Spatial Structural Change*

Fabian Eckert† Michael Peters‡

January 2018

Abstract

This paper studies the spatial implications of structural change. By shifting demand towards non-agricultural goods, the structural transformation benefits workers in urban centers and hurts rural locations. At the aggregate level, the economy can respond either through a reallocation of labor from rural to urban areas or through a reduction in agricultural employment within a given location. Using detailed spatial data for the U.S. between 1880 and 2000, we show that spatial reallocation accounts for almost none of the aggregate decline in agricultural employment. We interpret this fact through the lens of a novel quantitative theory of spatial structural change, and show that the absence of the spatial reallocation channel is due primarily to regional productivity shocks, which almost entirely offset the urban bias of the structural transformation. Frictions to labor mobility meanwhile are quantitatively unimportant. The model implies that spatial welfare differences declined substantially during the United States' structural transformation and that the spatial reallocation of factors can account for about 15% of aggregate growth since 1880.

*We thank seminar participants at Boston University, Stanford, Penn State, UNC, the SED, the EEA, the NBER Macroeconomic Across Time and Space Conference, the Edinburgh Workshop on Agricultural Productivity and the Yale International Lunch for very helpful comments. We are particularly thankful to Simon Alder, Timo Boppart, David Lagakos, Doug Gollin, Kjetil Storesletten and Fabrizio Zilibotti for their suggestions. We also thank Costas Arkolakis, Pete Klenow, and Aleh Tsyvinski for their comments. Zara Contractor, Andrés Gvirtz and Adam Harris provided outstanding research assistance.

†Yale University. fabian.eckert@yale.edu

‡Yale University and NBER. m.peters@yale.edu

1 Introduction

Structural change is a key feature of long-run economic growth. As countries grow richer, aggregate spending shifts towards non-agricultural goods and the share of employment in the agricultural sector declines. This sectoral bias of the growth process also implies that economic growth is unbalanced across space. In particular, by shifting expenditure away from the agricultural sector, the structural transformation is biased against regions, that have a comparative advantage in the production of agricultural goods. This "urban bias" of the structural transformation can affect the aggregate economy in two ways. It can either induce a process of spatial reallocation whereby individuals migrate from rural, agricultural areas to centers of non-agricultural production or the process of structural change can trigger a transformation at the local level, whereby agricultural employment shares decline within each locality.

In this paper, we start by documenting a striking - and to the best of our knowledge - new empirical fact: the spatial reallocation of people from high to low agricultural regions accounts for essentially none of the aggregate decline in agricultural employment since 1880.[1] In contrast, the entire structural transformation is due to a decline in agricultural employment, which occurs within localities. We also show that this fact is consistent with the secular trend of the urbanization rate, which increased from 25% to 75% between 1880 and 2000. In fact, we explicitly show that - like the change in agricultural employment - urbanization is also predominantly a within-region phenomenon.

To understand the determinants of these joint long-run spatial and sectoral adjustments and quantify their implications for the spatial distribution of welfare and aggregate productivity, we propose a novel quantitative theory of spatial structural change.[2] To do so, we combine an economic geography model with intra-national trade and frictional labor mobility with an otherwise standard, neoclassical model of the structural transformation. Regions differ in their sectoral productivities which may induce migration, subject to moving costs, as aggregate demand shifts towards non-agricultural goods (and hence regions with a non-agricultural comparative advantage).

The model highlights three key aspects that determine whether the economy accommodates the secular changes in demand conditions through spatial reallocation or intra-regional sectoral shifts: (1) the size of frictions to spatial mobility, (2) the spatial correlation between current agricultural employment shares and future earnings and (3) the elasticity of migration to spatial wage differentials. If moving costs were prohibitively high, spatial reallocation would (by construction) be unable to account for the decline in agricultural employment. US Census data reveals, however, that over the last century, 30% of the working-age population lived in states different from their state of birth. We show that this implies that prohibitively high moving frictions cannot explain the quantitative insignificance of the spatial

[1] This holds true when a region is defined as a US state, commuting zone (Tolbert and Sizer (1996)) or county. There are roughly 700 commuting zones and 3000 counties. For the remainder of the paper we focus on commuting zones for reasons outlined below.

[2] By "Spatial Structural Change" we refer to the simultaneous changes in sectoral employment and the spatial organization of economic activity. The father of the study of structural change, Simon Kuznets, was maybe the first to highlight the importance of studying spatial and sectoral reallocation in one unified framework (Lindbeck, ed (1992)).

reallocation channel, as there were more than enough people crossing state boundaries.

Such interstate migration flows are, however, only weakly correlated with the initial level of agricultural specialization. Whether people leave rural areas depends *both* on the extent to which rural areas tend to have low wages in the future and on the migration elasticity, i.e. the sensitivity of migration flows to spatial wage differences. Our model highlights that the structural transformation provides an endogenous mechanism pushing towards a negative correlation between future wages and initial agricultural employment shares: as demand for non-agricultural goods increases, wages increase in urban regions (those which have low initial agricultural employment). This "urban bias" of the structural transformation provides a strong force towards a net reallocation of workers towards urbanized areas as the country develops.

At the same time, regional productivities are not necessarily constant, but rather are subject to stochastic shocks. We show that, empirically, such local productivity shocks almost entirely offset the urban bias of the structural transformation. As a result, future earnings are only weakly correlated with initial agricultural employment shares. In addition, we also find that the elasticity of lifetime migration flows to spatial wage differences is relatively small. In particular, we show that spatial population *gross* flows are much larger than population *net* flows. This suggests that idiosyncratic, non-monetary preference shocks are an important determinant of migration decisions. Taken together, these facts imply, that there was too little net reallocation directed towards low agricultural regions to make spatial reallocation an important margin of adjustment for the structural transformation.

Finally, we use the model to study the implications of spatial structural change for aggregate economic performance and the spatial distribution of welfare. We first focus on the role of spatial reallocation for aggregate productivity. Because we estimate that rural, agricultural-intensive regions are - on average - less productive and generate less value added per worker than non-agricultural areas, the lack of spatial reallocation suggests that the US economy potentially missed out on substantial productivity improvements. Quantitatively, we find such productivity losses were modest. If spatial mobility was costless, aggregate income would only have been 4% higher in the year 2000. In contrast, if moving costs had been prohibitively high, income per capita would have been 15% lower. The observed process of spatial arbitrage in the US therefore seemed to have captured a large share of potential efficiency gains.

We then analyze the evolution of spatial welfare inequality. We find that welfare inequality across US commuting zones declined substantially between 1910 and 2000. While in 2000 the interquartile range of the distribution of spatial welfare corresponded to a doubling of regional income, in 1910 one would have only needed to increase regional income by about 160% to increase utility from 25% to 75% quantile of the distribution of welfare. Importantly, and similar to aggregate productivity, the possibility of spatial mobility was an important determinant of the reduction in spatial welfare inequality. If spatial mobility had been prohibitively costly, the spatial dispersion of welfare would had declined much less. In particular, unskilled workers, who have a comparative advantage in the agricultural sector and are hence particularly exposed to the urban bias of the structural transformation, would have seen no decline in

spatial inequality.

As a theoretical contribution, we propose a model that combines the basic ingredients from an economic geography model (spatial heterogeneity, intra-regional trade and costly labor mobility) with the usual features of neoclassical models of structural change (non-homothetic preferences, unbalanced technological progress, aggregate capital accumulation). Despite this richness, the theory remains highly tractable. Building on recent work by Boppart (2014), we first show that by combining a price independent generalized linear (PIGL) demand system with the commonly-used Frechet-distribution of individual skills, one can derive closed-form solutions for most aggregate quantities of interest. We then show how this structure can be embedded in an otherwise standard overlapping-generation model. Doing so allows us to tractably accommodate both individual savings (and hence aggregate capital accumulation) and costly spatial mobility. In particular, while individuals are forward looking in terms of their savings behavior, we show that the spatial choice problem reduces to a static one. As a result, we do not have to keep track of individuals' expectations about the entire distribution of future wages across locations - the aggregate interest rate is sufficient. Moreover, because our model essentially nests a version of the neoclassical model of growth as the number of regions collapses to one, we view our model as a natural spatial extension of neoclassical theories of the structural transformation.

Related Literature In this paper, we combine insights from the macroeconomic literature on the structural transformation with recent advances in quantitative models of economic geography. The literature on the process of structural change has almost exclusively focused on the time series properties of sectoral employment and value added shares - see Herrendorf et al. (2014) for a survey of this large literature.[3] In contrast, the recent generation of quantitative spatial models in the spirit of Allen and Arkolakis (2014) are mostly static in nature and focus on the spatial reallocation of workers across heterogeneous locations.[4] We show that these two aspects interact in a natural way. The urban bias, which is inherent in the structural transformation, induces changes in demand, which are non-neutral across space and hence affect the spatial equilibrium of the system. Conversely, the spatial topography, in particular the extent to which individuals are spatially mobile, has macroeconomic implications by determining equilibrium factor prices, capital accumulation and aggregate productivity.

Relatively few existing papers explicitly introduce a spatial dimension into an analysis of the structural

[3] Authors such as Kuznets (1957) and Chenery (1960) have been early observers of the striking downward trend in the aggregate agricultural employment share and the simultaneous increase in manufacturing employment in the United States. To explain these patterns, two mechanism have been proposed. Demand side explanations stress the role of non-homotheticities, whereby goods differ in their income elasticity (see e.g. Kongsamut et al. (2001), Gollin et al. (2002), Comin et al. (2015) and Boppart (2014)). Supply-side explanations argue for the importance of unbalanced technological progress across sectors and capital-deepening (see e.g. Baumol (1967), Ngai and Pissarides (2007), Acemoglu and Guerrieri (2008), and Alvarez-Cuadrado and Long (2011)).

[4] This literature has addressed questions of spatial misallocation (Hsieh and Moretti (2015), Fajgelbaum et al. (2015)), the regional effects of trade opening (Fajgelbaum and Redding (2014), Tombe et al. (2015)), the importance of market access (Redding and Sturm (2005)) and the productivity effects of agglomeration economies (Ahlfeldt et al. (2015)). See Redding and Rossi-Hansberg (2016) for a recent survey of this growing literature.

transformation. An early contribution is Caselli and Coleman II (2001), who argue that spatial mobility was an important by-product of the process of structural change in the US. Michaels et al. (2012) also study the relationship between agricultural specialization and population growth across US counties. Their analysis is more empirically oriented and does not use a calibrated structural model. More recently, Desmet and Rossi-Hansberg (2014) propose a spatial theory of the US transition from manufacturing to services and Nagy (2016) examines the process of city formation in the United States before 1860.

In allowing for a spatial microstructure, this paper also sheds new light on the so-called "agricultural productivity gap, i.e. the observation that value added per worker is persistently low in the agricultural sector. Gollin et al. (2013) use micro data for numerous countries to provide cross-sectional evidence of this pattern. Buera and Kaboski (2009) use data for the US to provide the time-series analogue. This discrepancy between agricultural employment and value added shares is at odds with most standard macroeconomic models of the structural transformation, which predict that value added per worker, i.e. marginal products, should be equalized (Herrendorf and Schoellman, 2015). Our model endogenously generates an "agricultural productivity gap", without resorting to labor market frictions across sectors, as spatial mobility costs keep wages in agricultural areas low.[5] Moreover, the urban bias of the structural transformation is a secular force, which tends to exert further downward pressure on agricultural productivity.

On the theoretical side, we build on Boppart (2014) by assuming a price independent generalized linear (PIGL) demand structure. This demand structure has more potent income effects than the widely-used Stone-Geary specification, a feature which is required to generate declines in agricultural employment of the magnitude observed in the data.[6] The non-homothetic CES demand system, recently proposed by Comin et al. (2015), has similarly favorable time-series properties. It is, however, less tractable for aggregating and hence not as easy to embed in a spatial structure.

The remainder of the paper is structured as follows. In Section 2, we document the empirical fact that spatial reallocation accounts for essentially none of the aggregate decline in agricultural employment over the last 120 years. Section 3 presents our model. In Section 4, we calibrate the model to time-series and spatial data from the US. In Sections 4.3 and 5, we explain why the spatial reallocation component of the structural transformation is small and we quantify the implications for aggregate productivity and spatial inequality. Section 6 concludes. An Appendix contains the majority of our theoretical proofs and further details on our empirical results.

[5]The existence of spatial gaps and their implications for aggregate productivity and welfare has been subject to a recent active literature (see e.g. Bryan et al. (2014), Lagakos et al. (2015) or Young (2013)). Bryan and Morten (2015) and Hsieh and Moretti (2015) use spatial models related to our to study the aggregate effects of spatial misallocation. In contrast to us, their models are static and there is no relation with the structural transformation. In an empirical exercise, Gollin et al. (2017) use direct measures of differences in spatial amenities and show that amenities in agricultural, rural localities are much lower than in urban centers. This finding is consistent with our structural estimates of spatial amenities in the US of the 19th century.

[6]Between 1880 and 2000 the aggregate agricultural employment share declines from around 50% to 2%. Alder et al. (2017) show that the PIGL demand system provides a good fit to the data since 1900. A model with Stone-Geary preferences can match the post-war data (see e.g. Herrendorf et al. (2013)), but has difficulties at the longer time horizon as income effects vanish asymptotically.

2 Spatial Reallocation and Structural Change

The structural transformation is not only unbalanced across sectors but also unbalanced across space. In particular, the reallocation of spending across sectors is biased agains rural regions, which specialize in agriculture. From an accounting perspective, there are two margins through which the economy can accommodate such changes. First, the structural transformation can induce a process of *spatial reallocation,* whereby labor reallocates from agriculturally intensive regions to places which are more specialized in the production of non-agricultural goods. Second, it can lead to a *regional transformation,* whereby regional agricultural employment shares decline in every locality.Formally, the aggregate decline in the agricultural employment share since 1880 can be decomposed as

$$s_{A1880} - s_{At} = \sum_r s_{rA1880} l_{r1880} - \sum_r s_{rAt} l_{rt} = \underbrace{\sum_r s_{rA1880} \left(l_{r1880} - l_{rt}\right)}_{\text{Spatial Reallocation}} + \underbrace{\sum_r \left(s_{rA1880} - s_{rAt}\right) l_{rt}}_{\text{Regional Transformation}}, \tag{1}$$

where s_{At} is the aggregate agricultural employment share at time t, l_{rt} denotes the share of the population living in region r at time t and s_{rAt} is the regional employment share in agriculture. As highlighted by (1), the spatial reallocation margin is important for the decline in agricultural employment, if net population growth, $l_{r1880} - l_{rt}$, and the initial agricultural employment share, s_{rA1880}, are negatively correlated.

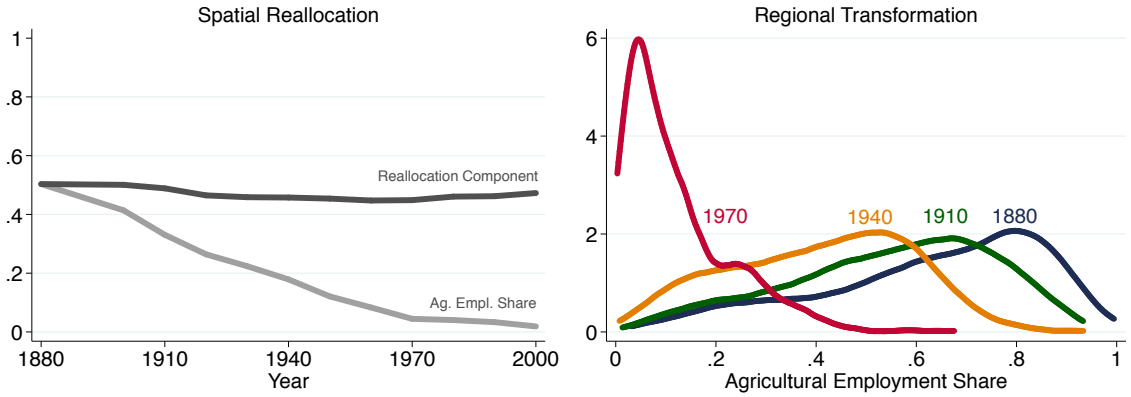
For the case of the U.S., the relative importance of the reallocation and transformation margins is striking: the spatial reallocation of labor accounts for essentially none of the structural transformation observed in the aggregate. To see this, consider Figure 1, where we implement (1) at the commuting zone level.[7] In particular, we conduct a "shift-share"-analysis and compare the actual employment share, s_{At}, to the implied agricultural employment share, which emerged *solely* from spatial reallocation, $\sum_r s_{rA1880} l_{rt}$. It is clear that - in an accounting sense - spatial reallocation accounts for none of the aggregate decline in agricultural employment. To put it differently, the process of the structural transformation is *not* driven by a reallocation of people from high to low agricultural places. Out of the total decline of about 48%, only 3% is due to the reallocation of workers across commuting zone boundaries.

It follows that most of structural change takes place *within* regions through transformations of the local structure of employment. This is seen in the right panel of Figure 1, where we display the cross-sectional distribution of regional agricultural shares for different years. First of all, note that there is substantial cross-sectional dispersion. While the majority of commuting zones have agricultural employment shares exceeding 75% in 1880, many regions are already much less agriculturally specialized and have agricultural employment shares below 25%. Secondly, there is a marked leftwards shift, whereby *all* regions see

[7] In the paper, we use the definition of a commuting zone as our definition of a region. There are roughly 700 commuting zones in the US. We describe our data in more detail in Section 4 below. We have replicated Figure 1 on the county level, it looks effectively identical. In addition we have computed the importance of regional transformation versus spatial reallocation for countries around the world using publicly available data from the Integrated Public Use Microdata Series (IPUMS) International data base. We find that regional transformation appears to be an integral part of the structural transformation across countries and the size of the spatial component is always below 5%.

a decline in agricultural employment and undergo a regional transformation as the US economy develops. This shows that structural change is not a process that induces regional specialization, but rather features a fractal property whereby all local economies undergo changes in sectoral structure akin to the aggregate economy.

These patterns are seemingly at odds with the process of urbanization, whereby cities, i.e. centers of non-agricultural production, grow.[8] This, however, is not the case. While inconsistent with a view of urbanization as a movement of workers towards few large population centers, it is consistent with a world in which workers move towards urban centers *within* labor markets. This is exactly, what happened in the US. As we explicitly show in Section A in the Appendix, the increase in urbanization is also a within-region phenomenon, which is very local in nature. In particular, like for the change in the agricultural employment, the reallocation of individuals from rural to urbanized counties explains almost none of the sharp rise in urbanization. There we also show that this pattern is not exclusive to the US but present in many countries around the world undergoing the structural transformation.



Notes: In the left panel, we show the aggregate agricultural employment share (light grey line) and the predicted agricultural share holding regional agricultural shares at their 1880 level, i.e. $\sum_r s_{rA1880} \times l_{rt}$, where s_{rAt} and l_{rt} are the agricultural employment share and the population share of region r at time t(dark grey line). In the right panel we show the cross-sectional distribution of agricultural employment shares between 1880 and 1970. We omit the 2000 cross-section from the left panel for expositional purposes only: the agricultural employment share does not change much between 1970-2000. For a detailed description of the construction of the regional data we refer to Section 4.

Figure 1: Spatial Structural Change: Spatial Reallocation vs. Regional Transformation

In the remainder of this paper, we develop a theory of spatial structural change, which can speak to these patterns. More specifically, we will use the calibrated model to answer (i) why the spatial reallocation component is small and (ii) whether its absence has important implications for the spatial distribution of welfare and aggregate productivity during the structural transformation.

[8] Over the 1880-2000 period that is the focus of this paper, the fraction of the US population living in cities tripled from about 25% to 75%. See Figure 10 in the Appendix for a time series of the fraction of the US population residing in an urban environment.

3 A Quantitative Theory of Spatial Structural Change

In this section we present a model of spatial structural change. The theory rests on three pillars. We start with what is essentially a neoclassical model of the structural transformation, where the process of structural change is generated from non-homothetic preferences and unbalanced technological progress. We introduce a spatial dimension by embedding this structure into an economic geography model of heterogeneous locations and costly spatial mobility. Finally, we allow for skill-based selection across locations and sectors of production by assuming that individuals differ in their human capital, with skilled workers have both an absolute advantage and a comparative advantage in the non-agricultural sector.

3.1 Environment

We consider an economy consisting of R regions indexed by r. Each region features an agricultural (A) and a nonagricultural (NA) sector , which are indexed by $s = A, NA$ throughout the paper. Since the model is dynamic we additionally index most quantities by t. For expositional simplicity, we sometimes omit the time subscript if where there is no risk of confusion.

Technology We consider an economy with two goods, an agricultural good and a non-agricultural good. Each good is a CES composite of differentiated regional varieties with a constant elasticity of substitution σ. In particular,

$$Y_s = \left(\sum_{r=1}^{R} Y_{rs}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where Y_{rs} is the amount of goods in sector s produced in region r and σ is the elasticity of substitution. For analytical tractability, we assume that goods are freely traded so that prices are equalized across locations.[9]

Regional production functions are fully neoclassical and given by

$$Y_{rst} = A_{rst} K_{rst}^{\alpha} H_{rst}^{1-\alpha},$$

where K_{rst} and H_{rst} denotes capital and labor (in efficiency units) in region r, sector s and time t. Building on the work of Herrendorf et al. (2015), we assume that capital shares are identical across sectors.[10] It is

[9] This assumption considerable simplifies workers' spatial choice problem, as outlined below.

[10] Herrendorf et al. (2015) find that quantitatively speaking differential technical progress is the main technological driver of the structural transformation, whereas sectoral differences in the capital shares and elasticities of substitution are of second order importance only. They explicitly note that Cobb–Douglas sectoral production functions that differ only in technical progress capture the main features on the technological side of the US structural transformation well.

conceptually useful to decompose regional productivity A_{rst} as

$$A_{rst} \equiv Z_{st} Q_{rst} \quad \text{with} \quad \sum_r Q_{rst}^{\sigma-1} = 1. \tag{3}$$

Here Z_{st} is an *aggregate* TFP shifter in sector s, which affects all regions proportionally. Additionally, there are *idiosyncratic* sources of productivity. The vector of $\{Q_{rs}\}_{rs}$ flexibly parameterizes the distribution of regional sectoral productivity differences. The common component of Q_{rs} across sectors within region r captures differences in *absolute* advantage. Regional differences in $Q_{rs}/Q_{rs'}$ capture differences in *comparative* advantage. Given the normalization embedded in (3), the vector $\{Q_{rs}\}_{rs}$ can be thought of as describing the heterogeneity in productivity across space.

Capital accumulates according to the usual law of motion

$$K_{t+1} = (1-\delta) K_t + I_t,$$

where I_t denotes the amount of investment at time t and δ is the depreciation rate. We assume that the investment good is a Cobb-Douglas composite of the agricultural and non-agricultural good given in (2). Letting ϕ be the share of the agricultural good in the production of investment goods, the price of the investment good is given by $P_{It} = P_{At}^{\phi} P_{NAt}^{1-\phi}$.[11] For the remainder of the paper the investment good will serve as the numeraire of our economy.

Selection, Human Capital and Labor Supply Individuals differ in their human capital. Allowing such differences is important for capturing differential spatial sorting by skill-group. In particular, we suppose that individuals can be of two types - high skilled and low skilled. Their skill type $h \in \{L, H\}$ determines the distribution of their sector-specific efficiency units $z^i = \left(z_A^i, z_{NA}^i\right)$. For tractability, we assume that for each worker i, z^i is drawn from the Frechet distribution

$$F_s^h(z) = e^{-\Psi_s^h z^{-\zeta}}, \tag{4}$$

where Ψ_s^h parametrizes the average level of human capital of individuals of skill type h in sector s and ζ governs the dispersion of skills. Without loss of generality we parameterize Ψ_s^h as[12]

$$\begin{bmatrix} \Psi_A^L & \Psi_{NA}^L \\ \Psi_A^H & \Psi_{NA}^H \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ q & q\mu \end{bmatrix}.$$

Hence, q measures the absolute advantage of skilled individuals and μ governs the complementarity between skills and non-agricultural employment. We denote the share of the aggregate labor force that is

[11] The accompanying production function for the investment good is given by $I_t = \phi^{\phi}(1-\phi)^{\phi} X_A^{\phi} X_{NA}^{1-\phi}$, where X_s is the amount of sector s goods used in the investment good sector.

[12] Note that $\Psi_s^L = 1$ is a normalization given the regional technologies Q_{rs}.

skilled by λ_t and the share of skilled workers in region by λ_{rt}. We assume that λ_t is exogenous, i.e. there is no human capital accumulation. In contrast, the spatial allocation of human capital, λ_{rt}, is endogenous and determined by peoples' migration decisions.

The Frechet assumption in (4) gives convenient aggregation properties. In particular, total earnings of individual i residing in region r are given by

$$y_r^i = \max\left\{w_{rA}z_A^i, w_{rNA}z_{NA}^i\right\},$$

where w_{rs} denotes the prevailing equilibrium wage per efficiency unit in region r. The Frechet distribution then implies that average earnings of individuals in skill group h are given by

$$E\left[y_r^{ih}\right] = \Gamma_\zeta \Theta_r^h,$$

where $\Gamma_\zeta = \Gamma(1-1/\zeta)$, $\Gamma(.)$ is the gamma function and

$$\Theta_r^h = \left(\Psi_A^h w_{rA}^\zeta + \Psi_{NA}^h w_{rNA}^\zeta\right)^{1/\zeta}. \tag{5}$$

Note that Θ_r^h differs across skill-types, but is equalized across sectors and can be directly calculated from regional wages (w_{rA}, w_{rNA}). The vector $\{\Theta_r^h\}_{rh}$ will turn out to be a key endogenous object in our analysis and we will refer to it as *expected regional income* (or for brevity *regional income*).[13]

Demographics In terms of preferences and demographics, ours is an overlapping generations (OLG) economy. Individuals live for two periods, work when they are young and save to be able to consume when they are old. The OLG structure is analytically extremely convenient because it generates a motive for savings (and hence capital accumulation), while still being sufficiently tractable to allow for spatial mobility. It is also empirically attractive in that it captures the importance of cohort effects in accounting for the structural transformation. In particular, both Hobijn et al. (2017) and Porzio and Santangelo (2017) show that a large share of the decline in agricultural employment due to young cohorts have much lower propensities to work in the agricultural sector.

In our model, individuals have three economic choices to make: (i) how much to save and consume when they are young, (ii) how to allocate their spending optimally across the two consumption goods in each period of life and (iii) and where to live and work. In terms of timing, we assume that individuals are born with a given skill type h, decide on their preferred location when they are young and then remain in that location for their entire life. For simplicity, we abstract from endogenous human capital accumulation

[13] Given (w_{rA}, w_{rNA}) we can also calculate the remaining allocations. The share of people of skill group h employed in sector s is given by $s_{rs}^h = \Psi_s^h \left(w_{rs}/\Theta_r^h\right)^\zeta$, so that the sectoral labor supply elasticity is governed by ζ. Note also that our model features the consequences of worker selection stressed by Lagakos and Waugh (2011): the *average* amount of efficiency units provided to sector s by individuals of skill group h is given by $\frac{H_{rs}^h}{L_{rs}s_{rs}^h} = \left(s_{rs}^h\right)^{-\frac{1}{\zeta}} \left(\Psi_s^h\right)^{\frac{1}{\zeta}}$, i.e. is decreasing in the sectoral employment share s_{rs}^h as individuals' sorting implies that the marginal worker in sector s is worse than the average worker.

and assume that skills are perfectly inherited, i.e. parents with skill h have children of skill h.

Letting $V(e_t, P_t)$ be the *indirect* utility of spending an amount e_t at prices $P_t = (p_{At}, p_{NAt})$, life-time utility of individual i after having moved to region r is given by

$$U_{rt}^i = \max_{[e_t, e_{t+1}, s]} \{V(e_t, P_t) + \beta V(e_{t+1}, P_{t+1})\}, \tag{6}$$

subject to

$$\begin{aligned} e_t + s_t &= y_{rt}^i \\ e_{t+1} &= (1 + r_{t+1}) s_t. \end{aligned}$$

Here, y_{rt}^i is individual i's real income in region r, s_t denotes the amount of savings and r_t is the real interest rate.

Preferences We follow the existing literature on the structural transformation by allowing for non-homothetic preferences. This preference-based explanations of the structural transformation, whereby consumers reduce their relative agricultural spending as they grow richer, has been shown to be quantitatively important for the process of structural changes (see e.g. Herrendorf et al. (2014), Boppart (2014) or Comin et al. (2015)). We follow Boppart (2014) and assume that individual preferences can be represented by the following indirect utility function

$$V(e, P) = \frac{1}{\eta} \left(\frac{e}{p_A^{\phi} p_{NA}^{1-\phi}} \right)^{\eta} - \frac{\nu}{\gamma} \left(\frac{p_A}{p_{NA}} \right)^{\gamma} + \frac{\nu}{\gamma} - \frac{1}{\eta}. \tag{7}$$

This is a slight generalization of the PIGL demand system employed by Boppart (2014).[14]

This demand system has two convenient properties. Firstly, it incorporates both income effects (governed by η) and price effects (governed by γ) in a flexible way. In particular, Roy's Identity implies that the expenditure share on the agricultural good, $\vartheta_A(e, p)$, is given by[15]

$$\vartheta_A(e, p) \equiv \frac{x_A(e, p) p_A}{e} = \phi + \nu \left(\frac{p_A}{p_{NA}} \right)^{\gamma} e^{-\eta}. \tag{8}$$

For $\eta > 0$, the expenditure share on agricultural goods is declining in total expenditure. This captures the income effect of non-homothetic demand, whereby higher spending reduces the relative expenditure

[14] For $V(e, p)$ to be well-defined, we have to impose additional parametric conditions. In particular, we require that $\eta < 1$, that $\gamma \geq \eta$. These conditions are satisfied in our empirical application. See Section B.5 of the Appendix for a detailed discussion. Boppart (2014) uses this demand system to study the the evolution of service sector. In terms of (7) he assumes that P_A is the price of goods and P_{NA} is the price of services and considers the case of $\phi = 0$.

[15] Roy's Identity implies that $\vartheta_A(e, p) = x_A^M(p, e) \times \frac{p_A}{e} = -\frac{\partial V(p, e(p,u))/\partial p_A}{\partial V(p, e(p,u))/\partial e} \frac{p_A}{e}$, where $x_A^M(p, e)$ denotes the Marshallian demand function. For $V(p, e)$ given in (7), this expression reduces to (8). See Section B.2 in the Appendix for details.

share on agricultural goods. Holding real income e constant, the expenditure share is increasing in the relative price of agriculture if $\gamma > 0$. The preferences in (7) also nest particular special cases. The case of $\eta = 0$ corresponds to a homothetic demand system, where expenditure shares only depend on relative prices. The case of $\eta = \gamma = 0$ is the Cobb Douglas case where expenditure shares are constant and equal to ϕ. Note that ϕ is also the agricultural share in the investment good sector. Hence, this case is akin to neoclassical growth model, where the consumption and investment good is identical. Importantly, Alder et al. (2017) have shown that a preference specification in the PIGL class provides a good fit to the long-run data. This is in contrast to the more popular Stone-Geary specification, which is unable to quantitatively account for the long-run process of structural change between 1880 and 2000.

The second advantage of the specification in (7) are its aggregation properties. We show below that these preferences allow for very tractable aggregation despite the fact that they fall outside the Gorman class. Exploiting these properties allows us to incorporate them into a general equilibrium trade model in a tractable way.[16]

Spatial Mobility While individuals know their skills, i.e. $h \in \{L, H\}$ prior to their mobility decision, they only learn the actual realization of their labor efficiency z^i thereafter. This structure has two convenient properties. First, individuals differ in their spatial mobility choice only by their skill. Allowing mobility to depend on the realization of their efficiency bundle z^i would be less tractable as we would need to keep track of continuum of ex-ante heterogenous individuals. Secondly, this structure retains the convenient aggregation properties of the Frechet distribution in (4). If workers' spatial choice was conditional on z^i, the distribution of skills within a location would no longer be of the Frechet form.

Now consider the decision for individual i with skill h to move from j to r. We follow the literature on discrete choice models and assume the value of doing so can be summarized by

$$\mathscr{U}_{jr}^{ih} = E^h\left[U_{rt}^i\right] - MC_{jr} + A_r + \kappa \nu_j^i,$$

where $E^h[U_r]$ is the expected utility of living in region r (which is conditional on the skill level h) and MC_{jr} denotes the cost of moving from j to r. A_r is akin to a location amenity, which summarizes the attractiveness of region r and is common to all individuals and ν_j^h is an idiosyncratic error term, which is independent across locations and individuals. Furthermore, κ parametrizes the importance of the idiosyncratic shock, i.e. the extent to which individuals sort based on their idiosyncratic tastes relative to the systematic attractiveness of region r. The higher κ, the less responsive are individuals to the fundamental value of a location r.

