Spatial Structural Change and Agricultural Productivity*

Fabian Eckert[†] Michael Peters[‡]

September 2017

Abstract

Standard models of structural change predict that the share of agricultural value added and agricultural employment are equalized. In the data they are not. While both decline as the economy develops, value added per worker in agriculture is substantially lower than in non-agriculture. Moreover, this agricultural productivity gap is remarkably persistent despite the large reallocation of production factors across sectors. In this paper, we argue that this sectoral productivity gap might to a large extent be a spatial gap. Using a novel dataset for more than 700 US commuting zones between 1880 and 2000, we document that agricultural employment shares are strongly negatively correlated with average earnings and uncorrelated with subsequent net population outflows. These facts are consistent of substantive frictions to spatial mobility, which prevent the spatial equalization of marginal products. To quantify the strength of this mechanism, we construct a novel theory of spatial structural change by embedding an economic geography model in a dynamic, neoclassical model of the structural transformation. We show that spatial frictions can account for more than 50% of the observed productivity gap. This implies that the direct productivity gains from reallocating workers across sectors are modest.

*We thank seminar participants at Stanford, Penn State, 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 and Doug Gollin 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

Standard models of the process of structural change imply that sectoral employment shares and sectoral value added shares are equalized. In the data they are not. Not only do agricultural employment shares consistently exceed the share of agricultural value added, but this "agricultural productivity gap" stays stubbornly high as countries undergo the structural transformation. A case in the point is the historical experience of the US. While the employment shares declined from 60% to essentially nil since 1850, value added per worker in the non-agricultural sector is about twice as high as its counterpart in the agricultural sector and varies little over time. Taken at face value, this implies that over the last 120 years of US growth, the marginal product of agricultural workers is about half as high as the marginal product in non-agriculture - despite the fact that more than 50% of the workforce reallocates.

In this paper, we argue that the spatial allocation of economic activity goes a long way to quantitatively explaining this pattern. Our argument is simple: If regions differ in their comparative advantage, individuals need to relocate as the economy develops and aggregate spendings shifts away from agriculture. If mobility is costly, the process of structural change puts downward pressure on wages in rural regions, which specialize in agriculture. Both the size and the persistence of these wage differences across locations depend on the speed of spatial reallocation. If spatial reallocation is subject to frictions, wage gaps emerge. Moreover, if the process of structural change evolves slowly, such wage gaps persist. In the aggregate, this spatial productivity gap manifests itself as an agricultural productivity gap, even though the marginal product of labor might be equalized across sectors within a location at each point in time.

Two empirical regularities from the US experience are suggestive that this mechanism might be important. First of all, using a novel dataset on historical employment patters and manufacturing earnings across all US counties starting in 1880, we document a sizable spatial wag gap across locations. Importantly, there is a strong negative correlation with the share of agricultural employment and this correlation remains very stable between 1880 and 2000. Hence, regions specializing in agriculture are and remain low wage regions. Our second fact suggests why such wage differentials are not arbitraged away through spatial mobility: empirically, the extent of net migration is only weakly correlated with agricultural specialization. While gross flows are substantial, the agricultural employment share is not the dominant predictor of the direction of net flows. We for example show that the reallocation of workers from high to low agricultural places during the 20th century has essentially zero explanatory power for the decline in the aggregate agricultural employment share. This implies that the entirety of the structural transformation is a within-region phenomenon and there might be very limited arbitrage (and hence aggregate reallocation gains) at the relevant, i.e. spatial, margin.

To quantify the importance of this mechanism, we combine an economic geography model with intranational trade and labor mobility and an otherwise standard, neoclassical model of structural change. As far as the process of structural change is concerned, we follow the macroeconomic literature and allow for demand side forces (i.e. non-homothetic preferences) and supply side forces (i.e. non-balanced technological progress across sectors). To generate a need for spatial reallocation, we assume that regions differ in their sectoral comparative advantage and workers can reallocate spatially subject to moving costs. In order to expand, regions need to pay higher wages to attract individuals. Because the process of structural change requires non-agricultural intensive regions to grow, marginal products across space are not equalized and agricultural value added shares shrink relative to the share of agricultural employment.

Moving costs are of course not the only plausible reason for the persistence of spatial wage gaps. Most importantly, it might be that differences in the average products are uninformative about a dispersion in marginal products. If workers for example select on unobserved skills and skilled workers have a comparative advantage in non-agricultural regions, the empirically observed persistent productivity gap might simply be a reflection of sorting behavior, whereby skilled individuals locate in non-agricultural regions. If that was the case, average value added per worker might very well systematically differ across sectors despite the fact that the marginal product of labor is equalized. Secondly, as for example stressed by Lagakos and Waugh (2011), sectoral specialization itself might be a reason why output per worker might be low - if individuals are heterogeneous in the skills they can provide to different industries, the quality of the marginal worker declines, the higher is the employment share of the industry. Finally, to the extent that rural regions provide other utility-relevant amenities, the agricultural productivity gap might simply be a compensating differentials gap. In our theory, we take these aspects explicitly into account and we show that (and why) one would understate the spatial gap if one were to abstract from these features.

Two modeling choices are crucial to make the analysis tractable, while still quantitatively meaningful. First of all, we follow the work of Boppart (2014) and assume that preferences are in the class of *price independent generalized linear (PIGL)* preferences. This preference specification has much more flexibility in the strength of income effects compared to the widely-used Stone-Geary specification. This is important when trying to take the model to the long-run data. In the Stone-Geary case, income effects vanish asymptotically (see e.g. Comin et al. (2015) or Alder et al. (2017)). This makes it difficult to quantitatively explain the observed decline in agricultural employment. The PIGL specification does a much better job to match the long-run data. At the same time, it is still the case that the PIGL preference specification has convenient aggregation properties. While the preferences are not in the Gorman class and hence do not permit a representative household, we show that these preferences together with the commonly-used Frechet-distribution of individual skill heterogeneity delivers tractable closed form solutions for the main objects of interests. Secondly, we frame our analysis in terms of an overlapping-generation model. This structure is key to allow for both individual savings (and hence capital accumulation) and costly spatial mobility. In particular, we show that individuals are forward looking in terms of their savings behavior but that their spatial choice problems reduces to a static problem. Hence, we do not have to keep track of individuals' expectations about the entire distribution of future wages in different locations - the aggregate interest rate is sufficient.

We apply our theory to the aggregate and regional pattern of the process of development of the US from 1880 to 2000. To do so we combine standard macroeconomic time-series data on the evolution of

GDP pc and relative prices with detailed spatial data (at the level of more than 700 commuting zone) on earnings, employment shares and employment. We show that the calibrated model can rationalize about 60% of the observed agricultural productivity gap without any frictions operating at the sectoral level. We then ask to what extent moving costs are the fundamental cause underlying this productivity gap. To do so, we analyze a counterfactual, where we assume no costs of spatial reallocation. While this would naturally increase the extent of gross workers flows, the implied agricultural productivity gap would - surprisingly - not be substantially different. The reason is that our model - as estimated from the data - implies that workers move for a variety of reasons. While higher wages are one component, regional amenities and idiosyncratic locational preferences also affect moving flows. If spatial mobility was free, individuals would move more for all of these reasons. As the latter two are not correlated with agricultural productivity, the relationship between agricultural specialization and net outflows would not markedly change.

Related Literature Our paper builds heavily on the macroeconomic literature on structural change and the recent literature on models of economic geography. The literature on the process of structural change has almost exclusively focused on the time series implications. 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. Later the same facts were documented across developed countries by Herrendorf et al. (2014).

As an explanation of these aggregate trends, two mechanism have been proposed. First of all, there are models of non-homothetic demand, where non-agricultural goods are income elastic. Early examples of this line of work are Kongsamut et al. (2001) and Gollin et al. (2002), who assume that subsistence requirements imply a low income elasticity of agricultural demand. Recently, Boppart (2014) and Comin et al. (2015) consider alternative preference structures. While Comin et al. (2015) proposes a non-homothetic CES demand system, Boppart (2014) introduces the PIGL demand structure mentioned above. In this paper, we follow Boppart (2014) in his choice of preference specification. This is mostly for analytical convenience, in particular its tractable aggregation properties.

An alternative supply-side explanation for the secular reallocation of resources across sectors is based on unbalanced technological progress or capital deepening. Originating with Baumol (1967), this mechanism has been formalized in Ngai and Pissarides (2007) and Acemoglu and Guerrieri (2008). Herrendorf et al. (2013) and Alvarez-Cuadrado and Long (2011) are recent example of empirically oriented papers, trying to distinguish these explanations. In our model, we allow for both unbalanced technological progress and non-homothetic demand.

We combine this strand of the literature with the recent literature on quantitative economic geography models following Allen and Arkolakis (2014). This literature is mostly static in nature and focuses on the spatial reallocation of workers across heterogeneous locations (see e.g the recent survey in Redding and Rossi-Hansberg (2016)). 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)) or the productivity effects of agglomeration economies (Ahlfeldt et al. (2015)). Bryan and Morten (2015) also stress the importance of moving costs on wage differences across space. In contrast to us, they only consider a static environment and do not focus on the structural transformation or, more generally, sectoral employment patterns across space.

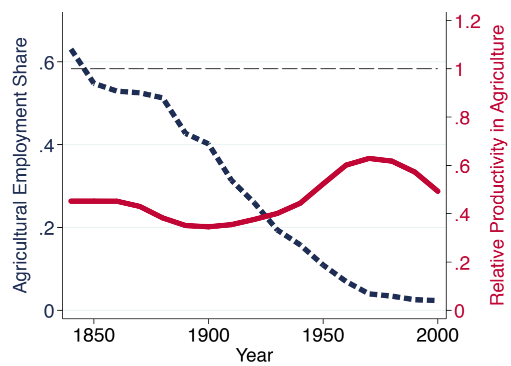
There are few papers that embed such models with spatial reallocation into dynamic macroeconomic environments. An early contribution is Caselli and Coleman II (2001), who - using a two region model with endogenous skill acquisition - argue that spatial mobility was an important by-product of the process of structural change in the US. Michaels et al. (2012) also study the relationship between structural change and spatial mobility but do not study the implications on agricultural productivity. Recent papers are Desmet et al. (2015), Nagy (2016) and Desmet and Rossi-Hansberg (2014). While Desmet and Rossi-Hansberg (2014) are concerned with the latter aspects of the structural transformation (between manufacturing and service employment) and Nagy (2016) studies the process of city formation in the time-period we are interested in (i.e. the US in the 19th century), none of them is concerned with agricultural productivity gap, the main focus of our work.

This productivity gap, is also the object of interest of a sizable empirical literature. Gollin et al. (2013) for example measure this agricultural productivity gap for a large cross-sections of countries using micro-data. The find results, which are comparable to the aggregate numbers in the US cited above, i.e. a relative difference of a factor of 2. Lagakos and Waugh (2011) argue that sectoral selection might be an important reason for differences in physical productivity across sectors. As we will show explicitly below, such selection effect have no consequences for the agricultural productivity gap as measured by value added. Similarly, there are is a set of paper about the importance of spatial wage gaps. Young (2013) argues that the observed wage differences across space are consistent individuals selecting on unobserved skills. Bryan et al. (2014) present direct experimental evidence on the existence of spatial wage gaps in Bangladesh. Lagakos et al. (2015) use this experimental evidence within a macroeconomic model of incomplete risk-sharing to gauge the welfare implications. To the best of our knowledge, we are the first to quantify to what extent spatial wage gaps could be the culprit of the agricultural productivity gap.

The remainder of the paper is structured as follows. In Sections 3 and 4 we describe our model and the relationship between spatial frictions and agricultural productivity. Section 5 contains our application to the structural transformation in the US. Section 7 concludes. An Appendix contains the majority of our theoretical proofs and further details and robustness checks for our empirical results.

2 Spatial Structural Change and Agricultural Productivity: Three Empirical Facts

In this Section we provide three important empirical regularities, which suggest why the agricultural productivity gap could be a spatial gap. Consider first Figure 1, where we use aggregate data to display the time series of the agricultural employment share (blue line) and relative agricultural productivity, i.e. value added per worker in agricultural relative to non-agriculture, for the US economy since 1850. While agricultural employment declines sharply, relative agricultural productivity is essentially constant and only half as large as productivity in the non-agricultural sector. This is inconsistent with most macroeconomic models of the structural transformation, where sectoral value added per worker is proportional to the wage, which is equalized across sectors - see e.g. Herrendorf et al. (2014).[1]



Notes: The figure shows the aggregate agricultural employment share (blue line) and value added per worker in agricultural relative to value added per worker in non-agriculture (red line).

