APPENDIX

A. THEORY

The theory appendix contains additional proofs and derivations omitted in the body of the paper.

A.1 The PIGL Demand Function

Consider the indirect utility function given in equation (1). Roy's Identity implies that sectoral expenditure shares are given by the following formula:

$$\text{(A.1)} \qquad \vartheta_s \equiv \vartheta_s\left(y, P_{rA}, P_{rM}\right) = -\frac{\frac{\partial V(y, P_{rA}, P_{rM})}{\partial p_s} P_{rs}}{\frac{\partial V(y, P_{rA}, P_{rM})}{\partial y} y}.$$

We compute the numerator and denominator separately and then combine them. The numerator can be written:

$$\frac{\partial V\left(y, P_{rA}, P_{rM}\right)}{\partial P_{rA}} P_{rA} = \frac{\partial}{\partial P_{rA}}\left[\frac{1}{\eta}\left(\frac{y}{P_{rA}^{\phi} P_{rM}^{1-\phi}}\right)^{\eta} - \nu \ln\left(\frac{P_{rA}}{P_{rM}}\right)\right] P_{rA} = -\phi\left(\frac{y}{P_{rA}^{\phi} P_{rM}^{1-\phi}}\right)^{\eta} - \nu,$$

while the denominator has the following expression:

$$\frac{\partial V\left(y, P_{rA}, P_{rM}\right)}{\partial y} y = \left(\frac{y}{P_{rA}^{\phi} P_{rM}^{1-\phi}}\right)^{\eta}.$$

Combining the two previous derivatives using the expression in equation A.1 yields the following expressions of the sectoral expenditure shares:

$$\text{(A.2)} \qquad \vartheta_A = \phi + \nu\left(\frac{y}{P_{rA}^{\phi} P_{rM}^{1-\phi}}\right)^{-\eta} \qquad \vartheta_M = (1-\phi) - \nu\left(\frac{e}{P_{rA}^{\phi} P_{rM}^{1-\phi}}\right)^{-\eta}.$$

The Allen-Uzawa elasticity of substitution is given by

$$\varrho = \frac{\frac{\partial^2 e(P_{rA}, P_{rM}, V)}{\partial P_{rA} \partial P_{rM}} e\left(P_{rA}, P_{rM}, V\right)}{\frac{\partial e(P_{rA}, P_{rM}, V)}{\partial P_{rA}} \frac{\partial e(P_{rA}, P_{rM}, V)}{\partial P_{rM}}},$$

where $e(P_{rA}, P_{rM}, V)$ is the expenditure function given by

$$e(P_{rA}, P_{rM}, V) = (V + \nu \ln(P_{rA}/P_{rM}))^{1/\eta} \eta^{1/\eta} P_{rA}^{\phi} P_{rM}^{1-\phi}. \tag{A.3}$$

Using equation (A.3), one can show that

$$\varrho = 1 - \eta \frac{(\vartheta_A - \phi)(\vartheta_M - (1-\phi))}{\vartheta_A \vartheta_M} = 1 + \eta \frac{(\vartheta_A - \phi)^2}{\vartheta_A (1 - \vartheta_A)}.$$

A.2 Labor Supply

We denote total payments per efficiency unit of labor in region r and sector s by w_{rs}. In non-agriculture, these payments reflect only the wage per efficiency unit. In agriculture, they also include the payments to land which are redistributed to workers, so:

$$w_{rA} \equiv \tilde{w}_{rA} + \frac{\alpha}{1-\alpha}\tilde{w}_{rA} = \frac{1}{1-\alpha}\tilde{w}_{rA},$$

where $\tilde{w}_{rA}$ denotes the wage per efficiency unit in agriculture in region r.

Individual workers learn the amount of efficiency units of labor they can supply to either sector once they arrive in a location. We denote the efficiency units individual i can supply to each sector by z_A^i and z_M^i. Individuals draw their efficiency units from a sector-specific Fréchet distribution, $P(z_s^i \leq z) = F_s(z) = e^{-h_s z^{-\zeta}}$, where the term h_s determines the efficiency units of the average worker and ζ the dispersion of efficiency units across workers in sector s. Since h_s is not separately identified from sector s productivity, i.e., Z_{rs}, we set $h_s = 1\,\forall s$ without loss of generality.

Worker i then chooses their sector so as to maximize their labor income, $y_r^i = \max_s \{w_{rs} z_s^i\}$. A standard set of arguments implies the following analytical expressions for the key objects of our theory.

1. Sectoral employment shares are

$$s_{rs} = h_s \left(\frac{w_{rs}}{\overline{w}_r}\right)^{\zeta} \quad \text{where} \quad \overline{w}_r = \left(\sum_s w_{rs}^{\zeta}\right)^{1/\zeta}. \tag{A.4}$$

2. The aggregate amounts of sectoral human capital are

$$H_{rs} = \Gamma_\zeta L_r \left(\frac{w_{rs}}{\overline{w}_r}\right)^{\zeta-1} = \Gamma_\zeta L_r s_{rs}^{\frac{\zeta-1}{\zeta}},$$

where $\Gamma_x \equiv \Gamma(1-1/x)$ and Γ denotes the Gamma function.

3. Total sectoral earnings are

$$w_{rs}H_{rs} = w_{rs}\Gamma_\zeta L_r \left(\frac{w_{rs}}{\overline{w}_r}\right)^{\zeta-1} = \overline{w}_r\Gamma_\zeta L_r \left(\frac{w_{rs}}{\overline{w}_r}\right)^{\zeta} = \overline{w}_r\Gamma_\zeta L_r s_{rs}.$$

4. The distribution of realized labor income, y_r^i, inherits the Fréchet distribution of the underlying efficiency units of labor and is given by

$$F_r(y) \equiv P\left(y_r^i \leq y\right) = e^{-\left(\sum_s w_{rs}^\zeta\right)y^{-\zeta}} = e^{-(y/\overline{w}_r)^{-\zeta}}. \tag{A.5}$$

Hence, a worker's expected income in region r prior to moving there is given by $E\left[y_r^i\right] = \Gamma_\zeta \overline{w}_r$. Due to the law of large numbers this also corresponds to the ex-post average income in location r, so that, $Y_r = \overline{w}_r\Gamma_\zeta L_r$.

A.3 PIGL Aggregation

In this section we derive the aggregate demand system and the expression for special welfare introduced in Section 2.2.

Aggregate Demand

Let $F_r(y)$ be the distribution of income derived in equation (A.5). Integrating over the sectoral expenditure shares of individual workers in region r in equation A.2 yields an expression for a region's aggregate expenditure share.

$$\vartheta_{rs} \equiv \vartheta_{rs}(\bar{w}_r, P_{rA}, P_{rM}) = \frac{\int \vartheta_A(y, P_{rA}, P_{rM})\, y dF_r(y)}{\int y dF_r(y)} = \phi + \nu\left(\frac{1}{P_{rA}^\phi P_{rM}^{1-\phi}}\right)^{-\eta}\frac{\int y^{1-\eta} dF_r(y)}{\int y dF_r(y)}.$$

Given that $F_r(y) = e^{-(y/\overline{w}_r)^{-\zeta}}$, we have that

$$P\left(y^{1-\eta} < m\right) = P\left(y < m^{\frac{1}{1-\eta}}\right) = e^{-\left(\frac{m^{\frac{1}{1-\eta}}}{\overline{w}_r}\right)^{-\zeta}} = e^{-\left(\frac{m}{\overline{w}_r^{1-\eta}}\right)^{-\frac{\zeta}{1-\eta}}}.$$

Hence,

$$\frac{\int y^{1-\eta} dF_r(y)}{\int y dF_r(y)} = \frac{\Gamma_{\frac{\zeta}{1-\eta}} \overline{w}_r^{1-\eta}}{\Gamma_\zeta \overline{w}_r} = \frac{\Gamma_{\frac{\zeta}{1-\eta}}}{\Gamma_\zeta} \overline{w}_r^{-\eta},$$

so that

$$\text{(A.6)} \qquad \vartheta_{rA} = \phi + \nu \frac{\Gamma_{\frac{\zeta}{1-\eta}}}{\Gamma_\zeta} \left(\frac{\overline{w}_r}{p_{rA}^{\phi} p_{rM}^{1-\phi}}\right)^{-\eta} = \phi + \nu^{RC} \left(\frac{\overline{w}_r}{p_{rA}^{\phi} p_{rM}^{1-\phi}}\right)^{-\eta},$$

where we defined the composite parameter $\nu^{RC} \equiv \nu \frac{\Gamma_{\frac{\zeta}{1-\eta}}}{\Gamma_\zeta}$.

