Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 144 (2009) 2477–2502
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

Spatial growth and industry age

Klaus Desmet [a], Esteban Rossi-Hansberg [b,*]

[a] *Universidad Carlos III, Department of Economics, Getafe (Madrid), Spain*
[b] *Princeton University, Department of Economics and WWS, Princeton, United States*

Received 27 July 2007; final version received 28 July 2008; accepted 20 November 2008

Available online 28 November 2008

Abstract

Between 1970 and 2000 employment growth across U.S. counties exhibited very different patterns in manufacturing and services. Whereas manufacturing employment growth was negatively related to initial manufacturing employment across the entire distribution of counties, service employment growth was positively related to initial service employment for intermediate sized counties. This paper presents a theory to rationalize these facts. Local sectoral growth is driven by technological diffusion across space and depends on the age of the sector. The theory correctly predicts the relation between county employment growth and initial county employment in manufacturing at the turn of the 20th century.

JEL classification: E2; O3; O4; R1

Keywords: Spatial growth; Industry age; Technology diffusion; Spillovers; Distribution of economic activity; Scale-dependent growth

1. Introduction

We start this paper with one observation: between 1970 and 2000 the scale dependence of U.S. county employment growth exhibited very different patterns in manufacturing and services (see also Desmet and Fafchamps [11]). County employment growth in manufacturing between 1970 and 2000 decreased with the level of manufacturing employment in 1970. Fig. 1 shows

* Corresponding author.
E-mail addresses: klaus.desmet@uc3m.es (K. Desmet), erossi@princeton.edu (E. Rossi-Hansberg).

0022-0531/$ – see front matter
doi:10.1016/j.jet.2008.11.001

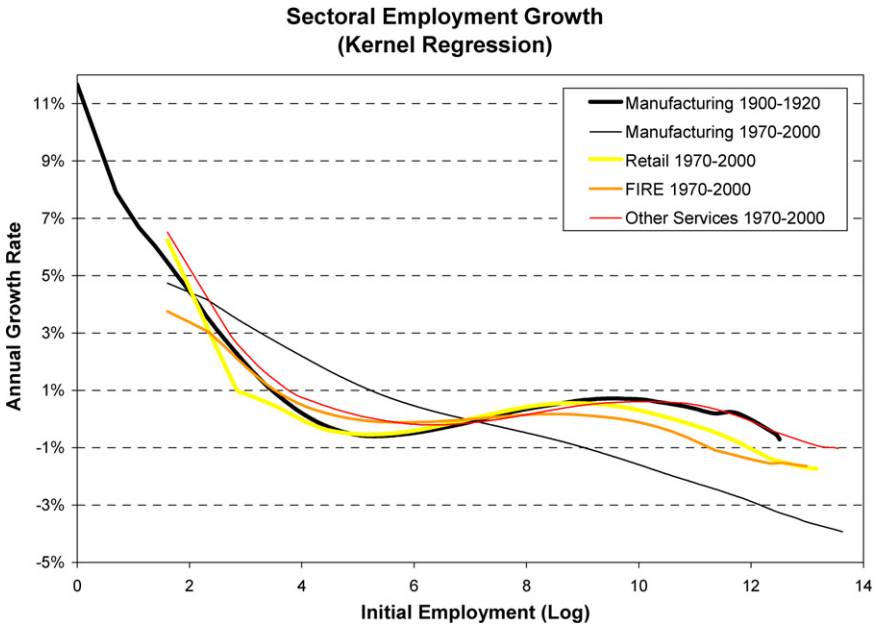


Fig. 1.

a non-parametric estimate of this relation[1]: it is clearly negative across the entire distribution. Small counties, in terms of manufacturing employment, grew faster than large ones. This implies that manufacturing employment dispersed in space. In contrast, when we look at services in Fig. 1, we see an S-shaped pattern. The relation between service employment growth between 1970 and 2000 and service employment in 1970 was negative for small counties, reversed and became positive for intermediate counties, and then turned negative again at the high end of the distribution.

What can account for this striking difference in the evolution of the distribution of economic activity across industries? Our take is that the age of an industry — the time elapsed since the last general purpose technology (GPT) affected the industry — can explain this disparity across manufacturing and services.

In a mature industry the benefits from agglomeration have been mostly exploited, and knowledge spillovers have lost much of their importance (Duranton and Puga [12]). This makes concentration of economic activity costly if land is a factor of production. Hence, as spillovers weaken, small counties grow faster than large ones. This is consistent with the manufacturing experience between 1970 and 2000. As argued by Jovanovic and Rousseau [23] and David and Wright [10], the last major GPT that mainly affected manufacturing was electricity and dates back to the turn of the 20th century. By the middle of the century this technology had been implemented and manufacturing had become mature. This explains the downward sloping pattern in manufacturing for the period 1970–2000 in Fig. 1.

[1] In particular, we use a non-linear kernel to regress detrended annual employment growth between 1970 and 2000 on the log of employment in 1970. Details of the calculation of these growth rates, as well as confidence intervals, are presented in Section 2.

In contrast, the other major GPT of the 20th century — information and communication technology — started somewhere in the middle of the 1970s and is having its greatest impact on services (Hobijn and Jovanovic [21]). Using our definition, this has made the service industry young. Many firms are still adapting to the new technological paradigm, and much of its potential has yet to be exploited. Because of trade, some counties with initially very low service employment shift away from manufacturing into services, as their pattern of specialization changes. Starting off with little service employment, those counties experience high growth in service jobs. Counties with intermediate levels of service employment behave differently. Instead of switching specialization patterns, services agglomerate to take advantage of the benefits from spatial concentration. As a result, larger counties, in terms of service employment, grow faster than smaller ones. Although the largest counties continue to benefit from this, they also suffer from increasing land congestion, so that dispersion re-emerges. The result is the observed S-shaped pattern in Fig. 1 for the different service industries between 1970 and 2000.

