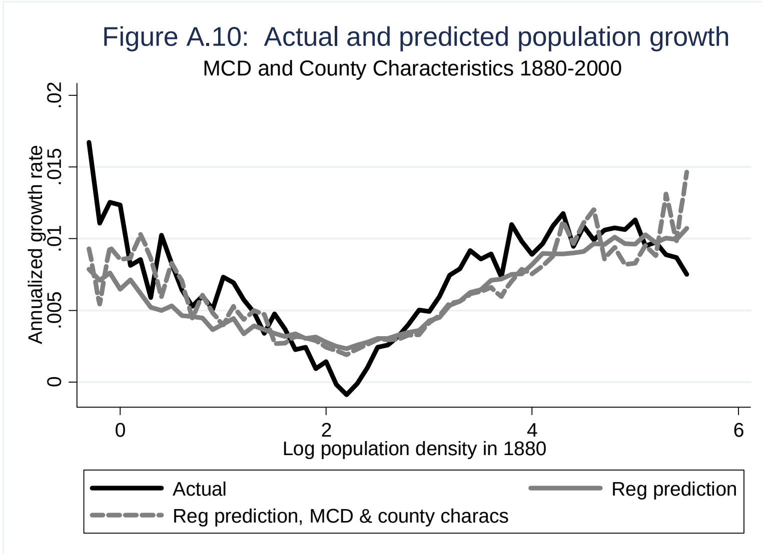


Note: The black line shows mean population growth from 1880-2000 within each population density bin based on estimating equation (2) in the paper for MCDs in the A and B states. "Reg prediction (A, M, S)" is the same as the Regression prediction, except that it separates non-agriculture into manufacturing and services. Population density bins are defined by rounding down log population 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 few observations, the figure (but not the estimation) omits the 1 percent most and least dense MCDs in 1880. See the text of the paper and this web appendix for further discussion of the construction of the data.