Indirect Utility

Using the indirect utility function in equation (1), we derive the following expression for the, expected utility in region r:

$$E\left[V\left(y, P_{rA}, P_{rM}\right)\right] = \frac{1}{\eta}\left(\frac{1}{P_{rA}^{\phi} P_{rM}^{1-\phi}}\right)^{\eta} \int y^{\eta} dF_r(y) - \nu \ln\left(\frac{P_{rA}}{P_{rM}}\right).$$

Workers effectively draw their income from the Fréchet distribution in equation (A.5) upon arriving in their region of choice. We use the properties of the Fréchet distribution to show that y^η itself is drawn from a Fréchet distribution with a shape parameter ζ/η and scale $\overline{w}_r^{\eta}$:

$$P\left(y^{\eta} < m\right) = P\left(y < m^{\frac{1}{\eta}}\right) = e^{-\left(\frac{m^{\frac{1}{\eta}}}{\overline{w}_r}\right)^{-\zeta}} = e^{-\left(\frac{m}{\overline{w}_r^{\eta}}\right)^{-\frac{\zeta}{\eta}}}.$$

By implication, $\int y^{\eta} dF_r(y) = \Gamma\left(1 - \frac{1}{\zeta/\eta}\right) \overline{w}_r^{\eta} = \Gamma_{\frac{\zeta}{\eta}} \overline{w}_r^{\eta}$,so that

$$\text{(A.7)} \qquad E\left[V\left(y, P_{rA}, P_{rM}\right)\right] = \frac{1}{\eta} \Gamma_{\frac{\zeta}{\eta}} \left(\frac{\overline{w}_{rt}}{P_{rA}^{\phi} P_{rM}^{1-\phi}}\right)^{\eta} - \nu \ln\left(\frac{P_{rA}}{P_{rM}}\right).$$

This is the expression in equation (13).

A.4 Characterization of Equilibrium

The equilibrium is characterized by the following system of equations:

1. **Spatial labor supply:** The spatial labor supply function is given in the law of motion for the local population in equation (A.8) given by

$$L_{jt} = \sum_r m_{rjt} L^Y_{rt} = \sum_r m_{rjt} n_{rt-1} L_{rt-1}, \tag{A.8}$$

where m_{rjt} is given in equation (10). Together with the expression for expected utility $\mathcal{V}_{rt}$ given in (13), equation A.8 determines the spatial supply function as a function of local wages $\overline{w}_{rt}$ and local prices $\{P_{rAt}, P_{rMt}\}_r$.

2. **Labor market clearing in agriculture:** the agricultural labor market clears when labor demand (LHS) equals labor supply (RHS)

$$w_{rAt}^{-\frac{1}{\alpha}} Z_{rAt}^{\frac{1}{\alpha}} T_r = \Gamma_\zeta L_{rt} \left(\frac{w_{rAt}}{\overline{w}_{rt}}\right)^{\zeta-1}. \tag{A.9}$$

Equation A.9 determines the scaled skill prices in the agricultural sector, $w_{rA} = \frac{1}{1-\alpha}\tilde{w}_{rA}$, as

$$w_{rAt}^{\zeta-1+\frac{1}{\alpha}} = \overline{w}_{rt}^{\zeta-1} Z_{rAt}^{\frac{1}{\alpha}} \frac{T_r}{\Gamma_\zeta L_{rt}} \tag{A.10}$$

Market clearing for non-agricultural products: For non-agricultural products, sales of firm ω located in region r are given by

$$p_{rt}(\omega)\, y_{rt}(\omega) = \sum_j \left(\frac{\tau_{rjM} p_{rrt}(\omega)}{P_{jMt}}\right)^{1-\sigma} \vartheta_{jMt} \Gamma_\zeta L_{jt} \overline{w}_{jt}.$$

The mass of non-agricultural firms that enter a location, N_{rt}, is also equal to the number of varieties produced in region r. Aggregating over the measure of varieties, N_{rt} yields:

$$\frac{\sigma}{\sigma-1} w_{rMt} H_{rPt} = N_{rt} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} w_{rMt}^{1-\sigma} Z_{rt}^{\sigma-1} \sum_j \left(\frac{\tau_{rjM}}{P_{jMt}}\right)^{1-\sigma} \vartheta_{jMt} \Gamma_\zeta L_{jt} \overline{w}_{jt},$$

where we used that total payments to production workers are a constant fraction,

$\frac{\sigma-1}{\sigma}$, of total sales. We denote by H_{rPt} and H_{rEt} the total mass of non-agricultural workers engaged in production and entry, respectively, so that $H_{rPt} + H_{rEt} = H_{rMt}$. The mass of local varieties, N_{rt}, itself is determined from free entry as $N_{rt} = \frac{1}{f_E} H_{rEt} = \frac{1}{\sigma f_F} H_{rMt}$ and $H_{rPt} = \frac{\sigma-1}{\sigma} H_{rMt}$. Hence,

$$w_{rMt} = \frac{1}{\sigma f_F} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} w_{rMt}^{1-\sigma} Z_{rMt}^{\sigma-1} \sum_j \left(\frac{\tau_{rjM}}{P_{jMt}} \right)^{1-\sigma} \vartheta_{jMt} \Gamma_\zeta L_{jt} \overline{w}_{jt},$$

which implies that

$$w_{rMt}^{\sigma} = \frac{1}{\sigma f_F} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} Z_{rMt}^{\sigma-1} \mathcal{D}_{rt}, \tag{A.11}$$

where $\mathcal{D}_{rt} = \sum_j \tau_{rjM}^{1-\sigma} P_{jMt}^{\sigma-1} \left(1 - \vartheta_{jAt}\right) \Gamma_\zeta L_{jt} \overline{w}_{jt}$ was defined in equation (14) and the non-agricultural spending share ϑ_{rMt} is given in equation A.6.

These equations fully determine the equilibrium. In particular, upon substituting for ϑ_{rMt} and $\mathcal{V}_{rt}$, equations (A.8), (A.9), and (A.11) are $3 \times R$ equations in the $3 \times R$ unknowns $\{w_{rAt}, w_{rMt}, L_{rt}\}$.

A.5 Additional Proofs and Derivations

A.5.1 Proof of Proposition 1

First, rewrite equation (A.11) as follows

$$w_{rMt} = \left(\frac{1}{f_E} \right)^{\frac{1}{\sigma}} \left(\frac{1}{\sigma} \right) (\sigma - 1)^{\frac{\sigma-1}{\sigma}} Z_{rMt}^{\frac{\sigma-1}{\sigma}} \mathcal{D}_{rt}^{\frac{1}{\sigma}} \equiv \mathscr{Z}_{rMt}, \tag{A.12}$$

which is the first result in Proposition (15). Furthermore, upon defining $\mathscr{Z}_{rAt} \equiv Z_{rAt} \left(\Gamma_\zeta \ell_{rt}\right)^{-\alpha}$ where $\ell_{rt} \equiv \frac{L_{rt}}{T_r}$ is population density, equation (A.10) reads

$$w_{rAt}^{\zeta-1+\frac{1}{\alpha}} = \overline{w}_{rt}^{\zeta-1} \mathscr{Z}_{rAt}^{\frac{1}{\alpha}} = \left(w_{rAt}^{\zeta} + w_{rMt}^{\zeta} \right)^{\frac{\zeta-1}{\zeta}} \mathscr{Z}_{rAt}^{\frac{1}{\alpha}} = \left(w_{rAt}^{\zeta} + \mathscr{Z}_{rMt}^{\zeta} \right)^{\frac{\zeta-1}{\zeta}} \mathscr{Z}_{rAt}^{\frac{1}{\alpha}}.$$

Rearranging terms yields

$$1=\left(1+\left(\frac{\mathscr{Z}_{rMt}}{w_{rAt}}\right)^{\zeta}\right)^{\frac{\zeta-1}{\zeta}}\left(\frac{\mathscr{Z}_{rAt}}{w_{rAt}}\right)^{\frac{1}{\alpha}},$$

which is the second result in Proposition 1. To derive the last result in Proposition 1, note that sectoral employment shares satisfy $s_{rAt}/\left(1-s_{rAt}\right)=\left(w_{rAt}/w_{rMt}\right)^{\zeta}$. As a result, equation (A.10) can be written as $w_{rAt}=s_{rAt}^{-\frac{\zeta-1}{\zeta}\alpha}\mathscr{Z}_{rAt}$, so that

$$\frac{s_{rAt}}{1-s_{rAt}}=\left(\frac{s_{rAt}^{-\frac{\zeta-1}{\zeta}\alpha}\mathscr{Z}_{rAt}}{\mathscr{Z}_{rMt}}\right)^{\zeta}=\left(\frac{\mathscr{Z}_{rAt}}{\mathscr{Z}_{rMt}}\right)^{\zeta}s_{rAt}^{-(\zeta-1)\alpha}.$$

Rearranging terms yields $\frac{s_{rAt}^{1+(\zeta-1)\alpha}}{1-s_{rAt}}=\left(\frac{\mathscr{Z}_{rAt}}{\mathscr{Z}_{rMt}}\right)^{\zeta}$.