If industry age is the key determinant of the scale dependence of an industry's employment growth, then the pattern of manufacturing employment growth at the turn of the 20th century, when manufacturing was young due to the electricity revolution, should look similar to the pattern we observe in the service industry at the end of the 20th century. This prediction is borne out in the data. Fig. 1 shows that the shape of manufacturing growth for 1900–1920 is almost identical to the shape of service employment growth for 1970–2000.[2] Both sectors exhibit the same type of S-shaped pattern for these very different periods in time. This novel finding is what underlies and motivates our theoretical model.

We present a two-sector spatial theory that formalizes the link between the age of an industry and its spatial growth pattern. The theory inherits many characteristics present in Lucas and Rossi-Hansberg [25] and Rossi-Hansberg [30]. The main difference is that our theory adds technological change through a combination of spatial knowledge spillovers and technology diffusion. Regions that do not benefit from strong enough knowledge spillovers obtain the best technologies invented in more dense areas through diffusion.[3] This process of technological change, together with the presence of land as a factor of production, gives rise to three forces: technology diffusion leads to dispersion; knowledge spillovers favor concentration; and land congestion implies dispersion. In young industries, diffusion dominates in low employment areas, knowledge spillovers dominate in medium employment areas, and congestion dominates in high employment areas. This explains the observed S-shaped growth pattern in young industries. In contrast, in mature industries knowledge spillovers are weaker, so that dispersion dominates across the entire distribution. This explains the observed downward sloping pattern in mature industries.

To understand the logic of the theoretical model, it is useful to tie it back to the differing growth patterns across time and sectors of Fig. 1. At the turn of the 20th century, manufacturing in the U.S. experienced rapid innovation, prompted by the advent of electricity. Knowledge

[2] Since we detrend all data and growth rates are annual, all curves are comparable even though they include intervals of different lengths.

[3] There has been little work in dynamic models of spatial industry location. Rossi-Hansberg [29] introduces a similar framework with capital accumulation. Holmes [22] provides a dynamic model where cluster location changes across time as firms take advantage of location-specific factors. There are a variety of frameworks that study the distribution of city sizes using dynamic models (Gabaix [16]; Duranton [13]; Rossi-Hansberg and Wright [31]) or focusing on rural to urban migration in a dynamic setting (Lucas [26]; Henderson [19]). But none of these papers uses a spatial theory and so they have no prediction for sectoral employment growth across regions.

Henderson and Venables [20] present a dynamic model of city evolution but, likewise, without a spatial component.

spillovers made geographic concentration of manufacturing employment useful. In contrast, by the end of the century, manufacturing had matured and standardized. There were fewer benefits to be reaped from agglomeration. At the same time, however, knowledge spillovers gained in importance in the service industry, as improvements in information and communication technology caused a wave of product and process innovation in that sector. The similarity between manufacturing at the turn of the 20th century and services 70 years later is the result of innovation and spillovers being important for these young industries. In contrast, these effects had lost much of their importance for manufacturing at the end of the 20th century.

The empirical literature on technology diffusion goes back to the seminal work of Griliches [17], who studied the spread of the use of hybrid corn in the U.S. In the specific case of diffusion of GPTs, not much attention has been given to the spatial dimension. Rosenberg and Trajtenberg [28] describe how the replacement of waterwheels by steam engines allowed manufacturing activity to relocate from rural to urban areas. By removing the geographic constraint of proximity to water, firms could move to densely populated areas, where they could take advantage of agglomeration economies. Whether this clustering had anything to do with knowledge spillovers is unclear, though. In the case of the Internet, Forman, Goldfarb, and Greenstein [14] find evidence pointing in that direction. While mere participation in the Internet spread rapidly across locations, the more complex applications, such as e-commerce, located predominantly in urban areas, where there was access to complementary inventions and activities.

A further question of relevance raised by our work is the way in which IT is similar to electricity. Hobijn and Jovanovic [21] and Jovanovic and Rousseau [23] have pointed out the many similarities between both GPTs, ranging from their diffusion patterns to the behavior of IPOs, patents and the stock market. In the midst of disappointment about the computer revolution, David [9] used the experience of the electric dynamo to argue that there was nothing surprising about the productivity slowdown paradox. In a recent contribution Atkeson and Kehoe [1] are more cautious. Based on a calibrated model of the electricity revolution, they find that slow diffusion depends crucially on agents having built up a large stock of knowledge about the old technology. For lack of data on the IT era, they conclude that it remains to be seen whether the computer will be a simple replay of the dynamo. Our paper suggests that at least along the dimension of spatial growth, electricity and IT exhibit similar behavior.

The rest of the paper is organized as follows. Section 2 provides evidence on scale dependence in growth rates. Section 3 presents the model, and Section 4 gives numerical results that characterize an equilibrium. Section 5 concludes.

2. Empirical evidence

In this section we document that spatial employment growth follows different patterns for young and mature industries. In young industries spatial employment growth is non-monotonic: locations with low or high employment exhibit employment growth rates that decline with initial employment size (spatial convergence or deconcentration), whereas locations with intermediate levels of employment exhibit employment growth rates that increase with initial employment size (spatial divergence or concentration). In mature industries spatial employment growth becomes monotonic: all locations, independent of their size, exhibit growth rates that decline with initial employment size (spatial convergence or deconcentration).

2.1. Industry age

We define an industry as young when it benefits from the diffusion of a new technology and knowledge spillovers are strong. In contrast, we define an industry as mature when the technology it uses has become standardized and knowledge spillovers have lost much of their importance. By a new technology we do not mean a marginal improvement over an already existing technology, but rather a radical innovation that represents a drastic change from the previously used technology. Our notion is akin to the introduction of a new general purpose technology (GPT).