Figure A.9: Actual and predicted population growth

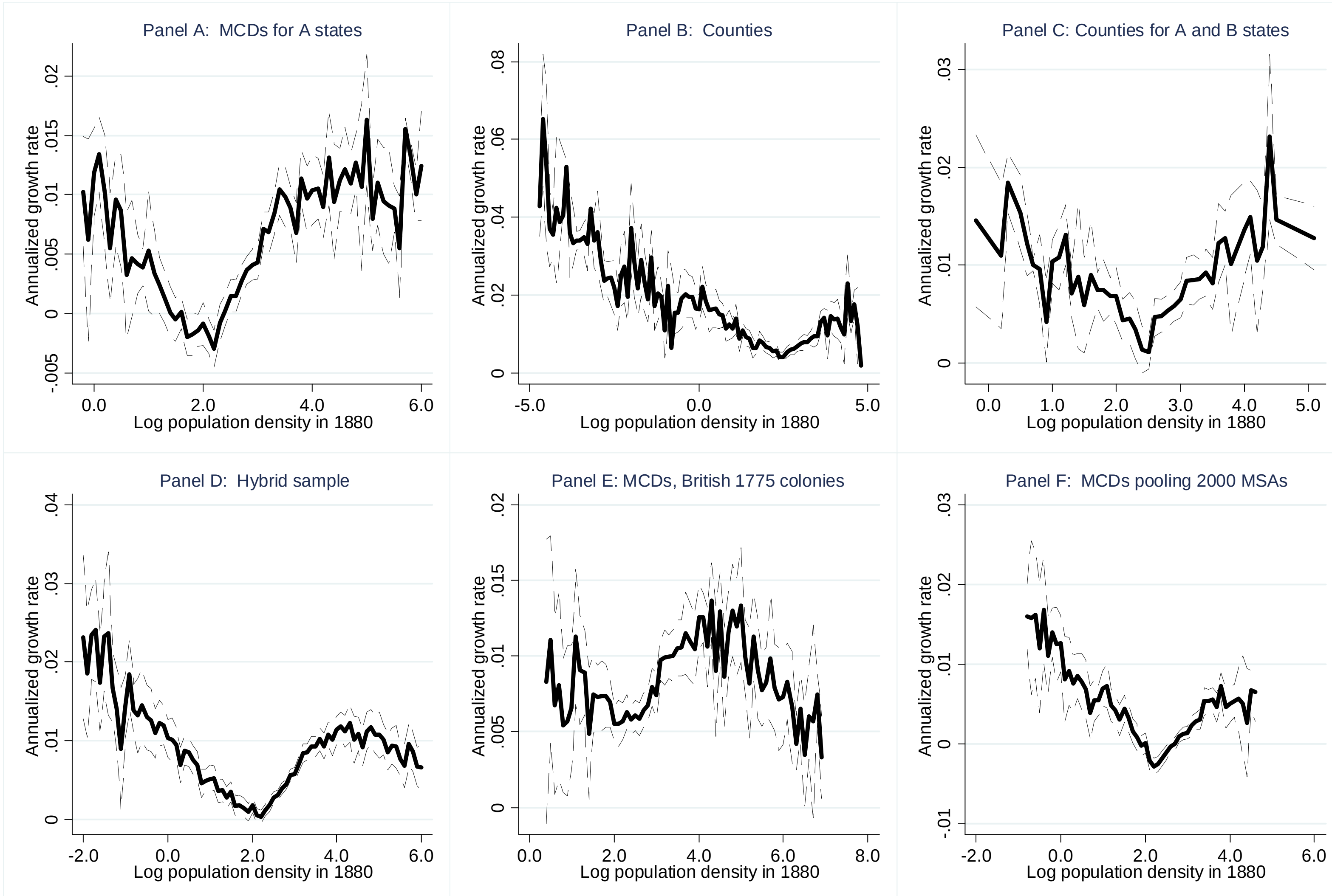


Note: Panels A-C repeat the exercise of Figure A7 using our county sub-periods data. Panels D-F show the difference between our baseline Employment Shares prediction using agriculture and non-agriculture and the augmented Employment Shares prediction that disaggregates non-agriculture into manufacturing and services. See the text of the paper and this web appendix for further details about the construction of the data.