A.5.2 Proof of Proposition 2

The wage exposure elasticity $\phi\left(s_{rA}\right)$ The first result in Proposition Proposition 1, directly implies the following:

$$d\ln w_{rMt}=d\ln\mathscr{Z}_{rMt}. \tag{A.13}$$

Taking the total derivative of the expression for the agricultural wage in Proposition 1yields:

$$d\ln w_{rAt}=\frac{\left(1-s_{rAt}\right)\gamma}{1+\left(1-s_{rAt}\right)\gamma}d\ln\mathscr{Z}_{rMt}+\frac{1}{1+\left(1-s_{rAt}\right)\gamma}d\ln\mathscr{Z}_{rAt}. \tag{A.14}$$

Finally, we can take the total derivative for the expression for the average income in a location in equation (A.4), and combine it with equations (A.13) and (A.14) to obtain:

$$\begin{aligned}
d\ln\overline{w}_{rt} &= s_{rAt}d\ln w_{rAt}+\left(1-s_{rAt}\right)d\ln w_{rMt}\\
&= \frac{s_{rAt}\left(1-s_{rAt}\right)\gamma}{1+\left(1-s_{rAt}\right)\gamma}d\ln\mathscr{Z}_{rMt}+\frac{s_{rAt}}{1+\left(1-s_{rAt}\right)\gamma}d\ln\mathscr{Z}_{rAt}+\left(1-s_{rAt}\right)d\ln\mathscr{Z}_{rMt}\\
&= \frac{\left(1+\gamma\right)\left(1-s_{rAt}\right)}{1+\left(1-s_{rAt}\right)\gamma}d\ln\mathscr{Z}_{rMt}+\frac{s_{rAt}}{1+\left(1-s_{rAt}\right)\gamma}d\ln\mathscr{Z}_{rAt}\\
&\equiv \phi\left(s_{rAt}\right)d\ln\mathscr{Z}_{rMt}+\left(1-\phi\left(s_{rAt}\right)\right)d\ln\mathscr{Z}_{rAt}, \qquad \text{(A.15)}
\end{aligned}$$

where $\phi\left(s_{rAt}\right) = \frac{(1+\gamma)(1-s_{rAt})}{1+(1-s_{rAt})\gamma}$.

The industrialization elasticity $\psi\left(s_{rA}\right)$ To derive the change in s_{rAt}, we take the total derivative of equation (A.4) and combined it with (A.15) to obtain:

$$d\ln s_{rAt} \quad = \quad \zeta\left(d\ln w_{rAt} - d\ln \overline{w}_{rt}\right) = \zeta\left(1 - s_{rAt}\right)\left(d\ln w_{rAt} - d\ln w_{rMt}\right).$$

Using the expressions in equations (A.14) and (A.13), this implies that

$$d\ln s_{rAt} = \frac{\left(1 - s_{rAt}\right)\zeta}{1 + \left(1 - s_{rAt}\right)\gamma}\left(d\ln \mathscr{Z}_{rAt} - d\ln \mathscr{Z}_{rMt}\right). \tag{A.16}$$

Finally, using that $ds_{rAt} = s_{rAt} d\ln s_{rAt}$,yields the expression in Proposition 2.

B. Data and Calibration

The material presented in this section complements the quantification section of the main paper. It contains a detailed description of the data, additional figures and tables, and details of our estimation procedure.

B.1 Description of Data Sources and Data Construction

The spatial unit of observation used throughout the paper is the "commuting zone" defined by Tolbert and Sizer [1996]. These were introduced into the economics literature by Autor and Dorn [2013]. We choose these units since they capture integrated labor market areas within which migration frictions are unlikely to play a role. During the period of our study, county boundaries were subject to substantial changes. To ensure consistent treatment, we use the crosswalk described in Eckert et al. [2020b] to map historical county boundaries to the time-invariant commuting zone delineations of Tolbert and Sizer [1996]. Since data collection by the US Statistical Office only occurred systematically in states that formed part of the Union, we drop data from states that joined the Union after 1870. We chose a cutoff a decade before the start of our period of analysis since the collection of the decennial census took considerable time, so data in the 1880 Census may be incomplete for States that joined the Union less than ten years earlier. As a result, we exclude the following states from our sample and drop them from all our data sets (year of eventual accession to the Union in parentheses): Colorado (1876), North Dakota (1889), South Dakota (1889), Montana (1889), Washington (1889), Idaho (1890), Wyoming (1890), Utah (1896), Oklahoma (1907), New Mexico (1912), Arizona (1912), Alaska (1959), Hawaii (1959). Figure A.1 shows agricultural employment shares in 1880 for each commuting zone in our final sample.

B.1.1 Full Count Decennial Census, 1880-1920

Source and Description We obtained the full count decennial census micro-data files for the years 1880, 1900, and 1920 from the IPUMS database (see Ruggles et al. [2017]). We selected the following variables: state, county, age , school attendance ("school"), years since immigration ("yrimmig"), state of birth (if applicable), and industry of employment using 1950 Census codes (ind1950),. We use the county and state identifiers included in the data to assign each observation to a commuting zone.

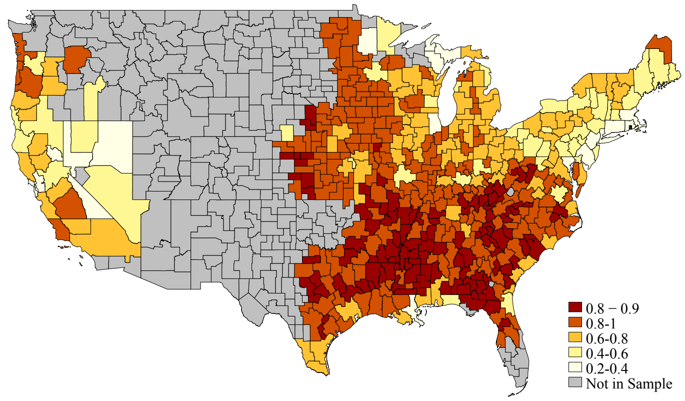
Sample Selection, Processing, and Use In the data, we define different groups of observations used in various parts of the paper. We define "workers" as observations with an industry identifier and age between 20 and 60 years. We define "agricultural workers" as workers who work in Agriculture, Forestry, and Fishing, corresponding to ind1950 codes 105, 116, and 126. For each commuting zone, dividing the total agricultural worker count by the total number of workers yields the agricultural employment share we use throughout the paper. In Figure A.1, we depict a map of the agricultural employment share in 1880.

"Immigrant workers" are workers who immigrated within the last 20 years. "Old workers" are workers between the ages of 40 and 60. "Young workers" are workers between the ages of 20 and 40. We use these groups of observations to inform the location- and decade-specific labor force growth rate n_{rt}.

"Adolescents" as observations with age between 6 and 18 years who do not work, are white, and are male. We define "adolescents in school" as adolescents who are currently attending school. We use these two groups of workers to compute a measure of the rate of school attendance for each commuting zone.

For each state, we also compute the number of workers born in any state. We use the resulting "lifetime state-to-state migration matrix" to estimate the elasticity of migration flows to distance.

FIGURE A.1: AGRICULTURAL EMPLOYMENT SHARES ACROSS COMMUTING ZONES, 1880



Notes: The map shows all commuting zones in the United States. The colors reflects the agricultural employment share bin into which individual commuting zones fall. Darker shades correspond to higher agticultural employment shares. Grey commuting zones are not in our sample since their corresponding state was not part of the Union of US states by 1870.

B.1.2 Census of Manufacturing

Source and Description We obtained county-level tabulations of the Census of Manufacturing data for the years 1880, 1900, and 1920 from the NHGIS database (see Manson et al. [2017]). We selected the following variables: total manufacturing payroll, total manufacturing employment, number of manufacturing establishments, and capital (real and personal) invested in iron and steel manufacturing establishments.

Sample Selection, Processing, and Use We drop all counties for which manufacturing payroll or employment is zero or missing. We then compute average manufacturing wages in each county by dividing total manufacturing payroll by the number of manufacturing workers. Throughout the paper, we refer to this ratio simply as "average wage" or "earnings ".We compute commuting zone-level average wages by taking the payroll-weighted average across county-level average wages within each commuting zone. In our model average wages are the same in both sectors, so that the average manufacturing wage in the data correspond to the average commuting zone wage in the model, $\overline{w}_{rt}$. We compute average establishment size for each county by dividing total manufacturing employment by the number of manufacturing establishments.

B.1.3 Census of Agriculture

Source and Description We obtained county-level tabulations of the Census of Agriculture for the years 1880, 1900, and 1920 from the NHGIS database (see Manson et al. [2017]). We selected the following variables: average land value per acre, acres of improved farm land, total number of farms, and total value of farm implements and machinery.

Sample Selection, Processing, and Use We drop all counties for which average land value per acre is zero or missing. We compute commuting zone-level average land values per acre by taking the area-weighted average across the land values in all the counties contained in a given commuting zone. We interpret these data in 1880 as land rents in the model and use them to identify the supply of agricultural land in each commuting zone in 1880, T_r. We compute average farm size for each county by dividing the total number of improved acres in farms by the total numbers of farms.

B.1.4 Linked Census Files

Source and Description Economists have written algorithms to match workers across sequential Decennial Census waves based on their names and a variety of other characteristics. IPUMS itself provides a matched file that lists individuals that appear both in 1880 and 1900 (see Ruggles et al. [2017]). In addition, Abramitzky et al. [2022] provide linked files for various pairs of years. We use the 1880-1900 linked file from IPUMS and from Abramitzky et al. [2022].