As in the standard conditional logit model, we assume that ν_j^h is drawn from a Gumbel distribution. This

[16] This is in contrast to the non-homothetic CES demand system, which has recently been analyzed in Comin et al. (2015) and for which no closed form aggregation results exist.

implies that the share of people with skill type h moving from j to r is given by

$$\rho_{jrt}^{h} = \frac{exp\left(\frac{1}{\kappa}\left(E^{h}\left[U_{rt}^{i}\right] + A_{rt} - MC_{jr}\right)\right)}{\sum_{l=1}^{R} exp\left(\frac{1}{\kappa}\left(E^{h}\left[U_{lt}^{i}\right] + A_{lt} - MC_{jl}\right)\right)}. \tag{9}$$

From this we obtain the law of motion for the spatial reallocation of workers between t and $t=1$ as

$$L_{rt}^{h} = \sum_{j=1}^{R}\left(\rho_{jrt}^{h} L_{jt-1}^{h}\right), \tag{10}$$

i.e. the number of people of skill type h in region r is given by the total inflows children of skilled individuals from all other regions (including itself).[17] Since in the model workers only move once, we will discipline the model with data on *lifetime* migration, i.e. the fraction of people who live (and work) in a different location than where they were born (see Molloy et al. (2011)). Note also that the share of non-movers of skill type h is simply given by ρ_{jj}^{h}. We will show below that $E^{h}[U_r]$ has a tractable closed form expression, which makes (9) easy to solve.

3.2 Competitive Equilibrium

Given the environment above, we can now characterize the equilibrium of the economy. We proceed in three steps. First we characterize the household problem, i.e. the optimal consumption-saving decision and spatial choice. We then show that the solution to the household problem together with our distributional assumptions on individuals' skills delivers an aggregate demand system, which can be written as a function of a single endogenous variable, despite the fact that our economy does not admit a representative consumer. Finally, we show that the dynamic competitive equilibrium has a structure akin to the neoclassical growth model: given the sequence of interest rates $\{r_t\}_t$, we can solve the entire spatial equilibrium from simple static equilibrium condition. The equilibrium sequence of interest rates can then be calculated from households' savings decisions.

Individual Behavior

First consider the households' consumption-saving decision given in (6). The two-period OLG structure together with the specification of preferences in (7) has a tractable solution for both the optimal allocation

[17] With non perfect inheritance of skills expression 10 would take the following form:

$$L_{rt}^{h} = \sum_{j=1}^{R} \rho_{jrt}^{h}\left[\upsilon_{jt-1}^{Lh} L_{jt-1}^{L} + \upsilon_{jt-1}^{Hh} L_{jt-1}^{H}\right],$$

where $\{\upsilon_{jt-1}^{hh'}\}_{hh'}$ is a general transition matrix containing the fraction of region j children of type h parents that attain a skill type h'.

of expenditure and the consumers' total utility U_r. We summarize this solution in the following Lemma. Due to the OLG structure, a young and an old generation co-exists in each time period. We therefore denote the optimal level of expenditure when young (old) of the generation that is born at time $t$as e_t^Y $\left(e_{t+1}^O\right)$.

Lemma 1. *Consider the maximization problem in (6) where $V(e,P)$ is given in (7). The solution to this problem is given by*

$$\begin{aligned} e_t^Y(y) &= \psi(r_{t+1})\, y && (11)\\ e_{t+1}^O(y) &= (1+r_{t+1})(1-\psi(r_{t+1}))\, y && (12)\\ U_r &= \frac{1}{\eta}\psi(r_{t+1})^{\eta-1} y^{\eta} + \Lambda_{t,t+1} \end{aligned}$$

where

$$\begin{aligned} \psi(r_{t+1}) &= \left(1+\beta^{\frac{1}{1-\eta}}(1+r_{t+1})^{\frac{\eta}{1-\eta}}\right)^{-1} && (13)\\ \Lambda_{t,t+1} &= -\frac{\nu}{\gamma}\left(\left(\frac{p_{At}}{p_{NAt}}\right)^{\gamma} + \beta\left(\frac{p_{At+1}}{p_{NAt+1}}\right)^{\gamma}\right) + (1+\beta)\left(\frac{\nu}{\gamma}-\frac{1}{\eta}\right). \end{aligned}$$

Proof. See Section B.1 in the Appendix. □

Lemma 1 characterizes the solution to the household problem. Three properties are noteworthy. First of all, the policy functions for the optimal amount of spending are linear in earnings. This will allow for a tractable aggregation of individuals' demands. Secondly, these expenditure policies resemble the familiar OLG structure, where the individual consumes a share

$$\frac{e_t^Y(y)}{y} = \psi(r_{t+1}) = \frac{1}{1+\beta^{\frac{1}{1-\eta}}(1+r_{t+1})^{\frac{\eta}{1-\eta}}}$$

of his income when young and consumes the remainder (and the accrued interest) when old. Note that, if $\eta = 0$, i.e. if if demand is homothetic, we recover the canonical OLG solution for log utility where the consumption share is simply given by $1/(1+\beta)$. Importantly, this consumption share only depends on the interest rate r_{t+1} and not on relative prices P_t or P_{t+1}. This is due to our assumption that nominal income e is deflated by the same price index as the investment good. This is convenient for tractability and akin to the single-good neoclassical growth model, where the consumption good and the investment good uses all factors in equal proportions. For our purposes, this ensures that an increase in the price of investment good, p_{It}, makes savings more attractive but at the same reduces the marginal utility of spending. Finally, overall utility U_r is additively separable in income y_t and current and future prices P_t and P_{t+1} (which determine $\Lambda_{t,t+1}$). As we show below, this property will be essential for the tractability of agents' optimal migration choices.

To solve agents' spatial choice problem, we have to calculate their expected life-time utility $E^h\left[U_{rt}^i\right]$.

Given that life-time utility is a power function of individual income y_r^i and individual income is Frechet distributed with shape ζ and mean $\Gamma_\zeta \Theta_r^h$, expected life-time utility is given by

$$E^h[U_r] = \left[\frac{1}{\eta}\psi(r_{t+1})^{\eta-1} y^\eta + \Lambda_{t,t+1}\right] = \frac{\Gamma_{\eta/\zeta}}{\eta}\psi(r_{t+1})^{\eta-1}\left(\Theta_r^h\right)^\eta + \Lambda_{t,t+1}.$$

Substituting this expression into (9) yields the equilibrium spatial choice probabilities as

$$\rho_{jrt}^h = \frac{exp\left(\frac{1}{\kappa}\left(\mathscr{W}_{rt}^h - MC_{jr}\right)\right)}{\sum_{l=1}^R exp\left(\frac{1}{\kappa}\left(\mathscr{W}_{rt}^h - MC_{jl}\right)\right)}, \tag{14}$$

where

$$\mathscr{W}_{rt}^h = E^h[U_r] + A_{rt} = \frac{\Gamma_{\eta/\zeta}}{\eta}\psi(r_{t+1})^{\eta-1}\left(\Theta_r^h\right)^\eta + A_{rt}, \tag{15}$$

i.e. $\mathscr{W}_{rt}^h$ is a measure of "regional attractiveness" (in particular, the average utility of an individual in region r). Note that the cross-sectional variation in $\mathscr{W}_{rt}^h$ stems from differences in average earnings, Θ_r^h, and in amenities A_{rt}. The former is endogenous and depends on the extent of spatial sorting and aggregate demand conditions. The latter is fully exogenous.

The expression in (14) has two important properties. First, note that it does *not* feature $\Lambda_{t,t+1}$, which is constant across locations and hence does not determine spatial labor flows. This additive separability of future prices embedded in $\Lambda_{t,t,+1}$ is crucial, because it turns individuals' optimal location choices into a static problem.[18] Moreover, for a given interest rate r_{t+1}, spatial flows *only* depend on the endogenous vector of regional income $\{\Theta_{rt}^h\}_r$, which are determined from static labor market clearing conditions. This structure allows us to calculate the transitional dynamics in the model with a realistic geography, i.e. with about 700 regions.

Finally, this expression formalizes the three determinants of the spatial reallocation component of the structural transformation. The extent to which individuals leave agricultural areas depends on (1) the size of the moving costs MC_{jr}, (2) the correlation between initial agricultural employment shares s_{rAt}^h and future regional attractiveness $\mathscr{W}_{rt+1}^h$ and (3) the importance of idiosyncratic shocks, parametrized by κ. The correlation between agricultural employment and population outflows will be large if $\mathscr{W}_{rt+1}^h$ and s_{rAt}^h are negatively correlated and the elasticity of moving flows with wage differences is large, i.e. κ is small.

Equilibrium Aggregation and Aggregate Structural Change

The spatial equilibrium of economic activity is of course driven by aggregate demand and supply conditions. Our economy does not admit a representative consumer, since the PIGL preference specification in (7) falls outside of the Gorman class. In particular, consider a set of individuals $i \in \mathscr{S}$, with spending

[18] Note that our assumption of frictionless trade is important for this result - with unrestricted trade costs, future prices would be location specific and $\Lambda_{t,t+1}$ would not be constant across space.

e_i. The aggregate demand for agricultural products of this set of consumers is given by

$$PC^A_{\mathscr{S}} = \int_{i\in\mathscr{S}} \vartheta_A(e_i,P)\, e_i di = \left(\phi + \nu\left(\frac{p_A}{p_M}\right)^{\gamma} \int_{i\in\mathscr{S}} e_i^{-\eta}\omega_i di\right) E_{\mathscr{S}},$$

where $E_{\mathscr{S}} = \int_{i\in\mathscr{S}} e_i di$ denotes aggregate spending and $\omega_i = e_i/E_{\mathscr{S}}$ is the share of spending of individual i. Hence, as long as preferences are non-homothetic, i.e. as long as $\eta > 0$, aggregate demand does not only depend on aggregate spending $E_{\mathscr{S}}$ and relative prices, but rather the entire distribution of spending $\{e_i\}_i$ matters. Characterizing the aggregate demand function in our economy, which features ample heterogeneity through individuals' skills, their location choice (which determines the factor prices they face) and the actual realization of the skill vector z^i, is therefore in principle non-trivial.

Our model, however, delivers tractable expressions for the economy's aggregate quantities. The distributional assumption on individual skills implies that individual income y^i is Frechet distributed. Combined with Lemma 1, which showed that individuals' expenditure policy functions are linear, this implies that individual spending e is *also* Frechet distributed. This allows us to solve for the aggregate demand system explicitly as a function of fundamentals.

In particular, it easy to verify that the aggregate expenditure share on agricultural good of this set of consumers $\mathscr{S}$ is given by[19]

$$\vartheta_A\left([e_i]_{i\in\mathscr{S}}, P\right) \equiv \frac{PC^A_{\mathscr{S}}}{E_{\mathscr{S}}} = \phi + \tilde{\nu}\left(\frac{p_A}{p_M}\right)^{\gamma} E_{\mathscr{S}}^{-\eta}, \tag{16}$$

where $E_{\mathscr{S}}$ is mean spending and $\tilde{\nu} = \nu\Gamma\left(1-\frac{1-\eta}{\zeta}\right)/\Gamma\left(1-\frac{1}{\zeta}\right)^{1-\eta}$ is a constant. Hence, the *aggregate* demand system resembles that of a representative consumer with mean spending $E_{\mathscr{S}}$ and an adjusted preference parameter $\tilde{\nu}$.

We exploit this "almost" aggregation property intensely in computing the model. Recall that our economy, consists of $2\times 2\times R$ distinct sets of the consumers - two generations, two skill types and R locations. Letting $E^{h,Y}_{rt}$ and $E^{h,O}_{rt}$ denote the mean spending of the young and generation with skills h in region r, Lemma 1 implies that

$$E^{h,Y}_{rt} = \psi(r_{t+1})\Gamma_\zeta \Theta^h_{rt} \text{ and } E^{h,O}_{rt} = (1+r_t)(1-\psi(r_t))\Gamma_\zeta \Theta^h_{rt-1}.$$

Hence, the aggregate amount of consumer spending on agricultural goods is given by

$$PC^A_t = \sum_{r,h}\left(\vartheta_A\left(E^{h,Y}_{rt},P\right)E^{h,Y}_{rt}\lambda^h_{rt}L_{rt} + \vartheta_A\left(E^{h,O}_{rt},P\right)E^{h,O}_{rt}\lambda^h_{rt-1}L_{rt-1}\right), \tag{17}$$

where $\vartheta_A(E_S,P)$ is the aggregate spending share given in (16) and $\lambda^h_{rt}L_{rt}$ denotes the total number of individuals of skill-type h in region r at time t. Equations (16) and (17) highlight three properties. First,

[19] See Section B.3 in the Appendix.

despite the fact that there is no strong aggregation in the Gorman sense, we can calculate aggregate consumer demand side directly from current and past average regional incomes $\{\Theta^h_{rt}, \Theta^h_{rt-1}\}_r$ and the interest rate r. Secondly, these equations highlight the usual demand side forces of structural change: to the extent that $\eta > 0$, i.e. preferences are non-homothetic, the agricultural consumption share, PC^A_t/E_t will decline as regional incomes, Θ^h_{rt}, rise. Finally, our model captures a distinct role for the *supply* of agricultural labor. Because labor markets are local in nature and regions differ in terms of their comparative advantage, the geographic distribution of workers determines the aggregate supply of agricultural products. A shift of people towards urban areas will for example reduce the relative supply of agricultural goods at given prices. Spatial mobility costs are therefore one example of cohort-specific "sectoral moving costs" highlighted by Hobijn et al. (2017) and Porzio and Santangelo (2017).

Equilibrium Conditions

As highlighted above, the key properties of our model are that (i) individual moving decisions are static and (ii) that our economy generates an aggregate demand system as a function of regional incomes $\{\Theta^h_{rt}\}_r$. This implies that, for a given path of interest rates $\{r_t\}_t$, we can calculate the equilibrium by simply solving a set of static equilibrium conditions.

Consider first the goods market. The market clearing condition for the agricultural good is given by

$$\sum_h w_{rA} H^h_{rA} \quad = \quad (1-\alpha)\,\pi_{rAt}\left(PC^A_t + \phi I_t\right), \tag{18}$$

i.e. total agricultural labor earnings in region r are equal to a share $1-\alpha$ of total agricultural revenue in region r.[20] This in turn is equal to region r's share, π_{rAt}, in aggregate spending on agricultural goods $PC^A_t + \phi I_t$. The CES structure of consumers' preferences implies that regional trade shares π are given by

$$\pi_{rAt} = \left(\frac{P_{rAt}}{P_{At}}\right)^{1-\sigma} = \frac{\left(Q_{rAt} w^{\alpha-1}_{rAt}\right)^{\sigma-1}}{\sum_{j=1}^R \left(Q_{jAt} w^{\alpha-1}_{jAt}\right)^{\sigma-1}}, \tag{19}$$

i.e. they neither depend on the identity of the sourcing region, nor the equilibrium capital rental rate R_t or the common component of productivity Z_{st}. Rather, a region r's agricultural competitiveness only depends on its productivity Q_{rAt} and the equilibrium price of labor in the agricultural market.

Now recall that spatial labor supply and aggregate agricultural demand PC^A_t is only a function of the equilibrium levels of regional income $\{\Theta^h_{rt}\}_r$ (see (9) and (17)). Similarly, we can derive the equilibrium

[20] It is easy to verify that regional agricultural labor income is given by

$$\sum_h w_{rAt} H^h_{rAt} = L_{rt}\Gamma\left(1-\frac{1}{\zeta}\right) w^{\zeta}_{rAt} \sum_h \lambda^h_r \Psi^h_s \left(\Theta^h_r\right)^{1-\zeta}.$$

level of investment spending I_t from the dynamic accumulation of capital. Because the future capital stock is only given by the savings of the young generation, we get that capital accumulates according to

$$K_{t+1} = (1-\psi(r_{t+1}))\sum_{r,h} Y_{rt}^h \lambda_{rt}^h L_{rt} = (1-\psi(r_{t+1}))(1-\alpha) PY_t. \tag{20}$$

Hence, future capital is simply a fraction $1-\psi(r_{t+1})$ of aggregate labor earnings, which in turn can directly be calculated from regional earnings.[21] This proportionality between the aggregate capital stock and aggregate GDP is a consequence of the linearity of agents' consumption policy rules. Investment spending is then given as the residual $I_t = K_{t+1} - (1-\delta)K_t$. Hence, together with the corresponding labor market clearing conditions for the non-agricultural sector, these equations directly determine the equilibrium levels of regional income $\{\Theta_{rt}^h\}_r$.

It is worthwhile to point out that our model retains many features of the baseline neoclassical growth model. In particular, for given initial conditions $\{K_0, L_{r,-1}, w_{r-1}\}$ and a path of interest rates $\{r_t\}_t$, the equilibrium evolution of wages and people (by skill type) are solutions to the static equilibrium conditions highlighted above. Given these allocations, the model predicts the evolution of the capital stock according to (20). A dynamic equilibrium requires that the set of interest rates $\{r_t\}_t$ is consistent with the implied evolution of the capital stock. More formally, a competitive equilibrium in our economy is defined in the usual way.

Definition 2. *Consider the economy described above. Let the initial capital stock K_0, the initial spatial allocation of people $\{L_{r,-1}\}$ and the vector of past wages $\{w_{r,-1}\}$ be given. A dynamic competitive equilibrium is a set of prices $\{P_{rst}\}$, wages $\{w_{rt}\}$, capital rental rates $\{R_t\}$, labor and capital allocations $\{L_{rst}, K_{rst}\}$, consumption and saving decisions $\{e_{rt}^Y, e_{rt}^O, s_{rt}\}$, and demands for regional varieties $\{c_{rst}\}$ such that consumers' choices $\{e_{rt}^Y, e_{rt}^O, s_{rt}\}$ maximize utility, i.e. are given by (11) and (12), the demand for regional varieties follows (19), firms' factor demands maximize firms' profits, markets clear, the capital stock evolved according to (20) and the allocation of people across space $\{L_{r,t}^Y\}$ is consistent with individuals' migration choices in (9).*

3.3 The Urban Bias of the Structural Transformation

A key property of our theory is the spatial bias of the structural transformation: rural areas are hurt by the secular change in spending patterns. This bias is of course at the heart of the induced spatial reallocation, whereby declining wages push people out of agricultural areas. At the same time, local productivities are time-varying. Such regional productivity shocks have the potential to reinforce or reduce the correlation between agricultural employment shares and future earnings. To see these two mechanisms more

[21] In particular, $(1-\alpha)PY_t = \sum_{r,h} \Gamma_\zeta \Theta_{rt}^h \lambda_{rt}^h L_{rt}$.

formally, consider a special case of our model where there are no moving costs, individuals have no locational preferences and there is no heterogeneity in sector-specific skill supply.[22] This parametrization of our model implies that equilibrium wages are equalized across space and sectors. The equilibrium conditions for the agricultural good (18) are given by

$$w_t s_{rAt} L_{rt} = (1-\alpha)\,\pi_{rAt}\,\vartheta_t^A\,PY_t,$$

where PY_t denotes aggregate GDP and $\vartheta_t^A \equiv \frac{PC_t^A + \phi I_t}{PY_t}$ is the aggregate spending share on agricultural goods. Note also that this specification of the model implies that equilibrium sectoral trade shares, π_{rst} are given by $\pi_{rst} = Q_{rst}^{\sigma-1}$, i.e. are fully exogenous and only depend on regional productivity (see (19)). Using the analogous equilibrium condition for the non-agricultural sector, it is easy to show that the equilibrium allocations satisfy

$$\begin{aligned} L_{rt} &= Q_{rAt}^{\sigma-1}\vartheta_t^A + Q_{rNAt}^{\sigma-1}\left(1-\vartheta_t^A\right) \\ \frac{s_{rAt}}{1-s_{rAt}} &= \frac{\pi_{rAt}\,\vartheta_t^A}{\pi_{rNAt}\left(1-\vartheta_t^A\right)} = \left(\frac{Q_{rAt}}{Q_{rNAt}}\right)^{\sigma-1}\frac{\vartheta_t^A}{1-\vartheta_t^A}. \end{aligned}$$

Hence, the local population size is a *demand-weighed* average of local sectoral productivities and regional (scaled) agricultural shares $\frac{s_{rAt}}{1-s_{rAt}}$ are proportional to the aggregate (scaled) agricultural expenditure share $\frac{\vartheta_t^A}{1-\vartheta_t^A}$. Importantly, the region-specific factor of proportionality is fully determined from the technological comparative advantage in agricultural production. Also, like in standard aggregate models of the structural transformation, aggregate value added and employment shares are equalized, i.e. $\vartheta_t^A = \sum_r s_{rAt} L_{rt} \equiv s_{At}$.

These equations concisely illustrate the urban bias of the structural transformation and the determinants of the spatial reallocation component. In particular, the structural transformation can be summarized by a decline in the sectoral spending share, $d\vartheta_t^A < 0$. The time series change in the local population is then given by

$$dL_{rt} = \underbrace{\left(Q_{rAt}^{\sigma-1} - Q_{rNAt}^{\sigma-1}\right) d\vartheta_t}_{\text{Urban Bias}} + \underbrace{\sum_s \vartheta_t^s\, dQ_{rst}^{\sigma-1}}_{\text{Local Produtivity Growth}}. \tag{21}$$

The first term on the right is what we refer to as the urban bias of the structural transformation: A decline in the spending share on agricultural goods reduces population in region r if and only if region r has a

[22] Formally, this parametrization is nested in our full model by assuming that $MC_{jr} = 0$, $A_r = 0$, $\kappa = 0$, $\zeta \to \infty$ and $q = \mu = 1$.

comparative advantage in the agricultural sector. In fact, it is easy to show that[23]

$$Q_{rAt}^{\sigma-1} > Q_{rNAt}^{\sigma-1} \quad \text{if and only if} \quad s_{rAt} > \vartheta_t^A = s_{rAt}.$$

Hence, the structural transformation is biased against regions in which the agricultural employment share *exceeds* the aggregate share of agricultural employment.

The second term of equation 21 clarifies the role of local productivity growth. Whether or not the *change* in productivity dQ_{rst} is correlated with the initial agricultural employment share depends on the stochastic process of productivity. Empirically, we find that this relationship is quite noisy: The local productivity growth channel largely offsets the urban bias channel to make the correlation between initial agricultural employment shares and population growth weak (as they are in the data).

The expressions above also show the extent to which the structural transformation affects the local economy as the change in (scaled) agricultural employment shares is given by

$$d\ln\left(\frac{s_{rAt}}{1-s_{rAt}}\right) = d\ln\left(\frac{\vartheta_t^A}{1-\vartheta_t^A}\right) + (\sigma-1)\, d\ln\left(\frac{Q_{rAt}}{Q_{rNAt}}\right).$$

In the absence of local productivity shocks, *all* regions' degree of sectoral specialization moves in the same direction, i.e. the process of structural change indeed transforms all regions in the economy. Hence, even if local productivity was constant and all changes were due to the structural transformation, the economy would both reallocate people *and* change the extent of regional specialization in all regions in the economy. Idiosyncratic shocks to comparative advantage break this tight relation across space as increases in relative agricultural productivity Q_{rA}/Q_{rNA} will increase agricultural employment despite the secular agricultural decline in the aggregate.

In our full model, three additional channels are at work. First , costly spatial mobility prevent wages from being equalized. Second, local amenities are an additional determinant of moving flows. Finally, idiosyncratic shocks reduce the sensitivity of moving flows to wage differentials. In the next section, we quantify the importance of all these different channels for the experience of the US economy.

4 Spatial Structural Change in the US: 1880 - 2000

We now calibrate the model to key moments of the spatial growth experience of the United States over the last 120 years. We do so by exploiting a novel panel data set on the regional development of the United States between 1880 and 2000.

[23] To see this, note that

$$Q_{rAt}^{\sigma-1} - Q_{rNAt}^{\sigma-1} = \pi_{rAt} - \pi_{rNAt} = \frac{s_{rAt}L_{rt}}{\vartheta_t^A} - \frac{s_{rNAt}L_{rt}}{1-\vartheta_t^A} = \frac{s_{rNAt}}{1-\vartheta_t^A}\left(\frac{s_{rAt}/\vartheta_t^A}{(1-s_{rAt})/(1-\vartheta_t^A)} - 1\right)L_{rt}$$

4.1 Data

To compose our panel we draw on various data sets published by the US Census Bureau. In particular, we use information from the Census of Manufacturing for 1880 and 1910, the Population Census for 1880-2000 and the County and City Data Books for 1940-2000. From these sources we construct a panel data set of total workers $\{L_{rt}\}_{rt}$, average manufacturing wages $\{w_{rt}\}_{rt}$, and sectoral employment shares $\{s_{rAt}\}_{rt}$ for all US counties at 30 year intervals between 1880 and 2000.[24] We define the agricultural sector to comprise agriculture, fishing and mining industries following the 1950 Census Bureau industrial classification system (outlined in Ruggles et al. (2015)). All remaining employed workers are assigned to the non-agricultural sector. We construct average manufacturing wages from county level total manufacturing payroll data and manufacturing head counts obtained from the same source. Table (17) in Appendix (C) contains a comprehensive list of all data sources and more details on the construction of the data set.

We aggregate this county-level data to the level of commuting zones as constructed by Tolbert and Sizer (1996).[25] We do so for three reasons. First, we need stable regional boundaries over time. Second we want the union of all regions to cover the entire continental United States. Lastly, regions should be large enough such that both an agricultural and a non-agricultural sector is present in each region. Commuting zones partition the territory of the United States into 741 polygons, all of which exhibited non-zero non-agricultural employment shares in 1880.

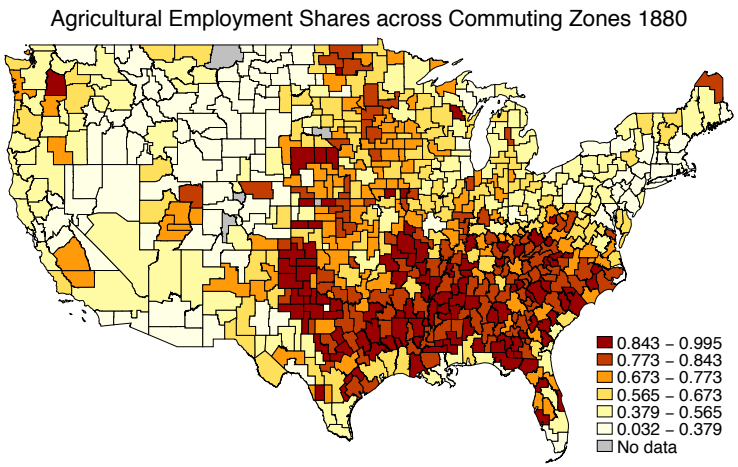
This procedure leaves us with a panel data set, which features sectoral employment shares, total employment and average manufacturing earnings for all continental commuting zones from 1880 to 2000. As a visual example, consider Figure 2, where we depict the distribution of agricultural employment shares across commuting zones. While some regions (in particular, commuting zones in Northeastern states like Massachusetts or New York) already have agricultural employment shares of less than 10%, many commuting zones in the South have more than 75% of their population employed in the agricultural sector. For the main calibration of the model we employ the cross-sections 1880, 1910, 1940, 1970 and 2000 only and normalize the size of the total US workforce to unity in each period.[26]

In addition, we rely heavily on the 1940 edition of the decennial Micro Census by the US Census Bureau (see Ruggles et al. (2015)). This is the most recent Census which contains individual identifiers for all

[24] Twelve states have not obtained statehood in 1880. We list them here and give the year of their accession to the Union: 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) and Hawaii (1959). We exclude Alaska, Hawaii and Washington D.C. The 1880 Census does report data for counties in all states, even those that had not yet officially obtained statehood in 1880, with two exceptions: Oklahoma and Hawaii. We impute 1880 data for Oklahoma's counties using a procedure described in Appendix (C).

[25] To do so, we construct a crosswalk between counties and 1990 commuting zones for every decade between 1880 and 2000. Using this we re-aggregate the county level data for the various years to commuting zones, employing area weights to allocate workers wherever counties are split.

[26] A detailed description of the construction of the county to commuting zone cross-walk for 1790-2000 as well as a panel of the populations of US commuting zones for that period is made available on the authors' website (Eckert et al. (2018)).



Notes: The figure shows the agricultural employment share across US commuting zones in 1880.

Figure 2: Agricultural Employment Share across US Commuting Zones in 1880

US counties and it is the first Census for which earnings and education variables are available.[27] We use this information to calibrate the spatial distribution of skilled workers.

To estimate the moving cost parameters we exploit the information in the Population Census. In the model, individuals move once in their lifetime to access their preferred labor market. In the data we therefore look at patterns of lifetime mobility. All censuses between 1880 and 2000 contain information on the state of residence and the state of birth. Focusing on workers between 26 and 50 years of age, we can therefore calculate state-by-state life-time mobility matrices for the periods 1880-1910, 1910-1940, 1940-1970 and 1970-2000.

Finally, we use micro-data on expenditure patterns from the 1930s to estimate consumer preferences. The Consumer Expenditure Survey in 1936 ("Study of Consumer Purchases in the United States, 1935-1936") contains detailed information on individual expenditure and allows us to calculate the expenditure share of food. We exploit this cross-sectional information on expenditure shares and total expenditure to estimate the extent of non-homotheticities in demand. Our information for the time-series of relative prices it taken from Alder et al. (2017).

4.2 Calibration

Calibration Strategy

In this section, we outline our calibration strategy. Section B in the Appendix provides considerably more detail.

[27] We use the Public-use Micro Census data compiled and maintained Ruggles et al. (2015). In the publicly available samples counties are censored and only become available after 70 years. As a result we cannot identify all counties in Census cross-sections beyond 1940, which is the cross-section most recently de-censored (2010).

Aggregate time series We calibrate the model such that *aggregate* income per capita grows at a constant rate and that the capital-output ratio is constant. In Section B.4 in the Appendix, we show that this implies that interest rates are constant and have a closed-form expression.[28] We calibrate the time series of aggregate sectoral productivities, $\{Z_{At}, Z_{Mt}\}_t$, to match the evolution of relative prices and a GDP growth rate of 2%.

Skill Supply To parametrize the skill supply, we need values for the supply elasticity (ζ), the comparative and absolute advantage of skilled workers (q and μ) and the initial distribution of skilled workers across space in 1880, $\{\lambda_{r1880}\}_r$. We define skilled individuals as workers who completed at least a high school education in 1940 and hold the aggregate share of skilled workers fixed.[29] This choice yields an aggregate skilled employment share of about 0.3. We then calibrate $\{\lambda_{r1880}\}_r$ for the model to exactly replicate the the spatial skill distribution in 1940, which is the first year for which we can calculate educational attainment at the commuting zone level. We calibrate the parameter ζ, i.e. the dispersion of individual productivity, to match the dispersion of earnings in the 1940 Census data. In particular, the model implies that the variance of log earnings within region-skill cells is given by $\frac{\pi^2}{6}\zeta^{-2}$. We therefore identify ζ from $\zeta = \frac{\pi}{6^{1/2}}\text{var}\left(\hat{u}^i_{rsh}\right)^{-1/2}$,where $\text{var}\left(\hat{u}^i_{rsh}\right)$ is the variance of the estimated residuals from a regression of log earnings on region, sector and skill-group fixed effects in the 1940 Census Data.

The two parameters q and μ in turn are chosen to match the aggregate skill premium and the aggregate relative manufacturing employment share of skilled workers in 1940. We calculate the skill premium simply as the ratio of average labor earnings of skilled relative to unskilled individuals. Similarly, we compute the relative manufacturing employment share of skilled workers as the non-agricultural employment share of skilled workers relative to the one of unskilled workers. Note that these measures already incorporate the unbalanced spatial sorting of skilled and unskilled individuals, i.e. they take into account that skilled workers live in high-wage and manufacturing intensive localities.

Moving Costs and Idiosyncratic Location Preferences We specify the costs of moving, denominated in utils, as a quadratic function of distance, d_{jr}, i.e.

$$MC_{jr} = \tau + \delta_1 d_{jr} + \delta_2 d_{jr}^2$$

whenever $j \neq r$ and zero otherwise.Because in the theory workers move only once (at the beginning of their working life), for the remainder of the paper all mention of migration refers to lifetime migration.[30] Here,

[28] In the counterfactual analysis, interest rates will of course be free to vary over time.

[29] Because we focus on the spatial aspects of the structural transformation, we abstract from skill deepening. The model could easily be extended to allows for changes in the aggregate supply of human capital.

[30] The aggregate lifetime migration rate refers to the fraction of people who live in a different location than where they were born (see Molloy et al. (2011)).

$\tau > 0$ corresponds to a fixed cost of moving and δ_1 and δ_2 flexibly parameterize how mobility costs vary with distance. We normalize d_{jr} so that the maximum distance in the US is 1.[31]

In the Census data we observe the state of birth and the current state of residence of every worker in every decade. In the model we aggregate commuting zone flows to the state level and choose $(\kappa, \tau, \delta_1, \delta_2)$ so as to make the model fit the state level flows between 1910 and 1940 as closely as possible.[32] We chose τ so as to match the aggregate lifetime interstate migration rate between 1910 and 1940 exactly.[33] We then chose $(\kappa, \delta_1, \delta_2)$ to minimize the distance between spatial mobility rates in the data and the model.[34]

Productivities and Amenities We calibrate local productivities and amenities $\{Q_{rst}, A_{rt}\}_{rst}$ as structural residuals, i.e. we force the model to match the spatial data on agricultural employment shares, populations and average manufacturing wages perfectly with $\{Q_{rst}, A_{rt}\}_{rst}$ absorbing any residual variation.