To determine whether a given industry is young at a given point in time, we therefore rely on the literature on GPTs. Jovanovic and Rousseau [23] and David and Wright [10] argue that the two major GPTs in the 20th century were electricity and information technology (IT). Jovanovic and Rousseau [23] define the starting point of a GPT as the date at which it reaches 1% diffusion in the median sector. This gives a date of 1894 for electricity, coinciding with the first hydroelectric facility at Niagara Falls, and a date of 1971 for IT, coinciding with Intel's 4004 microprocessor. As further evidence for the starting point of IT, Hobijn and Jovanovic [21] argue that the stock market decline in the early 1970s coincided with the arrival of 'good news' about IT. In particular, they find that stock prices dropped most in those sectors that had the largest post-1973 investments in IT, since in those sectors part of the capital stock became obsolete. As the ending point of a GPT, Jovanovic and Rousseau [23] use the date at which the diffusion curve flattens. This gives a date of 1929 for electricity, whereas, for IT, the curve has not plateaued yet. Focusing on the first and the last decades of the 20th century as the two time periods of major technological change is consistent with many other events associated with GPTs. For example, in the 20th century there were two surges in patents and trademarks, the first one between 1900 and 1930 and the second one after 1977. Similarly, IPOs increased between 1895 and 1929 and again after 1977.

This evidence suggests that the diffusion of electricity started in earnest somewhere between 1894 and 1900 and had ended by 1930, whereas the diffusion of IT started some time between 1971 and 1977, with the end still not in sight. Given that the diffusion of IT is still ongoing, we cannot compare the ending date of both GPTs. To make the timing of both technologies comparable, we focus on 1900–1920 and 1980–2000 as the periods when electricity and IT were young.

Although GPTs are pervasive in the sense that they tend to spread to the entire economy,[4] their effect may differ depending on the sector. In the case of electricity, David and Wright [10] argue that it affected mainly the manufacturing sector. In the decade after World War I, Kendrick [24] estimates that economy-wide TFP grew by 22%, whereas in manufacturing TFP grew by 76%. In the case of IT, the evidence points to the service sector being the big beneficiary. Hobijn and Jovanovic [21] compute IT intensity — the share of IT equipment in the total stock of equipment — in different sectors. In 1996 IT intensity stood at 42.4% in services, compared to 17.9% in manufacturing.[5] Within the broad category of services, the subsectors that have invested most in IT are FIRE (finance, insurance and real estate), communications, business services, and wholesale. In a growth accounting exercise of 60 industries, Bartelsman and Beaulieu [2] find that the contribution of IT to growth has been most prevalent in credit institutions. Basu and

[4] Pervasiveness is one of the fundamental characteristics of GPTs, according to the definition of Bresnahan and Trajtenberg [5].

[5] Similar results are found by Chun et al. [8], Triplett and Bosworth [32] and Basu and Fernald [3].

Fernald [3], for their part, suggest that the most IT-intensive sectors are communications, finance and insurance, and business services. Chun et al. [8] add wholesale to that list. Not all of these sectors were equally quick to adopt IT. According to Hobijn and Jovanovic [21] and Bartelsman and Beaulieu [2], the early investors in IT were FIRE and communications.

Based on the timing and the differential sectoral effects, our empirical implementation takes the following stylized view. The manufacturing sector was young in the period 1900–1920 and mature after the 1950s, whereas the service sector was mature before 1970 and became young some time in the period between 1970 and 1980. We therefore take both a time series approach (by following industries over time) and a cross-sectional approach (by comparing different industries) in studying spatial growth patterns. To guarantee that our evidence is not specific to the United States, we also analyze European data.

2.2. *Manufacturing and services in the United States*

Our empirical analysis for the United States takes counties as the unit of observation. There are several reasons for doing so. First, given our focus on spatial growth, counties provide an appropriate level of geographic detail. Second, counties cover the entire U.S., in contrast to, for instance, metropolitan areas or cities. Third, county data allow us to go sufficiently back in time. Our data on county-level employment come from a variety of sources. Data until 1930 come from the Historical Census Browser at the University of Virginia, and data from 1969 onward come from the Regional Economic Information System (REIS) compiled by the U.S. Bureau of Economic Analysis (BEA). One obvious concern is changing definitions of counties and county borders. By using information on years in which county definitions changed,[6] our regressions exclusively focus on counties whose definitions have not changed over the period of interest.[7]

Given our focus on possible non-monotonicities in spatial employment growth, we run non-linear kernel regressions of the form:

$$L_{t+s}^{i} = \phi\left(L_{t}^{i}\right) + e_{t}^{i}$$

where L_t^i is log employment in year t and county i. The estimation uses an Epanechnikov kernel with optimal bandwidth.[8] Because the distribution of employment levels is approximately log-normal, we focus on the log of employment.

To facilitate interpretation, we plot annual employment *growth* as a function of initial log employment in the same industry. In this case, a negative slope indicates deconcentration (convergence) and a positive slope indicates concentration (divergence).

Fig. 2 plots de-trended annual employment growth in the manufacturing sector for the periods 1900–1920 and 1980–2000 together with the 95% confidence intervals.[9] As can be seen, for the period 1900–1920, when electricity was diffusing and manufacturing was young, employment

[6] Information about changes in county borders come from Forstall [15].

[7] In particular, depending on the regression, we leave out counties whose borders changed after 1900 or after 1980. To make sure that the different patterns between 1900–1920 and 1980–2000 are not due to different samples of counties, we re-ran all of our regressions on the sample of counties whose borders did not change after 1900. None of the results changed.

[8] This methodology is described in detail in Desmet and Fafchamps [11].

[9] Fig. 1 shows results starting in 1970, not 1980, which illustrates that these facts are robust to the precise choice of the latter period starting date.