Sample Selection, Processing, and Use Both samples only include men, since women's surnames changed frequently making it difficult to match them over time. We only keep observations who are workers according to our definition of workers in the full count Census files.

We use the linked data to compute the share of workers moving from commuting zone r to commuting zone r' between 1880 and 1900. We use the resulting "commuting-zone-to-commuting-zone-migration matrix" to estimate the elasticity of migration flows to distance.

B.1.5 Historical Statistics of the United States

Source and Description For aggregate time series data, we use the canonical "Historical Statistics of the United States" (see Carter et al. [2006]). We use the series on real GDP and the series for the price of farm goods and the prices of all commodities other than farm goods.

Sample Selection, Processing, and Use Moments from both the GDP and the price series serve as targets in our estimation. We interpret the price series for farm goods as the price series for agricultural prices in our model, and the series on non-farm commodities as that of manufacturing goods. We target the growth rate of real GDP and relative prices between 1880-1900 and 1900-1920 in our estimation.

B.1.6 Historical Bank Branches

Source and Description We obtained data on the number of private banks for each US county for 1880 and 1910 from Jaremski and Fishback [2018].[25]

Sample Selection, Processing, and Use We merge these data with our Census data on the number of workers in each commuting zone to compute the change in the log of the number of bank branches per worker in each commuting zone. Since there are no branches in many commuting zones in 1880, we add a 1 to each observation in the bank branch data. Our results are robust to simply dropping observations with zero branches in 1880 instead.

B.2 Spatial Structural Change Across Counties

Figure A.2 replicates the patterns of structural change across commuting zones in Section 1 across counties. Since we aggregated the underlying Census data from the county to the commuting zone level, the data source for the two sets of figures is identical and the only differences is the level of spatial aggregation. Overall, the patterns of spatial structural change across counties are qualitatively very similar to those patterns across commuting zones in Section 1.

[25] We thank Matt Jaremski for his work in compiling these data and his generosity in sharing them with us.

The top left panel shows the aggregate time series of the aggregate agricultural employment share, as before. The top right panel shows a gradual shift of the cross-county distribution of agricultural employment shares to the left. The bottom left panel shows that agricultural counties in 1880 were poor counties, again mirroring the patterns we documented across commuting zones. Finally, the bottom right panel shows that wage growth between 1880 and 1920 was faster in initially more agricultural counties, and industrialization was fastest in counties in the intermediate range of agricultural specialization.

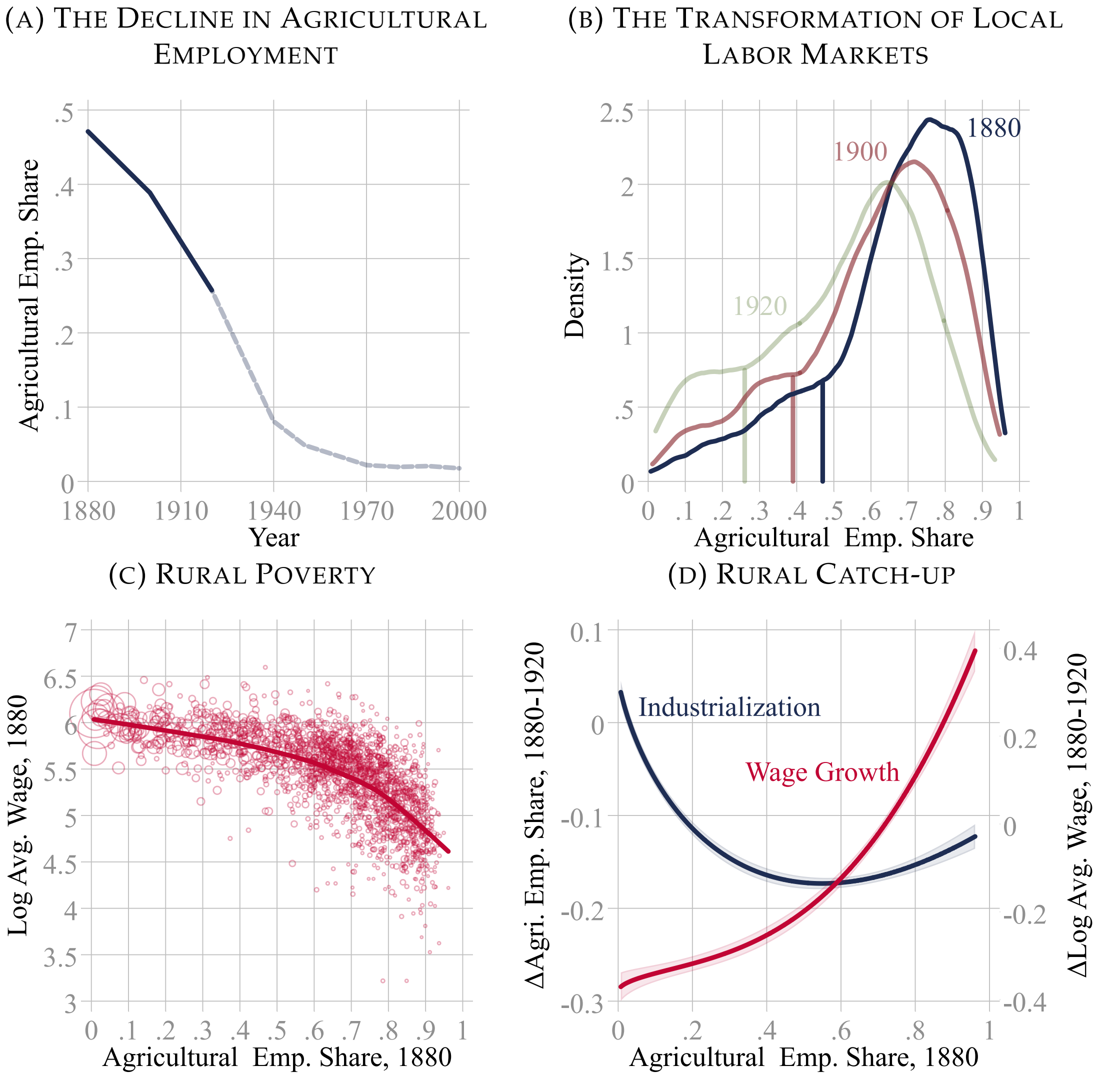
B.3 Details on Estimation Moments and Methods

B.3.1 Wage Growth, Industrialization, and Local Labor Supply

Four central parameters governing the impact of aggregate structural change are the extent of catch-up growth (λ_A, λ_M) , the sectoral labor supply elasticity ζ., and the spatial labor supply elasticity, ϵ. We estimate these parameters from the empirical relationships between initial regional specialization in agriculture in 1880 and the subsequent growth in regional wages, changes in agricultural employment shares, and total employment growth.. We take an indirect inference approach, that is we run the same regressions in model-generated data and the actual data and choose parameters to match the regression coefficients.

To implement our estimation strategy, we first document the patterns of structural change from Section 1 in regression form. The top panel of Table A.1 reports the facts from Figure 2 in a regression form and adds an additional regression for population growth. The first two columns show a regression of wage growth. The first column shows a regression of wage growth from 1880-1900 and 1900-1920 on 1880 and 1900 agricultural employment shares, respectively. The second column repeats the regression but adds state fixed effects. The two regressions corroborate the fact shown in Figure 2 in the main text: initially more agricultural regions saw faster wage growth than initially more industrialized regions. Columns 3 and 4 show the regression version of the "industrialization across space" fact in Figure 2 in the main text: industrialization exhibits a U-Shape when graphed against initial agricultural employment shares. We capture the U-Shape by regressing changes in agricultural employment shares between 1880-1900 and 1900-1920 on 1880 and 1900 agricultural employment shares in levels and squares, respectively. The last two columns run two similar regression for total employment growth as an outcome variable. As shown in the text, initially more agricultural regions see slower employment growth but at the

FIGURE A.2: SPATIAL STRUCTURAL CHANGE ACROSS COUNTIES



Notes: The top left panel shows the aggregate agricultural employment share in the US between 1880 and 2015. The top right panel shows the distribution of agricultural employment shares across counties in 1880, 1900, 1920, the years of our analysis. The bottom left panel shows a scatter plot between commuting zones' agricultural employment shares and average earnings in 1880 and a Lowess fit line. The size of the points is proportional to the total workforce in each commuting zone. The bottom right panel shows two fitted fractional polynomial curves along with 95% confidence intervals. They show the relationship between commuting zones' agricultural employment share in 1880 and (1) their average earnings growth between 1880 and 1920 (left axis), (2) the change in the agricultural employment share between 1880 and 1920 (right axis). In fitting the polynomials, we weight by commuting zones' total employment in 1880. The figure uses data from the US Decennial Census files for 1880, 1900, and 1920 and relies on data from the US Census of Manufacturing for the earnings data.

same time they experience faster wage growth.

All three patterns hold both across commuting zones, and across commuting zones within each state. The wage and employment growth patterns are stronger when estimated within state, whereas the U-Shape of industrialization appears very similar in regressions with and without state fixed effects.