In Appendix (B.1) we show that there is a unique mapping from the observed spatial data on agricultural employment shares, populations and average manufacturing to the vector of local productivity $\{Q_{rst}\}_{rst}$. Intuitively, the local employment shares contain information on $\frac{Q_{rAt}}{Q_{rNAt}}$ while the level of wages along with the total number of workers informs the level of Q_{rNAt}. The vector of amenities $\{A_{rt}\}_{rt}$ can then be inferred from the observed population flows together with evolution of skill shares.

Preference parameters We use the micro data from the Consumer Expenditure Survey in 1936 ("Study of Consumer Purchases in the United States, 1935-1936") to estimate the extent of non-homotheticities, i.e. η from the cross-sectional relationship between sectoral spending shares and the level of expenditure. We then use the time-series of the aggregate agricultural employment share to identify the remaining parameters (ϕ, ν, γ). As γ determines the price elasticity of demand, we discipline γ with the elasticity of substitution. Comin et al. (2015) estimate this elasticity to be around 0.7 in post-war data for the US.

Given our calibration strategy for the underlying distribution of productivity and amenities , internal consistency requires us to match the time series of agricultural employment shares exactly. As we discuss in detail in Section A.5 in the Appendix, the income effects as implied from the cross-sectional spending-food relationship are not strong enough to explain the entire decline in agricultural employment in the

[31] Before this normalization the largest distance between two continental commuting zone centroids is 2827.4 miles.

[32] The main reason we chose to calibrate to the 1910-1940 period is that we observe 1940 skill shares in the data which allows us to compute 1940 sectoral wages and human capital stocks and hence expected utility $E^h[U_j]$ without having to first back out skill share for 1910. Calibrating to any other pair of cross-sections, would entail simulating the model starting in 1940 to obtain skill shares in the respective end year.

[33] As shown in Section B.7 in the Appendix, the number of stayers in a commuting zone is a monotone function in τ given $(\kappa, \delta_1, \delta_2)$.

[34] More specifically, we chose $(\kappa, \delta_1, \delta_2)$ to minimize $\sum_i \sum_{j \neq i} L_{j,1940} \times \left(\log \rho_{ij,1940}^{DATA} - \log \rho_{ij,1940}^{MODEL}(\kappa, \delta_1, \delta_2)\right)^2$ conditional on always exactly matching the aggregate interstate migration rate through the choice of τ.

time-series.[35] We therefore allow the parameter ϕ to be time-specific to fully account for the residual decline in agricultural employment and choose ν to minimize the required time-variation in ϕ_t. Intuitively, ν is chosen for the model to explain as much of the aggregate process of structural change as possible, given the income and price elasticities η and γ. Recall that ϕ does not enter the household's decision problem directly. Finally, we chose the rate of time preference β to be consistent with the aggregate rate of investment.

Other parameters Finally, we need values of the capital share α, the rate of depreciation δ and the preference parameter σ. We do not estimate these parameters within our model, but rely on central values from the literature.

Calibration Results and Model Fit

In Table 1 we report the calibrated parameters and the main targeted moment, both in the data and the model. Naturally, the parameters are calibrated jointly.

The presence of *spatial mobility costs* is an important component of our theory. To get a sense of how well the model matches important features of the lifetime migration data, we report the distribution of stayers across states and the relationship between moving flows and distance in Figure 3. In the left panel we depict the cross-sectional distribution of the share of "stayers", i.e. ρ_{rr}, both in the data and the model. Because regions differ both in attractiveness, i.e. the utility they provide to their residents, and in distance to more attractive places, there is sizeable heterogeneity in the extent to which regions retain workers. Figure 3 shows that the model matches this cross-sectional heterogeneity, even though it is only calibrated to match the average rate of mobility. In the right panel, we show that the model also matches the distance gradient of moving flows. In particular, we run a gravity-type regression of moving flows and compare the model outcomes to those of the data. More specifically, we consider the specification

$$\log \frac{\rho_{jr}}{1-\rho_{jj}} = \alpha_j + \beta_r + u_{jr},$$

where α_j and β_r are origin and destination fixed effects. We then plot the estimated residual $\hat{u}_{jr}$ as a function of distance for the actual data and the data stemming from the model. The right panel of Figure 3 shows that the model captures this systematic pattern of lifetime migration flows well. In particular, spatial mobility is very local, i.e. spatial flows are steeply decreasing in distance.[36]

In terms of *consumers' preferences*, note that the demand system implies that expenditure shares at the individual level are given by $\vartheta_A(e,p) = \phi + \nu \left(\frac{p_A}{p_M}\right)^{\gamma} e^{-\eta}$. For $\phi \approx 0$, this implies that there is a log-linear

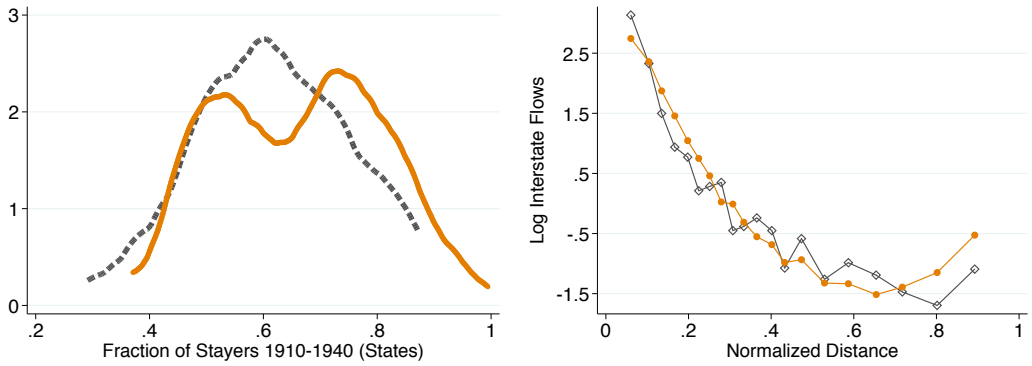
[35] This discrepancy between the cross-section and time-series is not particular to our application. For example, the results reported in Comin et al. (2015) also imply different estimates for the income elasticity stemming from the cross-section and the time-series. While reconciling this discrepancy between the cross-section and the time-series is an important open research question, it is not the main focus or our paper.

[36] To get a sense of the *economic* magnitude of our estimates of moving costs, consider an individual of skill h born in region

Parameter		Target	Value	Moments	
				Data	Model
		Skill Supply			
ζ	Skill heterogeneity	Residual Earnings variance in 1940	1.62	0.62	0.62
ξ	Share of skilled individuals	Share with more than high school degree in 1940	0.3	0.3	0.3
μ	Comparative advantage	Rel. non-ag. share of skilled workers in 1940	3.41	1.21	1.21
q	Absolute advantage	Skill premium in 1940	0.68	1.62	1.61
$[\lambda_{r,1880}]$	Initial distribution of skilled individuals	Skill distribution in 1940	.	$\{\lambda_{r,1940}\}$	match data perfectly
		Regional Fundamentals			
$[Q_{rAt}]_{rst}$	Agricultural productivity	Regional empl. shares and earnings	See Appendix	$\{e^{Man}_{r,t}, s_{A,rt}\}$	match data perfectly
$[Q_{rNAt}]_{rst}$	Non-agricultural productivity				
$[A_{rt}]_{rt}$	Amenities	Regional population		$\{L^h_{r,t}\}$	
		Time Series Implications			
$[Z_{NAt}]$	Non-agricultural productivity	Aggregate growth rate of GDP pc	See Appendix	2%	2%
$[Z_{At}]$	Agricultural productivity	Relative price of ag. goods	See Appendix	$\{P_{A,t}/P_{NA,t}\}$	match data perfectly
		Preference Parameters			
β	Discount rate	Investment rate along the BGP	0.29	0.15	0.15
ϕ	Ag. share in price index	Time series of ag. empl. share	See Appendix	$\{s_{A,t}\}$	match data perfectly
ν	PIGL Preference parameter	Time series of ag. empl. share	0.017	0.20	0.20
η	Non-homotheticity	Ag.share - expenditure relationship	0.35	0.04	0.04
γ	Price sensitivity	Elasticity of substitution in 2000	0.35	0.7	0.7
		Moving Costs and Mobility			
τ	Fixed costs of moving	Lifetime interstate migration rate	1.63	0.32	0.32
κ	Dispersion of idiosyncratic spatial tastes	Observed state-to-state flows	0.42	Estimated with NLS	
$[\delta_1, \delta_2]$	Distance (miles) elasticity of moving costs		$[8.44, -6.39]$	Estimated with NLS	
		Other parameters			
δ	Depreciation rate (over 30 years)	set exogenously	0.91 (0.08 annually)	.	
α	Capital share in production function	set exogenously	1/3	.	.

Notes: The table contains the calibrated parameters. See Section 4.2 for the calibration strategy.

Table 1: Calibrated structural parameters



Notes: In the left panel we plot the distribution of the share of people staying in their home state between 1910 and 1940 in the data (grey dotted line) and model (orange solid line). To construct the right panel, we run a gravity equation of the form

$$\log \frac{\rho_{jr}}{1-\rho_{jj}} = \alpha_j + \beta_r + u_{jr}$$

where ρ_{jr} denote the share of people moving from j to r and α_j and β_r denote origin and destination fixed effects. We run the regression both in the model (orange dots) and in the data (grey diamonds) and then plot the average $\hat{u}_{jr}$ by distance percentile.

Figure 3: Lifetime state-to-state migration: Model vs Data

relationship between expenditure shares and total expenditure. In Figure 4 we depict the cross-sectional distribution of the expenditure share for food (left panel) and the binned scatter plot between (the log of) expenditures and expenditure shares after taking our a set of regional fixed effects, to control for relative prices. The slope of the regression line is exactly the extent of the demand non-homotheticity η. From this plot, it is clear that there is substantial heterogeneity in the expenditure share on food in the cross-section of households. Moreover, the expenditure share is systematically declining in the level of expenditure and the cross-sectional relationship is essentially log-linear, as predicted by the theory. The slope coefficient implies that $\eta = 0.32$.[37]

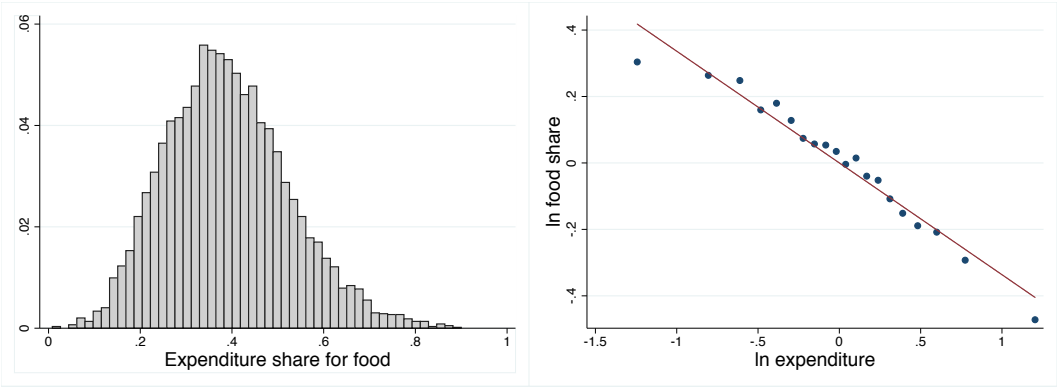
An important part of our calibration strategy is that we calibrate the cross-sectional distribution of sectoral productivities $\{Q_{rAt}, Q_{rNAt}\}_{rt}$ and amenities $\{A_{rt}\}_{rt}$ as structural residuals of the model in line with the new quantitative spatial economics literature (Redding and Rossi-Hansberg, 2016). Hence, given the remaining parameters reported in Table 1, the model matches the population distribution, the regional agricultural employment shares and average manufacturing earnings in each cross-section exactly. We

j. The expected value of moving (to her preferred location) is given by

$$\mathscr{U}_{jt}^{h,Mov} = E\left[\max_{r\neq j}\left\{\mathscr{U}_{jr}^{ih}\right\}\right] = \tilde{\gamma} - \tau + \ln\left\{\left[\sum_{r\neq j}\exp\left(\mathscr{W}_{rt}^{h} - f(d_{jr})\right)^{1/\kappa}\right]^{\kappa}\right\},$$

were τ denotes the fixed cost of moving, $f(d_{jr}) = \delta_1 d_{jr} + \delta_2 d_{jr}^2$ is the component of moving costs, which is distance-specific,and $\tilde{\gamma}$ is a constant, which depends on the Euler constant and $\Lambda_{t,t+1}$. Similarly, the expected value of staying in region j, $\mathscr{U}_{jt}^{h,Stay}$, is given by $\tilde{\gamma} + \mathscr{W}_{jt}^{h}$ as $\tau = f(d_{jj}) = 0$. In our calibrated model, we find that for the median commuting zone, the expected value of moving, $\mathscr{U}_{jt}^{h,Mov} - \mathscr{U}_{jt}^{h,Stay}$ is slightly positive and amounts to roughly 5% of the estimated fixed cost of moving. See also Section C.2 in the Appendix, where we discuss this in more detail.

[37] For our estimate, we of course do not impose the restriction that $\phi = 0$ and estimate the demand function using non-linear least squares. The parameter η is precisely estimated and - depending on the specification - between 0.3 and 0.34.



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. The underlying data stems from theData Source: Study of Consumer Purchases in the United States, 1935-1936.

Figure 4: Expenditure Shares on Food in 1936

find that productive regions have an absolute advantage in *both* sectors - the cross-sectional correlation between $\ln Q_{rA}$ and $\ln Q_{rNA}$ is about 0.5. Similarly, the correlation or productivity and amenities is also positive. Hence, productive places are *also* relatively pleasant places to live, in line with recent direct evidence for developing countries by Gollin et al. (2017). In Section C.2 in the Appendix, we provide additional details about the estimates of spatial fundamentals.[38]

4.3 Why is the reallocation component of the structural transformation small?

In Section 2, and especially Figure 1 we showed that the structural transformation in the US is entirely driven by a within-region transformation - the reallocation of individuals from agricultural regions to non-agricultural regions played only a minor role. This section explains why this is the case.

In our model, individuals spatially relocate to arbitrage away wage differences and to move towards high amenity regions. Hence, spatial mobility contributes to the decline in aggregate agricultural employment, if (i) moving frictions are not prohibitively large, (ii) regional employment shares are negatively correlated with *future* earning and *future* amenities and (iii) idiosyncratic taste shocks are of limited importance for moving decisions.

Observed gross and net moving flows are informative about (i) and (iii). The high level of gross lifetime interstate migration observed throughout the 20th century hints at the fact that moving frictions play a minor role in explaining the absence of the reallocation component in the structural transformation. Similarly, the high level of bi-directional flows in the data, leads the model to assign an important role to idiosyncratic preference shocks, i.e. there is a low elasticity of moving flows with respect to spatial wage

[38] In particular, we show empirically how our spatial fundamentals determine the distribution of agricultural employments shares. In Figure 11 in Appendix A, we also show directly how the productivities implied by the model relate to more direct measures of regional productivity. In particular we show, using their data, that the growth in market access for certain regions documented by Donaldson and Hornbeck (2016) is picked up as an increase in $\{Q_{rA}, Q_{rNA}\}_r$ by our model.

differences. Finally, turning to (ii), our theory highlights three forces, which shape the correlation between initial regional employment shares and future earnings. First of all, to the extent that wages show some persistence and population mobility is slow, individuals will leave rural areas if wages in such regions are low. Secondly, as discussed above, the urban bias of the structural transformation endogenously induces a negative correlation between agricultural employment and future earnings as consumer demand systematically shifts towards non-agricultural goods. Thirdly, the effect of regional productivity shocks depends on the correlation between local productivity growth with agricultural employment shares.

In Table 2 we provide reduced form evidence for each of these margins. In the first two columns, we report the results from a regression of log earnings on the (log of) agricultural employment share. There is a significant negative correlation, especially for high-skilled workers. In columns 3 and 4, we test for evidence of urban bias. In particular, we consider the specification

$$\ln \Theta^h_{rt} = \delta_t + \alpha \ln s_{rAt-1} + \beta \ln s_{rA,t-1} \Delta s_{At} + \gamma \ln \Theta^h_{rt-1} + u_{rt}, \tag{22}$$

where $\ln \Theta^h_{rt}$ denotes regional income, $\ln s_{rAt-1}$ is the lagged agricultural employment share and $\Delta s_{A,t}$ denotes the change in the aggregate agricultural employment share between t and $t-1$. The coefficient α captures the direct effect of agricultural specialization. The coefficient β captures the urban bias and we expect β to be positive: the larger the decline in the agricultural share, the more adversely will regions with a comparative advantage in agriculture be affected. Columns 3 and 4 show direct evidence for the urban bias: a decline in agricultural employment share particularly hurts regions with a large initial agricultural shareMoreover, this spatial bias is present for both low and high skilled workers. In columns 5 and 6 we regress the growth rate of local productivity, $d \ln Q_{rst}$, on the initial agricultural employment share and the respective current level of productivity, $\ln Q_{rst}$. Regions with a larger agricultural employment share tend to have higher (lower) productivity growth in the agricultural (non-agricultural) sector. However, even after controlling for the level of productivity, productivity shocks are quite noisy - the R^2 is only 0.4. Finally, the last column shows that rural areas tend to also have low future amenities. This is consistent with Gollin et al. (2017), who provide direct evidence that measured amenities in rural areas are much lower than in cities.[39]

The results in Table 2 show that there are multiple forces at play that determine the relationship between agricultural shares and future earnings. This suggests that the relationship between initial agricultural employment and future earnings might be weak. This is exactly what we find. If, for example, we run a regression of future earnings or amenities on a full set of 100 fixed effects for the percentiles of initial agricultural employment shares, we explain at most 1/3 of the cross-sectional variation.[40]

To quantify the importance of these different margins, let us go back to the "frictionless" parametrization of our model introduced in Section 3.3 and also assume that there are *no* changes in regional productivi-

[39] While column 7 in Table 2 reports the correlation between agricultural employment and future amenities, the correlation with current amenities is also negative.

[40] See Section C.3 in the Appendix for more details.

	Current Earnings		*The Urban Bias*		*Local Productivity Growth*		*Future Amenities*
	$\ln \Theta^L_{rt}$	$\ln \Theta^H_{rt}$	$\ln \Theta^L_{rt+1}$	$\ln \Theta^H_{rt+1}$	$d \ln Q_{rA}$	$d \ln Q_{rNA}$	A_{rt+1}
$\ln s_{rAt}$	-0.059***	-0.149***					
	(0.004)	(0.005)					
$\ln s_{rAt-1}$			-0.032***	-0.036***	0.175***	-0.365***	-0.223***
			(0.005)	(0.005)	(0.036)	(0.055)	(0.018)
$\ln s_{rAt-1} \times \Delta s_{At}$			0.516***	0.475***			
			(0.066)	(0.079)			
Year FE	✓	✓	✓	✓	✓	✓	✓
Lagged Earnings			✓	✓			
Lagged Fundamentals					✓	✓	✓
N	2580	2580	1935	1935	2580	2580	2580
R^2	0.894	0.918	0.908	0.924	0.428	0.409	0.088

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels.

Table 2: Income and Agricultural Specialization

ties.[41] This implies that the urban bias is the *only* driver of spatial mobility (see (21)). When we calibrate this model to the same aggregate time-series data, we find that this model has strikingly different implications for the link between initial agricultural employment shares and subsequent population flows. Consider for example Figure 5, where we depict the share of commuting zones experiencing population outflows between 1910 and 1940 within different deciles of their agricultural employment share in 1910. While the data shows a negative correlation between outflows and initial agricultural specialization, the relationship is noisy: even among the set of the regions with very high agricultural employment shares in 1910, about 20% see net population inflows. Similarly, 60% of the most non-agricultural areas see their population decline.[42] This is very different for the "urban bias" economy, where agricultural specialization and population outflows are perfectly aligned. In particular, this model implies that the *only* regions experiencing population inflows are the 15% of commuting zones with the lowest agricultural share in 1910. Hence, if only the urban bias mechanism was at play, the structural transformation would have induced much more population growth in non-agricultural, urban localities.[43]

This model therefore predicts a much bigger role for population mobility in the process of the structural transformation. This is seen in Figure 6, which has the exact same structure as Figure 1. The spatial reallocation of individuals now explains about one third of the aggregate decline in agricultural employment. Interestingly, as the model *overestimates* the role of spatial reallocation, it *underestimates* the extent of the structural transformation at the local level. In particular, there are more regions which remain dominated by the agricultural sector throughout the 20th century. Consider, for example, the two red densities

[41] This parametrization therefore assumes that there are (i) no moving costs ($MC_{jr} = 0$), (ii) no *changes* in regional fundamentals ($Q_{rst} = Q_{rs,1880}$), (iii) no amenities ($A_{rt} = 0$) and (iv) no idiosyncratic preferences for particular locations ($\kappa \to 0$).

[42] Note that Figure 5 reports the share of regions with net outflows, i.e. the extensive margin of population flows. In the aggregate, net population flows are by construction zero.

[43] Note that this strong form of sorting is implied by (21): only counties with $s_{rAt} < s_{At}$ see their population increase. And because population size and agricultural employment shares are strongly negatively correlated, far less than half of the regions are predicted to experience population inflows.



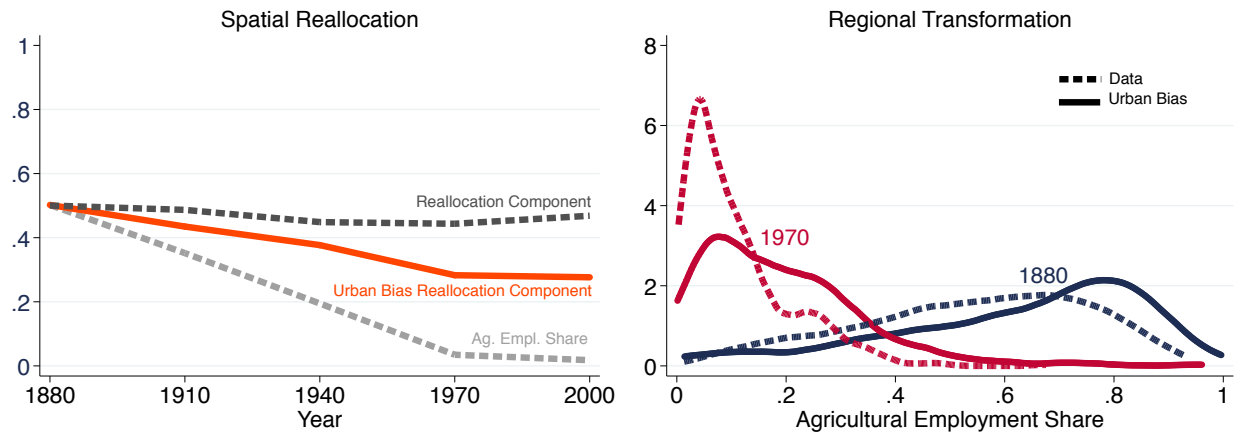
Notes: The figure reports the share of regions within the different deciles of the agricultural employment share in 1910, which experience net population outflows between 1910 and 1940. The first (last) decile refers the commuting zones with lowest (highest) agricultural employment share. We report the results for the data (dark grey) and the urban bias economy, which has no moving costs $\left(MC_{jr}=0\right)$, (ii) no changes in regional fundamentals $\left(Q_{rst}=Q_{rs1880}\right)$, (iii) no amenities $\left(A_{rt}=0\right)$ and (iv) no idiosyncratic preferences for particular locations $(\kappa \to 0)$.

Figure 5: Inflows and Outflows

on the far left, which corresponds to 1970. In the data (the dashed line), the vast majority of commuting zones have an agricultural employment share below 20%. In the model, there is a substantial number of regions, where 30%-40% of the workforce is still employed in the agricultural sector. The reason for this is that worker mobility and local structural transformation are substitutes - the easier it is to reallocate people across space, the more regional specialization can be sustained throughout the structural transformation.

Finally, in Table 3 we provide a decomposition into the different channels. In the first two rows, we again report the results for the data and the "urban bias" economy. While the data hardly shows any decline in the agricultural employment share, the urban bias model implies that the agricultural employment share would have declined from 50% to 35% only through the spatial reallocation of workers. In the next three rows, we shut off different components of our theory. Spatial productivity shocks itself accounts for a sizable part of the "lack of spatial reallocation". In particular, if productivity Q_{rst} was fixed at its level in 1880, the extent of spatial reallocation would have doubled from 3.2% to 6.5%. The reason is that productivity shocks introduce noise in the relationship between earnings growth and initial agricultural specialization.

The next two rows show that neither moving costs nor local amenities can explain the absence of reallocation. A reduction of moving costs would in fact *reduce* the role of spatial reallocation: the agricultural employment share would only fall by 0.6%, which is even less than the 3.2% in the data. The reason is that a reduction in moving costs would not only induce people to move towards regions with high earnings but also trigger more mobility for idiosyncratic reasons. The latter moves are by construction uncorrelated with agricultural specialization and hence do not contribute to the spatial reallocation component of structural change. Similarly, regional amenities are in fact a positive contributor to the spatial



Notes: In the left panel we show the aggregate agricultural employment share (light grey dashed line) and the predicted agricultural share holdings regional agricultural shares at their 1880 level in the urban bias economy (solid orange line). This line is calculated as $\sum_r s^M_{rA1880} l^M_{rt}$, where s^M_{rAt} and l^M_{rt} are the agricultural employment share and the population share of region r at time t in the urban bias economy. For comparison we also depict this object from the data (dark grey dashed line). In the right panel we show the cross-sectional distribution of agricultural employment shares between 1910 and 1970, both in the urban bias economy (solid lines) and in the data (dashed lines).

Figure 6: Spatial Reallocation without frictions

	Spatial growth	Moving costs	Amenities	Idio. shocks	Spatial Reallocation 1880-2000	Migration patterns: Net migration rate	Migration patterns: Turnover (gross/net)
Urban bias model					-15.2%	14.0%	1*
Baseline model	✓	✓	✓	✓	-3.2%	12.9%	2.8
Convergence		✓	✓	✓	-6.5%	10.7%	3.2
Moving Costs	✓		✓	✓	-0.6%	41.1%	2.4
Amenities	✓	✓		✓	+2.6%	6.0%	6.2

Notes: The table decomposes the spatial reallocation component into different margins. The first two rows contain the results for the frictionless and the baseline model. The third row abstracts from productivity convergence ($Q_{rst} = Q_{rs1880}$). The fourth row abstracts from moving costs $\left(MC_{jr} = 0\right)$. The fifth row abstracts from regional amenities $(A_{rt} = 0)$. The frictionless model has no idiosyncratic taste shocks (i.e. $\kappa \to 0$).

Table 3: Why is there no reallocation?

reallocation component. This is due to the fact that rural regions have (on average) low future amenities (see Table 2). Spatial amenities are therefore a force driving people out of agricultural areas.

The final difference between the baseline and the "urban bias" model is the importance of idiosyncratic shocks. The reason why our model infers that idiosyncratic shocks are empirically important is prevalence of *bi-directional flows*. If idiosyncratic shocks were absent, all individuals would agree on the ranking of potential destinations and we would not see people moving in both directions. Hence, the gross flows between any pair of locations would coincide with the net flows. The presence of idiosyncratic shock breaks this implication and gross flows exceed the amount of net flows. In the last two columns of Table 3 we therefore report the net migration rate and the rate of "turnover", i.e. the ratio of the gross and net migration rate. The latter can be interpreted as the number of bodies that have to be moved across space to reallocate one person "on net" between regions.

In the data and in our baseline model, this number is around 3, i.e. if two people move from region j to region r, one individual moves from r to j. Moreover, this number does not decrease dramatically across our different exercises and, if anything, increases. In the absence of moving costs for example, the net migration rate increases by 4, but so does the gross migration rate, keeping the turnover margin stable. If there were no regional amenities, the net migration rate would shrink to 6%, but the gross migration rate would stay roughy constant.[44] This is very different in the "urban bias" economy, where gross flows and net flows coincide. Importantly, the net migration rate between the baseline model and the frictionless economy is roughly the same. Hence, the extent of spatial reallocation is quite similar - but the correlation of net-reallocation and agricultural specialization is much higher in the frictionless economy.[45]

5 The Aggregate Implications of Spatial Structural Change

We now turn to the macroeconomic implications of spatial structural change. We focus on two aggregate outcomes: aggregate productivity and the spatial distribution of welfare. We are particularly interested in quantifying the role of spatial mobility in shaping aggregate outcomes. Both the urban bias of the structural transformation and the stochastic evolution of regional productivities induce changes in the marginal product of labor across locations over time. The ease of labor mobility therefore crucially determines the extent to which such marginal product differences can be kept in check. Labor mobility therefore tends to increase allocative efficiency and reduce spatial inequality.

To quantify the importance of spatial mobility, we compare our baseline model with two alternative parameterizations. First, we consider a model without any labor mobility by making the costs of moving prohibitively high. Secondly, we consider a model, where spatial sorting is costless and based only on local productivity and aggregate demand conditions.More specifically, we assume that both moving costs are absent $\left(MC_{jr} = 0\right)$ and agents do not derive utility from non-pecuniary features of particular locations ($A_r = 0$ and $\kappa = 0$), so that spatial sorting is only based on expected earnings.

We report the outcomes of these three different models in Table 4. In the first row, we verify that all models generate a similar decline in the aggregate agricultural share. Hence, the aggregate "size" of the structural transformation is constant across specifications. In the second row, we report the spatial reallocation component. Without spatial mobility, the reallocation component is naturally equal to zero. In the absence of moving costs and spatial tastes, the reallocation component is positive but still only slightly larger than the one in the data. This again shows the quantitative importance of regional productivity shocks: in Table 3, we showed that the reallocation component was 15% if regional technologies were

[44] Intuitively, amenities induce correlated moving flows across agents. This increases the net migration rate relative to the gross migration rate.

[45] To be precise, in the absence of moving costs, the equilibrium allocation of net flows is unique, but the gross flows are not determined. This indeterminacy vanishes for an arbitrarily small moving cost. In that case, the net migration rate is equal to the gross migration rate.

	Baseline	*Counterfactuals*	
		No Spatial Mobility	No Spatial Frictions
Decline in Agricultural Employment Share	-48.1%	-47.8%	-48.2%
Spatial Reallocation Component	-3.2%	0	-3.7%
Net migration rate	13%	0	22%
Increase in GDP pc (rel. to 1880)	10.78	8.89	11.09
Difference to baseline		-17.57%	2.87%
Increase in TFP (rel. to 1880)	4.88	4.27	4.99
Difference to baseline		-12.59%	2.22%
Agricultural productivity Gap (2000)	0.83	0.73	0.83
Spatial Income Inequality (2000)	0.19 / 0.21	0.58 / 0.58	0

Notes: The model with no spatial mobility corresponds to the baseline model except that individuals are not allowed to move (i.e. $MC_{jr} \to \infty$). The model without spatial frictions corresponds to the baseline model except that moving costs are zero (i.e. $MC_{jr} = 0$), that there are no amenities ($A_{rt} = 0$) and no idiosyncratic tastes (i.e. $\kappa = 0$). The net migration rate is the average migration rate between 1910 and 2000. TFP is measured as $TFP_t = Y_t / \left(K_t^{\alpha} L_t^{1-\alpha}\right)$. The "Agricultural Productivity Gap" is measured as aggregate value added per worker in agriculture relative the rest of the economy, i.e. $\left(VA_t^A / L_t^A\right) / \left(VA_t / L_t\right)$.Spatial income inequality is measured as the spatial IQR in log average earnings ($\ln \Theta_{rt}^h$). We report the results for low and high skilled individual.

Table 4: Reallocation and Aggregate Productivity: Sectors versus Space

held constant at their 1880 level. Conversely, even without spatial amenities and moving costs, the model generates a reallocation component, which is close to the one observed in the data.

In the lower panel, we report the implications for GDP growth and aggregate TFP.[46]Spatial mobility contributed heavily to aggregate productivity growth in the US during the 20th century. Without labor mobility, income per capita would have been lower by almost 17.5% and this decline is entirely accounted for by lower TFP, i.e. more misallocation.[47] [48] If, on the other hand, spatial mobility was only based on regional earnings and mobility was free, aggregate productivity and income per capita would have been higher by about 2.2% and 3% respectively. Compared to the losses from the absence of spatial mobility, these numbers are relatively small. Also note that the frictionless net migration rate is only 22%. This suggests that the 13%, which we see in the data, was sufficient to capture most of the productivity gains through spatial reallocation.[49]

These *spatial* productivity gains are in fact reflected in the aggregate *sectoral* productivity gaps. Both the baseline model and the model without spatial frictions, for example, imply an agricultural productivity gap of around 17% (Gollin et al., 2013). Hence, value added per worker in agriculture is about 17% lower

[46]We calculate TFP from the perspective of an aggregate production function, i.e. $TFP_t = Y_t / \left(K_t^{\alpha} L_t^{1-\alpha}\right)$.