Figure 1: The Structural Transformation and Agricultural Productivity in the US

In Figure 2 we report three empirical regularities about the spatial aspects of the structural transformation, which highlight why this agricultural productivity gap could be a spatial gap. In the left panel, we depict the cross-sectional correlation between agricultural employment shares and average earnings since 1880.[2] We see that this correlation is strongly negative in 1880 and that it remains negative for the entirety of the 20th century. The fact that the correlation between average earnings and agricultural employment stays negative despite the fact agricultural employment declines drastically suggests that spatial mobility is *not* the main driver for the process of structural change. The second panel in Figure 1

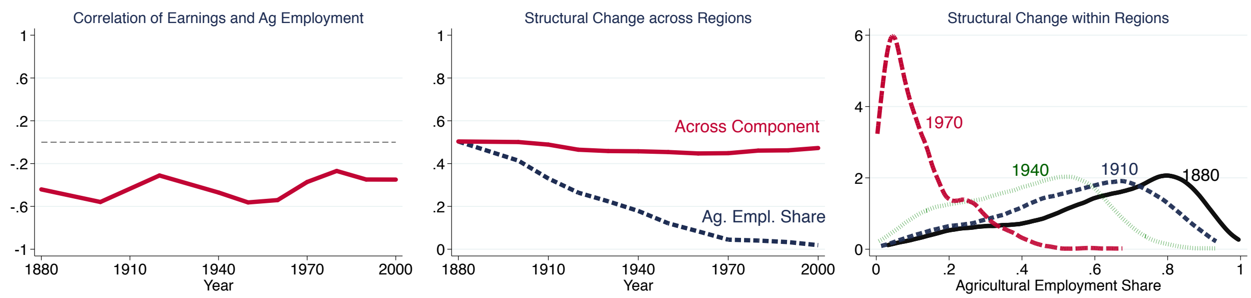
[1] To see this more formally, suppose that production in sector s takes place according to $Y_s = A_s F(k, l)$, where F has constant returns to scale and all markets are competitive. This implies that the capital intensity is equalized. Hence,

$$\frac{VA_s}{L_s} = \frac{P_s A_s F(L_s, K_s)}{L_s} = w \times \frac{F(k,1)}{\frac{\partial F(L_s,K_s)}{\partial L}} = w \times \frac{F(1,k)}{\frac{\partial F(1,k)}{\partial L}},$$

i.e. value added per worker is equalized across sectors.

[2] 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 5 below.

shows in what sense there indeed is an absence of "spatial arbitrage".In particular, we report the implied agricultural employment share, which emerged *solely* from spatial reallocation. More specifically, we conduct a "shift-share"-analysis, by fixing regional agricultural employment shares at their 1880 level and calculate the aggregate agricultural employment share using the population distribution from the data.[3] This is the red line in the middle panel of Figure 1. For comparison we again superimpose the actual agricultural employment share from Figure 1 in blue. It is clearly seen that - in an accounting sense - spatial reallocation accounts for essentially nothing 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. Conversely, most of structural change seems to take place *within* regions. That this is indeed the case is seen in the right panel of Figure 2, where we display the cross-sectional distribution of regional agricultural shares for different years. There is a marked leftwards shift, whereby *all* regions see a decline in agricultural employment. Hence, the structural transformation transforms places and is not merely a process which reallocates production factors across space.



Notes: In the left panel we show the correlation between the agricultural employment share and average earnings across US commuting zones. In the middle panel we show the aggregate agricultural employment share (blue line) and the predicted agricultural share holdings regional agricultural shares at their 1880 level, i.e. $\sum_r s_{A,r,1880} \times l_{r,t}$, where $s_{A,r,t}$ and $l_{r,t}$ are the agricultural employment share and the population share of region r at time t. In the right panel we show the cross-sectional distribution of agricultural employment shares in different years. For a detailed description of the construction of the regional data we refer to Section 5.

Figure 2: Spatial Structural Change: Three Facts

The patterns in Figure 1 are qualitatively consistent with an important role for the spatial allocation of resources in explaining the persistence of the agricultural productivity gap. Regions who specialize in agriculture are places with low wages *on average* and spatial arbitrage is too slow a process for such spatial wage gaps to disappear. Finally, the fact that the vast majority of labor reallocation takes place *within* regions, implies that the sharp decline in aggregate agricultural employment is perfectly consistent with persistent difference in average productivity if marginal products within locations are equalized. In this paper, we argue that the facts displayed in Figure 1 are also quantitatively consistent with the observed aggregate productivity gap displayed in Figure 1. This, of course, requires a structural model, which is where we turn now.

[3] More precisely, we calculate this series as $\sum s_{A,r,1880} \times \frac{L_{r,t}}{L_t}$, where $s_{A,r,1880}$ is the regional agricultural employment share in 1880 and $L_{r.t}$ is number of workers in region r.

3 Theory

In this section we present our theory of spatial structural change. The theory rests on three pillars. We start with an essentially neoclassical model of the structural transformation, where the process of structural change is generated from non-homothetic demand 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 and skilled workers have both an absolute advantage and a comparative advantage in the non-agricultural sector.

3.1 Environment

Technology We consider an economy with two goods, an agricultural good and a non-agricultural good. For simplicity we also sometimes refer to the latter as the manufacturing good. Each good is a CES composite of differentiated regional varieties with a constant elasticity of substitution σ. In particular,

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

where Y_{rs} is the amount of goods in sector s stemming from region $r$and σ is the elasticity of substitution. 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. For expositional simplicity, we suppose that capital shares are identical across sectors. We will allow for differences in our empirical application. It is useful to express productivity A_{rst} as

$$A_{rst} = Z_{st} \times Q_{rst} \text{ with } \sum_{r} Q_{rst}^{\sigma-1} = 1. \tag{2}$$

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 $[Q_{rs}]$ describes the distribution of regional productivity differences, that is the extent to which some regions are more efficient at producing sector s goods compared to other locations. The common components of Q_{rs} across industries captures differences in *absolute* advantage, i.e. some location might be more efficient to produce all goods. Similarly, regional differences in $Q_{rs}/Q_{rs'}$ capture differences in *comparative* advantage. Given the normalization embedded in (2), we also refer to the Q_{rs} as measuring the heterogeneity in productivity across space.

Capital accumulates in the usual way, i.e. according to

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

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 (1). Letting ϕ be the share of the agricultural good in the production of investment goods, the price of the investment good is given by$P_{I,t} = P_{A,t}^{\phi} P_{M,t}^{1-\phi}$. For the remainder we take the investment good to be the numeraire.

Selection, Human Capital and Labor Supply We allow individuals to differ in their human capital. Doing so is important to credibly measure the agricultural productivity gap if individual skills are correlated with sectoral sorting.[4] In particular, suppose that individuals can be of two types - high skilled and low skilled. Their skill type 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

$$F_{z_s^i}^h(z) = e^{-\Psi_s^h z^{-\zeta}}, \tag{3}$$

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. The empirically relevant case is one where skilled individuals have an *absolute* advantage, i.e. $\Psi_s^H > \Psi_s^L$ for all s and a comparative advantage in the manufacturing sector, i.e. $\Psi_{NA}^H / \Psi_A^H > \Psi_{NA}^L / \Psi_A^L$. A convenient parametrization of these assumption is that[5]

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

Hence, q parametrizes the absolute advantage of skilled individuals and μ governs the complementarity between skills and non-agricultural employment.

We assume that - at the aggregate level - a fraction λ of the population is skilled. How people of different skills are distributed across space is of course endogenous and will be determined endogenously from peoples' migration decisions. While individuals know their skills, i.e. $h \in \{L, H\}$ prior to their mobility decision, they only learn the actual realization of their efficiency bundle z^i afterwards. This structure has two convenient properties. First of all, 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 (3). If workers'

[4]Not surprisingly, such sorting behavior is going to be relevant for our application - we find strong evidence that unskilled individuals are overrepresented in the agricultural sector.

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

spatial choice was conditional on z^i, the distribution of skills within a location would no longer be of the Frechet form.

Given these assumptions, we can now characterize the individual earnings and aggregate labor supply in location r. Let λ_{rt} be the endogenous share of skilled individuals working in region r. Total earnings of individual i residing in region r are given by

$$y_r^i = \max\left\{w_{r,A} \times z_A^i, w_{r,NA} \times z_{NA}^i\right\}, \tag{4}$$

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

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

where $\Gamma_\zeta = \Gamma\left(1-\zeta^{-1}\right)$ and $\Gamma(.)$ is the gamma function and

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

Note that Θ_r^h is equalized across sectors and can be directly calculated from the regional wages $\left(w_{r,A}, w_{r,NA}\right)$ but differs across skill-types. This endogenous scalar Θ_r^h 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). In particular, given Θ_r^h, the share of people of skill group h employed in the two industries is given by

$$s_{s,r}^h = \Psi_s^h \times \left(\frac{w_{r,s}}{\Theta_r^h}\right)^\zeta, \tag{6}$$

so that the sectoral labor supply elasticity is governed by ζ. Using (6) is also easy to show that our model incorporates 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

$$\frac{H_{r,s}^h}{L_r \times s_{s,r}^h} = \left(s_{s,r}^h\right)^{-\frac{1}{\zeta}} \left(\Psi_s^h\right)^{\frac{1}{\zeta}},$$

i.e. is decreasing in the sectoral employment share $s_{s,r}^h$ as individual sorting implies that the marginal worker in sector s is worse than the average worker.

Demographics In terms of preferences and demographics, we consider an 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 is structure is convenient because it generates a motive for savings (and hence capital accumulation), while still being sufficiently tractable to allow for spatial mobility.

In our model, individuals have three economic choices to make: (i) they decide how much to save and consume, (ii) they allocate their spending optimally across the two consumption goods and (iii) and they decide where to work and live. In terms of timing, we assume that individuals are born, decide on their preferred location when they are young and then remain in that location for their entire life. Their offsprings are born in the location where the old generation resides. For simplicity, we abstract from endogenous human capital accumulation 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 function of spending an amount e_t at prices $P_t = (P_{A,t}, P_{NA,t})$, life-time utility of individual i after having moved to region r is given by

$$U_r^i = \max_{[e_t, e_{t+1}, s]} \{V(e_t, P_t) + \beta V(e_{t+1}, P_{t+1})\}, \tag{7}$$

subject to

$$e_t + s_t = y_{rt}^i \tag{8}$$

$$e_{t+1} = (1 + r_{t+1}) s_t. \tag{9}$$

Here, y_{rt}^i is individual i's real income in region r (see (4)), s_t denotes the amount of savings and $r_t = R_t - \delta$ is the real interest rate.

Preferences For our model to induce the process of structural change, we have to move away from homothetic preferences. In particular, we require a specification of preferences, where consumers reduce their relative agricultural spending as they grow richer. To do so, we follow Boppart (2014) and assume that individual preferences can be represented by the indirect utility function

$$V(e, P) = \frac{1}{\eta}\left(\frac{e}{p_A^{\phi} p_{NA}^{1-\phi}}\right)^{\eta} - \frac{\nu}{\gamma}\left(\frac{p_A}{p_{NA}}\right)^{\gamma} + \frac{\nu}{\gamma} - \frac{1}{\eta}, \tag{10}$$

This is a slight generalization of the PIGL demand system employed by Boppart (2014).[6] This demand system has two convenient properties. First of all, 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

[6]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 8.6 of the Appendix for a detailed discussion. Boppart (2014) uses this demand system to study the the evolution of service sector. In terms of (10) he assumes that P_A is the price of goods and P_{NA} is the price of services and considers the case of $\phi = 0$.

share on the agricultural good, $\vartheta_A(e,p)$, is given by[7]

$$\vartheta_A(e,p) \equiv \frac{x_A(e,p)\,p_A}{e} = \phi + \nu\left(\frac{p_A}{p_M}\right)^{\gamma} e^{-\eta}. \tag{11}$$

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 share on agricultural goods. The long-run secular decline in agricultural employment shown above is to a large extent driving by the increase in income per capita, which shifts aggregate demand away from agriculture. Holding real income e constant, the expenditure share is increasing in the relative price of agriculture if $\gamma > 0$. 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.

We opted for the the preferences specification in (10) for two reasons. First of all, Alder et al. (2017) have shown that while the more popular Stone-Geary specification is unable to quantitatively account for the long-run process of structural change between 1880 and 2000, a preference specification in the PIGL class provides a good fit to the long-run data. Secondly, we show below that these preferences allow for a very tractable aggregation despite the fact that they fall outside the Gorman class. Hence, they can be incorporated into a general equilibrium trade model in a tractable way.[8]

Spatial Mobility 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

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

where $E^h[U_r]$ is the expected utility of living in region r (which is conditional on the skill level h), U_r is characterized in (15), 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 which 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 embedded in $E^h[U_r]$.

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

[7] As we show in detail in Section 8.2 in the Appendix, Roy's Identity implies that

$$\vartheta_A(e,p) = x_A^M(p,e) \times \frac{p_A}{e} = -\frac{\frac{\partial V(p,e(p,u))}{\partial p_A}p_A}{\frac{\partial V(p,e(p,u))}{\partial e}e},$$

where $x_A^M(p,e)$ denotes the Marshallian demand function. For $V(p,e)$ given in (10), this expression reduces to (11).

[8] This is in contrast to the non-homothetic CES demand system, which has recently been analyzed in Comin et al. (2015).

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

$$\rho_{jrt}^{h}=\frac{exp\left(\frac{1}{\kappa}\times\left(E^{h}\left[U_{r}\right]+A_{rt}-MC_{jr}\right)\right)}{\sum_{l=1}^{R}exp\left(\frac{1}{\kappa}\times\left(E^{h}\left[U_{l}\right]+A_{lt}-MC_{jl}\right)\right)}, \tag{12}$$

so that the total number of workers of skill type h in region r is simply

$$L_{r,t}^{h} \;=\; \sum_{j=1}^{R}\left(\rho_{jrt}^{h}\times L_{j,t-1}^{h}\right),$$

i.e. given by the inflows from all children of skilled individuals in all other regions. Note also that the share of non-movers of skill type h is simply given by ρ_{jj}^{h}. We will show below that $E^{h}\left[U_{r}\right]$ has a tractable closed form expression, which makes (12) 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. We first 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 write 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

Consider first the households' consumption-saving decision given in (7). The two-period OLG structure together with the specification of preferences in (10) has a tractable solution for both the optimal allocation of expenditure and the consumers' total utility U_r. We summarize this solution in the following Lemma.