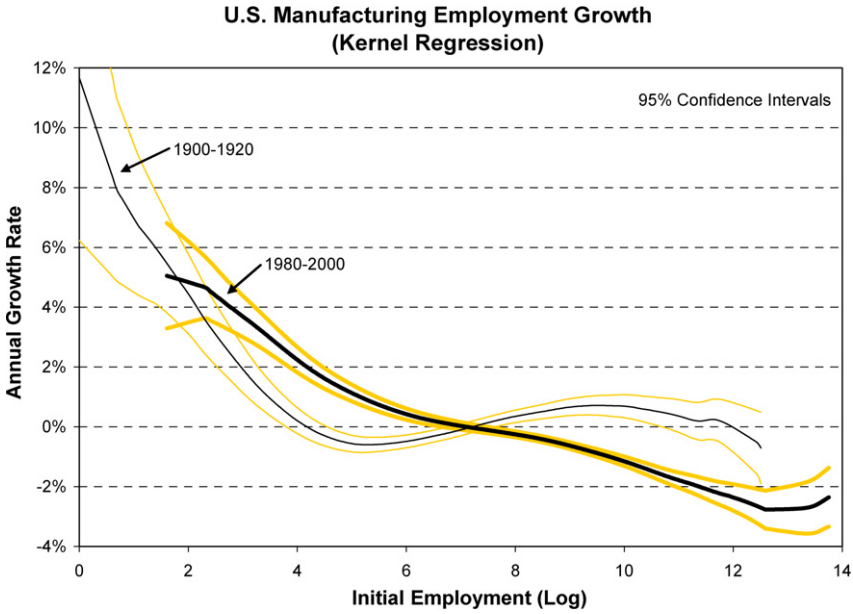


Fig. 2.

growth was non-monotonic. At that time there was deconcentration or convergence in the lower and the upper tail of the distribution and concentration or divergence in the middle part of the distribution. In contrast, for the period 1980–2000, when manufacturing was mature, there was deconcentration or convergence across the entire distribution. Though not reported in the paper, when analyzing the years between 1920 and 1980, one can observe how the S-shaped curve gradually changes into a downward sloping curve.

A picture similar to the one in manufacturing at the beginning of the 20th century emerges when analyzing the service sector at the end of the century, when it became young. Fig. 3 shows de-trended annual growth rates of employment in service industries from 1980 to 2000 as a function of employment in 1980. Comparing the growth rate in services in 1980–2000 to the same curve in manufacturing in 1900–1920 one can observe that the main difference is that small counties in manufacturing grew very fast. In 1900 there were some very small counties in terms of manufacturing employment. In contrast, in 1980 almost all counties had a basic employment level in services of about 50 employees. Apart from this, both figures exhibit the exact same pattern of scale dependence.

If the IT revolution started somewhere in the middle of the 1970s, making services young, then during the decades before the 1970s services should have been mature and thus should have exhibited negative scale dependence across the distribution. Although we do not have comparable data for that period, we checked for both retail and FIRE for the period 1950–1970 and found this prediction to hold up in the data.

One potential concern is that the patterns we describe are mostly about metropolitan counties. Maybe what we are witnessing is simply industries moving in and out of metropolitan areas. To address this concern, it is useful to separate locations between metropolitan and non-metropolitan counties as defined by the Office of Management and Budget. The main criterion for a county to

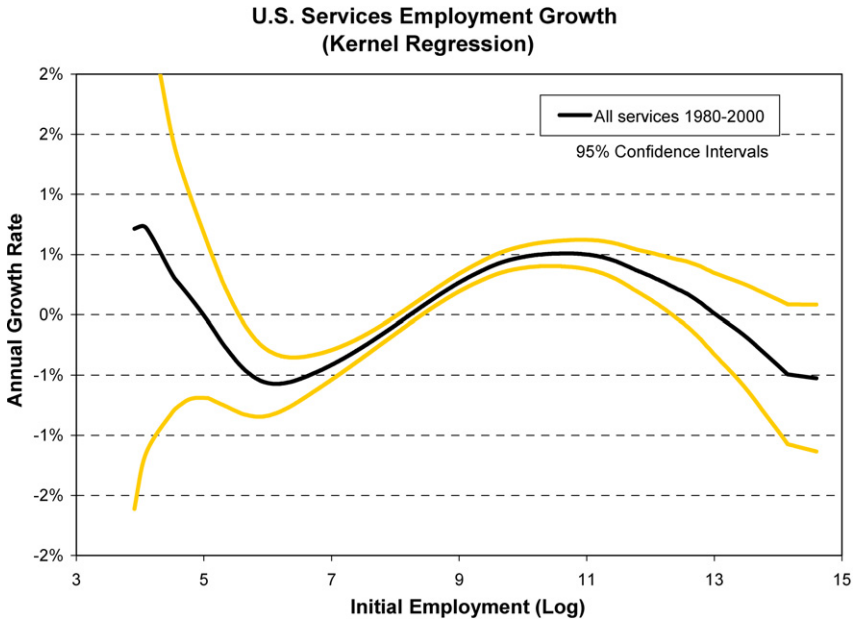


Fig. 3.

be classified as metropolitan is that it is part of an urban area of at least 50,000 residents.[10] For 1980 we find that 90% of the counties with employment in manufacturing above 15,214 (or 9.63 in logs) are metropolitan counties, whereas 90% of counties with employment in manufacturing below 4272 (or 8.36 in logs) are non-metropolitan counties. Comparing these numbers with the curve for 1980–2000 in Fig. 2 makes it clear that the pattern we document is not only a shift of manufacturing employment from cities to rural areas but a continuous dispersion throughout all county sizes. Similarly, we can compute the same thresholds for employment in services. For 1980 we find that 90% of counties with service employment above 22,248 (or 10.01 in logs) are classified as metropolitan and 90% of counties with service employment levels below 16,155 (9.69 in logs) are classified as non-metropolitan. Hence it is clear that the negative scale dependence at the top of the distribution in Fig. 3 includes only counties that form integral parts of cities. This is consistent with our argument that the negative scale dependence observed for large counties is the result of congestion costs in urban areas. In addition, the positive scale dependence in the middle part of the distribution includes both metro and non-metro counties.