The second panel of Table A.1 repeats all regressions across counties instead of commuting zones. We use the exact same data in both panels, with the only difference that for the first panel we aggregate the data from the county-level (at which much of it is reported) to the commuting zone level. The patterns of spatial structural change across counties are quantitatively and qualitatively very similar to those across commuting zones, especially in regressions with state fixed effects. The U-Shape of industrialization is slightly less pronounced, reflecting maybe that part of the sectoral reallocations that gives rise to it occurred across counties within commuting zones, and are weaker within counties themselves.

As part of our calibration, we run the regression in odd-numbered columns of the top panel of Table A.1 in our model. While all parameters are calibrated jointly, the coefficient on the agricultural employment share in Column 1 is informative about the strength of agricultural catch-up growth, λ_A. Since agricultural regions, on average, have low agricultural productivity to start with but have the majority of their population in the agricultural sector, the larger λ_A the more wage growth these regions experience.

Similarly, the strength of catch-up growth in non-agriculture, λ_M, informs the coefficient on the agricultural employment share in Column 3: if agricultural regions only saw catch-up un agricultural productivity their wages would grow, but they would not industrialize, maybe even de-industrialize. If they see their non-agricultural productivity grow faster than that of other regions, reallocating workers out of agriculture becomes profitable. Lastly, the sectoral labor supply elasticity, ζ, generates the U-shape of industrialization in the model. Intuitively, if labor supply is very inelastic, the U-shape is less pronounced and regional differences in sectoral reallocation are small since reallocation is hard. We use the decline in agricultural employment shares among regions with initial agricultural employment shares of above 80% as an additional moment for ζ.

Lastly, the spatial labor supply elasticity, ϵ, determines the relationship between wage growth and employment growth in the model. If supply is inelastic, large wage growth differences across regions lead to small differences in population growth. Since agricultural regions, on average, see faster wage growth, but slower employment growth, we

TABLE A.1: WAGE GROWTH, INDUSTRIALIZATION, AND EMPLOYMENT GROWTH ACROSS REGIONS: 1880-1900 AND 1900-1920

	PANEL A: COMMUTING ZONES					
	$\Delta \log \bar{w}_{rt}$		Δs_{rAt}		$\Delta \log L_{rt}$	
s_{rAt}	0.251***	0.357***	-0.484***	-0.466***	-0.360***	-0.753***
	(0.0220)	(0.0381)	(0.0279)	(0.0342)	(0.0309)	(0.0509)
s^2_{rAt}			0.451***	0.427***		
			(0.0317)	(0.0386)		
Observations	990	990	990	990	990	990
Adjusted R^2	0.839	0.855	0.309	0.362	0.167	0.315
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
State FEs		Yes		Yes		Yes
	PANEL B: COUNTIES					
	$\Delta \log \bar{w}_{rt}$		Δs_{rAt}		$\Delta \log L_{rt}$	
s_{rAt}	0.316***	0.334***	-0.371***	-0.379***	-0.510***	-0.765***
	(0.0132)	(0.0187)	(0.0145)	(0.0159)	(0.0164)	(0.0221)
s^2_{rAt}			0.344***	0.344***		
			(0.0167)	(0.0187)		
Observations	3918	3918	3918	3918	3918	3918
Adjusted R^2	0.704	0.720	0.194	0.237	0.221	0.328
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
State FEs		Yes		Yes		Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The outcome variables are average annual wage growth (Columns 1 and 2), change in agricultural employment shares (Columns 3 and 4), and growth in total employment for US commuting zones between 1880-1900 and 1900-1920. The regressor in all regressions is the agricultural employment share in 1880 and 1900, respectively, in all regressions. In Columns 3 and 4, the additional regressor is the squared agricultural employment share in 1880 and 1900, respectively.All regression are weighted by initial total employment of the commuting zone. and include decade fixed effects. Even-numbered Columns also include a state fixed effect.Data on wages from the Census of Manufacturing, all other data from the full-count US Decennial Census files. Robust standard errors in parentheses. * , **, and *** denote statistical significance at the 10%, 5% and 1% level respectively.

estimate labor supply to be relatively inelastic to local wage growth. As an additional moment informing the labor supply elasticity, we target estimates of the labor supply elasticity from the literature of 2 (see, e.g., Monte et al. [2018]).

B.3.2 Exogenous Growth in the Local Labor Force: The Bias of Immigration and Fertility Across Space

Our model accounts for growth in the local labor force through interregional migration. Empirically, other factors affecting the size of the local labor force are births, immigration, and deaths, all of which likely differ across commuting zones.These aspects of labor force entry and exit also generate aggregate population growth. In this section, we show which determinants of local labor force growth vary substantially across commuting zones and how we account for them in our analysis.

Figure A.3 shows proxies for the three most important sources of local population growth: births, immigration, and deaths. The rightmost panel shows the number of children per adult ("birth rates"). We measure local "birth rates" as the fraction of children between 0 and 20 relative to the number of working adults aged 20-60. Rural locations have substantially higher birth rates and hence experience faster innate population growth.

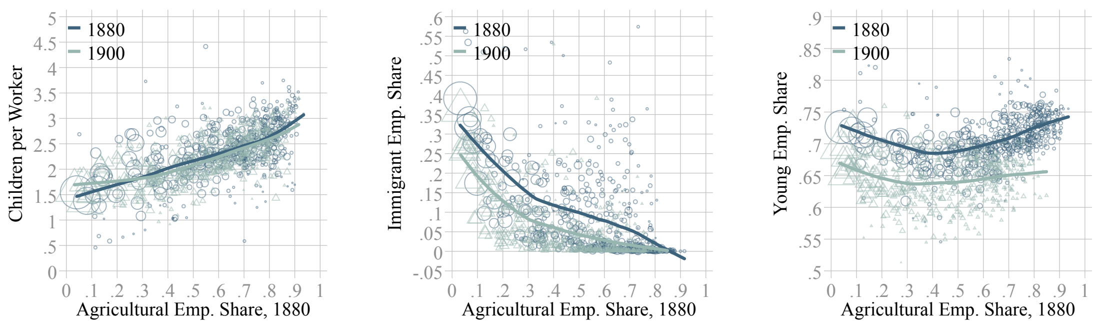
The middle panel shows the the correlation of the share of immigrants in the local workforce and initial agricultural employment shares. We measure immigrants as the share of workers that immigrated in the last 20 years from a foreign country. Immigrants predominantly settled in urban locations, and thus raised the population of such non-agricultural locations.

The rightmost panel of Figure A.3 provides evidence that - compared to births and immigration - labor force exit rates do not vary systematically across space. If death and retirement rates varied substantially across regions, the fraction of young workers (20-40 years old) in the total workforce should vary a lot, too. However, the figure shows that the fraction of young workers is essentially uncorrelated with the local agricultural employment share. We thus assume that the rate of labor force exit is constant across locations.

We now show how we use these data to estimate the exogenous component of local population growth n_{rt} in each region. To do so, recall that we denote by L_{rt}^{Y} the number of workers in a location at the beginning of period t, i.e., before making their moving decisions. L_{rt} is the number of workers working in region r during period t, i.e., the mass of workers that chose to move (or remain in) location r during period t. The local rate of

FIGURE A.3: IMMIGRATION, FERTILIY, AND AGE STRUCTURE ACROSS COMMUTING ZONES

(A) FERTILITY (B) IMMIGRATION (C) AGE



Notes: The left panel shows a proxy for the local birth rate, i.e., the share of children between 0 and 20 relative to the number of working adults aged 20-60, as a function of the initial agricultural employment share. The middle panel shows the share of immigrants in the local work force as function of the initial agricultural employment share. The right panel shows the fraction of young workers among the total workforce in each commuting zone. Total workers are defined as all individuals aged 20-60 that have an industry identifier. Young workers are defined as all individuals aged 20-40 that have an industry identifier. The underlying data source for all three panels are the US Decennial Census files for 1880 and 1900. The size of the symbols is proportional to a regions total employment. The graph shows the fit line of a local polynomial regression.

exogenous labor force growth, n_{rt}, is thus defined by $L^Y_{rt+1} = n_{rt}L_{rt}$. To calibrate n_{rt}, note that the following accounting identity describes the law of motion of the total labor force in region r at the beginning of period t:

$$L^Y_{rt} = L_{rt-1} - Exit_{rt-1,t} + Entry_{rt-1,t} = L_{rt-1}\left(1 - \frac{Exit_{rt-1,t}}{L_{rt-1}} + \frac{Entry_{rt-1,t}}{L_{rt-1}}\right),$$

where $Exit_{rt-1,t}$ is the number of workers exiting the labor force between periods $t-1$ and t but do not leave the location to work elsewhere. Similarly, $Entry_{rt-1,t}$ is the number of workers entering the labor force between periods $t-1$ and t that did not immigrate from another domestic region between $t-1$ and t.