Lemma 1. *Consider the maximization problem in (7), (8) and (9) where $V\left(e,P\right)$ is given in (10). The solution to this problem is given by*

$$e_{t}^{Y}\left(y\right) \;=\; \psi\left(r_{t+1}\right)\times y \tag{13}$$

$$e_{t+1}^{O}\left(y\right) \;=\; \left(1+r_{t+1}\right)\times\left(1-\psi\left(r_{t+1}\right)\right)\times y \tag{14}$$

$$U_{r} \;=\; \frac{1}{\eta}\psi\left(r_{t+1}\right)^{\eta-1}\times y^{\eta}+\Lambda_{t,t+1} \tag{15}$$

where

$$\psi(r_{t+1}) = \left(1+\beta^{\frac{1}{1-\eta}}(1+r_{t+1})^{\frac{\eta}{1-\eta}}\right)^{-1} \tag{16}$$

$$\Lambda_{t,t+1} = -\frac{\nu}{\gamma}\left(\left(\frac{p_{A,t}}{p_{M,t}}\right)^{\gamma}+\beta\left(\frac{p_{A,t+1}}{p_{M,t+1}}\right)^{\gamma}\right)+(1+\beta)\left(\frac{\nu}{\gamma}-\frac{1}{\eta}\right).$$

Proof. See Section 8.1 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 when young ($e_t^Y(y)$) and old ($e_{t+1}^O(y)$) 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. In particular, if $\eta = 0$, which is the case if demand is non-homothetic (again see Section 3.2 below), we recover the well-known OLG formulation with 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} but not on the 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, $p_{I,t}$, makes savings more attractive but at the same reduces the marginal utility of spending. Finally, overall utility U_r is additive separable between income y and current and future prices P_t and P_{t+1} (which determine $\Lambda_{t,t+1}$). This property will be essential to characterize agents' optimal spatial choice in a tractable way.

To solve agents' spatial choice problem, we have to calculate their expected life-time utility $E^h[U_r]$ - see (12). 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 $\Gamma_\zeta \times \Theta_r^h$, expected life-time utility can be calculated as

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

Substituting this expression into (12) yields the equilibrium spatial choice probabilities as

$$\rho_{jrt}^h = \frac{exp\left(\frac{1}{\kappa}\times\left(\frac{\Gamma_{\eta/\zeta}}{\eta}\psi(r_{t+1})^{\eta-1}\times\left(\Theta_r^h\right)^{\eta}+A_{rt}-MC_{jr}\right)\right)}{\sum_{l=1}^{R} exp\left(\frac{1}{\kappa}\times\left(\frac{\Gamma_{\eta/\zeta}}{\eta}\psi(r_{t+1})^{\eta-1}\times\left(\Theta_l^h\right)^{\eta}-MC_{jl}\right)\right)}.$$

This expression has two important properties. First of all, 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.[9] Moreover, for a given interest rate r_{t+1}, spatial flows *only* depend on the endogenous vector of regional income $\left[\Theta_{rt}^{h}\right]_{r}$, 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.

Equilibrium Aggregation and Aggregate Structure Change

The spatial equilibrium of economic activity is of course driven by aggregate demand conditions. Our economy does not admit a representative consumer, because the PIGL preference specification in (10) falls outside of the Gorman class. In particular, consider a set of individuals $i \in \mathscr{S}$, with spending e_i. The aggregate demand for agricultural products stemming from this set of consumers is given by

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

where $E_{\mathscr{S}} = \int_{i\in\mathscr{S}} e_i di$ denotes aggregate spending. Hence, as long as preferences are non-homothetic, i.e. as long as $\eta > 0$, aggregate demand does not only depend on aggregate spending $E_{\mathscr{S}}$, but the entire distribution of spending $[e_i]_i$ matters. Hence, 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 in principle non-trivial.

It turns out that our model delivers very tractable expressions for the economy's aggregate quantities. This is due to three reasons. The distributional assumption on individual skills implies that individual income y^i is Frechet distributed. In Lemma 1 we showed that individuals' expenditure policy functions are linear. Hence, individual spending $e$is *also* Frechet distributed. And because individuals' sectoral demands depend on spending via a power function, we can solve for the term capturing the inequality in spending explicitly. In particular, suppose that e_i is distributed Frechet with parameter A and shape ζ. It is then easy to very that (see Section 8.3 in the Appendix)

$$\int_{i\in\mathscr{S}} e_i^{-\eta}\frac{e_i di}{\int_i e_i di} = \frac{E\left[e^{1-\eta}\right]}{E\left[e\right]} = \upsilon \times A^{-\eta/\zeta},$$

where $\upsilon = \Gamma\left(1-\frac{1-\eta}{\zeta}\right)/\Gamma\left(1-\frac{1}{\zeta}\right)$ is a constant. Hence, aggregate demand is an explicit function of the dispersion in spending (ζ) and the level of income (A). Because equilibrium factor prices only determine individuals' income (and hence spending) via the parameter A, we can solve for the aggregate demand side of the economy explicitly. In particular the aggregate expenditure share of the set of consumers $\mathscr{S}$

[9]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.

is given by

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

where $E_{\mathscr{S}}=\Gamma_\zeta A^{1/\zeta}$ is mean spending and $\tilde{\nu}=\nu\Gamma\left(1-\frac{1-\eta}{\zeta}\right)/\Gamma\left(1-\frac{1}{\zeta}\right)^{1-\eta}$. Hence, the aggregate demand system is *as if* it stems from 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 $2\times2\times R$ distinct sets of the consumers - two generations, two skill types and R locations. Letting $E_{r,t}^{h,Y}$ and $E_{r,t}^{h,O}$ denote the mean spending of the young and generation with skills h in region r, Lemma 1 implies that

$$E_{r,t}^{h,Y}=\psi(r_{t+1})\Gamma_\zeta\Theta_{r,t}^{h}\text{ and }E_{r,t}^{h,O}=\left[(1+r_t)(1-\psi(r_t))\right]\Gamma_\zeta\Theta_{r,t-1}^{h}.$$

Note that the amount of spending of the old generation in year t depends on their income earned in period $t-1$, their savings rate $(1-\psi(r_t))$ and the accrued interest. The aggregate amount of consumer spending on agricultural goods is therefore given by

$$PC_t^A=\phi\times E_t+\tilde{\nu}\left(\frac{p_A}{p_M}\right)^{\gamma}\times\sum_{r,h}\left(\left(E_{r,t}^{h,Y}\right)^{1-\eta}\lambda_{r,t}^{h}L_{r,t}+\left(E_{r,t}^{h,O}\right)^{1-\eta}\lambda_{r,t-1}^{h}L_{r,t-1}\right),\tag{17}$$

where $E_t=\sum_{r,h}\left(E_{r,t}^{h,Y}\lambda_{r,t}^{h}L_{r,t}+E_{r,t}^{h,O}\lambda_{r,t-1}^{h}L_{r,t-1}\right)$ denotes aggregate consumer spending.
Because a fraction ϕ of total investment spending is spent on agricultural goods and total value added (or GDP) is given by $PY_t=I_t+PC_t$, the agricultural share in value added is given by

$$\vartheta_{A,t}=\frac{\phi I_t+PC_t^A}{PY_t}=\phi+\tilde{\nu}\left(\frac{p_A}{p_M}\right)^{\gamma}\times\frac{\sum_{r,h}\left(\left(E_{r,t}^{h,Y}\right)^{1-\eta}\lambda_{r,t}^{h}L_{r,t}+\left(E_{r,t}^{h,O}\right)^{1-\eta}\lambda_{r,t-1}^{h}L_{r,t-1}\right)}{PY_t}.\tag{18}$$

Moreover, because workers receive a fraction $1-\alpha$ of aggregate output, we can express aggregate GDP as

$$PY_t=\frac{1}{1-\alpha}\times\sum_{r,h}\Gamma_\zeta\Theta_{r,t}^{h}\lambda_{r,t}^{h}L_{r,t}.\tag{19}$$

Equations (18) and (19) highlight two properties. First of all, for a given allocation of people across space, both aggregate income PY_t and the agricultural share $\vartheta_{A,t}$ only depend on the vector of current and past average regional incomes $\left[\Theta_{r,t}^{h},\Theta_{r,t-1}^{h}\right]_r$. Secondly, these equations highlight the usual demand side forces of structural change. In particular, consider an allocation where the distribution of individuals by skill is stationary (i.e. $\lambda_{rt}^{h}=\lambda_r^h$ and $L_{r,t}=L_r$) and where regional incomes Θ_{rt} grow at rate g. Because spending $E_{r,t}^{h,g}$ and GDP are then proportional, (18) implies

$$\vartheta_{A,t} = \phi + \nu \left(\frac{p_{A,t}}{p_{M,t}} \right)^{\gamma} \times \Xi \times PY_t^{-\eta},$$

where Ξ is an (endogenous) constant. Hence, rising aggregate income PY_t will reduce the agricultural spending share as long as demand is non-homothetic, i.e. $\eta > 0$. At the same time, changes in relative prices will also affect $\vartheta_{A,t}$. In our model, the spatial allocation of resources will of course not be stationary as the economy undergoes the structural transformation. Similarly, regional incomes Θ_{rt} will also not grow at a constant rate. These features of "unbalanced spatial growth" will also affect the agricultural spending share directly.

Equilibrium

Given the physical environment above, we can now characterize the dynamic competitive equilibrium in our economy. Our assumptions ensure that the analysis remains very tractable. As highlighted above, the key properties of our theory are that (i) individual moving decisions are static and (ii) that our economy generates an aggregate demand system as a function of regional incomes $\left[\Theta_{rt}^{h}\right]_h$. This implies that, for a given path of interest rates $\{r_r\}_t$, we can calculate the equilibrium by simply solving a set of equilibrium conditions.

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

$$\sum_h w_{r,A} H_{r,A}^{h} \quad = \quad (1-\alpha) \times \pi_{rAt} \times \vartheta_{A,t}^{VA} \times PY_t, \tag{20}$$

i.e. total agricultural labor earnings in region $r$are equal to a constant share of total agricultural revenue in region r. This in turn is equal to region r's share, π_{rAt}, in aggregate spending on agricultural goods, $\vartheta_{A,t}^{VA} \times PY_t$. Standard arguments imply that regional trade shares π are given by

$$\pi_{rAt} = \left(\frac{P_{rAt}}{P_{At}} \right)^{1-\sigma} = \frac{\left(\frac{Q_{rAt}}{w_{rAt}^{1-\alpha}} \right)^{\sigma-1}}{\sum_{r=1}^{R} \left(\frac{Q_{rAt}}{w_{rAt}^{1-\alpha}} \right)^{\sigma-1}}, \tag{21}$$

i.e. they do neither depend on the identity of the sourcing region, nor the equilibrium interest 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. Similarly, it is easy to verify that regional agricultural labor income is given by

$$\sum_h w_{rAt} H_{rAt}^{h} = L_{rt} \Gamma \left(1 - \frac{1}{\zeta} \right) w_{rAt}^{\zeta} \times \sum_h \lambda_r^h \Psi_s^h \left(\Theta_r^h \right)^{1-\zeta}.$$

Together with the spatial labor supply equation (12) and the corresponding labor market clearing conditions for the non-agricultural sector, these equations directly determine the equilibrium levels of regional income $\left[\Theta_{rt}^{h}\right]_{h}$ (or alternatively the equilibrium wages).

Finally, we can also directly calculate the remaining macroeconomic aggregates, in particular the dynamic accumulation of capital. Because the future capital stock is only given by the savings of the young generation, we get that

$$K_{t+1} \quad = \quad (1-\psi(r_{t+1})) \times \sum_{r,h} Y_{r,t}^{h} \lambda_{r,t}^{h} L_{r,t} = (1-\psi(r_{t+1}))(1-\alpha) P Y_{t}. \tag{22}$$

Hence, future capital is simply a fraction $1-\psi(r_{t+1})$ of aggregate labor earnings. This proportionality between the aggregate capital stock and aggregate GDP is a consequence of the linearity of agents' consumption policy rules.

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 (22). 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 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 $\left[e_{rt}^{Y}, e_{rt}^{O}, s_{rt}\right]$, sectoral spending shares $\left[\vartheta_{rt}^{Y}, \vartheta_{rt}^{O}\right]$ and demands for regional varieties $[c_{rst}]$ such that*

1. *consumers' choices $\left[e_{rt}^{Y}, e_{rt}^{O}, s_{rt}\right]$ maximize utility, i.e. are given by (13) and (14),*
2. *the sectoral allocation of spending $\left[\vartheta_{rt}^{Y}, \vartheta_{rt}^{O}\right]$ is optimal, i.e. given by (11),*
3. *the demand for regional varieties follows (21) and firms' factor demands maximize firms' profits*
4. *markets clear,*
5. *the capital stock evolved according to (22),*
6. *the allocation of people across space $\left[L_{r,t}^{Y}\right]$ is consistent with individuals' migration choices in (12).*

4 The Spatial Gap

The above model is a model of spatial structural change. It delivers a concise framework to connect the allocation of factors across space and sectors along the structural transformation. Besides being of interest in itself, this paper argues that novel spatial dimension of structural change can in fact explain why agricultural productivity at the *aggregate* level is likely to be persistently low during the structural transformation.