In looking at Fig. 3, one may argue that although the service sector as a whole exhibits the aforementioned S-shaped pattern in 1980–2000, particular industries within the service sector may not. The empirical evidence suggests that finance, insurance and real state (FIRE) is a sector where IT has been particularly important and so we expect to see the pattern there. Other important industries are retail and other services. Fig. 4 presents the kernel regressions for these

[10] To be precise, before 2003 metro areas were defined to include central counties with one or more cities of at least 50,000 residents or with an urbanized area of 50,000 or more and total area population of at least 100,000. Outlying counties were included if they were economically tied to the central counties. See http://www.ers.usda.gov/Briefing/Rurality/NewDefinitions/ for more details.

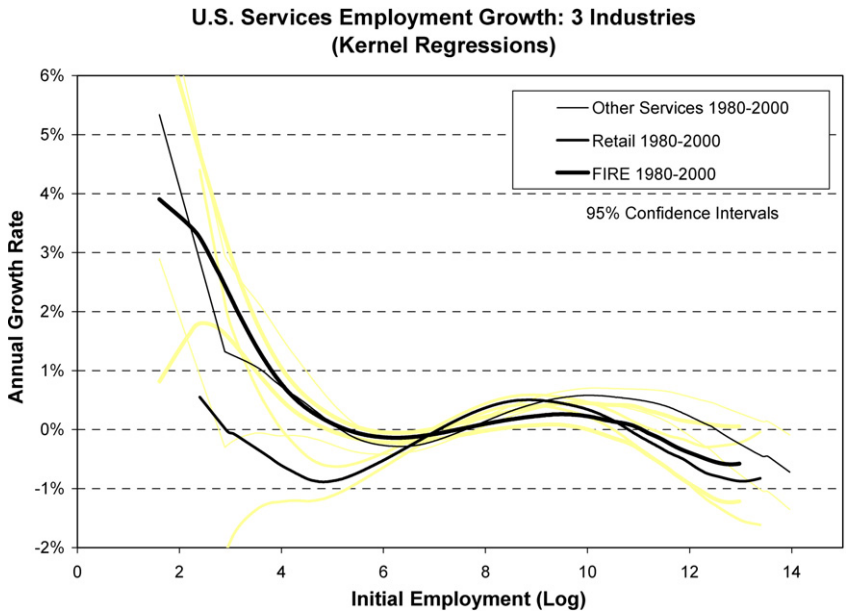


Fig. 4.

three industries. Clearly the S-shaped pattern is present in all of them, perhaps somewhat more pronounced in FIRE and other services. Unfortunately, service employment data are not available for the beginning of the 20th century, so we cannot contrast the predictions of the model for services in this earlier period.[11]

2.3. Manufacturing and services in Europe

To make sure our findings are not specific to the United States, we also analyze spatial employment growth across European regions. Data on sectoral employment come from the Cambridge Econometrics Regional Database, which covers 236 Western European regions from 1975 to 2000.

Although we are unable to study the effect of electricity, our time series is long enough to analyze the effect of IT on different sectors of the economy. Before doing so, we need to compare the European experience to the U.S., both in terms of the sectoral impact and in terms of the timing.

Regarding the sectoral impact of IT, results are similar to those in the United States. Manufacturing is not benefiting to the same extent as services. As pointed out by Basu et al. [4], although manufacturing accounts for about one-fifth of GDP, it has less than one-fifth of the computers in both the United States and the United Kingdom. Within the broad category of services, those subsectors that have invested most heavily in IT in the U.K. are finance/insurance, business ser-

[11] All the same patterns for scale dependence in the growth rates are preserved if we focus on employment density (employment over county area). The same patterns are also present if we plot sectoral growth rates as a function of total, rather than sectoral, employment. This finding is important, since, in the theory, some forces are sector specific (e.g., knowledge spillovers), whereas others are location specific and apply to all sectors (e.g., land rents).

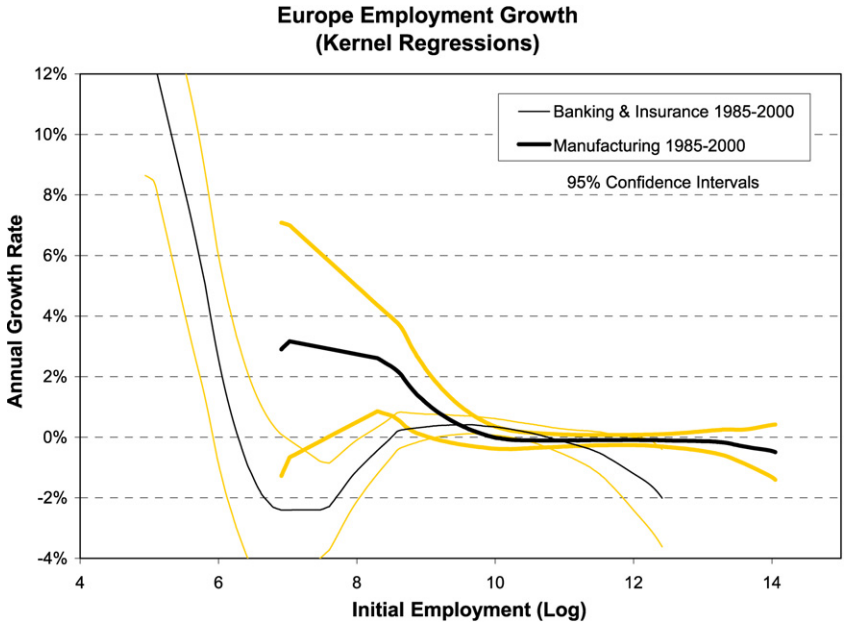


Fig. 5.

vices/real estate, and communications (Basu et al. [4]). Given that our database does not have disaggregated data on all of these sectors, in our empirical analysis we focus on 'banking and insurance.'