Given our assumption of a constant labor force exit rate across regions, we set the exit rate equal to a common constant, δ, so that:

$$\frac{Exit_{rt-1,t}}{L_{rt-1}} = \delta.$$

The gross rate of local labor force growth prior to workers making their migration decisions is thus given by

$$n_{rt-1} = \frac{L^Y_{rt}}{L_{rt-1}} \quad = \quad 1 - \delta + \frac{Entry_{rt-1,t}}{L_{rt-1}}.$$

Let C_{rt} denote the number of children in r at time $t-1$ and I_{rt} denote the number of working immigrants in region r that arrived between $t-1$ and t. Let ι be the fraction of children that join the labor force. Since we assume differences in entry rates to be due to differences in fertility rates and immigration only, we relate C_{rt} and I_{rt} to $Entry_{rt-1,t}$ according to

$$\frac{Entry_{rt-1,t}}{L_{rt-1}} = x \times \frac{\iota C_{rt} + I_{rt}}{L_{rt-1}},$$

where x is a scalar that reflects measurement error, e.g., some children die, time is not discrete (i.e., the 16 year old children enter the labor market earlier than the 5 yr old children), or immigrants might move across locations within the US in-between Census years. Then

$$n_{rt-1} \quad = \quad 1 - \delta + x \times \frac{\iota C_{rt} + I_{rt}}{L_{rt-1}},$$

where C_{rt}, I_{rt} and L_{rt-1} are observed in the data.

We choose the scalar x to ensure that this accounting equation satisfies the *aggregate* rate of employment growth in the Census, that is we ensure that the following equation holds in the data:

$$\text{Total employment at } t = \sum_r \left(1 - \delta + x\frac{\iota C_{rt} + I_{rt}}{L_{rt-1}}\right) L_{rt-1}.$$

Rearranging terms implies that

$$x = \frac{\text{Total employment in t} - (1-\delta)\,\text{Total employment in t-1}}{\sum_r \left(\iota C_{rt} + I_{rt}\right)}.$$

Hence, for a given exit rate δ and labor force participation rate ι we pick the scale x for the aggregate birth and immigration inflow to account for all employment growth. And then we use this x to calculate - in the model - the number of workers in region r prior to their migration choices as

$$L^Y_{rt} \quad = \quad n_{rt-1} L_{rt-1} = L_{rt-1}\left(1-\delta\right) + x\left(\iota C_{rt} + I_{rt}\right).$$

Hence, local labor force growth prior to worker's migration choices depends on the observable $(\iota C_{rt} + I_{rt})$ and it has the correct slope for our model to be consistent with aggregate population growth.

To pick the exit rate δ, note that the fraction of old workers at time t is given by

$$\text{(A.17)} \qquad \text{Share of old workers}_t \quad = \quad \frac{(1-\delta)\sum_r L_{rt-1}}{\sum_r L_{rt}^Y} = (1-\delta)\frac{\sum_r L_{rt-1}}{\sum_r L_{rt}}.$$

Because $\frac{\sum_r L_{rt-1}}{\sum_r L_{rt}}$ is simply the ratio of the total labor force at $t-1$ divided by the total labor force at t, which are both observed, we can calculate δ for any target of the share of old workers. A generation in our model corresponds to 20 years in the data. In calibrating δ, we think of 0-20 year olds as not working, of 20-40 year olds as "young" workers, and of "40-60" year olds as "old workers." The share of old workers in our data is 0.34, 0.35 and 0.37 in 1880, 1900, and 1920, respectively. Because, empirically, some people above 60 are still in the workforce, we take a number of 0.45. Together with a rate of population growth of about 35% observed in the data (at the 20 year horizon), equation (A.17) implies that δ is given by 0.4.

To calibrate ι, we combine the mortality rate of children with the rate of labor force participation among 20-40 year olds in 1900. which is about 0.5 in our data, which comprises both men and women. As a result, we set $\iota = 0.5$.

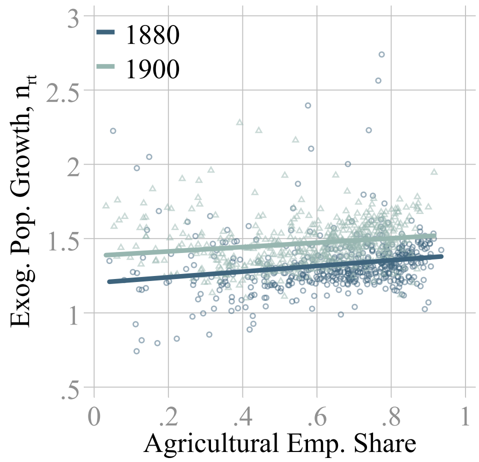
In Figure A.4, we show the calibrated exogenous rate of population growth, n_{rt}, for the two time periods 1880-1900 and 1900-1920. The figure shows that, on net, exogenous population growth was slightly higher in agricultural regions. The relationship between agricultural specialization and subsequent exogenous population growth weakens somewhat over these periods suggesting population growth became somewhat more balanced as fertility rates in more rural regions started to decline.

B.3.3 Migration Gravity Equations: Estimating the Distance Elasticity of Moving Costs

In this section, we describe our estimation of the distance elasticity of migration costs, κ. In the model, the mass of workers migrating from region r to region r' between two periods is given by:

$$M_{rr't} = m_{rr't} L_{rt}^Y = \frac{\left(\mu_{rr'} \mathcal{V}_{r't} \mathcal{B}_{r't}\right)^{\varepsilon}}{\sum_j \left(\mu_{rj} \mathcal{V}_{jt} \mathcal{B}_{jt}\right)^{\varepsilon}} L_{rt-1} n_{rt-1}.$$

FIGURE A.4: EXOGENOUS POPULATION GROWTH



Notes: This figures shows the calibrated rate of exogenous population growth across commuting zones between 1880-1900 and 1900-1920. In each year there is one region with a rate above 3 which we drop to shows the variation among the remaining observations in more detail.

We project the moving cost between two regions on the physical distance between them, i.e., we set $\mu_{rr'} = d_{rr'}^{-\kappa}$., where the parameter κ parameterizes the distance cost of migration. The larger κ, the more the destination utility of areas furhter away is discounted. In our empirical estimation, we set $d_{rr'}\ \forall r = r'$ to the average distance between county centroids within a commuting zone, and $d_{rr'}\ \forall r \neq r'$ to the distance between commuting zone centroids.

Taking logs on both sides and grouping terms then yields:

$$\log M_{rr't} = \alpha_{r't} + \beta_{rt} - \kappa\epsilon \log d_{rr'} \tag{A.18}$$

where

$$\alpha_{r't} = \varepsilon \log(\mathcal{V}_{r't}\mathcal{B}_{r't}) \quad \text{and} \quad \beta_{rt} = \log(L_{rt-1}n_{rt-1}) - \log(\sum_{r''} \left(d_{rr''}^{-\kappa}\mathcal{V}_{r''t}\mathcal{B}_{r''t}\right)^{\varepsilon}).$$

Since we calibrate the model at the commuting zone level, the indices r and r' refer to commuting zones. Equation A.18 suggests a fixed effect regression of commuting zone migration flows to recover to recover the elasticity of migration flows to distance, $\kappa\epsilon$, relevant in our model.

The Decennial Census files do not contain information on workers' migration history at the commuting zone or county level. Hence, it is impossible to directly construct cross-commuting zone migration flows. We therefore rely on information from the linked Cen-

TABLE A.2: MIGRATION GRAVITY EQUATIONS

	LN BILATERAL MIGRATION PROBABILITY				
	1880-1900		1880(20)1920		
Log Distance	-2.922***	-2.632***	-3.925***	-2.262***	-2.291***
	(0.0327)	(0.0111)	(0.0572)	(0.0306)	(0.0311)
Observations	254762	349230	3983	3983	3935
(Pseudo) R^2	0.8135	0.8989	0.9296	0.806	0.799
Year FE			✓	✓	✓
Origin+Dest. FE	✓	✓	✓	✓	✓
Estimator	PPML			OLS	
Spatial Unit	CZ	CZ	State	State	State
Linking Method	IPUMS	ABE	NA	NA	NA

Notes: (1) PPML with census data from IPUMS linked by IPUMS. (2) PPML with census data from IPUMS linked by Abramitzky Boustan Eriksson. (3) PPML in state flow data from IPUMS, pooled across all years. (4) OLS regression in state flow data from IPUMS adding a 1 to all flows, pooled across all years. (5) OLS regression in state flow data from IPUMS dropping zero flow observations, pooled across all years.

sus files described in our data section above. Since linking rates are relatively low, the majority of bilateral commuting zone pairs exhibit *no* migration flows between 1880 and 1900. We hence estimate equation (A.18) using Poisson Pseudo Maximum Likelihood (PPML), as proposed by Silva and Tenreyro [2006]. More specifically, we estimate the following equation using PPML:

$$M_{rr't} = \exp\left(\alpha_{r't} + \beta_{rt} - \kappa\epsilon \log d_{rr'}\right) + \epsilon_{rr't}. \tag{A.19}$$

Columns (1) and (2) in Table A.2 report the estimates based on two different linked-Census files by Ruggles et al. [2015] ("IPUMS") and Abramitzky et al. [2021] ("ABE"). These files differ slightly in their technique to link individuals across census years. Reassuringly, both produce similar estimates: we estimate an elasticity of migration flows with respect to geographic distance ($\kappa\varepsilon$) of around 2.75.