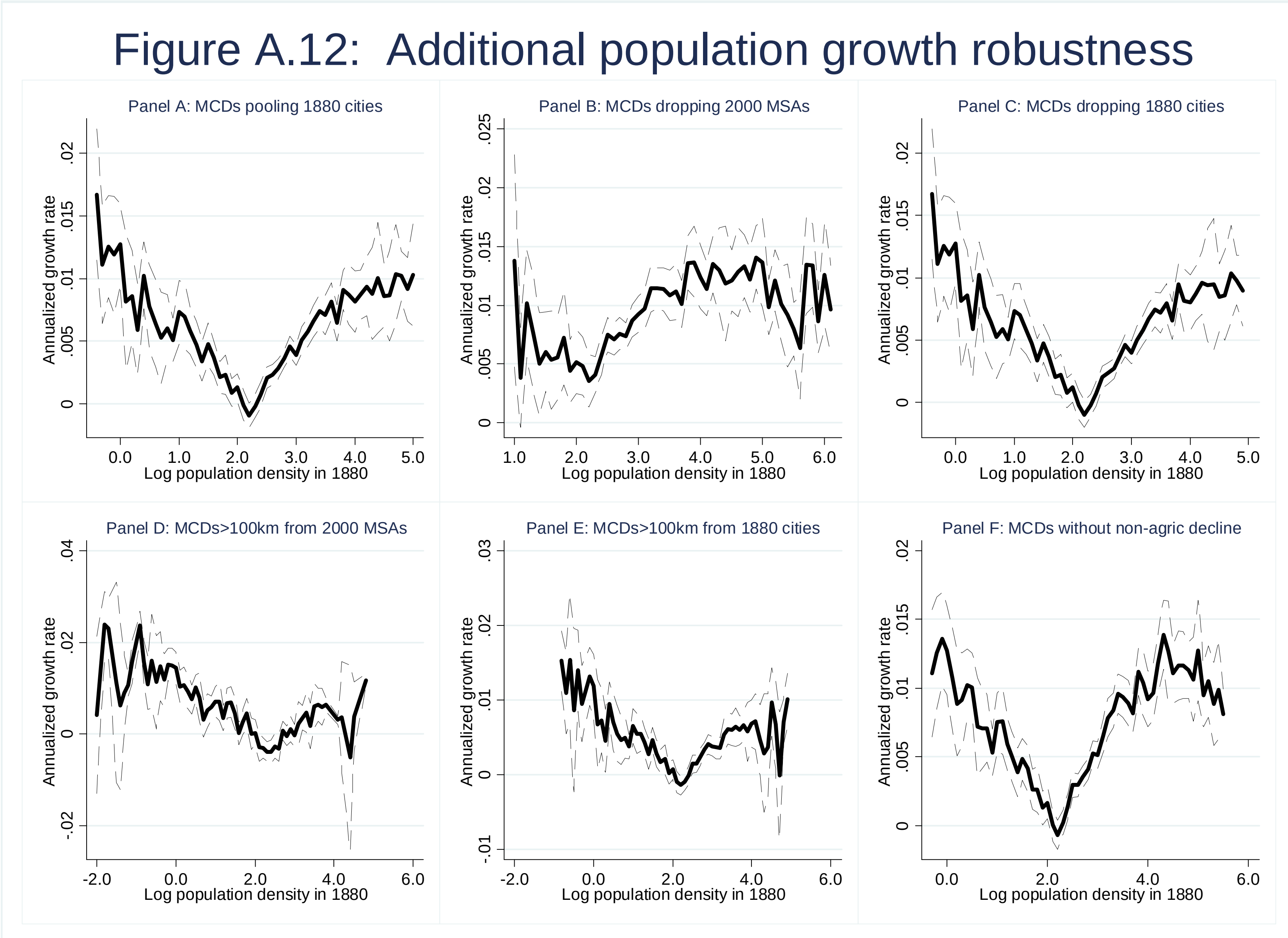


Note: The black line shows mean population growth from 1880-2000 within each population density bin based on estimating equation (2) in the paper for the sample of A and B states. The Regression prediction is based on the 1880 agricultural employment share, 1880 log population density and their interaction, as discussed in the paper. The Regression prediction with MCD and county characteristics augments this specification with the 1880 agricultural employment share and 1880 log population density for the county of which the MCD is part as well as their interaction, as discussed in this appendix. Population density bins are defined by rounding down log population 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 few observations, the figure (but not the estimation) omits the 1 percent most and least dense MCDs in 1880. See the text of the paper and this web appendix for further discussion of the construction of the data.