We define the agricultural productivity gap as the share of agricultural value added relative to the share of agricultural employment. In our spatial model, this productivity gap can be written as

$$GAP_t \equiv \frac{\vartheta^{VA}_{A,t}}{L_{A,t}/L_t} = \frac{\sum_{r=1}^{R} s_{Art} \times \frac{\bar{y}^A_{rt}}{\bar{y}_{rt}} \times \frac{\bar{y}_{rt} \times L_{rt}}{\sum_{r=1}^{R} \bar{y}_{rt} \times L_{rt}}}{\sum_r s_{Art} \times \frac{L_{rt}}{\sum_r L_{rt}}}. \tag{23}$$

Here, $s_{A,r}$ and L_{rt} denote the regional agricultural employment share and the regional population and $\bar{y}_{rt}$ and $\bar{y}^A_{rt}$ denote regional *average* earnings and agricultural earnings, respectively, i.e.

$$\bar{y}^A_r = \frac{\lambda_r s^H_{A,r} \Theta^H_{rt} + (1-\lambda_r)\,\Theta^L_r s^L_{A,r}}{s_{A,t}} \text{ and } \bar{y}_r = \lambda_{rt} \Theta^H_{rt} + (1-\lambda_{rt})\,\Theta^L_{rt},$$

where all the variables are defined as above. Hence, the *aggregate* agricultural productivity gap depends on the joint distribution of regional agricultural employment shares s_{Art}, relative agricultural earnings $\bar{y}^A_{rt}/\bar{y}_{rt}$, average regional earnings $\bar{y}_{rt}$ and the size of the population L_{rt}. While all these ingredients are of course endogenous to the process of structural change, we will show that the gap implied by (23) can quantitatively go a long way to explain the aggregate productivity gap observed in the US.

To build some intuition why this is the case, suppose there was only *absolute* advantage but no comparative advantage, i.e. $\mu = 1$. This implies that high and low skilled earnings are proportional ($\Theta^H_{rt} = q \times \Theta^L_{rt}$) and that sectoral employment shares are equalized across skill groups (i.e. $s^H_{A,r} = s^L_{A,r} = s_{A,r}$). Hence, relative agricultural earnings within regions are equalized $\left(\bar{y}^A_{rt} = \bar{y}_{rt}\right)$ and the aggregate productivity gap in (23) reduces to

$$GAP_t = \frac{\sum_{r=1}^{R} s_{A,r} \times \frac{L_r \bar{y}_r}{\sum_{r=1}^{R} L_r \bar{y}_r}}{\sum_r s_{Art} \times \frac{L_{rt}}{L_t}}, \tag{24}$$

where regional income per capita $\bar{y}_r$ is given by

$$\bar{y}_r = \underbrace{[\lambda_r \times q + (1-\lambda_r)]}_{\text{Human Capital}} \times \underbrace{\Theta^L_{rt}}_{\text{MPL}}, \tag{25}$$

i.e. reflects both the skill composition of the regional workforce and average earnings, i.e. the marginal productivity of labor in region r.

These equations are instructive. In particular, they highlight that agricultural productivity is relatively *low* if the cross-sectional correlation between agricultural employment shares $s_{A,r}$ and regional income per capita $\bar{y}_r$ is *negative*. Hence, the low aggregate productivity of the agricultural sector might simply reflect the fact that agricultural intensive regions attract low-skilled individuals $\left(corr\left(s_{A,r}, \lambda_r\right) < 0\right)$ or that they have low marginal productivity in *all* industries, i.e. a low level of productivity $Q_{A,r}$ and $Q_{NA,r}$. If $\bar{y}_r$ was equalized across space, there would be no productivity gap. This would for example be the case if all regions were symmetric.

We can further decompose the variation in earnings Θ^L_{rt}. In particular, individual's sectoral labor supply function implies that (see (6)) that

$$\Theta^L_{rt} = \underbrace{w_{rA,t}}_{\text{Skill prices}} \times \underbrace{s_{rAt}^{-1/\zeta}}_{\text{Selection}},$$

i.e. total earnings in region r could be low because agricultural skill prices are low or because the share of workers in the agricultural sector is large. This last term is the selection effect highlighted by Lagakos and Waugh (2011), whereby sectoral specialization reduces the average product of labor through sorting based on comparative advantage. Because the marginal worker in the agricultural sector is worse than the average worker, the larger the share of people working in agriculture, the lower average warnings, holding agricultural skill prices constant.

Finally, equation (24) also highlight why it is appropriate to refer to the agricultural productivity gap as a *spatial gap*: in case there is only a single region, i.e. $R = 1$, (24) implies that $GAP_t = 1$, i.e. productivity is equalized across sectors as in the baseline model of structural change. While this is seemingly inconsistent with Lagakos and Waugh (2011), note that they look at *physical* productivity across sector. The agricultural productivity gap, however, is evaluated at sectoral prices. If worker selection takes the Frechet form, the productivity and price effect *exactly* cancel out.[10] If, however, there is spatial variation in the extent to which workers sort into different sectors and wages differ across regions, the extent of selection does directly determine the agricultural productivity gap. To what extent this spatial friction is quantitatively important depends on the calibrated model. This is where we turn now.

5 Quantitative Exercise

We now take the framework above to the data. To do so, we exploit a novel dataset on regional economic development of the US between 1880 and 2000. We first describe the dataset and provide three basic empirical regularities, which suggest the importance of spatial frictions for the productivity gap in agriculture. After describing our calibration strategy we then turn to our two main results. First we measure the spatial gap through the lens of the model. We then consider our counterfactual exercise, where we gauge the importance of the costs of spatial mobility for the agricultural productivity gap.

[10] We show this formally in Section 9.2 in the Online Appendix.

5.1 Data

Our main data sources are the Census of manufacturing for 1880 and 1910, the Micro Census for 1880-2000 and County and City Data Books for 1940-2000. From these sources we construct a data set of total workers $[L_{rt}]_{rt}$, average manufacturing wages $[w_{rt}]$, and sectoral employment shares $\left[s_{A,rt}\right]_{rt}$ for all US counties at 30 year intervals between 1880 and 2000.[11] We define the agricultural sector to comprise agriculture, fishing and mining industries defined according to the 1950 Census Bureau industrial classification system (outlined in Ruggles et al. (2015)). All remaining employed workers are classified as non-agricultural workers. We construct average manufacturing wages from county level total manufacturing payroll data and manufacturing head counts obtained from the same source.[12][13] In Appendix 9.14 table 7 contains a comprehensive list of data sources.

We are interested in structural change within and across regions of the continental US economy. This poses the question of what constitutes a meaningful spatial aggregation for local economies in the United States, which have the potential to experience structural change within them and together cover all US territory.[14] Commuting zones as constructed by Tolbert and Sizer (1996) constitute a meaningful definition of such local economies. They partition the US economy into 741 local labor markets that in 1990 maximized commuting flows within and minimized them across. Contrary to counties all of these commuting zones exhibited non-zero non-agricultural employment shares in 1880 as can be seed in figure 3. Note that in 1880 around half of the workforce is employed in agriculture in the aggregate.

To map our county data to commuting zones we use ARCGis software to construct a crosswalk between counties and 1990 commuting zones for every decade between 1880 and 2000. Using this we re-aggregate county level data to commuting zones, employing area weights to allocate county level allocations wherever counties are split.[15]

The result of this procedure is a panel data set for 717 continental commuting zones from 1880 to 2000 which features sectoral employment shares, total employment and average manufacturing wages. For the

[11] 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). 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 (9.14).

[12] 1880 wages: 1880 Census of Manufacturing; 1910 wages: 1920 Census of Manufacturing; 1940 wages: 1947 County and City Data Book; 1970 wages: 1970 County and City Data Book; 2000 wages: 2000 County and City Data Book.

[13] 1940 is the only year for which we have wage data from the micro census and on the county level from the Census of Manufacturing. Aggregating and averaging the former to the county level leads a worker weighted correlation coefficient of about 0.7.

[14] The requirement for this is a somewhat diversified industrial structure. In the 1880 data there are counties with agricultural employment shares of almost 1, making counties a seemingly unsuitable aggregation.

[15] The United States moved from 2473 counties in 1880 to 3142 counties in 2000. In this process some new counties were created by splitting in half an existing one and other new ones created by joining together parts of others. In order to map 1880 county level data to 1990 counties and from there to commuting zones, it hence becomes necessary to work with fractions of counties. We use area weights and the assumption of a uniform distribution of industrial activity across space for this aggregation. More details in the Appendix.

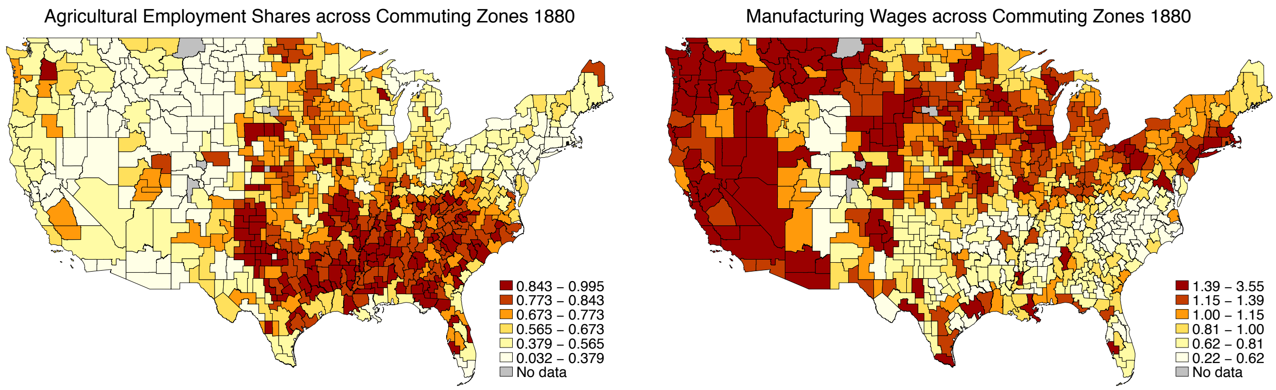


Figure 3: Left Panel: Agricultural Employment Share across US Commuting Zones in 1880. Right Panel: Ration of Local Manufacturing Wages relative to US Median across US Commuting Zones in 1880.

main calibration of the model we employ the cross-sections 1880, 1910, 1940, 1970 and 2000 only and normalize the size of the workforce to unity in each period. We scale the level of wages to ensure that income per capita grows at a constant rate. More details on the the construction of this panel can be found in online Appendix 9.14.

Beyond this panel, the 1940 version of the decennial Micro Census is a data source of particular importance for the paper: it is the most recent Census for which all US counties are available and it is the first Census for which earnings and education variables are available.[16] We define a skilled worker as one with a completed high school degree.[17] This pins down the skilled employment share in 1940, λ_{r1940}. The next section demonstrates that the structure of the model together with other data inputs is such that a single cross-section of skill shares is sufficient to infer local skill shares for the remaining cross-sections in the sample. Finally, we use the 1940 Census to compute the US wide skill premium and a manufacturing skilled employment premium which are used in the calibration below. To that end we use total pre-tax annual wage and salary income together with the same sector definitions as in the construction of the panel to construct the average annual wage of skilled workers relative non-skilled ones. The manufacturing skilled employment premium is simply the ratio of the fraction of skilled people working in manufacturing over the fraction of unskilled workers working in manufacturing out of all unskilled workers.

In the model agents move once in their life time in order to choose a labor market. We use the US decennial Census Micro data obtained from IPUMS (see Ruggles et al. (2015)) to construct an interstate migration flow matrix for the cross-sections 1880, 1910, 1940, 1970 and 2000. In a given Census we choose workers between 26 and 50 years of age, and compute the number of them living in a different

[16]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).

[17]We assume a direct mapping between skilled employment and education. Choosing a higher education cutoff would result in a overall skilled worker share in the US below 0.29, which seems unrealistic. We do robustness with lower cutoffs.

state from their state of birth. Where we use them state level wages are constructed by aggregating county level manufacturing wages from above panel.

We use the data constructed by Alvarez-Cuadrado and Poschke (2011) on the relative price of manufacturing commodities to agricultural goods directly infer $\left[\frac{P_{A,t}}{P_{M,t}}\right]_t$ from the data. To map the model more directly to the data a combined price index for manufacturing and the service sector would be needed. We do not know of the existence of a reliable such series.

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, i.e. the parameter η.

5.2 Calibration

In this section we outline our calibration procedure. In Section 9.3 in the Appendix we describe our calibration in much more detail. 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 8.5 in the Appendix, we show that this implies that interest rates are constant and have a closed form expression. For our counterfactual analysis, interest rates will of course not be constant.

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 $\lambda_{r,1880}$. We define skilled individuals as workers with at least a completed high-school education in 1940 and hold the aggregate share of skilled workers fixed. For the theory we need the spatial distribution of skilled workers in 1880, i.e. $\lambda_{r,1880}$. In the data this object is unobserved, as the 1880 micro census only has information on literacy, but not on schooling. We calibrate this initial allocation by requiring the model to endogenously perfectly replicate the cross-sectional distribution of skill supplies in 1940.

The 1940 cross-section of data is the only one for which we observe average manufacturing wages, sectoral employment and the commuting zone level skill share $\lambda_{r,1940}$. Section 9.3.1 in the Appendix shows how these data inputs can be translated into regional sectoral wages, w_{rs1940}, and human capital stocks, H_{rs1940}, conditional on the parameters ζ, μ, q. This in turn allows us to construct $\Theta^x_{r,1940}(\zeta, \mu, q)$ and $s^x_{r,s}(\zeta, \mu, q)$ for $x = H, L$.