[47]Income per capita is not proportional to TFP because of capital accumulation. As the capital-output ratio is essentially constant in all parametrizations, we have that $Y \propto TFP^{\frac{1}{1-\alpha}}$.

[48]

[49]To put these results in perspective, Bryan and Morten (2015), using a static spatial equilibrium model, estimate higher productivity losses from limited labor mobility for Indonesia. In their specifications without agglomeration and endogenous amenities, which we for simplicity abstract from in our model, they find that productivity would increase by about 17% if moving costs were zero and there were no amenity differences. There are two main reasons, why our results are small. First of all, by explicitly considering capital as a factor of production (which can be traded without frictions) our economy allows for some factor other than labor to adjust. Secondly, Bryan and Morten (2015) argue that the US economy has lower mobility costs than Indonesia.

than for the average worker even though there are no sectoral frictions. These productivity differences are due to skill-based sorting, whereby high-skilled individuals both work in the non-agricultural sector within regions and move towards urban locations spatially. Without spatial mobility this gap would increase to 27%. Hence, even though the spatial reallocation component is seemingly small, it has important aggregate effects for sectoral productivity differences by reducing the spatial dispersion in the marginal product of factors.

One implication of these spatial differences is the regional dispersion in average earnings. In the last line of Table 4, we report the interquartile range of log earnings for low and high skilled individuals across US commuting zones. In the baseline model, this difference amounts to roughly 0.2 log points. Without spatial mobility, labor is inelastically supplied at the region level and spatial inequality would increase by a factor of three and amount to almost 0.6 log points. If, on the other hand, spatial mobility were costless and there were no amenities, utility equalization implies that regional earnings would be equalized within skill groups at each point in time.

These differences in spatial earnings, translate into spatial welfare differences. We focus on $\mathscr{W}_{rt}^h = E^h[U_r] + A_{rt}$ (see (15)) as our measure of regional welfare for skill group h. We showed above that regional income Θ_r^h (which, recall, is the sole determinant of $E^h[U_r]$) and regional amenities are negatively correlated with the agricultural employment share. Hence, welfare is systematically lower in agriculturally intensive places. It is also the case that $\mathscr{W}_{rt}^h$ is affected by the spatial bias of the structural transformation: similar to average income (as reported in Table 2), welfare grows less the more agricultural a place, conditional on a given decline of the aggregate agricultural share.[50]

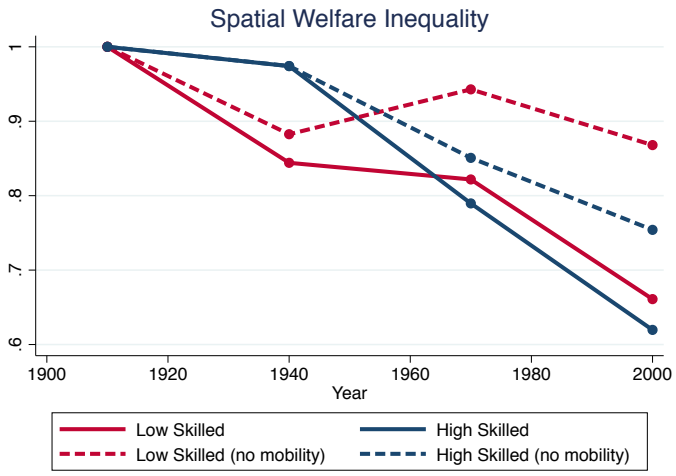
To quantify to the evolution of spatial welfare inequality during the 20th century, we convert utility differences in "life-time-income" equivalents. Specifically, let $T_{rt}^h(\Delta)$ be the increase of expected lifetime income, an individual with skill h in region r requires to increase the average utility of living in region r by Δ, i.e.

$$\mathscr{W}_{rt}^h\left(y \times T_{rt}^h(\Delta)\right) = \mathscr{W}_{rt}^h(y) + \Delta.$$

Letting $\Delta_t^{h,IQR}$ denote the interquartile range of $\mathscr{W}_{rt}^h$ at time t, our measure of spatial inequality for individuals of skill h is then given as the regional average of $T_{rt}^h\left(\Delta_t^{L,IQR}\right)$, i.e. $T_t^h\left(\Delta_t^{L,IQR}\right) = \frac{1}{R}\sum T_{rt}^h\left(\Delta_t^{h,IQR}\right)$. In Figure 7, we depict the evolution spatial welfare inequality, relative to the level in 1910. The solid lines refer to our baseline calibration. Spatial welfare inequality decreased substantially during the 20th century. For both high and low skill workers, the required increase in lifetime income to raise utility from the 25% quantile to the 75% quantile of the spatial distribution of welfare decreased by about 35% since 1910. In terms of levels, in 1910 the average region would have required an increase in average lifetime income by about 160% to increase utility by the interquartile range of regional welfare. In 2000, lifetime income would only have to be doubled.[51]

[50] For completeness we include a version of Table **??** using $\mathscr{W}_{rt}^h$ as a dependent variable in Section A.7 in the Appendix.

[51] The decline in inequality is not driven by changes in the distribution of regional amenities. In Section **??** in the Online Appendix we show that the dispersion in amenities is constant between 1910 and 2000 and that spatial welfare differences



Notes: The figure shows average increase in life-time income required to increase utility from the 25% quantile to the 75% quantile of the cross-sectional welfare distribution at time t. See Section C.1 in the Appendix for details.

Figure 7: The Evolution of Spatial Welfare Inequality

Spatial mobility is an important factor for this decline in spatial welfare inequality. Without labor mobility to prevent the regional distribution of income from fanning out during the structural transformation, spatial welfare would have declined much less. To see this, Figure 7 also depicts the evolution of spatial inequality in the model without labor mobility. Two features are noteworthy. First of all, the decline in spatial inequality is much lower compared to the model with spatial mobility. Because amenities evolve exogenously, the only difference between these two models is solely the endogenous distribution in spatial earnings. Hence, spatial arbitrage through labor mobility goes a long a way to compressing the spatial distribution of welfare.

Secondly, labor mobility is a particularly important adjustment mechanism for low-skilled workers. This discrepancy between workers of different skills is the consequence of the urban bias of the structural transformation. The secular shift away from agriculture hurts agricultural regions and, in particular, unskilled workers who have a comparative advantage in the agricultural sector. Moving out of rural locations as a response to this decrease in labor demand is an important mechanism to prevent regional wages, and in particular, agricultural wages, from falling. Spatial mobility was therefore a crucial adjustment mechanism for low-skilled workers to weather the first structural transformation away from agriculture. Without it, spatial welfare inequality would have been substantially higher.

6 Conclusion

The structural transformation, i.e. the systematic reallocation of employment out of the agricultural sector, is key feature of long-run economic growth. This sectoral bias of the growth process naturally

had declined even in the absence of amenities. The level of spatial welfare differences is, however, substantially lower.

affects the spatial allocation of economic activity. In particular, by shifting expenditure away from the agricultural sector, the structural transformation benefits urban, non-agricultural regions and hurts rural locations. In this paper, we use a novel theory of spatial structural change and detailed regional data to analyze the spatial nature of the structural transformation of the US between 1880 and 2000.

We first document empirically that the process of spatial reallocation towards non-agricultural regions explains essentially none of the aggregate decline in the agricultural employment share from 50% to essentially nearly over the last 120 years. In contrast, the entire decline is accounted for by within-region changes, whereby agricultural employment declines in each locality. While these patterns seem to contradict the large increase in urbanization over the same time-period, we show that this is not the case: like the change in agricultural employment, the increase in the proportion of urban dwellers is also very local in nature.

To understand why this is the case and wether this mode of adjustment has important aggregate implications, we construct a new quantitative model of the structural transformation that explicitly incorporates a spatial layer. The model combines basic features from an economic geography model featuring costly labor mobility with the canonical ingredients of neoclassical models of structural change, i.e. non-homothetic preferences, unbalanced technological progress and aggregate capital accumulation. Despite this richness, we show that the analysis remains very tractable and can be applied to a realistic geographical structure.

Our analysis yields three main results. First of all, we show that the insignificance of the spatial reallocation channel is mostly due to the evolution of regional productivities. Because local productivities are subject to shocks and regional absolute and comparative advantage are imperfectly correlated, the relationship between agricultural specialization and future earnings (and hence net population outflows) is only weakly negative - *despite* the urban bias of the structural transformation. Secondly, we show that spatial mobility was an important contributor of aggregate productivity growth in the US. Without labor mobility, aggregate income would have been 17% smaller. Finally, spatial welfare inequality declined substantially during the structural transformation and spatial mobility played an important role in that decline by keeping the regional income distribution compressed.

References

Acemoglu, Daron and Veronica Guerrieri, "Capital deepening and nonbalanced economic growth," *Journal of political Economy*, 2008, *116* (3), 467–498.

Adao, Rodrigo, Costas Arkolakis, and Federico Esposito, "Trade, Agglomeration Effects, and Labor Markets: Theory and Evidence," *Mimeo, Yale University*, 2017.

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf, "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Alder, Simon, Timo Boppart, and Andreas Muller, "A Theory of Structural Change that Can Fit the Data," 2017.

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

Alvarez-Cuadrado, Francisco and Ngo Van Long, "Capital-labor substitution, structural change and growth," *CIRANO-Scientific Publications 2011s-68*, 2011.

Baumol, William J, "Macroeconomics of unbalanced growth: the anatomy of urban crisis," *The American Economic Review*, 1967, pp. 415–426.

Boppart, Timo, "Structural Change and the Kaldor Facts in a Growth Model With Relative Price Effects and Non-Gorman Preferences," *Econometrica*, 2014, *82* (6), 2167–2196.

Bryan, Gharad and Melanie Morten, "Economic development and the spatial allocation of labor: Evidence from Indonesia," *Manuscript, London School of Economics and Stanford University*, 2015.

_ , Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh," *Econometrica*, 2014, *82* (5), 1671–1748.

Buera, Francisco J and Joseph P Kaboski, "Can traditional theories of structural change fit the data?," *Journal of the European Economic Association*, 2009, *7* (2-3), 469–477.

Caselli, Francesco and Wilbur John Coleman II, "The US structural transformation and regional convergence: A reinterpretation," *Journal of Political Economy*, 2001, *109* (3), 584–616.

Chenery, Hollis B, "Patterns of industrial growth," *The American Economic Review*, 1960, *50* (4), 624–654.

Comin, Diego, Danial Lashkari, and Martı Mestieri, "Structural Change with Long-Run Income and Price Effects," 2015. Working Paper.

Desmet, Klaus and Esteban Rossi-Hansberg, "Spatial Development," *American Economic Review*, 2014, *104* (4), 1211–43.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A Market Access Approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Eckert, Fabian, Andres Gvirtz, and Michael Peters, "Constructing a population panel for US Commuting Zones 1790-2000," *Mimeo, Yale University*, 2018.

Fajgelbaum, Pablo and Stephen J Redding, "External integration, structural transformation and economic development: Evidence from argentina 1870-1914," 2014. Working Paper.

Fajgelbaum, Pablo D, Eduardo Morales, Juan Carlos Suarez Serrato, and Owen M Zidar, "State taxes and spatial misallocation," Technical Report, National Bureau of Economic Research 2015.

Gollin, Douglas, David Lagakos, and Michael E Waugh, "The Agricultural Productivity Gap," Technical Report, Quarterly Journal of Economics 2013.

_ , Martina Kirchberger, and David Lagakos, "In Search of a Spatial Equilibrium in the Developing World," 2017.

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

Herrendorf, Berthold and Todd Schoellman, "Why is measured productivity so low in agriculture?," *Review of Economic Dynamics*, 2015, *18* (4), 1003–1022.

_ , Christopher Herrington, and Akos Valentinyi, "Sectoral technology and structural transformation," *American Economic Journal: Macroeconomics*, 2015, *7* (4), 104–133.

_ , Richard Rogerson, and Akos Valentinyi, "Two perspectives on preferences and structural transformation," *The American Economic Review*, 2013, *103* (7), 2752–2789.

_ , _ , and Ákos Valentinyi, "Growth and Structural Transformation," *Handbook of Economic Growth*, 2014, *2*, 855–941.

Hobijn, Bart, Todd Schoellman, and Alberto J. Vindas Q., "Structural Transformation:A Race between Demographics and Technology," 2017. Working Paper.

Hsieh, Chang-Tai and Enrico Moretti, "Why do cities matter? Local growth and aggregate growth," Technical Report, National Bureau of Economic Research 2015.

Kongsamut, Piyabha, Sergio Rebelo, and Danyang Xie, "Beyond balanced growth," *The Review of Economic Studies*, 2001, *68* (4), 869–882.

Kuznets, Simon, "Quantitative aspects of the economic growth of nations: II. industrial distribution of national product and labor force," *Economic Development and Cultural Change*, 1957, pp. 1–111.

Lagakos, David and Michael E Waugh, "Selection, Agriculture and Cross-Country Productivity Differences," *Agriculture and Cross-Country Productivity Differences (November 1, 2011)*, 2011.

_ , Mushfiq Mobarak, and Michael E Waugh, "The Macroeconomics of Rural-Urban Migration," 2015.

Lindbeck, Assar, ed., *Nobel Lectures, Economics 1969-1980*, World Scientific Publishing Co., Singapore, 1992.

Michaels, Guy, Ferdinand Rauch, and Stephen J Redding, "Urbanization and structural transformation," *The Quarterly Journal of Economics*, 2012, *127* (2), 535–586.

Molloy, Raven, Christopher L Smith, and Abigail Wozniak, "Internal migration in the United States," *The Journal of Economic Perspectives*, 2011, *25* (3), 173–196.

Nagy, Dávid Krisztián, "City location and economic development," 2016. Working Paper.

Ngai, L Rachel and Christopher A Pissarides, "Structural change in a multisector model of growth," *The American Economic Review*, 2007, *97* (1), 429–443.

Porzio, Tommaso and Gabriella V. Santangelo, "Structural Change and the Supply of Agricultural Workers," 2017. Working Paper.

Redding, Stephen and Esteban Rossi-Hansberg, "Quantitative Spatial Economics," 2016. Working Paper.

Redding, Stephen J and Daniel M Sturm, "The costs of remoteness: Evidence from German division and reunification," *LSE STICERD Research Paper No. PEPP11*, 2005.

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander, "Integrated Public Use Microdata Series: Version 6.0 [dataset]. Minneapolis: University of Minnesota, 2015," 2015.

Tolbert, Charles M and Molly Sizer, "US commuting zones and labor market areas: A 1990 update," 1996.

Tombe, Trevor, Xiaodong Zhu et al., "Trade, migration and productivity: A quantitative analysis of china," *Manuscript, University of Toronto*, 2015.

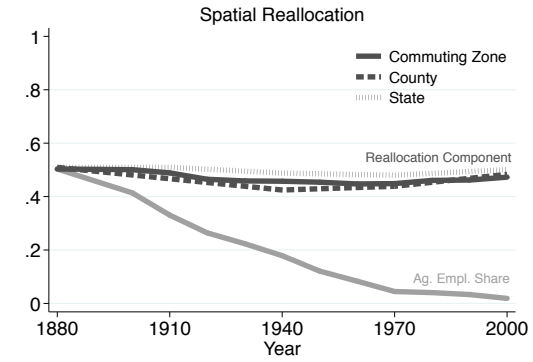
Young, Alwyn, "Inequality, the urban-rural gap, and migration," *The Quarterly Journal of Economics*, 2013, *128* (4), 1727–1785.

APPENDIX

A Spatial Reallocation and Structural Change: Additional Results

In Section 2 we showed that the spatial reallocation of individuals cannot account for the observed decline of the agricultural employment share in the US. In this section we provide additional details for this empirical result.

First we consider other definitions of regions besides the commuting zone definition used throughout the main body of the paper. Figure 8 replicates Figure 1 on the county and state level. We use the same underlying data and simply aggregate it differently. Figure 8 reveals that the reallocation component is quantitatively unimportant regardless which of the three region definitions is chosen. Note that the reallocation component is zero by construction if we consider the entire US as one region. The ordering of the lines is indicative of that: for most years the reallocation component is largest for counties (the smallest aggregation we consider) and smallest for state (the largest aggregation we consider). All in all Figure 8 reinforces a view of reallocation and urbanization as highly local phenomena, predominantly operational on a intra-county level.



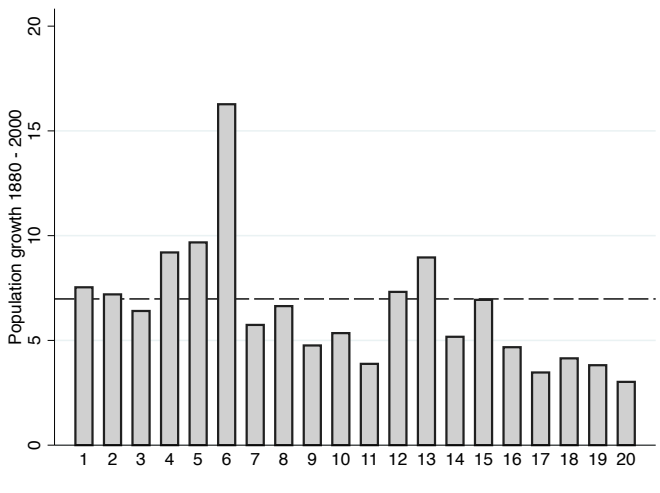
Notes: We show the aggregate agricultural employment share (light grey line) and the predicted agricultural share holding regional agricultural shares at their 1880 level, i.e. $\sum_r s_{rA1880} l_{rt}$, where s_{rAt} and l_{rt} are the agricultural employment share and the population share of region r at time t(dark grey line). We do this for three different definitions of r: state, commuting zone or county. The underlying data is the same as described in the data appendix and simply aggregated differently for each of the three scenarios.

Figure 8: Spatial Reallocation Across States, Commuting Zones and Counties

Next we provide additional evidence for this (non-)relationship between past agricultural specialization and subsequent population growth. In particular, we consider a regression of the form

$$g_{rL}^{1880-2000} = \alpha + \beta s_{rA}^{1880} + u_r,$$

where $g_{rL}^{1880-2000}$ denotes regional population growth between 1880 and 2000 and s_{rA}^{1880} denotes the agricultural employment share in 1880. The results are contained in Table 5. Columns 1 to 3 show that there is no significant relationship between agricultural specialization in 1880 and population growth between 1880 and 2000. Columns 2 and 3 weigh each regression by their initial population in 1880. In column



Notes: The figure shows the average rate of population growth between 1880 and 2000 for 20 quantiles of the agricultural employment share in 1880.

Figure 9: Agricultural Specialization in 1880 and Population growth 1880-2000

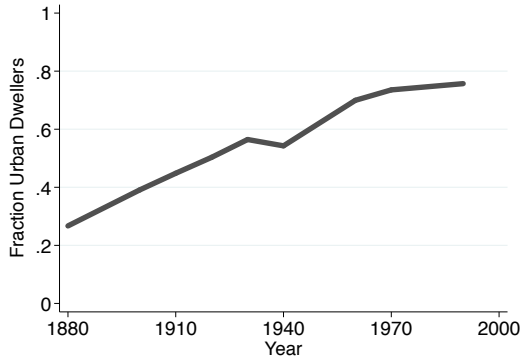
4 we include a set of twenty fixed effects of the initial agricultural share quantiles. While these fixed effects are jointly statistically significant, their explanatory power is still very small. Figure 9 shows this relationship graphically. More specifically, we report average population growth between 1880 and 2000 for twenty quantiles of the agricultural employment share.While population growth tends to be slightly smaller in regions with a high agricultural employment share in 1880, the relationship is not particularly strong and certainly not monotone.

	Dep. variable: Population growth 1880 - 2000			
Agricultural share 1880	-26.241	-3.702*		
	(40.829)	(1.917)		
log Agricultural share 1880			-0.548	
			(0.658)	
Ag quantile FE	No	No	No	Yes
Weights	No	Yes	Yes	Yes
Observations	717	717	717	717
R^2	0.000	0.002	0.000	0.014

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. Column 4 contains a whole set of 20 fixed effects for the different quantiles of agricultural employment shares in 1880.

Table 5: Agricultural Specialization in 1880 and Population growth 1880-2000

The absence of the spatial reallocation channel seems to be inconsistent with sharp increase in the rate of urbanization shown in Figure 10. We now show that this is not the case. In particular, we show that both the secular decrease in agricultural employment and the increase in urbanization are very local phenomena. In the left panel of Figure 11, we show the increase in urbanization since 1880. The share



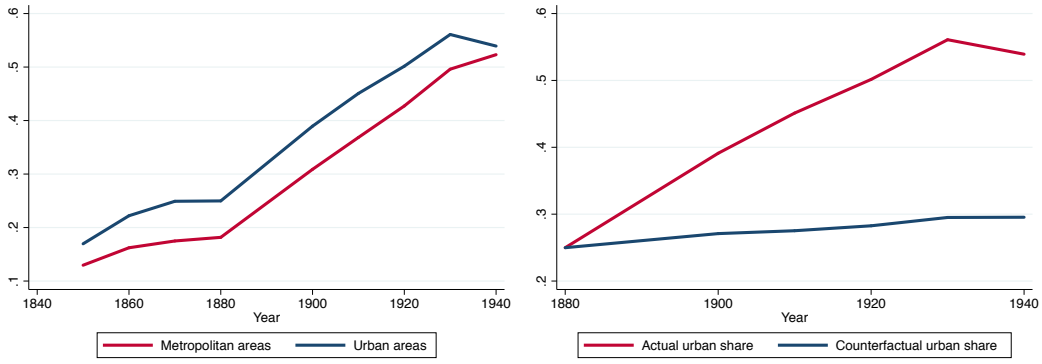
Notes: The graph is constructed from the decennial US micro census available on IPUMS. The line graphs the fraction of Americans who live in an urban environment rather than a rural one over time. An urban environment is defined as a locality with more than 2500 inhabitants.

Figure 10: Fraction of US population living in urban environment

of people living in urban areas (i.e. cities with more than 2500 inhabitants) increases from just shy of 20% in 1850 to more than 50% of the population in 1940. A different measure, the share of people living in metropolitan areas, shows a similar pattern. In the right panel we again decompose this time-series evolution into a within and across commuting zone component. In particular, we calculate the increase in the rate of urbanization due to spatial reallocation as

$$u_t^{CF} = \sum_r u_{r1880} \frac{L_{rt}}{\sum_r L_{rt}},$$

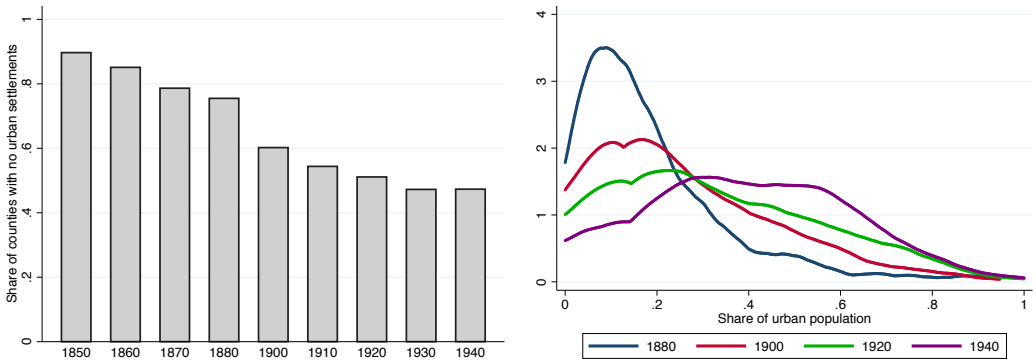
where $u_{r,1880}$ is the urbanization rate in commuting zone r in 1880. If the increase in urbanization stemmed from individuals migrating into high-urbanization commuting zone, this counterfactual urbanization rate would be close to the actual time series.



Notes: In the left panel we show the aggregate share of people living in urban areas and metropolitan areas between 1850 and 1940. In the right panel, we again show the share of people living in urban areas and counterfactual urban share calculated according to $u_t^{CF} = \sum_r u_{r1880} \times \frac{L_{rt}}{\sum_r L_{rt}}$, where $u_{r,1880}$ is the urban share in 1880.

Figure 11: Urbanization within and across commuting zones

Figure 11 shows that this is not the case: as for the agricultural employment share, the cross-commuting zone population flows explain a minor share of actual increase observed in the data. Figure 12 provides an alternative way to see this within-commuting zone pattern of urbanization. In the left panel we display the "extensive" margin of urbanization, i.e. the share of counties, which have no urban centers. Expectedly, this number is declining. Similarly, the right panel shows the distribution of the share of the urban population across commuting zones conditional on this share being positive. As for the patterns of agricultural employment depicted in Figure 1 these densities shift to the right, indicating that urbanization takes place within all counties in the US.



Notes: In the left panel we depict the share of counties without any urban areas. In the right panel we show the cross-sectional distribution of the share of the urban population across counties.

Figure 12: Urbanization within commuting zones

Finally we ask whether the almost absence of the reallocation component of structural change is a more general feature of structural transformations or particular to the United States. Indeed a look at other countries around the world suggests that the highly localized nature of reallocation and urbanization is a feature of the structural transformation of countries around the world. To see this consider Table 6 which uses data from IPUMS (see Ruggles et al. (2015)) to compute the reallocation component of the decline in the aggregate agricultural share for seven additional countries and finds it to be generally very small.

Country	USA	Argentina	China	India	Mali	Mexico	Spain	Venezuela
Time Period	1880-2000	1970-2001	1982-2000	1987-2009	1987-2009	1970-2010	1991-2011	1981-2001
Number of Regions	712	312	198	413	47	2316	52	157
ΔAg. Empl. Share	-0.48	-.076	-.10	-.09	-.12	-.29	-.12	-.02
Δ implied by reallocation	-0.03	0	0	-0.01	-0.05	-0.03	-0.01	0.01

Notes: These regions are in general larger than counties in the United States. All numbers rounded to two decimal places. The US numbers are based only on continental commuting zones. Source: for all countries except the US we use IPUMS international as a data source. Sources for the US data are discussed in detail below.

Table 6: Reallocation Component in Structural Change Across the World

The regions available for these countries in IPUMS are generally larger than commuting zones in the United States, but smaller than US states. Given Figure 8 for the United States, we take this quasi-absence

of the reallocation component at the sub-state resolution as indicative of the local nature of urbanization and reallocation being an important feature of the process of structural change more generally.

B Theory: Proofs and Derivations

B.1 Proof of Lemma 1

Suppose that the indirect utility function falls in the PIGL class, i.e. $V(e,P) = \frac{1}{\eta}\left(\frac{e}{B(p)}\right)^{\eta} + C(P) - \frac{1}{\eta}$. The maximization problem is

$$U_r^i = \max_{[e_t, e_{t+1}, s]} \left\{ V(e_t, P_t) + \beta V(e_{t+1}, P_{t+1}) \right\},$$

subject to

$$\begin{aligned} e_t + s_t p_{I,t} &= y_{rt}^i \\ e_{t+1} &= (1 + r_{t+1}) s_t p_{I,t+1}. \end{aligned}$$

Substituting for e_{t+1} yields

$$U_r^i = \max_{e_t} \left\{ V(e_t, P_t) + \beta V\left((1 + r_{t+1}) \left(y_{rt}^i - e_t \right) \frac{p_{I,t+1}}{p_{I,t}}, P_{t+1} \right) \right\}.$$

The optimal allocation of spending is determined from the Euler equation

$$\frac{\partial V(e_t, P_t)}{\partial e} = \beta (1 + r_{t+1}) \frac{p_{I,t+1}}{p_{I,t}} \frac{\partial V(e_{t+1}, P_{t+1})}{\partial e}.$$

From above (24) we get that this equation is

$$\begin{aligned} e_t^{\eta-1} B(p_t)^{-\eta} &= \beta (1 + r_{t+1}) \frac{p_{I,t+1}}{p_{I,t}} e_{t+1}^{\eta-1} B(p_{t+1})^{-\eta} \\ &= \beta \left((1 + r_{t+1}) \frac{p_{I,t+1}}{p_{I,t}} \right)^{\eta} \left(\left(y_{rt}^i - e_t \right) \right)^{\eta-1} B(p_{t+1})^{-\eta} \end{aligned}$$

This yields

$$\frac{y_{rt}^i - e_t}{e_t} = \left(\left(\frac{1}{1 + r_{t+1}} \frac{B(p_{t+1}) / B(p_t)}{p_{I,t+1} / p_{I,t}} \right)^{\eta} \frac{1}{\beta} \right)^{\frac{1}{\eta-1}},$$

so that

$$\begin{aligned} e_t &= \frac{1}{1+\left(\phi_{t,t+1}^{\eta}\frac{1}{\beta}\right)^{\frac{1}{\eta-1}}}y_{rt}^{i} \\ e_{t+1} &= \frac{(1+r_{t+1})\,p_{I,t+1}}{p_{I,t}}\frac{\left(\phi_{t,t+1}^{\eta}\frac{1}{\beta}\right)^{\frac{1}{\eta-1}}}{1+\left(\phi_{t,t+1}^{\eta}\frac{1}{\beta}\right)^{\frac{1}{\eta-1}}}y_{rt}^{i} \end{aligned}$$

where

$$\phi_{t,t+1} \equiv \frac{1}{1+r_{t+1}}\frac{B(p_{t+1})/B(p_t)}{p_{I,t+1}/p_{I,t}}.$$

Hence, (6) implies that

$$\begin{aligned} U_{jr} &= V_{jr}(e_t,P_t)+\beta V_{jr}(e_{t+1},P_{t+1}) \\ &= A_{jr}\frac{1}{\varepsilon}\left(\frac{e_t}{B(p_t)}\right)^{\varepsilon}+C(P_t)-\frac{1}{\varepsilon}+\beta\left(A_{jr}\frac{1}{\varepsilon}\left(\frac{e_{t+1}}{B(p_{t+1})}\right)^{\varepsilon}+C(P_{t+1})-\frac{1}{\varepsilon}\right) \\ &= A_{jr}\frac{1}{\varepsilon}w_{rt}^{\varepsilon}\left[B(p_t)^{-\varepsilon}\left(1+\left(\frac{1}{\beta}\phi_{t,t+1}^{\varepsilon}\right)^{\frac{1}{\varepsilon-1}}\right)^{1-\varepsilon}\right]+C(P_t)+\beta C(P_{t+1})-\frac{1+\beta}{\varepsilon}. \end{aligned}$$

This can be written as

$$U_{jr} = A_{jr}w_{rt}^{\varepsilon}\Phi_{t,t+1}+\Lambda_{t,t+1},$$

where

$$\begin{aligned} \Phi_{t,t+1} &= \frac{1}{\varepsilon}B(p_t)^{-\varepsilon}\left(1+\left(\frac{1}{\beta}\phi_{t,t+1}^{\varepsilon}\right)^{\frac{1}{\varepsilon-1}}\right)^{1-\varepsilon} \\ \Lambda_{t,t+1} &= C(P_t)+\beta C(P_{t+1})-\frac{1+\beta}{\varepsilon}. \end{aligned}$$

For our specification we have that $\varepsilon=\eta$ and $B(p_t)=p_{A,t}^{\phi}p_{NA,t}^{1-\phi}=1$. Hence,

$$\phi_{t,t+1} \equiv \frac{B(p_{t+1})}{B(p_t)(1+r_{t+1})} = \frac{1}{1+r_{t+1}} = \phi_{t+1}$$

and

$$\Phi_{t,t+1} = \frac{1}{\eta}\left(1+\left(\frac{1}{\beta}\left(\frac{1}{1+r_{t+1}}\right)^{\eta}\right)^{\frac{1}{\eta-1}}\right)^{1-\eta} = \frac{1}{\eta}\left(1+\beta^{\frac{1}{1-\eta}}(1+r_{t+1})^{\frac{\eta}{1-\eta}}\right)^{1-\eta}.$$

Note also that

$$e_t = \frac{1}{1+\left(\phi_{t,t+1}^{\eta}\frac{1}{\beta}\right)^{\frac{1}{\eta-1}}}w_t = \frac{1}{1+\beta^{\frac{1}{1-\eta}}(1+r_{t+1})^{\frac{\eta}{1-\eta}}}w_t.$$

This proves the results for Lemma 1.