Figure A.11: Population growth robustness

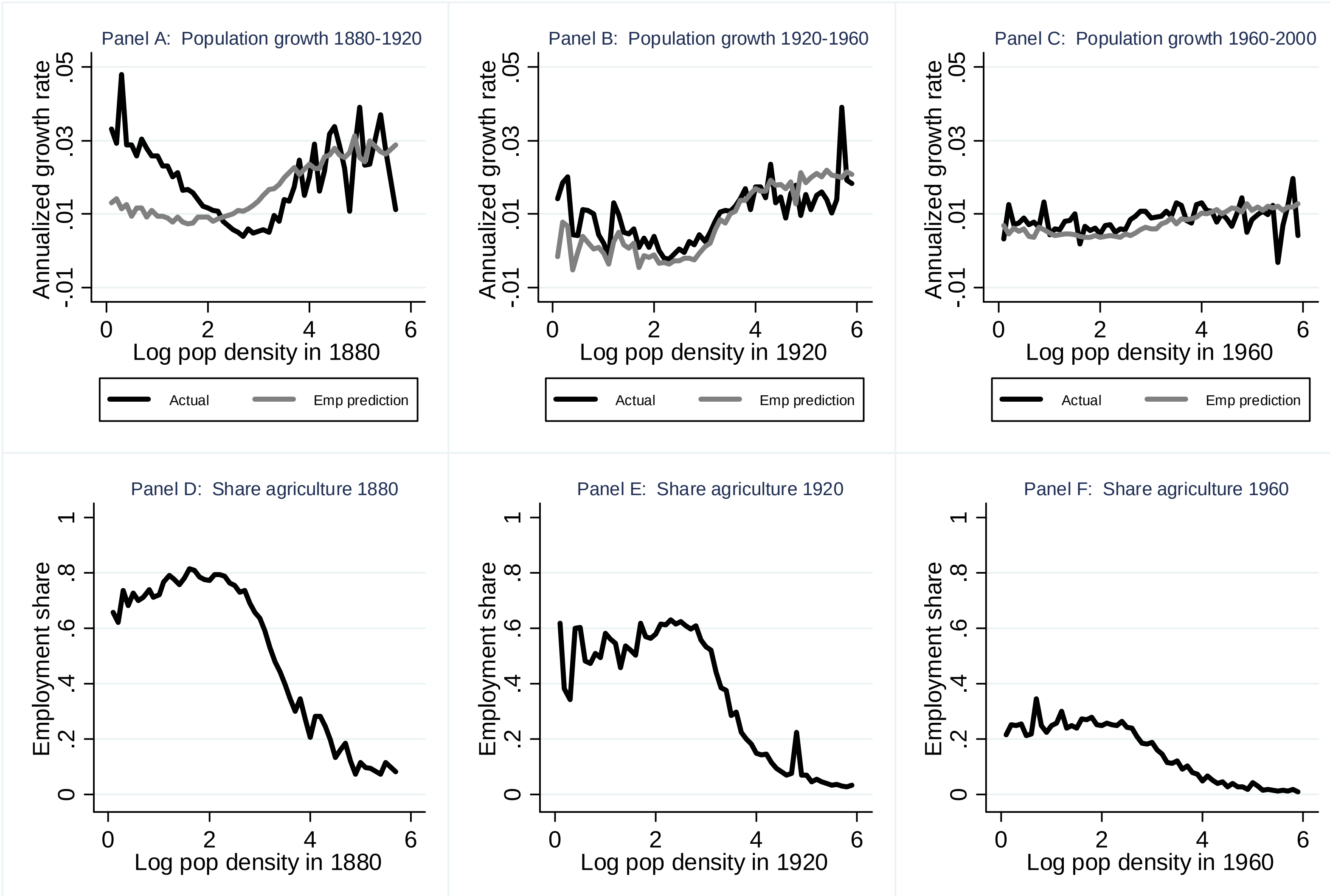


Note: This figure shows the robustness of stylized fact 2 (Figure 1, Panel B in the paper) for population growth by reproducing it for other samples. Since population density bins at the extreme ends of the distribution typically contain few observations, the figures (but not the estimations) omit the 1 percent most and least dense MCDs in 1880. See the text of the paper and this web appendix for further details on the samples used in each panel and the construction of the data.

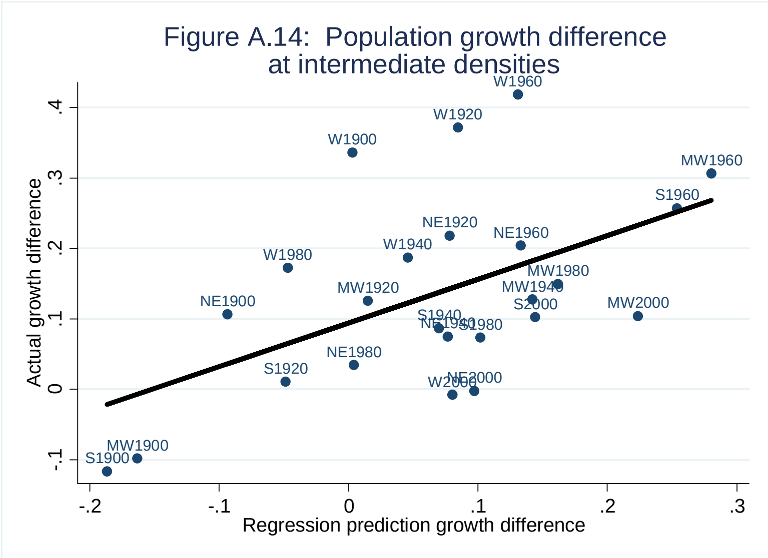


Note: This figure shows the robustness of stylized fact 2 (Figure 1, Panel B in the paper) for population growth by reproducing it for other samples. Since population density bins at the extreme ends of the distribution typically contain few observations, the figures (but not the estimations) omit the 1 percent most and least dense MCDs in 1880. See the text of the paper and this web appendix for further details on the samples used in each panel and the construction of the data.