We calibrate the parameter ζ, i.e. the dispersion in individual productivity, to match the dispersion of earnings in the 1940 Census data. In particular, the model implies that the variance of log earnings with region-skill cells is given by $\frac{\pi^2}{6} \times \zeta^{-2}$. We therefore identify ζ from

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

where 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.

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 1940 Micro Census data we calculate the unconditional skill premium by taking the average yearly earnings of all individuals with at least high school education and dividing it by the equivalent number for all workers with less than high school education, which yields a skill premium of 1.62. 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. Empirically, this number is equal to 1.21, i.e. high-skilled individuals are 20% more likely to work outside of agriculture. 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.

Aggregate and Local Productivities and Amenities We calibrate local productivities and amenities $[Q_{rst}, A_{rt}]$ as structural residuals, i.e. we force the model to match the spatial data perfectly with $[Q_{rst}, A_{rt}]$ absorbing any residual variation. In the previous section we showed how to obtain sectoral wages and human capital conditional on regional skill shares and a range of other available data and already calibrated parameters. In section (9.3.1) of the appendix we show that these objects can be used to calculate trade shares as follows:

$$\pi_{rst} = \frac{w_{r,s}H_{r,s}}{\sum_r w_{r,s}H_{r,s}}.$$

As a result we can treat π_{rst} as data to obtain Q_{rst} from the expression for trade shares given in (21) which we restate here for convenience

$$\pi_{rst} = \frac{\left(\frac{Q_{rst}}{w_{rs}^{1-\alpha}}\right)^{\sigma-1}}{\sum_{J=1}^{R}\left(\frac{Q_{jst}}{w_{js}^{1-\alpha}}\right)^{\sigma-1}} \quad \text{with} \quad \sum_r^R Q_{rst}^{\sigma-1} = 1$$

Together with sectoral wages and the parameters (σ, α), this equation along with our normalization identifies Q_{rst} exactly.[18]

[18]To see the role of Q_{rst} as structural residuals more directly note we can rewrite (21) as follows:

$$\log \frac{\pi_{rst}}{\pi_{1st}} = (\sigma-1)\log(Q_{rst}/Q_{1st}) + (1-\sigma)(1-\alpha)\log(w_{rst}/w_{1st})$$

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 equation (12) which allows for the construction of a useful mapping

$$A_{rt}^{k+1} = f(\{\Theta_{rt}^h, L_{rt}^h, L_{rt-1}^h, A_{rt}^k\}_{h=L,H}) \quad \text{with} \quad \sum_r A_r = 0$$

The algorithm to jointly back out skill shares and amenities is then as follows: (1) Use 1940 skill shares from the data to calculate Θ_{rt}^h. (2) Guess skill shares in 1910 and use worker stocks in 1910 to calculate L_{rt-1}^h. Then we iterate on f to get implied amenities A_{rt}. 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. We iterate between updating skill shares and amenities until convergence. (3) We then guess skill shares for 1970 and use the f mapping forward to obtain 1970 and 2000 skill shares and amenities: this time guessing skill shares three decades ahead, calculating implied amenities by iterating on f and then updating the skill share guess. This procedure leaves us with a panel of skill shares and amenities for all commuting zones and cross-sections. More details can be found in section 9.3.1 in the Appendix. In figure 14 in the Appendix we plot the densities of the calibrated $[Q_{rst}, A_{rt}]$ for every cross-section.

Lastly, we calibrate the time series of aggregate sectoral productivities, $[Z_{At}, Z_{Mt}]$, to match the evolution of relative prices and a GDP growth rate of 2%. We plot these two implied series in figure 15 in the Appendix.

Moving Costs We specify moving frictions as a quadratic function of distance as follows:

$$\tau_{ij} = \begin{cases} \tau & \text{if } j = i \\ \delta_1 d_{ij} + \delta_2 d_{ij}^2 & \text{if } j \neq i \end{cases}$$

$\tau > 0$ corresponds to a fixed cost of moving: if a worker decides to move away from his commuting zone of birth he forgoes the place utility τ. Additionally movers have to pay a cost, denominated in utils, that varies with the distance travelled. We normalize d_{ij} so that the maximum distance in the US is 1. Before this normalization the largest distance between two continental commuting zone centroids is 2827.4 miles.

In the Census data we observe for every decade and for every individual her state of birth and her current state of residence. Using this we can construct a lifetime state-to-state migration flow matrix. For our baseline estimates we would like to choose $(\kappa, \tau, \delta_1, \delta_2)$ so as to match these state to state flows as closely as possible for the 1910-1940 period, while still matching local populations, employment shares and wages exactly.

Regressing trade shares (now observed) on wages yields the log ratio of Q_{rst} as a structural residual. The necessity of an additional normalization then also becomes clear. We choose $\sum_r^R Q_{rst}^{\sigma-1} = 1$ since this choice helps to simplify certain analytic expressions.

Since our model features commuting zones while our data records state to state flows we will aggregate commuting zone flows to the state level in the model and calibrate $(\kappa,\tau,\delta_1,\delta_2)$ so as to make the model fit state to state flows as closely as possible.

Consider the flow equation from the model:

$$L^h_{r,t} = \sum_j \rho^h_{jr,t} \times L^h_{j,t-1}$$

where ρ^h_{jr} is given in equation 12. Also note:

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

So that the number of stayers relative to movers increases monotonically in τ for fixed A_l. Note that $const^h_j$ is independent of moving cost parameters given that we always match local labor supply exactly. The procedure for estimating $(\kappa,\tau,\delta_1,\delta_2)$ is then to guess $(\kappa,\delta_1,\delta_2)$, compute commuting zone flows and aggregate them to the state level. For each guess τ is chosen so as to ensure the model matches the aggregate interstate migration rate of 0.3 exactly. We then search over combinations of $(\kappa,\delta_1,\delta_2)$ to identify the tupel that minimizes the following objective:

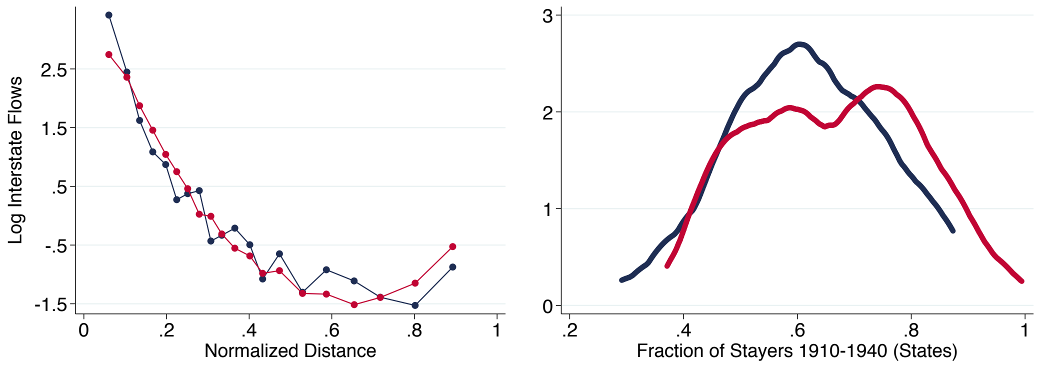
$$\Lambda = \sum_i \sum_{j\neq i} L_{i,1940} \times \left(\log \rho^{DATA}_{ij,1940} - \log \rho^{MODEL}_{ij,1940}(\kappa,\delta_1,\delta_2)\right)^2$$

The objective function Λ is a destination population weighted sum of the percentage difference between model and data of the inflows from all states except the destination itself. 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.

Note that for any guess of $(\kappa,\delta_1,\delta_2)$ we resolve 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.

Figure 4 shows the density of state level stayers in model and data as well as the distance-flow relationship. The model matches the pattern observed in the data quite well. In particular, the model matches the cross-sectional heterogeneity in the share of stayers within a location.

Preferences I: Estimating the non-homotheticity η We use the microdata from the Consumer Expenditure Survey in 1936 ("Study of Consumer Purchases in the United States, 1935-1936") to estimate the extent of non-homotheticities. 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 relationship



Notes: In the right panel we plot the distribution of the share of people staying in their home state in the data (blue) and model (red). To construct the left panel, we run a gravity equation of the form

$$\frac{\rho_{jr}}{1-\rho_{jj}} = \alpha_j + \alpha_r + \delta_1 \times d_{jr} + \delta_2 \times d_{dj}^2 + u_{jr}$$

where ρ_{jr} denote the share of people moving from j to r, α_j and α_r denote origin and destination fixed effects and d_{jr} denotes the distance between j and r. We run the regression both in the model (red line) and in the data (blue line) and then plot the predicted values.

Figure 4: State-to-state migration: Model vs Data

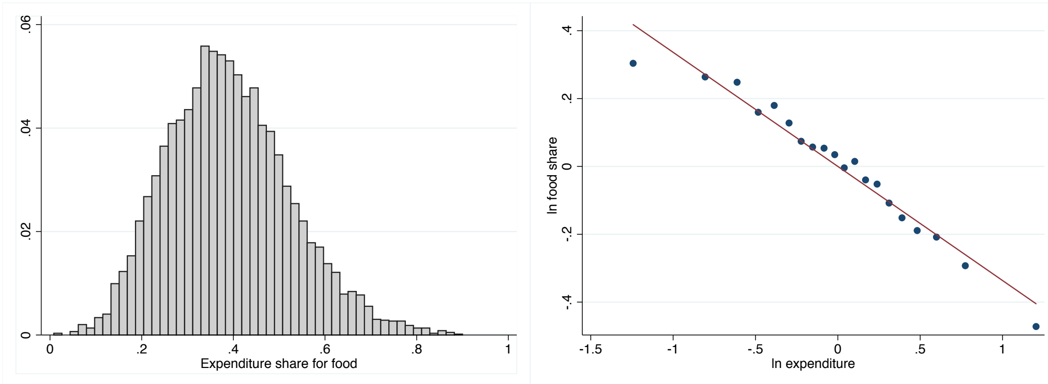
between expenditure shares and total expenditure, i.e.

$$\ln \vartheta_A(e,p) = f(p) - \eta \times \ln e. \tag{26}$$

Note that the intercept $f(p)$ is a function of prices but does not vary in the cross-section of households. In Figure 5 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, which are supposed to proxy for $f(p)$. The slope of the regression line is exactly the extent of the demand non-homotheticity η. It is clearly seen 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 implied that $\eta = 0.32$. In Section 9.3 in the Appendix we also report the regression results, when we 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.

Preferences II: Estimating (ϕ, ν, γ) In contrast to η, which is estimated from the cross-sectional relationship, we 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. In particular, the model implies that the elasticity of substitution of individual i is given by

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



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 as in (26) (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 the Study of Consumer Purchases in the United States, 1935-1936.

Figure 5: Expenditure Shares on Food in 1936

Hence, the elasticity of substitution varies across households and across time. Comin et al. (2015) estimate this elasticity to be around 0.7 in post-war data for the US. We therefore chose γ so that the model implies an average elasticity of substitution of 0.7 in the year 2000.

Given our calibration strategy for the underlying distribution of productivity and amenity fundamentals, internal consistency requires us to match these four agricultural employment shares exactly. As we discuss in detail in Section 9.13 in the Appendix, the income effects as implied from the cross-sectional spending-food relationship displayed in Figure 5 are not strong enough to explain the entire decline in agricultural employment in the time-series.[19] We therefore allow the parameter ϕ to be time-specific to fully account for the residual decline in agricultural employment and chose 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. Note that ϕ does not enter the household's decision problem directly. This strategy, which we explain in detail in Section 9.10 of the Appendix, yields parameter estimates, which are reported in Table 1.

Endogenous Amenities 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:

$$A_{r,t} = \rho L_{r,t}^{\theta} + \bar{A}_{r,t} \tag{27}$$

where $\bar{A}_r$ is the exogenous component of the amenity inherent to location r and $\rho, \theta \in \mathbb{R}$ flexibly parameterize the endogenous effect of local population on amenities. It is natural to think about a congestive

[19]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.

effect of local population on local amenities as in Ahlfeldt et al. (2015) and Allen and Arkolakis (2014) and our estimates will imply this.

Similar to Ahlfeldt et al. (2015), notice that for each set of parameters of the model, there is a unique vector $A_{r,t}$ 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 back out the full panel of amenities $A_{r,t}$ for $t = 1910, 1940, 1970, 2000$. We then rewrite 27 in differences:

$$\Delta A_{r,t} = \rho \Delta L^{\theta}_{r,t} + \Delta \bar{A}_{r,t} \tag{28}$$

In equation 28 $\{\Delta A_{r,t}, \Delta L_{r,t}\}_{r=1,\ldots,R,t=1910(30)2000}$ are observed independently of (ρ,θ). With an instrument $\varphi_{r,t}$ for $\Delta L^{\theta}_{r,t}$ that induces variation in local population inflows between t and $t-1$ orthogonal to $\Delta \bar{A}_{r,t}$ the two parameters (ρ,θ) could be obtained by minimizing the sample equivalent of the following moment condition:

$$E\left[\varphi_{r,t}\Delta \bar{A}_{r,t}\right] = E\left[\varphi_{r,t}\left(\Delta A_{r,t} - \rho \Delta L^{\theta}_{r,t}\right)\right] = 0$$

To construct such an instrument we turn to 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) holding the implied $\bar{A}_{r,1880}$ fixed throughout but letting all other fundamentals vary as calibrated. This generates a series of population changes $\Delta L_r(\Delta Q_{r,s,t}, \Delta \bar{A}_{r,t} = 0, \Delta Z_{s,t} \mid \theta, \rho)$ which is by construction orthogonal to changes in $\bar{A}_{r,t}$.