Regarding the timing, there is ample evidence that Europe has lagged behind the U.S. in the adoption of IT.[12] We therefore take 1985–2000 as the relevant period of IT diffusion in Europe. In Fig. 5 we pooled the data for five-year intervals between 1985 and 2000 to increase sample size. As in the U.S., for Europe we observe a declining curve for manufacturing and an S-shaped curve for services (in the case of the banking and insurance industry). This is consistent with a mature (manufacturing) and a young (services) industry.

2.4. Scale dependence in productivity growth

In the theory we present below, the driving force behind the growth of employment is the growth of total factor productivity (TFP). As a robustness check, we therefore re-run our kernel regressions for the U.S. using TFP, instead of employment. To compute sectoral TFP, we need empirical counterparts of land, sectoral output, and sectoral employment.[13] For land we use data

[12] Gust and Marquez [18] and van Ark et al. [33] document this for slightly different sets of European countries. Both studies include Finland, France, Germany, Italy, Netherlands, Norway, Spain, Sweden, and the U.K. Each of these countries had lower investments in IT than the U.S. in the 80s and 90s.

[13] The theoretical model we present below has two factors, land and labor. Production per unit of land in manufacturing will be written as $M = Z_M L_M^{\mu}$, where Z_M is manufacturing TFP. Solving for manufacturing TFP gives us $Z_M = M/L_M^{\mu}$. Similarly, production per unit of land in services will be written as $S = Z_S L_S^{\sigma}$, so that services TFP is $Z_S = S/L_S^{\sigma}$. This calculation of TFP does not account for capital or other factors of production, but it has the advantage of being exactly consistent with our theory.

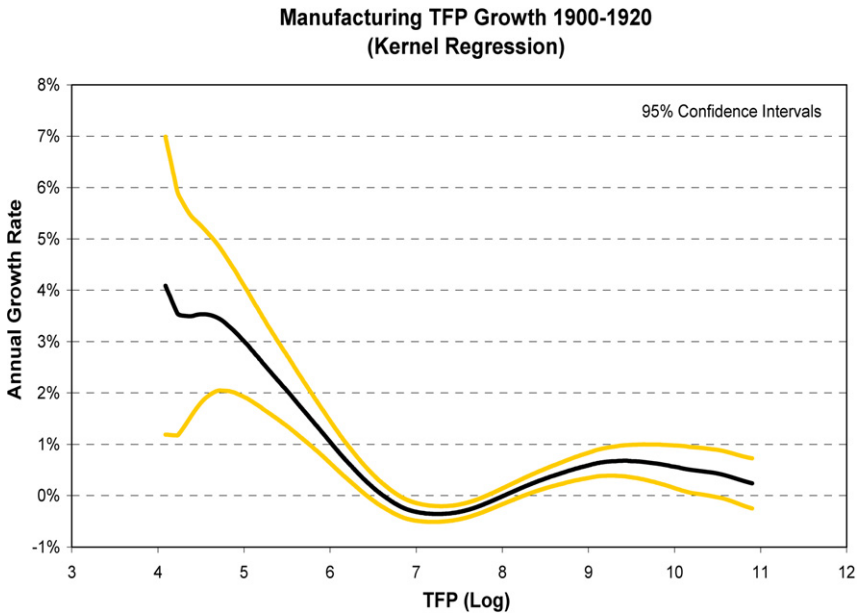


Fig. 6.

on county area from the U.S. Geological Survey. Manufacturing output for the years 1900–1920 come from the Historical Census Browser at the University of Virginia, which provides county-level data on the value of manufacturing production. Manufacturing and services output for the years 1980–2000 comes from the Bureau of Economic Analysis, which collects county-level data on total earnings per sector. Sectoral employment variables are the ones used in the previous section. Substituting these variables into our production functions allows us to compute TFP for manufacturing and services in each county.[14]

Fig. 6 plots annual de-trended TFP growth in manufacturing on the log of initial TFP for the period 1900–1920. Given that manufacturing is young during this period, TFP growth exhibits the expected S-shaped pattern. By the last decades of the 20th century manufacturing has become a mature industry. Spillovers have lost their importance. Fig. 7 shows how, for the period 1980–2000, the S-shaped pattern has disappeared, with convergence now dominating across the entire distribution.

If the S-shaped pattern is related to the youth of an industry, it should also apply to the service industry at the end of the 20th century. Fig. 8 plots de-trended TFP growth for the period 1980–2000 for three different service sectors: retail, other services and FIRE. As expected, the S-shape re-emerges. Note one slight difference with the manufacturing industry during the period 1900–1920: the convergence at the top part of the distribution is less pronounced. This suggests that spillovers were sufficiently strong to cause high TFP growth in most of the highly productive counties. Of course, when we look at employment, instead of TFP, land congestion costs for

[14] These measures of TFP depend on the value of the parameters μ and σ, which we set to be equal to 0.5 and 0.55, as in the numerical simulations in Section 4. The results do not change qualitatively if we increase the values of μ and σ to, say, $2/3$.