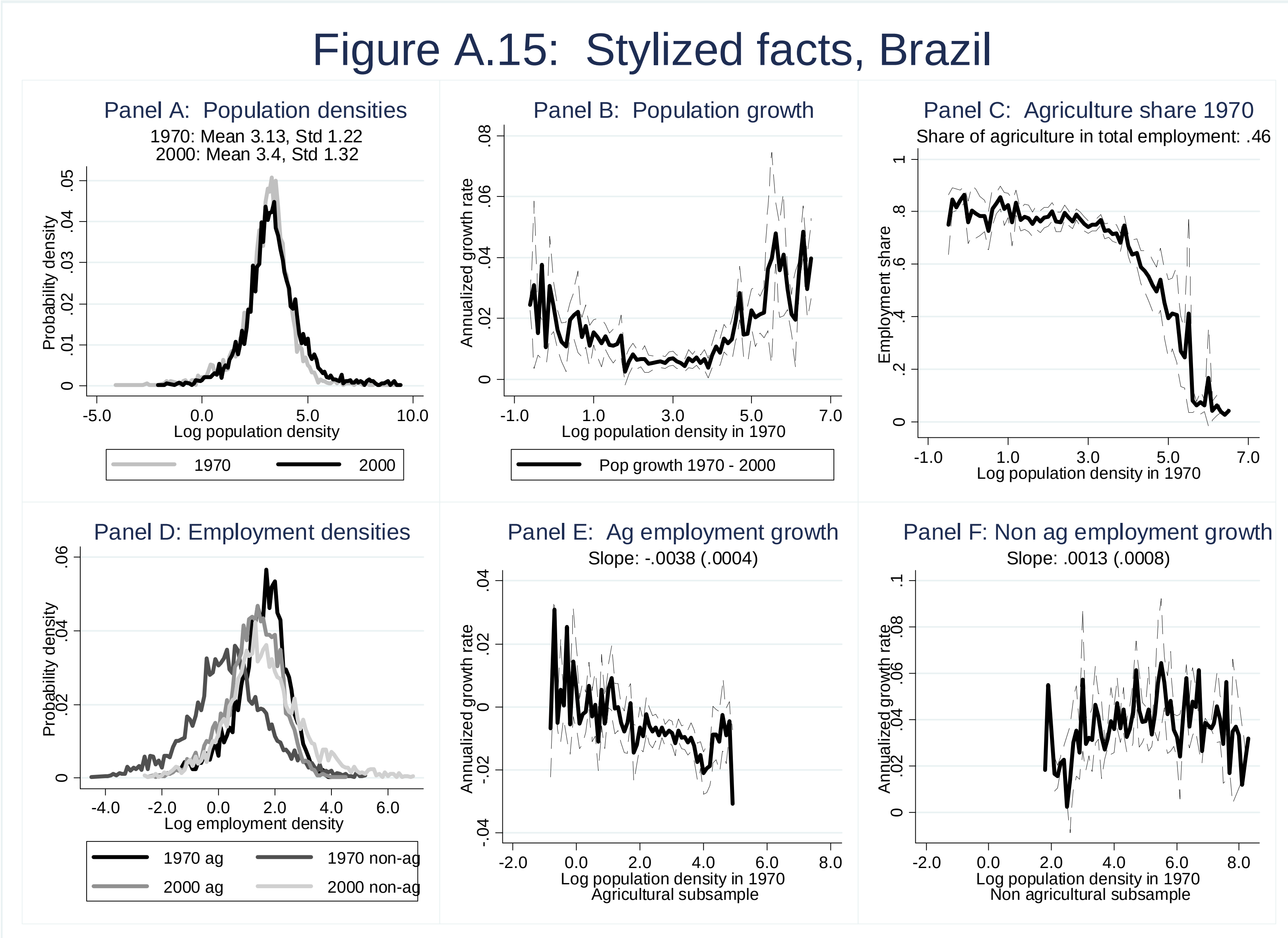
Figure A.13: Actual population growth and prediction



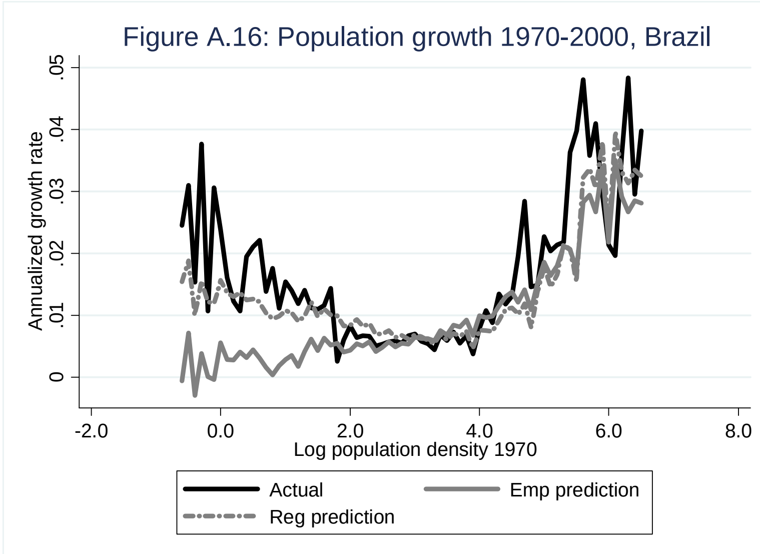
Note: The black (grey) lines in Panels A-C show mean actual population growth (mean predicted population growth from the Employment Shares prediction) for each population density bin based on estimating equation (2) in the paper for three 40-year sub-periods. Panels D-F show the mean share of agriculture in employment for each population density bin in the initial years of the same sub-periods. Log population density bins are defined by rounding down log population 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 few observations, the figures (but not the estimations) focus on the density range 0-6. See the text of the paper and this web appendix for further discussion of the construction of the data.