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

Other parameters Finally, we need values of the capital share α, the rate of depreciation δ, the preference parameter σ and the consumers' discount rate β. For (α,σ,δ) we we take central values from the literature. We take a capital share of $1/3$. This implies that our model is consistent with the aggregate labor share of 2/3, which is relatively constant during our time-period. We assume an annual rate of depreciation of 8%. Finally, we chose $\sigma = 4$. We also target an investment rate of 15% and a growth rate of 2%, both of which are consistent with the aggregate US experience in the 20th century. These moments directly determine β.

5.3 Calibration Results

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.

An important part of our calibration strategy is that we calibrate the cross-sectional distribution of sectoral productivities $\{Q_{A,rt}, Q_{NA,rt}\}_r$ and amenities $\{A_{rt}\}_r$ as structural residuals. Hence, given the remaining

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 substition in 2000	0.35	0.7	0.7
		Moving Costs and Mobility			
τ	Fixed costs of moving	Interstate migration rate	1.63	0.3	0.3
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	.	
α	Capital share in production function	set exogenously	1/3	.	.

Notes: This table contains the calibrated parameters.

Table 1: Calibrated structural parameters

	Dependent Variable		
	log Pop. ($\ln L_r$)	Agr. empl. share ($s_{A,r}$)	log manufac. earnings
Productivity ($\ln Q_{NA,r}$)	0.522***	-0.019***	0.138***
	(0.032)	(0.004)	(0.007)
Comparative Adv. ($\ln Q_{A,r}/Q_{NA,r}$)	-0.042	0.022***	0.038***
	(0.036)	(0.004)	(0.007)
Amenities (A_r)	0.567***	0.007	-0.164***
	(0.080)	(0.011)	(0.017)
Year FE	Yes	Yes	Yes
Observations	2580	2580	2580
R^2	0.916	0.783	0.995

Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. The dependent variables are the structural residuals "productivity" ($\ln Q_{NA,r}$), "comparative advantage ($\ln Q_{A,r}/Q_{NA,r}$)" and "Amenities" (A_{rt}), which are inferred from the calibrated model. The regressions are run for the years 1910, 1940, 1970 and 2000.

Table 2: Fundamentals and Endogenous Outcomes

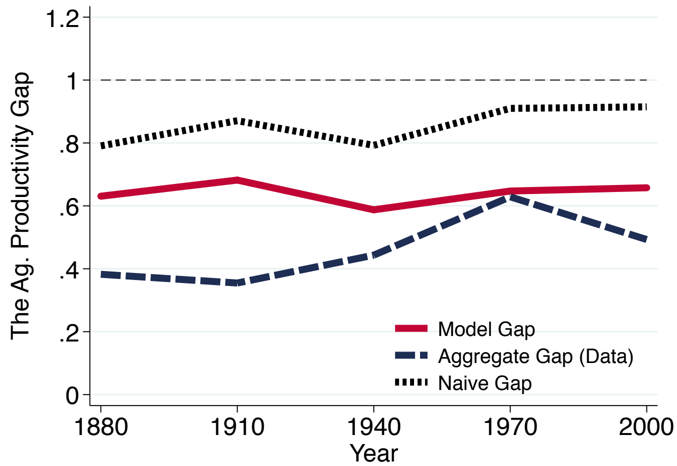
parameters reported in Table 1, the model matches the population distribution, the regional agricultural employment shares and average manufacturing earnings exactly. In Table 2 we report three descriptive regressions to illustrate the model's mapping form unobserved fundamentals to observed endogenous outcomes. In particular, we run regression of the form

$$y_{rt} = \delta_t + \gamma \times \ln Q_{NA,rt} + \beta \times \ln\left(\frac{Q_{A,rt}}{Q_{NA,rt}}\right) + \alpha \times A_{rt} + u_{rt},$$

where y_{rt} denotes the endogenous outcomes at the regional level, i.e. population size, agricultural employment shares or manufacturing earnings. For ease of interpretation, we project these outcomes on the *level* of productivity $\left(Q_{NA,rt}\right)$, the comparative advantage in agriculture and regional amenities. We find that (i) regional size is mostly driven by the level of productivity and regional amenities, that (ii) a regions' agricultural employment share is negatively correlated with regional productivity and positively correlated with comparative advantage in agriculture and that (iii) manufacturings earnings are positively correlated with productivity and comparative advantage but negatively correlated with amenities.

6 Results

With the calibrated model at hand, we can now answer the two questions we set out to address. In Section 6.1 we ask to what extent the calibrated model can account for the observed agricultural productivity gap at the aggregate level. We also provide a decomposition into the different regional fundamentals. In Section 6.2 we then quantify the importance of spatial mobility costs. In particular, we show that the spatial gap would not have been markedly lower in the absence of moving costs and we also explain why



Notes: The figure shows the agricultural productivity gap as implied by the calibrated model (red line), the agricultural productivity gap from the aggregate data (blue line) and the "naive spatial gap", which we calculate as $GAP_t^{Naivs} = \left(\sum_r s_{A,rt} \times \frac{e_{rt}^M L_{rt}}{\sum_r e_{rt}^M L_{rt}}\right) / L_{A,t}$, where e_{rt}^M denotes the average manufacturing earnings from the data.

Figure 6: The Spatial Agricultural Productivity Gap in the US from 1880 - 1910

this is the case.

6.1 The Spatial Gap in the US from 1880 to 2000

With the quantitative model we can now calculate the implied spatial gap by evaluating (23). The result is contained in Figure 6 below. In red we display the implied agricultural gap by the model. It is seen that the model implies that value added per worker is substantially lower in agriculture than in non-agriculture, even though there are - by construction - no sectoral labor frictions. For comparison we also depict the productivity gap as measured in the aggregate data (and reported in Figure 1) as the blue line. Hence, Figure 6 shows that our model can comfortably explain between 50% and 70% of productivity differences between sectors without resorting to frictions across industries.

Because our model is calibrated to perfectly rationalize the data on employment shares and manufacturing earnings, the question is to what extent Figure 6 is informative about the mechanism of the model. To see that the model indeed adds explanatory power for the implied agricultural gap, the black line in Figure 6 shows the implied productivity gap from a "naive" exercise, which abstracts from human capital differences and treats observed manufacturing earning as skill-adjusted wages.In particular, suppose one calculates the productivity gap

$$GAP_t^{\text{Naive}} = \frac{\sum_r s_{A,rt} \times \frac{e_{rt}^M L_{rt}}{\sum_r e_{rt}^M L_{rt}}}{\sum_r s_{A,rt} \times \frac{L_{rt}}{\sum_r L_{rt}}}, \tag{29}$$

where e_{rt}^M denotes the observed average manufacturing earnings in region r. Note that (29) can be calculated directly from the data. As can be seen from Figure 6, the implied productivity gap is much

	1880	1910	1940	1970	2000
Full Spatial Gap	0.63	0.70	0.62	0.69	0.70
No Change in Spatial Fundamentals					
Constant Amenities	0.63	0.70	0.59	0.65	0.68
Constant Productivities	0.63	0.60	0.54	0.48	0.48

Notes: In the first row we report the spatial gap as implied from our baseline calibration displayed in Figure 6. In the second row we report the implied spatial gap if we kept all spatial fundamentals, i.e. amenities A_{rt} and regional productivities Q_{srt} constant at their 1880 level. In row three (four) we keep amenities A (productivities Q) constant and only vary productivities (amenities).

Table 3: Decomposing the Spatial Gap

lower. This implies that the correlation between observed manufacturing earnings and agricultural employment shares understates the required one with skill-adjusted wages substantially. In particular, the model generates a negative correlation between skill shares and agricultural employment across space, allows for differences in comparative advantage, where by skilled individuals sort into the manufacturing sector and accounts for selection effects, whereby sectoral specialization reduces average physical labor productivity.

To see the relative importance of the different sources of spatial heterogeneity, we report a simple decomposition of the spatial gap displayed in Figure 6 in Table 3. In the first row we report the implied spatial gap from our baseline model. As shown in Figure 6, our model predicts the relative value added share in the agricultural sector to be between 60% and 70%. The remaining rows calculate this gap when shut down the variation in the different regional fundamentals.

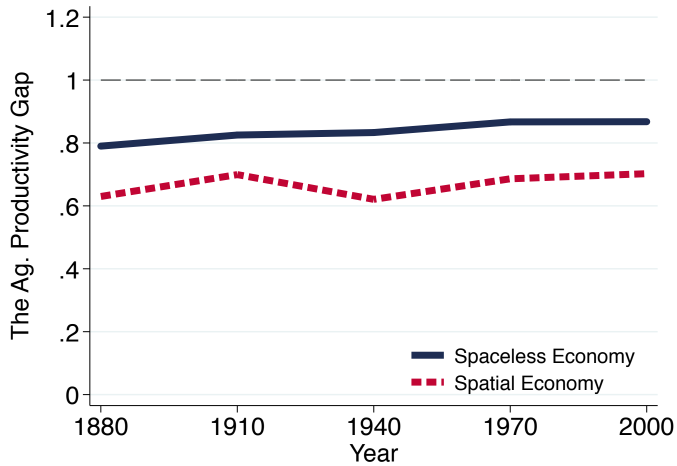
The Importance of Spatial Heterogeneity

In Section 4 we showed that for a special case of our model there would be *no* agricultural productivity gap in the absence of space. In particular, if there were only differences in absolute advantage between workers so that skills were "sector-neutral", value added per worker would be equalized across sectors. Hence, in this case, our model reduces to a standard macroeconomic model of the structural transformation.

Quantitatively, we find that skills are not sector-neutral. In particular, skilled individuals are *more* likely to work in the manufacturing sector. Such sectoral sorting, would also generate an agricultural productivity gap in the "space-less economy", where $R = 1$. In particular, the expression in (23) implies that the implied productivity gap in the space-less economy is given by

$$GAP_t^{\text{No Space}} = \frac{\frac{s_{A,t}^H}{s_{A,t}}(\lambda \times SP_t) + \frac{s_{A,t}^L}{s_{A,t}}(1-\lambda)}{(\lambda \times SP_t + (1-\lambda))}, \tag{30}$$

where SP_t is the skill-premium, i.e. average earnings of skilled individuals relative to unskilled individuals, λ is the aggregate share of skilled individuals and $s_{A,t}^j$ and $s_{A,t}$ denote the skill-specific and aggregate



Notes: The figure shows the agricultural productivity gap for the spatial model (red line) and the spaceless economy (blue line). As explained in the text, both models are calibrated to the same aggregate time-series data.

Figure 7: The Importance of Space

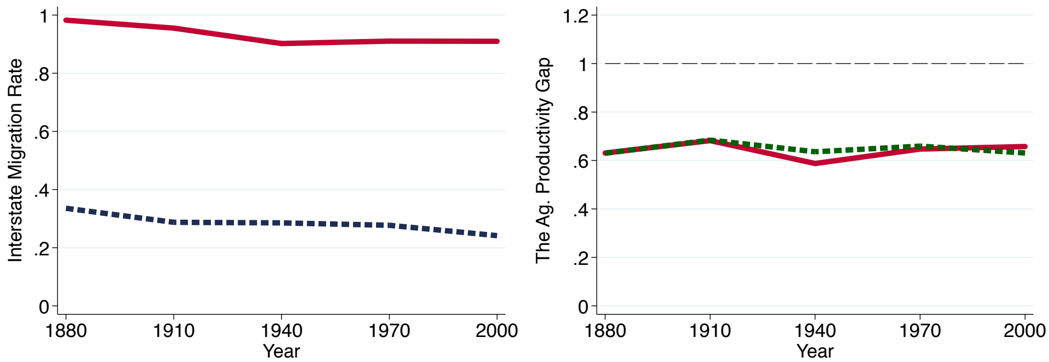
agricultural employment shares. Hence, if high skilled individuals have higher earnings $(SP_t > 1)$ and are less likely to work in agriculture, there will be a gap as relative value added per worker in agriculture is relatively low. Again that there would be no gap if sectoral employment shares were equalized across skill-groups.

It is useful to express the skill-specific relative agricultural employment shares as a function of the share of skilled individuals λ, the aggregate agricultural share $s_{A,r}$ and the manufacturing share of high-skilled individual relative to low-skilled individuals. As we show in Section 9.12 in the Appendix, this yields

$$\frac{s^L_{A,t}}{s_{A,t}} = \frac{s_{A,t} + \lambda \times (MSH_t - 1)}{(\lambda \times MSH_t + (1-\lambda))\, s_{A,t}} \text{ and } \frac{s^H_{A,t}}{s_{A,t}} = \frac{s_{A,t} - (1-\lambda)\,(MSH_t - 1)}{(\lambda \times MSH_t + (1-\lambda))\, s_{A,t}}, \tag{31}$$

where $MSH_t = s^H_{NA,t}/s^L_{NA,t}$ is the relative non-agricultural employment share of high-skilled individuals. Hence, the productivity gap in the space-less economy is fully determined by $\left\{s_{A,r}, SP_t, MSH_t\right\}_t$.