B.2 Consumption expenditure shares

We can derive the expenditure shares from the indirect utility function $V_{jr}(e,p)$ in (7) from Roy's identity. The indirect utility function is defined by

$$V(p,e(p,u)) = u.$$

Hence,

$$\frac{\partial V(p,e(p,u))}{\partial p_j} + \frac{\partial V(p,e(p,u))}{\partial e}\frac{\partial e(p,u)}{\partial p_j} = 0.$$

The expenditure function is given by

$$e(p,u) = \min_x \sum p_j x \text{ s.t. } u(x) \geq u.$$

Hence,

$$\frac{\partial e(p,u)}{\partial p_j} = x_j(p,u),$$

where $x_j(p,u)$ is the Hicksian demand function. We also have that the Hicksian and the Marshallian demands are linked via

$$x^H(p,u) = x^M(p,e(p,u)).$$

Hence,

$$\frac{\partial V(p,e(p,u))}{\partial p_j} + \frac{\partial V(p,e(p,u))}{\partial e} x_j^M(p,e(p,u)) = 0$$

Rearranging terms yields

$$x_j^M(p,e(p,u)) = -\frac{\frac{\partial V(p,e(p,u))}{\partial p_j}}{\frac{\partial V(p,e(p,u))}{\partial e}}.$$

The expenditure share on good j is therefore given by

$$\vartheta_j(e,p) = x_j^M(p,e)\frac{p_j}{e} = -\frac{\frac{\partial V(p,e(p,u))}{\partial p_j} p_j}{\frac{\partial V(p,e(p,u))}{\partial e} e}. \tag{23}$$

We consider an indirect utility function from the PIGL class

$$V(e,P) = \frac{1}{\varepsilon}\left(\frac{e}{B(p)}\right)^{\varepsilon} + C(P) - \frac{1}{\varepsilon}.$$

We get that

$$\begin{aligned}\frac{\partial V_{rj}}{\partial e} &= e^{\varepsilon-1}B(p)^{-\varepsilon} \\ \frac{\partial V_{rj}}{\partial p_j} &= -e^{\varepsilon}B(p)^{-\varepsilon-1}\frac{\partial B(p)}{\partial p_j}+\frac{\partial C(p)}{\partial p_j}.\end{aligned} \tag{24}$$

(23) therefore implies that

$$\begin{aligned}\vartheta_j(e,p) &= \frac{e^{\varepsilon}B(p)^{-\varepsilon-1}\frac{\partial B(p)}{\partial p_j}p_j-\frac{\partial C(p)}{\partial p_j}p_j}{e^{\varepsilon-1}B(p)^{-\varepsilon}e} \\ &= \frac{\partial B(p)\,p_j}{\partial p_j B(p)}-\frac{\partial C(p)\,p_j}{\partial p_j C(p)}C(p)\left(\frac{e}{B(p)}\right)^{-\varepsilon} \\ &= \eta_j^B-\eta_j^C\left(\frac{e}{B(p)}\right)^{-\varepsilon}C(p),\end{aligned}$$

where η_j^B and η_j^C are the price elasticities of the B and C function respectively.

The specification we consider is

$$V(e,P)=\frac{1}{\eta}\left(\frac{e}{p_A^{\phi}p_M^{1-\phi}}\right)^{\eta}-\frac{\nu}{\gamma}\left(\frac{p_A}{p_M}\right)^{\gamma}-\frac{1}{\eta}+\frac{\nu}{\gamma}.$$

Hence, we have that

$$\begin{aligned}\frac{\partial V}{\partial e} &= \left(\frac{e}{p_A^{\phi}p_M^{1-\phi}}\right)^{\eta}\frac{1}{e} \\ \frac{\partial V}{\partial p_A} &= -\phi\left(\frac{e}{p_A^{\phi}p_M^{1-\phi}}\right)^{\eta}\frac{1}{p_A}-\nu\left(\frac{p_A}{p_M}\right)^{\gamma}\frac{1}{p_A}.\end{aligned}$$

Hence,

$$x_A(e,p) = \phi\frac{e}{p_A}+\nu\left(\frac{p_A}{p_M}\right)^{\gamma}\frac{1}{p_A}\left(p_A^{\phi}p_M^{1-\phi}\right)^{\eta}e^{1-\eta}. \tag{25}$$

The expenditure share is

$$\vartheta_A(e,p) = \frac{x_A(e,p)\,p_A}{e}=\phi+\nu\left(\frac{p_A}{p_M}\right)^{\gamma}\left(\frac{e}{p_A^{\phi}p_M^{1-\phi}}\right)^{-\eta}.$$

B.3 Earning, Expected Earnings and Aggregate Demand

Consider individual i in region r. Given her optimal occupational choice, the earnings of individual i are given by

$$y^i = \max_s \left\{ w_{rs} z_s^i \right\}.$$

We assumed that individual productivities are Frechet Distributed, i.e.

$$F_s^h(z) = e^{-\Psi_s^h z^{-\zeta}}, \tag{26}$$

where Ψ_s^h parametrizes the average level of productivity of individuals of skill h in region r in sector s and ζ the dispersion of skills. Hence, the distribution of sectoral earning $y_{sr}^i \equiv w_{s,r} z_s^i$ is also Frechet and given by

$$F_{y_{rs}}(y) = P\left(z \leq \frac{y}{w_{rs}} \right) = e^{-\Psi_s^h w_{rs}^{\zeta} y^{-\zeta}}.$$

Using standard arguments about the max stability of the Frechet, the distribution of total earnings y^i is also Frechet and given by

$$F_{y_r}^h(y) = e^{-\left(\Theta_r^h\right)^{\zeta} y^{-\zeta}} = e^{-\left(\frac{y}{\Theta_r^h}\right)^{-\zeta}} \tag{27}$$

where

$$\Theta_r^h = \left(\sum_s \Psi_s^h w_{rs}^{\zeta} \right)^{1/\zeta} = \left(\Psi_A^h w_{rA}^{\zeta} + \Psi_{NA}^h w_{rNA}^{\zeta} \right)^{1/\zeta}.$$

Hence, average earnings of individual i with skill type h in region r are given by

$$E\left[y_{r,h}^i \right] = \Gamma\left(1 - \frac{1}{\zeta} \right) \Theta_r^h.$$

From (27) we can derive the distribution of $y^{1-\eta}$. As $\eta < 1$, we have that

$$\begin{aligned} F_{y^{1-\eta}}(q) = P\left(y^{1-\eta} \leq q \right) = P\left(y \leq q^{1/(1-\eta)} \right) &= e^{-\Theta^{\zeta}\left(q^{\frac{1}{1-\eta}} \right)^{-\zeta}} \\ &= e^{-\left(\frac{q}{\Theta^{1-\eta}} \right)^{-\frac{\zeta}{1-\eta}}}. \end{aligned}$$

Hence, $y^{1-\eta}$ is still Frechet. Therefore

$$\int_i y_i^{1-\eta} di = L_r^h E\left[y_i^{1-\eta} \right] = L_r^h \Gamma\left(1 - \frac{1-\eta}{\zeta} \right) \times \left(\Theta_r^h \right) = L_r^h \Gamma\left(1 + \frac{\eta - 1}{\zeta} \right) \left(\Theta_r^h \right)^{1-\eta}.$$

B.4 Balanced Growth Path Relationships

Consider a dynamic allocation where GDP grows at a constant rate and the capital output ratio is constant. Static optimality requires that $R_t = \frac{\alpha PY_t}{K_t}$. Hence, a constant capital output ratio directly implies that the return to capital R_t has to be constant. Hence, the real interest on saving is also constant and given by $r = R - \delta$. This also implies that the consumption rate in (13) is constant and given by

$$\psi = \left(1 + \beta^{\frac{1}{1-\eta}} (1+r)^{\frac{\eta}{1-\eta}}\right)^{-1}. \tag{28}$$

To solve for the interest rate, note that (20) implies that

$$\frac{K_{t+1}}{K_t} = \frac{(1-\psi)(1-\alpha) PY_t}{\alpha PY_t / R} = (1-\psi)\frac{(1-\alpha)}{\alpha}(r+\delta) = 1+g,$$

where g is the growth rate of the economy. Hence,

$$1+g = (1-\psi)\frac{(1-\alpha)}{\alpha}(r+\delta) = \left(\frac{\beta^{\frac{1}{1-\eta}} (1+r)^{\frac{\eta}{1-\eta}}}{1+\beta^{\frac{1}{1-\eta}} (1+r)^{\frac{\eta}{1-\eta}}}\right)\frac{(1-\alpha)}{\alpha}(r+\delta). \tag{29}$$

This equation determines r as a function of parameters. Along the BGP the consumption and investment rate is equal to

$$\begin{aligned} PC_t &= \psi(1-\alpha) PY_t + \alpha PY_t + (1-\delta)\frac{\alpha}{R} PY_t = \left[\psi(1-\alpha) + \alpha + (1-\delta)\frac{\alpha}{R}\right] PY_t \\ PI_t &= (1-\psi)(1-\alpha) PY_t - (1-\delta)\frac{\alpha}{R} PY_t = \left[(1-\psi)(1-\alpha) - (1-\delta)\frac{\alpha}{R}\right] PY_t \end{aligned}$$

Using (29) yields

$$\frac{PI_t}{PY_t} = (g+\delta)\frac{\alpha}{R} \quad \text{and} \quad \frac{PC_t}{PY_t} = 1 - (g+\delta)\frac{\alpha}{R}.$$

From (29) we also get that

$$\psi = 1 - \frac{\alpha}{1-\alpha}\frac{1+g}{R}. \tag{30}$$

B.5 Regularity conditions for PIGL preferences

In our model, expenditure share on the two goods are given by

$$\begin{aligned} \vartheta_A(e,p) &= \phi + \nu\left(\frac{p_A}{p_M}\right)^{\gamma} e^{-\eta} \\ \vartheta_{NA}(e,p) &= 1 - \phi - \nu\left(\frac{p_A}{p_M}\right)^{\gamma} e^{-\eta}. \end{aligned}$$

For these expenditure shares to be positive, we need that

$$\vartheta_A(e,p) \geq 0 \Rightarrow e^{\eta} \geq -\frac{\nu}{\phi}\left(\frac{p_A}{p_M}\right)^{\gamma}, \tag{31}$$

and

$$\vartheta_{NA}(e,p) \geq 0 \Rightarrow e^{\eta} \geq \frac{\nu}{1-\phi}\left(\frac{p_A}{p_M}\right)^{\gamma}.$$

Note first that (31) is trivially satisfied as long as $\nu > 0$. Also note that satisfying both of these restrictions automatically implies that $\vartheta_s(e,p) \leq 1$. In addition, as we show in Section A.2 in the Online Appendix, for the Slutsky matrix to be negative semi definite, we need that

$$\nu(1-\eta)\left(\frac{p_A}{p_M}\right)^{\gamma} - \frac{(1-\phi)\phi}{\nu}\left(\frac{p_A}{p_M}\right)^{-\gamma} e^{2\eta} \leq (1-2\phi-\gamma)e^{\eta}.$$

Hence, for our preferences to be well-defined, we require that

$$e^{\eta} \geq \frac{\nu}{1-\phi}\left(\frac{p_A}{p_M}\right)^{\gamma} \tag{32}$$

$$(1-2\phi-\gamma)e^{\eta} + \frac{(1-\phi)\phi}{\nu}\left(\frac{p_A}{p_M}\right)^{-\gamma} e^{2\eta} \geq \nu(1-\eta)\left(\frac{p_A}{p_M}\right)^{\gamma}. \tag{33}$$

Lemma 3. *A sufficient condition for (32) to be satisfied is that (33) holds and that*

$$\gamma > (1-\phi)\eta. \tag{34}$$

Proof. Equation (33) can be written as

$$\left(\frac{e^{\eta}}{\frac{\nu}{1-\phi}\left(\frac{p_A}{p_M}\right)^{\gamma}} - 1\right) + \phi\frac{e^{\eta}}{\frac{\nu}{1-\phi}\left(\frac{p_A}{p_M}\right)^{\gamma}}\frac{e^{\eta}}{\frac{\nu}{1-\phi}\left(\frac{p_A}{p_M}\right)^{\gamma}} \geq [(1-\phi)(1-\eta)-1] + (2\phi+\gamma)\frac{e^{\eta}}{\frac{\nu}{1-\phi}\left(\frac{p_A}{p_M}\right)^{\gamma}}.$$

Letting $x = \frac{e^{\eta}}{\frac{\nu}{1-\phi}\left(\frac{p_A}{p_M}\right)^{\gamma}}$, this yields

$$(x-1) + (\phi x - (2\phi+\gamma))x \geq -(1-(1-\phi)(1-\eta)).$$

Now let $h(x) = (x-1) + (\phi x - (2\phi+\gamma))x$. For (33) to be satisfied we need that

$$h(x) \geq -(1-(1-\phi)(1-\eta)).$$

Note that h is strictly concave with a minimum at

$$h'(x^*) = 1 + \phi x^* - (2\phi+\gamma) + \phi x^* = 0.$$

Hence,

$$x^* = 1 - \frac{1-\gamma}{2\phi} < 1.$$

Also note that

$$h(0) = -1 < -(1-(1-\phi)(1-\eta)).$$

Hence, for (33) to be satisfied, it has to be the case that $x > x^* = 1 - \frac{1-\gamma}{2\phi}$ as $h(x^*) < h(0)$. Hence, condition (33) implies (32) if

$$h(1) = \phi - 2\phi - \gamma < -(1-(1-\phi)(1-\eta)).$$

Rearranging terms yields $(1-\phi)\eta < \gamma$, which is (34). □

Hence, the preferences are well defined as long (33) is satisfied and (34) holds. Because the RHS of (33) is increasing in e in the relevant range, i.e. as long as (33) is satisfied, this implies that the preferences are well defined as long as e is high enough.

Now note that $e_{it} = \psi_{t+1} \times y_{it}$, where y_{it} denotes total earnings of individual i. From (27) we know that

$$P(e_{it} \leq \kappa) = P\left(y_{it} \leq \frac{\kappa}{\psi_{t+1}}\right) = e^{-\left(\Theta_r^h \psi_{t+1}\right)^{\zeta} \kappa^{-\zeta}},$$

where

$$\left(\Theta_{rt}^h \psi_{t+1}\right)^{\zeta} = \left(\Psi_A^h w_{rAt}^{\zeta} + \Psi_{NA}^h w_{rNAt}^{\zeta}\right) \psi_{t+1}^{\zeta} > \left(\Psi_A^h w_{rAt}^{\zeta} + \Psi_{NA}^h w_{rNAt}^{\zeta}\right) \left(\frac{1}{\eta}\left(1+\beta^{\frac{1}{1-\eta}}\right)^{1-\eta}\right)^{\zeta}.$$

Hence, as long as aggregate productivity in 1880 (and hence w_{rs1880}) is high enough, we can make $P(e_{it} \leq \kappa)$ arbitrarily small.

C Additional Empirical Results

C.1 Spatial Welfare Inequality

In Figure 7 in the main text we depict the evolution of spatial inequality. We construct this figure in the following way. From (48) we have that expected utility in region r is given by

$$\mathscr{W}_{rt}^h\left(\Theta_r^h\right) = \frac{\Gamma_{\eta/\zeta}}{\eta} \psi (r_{t+1})^{\eta-1} \left(\Theta_r^h\right)^{\eta} + \Lambda_{t,t+1} + A_{rt}. \tag{35}$$

	Expected value of moving relative to fixed costs Δ^h_{jt}				
	10%	25%	50%	75%	90%
Low Skilled	-0.2543	-0.0937	0.0571	0.1900	0.2871
High Skiled	-0.2464	-0.0654	0.1065	0.2543	0.3619

Notes: The table reports different quantiles of the distribution of Δ^h_{jt} calculated as (36).

Table 7: The Economic Magnitude of Moving Costs

Let $T^h_{rt}(\Delta)$ be the increase in average income Θ^h_r required to increase utility by Δ, i.e. $\mathscr{W}^h_{rt}\left(\Theta^h_r T^h_{rt}(\Delta)\right) = \mathscr{W}^h_{rt}\left(\Theta^h_r\right) + \Delta$. (35) implies that

$$T^h_{rt}(\Delta) = \left(1 + \frac{\Delta}{\frac{\Gamma_{\eta/\zeta}}{\eta}\psi\left(r_{t+1}\right)^{\eta-1}\left(\Theta^h_r\right)^{\eta}}\right)^{1/\eta}.$$

Let $\Delta^{h,IQR}_t$ be the interquartile range in regional welfare $\mathscr{W}^h_{rt}$ at time t, i.e. $\Delta^{h,IQR}_t = \mathscr{W}^{h,75}_t - \mathscr{W}^{h,25}_t$, where $\mathscr{W}^{h,x}_t$ is the x-quantile of the distribution of $\mathscr{W}^h_{rt}$. Figure 7 shows the time series evolution of $\overline{T}^h_t\left(\Delta^{h,IQR}_t\right) = \frac{1}{R}\sum_r T^h_{rt}(\Delta)$ for both high and low skilled individuals.

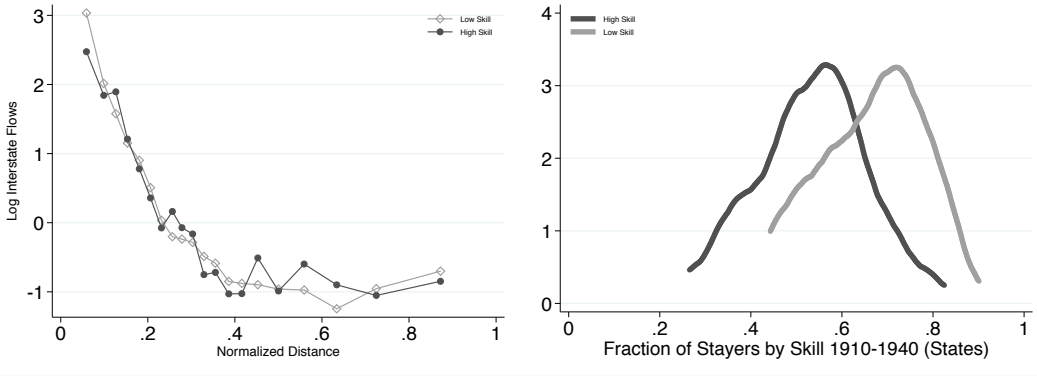
C.2 Additional Properties of the Calibrated Model

Moving costs To quantify the economic magnitude of our estimated fixed costs, we calculate

$$\Delta^h_{jt} \equiv \frac{\mathscr{U}^{h,Mov}_{jt} - \mathscr{U}^{h,Stay}_{jt}}{\tau}, \tag{36}$$

i.e. the average increase in utility by moving *relative* to the fixed cost of moving as a norm for the economic magnitude of fixed costs. Because utility is not equalized, these gains differ by commuting zone. In Table 7 we report some statistics of the distribution of Δ^h_{jt}. In the top row we report these statistics for low skilled individuals, in the bottom row for high skilled individuals. Table 7 shows that for the median commuting zone the expected value of moving is slightly positive and amounts to roughly 5% (10% for high skilled individuals) of the estimated fixed cost of moving.

Migration by skill type In the model low and high skill workers are subject to the same distance costs, which yields an elasticity of moving flows to distance that is very similar for both groups. As the left panel of the below figure shows this is in line with the data. In the model high skill types are slightly more likely to move but not a lot. In the data low skilled people are substantially less likely to leave their birth state. The model can match if we allowed τ to differ by skill type, since as outlined in the calibration section, the total number of stayers is monotone in τ. We chose to abstract from this for simplicity in the main part of the paper.



Notes: To construct the left panel, we run a gravity equation of the form where origin and destination fixed effects are skill specific:

$$\log \frac{\rho_{jr}}{1-\rho_{jj}} = \alpha_j^h + \alpha_r^h + \hat{u}_{jr}^h$$

The red diamonds line plots the resulting $\hat{u}_{jr}^H$ by distance percentile, the blue line does the same for low skill types. The right panel we plot the distribution of the share of people staying in their home state between 1910 and 1940 for low and high skilled people.

Figure 13: Lifetime interstate Migration by Skill in 1940 in the Data

Regional Fundamentals In this section, we describe additional details for the estimated spatial productivities $\{Q_{rAt}, Q_{rNAt}\}_{r,t}$ and amenities $\{A_{rt}\}_{r,t}$. This allows us to analyze the the fundamental determinants of agricultural specialization by projecting the endogenous agricultural employment share on regional fundamentals and the two state variables of the system, namely the distribution of skills and population size in 1880. We do so in Table 8.We start by analyzing the cross-sectional relationship between income and agricultural specialization. In the first column, we report the relationship between the agricultural employment share and average manufacturing earnings. Both of these variables are directly observed in the data and strongly negatively correlated. Filtered through the lens of the structural model, this implies that average regional earnings for low skilled workers, Θ_{rt}^L, are also lower in rural areas. In column three we show that this gradient is even stronger for high-skilled individuals: holding unskilled income fixed, regions with a higher agricultural employment share offer particularly bad earnings opportunities for high skilled workers.

In the remaining columns we directly focus on the relationship between agricultural employment and the underlying regional fundamentals. In particular, we report the results from the specification

$$\ln s_{rAt} = \delta_t + \beta_{NA} \ln Q_{rNAt} + \beta_A \ln Q_{rAt} + \gamma A_{rt} + \zeta \lambda_{r1880} + \eta \ln \text{pop}_{r1880} + u_{rt},$$

where δ_t is a year fixed effect. In columns four to six we report the bivariate partial correlations. Rural, agricultural regions are regions with low productivity in the non-agricultural sector, a comparative advantage in the production of agricultural goods and low amenities. The last column reports the respective partial correlations. In particular, the coefficient on regional amenities and the regional skill share drops by a factor of ten. This reflects the existing cross-sectional correlation with regional productivity, in particular non-agricultural productivity Q_{rNA}.

	e_{rt}^{Man}	$\ln \Theta_{rt}^{L}$	$\ln \Theta_{rt}^{H}$	log Ag. empl. share ($\ln s_{rA}$)			
Ag. empl share ($\ln s_{rA}$)	-0.098***	-0.050***	-0.045***				
	(0.012)	(0.011)	(0.005)				
Low-skilled earnings ($\ln \Theta_{rt}^{L}$)			1.051***				
			(0.022)				
Non Ag. Productivity ($\ln Q_{rNA}$)				-0.331***			-0.344***
				(0.020)			(0.007)
Ag. Productivity ($\ln Q_{rA}$)					0.404***		0.408***
					(0.022)		(0.006)
Amenities (A_r)						-0.430***	-0.055***
						(0.075)	(0.013)
Skill share in 1880 (λ_{r1880})		0.096**	0.013	0.139	-1.594***	-0.466**	-0.117***
		(0.038)	(0.009)	(0.123)	(0.150)	(0.196)	(0.032)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ln pop 1880	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3585	2580	2580	3225	3225	2580	2580
R^2	0.989	0.941	0.991	0.891	0.902	0.810	0.989

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. All specification control for the population in 1880.

Table 8: Fundamental Determinants of Agricultural Specialization

In Table 9 we focus directly on the dynamics of spatial productivity and amenities. In particular, we consider a simple autoregressive specification

$$y_{rt} = \delta_t + \delta_r + \beta y_{r,t-1} + u_{rt}$$

where y denotes either log sectoral productivity, $\ln Q_{rst}$, or the level of amenities A_{rt} and δ_r is a region fixed effect. The first four columns show that productivities are mean-reverting and that there is an important fixed, region-specific component determining spatial productivity between 1880 and 2000. It is also interesting to note that these patterns differ slightly across sectors. In particular, there seems to be more mean-reversion in the non-agricultural sector and the regional fixed effects explain less of the cross-sectional variation. The last two columns show the same result for regional amenities. In contrast to productivity, amenities show more persistence.

	$\ln Q_{NA,r,t}$		$\ln Q_{A,r,t}$		$\ln A_{r,t}$	
Lag	0.700***	0.204***	0.438***	0.109***	0.877***	-0.080**
	(0.016)	(0.017)	(0.020)	(0.023)	(0.022)	(0.038)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Commuting Zone FE	No	Yes	No	Yes	No	Yes
N	2580	2580	2580	2580	1935	1935
R^2	0.689	0.879	0.277	0.580	0.630	0.856

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels.

Table 9: The Process of Spatial Fundamentals

C.3 The Correlation of Agricultural Employment and Future Earnings

	Earnings		Amenities
	Low-skilled	High-skilled	
Share of Variance Explained	32.7%	48.6%	27%

Notes: The table reports the average R^2 of a regression of regional log earning (by skill groups) and regional amenities on 100 fixed effects for the respective percentiles of the agricultural employment shares. Hence, the regression we run is $y_{rt} = \sum_{q=1}^{100} \theta_{rt-1}^{q} + u_{rt}$, where θ_{rt-1}^{q} is an indicator variable for the q percentile of the distribution of agricultural employment shares at time $t-1$. We run the regression for each period 1910, 1940, 1970 and 2000 and report the time average of the R^2 in the table. The individual R^2 for the four years are 24.9% 41.6% 35.2% and 29.0% for low skilled earnings, 48.2%, 61.5%, 48.0%, 36.8% for high-skilled earnings and 21.6% 34.2% 26.3% and 25.8% for amenities.

Table 10: Agricultural Employment and Future Local Attractiveness

C.4 Local Productivities and Market Access

As stated in the main part of the paper an important part of our calibration approach is that we calibrate the cross-sectional distribution of sectoral productivities $\{Q_{rAt}, Q_{rNAt}\}_{rt}$ and amenities $\{A_{rt}\}_{rt}$ as structural residuals of the model as is commonly done in the new quantitative spatial economics literature (see (Redding and Rossi-Hansberg, 2016) for a recent and excellent review of this literature).

The local sectoral productivities calibrated this way, $\{Q_{rAt}, Q_{rNAt}\}_{rt}$, like aggregate TFP in the Solow model are residuals necessary to fit the data on wages and sectoral employment conditional on the calibrated parameters of the model. While they pick up all factors that make one region more productive than another, some measurable others not, there is one factor that determines the productivity of a location in a very direct way: its integration into the national transportation network. In this section we use direct measures on the "market access cost" via various modes of transport by Donaldson and Hornbeck (2016) for each county in the US for 1880 and 1910.

We use the data on county to county transport cost to construct a commuting zone level market access index for 1880 and 1910. In particular we average the market access cost across all county to county pairs within two given commuting zones to obtain a measure for the ease of transporting goods between these two zones. Then we take the destination population share weighted sum of market access cost for each origin commuting zone to obtain our market access index.

We show that these market access costs are important ingredients of $\{Q_{rAt}, Q_{rNAt}\}_{rt}$. As Table 11 shows regional productivties are low in places that have a high market access cost. The table further shows that those commuting zones that see their market access cost decline the most are also those who see there sectoral productivites increase the most. We take this as evidence that transportation costs are one of the directly measurable ingredients in regional productivity shifters that our framework picks up as $\{Q_{rAt}, Q_{rNAt}\}_{rt}$.

	(1) Q_NA	(2) Q_A	(3) Q_NA	(4) Q_A	(5) Q_NA	(6) Q_A	(7) Q_NA_growth	(8) Q_A_growth	(9) Q_NA_growth	(10) Q_A_growth
log_cost	-1.897***	-2.077***	-3.937***	-3.760***	-5.205***	-4.291***				
	(0.224)	(0.125)	(0.353)	(0.309)	(0.290)	(0.195)				
cost_dh_growth							-3.937***	-3.760***	-4.005***	-2.395***
							(0.594)	(0.368)	(0.388)	(0.328)
_cons	-5.508***	-2.470***	-0.412	1.734*	2.753***	3.059***	0.346**	-0.238*	0.330**	0.0951
	(0.539)	(0.290)	(0.884)	(0.773)	(0.729)	(0.489)	(0.124)	(0.0994)	(0.114)	(0.0968)
Year FE	YES	YES	YES	YES	YES	YES				
CZ FE			YES	YES						
State FE					YES	YES			YES	YES
N	1242	1242	1242	1242	1242	1242	621	621	621	621
adj. R^2	0.133	0.286	0.789	0.576	0.538	0.455	0.166	0.192	0.337	0.399

Standard errors in parentheses

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

Notes: We use data on the cost of reaching any other county from a given county provided by Donaldson and Hornbeck (2016), take the destination population weighted sum and aggregate them to the commuting zone level. This gives an index of market access for 1880 and 1910 for every commuting zone. Here we relate changes in agricultural productivity implied by the model to changes in market access costs. The commuting zones that saw there market access cost decrease the most are the ones that saw there agricultural productivity increase the most.

Table 11: Agricultural Productivity and Market Access Cost

ONLINE APPENDIX

Not for Publication

A Theory: Additional Proofs and Derivations

A.1 Additional derivations on the model with selection

In this section we derive some convenient properties of the selection model. Recall from (4) that the distribution of individual skills is given by

$$F_s^h(z) = e^{-\Psi_s^h z^{-\zeta}}, \tag{37}$$

where Ψ_s^h parametrizes the average level of productivity of individuals of skill h in region r in sector s and ζ the dispersion of skills. The following result will turn out to be useful

Lemma 4. *Let $\{x_i\}_{i=1}^n$ be distributed iid according to*

$$F_{x_i}(x) = e^{-A_i x^{-\zeta}}.$$

Then

$$E\left[x_i | x_i = \max_i [x_i]\right] = \Gamma\left(1 - \frac{1}{\zeta}\right)\left(\sum_{i=1}^n A_i\right)^{1/\zeta}. \tag{38}$$

Note that this object does not depend on i.