Note: This figure shows a relationship similar to that in Panel C of Figure 3 in the paper, but uses the Regression prediction instead of the Employment Shares prediction. The figure uses our county sub-periods data. See the text of the paper and this web appendix for a discussion of the construction of the figure and Regression prediction.

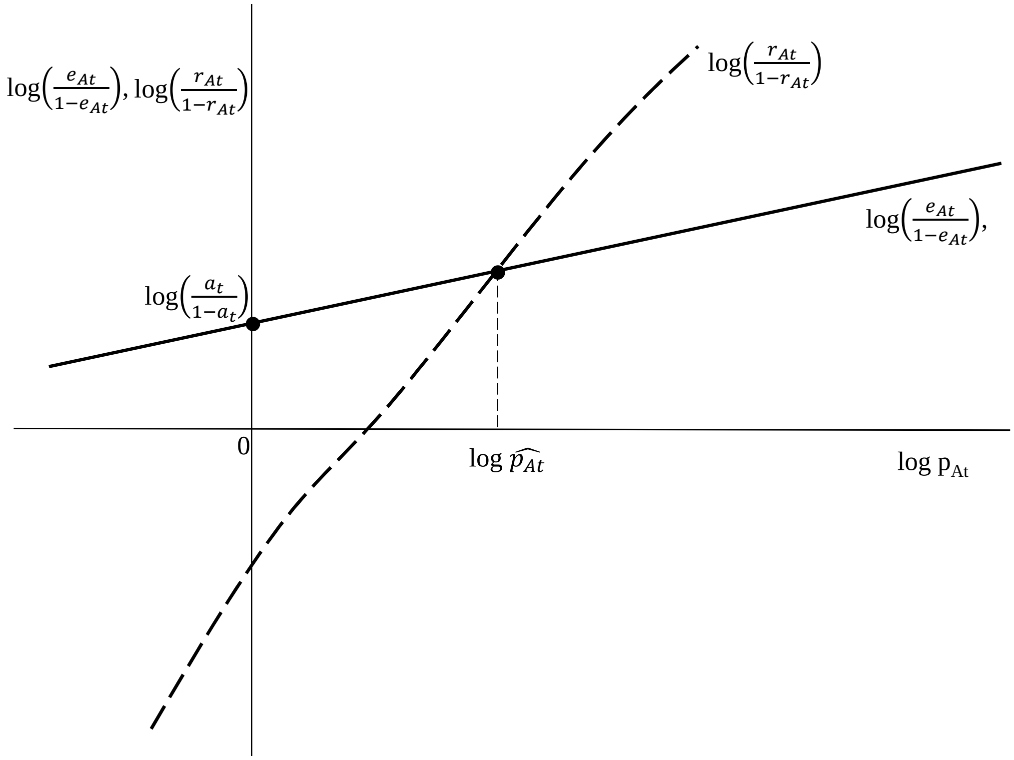


Note: This figure reproduces the stylized facts shown in Figure 1 in the paper using Brazilian instead of U.S. data. See the text of the paper and this web appendix for further discussion of the construction of the data. Since population density bins at the extreme ends of the distribution typically contain few observations, in panels B, C, E and F the figures (but not the estimations) omit the 1 percent most and least dense AMCs in 1970.



Note: This figure shows the explanatory power of the Employment Share and Regression prediction using Brazilian data. The x-axis uses population density bins, defined by rounding down log population density for each AMC to the nearest single digit after the decimal point. For example, all AMCs with log population density greater than or equal to 0.1 and less than 0.2 are grouped together in bin 0.1. Since population density bins at the extreme ends of the distribution typically contain few observations, the figure (but not the estimation) omits the 1 percent most and least dense AMCs in 1970. See the text of the paper and this web appendix for further discussion of the construction of the data.

Figure A.17: Existence of a unique equilibrium relative price of the agricultural good



Note: This figure shows the existence of a unique equilibrium relative price of the agricultural good in the model, as discussed in this web appendix.