To calculate the implied spatial gap in the spaceless economy, we calculate $GAP_t^{\text{No Space}}$ given the model-implied skill premium and relative manufacturing share from our spatial economy. Hence, the spatial and the spaceless economy are calibrated to the exact same data. In Figure 7 we compare the implied productivity gap of the spatial model with the spaceless economy. The red line is again the spatial gap depicted in Figure 6. The blue line shows the productivity gap in the spaceless economy, which is calibrated to the exact same aggregate time-series moments as the spatial economy. As Figure 7 clearly shows: the implied productivity gap from the spaceless economy is only half as large as the one implied by the spatial economy. Hence, the spatial gap displayed in Figure 6 is to a large extent due to space and not only a function of the sectoral skill-composition.



Notes: In the left panel we show the interstate migration rate in the baseline model (blue line) and in an economy without costs of spatial mobility (red line). In the left panel we display the spatial gap in the baseline model (blue line) and in an economy without costs of spatial mobility (red line).

Figure 8: The (Non)-Importance of Spatial Mobility Costs

6.2 The Importance of Spatial Moving Costs

The results above highlight that differences in factor prices across space are highly persistent, i.e. that there is little spatial arbitrage between agricultural-intensive and manufacturing-intensive locations. A natural culprit of such wage gaps are the existence of moving costs, MC_{jr}. To analyze to what extent such moving costs are the fundamental driver of the low productivity in agriculture, we now consider a counterfactual exercise and calculate the agricultural productivity gap, if spatial mobility was free (i.e. $\tau = \delta = 0$).

Doing so naturally increases the amount of spatial mobility substantially. In particular, the share of individuals, who do *not* change their location after birth drops essentially to zero. This is natural: in the presence of idiosyncratic regional taste shocks and with more than 700 locations to choose from, it is very unlikely that any given individual considers her original location the preferred one. One implication of these patterns of spatial mobility is a sharp increase in the interstate migration rate. In the left panel of Figure 8, we show the time-series of the interstate migration rate in the baseline economy (blue line) and in counterfactual economy without moving costs - if there were no moving costs, the interstate migration rate was close to 100%.

In the right panel of Figure 8 we display the implied spatial gap. Again we superimpose the results of the baseline calibration for comparison. These lines are strikingly similar. Hence, the sharp increase in spatial mobility does *not* raises relative agricultural productivity. This was only the case if spatial net flows tended to be negatively correlated with agricultural employment as such flows will increase agricultural wages relative to non-agricultural wages holding the fundamentals fixed. In our counterfactual economy with free mobility this does not occur in equilibrium.

Why do not lower moving costs reduce the spatial gap?

The underlying explanation for the patterns shown in Figure 8 is simple: in our economy (as apparently in the data) there are ample reasons move *other* than wage differences induced by differences in the sectoral structure. In particular, idiosyncratic tastes and regional amenities are powerful determinants of individual location choices. If mobility was costless, these consideration would still be powerful and dissociate the equilibrium net flow of people from the extent of agricultural specialization.

To see this more clearly, consider a streamlined version of our model. In particular, suppose (i) that there are no amenity differences ($A_{rt} = 0$ for all r,t), (ii) that there is no skill heterogeneity ($q = \mu = 1$), (iii) that regional productivity is constant at its 1880 level ($Q_{s,r,t} = Q_{s,r,1880}$) and (iv) that moving costs are only parametrized by the fixed costs of moving τ ($\delta = 0$). This structure implies that individuals' mobility incentives are determined from the fixed costs of moving (τ) and the dispersion of idiosyncratic preference shocks (κ). While the former parametrizes the ease of moving, the latter determines the direction of flows conditional on moving, i.e. the correlation between wage differences and regional flows.[20]

In Figure 9 we show the implied spatial gap for the case of free mobility (i.e. $\tau = 0$) and the estimated value of the fixed moving costs if individuals mostly care about earnings rather than locations per se, i.e. for $\kappa \approx 0$. Figure 9 clearly shows that in such environments, spatial mobility *is* an arbitraging force. If mobility was free, the agricultural gap would indeed decline. In fact, without spatial mobility acting as a balancing force, relative agricultural productivity would *decrease* along the structural transformation. This pattern is closely related to the pattern in the spaceless economy shown in Figure 7 above. As the structural transformation in itself is an aggregate force that puts pressure on agricultural regions through a decline in the demand for agricultural products, spatial mobility is an important margin for the aggregate economy to adjust.

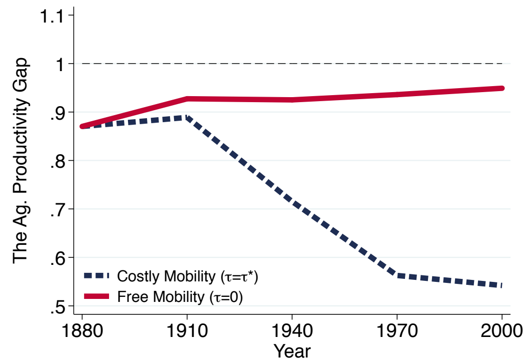
However, empirically, individuals have locational preferences other than regional factor prices and our structural estimation reveals that these considerations are important. Hence, the empirically relevant case is the one of $\kappa > 0$. If, however, such considerations are important, spatial mobility is no longer an arbitraging force. To see this concretely, consider the case of Memphis County. In 1880, Memphis County has a workforce of about 800.000 individuals, 75% of which worked in the agricultural sector. At the time, this put Memphis County at 67 percentile of the distribution of agricultural employment shares in the US.

How do migration patterns of the model-residents in Memphis county in 1880 look like? In Figure 10

[20] To see this formally, note that the share of people from skill group h moving from j to r, ρ^h_{jr}, satisfies

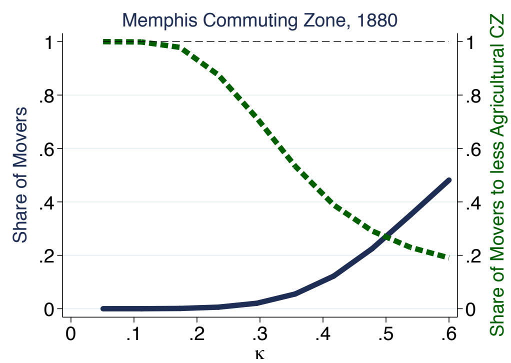
$$\frac{\rho_{jr}}{1-\rho_{jj}} = \frac{exp\left(\frac{1}{\kappa}\left(\frac{\Gamma_{\eta/\zeta}}{\eta}\psi\left(r_{t+1}\right)^{\eta-1}\left(\Theta^h_r\right)^{\eta}\right)\right)}{\sum_{l\neq r} exp\left(\frac{1}{\kappa}\left(\frac{\Gamma_{\eta/\zeta}}{\eta}\psi\left(r_{t+1}\right)^{\eta-1}\left(\Theta^h_l\right)^{\eta}\right)\right)},$$

i.e. the probability of moving from j to r conditional only depends on κ.



Notes: This figure shows the implied spatial gap in the simplified model for a high and low level of moving costs τ

Figure 9: Arbitrage if Spatial Tastes are Unimportant ($\kappa \approx 0$)



Notes: This figure shows the share of people moving out of Memphis county (blue line, left axis) and the share of movers who move to regions with an agricultural share exceeding the one in Daniels count (green line, right axis).

Figure 10: Mobility Out of Memphis County in 1880

we depict these patterns as a function of the importance of the idiosyncratic shock κ for the estimated level of fixed costs τ^*. The blue line displays the share of young residents of Memphis County who move to a new location. It is clearly seen that mobility out of Memphis County increases in the dispersion of idiosyncratic tastes. At the same time, less movers from Memphis County actually move towards regions, with a smaller agricultural sector. As the red line shows, for small levels of κ, everyone who leaves Memphis County heads towards a less agricultural location. As κ increases, more people move *and* more of these actually relocate to regions with a higher agricultural employment share. Our structural estimates suggest that the US in the 20th century resembled much more a high-κworld, where the correlation between net population flows and agricultural employment shares is small. In such a world, spatial arbitrage is limited and lower costs of moving are unlikely generate large economic gains of reallocation.

7 Conclusion

The process of structural change is characterized by two salient regularities: a marked decline in the agricultural employment share and a persistent agricultural productivity gap, whereby agricultural value added per worker is consistently lower than in the non-agricultural sector. These facts are inconsistent with the vast majority of models of structural change. Moreover, the persistence of the agricultural productivity gap despite the fact that the structural transformation induces a large reallocation of production factors, suggests that sectoral frictions might not be the major culprit.

In this paper, we argue that the agricultural productivity gap could in fact be a spatial gap. If regions differ in their comparative advantage, individuals need to spatially relocate as the economy develops and aggregate spendings shifts away from agriculture. If this process of spatial reallocation is costly, spatial wage gaps emerge. In particular, manufacturing-intensive regions will have to pay higher wages to attract agricultural workers from afar. This spatial correlation between agricultural specialization and equilibrium directly implies a sectoral productivity gap. Moreover, if agricultural specialization is only weakly correlated with net population outflows, this spatial agricultural productivity might be very persistent, even though the marginal product of labor might be equalized across sectors within a location at each point in time.

We show that this mechanism can go a long way to explain the agricultural productivity gap quantitatively. On the theoretical side, we introduce a novel model of spatial structural change by embedding an economic geography model with costly labor mobility into a dynamic, neoclassical model of the structural transformation. We rely on a non-homothetic demand system of the PIGL class as considered in Boppart (2014). Combined with the canonical Frechet-structure of skill heterogeneity, this preference structure remains highly tractable, despite falling outside the Gorman class.

As our application, we study the growth experience of the US between 1880 and 2000. Using a novel dataset on earnings, sectoral employment shares and the number of workers across US commuting zones, we find that the model can comfortably explain 50% of the agricultural productivity gap observed in the aggregate data. Because the sectoral reallocation of production factors takes place almost entirely within localities, this implies that the macroeconomic gains from worker reallocation are much more modest than the agricultural productivity gap as inferred from aggregate data suggests.

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 Markus Poschke, "Structural change out of agriculture: Labor push versus labor pull," *American Economic Journal: Macroeconomics*, 2011, *3* (3), 127–158.

_ 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.

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.

—, David Krisztian Nagy, and Esteban Rossi-Hansberg, "The Geography of Delevopment: Evaluating Migration Restrictions and Coastal Flooding," 2015. Working Paper.

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.

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

Herrendorf, Berthold, 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.

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.

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

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.

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.

8 Appendix

8.1 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 (33) 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

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

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, (7) 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

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

This proves the results for Lemma 1.

8.2 Consumption expenditure shares

We can derive the expenditure shares from the indirect utility function $V_{jr}(e, p)$ in (10) from Roy's identity. The indirect utility function is defined by

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

Hence,

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

The expenditure function is given by

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

Hence,

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

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

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

Hence,

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

Rearranging terms yields

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

The expenditure share on good j is therefore given by

$$\vartheta_j\left(e,p\right)=x_j^M\left(p,e\right)\times\frac{p_j}{e}=-\frac{\frac{\partial V\left(p,e\left(p,u\right)\right)}{\partial p_j}p_j}{\frac{\partial V\left(p,e\left(p,u\right)\right)}{\partial e}e}. \tag{32}$$

We consider an indirect utility function from the PIGL class

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

We get that

$$\begin{aligned}\frac{\partial V_{rj}}{\partial e} &= e^{\varepsilon-1}B\left(p\right)^{-\varepsilon} \\ \frac{\partial V_{rj}}{\partial p_j} &= -e^{\varepsilon}B\left(p\right)^{-\varepsilon-1}\frac{\partial B\left(p\right)}{\partial p_j}+\frac{\partial C\left(p\right)}{\partial p_j}.\end{aligned} \tag{33}$$

(32) therefore implies that

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

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,

$$x_A(e,p) = \phi\frac{e}{p_A} + \nu\left(\frac{p_A}{p_M}\right)^{\gamma}\frac{1}{p_A}\left(p_A^{\phi}p_M^{1-\phi}\right)^{\eta}e^{1-\eta}. \tag{34}$$

The expenditure share is

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

8.3 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

$$y^i = \max_s \left\{w_{s,r}z_s^i\right\}.$$

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

$$F_{z_s^i}^h(z) = e^{-\Psi_s^h \times z^{-\zeta}}, \tag{35}$$

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 $y_{sr}^i \equiv w_{s,r} \times z_s^i$ is also Frechet and given by

$$F_{y_{sr}^i}(y) = P\left(z \leq \frac{y}{w_{s,r}}\right) = e^{-\Psi_s^h w_{s,r}^{\zeta} \times y^{-\zeta}}.$$

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

$$F_{y_r^i}^h(y) = e^{-\left(\Theta_r^h\right)^{\zeta} \times y^{-\zeta}} = e^{-\left(\frac{y}{\Theta_r^h}\right)^{-\zeta}} \tag{36}$$

where

$$\Theta_r^h = \left(\sum_s \Psi_s^h w_{s,r}^{\zeta}\right)^{1/\zeta} = \left(\Psi_A^h w_{A,r}^{\zeta} + \Psi_{NA}^h w_{NA,r}^{\zeta}\right)^{1/\zeta}.$$

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

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

From (36) we can derive the distribution of $y^{1-\eta}$. As $\eta < 1$, we have that