Proof. Suppose that $i = 1$ and let us derive the conditional distribution of x_1, conditional that x_1 is the highest $\{x_j\}_j$. The joint distribution of $\{x_j\}_j$ is given by

$$F(x_1, x_2, ..., x_3) = \prod_{j=1}^n F(x_i) \tag{39}$$

because of independence. Hence, we get that

$$\begin{aligned}
P\left(x_1 \leq m | x_1 = \max_j [x_j]\right) &= \frac{1}{P\left(x_1 = \max_j [x_j]\right)} \int_0^m \prod_{j=2}^n P\left(x_j < x\right) dF_{x_1}(x) \\
&= \frac{1}{P\left(x_1 = \max_j [x_j]\right)} \int_0^m \prod_{j=2}^n e^{-A_j x^{-\zeta}} A_1 \zeta x^{-\zeta-1} e^{-A_1 x^{-\zeta}} dx \\
&= \frac{A_1}{P\left(x_1 = \max_j [x_j]\right)} \int_0^m \zeta x^{-\zeta-1} \prod_{j=1}^n e^{-A_j x^{-\zeta}} dx \\
&= \frac{A_1}{P\left(x_1 = \max_j [x_j]\right)} \int_0^m \zeta x^{-\zeta-1} e^{-(\Sigma_j A_j) x^{-\zeta}} dx.
\end{aligned}$$

Now let us derive $P\left(x_1 = \max_j [x_j]\right)$. We get that

$$P\left(x_1 = \max_j [x_j]\right) = P\left(x_j \leq x_1 \text{ for all } j \geq 2\right) = \int_{x_1} \frac{\partial F}{\partial x_1}(x_1, x_1, x_1, ..) dx_1.$$

Using (39), we get that

$$\begin{aligned}\frac{\partial F(x_1,x_2,...,x_3)}{\partial x_1} &= \frac{\partial F(x_1)}{\partial x_1}\left(\prod_{j=2}^{n}F(x_i)\right)=e^{-A_1x^{-\zeta}}A_1\zeta x_1^{-\zeta-1}\left(\prod_{j=2}^{n}e^{-A_ix_j^{-\zeta}}\right)\\ &= A_1\zeta x_1^{-\zeta-1}\left(\prod_{j=1}^{n}e^{-A_ix_i^{-\zeta}}\right)\\ &= A_1\zeta x_1^{-\zeta-1}e^{-\sum_i^n A_ix_i^{-\zeta}}.\end{aligned}$$

Hence,

$$\begin{aligned}\int_{x_1}\frac{\partial F}{\partial x_1}(x_1,x_1,x_1,..)\,dx_1 &= \int_{x_1}A_1\zeta x_1^{-\zeta-1}e^{-(\sum_i^n A_i)x_1^{-\zeta}}dx_1\\ &= \frac{A_1}{\sum_i^n A_i}\int_{x_1}\frac{\zeta}{(\sum_i^n A_i)^{1/\zeta}}\left(\frac{x_1}{(\sum_i^n A_i)^{1/\zeta}}\right)^{-\zeta-1}e^{-\left(\frac{x_1}{(\sum_i^n A_i)^{1/\zeta}}\right)^{-\zeta}}dx_1\\ &= \frac{A_1}{\sum_i^n A_i}.\end{aligned}$$

Substituting this above yields

$$\begin{aligned}P\left(x_1\leq m|x_1=\max_j\left[x_j\right]\right) &= \left(\sum_i^n A_i\right)\int_0^m\zeta x^{-\zeta-1}e^{-(\Sigma_jA_j)x^{-\zeta}}dx\\ &= \int_0^m\frac{\zeta}{\kappa}\left(\frac{x}{\kappa}\right)^{-\zeta-1}e^{-\left(\frac{x}{\kappa}\right)^{-\zeta}}dx,\end{aligned}$$

where $\kappa=\left(\sum_jA_j\right)^{1/\zeta}$. This is a Frechet distribution with shape ζ and scale κ, i.e.

$$F_{x_1|x_1=\max_j\left[x_j\right]}(m)=e^{-\left(\frac{x}{\left(\Sigma_jA_j\right)^{1/\zeta}}\right)^{-\zeta}}=e^{-\left(\Sigma_jA_j\right)x^{-\zeta}}.$$

This implies (38) □

Lemma 4 is useful because it allows us to calculate sectoral earnings and the sectoral supply of efficiency units. Consider first sectoral earnings. Let $y_{rs}^{i,h}=w_{rs}z_s^{i,h}$ be the earnings of individual i of skill h in region r working in sector s. The distribution of earnings is

$$P\left(y_{sr}^{i,h}<y\right)=P\left(z_s^{i,h}<\frac{y}{w_{rs}}\right)=e^{-\Psi_s^h\left(\frac{y}{w_{rs}}\right)^{-\zeta}}=e^{-\Psi_s^hw_{rs}^{\zeta}y^{-\zeta}}.$$

Hence, Lemma 4 implies that

$$E\left[y_{rs}^{i,h}|y_{rs}^{i,h}=\max_s\left\{y_{rs}^{i,h}\right\}\right]=\Gamma\left(1-\frac{1}{\zeta}\right)\left(\sum_s\Psi_s^h w_{rs}^{\zeta}\right)^{1/\zeta}.$$

Similarly, the average labor supply in sector s is given by

$$\begin{aligned}
E\left[z_s^{i,h}|y_{rs}^{i,h}=\max_s\left\{y_{rs}^{i,h}\right\}\right] &= E\left[z_s^{i,h}|w_{rs}z_s^{i,h}=\max_s\left\{w_{rs}z_s^{i,h}\right\}\right] \\
&= E\left[z_s^{i,h}|z_s^{i,h}=\max_k\left\{\frac{w_{rk}}{w_{rs}}z_k^{i,h}\right\}\right] \\
&= \Gamma\left(1-\frac{1}{\zeta}\right)\left(\Psi_s^h+\sum_{k\neq s}\Psi_k^h\left(\frac{w_{rk}}{w_{rs}}\right)^{\zeta}\right)^{1/\zeta} \\
&= \Gamma\left(1-\frac{1}{\zeta}\right)\frac{1}{w_{rs}}\left(\sum_s\Psi_s^h w_{rs}^{\zeta}\right)^{1/\zeta}.
\end{aligned}$$

Also, the share of people working in sector s is given by

$$s_{rs}^h=P\left(y_{rs}^{i,h}=\max_k\left\{y_{rk}^{i,h}\right\}\right)=\frac{\Psi_s^h w_{rs}^{\zeta}}{\sum_s\Psi_s^h w_{rs}^{\zeta}}=\Psi_s^h\left(\frac{w_{rs}}{\left(\sum_s\Psi_s^h w_{rs}^{\zeta}\right)^{1/\zeta}}\right)^{\zeta}.$$

It is useful to define the endogenous scalar of *average earning of skill group h in region r as*

$$\Theta_r^h=\left(\sum_s\Psi_s^h w_{rs}^{\zeta}\right)^{1/\zeta}.$$

Then we can write the sectoral employment share of skill group h, s_{rs}^h, as

$$s_{rs}^h = \Psi_s^h\left(\frac{w_{rs}}{\Theta_r^h}\right)^{\zeta},$$

and the aggregate amount of sectoral earnings of skill group h, $w_{rs}H_{rs}^h$, as

$$\begin{aligned}
w_{rs}H_{rs}^h &= L_r\lambda_r^h P\left(y_{rs}^{i,h}=\max_k\left\{y_{rk}^{i,h}\right\}\right)E\left[y_{rs}^{i,h}|y_{rs}^{i,h}=\max_s\left\{y_{rs}^{i,h}\right\}\right] \\
&= L_r\lambda_r^h\Gamma\left(1-\frac{1}{\zeta}\right)\left(s_{rs}^h\Theta_r^h\right).
\end{aligned}$$

Note that we can write

$$s_{rs}^h\Theta_r^h=\Psi_s^h w_{rs}^{\zeta}\left(\Theta_r^h\right)^{1-\zeta}.$$

Hence,

$$w_{rs}H_{rs}^{h} = L_r\Gamma\left(1-\frac{1}{\zeta}\right)w_{rs}^{\zeta}\left(\lambda_r^{h}\Psi_s^{h}\left(\Theta_r^{h}\right)^{1-\zeta}\right).$$

Note that the aggregate level of aggregate human capital of individuals of skill h working in sector s in region r is given by

$$\begin{aligned} H_{rh}^{h} &= L_r\Gamma\left(1-\frac{1}{\zeta}\right)w_{rs}^{\zeta-1}\left(\lambda_r^{h}\Psi_s^{h}\left(\Theta_r^{h}\right)^{1-\zeta}\right) = L_r\Gamma\left(1-\frac{1}{\zeta}\right)\left(\lambda_r^{h}\Psi_s^{h}\left(\frac{w_{rs}}{\Theta_r^{h}}\right)^{\zeta-1}\right) \\ &= L_r\Gamma\left(1-\frac{1}{\zeta}\right)\left(\lambda_r^{h}\Psi_s^{h}\left(\frac{s_{rs}^{h}}{\Psi_s^{h}}\right)^{\frac{\zeta-1}{\zeta}}\right) = L_r\Gamma\left(1-\frac{1}{\zeta}\right)\left(\lambda_r^{h}\left(\Psi_s^{h}\right)^{\frac{1}{\zeta}}s_{rs}^{h\,\frac{\zeta-1}{\zeta}}\right). \end{aligned}$$

This also implies that total earnings in sector s in region r are given by

$$Y_{rs} = \sum_h w_{rs}H_{rs}^{h} = L_r\Gamma\left(1-\frac{1}{\zeta}\right)w_{rs}^{\zeta}\sum_h \lambda_r^{h}\Psi_s^{h}\left(\Theta_r^{h}\right)^{1-\zeta}.$$

Finally, *aggregate* earnings in region r, Y_r, are given by

$$\begin{aligned} Y_r &= \sum_h\left[\sum_s w_{rs}H_{rs}^{h}\right] = \sum_h L_r\Gamma\left(1-\frac{1}{\zeta}\right)\left(\lambda_r^{h}\Theta_r^{h}\right) \\ &= L_r\Gamma\left(1-\frac{1}{\zeta}\right)\sum_h\left(\lambda_r^{h}\Theta_r^{h}\right). \end{aligned}$$

A.2 Regularity conditions for PIGL preferences: Details

The PIGL preferences are given by

$$V = \frac{1}{\eta}\left(\frac{e}{a(P)}\right)^{\eta} - b(P).$$

The corresponding expenditure function is

$$e(V,p) = \eta^{1/\eta}\left(V+b(p)\right)^{1/\eta}a(P). \tag{40}$$

We need to ensure sure that the Slutsky matrix, i.e. the Hessian matrix of the expenditure function, is symmetric, negative semidefinite and satisfies H · P = 0, where P is the vector of prices. Using (40), we get that

$$\frac{\partial e(V,p)}{\partial p_j} = e(V,p)\left[\frac{1}{\eta}\frac{b_j(p)}{V+b(p)} + \frac{a_j(p)}{a(p)}\right], \tag{41}$$

where $b_j(p) = \frac{\partial b(p)}{\partial p_j}$ and $a_j(p) = \frac{\partial a(p)}{\partial p_j}$. Note that we can also write (41) as

$$\begin{aligned}\frac{\partial e(V,p)}{\partial p_j} &= e(V,p)\left[\frac{1}{\eta}\frac{b_j(p)}{\frac{1}{\eta}\left(\frac{e(V,p)}{a(P)}\right)^{\eta}} + \frac{a_j(p)}{a(p)}\right] \\ &= \left(\frac{e(V,p)}{a(p)}\right)^{1-\eta}\left[b_j(p)a(p) + a_j(p)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]. \end{aligned} \tag{42}$$

Equation (41) also implies that

$$\begin{aligned}\frac{\partial^2 e}{\partial p_j^2} &= \frac{\partial e}{\partial p_j}\left[\frac{1}{\eta}\frac{b_j(p)}{V+b(p)} + \frac{a_j(p)}{a(p)}\right] + e(V,p)\frac{\partial}{\partial p_j}\left[\frac{1}{\eta}\frac{b_j(p)}{V+b(p)} + \frac{a_j(p)}{a(p)}\right] \\ &= e(V,p)\left[\frac{1}{\eta}\frac{b_j(p)}{V+b(p)} + \frac{a_j(p)}{a(p)}\right]^2 + e(V,p)\left[\frac{1}{\eta}\frac{b_{jj}(p)(V+b(p)) - b_j(p)^2}{(V+b(p))^2} + \frac{a_{jj}(p)a(p) - a_j(p)^2}{(a(p))^2}\right].\end{aligned}$$

Simplifying terms yields

$$\frac{\partial^2 e}{\partial p_j^2} = \left(\frac{e}{a(P)}\right)^{1-2\eta}\left(\frac{1}{p_j}\right)^2 \Lambda_j,$$

where

$$\Lambda_j \equiv (1-\eta)\left(b_j(p)p_j\right)^2 a(p) + \left(2b_j(p)p_j a_j(p)p_j + b_{jj}(p)p_j^2 a(p)\right)\left(\frac{e}{a(P)}\right)^{\eta} + a_{jj}(p)p_j^2\left(\frac{e}{a(P)}\right)^{2\eta}. \tag{43}$$

For the cross-derivative we get that

$$\begin{aligned}\frac{\partial^2 e}{\partial p_j p_k} &= \frac{\partial e}{\partial p_k}\left[\frac{1}{\eta}\frac{b_j(p)}{V+b(p)} + \frac{a_j(p)}{a(p)}\right] + e(V,p)\frac{\partial}{\partial p_k}\left[\frac{1}{\eta}\frac{b_j(p)}{V+b(p)} + \frac{a_j(p)}{a(p)}\right] \\ &= e(V,p)\left[\frac{1}{\eta}\frac{b_k(p)}{V+b(p)} + \frac{a_k(p)}{a(p)}\right]\left[\frac{1}{\eta}\frac{b_j(p)}{V+b(p)} + \frac{a_j(p)}{a(p)}\right] \\ &\quad + e(V,p)\left[\frac{1}{\eta}\frac{b_{jk}(p)(V+b(p)) - b_j(p)b_k(p)}{(V+b(p))^2} + \frac{a_{jk}(p)a(p) - a_j(p)a_k(p)}{a(p)^2}\right].\end{aligned}$$

Again, we can write this as

$$\frac{\partial^2 e}{\partial p_j p_k} = \left(\frac{e(V,p)}{a(p)}\right)^{1-2\eta}\frac{1}{p_j p_k}\Omega_{jk} \tag{44}$$

where

$$\begin{aligned}\Omega_{jk} &= (1-\eta)b_j(p)p_j b_k(p)p_k a(p) + \left(b_j(p)p_j a_k(p)p_k + b_k(p)p_k a_j(p)p_j + b_{jk}(p)p_j p_k a(p)\right)\left(\frac{e}{a(P)}\right)^{\eta} \\ &\quad + a_{jk}(p)p_j p_k\left(\frac{e}{a(P)}\right)^{2\eta}.\end{aligned}$$

In our application we have

$$\begin{aligned} b(p) &= \frac{\nu}{\gamma}\left(\frac{p_A}{p_M}\right)^{\gamma} - \frac{\nu}{\gamma} + \frac{1}{\eta} = \tilde{b}(p) - \frac{\nu}{\gamma} + \frac{1}{\eta} \\ a(p) &= p_A^{\phi} p_M^{1-\phi} \end{aligned}$$

Define the respective elasticities as

$$\begin{aligned} \varepsilon_j^1 &= \tilde{b}_j(p)\frac{p_j}{\tilde{b}(p)} \\ \varepsilon_j^2 &= \tilde{b}_{jj}(p)\frac{p_j}{\tilde{b}_j(p)} = \tilde{b}_{jj}(p)\frac{p_j^2}{\frac{\tilde{b}_j(p)p_j}{\tilde{b}(p)}}\frac{1}{\tilde{b}(p)} = \tilde{b}_{jj}(p)\frac{p_j^2}{\varepsilon_j^1}\frac{1}{\tilde{b}(p)} \end{aligned}$$

Similarly,

$$\begin{aligned} \tau_j^1 &= a_j(p)\frac{p_j}{a(p)} \\ \tau_j^2 &= a_{jj}(p)\frac{p_j^2}{\tau_j^1}\frac{1}{a(p)}. \end{aligned}$$

Hence, (43) reduces to

$$\begin{aligned} \Lambda_j &\equiv (1-\eta)\left(\varepsilon_j^1\tilde{b}(p)\right)^2 a(p) + \left(2\varepsilon_j^1\tilde{b}\tau_j^1 a(p) + \varepsilon_j^2\varepsilon_j^1\tilde{b}(p)a(p)\right)\left(\frac{e}{a(P)}\right)^{\eta} + \tau_j^2\tau_j^1 a(p)\left(\frac{e}{a(P)}\right)^{2\eta}. \\ &= a(p)\left\{(1-\eta)\left(\varepsilon_j^1\right)^2\left(\tilde{b}\right)^2 + \left(2\varepsilon_j^1\tau_j^1 + \varepsilon_j^2\varepsilon_j^1\right)\tilde{b}\left(\frac{e}{a(P)}\right)^{\eta} + \tau_j^2\tau_j^1\left(\frac{e}{a(P)}\right)^{2\eta}\right\}. \end{aligned}$$

In our application, we have

$$\tau = \begin{bmatrix} \tau_A^1 & \tau_{NA}^1 \\ \tau_A^2 & \tau_{NA}^2 \end{bmatrix} = \begin{bmatrix} \phi & 1-\phi \\ \phi-1 & -\phi \end{bmatrix} \text{ and } \varepsilon = \begin{bmatrix} \varepsilon_A^1 & \varepsilon_{NA}^1 \\ \varepsilon_A^2 & \varepsilon_{NA}^2 \end{bmatrix} = \begin{bmatrix} \gamma & -\gamma \\ \gamma-1 & -(\gamma+1) \end{bmatrix}.$$

Hence,

$$\begin{aligned} 2\varepsilon_A^1\tau_A^1 + \varepsilon_A^2\varepsilon_A^1 &= 2\gamma\phi + (\gamma-1)\gamma = \gamma(2\phi+\gamma-1) \\ 2\varepsilon_M^1\tau_M^1 + \varepsilon_M^2\varepsilon_M^1 &= -2\gamma(1-\phi) + \gamma(\gamma+1) = \gamma(\gamma+1-2+2\phi) = \gamma(2\phi+\gamma-1). \end{aligned}$$

This implies that

$$\Lambda_j = \Lambda_k = \Lambda(p) = a(p)\left\{(1-\eta)\gamma^2\left(\tilde{b}\right)^2 - \gamma(1-\gamma-2\phi)\tilde{b}\left(\frac{e}{a(P)}\right)^{\eta} - \phi(1-\phi)\left(\frac{e}{a(P)}\right)^{2\eta}\right\}.$$

To evaluate Ω_{jk}, note that

$$\begin{aligned} b_{jk}(p) &= \tilde{b}_{jk}(p) = -\gamma^2 \frac{\tilde{b}(p)}{p_j p_k} \\ a_{jk}(p) &= \phi(1-\phi)\frac{a(p)}{p_j p_k}. \end{aligned}$$

Hence,

$$\Omega_{jk} = a(p)\left\{(1-\eta)\varepsilon_j^1\varepsilon_k^1\tilde{b}(p)^2 + \left(\varepsilon_j^1\tau_k^1 + \varepsilon_k^1\tau_j^1 - \gamma^2\right)\tilde{b}(p)\left(\frac{e}{a(P)}\right)^{\eta} + \phi(1-\phi)\left(\frac{e}{a(P)}\right)^{2\eta}\right\}.$$

Note that

$$\begin{aligned} \varepsilon_A^1\tau_M^1 + \varepsilon_M^1\tau_A^1 &= \gamma(1-\phi) - \gamma\phi = \gamma - 2\phi\gamma \\ \varepsilon_M^1\tau_A^1 + \varepsilon_A^1\tau_M^1 &= -\gamma\phi + \gamma(1-\phi) = \gamma - 2\phi\gamma \end{aligned}$$

Hence,

$$\begin{aligned} \Omega_{jk} = \Omega_{kj} &= a(p)\left\{-(1-\eta)\gamma^2\tilde{b}(p)^2 + \left(\gamma - 2\phi\gamma - \gamma^2\right)\tilde{b}(p)\left(\frac{e}{a(P)}\right)^{\eta} + \phi(1-\phi)\left(\frac{e}{a(P)}\right)^{2\eta}\right\} \\ &= a(p)\left\{-(1-\eta)\gamma^2\tilde{b}(p)^2 + \gamma(1-2\phi-\gamma)\tilde{b}(p)\left(\frac{e}{a(P)}\right)^{\eta} + \phi(1-\phi)\left(\frac{e}{a(P)}\right)^{2\eta}\right\} \\ &= -a(p)\left\{(1-\eta)\gamma^2\tilde{b}(p)^2 - \gamma(1-\gamma-2\phi)\tilde{b}(p)\left(\frac{e}{a(P)}\right)^{\eta} - \phi(1-\phi)\left(\frac{e}{a(P)}\right)^{2\eta}\right\} \\ &= -\Lambda(p) \end{aligned} \tag{45}$$

This implies that the Hessian is given by

$$\begin{aligned} H &= \begin{bmatrix} \frac{\partial^2 e(p,V)}{\partial^2 p_A} & \frac{\partial^2 e(p,V)}{\partial p_A \partial p_{NA}} \\ \frac{\partial^2 e(p,V)}{\partial p_{NA}\partial p_A} & \frac{\partial^2 e(p,V)}{\partial^2 p_{NA}} \end{bmatrix} \\ &= \begin{bmatrix} \left(\frac{e}{a(P)}\right)^{1-2\eta}\left(\frac{1}{p_A}\right)^2\Lambda & -\left(\frac{e(V,p)}{a(p)}\right)^{1-2\eta}\frac{1}{p_A p_{NA}}\Lambda \\ -\left(\frac{e(V,p)}{a(p)}\right)^{1-2\eta}\frac{1}{p_A p_{NA}}\Lambda & \left(\frac{e}{a(P)}\right)^{1-2\eta}\left(\frac{1}{p_{NA}}\right)^2\Lambda \end{bmatrix} \\ &= \left(\frac{e(V,p)}{a(p)}\right)^{1-2\eta}\frac{\Lambda(p)}{p_A p_{NA}}\begin{bmatrix} \frac{p_{NA}}{P_A} & -1 \\ -1 & \frac{p_A}{p_{NA}} \end{bmatrix}. \end{aligned}$$

Hence, for the matrix to be negative semi-definite, we need that

$$\begin{aligned} 0 &\geq \Lambda(p) \\ &= (1-\eta)\gamma^2\left(\tilde{b}\right)^2 - \gamma(1-\gamma-2\phi)\tilde{b}\left(\frac{e}{a(P)}\right)^{\eta} - \phi(1-\phi)\left(\frac{e}{a(P)}\right)^{2\eta}. \end{aligned}$$

Substituting for $\tilde{b}$ and using that $a(p)=1$,this condition is

$$\begin{aligned} 0 &\geq (1-\eta)\gamma^2\left(\frac{\nu}{\gamma}\left(\frac{p_A}{p_M}\right)^{\gamma}\right)^2 - \gamma(1-\gamma-2\phi)\frac{\nu}{\gamma}\left(\frac{p_A}{p_M}\right)^{\gamma}e^{\eta} - \phi(1-\phi)e^{2\eta} \\ &= (1-\eta)\nu^2\left(\frac{p_A}{p_M}\right)^{2\gamma} - (1-\gamma-2\phi)\nu\left(\frac{p_A}{p_M}\right)^{\gamma}e^{\eta} - \phi(1-\phi)e^{2\eta}. \end{aligned}$$

Hence,

$$\frac{\phi(1-\phi)}{\nu\left(\frac{p_A}{p_M}\right)^{\gamma}}e^{2\eta} + (1-\gamma-2\phi)e^{\eta} \geq (1-\eta)\nu\left(\frac{p_A}{p_M}\right)^{\gamma}.$$

This is the condition contained in Section B.5 in the Appendix.

A.3 The elasticity of substitution

The elasticity of substitution is given by

$$\sigma^{ES} = \frac{\frac{\partial^2 e(p,V)}{\partial P_A \partial P_{NA}} e(p,V)}{\frac{\partial e(p,V)}{\partial P_A}\frac{\partial e(p,V)}{\partial P_{NA}}}.$$

From (44), (45) and (42) we know that

$$\frac{\partial^2 e(p,V)}{\partial P_A \partial P_{NA}} = -\left(\frac{e}{a(p)}\right)^{1-2\eta}\frac{1}{p_A p_{NA}}\Lambda$$

and that

$$\frac{\partial e(p,V)}{\partial P_j} = \left(\frac{e(V,p)}{a(p)}\right)^{1-\eta}\left[b_j(p)a(p) + a_j(p)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right].$$

Hence,

$$\begin{aligned}
\sigma^{ES} &= \frac{-\left(\frac{e}{a(p)}\right)^{1-2\eta}\frac{1}{p_A p_{NA}}\Lambda e(p,V)}{\left(\frac{e(V,p)}{a(p)}\right)^{1-\eta}\left[b_A(p)a(p)+a_A(p)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]\left(\frac{e(V,p)}{a(p)}\right)^{1-\eta}\left[b_{NA}(p)a(p)+a_{NA}(p)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]} \\
&= \frac{-a(p)\Lambda}{p_A p_{NA}\left[b_A(p)a(p)+a_A(p)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]\left[b_{NA}(p)a(p)+a_{NA}(p)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]} \\
&= \frac{-a(p)\Lambda}{\left[b_A(p)p_A a(p)+a_A(p)p_A\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]\left[b_{NA}(p)p_{NA}a(p)+a_{NA}(p)p_{NA}\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]}.
\end{aligned}$$

Now note that

$$\begin{aligned}
b_A(p)p_A &= \tilde{b}(p)\varepsilon_A^1 = \gamma\tilde{b}(p) \\
b_{NA}(p)p_{NA} &= \tilde{b}(p)\varepsilon_{NA}^1 = -\gamma\tilde{b}(p) \\
a_A(p)p_A &= \phi a(p) \\
a_{NA}(p)p_{NA} &= (1-\phi)a(p).
\end{aligned}$$

Hence,

$$\begin{aligned}
\sigma^{ES} &= \frac{-a(p)\Lambda}{\left[\gamma\tilde{b}(p)a(p)+\phi a(p)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]\left[-\gamma\tilde{b}(p)a(p)+(1-\phi)a(p)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]} \\
&= \frac{-\left\{(1-\eta)\gamma^2\left(\tilde{b}\right)^2-\gamma(1-\gamma-2\phi)\tilde{b}\left(\frac{e}{a(P)}\right)^{\eta}-\phi(1-\phi)\left(\frac{e}{a(P)}\right)^{2\eta}\right\}}{\left[\gamma\tilde{b}(p)+\phi\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]\left[-\gamma\tilde{b}(p)+(1-\phi)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]}
\end{aligned}$$

Note that the expenditure share on the agricultural good is given by

$$\vartheta_A(e,p) = \phi+\nu\left(\frac{p_A}{p_M}\right)^{\gamma}e^{-\eta} = \phi+\gamma\tilde{b}(p)\left(\frac{e(V,p)}{a(P)}\right)^{-\eta}.$$

Hence,

$$\gamma\tilde{b}(p) = (\vartheta_A(e,p)-\phi)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}.$$

Substituting this above, we get that

$$\begin{aligned}
\sigma^{ES} &= \frac{-(1-\eta)(\vartheta_A(e,p)-\phi)^2\left(\frac{e(V,p)}{a(P)}\right)^{2\eta}+(1-\gamma-2\phi)(\vartheta_A(e,p)-\phi)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\left(\frac{e}{a(P)}\right)^{\eta}+\phi(1-\phi)\left(\frac{e}{a(P)}\right)^{2}}{\left[(\vartheta_A(e,p)-\phi)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}+\phi\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]\left[-(\vartheta_A(e,p)-\phi)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}+(1-\phi)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]} \\
&= \frac{-(1-\eta)(\vartheta_A(e,p)-\phi)^2\left(\frac{e(V,p)}{a(P)}\right)^{2\eta}+(1-\gamma-2\phi)(\vartheta_A(e,p)-\phi)\left(\frac{e(V,p)}{a(P)}\right)^{2\eta}+\phi(1-\phi)\left(\frac{e}{a(P)}\right)^{2\eta}}{\left[\vartheta_A(e,p)\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]\left[(1-\vartheta_A(e,p))\left(\frac{e(V,p)}{a(P)}\right)^{\eta}\right]} \\
&= \frac{-(1-\eta)(\vartheta_A(e,p)-\phi)^2+(1-\gamma-2\phi)(\vartheta_A(e,p)-\phi)+\phi(1-\phi)}{[\vartheta_A(e,p)][(1-\vartheta_A(e,p))]} \\
&= \frac{-(1-\eta)(\vartheta_A-\phi)^2+(1-\gamma-2\phi)(\vartheta_A-\phi)+\phi(1-\phi)}{\vartheta_A(1-\vartheta_A)}.
\end{aligned}$$

Rearranging terms yields

$$\begin{aligned}
\sigma^{ES} &= 1-\eta+\frac{-(\gamma-\eta(1-\phi))\vartheta_A(e,p)(1-\phi)+[\gamma+\phi\eta]\phi(1-\vartheta_A(e,p))}{\vartheta_A(e,p)(1-\vartheta_A(e,p))} \\
&= 1-\eta-\frac{(\gamma-\eta(1-\phi))(1-\phi)}{1-\vartheta_A(e,p)}+\frac{[\gamma+\phi\eta]\phi}{\vartheta_A(e,p)}.
\end{aligned}$$

A.4 Selection and Space

In our model, the sectoral selection of workers does affect the agricultural productivity gap. Moreover, such selection effects are absent in a model without space, as for example considered in Lagakos and Waugh (2011). To see that explicitly, consider the following aggregate setup, which is essentially the one considered in Lagakos and Waugh (2011), but expressed in our notation. Suppose production takes place according to

$$Y_s = Q_s H_s, \tag{46}$$

where H_s is human capital and Q_s is productivity. If skills are drawn from a distribution $F_s(z) = e^{-\Psi_s z^{-\zeta}}$ and the sector wages are given by w_s, aggregate output in both sectors is given by

$$Y_s = Q_s(s_s)^{\frac{\zeta-1}{\zeta}}(\Psi_s)^{\frac{1}{\zeta}} = Q_s s_s(s_s)^{-\frac{1}{\zeta}}(\Psi_s)^{\frac{1}{\zeta}}, \tag{47}$$

where s_s is the sectoral employment share. Hence, the ratio of labor productivity in the two sectors is given by

$$\frac{Y_A/L_A}{Y_{NA}/L_{NA}} = \frac{Y_A/(\omega_A L)}{Y_A/((1-\omega_A)L)} = \frac{Q_A}{Q_{NA}}\left(\frac{\Psi_A}{\Psi_{NA}}\right)^{1/\zeta}\left(\frac{s_A}{1-s_A}\right)^{-1/\zeta}.$$

The first two terms are the technology and human capital differences. The last term is selection: the higher s_A, the *lower* labor productivity in agriculture. This is the effect highlighted in Lagakos and

Waugh (2011).

Note that $\frac{Y_A/L_A}{Y_{NA}/L_{NA}}$ is the the ratio of *physical* productivities. The agricultural productivity, in contrast, refers to sectoral *value added* per worker. Given the production function in (46), sectoral prices are given by $p_s = Q_s^{-1} w_s$. Hence, using (47), relative value added is given by

$$\frac{p_A Y_A}{p_{NA} Y_{NA}} = \left(\frac{Q_A}{Q_{NA}}\right)^{-1} \frac{w_A}{w_{NA}} \frac{Q_A s_A \left(s_A\right)^{-\frac{1}{\zeta}} \left(\Psi_A\right)^{\frac{1}{\zeta}}}{Q_{NA} s_{NA} \left(s_{NA}\right)^{-\frac{1}{\zeta}} \left(\Psi_{NA}\right)^{\frac{1}{\zeta}}} = \frac{s_A}{s_{NA}} \frac{w_A}{w_{NA}} \left(\frac{\Psi_A}{\Psi_{NA}}\right)^{1/\zeta} \left(\frac{s_A}{s_{NA}}\right)^{-1/\zeta}.$$

But now note that $\frac{w_A}{w_{NA}} \left(\frac{\Psi_A}{\Psi_{NA}}\right)^{1/\zeta} \left(\frac{s_A}{s_{NA}}\right)^{-1/\zeta} = 1$, so that

$$GAP = \frac{p_A Y_A / L_A}{p_{NA} Y_{NA} / L_{NA}} = \frac{s_A}{s_{NA}} \left(\frac{s_A}{s_{NA}}\right)^{-1} = 1.$$

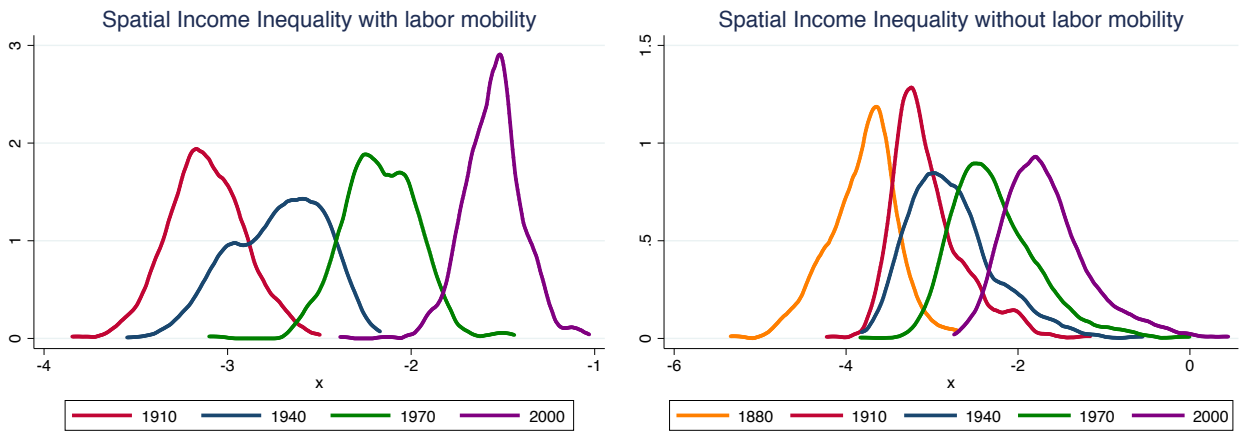
Hence, under the parametric structure of the Frechet distribution, selection does *not* show up as a productivity gap. While a high agricultural employment share does reduce physical productivity, this high employment share has to coincide with a high relative wage. In the case of the Frechet distribution, these two effects exactly cancel out. In our economy with spatial heterogeneity, this is not the case as seen.

A.5 Accounting for the aggregate process of structural change

Our model generates the structural transformation through a combination of income and price effects. For given preference parameters it is an empirical question whether the implied decline in agricultural employment shares is due demand side forces or driven by supply side factors. For the observed distribution and level of aggregate earnings in each cross-section and the price ratio of agricultural relative to non-agricultural output we estimate the preference parameters in Sections B.6 and B.8.

A.6 Spatial Income Inequality with and without mobility

Figure 14 displays the cross-sectional distribution of unskilled earnings, $\ln \Theta_{rt}^L$, in the baseline model (left panel) and in the economy without spatial mobility (i.e. where moving costs are prohibitively high). Spatial mobility - even if it is costly - is a potent arbitraging force. If spatial mobility is possible, the variance in the cross-sectional distribution of earnings decreases and aggregate growth raises all boats (or rather commuting zones). In particular, there is hardly any overlap in the distribution of earnings in 1910 and 1970. This is very different in the absence of spatial mobility. In that case, the distribution of earnings fans out and a large number of regions have lower earnings in 1970 than some other regions in 1910.



Notes: The figure displays the cross-sectional distribution of unskilled earnings, $\ln \Theta_{rt}^{L}$, in the baseline model (left panel) and in the economy without spatial mobility, i.e. where moving costs are prohibitively high (right panel).

Figure 14: Income Equality and Spatial Mobility

A.7 Welfare and Agricultural Specialization

We start by analyzing the cross-sectional relationship between welfare, income and agricultural specialization. As before we measure regional welfare for agents of skill group h in region r by

$$\mathscr{W}_{rt}^{h} = E^{h}\left[U_{r}\right] + A_{rt} = \frac{\Gamma_{\eta/\zeta}}{\eta}\psi\left(r_{t+1}\right)^{\eta-1}\left(\Theta_{r}^{h}\right)^{\eta} + \Lambda_{t,t+1} + A_{rt}. \tag{48}$$

Recall that equation (48) does not take the option value of individual mobility into account, i.e. $\mathscr{W}_{rt}^{h}$ is *not* the average utility of someone born in region r. Rather equation (48) is the average utility of an individual born in r if mobility was not allowed. We hence view $\mathscr{W}_{rt}^{h}$ as a measure of "regional attractiveness". Note that the cross-sectional variation in $\mathscr{W}_{rt}^{h}$ stems from differences in average earnings, Θ_{r}^{h}, and in amenities A_{rt}. The former reflects local productivity and sorting by skill, while, in the main body of the paper, the latter is taken to be follow fully exogenous stochastic process.