Linking data across census years requires a set of assumptions and large amounts of data processing. For robustness, we therefore repeat the estimation on a different data set that we can directly compute from the cross-sectional Census data but that only contains state-to-state flows. In particular, as discussed in the data section above, we use the information on the state of birth of each worker contained in the Decennial Census files to construct a matrix of lifetime state-to-state migration flows for all workers between 20 and 40. Column 3 of Table A.2 presents the PPML estimates of the distance elasticity in the state-to-state data. The estimate is larger than in the commuting zone data highlighting that there are, by construction, less flows across states than across commuting zones making distance appear as a larger impediment of migration.

The state-by-state migration matrix has very few pairs of states with zero flows. Across the two cross-sections of data for 1880-1900 and 1900-1920, about 50 pairs exhibit zero flows.As a result, we can also estimate the gravity regression using simple OLS. instead of PPML. Columns 4 and 5 in Table A.2 report estimates from a regression where we simply replace zeros with 1s and another where we omit all zero-valued pairs of states. Since states further apart are more likely to report zero flows *because* they are further apart, dropping them leads to a smaller estimate of the elasticity at 2.62, the lowest of all our estimates.

Note that our theory only produces an approximate gravity equation for the flows between groups of commuting zones (such as states) because of Jensen's inequality. As a result, the distance elasticities stemming from state-level data do not map directly to the structural parameters $\kappa\varepsilon$. Nevertheless, we find it re-assuring that the state-level regressions estimates are not too dissimilar from the regression estimates using commuting zone. Furthermore, since a fraction of moves in the model happen across commuting zones within the same state, we would expect the state-to-state distance coefficient to be larger than the commuting zone to commuting zone one in model-generated data, too.

B.3.4 Computing Macroeconomic Aggregates

Aggregate GDP Growth

We measure the growth rate of aggregate GDP using the Fisher chained index. The Fisher Index is defined by

$$g_t^F = \sqrt{\frac{P_{ct-1}C_t}{p_{ct-1}C_{t-1}} \times \frac{P_{ct}C_t}{P_{ct}C_{t-1}}},$$

where P_{ct} is the price of the consumption good at time t and C_t is the quantity.

In the context of our model with R regions and two sectors s, we construct the following auxiliary indices:

$$S_{t-1}\left(P_{t-1}\right)=\sum_{r=1}^{R}\sum_{s=1}^{S}P_{rst-1}C_{rst-1}\quad\text{and}\quad S_{t-1}\left(P_{t}\right)=\sum_{r=1}^{R}\sum_{s=1}^{S}P_{rst}C_{rst-1}.$$

and

$$S_{t}\left(P_{t-1}\right)=\sum_{r=1}^{R}\sum_{s=1}^{S}P_{rst-1}C_{rst}\quad\text{and}\quad S_{t}\left(P_{t}\right)=\sum_{r=1}^{R}\sum_{s=1}^{S}P_{rst}C_{rst}.$$

where $P_{rst}(C_{rst})$ denotes the prices (consumption quantities) of sector s goods in region r at time t.

We then combine these expressions into the corresponding Fisher Index as follows:

$$g_t^F=\sqrt{\frac{S_t\left(P_{t-1}\right)}{S_{t-1}\left(P_{t-1}\right)}\times\frac{S_t\left(P_t\right)}{S_{t-1}\left(P_t\right)}}.$$

In our model, C_{rst} can be computed as

$$\begin{aligned}C_{rAt}&=\frac{\vartheta_{rAt}\int ydF_r\left(y\right)}{P_{rAt}}=\frac{\vartheta_{rAt}\Gamma_{\zeta}\overline{w}_{rt}L_{rt}}{P_{rAt}}\\C_{rMt}&=\frac{\left(1-\vartheta_{rAt}\right)\int ydF_r\left(y\right)}{P_{rMt}}=\frac{\left(1-\vartheta_{rAt}\right)\Gamma_{\zeta}\overline{w}_{rt}L_{rt}}{P_{rMt}}.\end{aligned}$$

Relative Prices

To compute the time-series of the relative price of non-agricultural to agricultural goods, we compute chained sectoral price indices and then take their ratio. More specifically, consider sector s and time-period between $t-1$ and t. Let P_{st}^L and P_{st}^P denote the Laspeyres and Paasche indices, respectively. These are given by

$$P_{st}^L=\frac{\sum_r P_{rst}C_{rst-1}}{\sum_r P_{rst-1}C_{rst-1}}\quad P_{st}^P=\frac{\sum_r P_{rst}C_{rst}}{\sum_r P_{rst-1}C_{rst}}.$$

The Fisher Index for sector s is then given by $P_{st}^F=\sqrt{P_{st}^L\times P_{st}^P}$.. The time-series of the relative price is then given by $\mathcal{P}_t^{M-A}=P_{Mt}^F/P_{At}^F$.

TABLE A.3: DETERMINANTS OF AGRICULTURAL SPECIALIZATION

	BIVARIATE CORRELATION OF s_{rA1880} WITH ...				
	$\ln A_{r1880}$	$\ln Z_{r1880}$	$\ln A_{r1880}/Z_{r1880}$	$\ln \ell_{r1880}$	$\ln B_{r1880}$
	-0.231***	-0.257***	0.0368	-0.133***	-0.115***
	(0.0110)	(0.0114)	(0.0292)	(0.00630)	(0.0304)
Observations	495	495	495	495	495
Adjusted R^2	0.430	0.515	0.003	0.493	0.045

Notes: The table reports the results of a set of bivariate regression $s_{rA1880} = \alpha + \beta x_r + u_r$, where $x_r = \ln A_{r1880}$ (Column 1), $x_r = \ln Z_{r1880}$ (Column 2), $x_r = \ln\left(A_{r1880}/Z_{r1880}\right)$ (Column 3), $x_r = \ln \ell_{r1880}$ (Column 4) and $x_r = \ln B_{r1880}$ (Column 5). Robust standard errors in parentheses. * , **, and *** denote statistical significance at the 10%, 5% and 1% level respectively.

B.3.5 Agricultural Employment and Regional Fundamentals

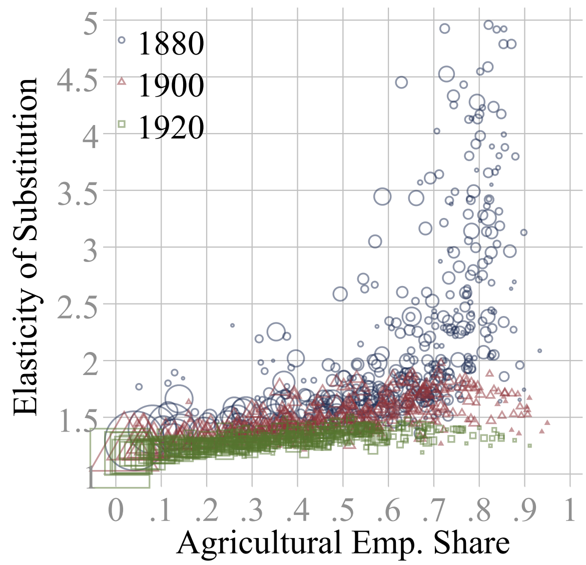
In Table A.3, we summarize the variation of the regional fundamentals by reporting their correlation with the agricultural employment share in 1880. Specifically, we run a set of bivariate regressions between s_{rA1880} and the (log of the) calibrated productivities $\ln A_{r1880}$ and $\ln Z_{rM1880}$ (columns 1 and 2), a region's technological comparative advantage $\ln\left(Z_{rA1880}/Z_{rM1880}\right)$ (column 3), a region's population density $\ln \ell_{r1880}$ (column 4), and regional amenities $\ln B_{r1880}$ (column 5). Table A.3 reports the regression coefficients on the regional agricultural employment share.

Most importantly, as shown in Columns 1 and 2, we find that agricultural regions in 1880 have *both* low agricultural productivity and low manufacturing productivity. Agricultural specialization is thus only a reflection of comparative advantage in agriculture, not absolute advantage. This is seen in Column 3, which shows a positive, albeit statistically insignificant, correlation between agricultural employment shares and the relative productivity of the agricultural sector. Column 4 shows that rural labor markets are land-abundant, that is their population density ℓ_{rt} is low. This pattern is implied by the fact that, empirically, agricultural land rents in rural regions are relatively low compared to more urban locations. Finally, amenities are lower in rural regions, indicating that such regions are sparsely populated even given their low wages (Column 5).