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

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

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

8.4 Agricultural Demand

In the main text we derived

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

where

$$E_t^{h,Y} = \psi_{t+1} \times \Gamma\left(1 - \frac{1}{\zeta}\right) \Theta_{r,t}^h = \psi_{t+1} \times Y_{rt}^h.$$

it is also that case that

$$E_t^{h,O} = (1+r_t)(1-\psi_t) \times Y_{rt-1}^h = (1+r_t)(1-\psi_t)\frac{1}{\psi_t} \times E_{t-1}^{h,Y}.$$

Hence, note that

$$\begin{aligned}
&\frac{\sum_{r,h}\left(\left(E_{r,t}^{h,Y}\right)^{1-\eta}\lambda_{r,t}^h L_{r,t} + \left(E_{r,t}^{h,O}\right)^{1-\eta}\lambda_{r,t-1}^h L_{r,t-1}\right)}{\frac{1}{1-\alpha}\times\sum_{r,h} Y_{rt}^h \lambda_{r,t}^h L_{r,t}} \\
= & \frac{1}{(1-\alpha)^{-1}}\left(\psi_{t+1}^{1-\eta}\sum_{r,h}\left(Y_{rt}^h\right)^{-\eta}\frac{Y_{rt}^h\lambda_{r,t}^h L_{r,t}}{\sum_{r,h} Y_{rt}^h\lambda_{r,t}^h L_{r,t}} + \left[(1+r_t)(1-\psi_t)\right]^{1-\eta}\sum_{r,h}\left(Y_{rt-1}^h\right)^{-\eta}\frac{Y_{rt-1}^h\lambda_{r,t-1}^h L_{r,t-1}}{\sum_{r,h} Y_{rt}^h\lambda_{r,t}^h L_{r,t}}\right) \\
= & \frac{1}{(1-\alpha)^{-1}}\left(\psi_{t+1}^{1-\eta}\sum_{r,h}\left(Y_{rt}^h\right)^{-\eta}\frac{Y_{rt}^h\lambda_{r,t}^h L_{r,t}}{\sum_{r,h} Y_{rt}^h\lambda_{r,t}^h L_{r,t}} + \left[(1+r_t)(1-\psi_t)\right]^{1-\eta}(1+g_t)^{-1}\sum_{r,h}\left(Y_{rt-1}^h\right)^{-\eta}\frac{Y_{rt-1}^h\lambda_{r,t-1}^h L_{r,t-1}}{\sum_{r,h} Y_{rt-1}^h\lambda_{r,t-1}^h L_{r,t-1}}\right) \\
= & \frac{1}{(1-\alpha)^{-1}}\left(\psi_{t+1}^{1-\eta}\sum_{r,h}\left(Y_{rt}^h\right)^{-\eta}\omega_{rt}^h + \left[(1+r_t)(1-\psi_t)\right]^{1-\eta}(1+g_t)^{-1}\sum_{r,h}\left(Y_{rt-1}^h\right)^{-\eta}\omega_{rt-1}^h\right).
\end{aligned}$$

This is the expression in the main text.

8.5 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

$$R_t = \frac{\alpha PY_t}{K_t} = \alpha\frac{\frac{1}{1-\alpha}\sum_{r=1}^R w_{r,t}L_{r,t}}{K_t}. \tag{37}$$

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 (16) is constant and given by

$$\psi = \left(1+\beta^{\frac{1}{1-\eta}}(1+r)^{\frac{\eta}{1-\eta}}\right)^{-1}. \tag{38}$$

To solve for the interest rate, note that (22) and (37) imply that

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

A constant capital output ratio implies that

$$\frac{K_{t+1}}{K_t} = \frac{PY_t}{PY_t} = 1+g,$$

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

$$1+g=(1-\psi)\frac{(1-\alpha)}{\alpha}(r+\delta)=\left(\frac{\beta^{\frac{1}{1-\eta}}(1+r)^{\frac{\eta}{1-\eta}}}{1+\beta^{\frac{1}{1-\eta}}(1+r)^{\frac{\eta}{1-\eta}}}\right)\frac{(1-\alpha)}{\alpha}(r+\delta). \tag{39}$$

This equation determines r as a function of parameters.

Along the BGP the consumption and investment rate is equal to

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

Using (39) 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 (39) we also get that

$$\psi=1-\frac{\alpha}{1-\alpha}\frac{1+g}{R}. \tag{40}$$

8.6 Regularity Conditions for the preferences to be well-defined

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

$$\vartheta_A(e,p)\geq 0 \Rightarrow e^{\eta}\geq -\frac{\nu}{\phi}\left(\frac{p_A}{p_M}\right)^{\gamma}, \tag{41}$$

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 (41) 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 10.1 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

$$\begin{aligned} e^{\eta} &\geq \frac{\nu}{1-\phi}\left(\frac{p_A}{p_M}\right)^{\gamma} && (42)\\ (1-2\phi-\gamma)\,e^{\eta}+\frac{(1-\phi)\,\phi}{\nu}\left(\frac{p_A}{p_M}\right)^{-\gamma}e^{2\eta} &\geq \nu\,(1-\eta)\left(\frac{p_A}{p_M}\right)^{\gamma}. && (43)\end{aligned}$$

Lemma 3. *A sufficient condition for (42) to be satisfied is that (43) holds and that*

$$\gamma > (1-\phi)\,\eta. \tag{44}$$

Proof. Equation (43) can be written as

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

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 $h(x)=(x-1)+(\phi x-(2\phi+\gamma))\,x$. Hence, 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 (43) to be satisfied, it has to be the case that $x>x^*=1-\frac{1-\gamma}{2\phi}$. Hence, condition (43) implies (42) if

and

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

Rearranging terms yields $(1-\phi)\,\eta<\gamma$, which is (44). □

Hence, the preferences are well defined as long (43) is satisfied and (44) holds. Because the RHS of (43) is increasing in e in the relevant range, i.e. as long as (43) 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 (36) 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

$$\left(\Theta_r^h \psi_{t+1}\right)^{\zeta} = \left(\Psi_A^h w_{A,r}^{\zeta} + \Psi_{NA}^h w_{NA,r}^{\zeta}\right) \times \psi_{t+1}^{\zeta} > \left(\Psi_A^h w_{A,r}^{\zeta} + \Psi_{NA}^h w_{NA,r}^{\zeta}\right) \times \left(\frac{1}{\eta}\left(1+\beta^{\frac{1}{1-\eta}}\right)^{1-\eta}\right)^{\zeta}.$$

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

8.7 Agricultural Specialization and Population growth

In Figure 2 we showed that the decline in the agricultural employment share in not driven by population mobility away from agricultural locations. In this section we provide further evidence for this (non-)relationship between past agricultural specialization and subsequent population growth. In particular, we consider a regression of the form

$$g_{L,r}^{1880-2000} = \alpha + \beta \times s_{A,r}^{1880} + u_r,$$

where $g_{L,r}^{1880-2000}$ denotes regional population growth between 1880 and 2000 and $s_{A,r}^{1880}$ denotes the agricultural employment share in 1880. The results are contained in Table 4. 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 4 we include a whole 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 11 shows this relationship graphically. 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.

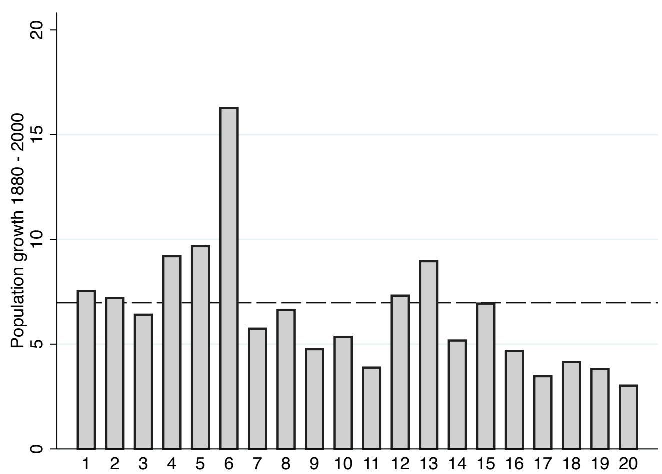
8.8 Urbanization within and across commuting zones

In Figure 2 we showed that the secular decrease in agricultural employment is a *within* commuting zone phenomenon. The same pattern holds true for the process of urbanization. The left panel in Figure 12 shows the increase in urbanization since 1880. The share 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

	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

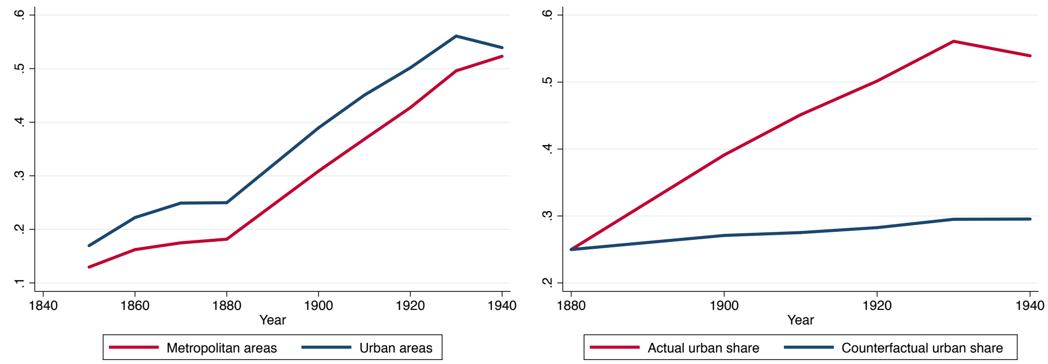
Notes: Robust standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. Column 4 contains a whole set of 20 fixed effects for the different quantiles of agricultural employment shares.

Table 4: Agricultural Specialization in 1880 and Population growth 1880-2000



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 11: Agricultural Specialization in 1880 and Population growth 1880-2000



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 $u_t^{CF} = \sum_r u_{r,1880} \times \frac{L_{rt}}{\sum_r L_{rt}}$, where $u_{r,1880}$ is the urban share in 1880.

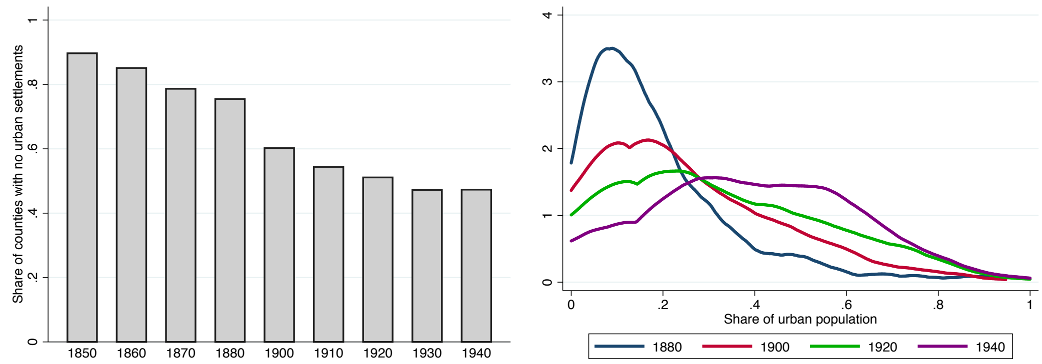
Figure 12: Urbanization within and across commuting zones

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 counterfactual urbanization as

$$u_t^{CF} = \sum_r u_{r,1880} \times \frac{L_{rt}}{\sum_r L_{rt}},$$

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. Figure 12 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.

To see this within-commuting zone patter of urbanization more directly, consider Figure 13. In the left panel we depict 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 2 these densities shift to the right, indicating that urbanization takes place in all regions 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 13: Urbanization within commuting zones

9 Online Appendix

9.1 Additional derivations on the model with selection

In this section we derive some other properties of the selection model, which are convenient. Recall from (3) that the distribution of individual skills is given by

$$F_{z_s^i}^h(z) = e^{-\Psi_s^h \times z^{-\zeta}}, \tag{45}$$

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*

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

Then

$$E\left[x_i | x_i = \max_i [x_i]\right] = \Gamma\left(1 - \frac{1}{\zeta}\right) \times \left(\sum_{i=1}^n A_i\right)^{1/\zeta}. \tag{46}$$

Note that this object does not depends on i.

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

$$F(x_1, x_2, ..., x_3) = \prod_{j=1}^n F(x_i) \tag{47}$$

because of independence. Hence, we get that

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

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

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

Using (47), we get that

$$
\begin{aligned}
\frac{\partial F(x_1, x_2, ..., x_3)}{\partial x_1} &= \frac{\partial F(x_1)}{\partial x_1} \times \left(\prod_{j=2}^n F(x_i)\right) = e^{-A_1 \times x^{-\zeta}} A_1 \zeta x_1^{-\zeta-1} \times \left(\prod_{j=2}^n e^{-A_i \times x_j^{-\zeta}}\right) \\
&= A_1 \zeta x_1^{-\zeta-1} \times \left(\prod_{j=1}^n e^{-A_i \times x_i^{-\zeta}}\right) \\
&= A_1 \zeta x_1^{-\zeta-1} \times e^{-\Sigma_i^n A_i \times x_i^{-\zeta}}.
\end{aligned}
$$

Hence,

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

Substituting this above yields

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

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

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

This implies (46) $\square$

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

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

Hence, Lemma 4 implies that

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

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

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

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

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

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

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

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

$$s_{sr}^{h} = \Psi_{s}^{h} \times \left(\frac{w_{rs}}{\Theta_{r}^{h}}\right)^{\zeta},$$

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

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

Note that we can write

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

Hence,

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

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

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

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

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