In Table 12 we study the cross-sectional relationship between welfare (and its individual components) and agricultural specialization as measured by the (log of) the agricultural employment share. In column 1 we show that there is a strong negative relationship between regional welfare of unskilled individuals and agricultural specialization. Column 2 shows that this gradient is particularly strong for high-skilled individuals: holding unskilled individuals' welfare, $\mathscr{W}_{rt}^{L}$, fixed, regions with a higher agricultural employment share provide particularly little utility for high skilled workers. Columns 3 and 4 show the same patterns for average earnings:[52] agricultural regions have low earnings and particularly so for high skilled individuals. Note also the strong *positive* correlation between skilled and unskilled earnings across re-

[52] Note that this object is not directly observed in the data. In the data we see average earnings in the manufacturing sector. This measure conflates skill prices with the endogenous sorting of skilled individuals across space and does not account for the sectoral selection within regions.

	$\mathscr{W}_{rt}^L$	$\mathscr{W}_{rt}^H$	$\ln\Omega_{rt}^L$	$\ln\Omega_{rt}^H$	A_{rt}	$\mathscr{W}_{rt}^L$	$\ln\Omega_{rt}^L$	A_{rt}
ln Ag. empl. share	-0.267***	-0.060***	-0.067***	-0.085***	-0.223***	-0.138***	-0.080***	-0.098***
	(0.015)	(0.001)	(0.004)	(0.003)	(0.015)	(0.013)	(0.008)	(0.012)
Spatial welfare low skilled		1.012***						
		(0.002)						
ln Avg earnings low skilled				1.011***				
				(0.008)				
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Commuting Zone FE	No	No	No	No	No	Yes	Yes	Yes
Observations	2580	2580	2580	2580	2580	2580	2580	2580
R^2	0.435	0.997	0.901	0.990	0.125	0.894	0.960	0.822

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels.

Table 12: Welfare, Income and Agricultural Specialization

gions. This is due to the positive correlation between agricultural and non-agricultural productivity. In column 5, we focus on location amenities A_{rt}: agricultural intensive places are *not* only locations with little earnings potential, but also places, where spatial amenities are low. Finally, in the last three columns, we show that all these results are qualitatively robust to the inclusion of commuting zone fixed effects. This is consistent with our findings in Table 8, where we showed a negative correlation between agricultural employment shares and regional productivity conditional on regional fixed effects.

Our theory particularly stresses structural change is biased against regions, which rely on agriculture. To analyze the effects of this differential exposure of the structural transformation in more detail, we now consider a specification along the lines of

$$y_{rt} = \delta_t + \alpha s_{rAt-1} + \beta s_{rAt-1} \times \Delta s_{At} + \gamma y_{rt-1} + u_{rt},$$

where y_{rt} is our regional outcome measure (i.e. regional welfare $\mathscr{W}_{rt}^h$ or earnings Θ_{rt}^h) , s_{rAt-1} is the lagged agricultural employment share and Δs_{At} denotes the decline in the aggregate agricultural employment share between t and $t-1$. The main coefficient of interest is β, which we expect to be negative: the larger the *decline* in the agricultural share, the more adversely will regions with a comparative advantage in agriculture be affected. In Table 13 we report the results for both welfare (columns 1 - 3) and earnings (columns 4 - 6) of low skilled individuals. Columns 1 and 4 first show that there is a negative relationship between the growth in welfare and earnings and the lagged agricultural employment share. Columns 2 and 5 are our main specification, where we show that a decline in agricultural employment share particularly hurts regions with a large initial agricultural share, i.e. $\beta < 0$. This spatial bias of the structural transformation is also unaffected when we introduce a full set of commuting zone fixed effects.

	$\mathscr{W}_{rt}^L$			$\ln \Theta_{rt}^L$		
ln lagged Ag. empl. share	-0.034**	0.047***	-0.032*	-0.072***	-0.034***	-0.084***
	(0.014)	(0.014)	(0.018)	(0.004)	(0.005)	(0.009)
ln lagged Ag. empl. share × Ag. Decline		-1.029***	-1.265***		-0.476***	-0.303***
		(0.206)	(0.166)		(0.065)	(0.058)
Lagged Spatial welfare low skilled	0.882***	0.869***	-0.078**			
	(0.023)	(0.022)	(0.038)			
Lagged ln Avg earnings low skilled				0.500***	0.487***	-0.094***
				(0.018)	(0.018)	(0.032)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Commuting Zone FE	No	No	Yes	No	No	Yes
N	1935	1935	1935	1935	1935	1935
R^2	0.761	0.768	0.918	0.911	0.913	0.958

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels.

Table 13: The Spatial Bias of the Structural Transformation

B Calibration and Accounting

In this section we describe the detail of our accounting and calibration strategy. We start be describing the mapping between the data and the spatial fundamentals $\{Q_{rAt}, Q_{rNAt}, A_{rt}\}_{rt}$.

B.1 Inferring Spatial Fundamentals

Here we describe how we can use the model to infer the whole vector of sectoral productivities $\{Q_{rs}\}_{rs}$ and amenities $\{A_r\}_r$ across space for each cross-section of the data. To do so we use three pieces of information:

- Data on average regional manufacturing *earnings* $\{e_{rMan}\}_r$
- Data on the distribution of workers across space $\{L_r\}_r$
- Data on regional sectoral employment shares $\{s_{rs}\}_{rs}$

We are going to show that our framework implies a unique mapping from this data to the set of fundamentals $\{Q_{rAt}, Q_{rNAt}, A_{rt}\}_{rt}$. Throughout this section we take the regional skilled employment shares λ_{rt} and all parameters as given.[53] Note that we drop time subscripts t since the described procedure is identical for each cross-section. Subsequent sections will outline our strategy for obtaining these objects in detail.

Step 1: From bodies to human capital ($\{L_r\}_r$ and $\{s_{rs}\}_{rs} \to H_{rs}$) Let the share of skilled people λ_r be given. Let $L_r^H \equiv \lambda_r L_r$ and $L_r^L \equiv (1-\lambda_r) L_r$. Given $\{L_r^L, L_r^H\}_r$ and $\{s_{rA}^H, s_{rA}^L\}$, we can directly calculate

[53] In practice of course, much of the calibration process of spatial fundamentals and the remaining parameters is necessarily carried out jointly. Here we describe the calibration strategy behind individual parameters separately.

the supply of sectoral human capital as

$$H_{rNA} = \Gamma\left(1-\frac{1}{\zeta}\right)\left[L_r^H\left(s_{rNA}^H\right)^{\frac{\zeta-1}{\zeta}}(q\mu)^{\frac{1}{\zeta}}+L_r^L\left(s_{rNA}^L\right)^{\frac{\zeta-1}{\zeta}}\right] \tag{49}$$

$$H_{rA} = \Gamma\left(1-\frac{1}{\zeta}\right)\left[L_r^H\left(s_{rA}^S\right)^{\frac{\zeta-1}{\zeta}}q^{\frac{1}{\zeta}}+L_r^L\left(s_{rA}^L\right)^{\frac{\zeta-1}{\zeta}}\right]. \tag{50}$$

Note that

$$\frac{s_{rNA}^H}{1-s_{rNA}^H}=\mu\left(\frac{w_{rNA}}{w_{rA}}\right)^{\zeta}=\mu\frac{s_{rNA}^L}{1-s_{rNA}^L}.$$

This implies that

$$s_{rNA}^H=\frac{\mu s_{rNA}^L}{1+(\mu-1)s_{rNA}^L}. \tag{51}$$

And therefore

$$\begin{aligned} s_{rA}^H &= 1-s_{rNA}^H=\frac{1-s_{rNA}^L}{1+(M-1)s_{rNA}^L}=\frac{1-s_{rNA}^L}{1+(M-1)\left(s_{rNA}^L-1\right)+M-1} \\ &= \frac{s_{rA}^L}{\mu-(\mu-1)s_{rA}^L}. \end{aligned}$$

Using this in (49) and (50) yields

$$H_{rNA} = \Gamma\left(1-\frac{1}{\zeta}\right)\left[L_r^H\left(\frac{\mu s_{rNA}^L}{1+(\mu-1)s_{rNA}^L}\right)^{\frac{\zeta-1}{\zeta}}(q\mu)^{\frac{1}{\zeta}}+L_r^L\left(s_{rNA}^L\right)^{\frac{\zeta-1}{\zeta}}\right] \tag{52}$$

$$H_{rA} = \Gamma\left(1-\frac{1}{\zeta}\right)\left[L_r^H\left(\frac{s_{rA}^L}{\mu-(\mu-1)s_{rA}^L}\right)^{\frac{\zeta-1}{\zeta}}q^{\frac{1}{\zeta}}+L_r^L\left(s_{rA}^L\right)^{\frac{\zeta-1}{\zeta}}\right]. \tag{53}$$

Equations (52) and (53) determines H_{rNA} and H_{rA} as a functions of observables $\{L_r^L, L_r^H\}_r$, parameters (q,μ,ζ) and s_{rA}^L, which we treat as unobserved. To link s_{rA}^L to observables, we use that the *regional* sectoral employment share in agriculture is given by

$$\begin{aligned} s_{rA} &= \frac{L_r^H s_{rA}^H+L_r^L s_{rA}^L}{L_r^H+L_r^L}=\frac{L_r^H}{L_r^H+L_r^L}\left(\frac{s_{rA}^L}{\mu-(\mu-1)s_{rA}^L}\right)+\frac{L_r^L}{L_r^H+L_r^L}s_{rA}^L \\ &= \lambda_r\left(\frac{s_{rA}^L}{\mu-(\mu-1)s_{rA}^L}\right)+(1-\lambda_r)s_{rA}^L. \end{aligned} \tag{54}$$

Equation (54) provides a mapping from s_{rA} to s_{rA}^L given parameters and the regional skilled employment share $\lambda_r = \frac{L_r^H}{L_r^H + L_r^L}$. In particular, the solution to this quadratic equation is given by[54]

$$s_{rA}^L = \frac{1}{2}\frac{\lambda_r + (1-\lambda_r)\mu + (\mu-1)s_{A,r}}{(1-\lambda_r)(\mu-1)} - \sqrt{\frac{1}{4}\left(\frac{\mu + (1-\lambda_r)\mu + (\mu-1)\omega_{rA}}{(1-\lambda_r)(\mu-1)}\right)^2 - \frac{\mu}{(1-\lambda_r)(\mu-1)}\omega_{rA}}$$

Hence, (54), (52) and (53) provide a mapping

$$\{L_r^L, L_r^H, s_{rA}\}_r + \text{parameters} \to \{H_{rA}, H_{rNA}\}_r.$$

From now on we treat the regional sectoral human capital supply, $\{H_{rA}, H_{rNA}\}_r$, as data.

Step 2: From earnings to skill prices In the data we observe $\{e_{rMan}\}_r$, i.e. earnings per manufacturing employee. In the theory this object is given by

$$e_{rMan} = \frac{w_{rNA}H_{rNA}}{L_r^H s_{rNA}^H + L_r^L s_{rNA}^L}, \tag{55}$$

where $H_{rNA} = H_{rNA}^H + H_{rNA}^L$. Again, we can infer from s_{rNA}^H and ω_{rNA}^L from (54) and (51). From above we already know H_{rNA}. Given average observed earnings e_{rMan}, (55) implies a mapping from the data to w_{rNA}. Hence, from now on we treat w_{rNA} as data.

We then use the model to predict agricultural wages (per efficiency unit) $\{w_{rA}\}_r$. Again: this is due both the fact that we do not have agricultural wages *and* the fact that - to properly use the accounting - we have to make sure that wages are consistent with the equilibrium conditions. The model implies that

$$w_{rA} = \left(\frac{s_{rA}^L}{1-s_{rA}^L}\right)^{1/\zeta} w_{rNA}. \tag{56}$$

Using the mapping from s_{rA} to s_{rA}^L given in (54), we can use (56) to infer w_{rA} from information on w_{rNA} and ω_{rA}. Hence, we get a mapping

$$\{L_r^L, L_r^H, \omega_{rs}, e_{rMan}\}_r + \text{parameters} \to \{w_{rNA}, w_{rA}\}_r.$$

From now on, we treat $\{w_{rNA}, w_{rA}\}_r$ as data.

Step 3: From human capital to productivities (given sectoral wages) Consider the sector s labor market clearing condition. These have the structure

$$w_{rs}H_{rs} = (1-\alpha)\pi_{rs}\Omega_s.$$

[54] The other root is not a solution to the equation as it implies $s_{rA}^L > 1$.

where

$$\Omega_s = \begin{cases} \vartheta_s PC & \text{if } s \neq I \\ S - (1-\delta)K & \text{if } s = I \end{cases}.$$

where S are aggregate savings, K is the current capital stock, ϑ_s is the sector s aggregate expenditure share and PC is total consumption spending. Summing across regions implies $\sum_r w_r H_{rs} = (1-\alpha)\Omega_s$, so that

$$\pi_{rs} = \frac{w_{rs}H_{rs}}{\sum_r w_{rs}H_{rs}}.$$

For the agricultural sector, we immediately get

$$\pi_{rA} = \frac{w_{Ar}H_{Ar}}{\sum_j w_{Aj}H_{Aj}}. \tag{57}$$

Hence, from now on we can treat π_{rs} as data. To infer regional productivities, note that the model implies that (see (19))

$$\pi_{rs} = \frac{\left(\frac{Q_{rs}}{w_{rs}^{1-\alpha}}\right)^{\sigma-1}}{\sum_{j=1}^{R}\left(\frac{Q_{js}}{w_{js}^{1-\alpha}}\right)^{\sigma-1}}.$$

Define the *relative* trade shares (relative to region 1 say) as

$$\frac{\pi_{rs}}{\pi_{1s}} = \left(\frac{Q_{rs}}{Q_{1s}}\right)^{\sigma-1}\left(\frac{w_{rs}}{w_1}\right)^{(1-\sigma)(1-\alpha)}.$$

Hence,

$$\left(\frac{Q_{rs}}{Q_{1s}}\right)^{\sigma-1} = \frac{\pi_{rs}}{\pi_{1s}}\left(\frac{w_{rs}}{w_1}\right)^{(\sigma-1)(1-\alpha)}, \tag{58}$$

so that we can calculate $\frac{Q_{rs}}{Q_{1s}}$ directly from the data. These are $R-1$ equations for the $R-1$ relative productivities in each sector. Finally we can use the normalization

$$1 = \sum_r (Q_{rs})^{\sigma-1} = (Q_{1s})^{\sigma-1} + \sum_{r=2}^{R}(Q_{rs})^{\sigma-1} = (Q_{1s})^{\sigma-1}\left[1 + \sum_{r=2}^{R}(\frac{Q_{rs}}{Q_{1s}})^{\sigma-1}\right],$$

so that

$$Q_{1s}^{\sigma-1} = \left[1 + \sum_{r=2}^{R}(\frac{Q_{rs}}{Q_{1s}})^{\sigma-1}\right]^{-1}. \tag{59}$$

Which then finally allows to back out all sectoral productivities in levels.

Note that these derivations took the regional skill shares λ_r as given. These are of course an equilibrium

object and hence have to be consistent with individual labor supply. The spatial supply equation is given by

$$L_{rt}^{h} = \sum_{j=1}^{R} \left(\rho_{jrt}^{h} L_{jt-1}^{h} \right), \tag{60}$$

where ρ_{jrt}^{h} is given in (9) as

$$\rho_{jrt}^{h} = \frac{exp\left(\frac{1}{\kappa} \left(\frac{\Gamma(1-\eta/\zeta)}{\eta} \psi \left(r_{t+1} \right)^{\eta-1} \left(\Theta_{r}^{h} \right)^{\eta} + A_{r} - MC_{jr} \right) \right)}{\sum_{l=1}^{R} exp\left(\frac{1}{\kappa} \left(\frac{\Gamma(1-\eta/\zeta)}{\eta} \psi \left(r_{t+1} \right)^{\eta-1} \left(\Theta_{r}^{h} \right)^{\eta} + A_{l} - MC_{jl} \right) \right)}. \tag{61}$$

Recall that we only observe skill shares on the county level in 1940. It turns out the structure of the model provides sufficient restrictions to imply skill shares, λ_{rt}, for every other cross-section given λ_{1940}. Skill shares and amenities are calibrated jointly using equations (60) and (61) which allows for the construction of two useful mappings:

$$A_{rt}^{k+1} = -\kappa \log \left[\sum_{j} \frac{exp\left(\frac{1}{\kappa} \left(\frac{\Gamma_{\eta/\zeta}}{\eta} \psi \left(r_{t} \right)^{\eta-1} \left(\Theta_{rt}^{h} \right)^{\eta} + A_{rt}^{k} - MC_{jr} \right) \right)}{\sum_{l=1}^{R} exp\left(\frac{1}{\kappa} \left(\frac{\Gamma_{\eta/\zeta}}{\eta} \psi \left(r_{t+1} \right)^{\eta-1} \left(\Theta_{lt}^{h} \right)^{\eta} - MC_{jl} \right) \right)} \frac{L_{jt-1}^{h}}{L_{rt}^{h}} \right] \quad \text{with} \quad \sum_{r} A_{rt} = 0$$

and

$$L_{jt-1}^{h} = \frac{L_{rt}^{h} - \sum_{j \neq r} \rho_{jr}^{h} L_{jt-1}^{h}}{\rho_{jj}^{h}} \quad \text{with} \quad \sum_{h} L_{rt}^{h} = L_{rt}$$

Where $\sum_{r} A_{rt} = 0$ is a normalization and $\sum_{h} L_{rt}^{h} = L_{rt}$ is an adding up constraint, forcing workers across skill groups to add up to total region r workforce whose data equivalent our model matches exactly for every cross-section. The procedure to jointly back out skill shares and amenities is then as follows:

1. Use 1940 skill shares from the data to calculate $\{\Theta_{r1940}^{h}\}_{rh}$. Guess skill shares in 1910 and use worker stocks in 1910 to calculate $\{L_{r1910}^{h}\}_{rh}$. Also guess a vector of amenities that satisfies the normalization, $\{A_{r1940}^{0}\}$. Then use the above mapping to obtain a new amenities guess A_{r1940}^{k+1}.

2. Since these amenities are not indexed by skill type they may imply a L_{rt-1}^{h} that yields skill shares for 1910 different from our initial guess. Using A_{r1940}^{k+1} we can construct ρ_{jr}^{h} and then use the second mapping to update skill shares in 1910.

3. We iterate between updating skill shares and amenities until convergence, then use the 1910 skill shares as data to obtain skill shares in 1880 and amenities in 1910 in the same way.

To obtain skill shares and amenities for 1970 by constructing similar but forward looking mappings. This procedure leaves us with a panel of skill shares and amenities for all commuting zones and cross-

sections.This concludes the proof that we there is a unique mapping

$$\{L_r, \omega_{rs}, e_{rMan}\}_r \to \{Q_{rA}, Q_{rNA}, A_r\}_r.$$

B.2 Matching the time series of aggregate GDP

In the data we observe the distribution of manufacturing earnings $\{e_{rMan}\}_r$. While we force the model to replicate *relative* earnings across space, we also want to ensure that aggregate GDP grows at a given constant rate. This can be enforced by simply scaling the level of observed manufacturing earnings. To see this, recall that observed earnings and unobserved skill prices were given in (55) as

$$e_{rMan} = \frac{w_{rNA} H_{rNA}}{L_r^H s_{rNA}^H + L_r^L s_{rNA}^L}.$$

We showed above that we can calculate H_{rNA} and s_{rs}^h directly from the data without using data on earnings. Hence, level shifts in e_{rMan} will simply imply level shifts in unobserved skill prices.

Aggregate GDP is proportional to labor earnings, i.e.

$$GDPpc = \frac{1}{1-\alpha} \sum_r \left(w_{rA} H_{rA} + w_{rNA} H_{rNA} \right).$$

Using (55) and (56) we get that

$$w_{rNA} H_{rNA} = e_{rMan} \left(L_r^H s_{rNA}^H + L_r^L s_{rNA}^L \right)$$

and

$$\begin{aligned} w_{rA} H_{rA} &= \left(\frac{s_{rA}^L}{1 - s_{rA}^L} \right)^{1/\zeta} w_{rNA} H_{rA} \\ &= \left(\frac{s_{rA}^L}{1 - s_{rA}^L} \right)^{1/\zeta} \frac{e_{rMan} \left(L_r^H s_{rNA}^H + L_r^L s_{rNA}^L \right)}{H_{rNA}} H_{rA} \end{aligned}$$

so that

$$GDPpc = \frac{1}{1-\alpha} \sum_r e_{rMan} \left(L_r^H s_{rNA}^H + L_r^L s_{rNA}^L \right) \left[1 + \left(\frac{s_{rA}^L}{1 - s_{rA}^L} \right)^{1/\zeta} \frac{H_{rA}}{H_{rNA}} \right].$$

Given the data we can there for just scale e_{rMan} to get a particular level of GDP. In particular, let the scaled earnings be

$$\tilde{e}_{rMan,t}(\kappa) \equiv \kappa e_{rMan,t}.$$

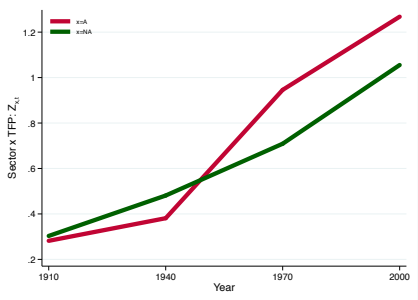


Figure 15: Aggregate sectoral productivity growth over time

Let $GDPpc(\kappa)$ be the level of GDP per capita corresponding to a scaling factor κ. Then

$$\begin{aligned} GDPpc(\kappa) &= \frac{1}{1-\alpha}\sum_r \tilde{e}_{rMan,t}(\kappa)\left(L_r^H s_{rNA}^H + L_r^L s_{rNA}^L\right)\left[1+\left(\frac{s_{rA}^L}{1-s_{rA}^L}\right)^{1/\zeta}\frac{H_{rA}}{H_{rNA}}\right] \\ &= \kappa\frac{1}{1-\alpha}\sum_r e_{r,Man,t}\left(L_r^H s_{rNA}^H + L_r^L s_{rNA}^L\right)\left[1+\left(\frac{s_{rA}^L}{1-s_{rA}^L}\right)^{1/\zeta}\frac{H_{rA}}{H_{rNA}}\right] \\ &= \kappa GDPpc(1). \end{aligned}$$

Hence, if we normalize GDPpc to unity in 1880, the scaling factor is given by

$$\kappa_t^{GDP} = \frac{1.02^{t-1880}}{\frac{1}{1-\alpha}\sum_r\left[e_{rMan,t}\left(1+\left(\frac{s_{rAt}^L}{1-s_{rAt}^L}\right)^{\frac{1}{\zeta}}\frac{H_{rAt}}{H_{rNAt}}\right)\left(L^H s_{rNAt}^H + L^L s_{rNAt}^L\right)\right]} = \frac{1.02^{t-1880}}{GDPpc(1)}.$$

This normalization implies that

$$GDPpc^{1880}\left(\kappa_{1880}^{GDP}\right) = 1$$

and that GDPpc grows at a rate of 2% a year.

B.3 Obtaining skilled employment for each cross-section

The US Population Census from 1940 is the first one with detailed education variables and the last one with a complete set of county identifiers (see data section for details). We define skilled employment as workers with at least a completed high school degree in 1940. This yields the vector of skilled employment shares $\{\lambda_{r1940}\}_r$ as a data object in 1940. In this section we demonstrate that the structure of the model implies that a single cross-section of skilled employment shares implies them for all other cross-section, given that we match local populations, sectoral employment and wages exactly. The assumptions that the aggregate skilled employment share is constant throughout the years and that skills are perfectly transmitted between parents and children also play a crucial role. We proceed in two steps. First we show show how to use the 1940 data to go backward in time and obtain $\{\lambda_{r1880}\lambda_{r1910}\}_r$ then we show how to

infer $\{\lambda_{r1970}\lambda_{r2000}\}_r$ from $\{\lambda_{r1940}\}_r$.From $\{\lambda_{r1940}\}_r$ to $\{\lambda_{r1970}\lambda_{r2000}\}_r$

For the forward direction we take as inputs $\{\lambda_{r1940}\}_r$ and $\{A_{rt}\}_{rt}$ so that we can construct ρ^h_{jr1970} from the data. The procedure to find $\{\lambda_{r1970}\lambda_{r2000}\}_r$ is then as follows:

1. Guess the spatial distribution of skills in 1970, i.e. λ_{r1970}. Note that the share of skilled in the population stays constant, i.e. we can already impose the condition

$$\sum_r \lambda_{r1940} L_{r1940} = \sum_r \lambda_{r1970} L_{r1970}.$$

2. Given $\lambda_{r,1940}$, we can use the exact same procedure we used before, i.e.

   (a) Use the data on earnings, population and employment shares to back out Q_{rs1970}

   (b) Given step (a) we can calculate Θ^H_{r1970} and Θ^L_{r1970}. As $\psi(r_{t+1})$ is known along the BGP, the spatial mobility probability ρ^H_{jr1970} and ρ^L_{jr1970} in (61) only depend on A_{r1970}. The evolution of population is given by - see (60) -

$$\begin{aligned} L_{r1970} &= L^H_{r1970} + L^L_{r1970} = \sum_{j=1}^{R} \left(\rho^H_{jr1970} L^H_{j1940}\right) + \sum_{j=1}^{R} \left(\rho^h_{jr1970} L^H_{j1940}\right) \\ &= \sum_{j=1}^{R} \left(\rho^H_{jr1970}\lambda_{j1940} + \rho^L_{jr1970}\left(1-\lambda_{j1940}\right)\right) L_{j1940}. \end{aligned}$$

   Hence, given $\{L_{j1940}\}_j$ and $\{\lambda_{j1940}\}_j$ this determines L_{r1970} as a function of $\{A_r\}_r$. We can therefore chose $\{A_{r1970}\}_r$ to match $\{L_{r1970}\}_r$.

3. The *implied* skill share in 1970 is then given by

$$\lambda^{k+1}_{r1970} = \frac{\sum_{j=1}^{R} \rho^H_{jr1970}\lambda_{j1940} L_{j1940}}{\sum_{j=1}^{R} \left(\rho^H_{jr1970}\lambda_{j1940} + \rho^L_{jr1970}\left(1-\lambda_{j1940}\right)\right) L_{j1940}}$$

   If $\lambda^{k+1}_{r1970} = \lambda^k_{r1970}$, we are done. If $\lambda^{k+1}_{r1970} \neq \lambda^k_{r1970}$, we need to update our guess for λ_{r1970} and start back at step 2.

4. Once we find a fixed point for λ_{r1970}, we can start the same procedure for 2000.

B.3.1 From $\{\lambda_{r1940}\}_r$ to $\{\lambda_{r1880}\lambda_{r1910}\}_r$

We can rewrite equation (10) as follows

$$L^h_{jt-1} = \frac{L^h_{rt} - \sum_{j\neq r} \rho^h_{jr} L^h_{jt-1}}{\rho^h_{jj}} \quad \text{with} \quad \sum_h L^h_{rt} = L_{rt}$$

Suppose for simplicity that we know $\{A_{rt}\}_{rt}$ so that we can construct ρ^h_{jr1940} from the data. Likewise $L^H_{r1940} = \lambda_{r1940} L_{r1940}$ is a data object. The procedure to find λ_{r1910} is then as follows

1. Guess $\lambda^0_{r1910} = \lambda_{r1910}$ and use the 1910 population count (which the model matches exactly) to construct $L^{H,0}_{r1910} = \lambda^0_{r1910} L_{r1910}$.

2. Use the following mapping to update $\{L^{H,0}_{r1910}\}_r$

$$L^{H,1}_{j1910} = \frac{L^H_{r1940} - \sum_{j \neq r} \rho^H_{jr1940} L^{H,0}_{j1910}}{\rho^H_{jj1940}} \quad \text{with} \quad \sum_h L^h_{r1910} = L_{r1910}$$

3. We iterate on this mapping until convergence, ie until $L^{H,k+1}_{j1910} = L^{H,k}_{j1910}$. Then we form $\lambda_{r1910} = \frac{L^{H,k+1}_{j1910}}{L_{j1910}}$.

4. Now we treat $\{\lambda_{r1910}\}_r$ as data and repeat the same exercise to find λ_{r1880}.

Note that in practice we do not know A_{r1910}. We then alternate between the procedure above and the procedure for finding $\{A_r\}_r$ outlined above to jointly solve for $\{A_{r1910}\}_r$ and $\{\lambda_{r1910}\}_r$.

The results of this subsection and the last is a panel of skill shares $\{\lambda_{rt}\}_{rt}$ for all cross-sections to which our model is calibrated, the 1940 cross-section of which coincides with the data.

B.4 Calibration: Skill Supplies

We calibrate the parameter ζ, i.e. the dispersion in individual productivity, to match earnings dispersion in the 1940 Census data. Consider the set of individuals in region r of skill type h in industry s at time t. The distribution of earnings is $F^h_r(y) = e^{-\left(\Theta^h_r\right)^\zeta \times y^{-\zeta}}$. Hence,

$$P\left(\ln y^i_{rs} \leq k\right) \quad = \quad P\left(y^i_{rs} \leq e^k\right) = e^{-\left(\Theta^h_r\right)^\zeta \left(e^k\right)^{-\zeta}} = e^{-e^{-\zeta\left(k + \ln\left(\Theta^h_r\right)\right)}},$$

i.e. log earnings are Gumbel distributed. This implies that $\text{var}\left(\ln y^i_{rs}\right) = \frac{\pi^2}{6}\left(\frac{1}{\zeta}\right)^2$, i.e. the cross-sectional variance across individuals within skill group-region-sector cells is constant. Hence, consider the regression

$$\ln y^i_{rsh} = \delta_r + \delta_s + \delta_h + u^i_{rsh}, \tag{62}$$

where y^i_{rsh} denotes individual earnings and δ_t, δ_s and δ_h are region, sector and skill-group fixed effects. We then set

$$\zeta = \left(\frac{\pi^2}{6} \frac{1}{\text{var}\left(\hat{u}^i_{rsh}\right)}\right)^{1/2}$$

where var$\left(\hat{u}^i_{rsh}\right)$ is the variance of the estimated residuals from (62). For the baseline calibration we obtain estimated of $\zeta = 1.62$ from 62. Alternative specifications all yielded estimates for ζ within the $(1.32 - 1.62)$ interval, see table 14.

Specification	A	B	C	D	E	F
(1) Sector FE	✓	✓	✓	✓	✗	✗
(2) Skill Group FE	✓	✓	✓	✓	✓	✓
(3) Age FE	✓	✗	✗	✗	✗	✗
(4) CZ FE	✓	✓	✓	✓	✓	✓
Interactions	✗	(1) and (2)	(1) and (2), (2) and (4)	(1) and (2), (1) and (4)	(1) and (2)	(2) and (4)
Comments	Sample Weights	Drop 1st & 99th pct	✗	✗	Agri only	Non-Agri only
Implied ζ	1.62	1.60	1.48	1.48	1.32	1.42

Table 14: Calibration of ζ

B.5 Calibrating comparative and absolute advantage

We calibrate q and μ to match the aggregate skill premium and the aggregate relative manufacturing employment share of skilled workers in 1940. In the model these two objects can be constructed as a function of data, the above calibrated ζ and the two parameters q and μ only. More specifically, we calculate in the 1940 cross-section

$$\text{skill premium} = \frac{\sum_{r=1}^{R} L_r \lambda_r \Theta_r^H(\zeta,\mu,q)/\xi}{\sum_{r=1}^{R} L_r (1-\lambda_r) \Theta_r^H(\zeta,\mu,q)/(1-\xi)} \tag{63}$$

and

$$\text{skilled manufacturing premium} = \frac{\sum_{r=1}^{R} L_r \lambda_r s_{rNA}^H(\zeta,\mu,q)/\xi}{\sum_{r=1}^{R} L_r (1-\lambda_r) s_{rNA}^L(\zeta,\mu,q)/(1-\xi)}. \tag{64}$$

where ξ is the total number of skilled workers in the economy. Recall that total population is normalized to 1 for each cross-section so that $1-\xi$ is the number of unskilled people. In 1940, we observe manufacturing earnings e_{rMan}, the skill shares λ_r and the allocation of people across space L_r. Following the same steps as in Section B, this data is sufficient to calculate the skill specific skill shares and Θ_r^h. Hence, we can directly calculate the two moments in (63) and (64) as a function of (q,μ). This also implies that we can directly calibrate (q,μ) without having to use any other information. Hence, we can calibrate (q,μ) once and then calibrate the remaining structural parameters. This calibration yields $q = 0.61$ and $\mu = 4.08$.