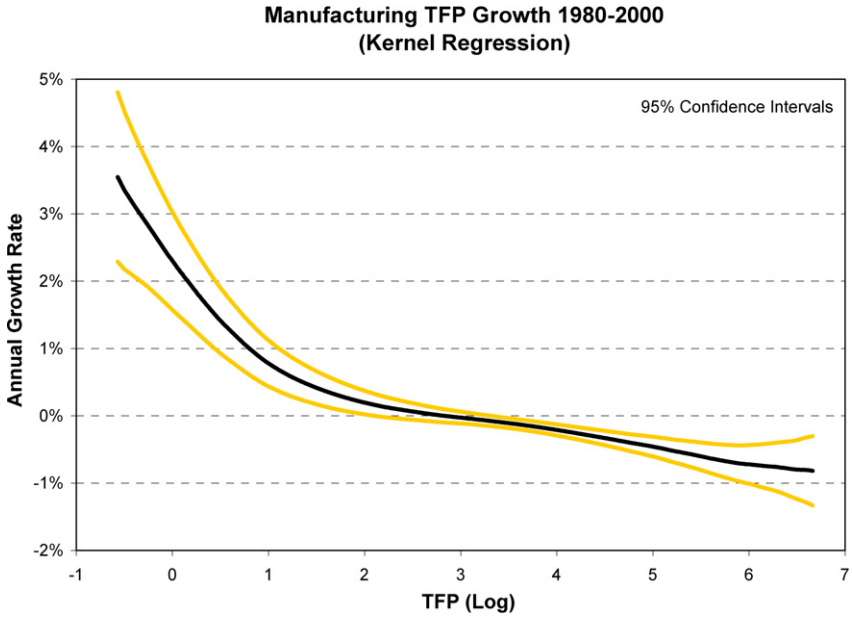


Fig. 7.

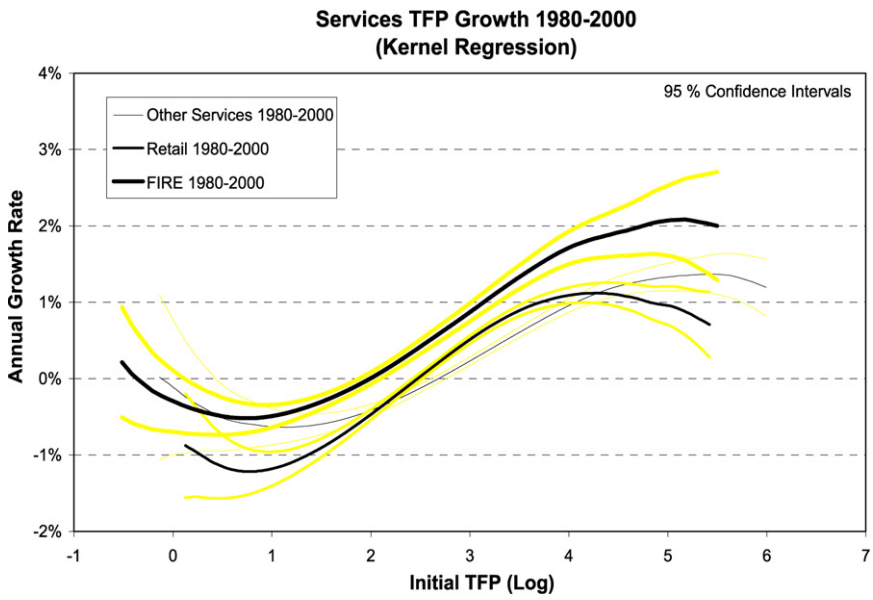


Fig. 8.

large agglomerations will yield more pronounced negative scale dependence at the top of the distribution.

Figs. 6–8 would not look qualitatively different if they were to plot TFP growth as a function of initial employment, rather than as a function of initial TFP. This is because employment and TFP

are positively correlated in the data. In our minds, the positive relation between employment and TFP suggests that the driving force behind the patterns of employment and TFP is the diffusion and adoption of technologies, as in the theory we propose.

2.5. Subsectoral analysis

Our empirical analysis so far has been based on rather broad sectors. In principle, this makes sense for two reasons. First, at higher levels of disaggregation, one would expect the behavior of particular sectors to become more idiosyncratic. Second, at higher levels of disaggregation, data availability at the county level becomes problematic because of disclosure and confidentiality issues. However, while taking into account these caveats, it may still be worthwhile to explore the relation between spatial growth and industry age for more detailed sectors.

The goal of this section, then, is to analyze whether our findings continue to hold for more disaggregated sectors. In particular, we look at the service sector (retail, FIRE and other services) and analyze whether in the last two decades those subsectors that were particularly affected by IT exhibited an S-shaped spatial growth pattern, and whether those subsectors that did not experience much effect from IT exhibited a monotonically decreasing spatial growth pattern.

We use employment data at the two-digit SIC level from the County Business Patterns dataset spanning the period 1977–1997.[15] For the 60 available sectors, on average, there are data for only about one-third of the slightly more than 3000 counties in the United States. As mentioned, this is mainly a problem of confidentiality: when employment in a certain sector and county is concentrated in a limited number of firms, the data are not disclosed. To limit this problem, we focus on those subsectors for which we have at least two-thirds of the counties. Within the three service categories (retail, FIRE, other services), this leaves us with 10 subsectors.

To decide which sectors to focus on, we turn to the empirical literature on IT intensity. There are three relevant studies that analyze IT intensity at the two-digit SIC level (Chun et al. [8]; Caselli and Paternò [7]; McGuckin and Stiroh [27]). In all of these studies IT intensity is defined as IT capital as a share of total capital.[16] For each of those studies, we choose the most IT-intensive sector and the least IT-intensive sector, within the subset of sectors that have observations for at least two thirds of the counties. In spite of the differences in definitions, in all three studies the most IT-intensive sector is legal services and the least IT-intensive sector is auto repair. For the case of legal services the IT intensity is estimated to be between 30% and 17%. For the case of auto repair the IT intensity is found to be very low in all three studies, between 2.3% and 3.8%.

Figs. 9 and 10 show employment growth between 1977 and 1997 across U.S. counties, using the same methodology as before. As can be seen, legal services exhibit the S-shaped spatial growth pattern. This is consistent with the importance of IT in that particular subsector of the economy. In contrast, auto repair looks like a mature sector, with convergence across the entire distribution. Again, this is consistent with IT being of little importance in that particular sector. Although doing a more in-depth analysis at the two-digit level is beyond the scope of this paper, these examples do suggest that our basic finding — an S-shaped pattern of spatial growth in

[15] These data are available on-line at the Geospatial and Statistical Datacenter at the University of Virginia.