B.3.6 Implied Substitution Elasticities between Sectoral Consumption

The model-implied elasticity of substitution between value added in the two sector sis given in equation (3). Because the agricultural expenditure share ϑ_A varies across indi-

FIGURE A.5: THE ELASTICITY OF SUBSTITUTON



Notes: The figure shows the elasticity of substitution $\varrho_{rt} = 1 + \eta \frac{(\vartheta_{rAt} - \phi)^2}{\vartheta_{rAt}(1 - \vartheta_{rAt})}$ as a function of the agricultural employment share. For ease of readibility we do not display localities with an elasticity exceeding 5. This restrictions is only binding for some regions in 1880. The size of the scatter symbols is proportional to a location's total employment.

viduals (and hence space and time), ϱ varies with the level of development. To visualize the variability in ϱ, we define the following "average" substitution elasticity in region r at time t,

$$\varrho_{rt} = 1 + \eta \frac{(\vartheta_{rAt} - \phi)^2}{\vartheta_{rAt} (1 - \vartheta_{rAt})},$$

where ϑ_{rAt} is the aggregate agricultural employment share defined in equation (12) above. In Figure A.5, we plot ϱ_r as a function of the local agricultural employment share in 1880, 1900, and 1920. In the cross-section, substitution elasticities are generally higher in rural labor markets, reflecting the positive relationship between agricultural spending shares and agricultural employment. Across time, the elasticity of substitution declines, because economic growth reduces the agricultural spending share.

B.3.7 Micro Estimates of the Engel Elasticity η

Targeting the time series of the agricultural employment share implied an estimate of $\eta = 0.93$. However, our model also implies a log-linear relationship between individuals' expenditure share on agricultural products and their total expenditure that can be used to estimate η from cross-sectional microdata:

$$\ln \vartheta_A \left(y, P_{r,M}\right) = \ln\left(-\nu P_{r,M}^{-1}\right) - \eta \ln y, \tag{A.20}$$

where we used our estimate $\phi \approx 0$. We use the 1936 Consumer Expenditure Survey (CEX) by the U.S. Bureau of Labor Statistics obtained from the Inter-university Consortium for Political and Social Research (ICPSR) to provide direct evidence on the log-linear relationship between expenditure shares and total expenditure.[26]

The CEX contains micro data on expenditure of individuals on a large variety of categories and a swath of individual characteristics. We use the household files. We obtain information on households' total expenditure, expenditure on food, urban/rural status, size, interview data, occupation and industry of household head, race of household head, and county of residence.

In Figure A.6, we show that this log-linear relationship is a good description of the data. In the left panel, we display the cross-sectional distribution of food shares. Empirically this variation is substantial, ranging from 5% to 80%. In the right panel, we show the empirical relationship between log expenditure and log food shares as a binned scatter plot. As implied by our theory, the elasticity between food shares and expenditure is indeed essentially constant across the entire range of the distribution of expenditure. The slope coefficient falls in between $\eta \in (0.315, 0.362)$ depending on which additional controls are chosen, implying that our "macroestimate" of $\eta = 0.93$ is higher then the microestimate that exploits cross-sectional variation.

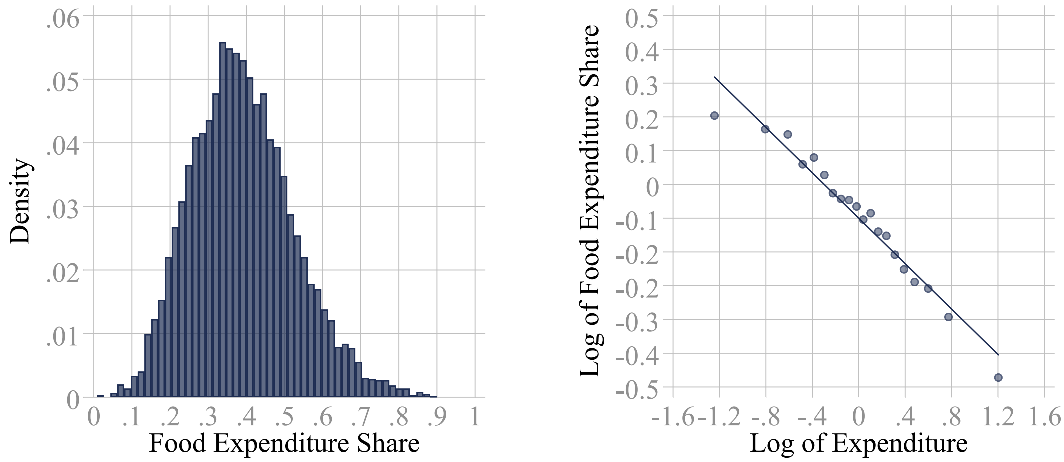
B.3.8 Validating the First-Order Approximation

In Figure 7, we reported the decomposition of local wage growth into the four components highlighted in Proposition 2. In the theory outlined in the paper, the underlying first-order approximation that decomposes wage growth into the four margins takes the following form:

[26] We note that our theory is written in terms of value added. The expenditure data is in terms of final expenditure data. Herrendorf et al. [2013] show that in general there is no direct mapping between the preference parameters of the value added and the final good demand system. However, Fan et al. [2022] show that for the class of PIGL preferences used here, the Engel elasticity η is portable between the final good and value added demand system.

FIGURE A.6: HETEROGENEITY IN FOOD EXPENDITURE SHARES

(A) 1880-1900 (B) 1900-1920



Notes: The figure shows the cross-sectional distribution of the individual expenditures shares on food (left panel) and the bin scattered relationship between the (log) expenditure share on food and (log) total expenditure (right panel). The relationship in the right panel is conditional on a set of location and family size fixed effects.

(A.21)
$$d\ln \overline{w}_{rt} = \phi_M\left(s_{rA}\right)\left(\frac{1}{\sigma}d\ln \mathcal{D}_t + \frac{\sigma-1}{\sigma}d\ln Z_{rMt}\right) + \phi_A\left(s_{rA}\right)\left(d\ln Z_{rAt} - \alpha d\ln \ell_{rt}\right).$$

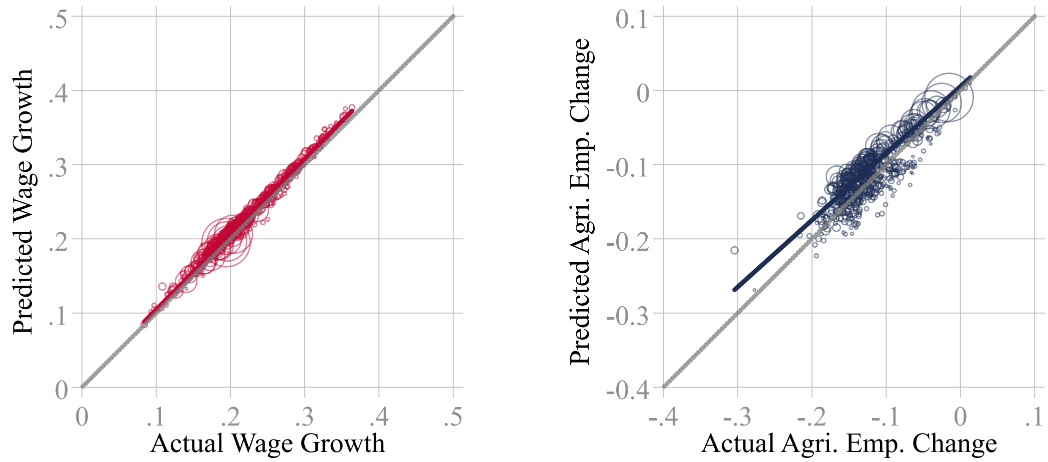
The theory also permits a similar first-order approximation can be derived for the change in local agricultural employment shares

$$ds_{rAt} = \psi\left(s_{rA}\right)\left(d\ln Z_{rAt} - \alpha d\ln \ell_{rt} - \frac{\sigma-1}{\sigma}d\ln Z_{rMt} - \frac{1}{\sigma}d\ln \mathcal{D}_{rt}\right), \tag{A.22}$$

In the quantitative model there is an additional "agricultural demand" term in these approximations that emerges once there are trade costs for agricultural goods, which we abstracted from in the main theory but re-introduce in the quantitative version of the model.

In Figure A.7, we show that equations A.21 and A.22 provide an excellent fit of the data despite being an approximation. This provides justification for using these equations to decompose local wage growth and industrialization in Section 5. Specifically, the left panel shows the correlation between local wage growth based on equation (A.21) and local wage growth stemming from the non-linear solution of the model. If the model

FIGURE A.7: ACCURACY OF THE FIRST-ORDER-APPROXIMATION OF THE MARGINS OF LOCAL WAGE GROWTH AND INDUSTRIALIZATION



Notes: The left panel shows the correlation between actual wage growth in the model and predicted wage growth based on equation A.21. The right panel shows the correlation between actual agricultural employment share changes in the model and predicted agricultural employment share changes based on a first order approximation in the model. Each dot a is a commuting zone. and the size of the dots is proportional to a commuting zone's total employment in 1880. The dashed grey line is a 45 degree line. The solid line in each panel is a weighted fit line using 1880 total employment as weights.

were to follow equation (A.21) exactly, the results should lie on a 45 degree line. Each red dot represents a commuting zone, the grey dashed line is a 45 degree line, and the solid red line represents the best fit through the data. The right panel compares the change in the local agricultural employment share based on equation A.22 to the change in the local agricultural employment in the simulated model. The linear fit line is again very close to the grey dashed 45 degree line providing support for using the first-order approximation in our analysis.