B.6 Estimation of η

In Table 15 we report both the OLS estimates of (**??**) and the results of the non-linear least squares of the actual expenditure system. Columns 1 to 3 show that η is precisely estimated at around 0.3. In particular, this estimate is almost constant across urban and village household and the farm population. This is also seen in the last three columns, where we report the non-linear estimate for η. Estimating η independently

	OLS Estimates $\ln \vartheta_{A,i} = \beta - \eta \times \ln e_i$ Full Sample			Non-linear least squares $\vartheta_{A,i} = \beta_0 + \beta_1 e_i^{-\eta}$		
				Urban	Villages	Farm pop.
ln expenditure	-0.314***	-0.336***	-0.325***	-0.312	-0.315	-0.315
	(0.007)	(0.009)	(0.011)			
ln expenditure × village			0.007			
			(0.020)			
ln expenditure × farm			-0.032*			
			(0.018)			
Rural/Urban FE	Yes	Yes	Yes			
Family Size	No	Yes	Yes			
Race FE	No	Yes	Yes			
Interview Date FE	No	Yes	Yes			
Observations	5945	5342	5342	2747	1280	1324
R^2	0.286	0.376	0.377			

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Estimates of η

for the three samples yields essentially the same estimate. Hence, for the remainder of the paper, we take $\eta = 0.32$.

B.7 Estimating mobility costs

Recall from the main part of the paper that we specify moving frictions as a quadratic function of distance as follows:

$$MC_{jr} = \begin{cases} 0 & \text{if } r = j \\ \tau + \delta_1 d_{jr} + \delta_2 d_{jr}^2 & \text{if } r \neq j \end{cases}$$

where $\tau > 0$ corresponds to a fixed cost of moving, while δ_1 and δ_2 flexibly parameterize how costs vary with distance. In the data we normalize d_{ij} so that the maximum distance in the US is 1. This is purely for convenience in the calibration procedure: with large numbers for d_{jr} estimates for δ_1, δ_2 are very small in magnitude, which can lead to difficulties in Matlab. Before this normalization the largest distance between two continental commuting zone centroids is 2827.4 miles.

Consider the spatial law of motion, equation 10 from the paper, here re-stated for convenience:

$$L_{rt}^h = \sum_j \rho_{jrt}^h L_{jt-1}^h$$

where ρ^h_{jr} is given in equation 9. Also note:

$$\frac{\rho^h_{jj}}{\rho^h_{jr}} = \frac{exp\left(\frac{1}{\kappa}\left(E^h\left[U_j\right]+A_j\right)\right)}{exp\left(\frac{1}{\kappa}\left(E^h\left[U_r\right]-\tau-\delta_1 d_{jr}-\delta_2 d^2_{jr}+A_r\right)\right)}.$$

As can be seen the number of stayers relative to movers increases monotonically in τ for fixed A_l. Note that $E^h\left[U_j\right]$ is independent of moving cost parameters given that we always match local labor supply exactly. To calibrate $(\kappa,\tau,\delta_1,\delta_2)$ we then guess $(\kappa,\delta_1,\delta_2)$, select τ to ensure the model matches an aggregate interstate migration rate of 0.3 exactly, compute commuting zone flows and aggregate them to the state level and across skill types. Next we compute the state-to-state equivalent of ρ^h_{jr}, aggregated across skill types, and denote it ρ^{MODEL}_{ij1940}. Let ρ^{DATA}_{ij1940} denote the same object computed in the Census Data. Then we evaluate the following objective:

$$\Lambda = \sum_i \sum_{j\neq i} L_{j,1940}\left(\log\rho^{DATA}_{ij1940} - \log\rho^{MODEL}_{ij1940}(\kappa,\delta_1,\delta_2)\right)^2$$

The objective function Λ is a destination j population weighted sum of the percentage difference between model and data of the inflows from all states. We exlude stayers, whose aggregate number we match exactly by adjusting τ for any guess of $(\kappa,\delta_1,\delta_2)$. We evaluate this objective function over a large grid of $(\kappa,\delta_1,\delta_2)$ and find that it is well behaved and exhibits a unique minimum.

Note that for any guess of $(\kappa,\delta_1,\delta_2)$ we re-solve for skilled employment shares in 1910 and the amenity vector in 1940 so as to match skilled employment in 1940 as well as total populations in 1910 and 1940 exactly.

The algorithm to calibrate $\kappa,\tau,\delta_1,\delta_2$ then proceeds as follows:

1. Guess κ,δ_1,δ_2.
2. Compute A_j and the skill shares in 1910 L^h_{r1910} so as to exactly match 1910 and 1940 population shares and 1940 skill shares.
3. Using A_l and the 1910 skill shares compute the implied ρ^h_{jj} in the model, aggregate to the state level and compute total lifetime interstate migration. Adjust τ so as to match exactly and lifetime interstate migration rate of 0.3. We see from the equation of $\frac{\rho^h_{jj}}{\rho^h_{jr}}$ that stayers relative to movers are monotone in τ holding A_l and L^h_r fixed.
4. With the new τ we move back to step 2. Iterate between 2. and 3. until convergence is achieved which yields A_l, L^h_l and τso as to match inter commuting zone net migration, skill shares in 1940 and the interstate migration rate between 1910 and 1940 exactly.

5. With this in hand we compute the following objective function on the state level:

$$\Lambda = \sum_i \sum_{j \neq i} L_{j1940} \left(\log \rho_{ij1940}^{DATA} - \log \rho_{ij1940}^{MODEL}(\kappa, \delta_1, \delta_2) \right)^2$$

Note that each objective function is associated with a unique τ which yields the exact interstate migration rate of 0.3 for the current guess of $(\kappa, \delta_1, \delta_2)$.

6. We evaluate this objective function over a wide grid for $(\kappa, \delta_1, \delta_2)$ and find that it is well behaved and in particular exhibits a unique minimum.

We obtain the following parameter estimates:

Specification	δ_1	δ_2	κ	τ
Value	8.44	-6.39	0.42	1.63

Table 16: Moving Cost Estimates

B.8 Calibrating the preference parameters ν, γ and ϕ

The model implies that the elasticity of substitution is given by

$$\sigma_i^{ES}\left(\vartheta_{A,i}\right) = 1 - \eta - \frac{\left(\gamma - \eta\left(1-\phi\right)\right)\left(1-\phi\right)}{1-\vartheta_{A,i}} + \frac{\left[\gamma + \phi\eta\right]\phi}{\vartheta_{A,i}},$$

where $\vartheta_{A,i}$ is expenditure share. We aim to match some aggregate moment of the distribution of substitution elasticities σ_i^{ES}. It is analytically convenient to target the average substitution elasticity of the young generation. Recall that the expenditure share of the young generation with skill h in region $r$is given by

$$\vartheta_{rAt}^{h,Y} = \phi + \tilde{\nu}\left(\frac{p_A}{p_M}\right)^{\gamma}\left(E_r^{h,Y}\right)^{-\eta},$$

where $\tilde{\nu} = \nu \frac{\Gamma\left(1-\frac{\eta}{\zeta}\right)}{\Gamma\left(1-\frac{1}{\zeta}\right)^{1-\eta}}$ is a constant and $E_{rt}^{h,Y} = \psi\left(r_{t+1}\right)\Gamma\left(1-\frac{1}{\zeta}\right)\Theta_{r,t}^{h}$. Hence, the average expenditure share of the young is

$$\vartheta_{At}^{Y} = \phi + \tilde{\nu}\left(\frac{p_A}{p_M}\right)^{\gamma}\overline{E}_{t,Y}^{-\eta},$$

where $\overline{E}_{tY}^{-\eta} = \sum_{r,h}\left(E_{rt}^{h,Y}\right)^{-\eta}\lambda_{rt}^{h}L_{rt}$. Hence,

$$\sigma_i^{ES}\left(\vartheta_{At}^{Y}\right) = 1 - \eta - \frac{\left(\gamma - \eta\left(1-\phi\right)\right)\left(1-\phi\right)}{1-\phi-\tilde{\nu}\left(\frac{p_A}{p_M}\right)^{\gamma}\overline{E}_{tY}^{-\eta}} + \frac{\left[\gamma+\phi\eta\right]\phi}{\phi+\tilde{\nu}\left(\frac{p_A}{p_M}\right)^{\gamma}\overline{E}_{tY}^{-\eta}}.$$

We calibrate γ to ensure that $\sigma_i^{ES}\left(\vartheta_{At}^{Y}\right) = 0.7$. Note that for given ϕ and γ, $\sigma_i^{ES}\left(\vartheta_{A,t}^{Y}\right)$ only depends on γ (as $\overline{E}_{t,Y}^{-\eta}$ can be directly calculated from the data).

We calibrate ϕ for the model to be consistent with the equilibrium agricultural value added shares $\vartheta_{A,t}^{VA}$, which are given by

$$\vartheta_{A,t}^{VA} = \phi + \tilde{\nu}\left(\frac{p_A}{p_M}\right)^{\gamma} \frac{\sum_{r,h}\left(\left(E_{rt}^{h,Y}\right)^{1-\eta}\lambda_{rt}^{h}L_{rt} + \left(E_{rt}^{h,O}\right)^{1-\eta}\lambda_{rt-1}^{h}L_{rt-1}\right)}{PY_t}. \tag{65}$$

Internal consistency requires to match ϑ_{At}^{VA} at each point in time. Hence, we allow for ϕ to be time-specific by solving (65). As (65) depends on ν, we chose ν to minimize the extent to which ϕ_t vary across time.

B.9 Calibrating the endogenous amenity parameters

In our counterfactuals we allow amenities to be a function of local population L_r. In particular we assume that the overall amenity in location r can be written as:

$$A_{rt} = \rho L_{rt}^{\theta} + \bar{A}_{rt} \tag{66}$$

where $\bar{A}_r$ is denotes the exogenous component of the amenities in location r and $\rho, \theta \in \mathbb{R}$ flexibly parameterize the endogenous part. It is natural to think about a congestion effect of local population on local amenities as in Ahlfeldt et al. (2015) and Allen and Arkolakis (2014) and our estimates will imply this with $\rho < 0$ and $\theta > 0$. Note that in the calibration total population in the United States is normalized to unity in each cross-section.

Similar to Ahlfeldt et al. (2015), notice that for each set of parameters of the model, there is a unique vector $\{A_{rt}\}_{rt}$ for each $t = 1910, 1940, 1970, 2000$ such that the observed vectors of wages, local populations and employment shares in each cross-section are an equilibrium of the model. This result is independent of (ρ, θ). In order to estimate (ρ, θ) we first fit the model to the data and obtain the full panel of amenities $\{A_{rt}\}_{rt}$ for $t = 1910, 1940, 1970, 2000$. Then we re-write (66) in differences:

$$\Delta A_{rt} = \rho \Delta L_{rt}^{\theta} + \Delta \bar{A}_{rt} \tag{67}$$

In equation 67 $\{\Delta A_{rt}, \Delta L_{rt}\}_{rt}$ are observed independently of (ρ, θ). With an instrument φ_{rt} for ΔL_{rt}^{θ} that induces variation in local population inflows between t and $t-1$ orthogonal to $\Delta \bar{A}_{rt}$ the two parameters (ρ, θ) could be obtained by minimizing the sample equivalent of the following moment condition:

$$E\left[\varphi_{rt}\Delta \bar{A}_{rt}\right] = E\left[\varphi_{rt}\left(\Delta A_{rt} - \rho \Delta L_{rt}^{\theta}\right)\right] = 0 \tag{68}$$

To construct such an instrument we employ a Model Implied Optimal IV (MOIV) general equilibrium estimator as recently proposed by Adao et al. (2017). We simulate the model for a given guess of (ρ, θ), setting $A_{rt} = \bar{A}_{r1910}$ for $t = 1910, 1940, 1970, 2000$, but letting the aggregate sectoral TFP series, $\{Z_{st}\}_{st}$, and the vectors of regional technology shifters $\{Q_{rAt}, Q_{rNAt}\}_{rt}$ vary as calibrated. This generates a series of population changes $\Delta L_r(\Delta Z_{st}, \Delta Q_{rst}, \Delta \bar{A}_{rt} = 0 \mid \theta, \rho)$ conditional on (ρ, θ) which is by construction orthogonal to changes in $\bar{A}_{rt}$. To make this hold in the model we search over a grid for (ρ, θ) to find the values that minimize the sample equivalent of the moment condition in equation (68):

$$\Lambda = \sum_{t=1940,1970,2000} \left[\frac{1}{N} \sum_{r} \Delta L_{rt}(\Delta Q_{rst}, \Delta \bar{A}_{rt} = 0, \Delta Z_{st} \mid \theta, \rho) \Delta \bar{A}_{rt} \right]^2 .$$

Note that for every new guess $\theta = \theta^0$ and $\rho = \rho^0$ we need to recompute the counterfactual panel of local population changes $\Delta L_r(\Delta Q_{rst}, \Delta \bar{A}_{r,t} = 0, \Delta Z_{st} \mid \theta = \theta^0, \rho = \rho^0)$. The objective Λ obtains a unique minimum at $\rho = -0.251$ and $\theta = 2.575$.

The procedure is then as follows:

1. Guess $\theta = \theta^0$ and $\rho = \rho^0$ and compute $\bar{A}_r$ such that $A_r = \bar{A}_r + \rho L_r^{\theta}$, where A_r is the calibrated amenity.

2. Simulate the model forward with $\theta = \theta^0$ and $\rho = \rho^0$, with all other parameters as calibrated and with all fundamentals varying as calibrated except for fixing $\bar{A}_{rt}$ and $\bar{A}_{r1880}$. In particular we compute

   $$\Delta L_r(\Delta Q_{rst}, \Delta \bar{A}_{rt} = 0, \Delta Z_{st} \mid \theta = \theta^0, \rho = \rho^0)$$

   which is the variation induced in local net migration rates induced by changes in the other fundamentals and by the full structure of the model for our current parameter guesses $\theta = \theta^0$ and $\rho = \rho^0$.

3. We then search for the θ and ρ that minimize the sample analog of the squared sum of following moment condition for each year:

   $$E\left[\Delta L_{rt}(\Delta Q_{rst}, \Delta \bar{A}_{rt} = 0, \Delta Z_{st} \mid \theta, \rho) \Delta \bar{A}_{rt}\right] = 0$$

   In particular then the minimize the following objective:

   $$\Lambda = \sum_{t=1940,1970,2000} \left[\frac{1}{N} \sum_{r} \Delta L_{rt}(\Delta Q_{rst}, \Delta \bar{A}_{rt} = 0, \Delta Z_{st} \mid \theta, \rho) \Delta \bar{A}_{rt} \right]^2$$

   Note that for every new guess $\theta = \theta^0$ and $\rho = \rho^0$ we need to recompute the counterfactual panel of local population changes $\Delta L_r(\Delta Q_{rst}, \Delta \bar{A}_{rt} = 0, \Delta Z_{st} \mid \theta = \theta^0, \rho = \rho^0)$.

An extensive grid search reveals this objective to be well-behaved with a unique minimum attained at $\rho = -0.251$ and $\theta = 2.575$.

C Data Sources and Construction

C.1 Detailed data sources for the panel

In

C.2 Wages

For all years we construct manufacturing wages as total payroll of all manufacturing establishments in a given county divided by the total number of manufacturing workers. The data source for 1880 is the census of manufacturing. Unfortunately the Census of Manufacturing in 1910 does not contain county level wage information and we use 1920 Census of Manufacturing data instead. Since we calibrated our model so as to match GDP per capita growth exactly, all that matters for us as far as the wage data is concerned is the dispersion of wages across counties, not their absolute level.

For the years 1940-2000 we use data obtained from County and City Data Books for 1947, 1977 and 2000 published by the US Census Bureau. Note that the 1947 edition report manufacturing payrolls for 1939 while the 1977 edition reports them for 1972, which we use in conjunction with manufacturing employment for those years to construct our wage measure for the years 1940 and 1970 respectively. Note that for all these years we use manufacturing employment from the respective data for consistency reasons.

It is important to note that we first collect total payroll and total manufacturing employment on the county level for the respective year, then re-aggregate both to commuting zones employing area weights where necessary, before constructing average manufacturing wages for each commuting zones as the ratio of the total manufacturing payroll divided by total manufacturing employment for all counties (or parts thereof) that constitute a given commuting zone.

To the best of our knowledge, for the time prior to 1940 there is no other data source that would allow for the construction of consistent county level wage measures for the entire US territory. For 1940 however, the recently released full version of the decennial population census reports individual wages and industries and allows the construction of average manufacturing wages by county. Figure 17 in the appendix scatters the manufacturing wage measures computed for each county from the two data sources against one another. Their worker weighted correlation is 0.77[55], which we take as a reassuring robustness check for our census of manufacturing based wage measure. Note that decennial population Censuses after 1940 do not include a full set of county identifiers and do not allow for the construction of county level averages, precluding a comparison in later years.

Figure 16 shows demeaned log wages for the five cross-sections used in the baseline calibration. As can be seen there is generally wage convergence with the war year constituting an outlier in that regard.

[55] Unweighted correlation 0.4.

Variable	Year	Source
Manufacturing Wages and Manufacturing Employment	1880	U.S. Bureau of the Census. Tenth Decennial Census of the United States, 1880. Volume II. Report on the Manufactures of the United States at the Tenth Census. (Washington, D.C.: Government Printing Office, 1883). Table IV
	1910[a]	U.S. Bureau of the Census. Fourteenth Decennial Census of the United States, 1920. Volume VIII. Manufactures: 1919. "General Report & Analytical Tables." (Washington, D.C.: Government Printing Office, 1923). Table 51.
	1940	Sources: U.S. Bureau of the Census. County Data Book. (Washington, DC: G.P.O, 1947). U.S. Bureau of the Census [1947], Tables 1 & 3.
	1970	United States Department of Commerce. Bureau of the Census. County and City Data Book, 1977. The manufactures statistics presented here are based on results of the 1972 Census of Manufactures, conducted by the Bureau of the Census as part of the 1972 economic censuses. Final data are presented in Census of Manufactures, 1972, Vol. I, Subject and Special Statistics; Vol. II, Industry Statistics; and Vol. III, Area Statistics.
	2000	Sources: U.S. Bureau of the Census. County and City Data Book: 2000 (Washington, DC: G.P.O, 2001). Table B, and additional variables included in the machine-readable data set.
(Non-)Agricultural Employment	1880	U.S. Bureau of the Census. 1880 Decennial Census. Obtained from the IPUMS repository of the University of Minnesota.
	1910	U.S. Bureau of the Census. 1910 Decennial Census. Obtained from the IPUMS repository of the University of Minnesota.
	1940	U.S. Bureau of the Census. 1940 Decennial Census. Obtained from the IPUMS repository of the University of Minnesota.
	1970[b]	U.S. Bureau of the Census.1970 Decennial Census. Tabulated by and obtained from National Historical Geographical Information System (NHGIS) website repository.
	2000[c]	Sources: U.S. Bureau of the Census. County and City Data Book: 2000 (Washington, DC: G.P.O, 2001). Table B, and additional variables included in the machine-readable data set. Tabulated by and obtained from National Historical Geographical Information System (NHGIS) website repository.

Table 17: Detailed Data Sources

[a] County level wages for 1910 are unavailable so we use the 1920 ones. Resultantly, we also use the manufacturing head count from 1929.
[b] Micro Census not available on county level.
[c] Micro Census not available on county level.

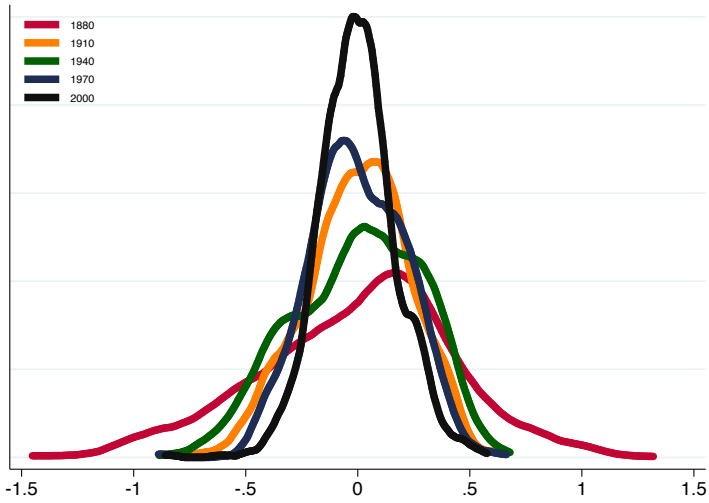


Figure 16: Normalized Wages in each year.

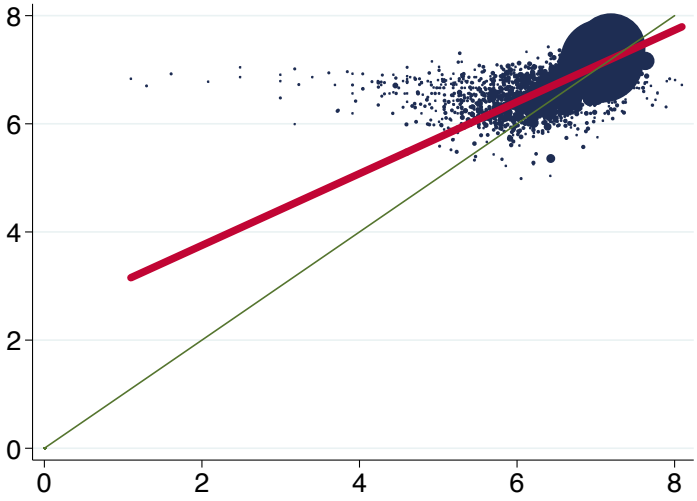


Figure 17: Log Average Manufacturing Wages in 1940.

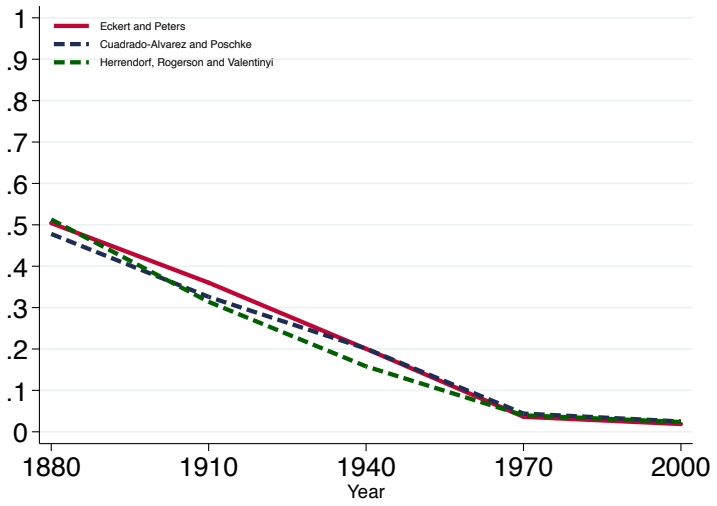


Figure 18: Aggregate Agricultural Employment Shares

C.3 Sectoral Employment

While the decennial population Censuses prior to 1940 do not report wages for individuals, they do report their sector of employment, which allows for the construction of county level employment counts. We define agricultural employment as people listed as working in Agriculture, Forestry and Fisheries[56] and non-agricultural workers as all others with reported industries. Workers without industry affiliation are discarded. We do not drop women, since for the years beyond 1940 only county level aggregates are available which do not allow to distinguish workers by gender.

For 1970 we obtain employment counts from the 1970 people Census tabulated on county level from the National Historical Geographical Information System (NHGIS) website. For 2000 we likewise obtain tabulated data from NHGIS. For 2000 only, agriculture also includes mining and hunting.

In figure 18 we construct a series of the US aggregate agricultural employment shares from our county level data and compare it to those from other aggregate sources. The series are very similar, which furthers our confidence in our data sources.

C.4 Manufacturing Premium and Skill Premium

We use the decennial micro Census from 1940 onwards to construct a skill premium based on different definitions of skill. We experiment with three cutoffs for skilled employment based on the standardized education measure provided by IPUMS (see Ruggles et al. (2015)). 19 shows the average wage for people in the United States in the respective year who have either high school or higher, some college or higher or a completed college education.

[56] 1950 Industry Classification Codes 105, 116 and 126.

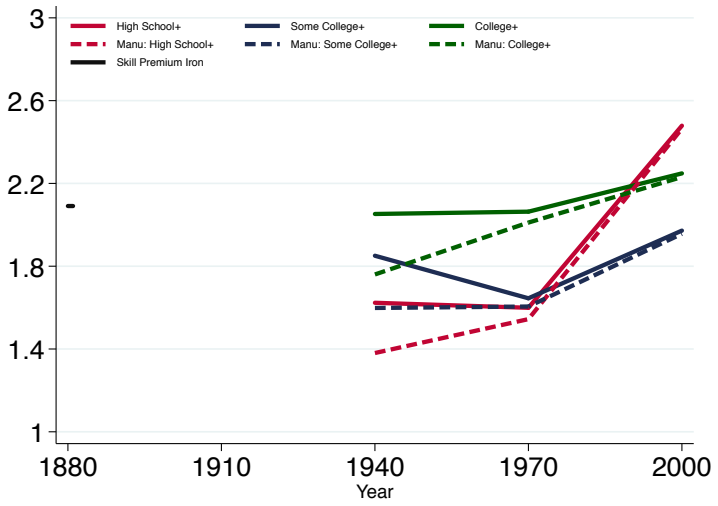


Figure 19: Skill Premium for different skill measures

The micro Census does not provide wages before 1940. The census of manufacturers for 1880 provides a daily wage for skilled and unskilled iron workers by county. Using this we can construct a measure of the skill premium in the iron industry in 1880. To be more comparable we also compute the skill premium for 1940 to 2000 within the manufacturing industry. In the baseline version of the model we calibrate the absolute advantage of skilled over unskilled workers so as to match this premium in 1940 with high school as the skill cutoff.

The moment used in calibrating the comparative advantage of workers is the manufacturing premium in equation 64. As for the skill premium there are different education cutoffs that can be used to define skill. Contrary to the skill premium which cutoff is chosen does only make a minor difference as can be seen in figure 20. In the baseline we target the manufacturing premium in 1940 based on the high school cutoff.

C.5 Lifetime interstate Migration Data

In the model agents move once in their life time in order to choose a labor market. We use US decennial Census Micro data obtained from IPUMS to construct an interstate migration flow matrix for the cross-sections 1880, 1910, 1940, 1970 and 2000. In a given Census we choose men between 26 and 50 years of age, and compute the number of them living in a different state from their state of birth. We can also compute the total number of men of this age born in each state, which then allows us to construct a state level fraction of "stayers", i.e.. men born in a given state and still living their during their working life. We plot this lifetime interstate migration measure from our data and from Molloy, Smith and Wozniak. They compute it as the total fraction of people living in a different state than they were born in in a given year. Since children are more likely to live where they were born, and old people often return to where they were born for retirement this explains the lower rate. Since we are interested in the location and

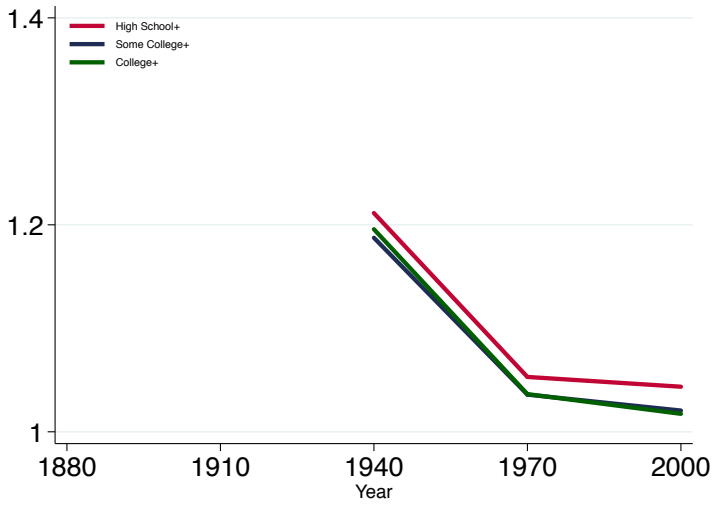


Figure 20: Manufacturing Premium for different skill measures

migration decisions of workers we prefer our measure.

Where we use them state level wages are constructed by aggregating county level total manufacturing payrolls and workers to the state level, before computing average wages as their ratio.

C.6 Distance Data

Interstate differences are computed as the straight line distances in miles between states' centroids. To obtain distances between commuting zones we averaged the longitude and latitude of centroids of counties contained in a given commuting zone and then computed straight line distances between these approximated commuting zone centroids.

C.7 Construction of 1880 Wage Cross-section

For 1880 we use the county level data tabulated in the Report on the Manufactures of the United States at the Tenth Census obtained from ICPSR. We calculate total manufacturing employment in each country from this data by counting together the total number of men, women and children employed in the manufacturing sector. With this in hand we use the total annual manufacturing payroll for each county to obtain an estimate of the average manufacturing wage for each county in the United States for 1880.

Working with state or even county level data from the 19th Century can be problematic. There are a twelve states which in 1880 were not part of the Union yet (had not obtained statehood), we list them here and give the year of the accession to the Union: 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) and Hawaii (1959).

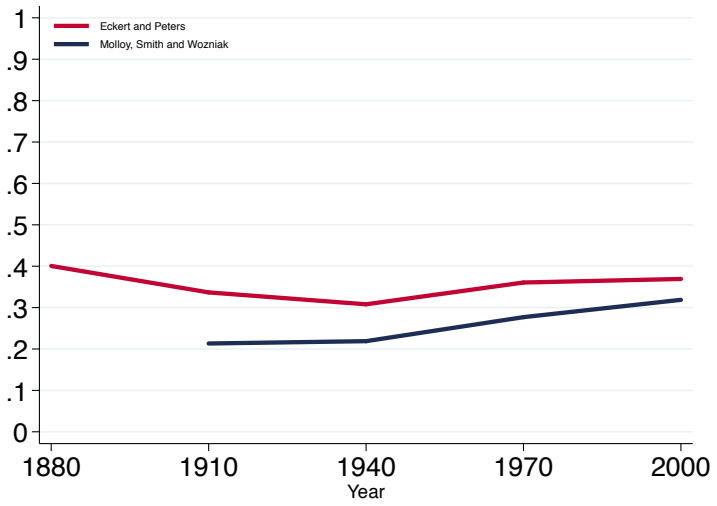


Figure 21: Lifetime interstate Migration Rate

The 1880 Census does report data for counties in all states, even those that had not yet officially obtained statehood in 1880, with two exceptions: Oklahoma and Hawaii. Since we focus our analysis on the continental United States we exclude data from Hawaii and Alaska from our entire data set. The missing data for Oklahoma is likely due to the fact that in 1880 Oklahoma remains the only state in the territory of what is today known as the United States that had not been settled. During the 19th Century thousands of Native Americans were re-settled into a territory dubbed Indian Territory within present day Oklahoma, with the remaining part of which was called Oklahoma Territory.[57] By 1890 over 30 Native American nations and tribes had been concentrated on land within Indian Territory. White settlers also more and more pushed into the territory and with the Dawes Act of 1887 were granted the right to purchase certain lands within in the Territory. Major land runs followed and Oklahoma was made a state in 1907. The 1890 Census includes data on counties within the Oklahoma Territory of present day Oklahoma, in particular total manufacturing payroll by county and manufacturing employment by county. In addition we have total population counts for both Indian territories and Oklahoma Territory in 1890. We also have the actual employment shares for agriculture, manufacturing and services in 1910 once Oklahoma obtained statehood, its entire territory is structured into counties and the Census starts to report on it regularly. With all this information in hand we want to reconstruct an estimated county level cross-section of data for 1880 for present day Oklahoma. We proceed as follows

1. We take the 1910 map of Oklahoma in which the entire territory is structured into counties as the base map.

2. We assign each county on the 1910 map to either Indian Territory or Oklahoma Territory in 1880.

[57] States, or in this case parts of states, that had been settled but not yet joined the Union were generally referred to as territories. In the 1880 data the territory of North and South Dakota appears simply as Dakota Territory.

3. We deflate the 1890 total manufacturing payroll for Oklahoma territory based on the wage data in 1890 and 1880 we have for the neighboring state Arkansas. Then we deflate the total 1890 manufacturing workers count by the growth of manufacturing employment in Arkansas between 1880 and 1890. Next we distributed the resulting estimate of manufacturing workers in 1880 Oklahoma Territory proportionally to 1910 manufacturing employment across 1910 counties. Likewise we distribute the estimated total manufacturing payroll for 1880 proportionally to the 1910 distributed of the manufacturing payroll. This allows us to compute an estimated per capita wage of manufacturing for Oklahoma counties in 1880. Lastly, we assume that all counties in Indian territory had a manufacturing wage which is equal to the minimum of manufacturing wages in the Oklahoma territory.

4. To get numbers for agricultural (A) and non-agricultural (NA) employment for each county in 1880 we look at 1910 employment in A and NA. We then look at Arkansas' population growth between 1880 an 1890 and its ratio of population to employment in 1880. We then use our numbers for total populations of the two territories in 1890, to first deflate by the population growth rate to obtain an estimate of the 1880 population and then multiply by the Arkansas employment to population ratio to obtain an estimate of employment for both territories. Then we use the 1910 employment , scale up agricultural employment in line with Arkansas between 1880 and 1910 and then distribute our estimate of 1880 workers proportionally across sectors and counties according to these numbers.

This leaves us with an estimate of sectoral employment and payroll for Oklahoma in 1880 and allows us to reconstruct the full 1880 map of the continental United States.