[16] The definitions of IT capital are slightly different across studies, though: in Chun et al. [8] and Caselli and Paternò [7] it is essentially defined as the sum of hardware and software, whereas in McGuckin and Stiroh [27] it refers to the sum of computer hardware and other high-tech equipment.

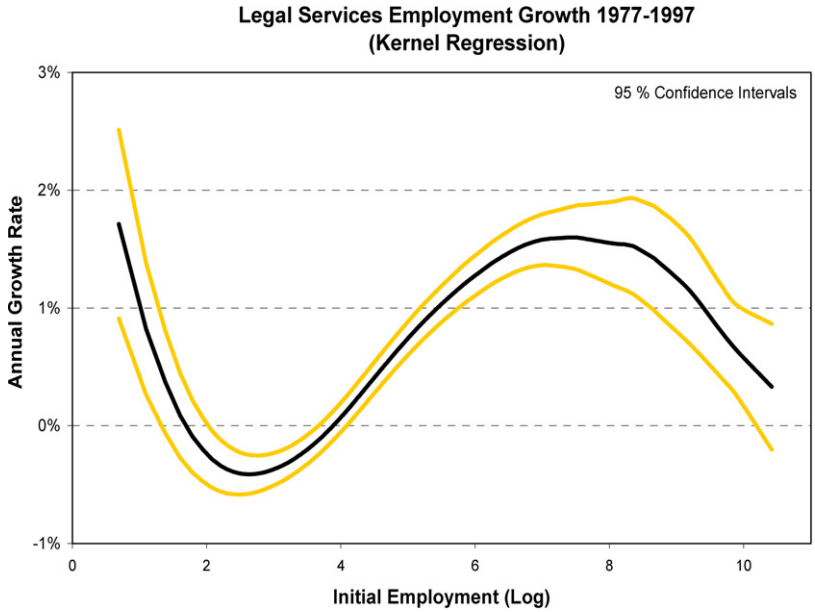


Fig. 9.

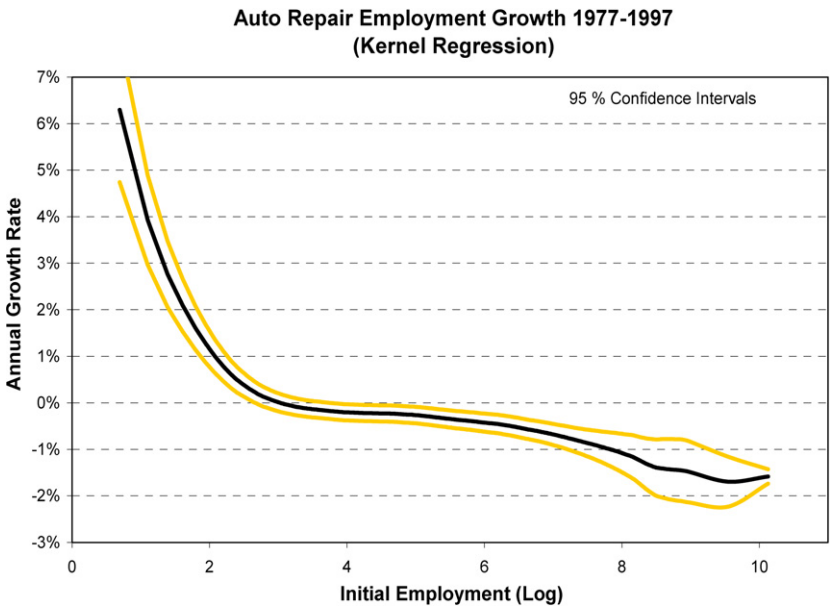


Fig. 10.

sectors affected by IT and a monotonically declining pattern of spatial growth in the rest of the economy — holds up when we look at more detailed sectors.

3. The model

The economy consists of land and people located in the closed interval [0, 1]. Throughout we refer to a location as a point in this interval, and we let the density of land at each location ℓ be equal to one. Hence, the total mass of land in the economy is equal to one. A region is then a connected subset of the unit interval.[17] The total number of agents is given by $\bar{L}$, and each of them is endowed with one unit of time.

3.1. Preferences

Agents live where they work and they derive utility from the consumption of two goods: manufactures and services. Labor is freely mobile across locations and sectors. Agents supply their unit of time inelastically in the labor market. They order consumption bundles according to a utility function $U(c_M, c_S)$ with standard properties, where c_i denotes consumption of good $i \in \{M, S\}$. We also assume that $U(\cdot)$ is homogeneous of degree one. Agents own land where they work and live. Apart from land, there is no other saving technology.[18]

The problem of a consumer at a particular location ℓ is given by

$$\begin{aligned} &\max_{c_i} U(c_M, c_S) \\ &\quad \text{s.t. } w(\ell) + R(\ell)/L(\ell) = p_M(\ell)c_M(\ell) + p_S(\ell)c_S(\ell) \end{aligned} \tag{1}$$

where $p_i(\ell)$ denotes the price of good i at location ℓ, $R(\ell)$ denotes land rents per unit of land at location ℓ (so $R(\ell)/L(\ell)$ is the dividend from land ownership, since $L(\ell)$ is total employment at ℓ and all agents are identical), and $w(\ell)$ denotes the wage at ℓ. The first-order conditions of this problem yield $U_i(c_M, c_S) = \lambda(\ell)p_i(\ell)$, for all $i \in \{M, S\}$, where $U_i(\cdot)$ is the marginal utility of consuming good i and $\lambda(\ell)$ is a location-specific Lagrange multiplier. Denote by $\bar{U}(p_M(\ell), p_S(\ell), w(\ell) + R(\ell)/L(\ell))$ the indirect utility function of an agent at location ℓ. Because of free mobility of labor, it must be the case that

$$\bar{U}\big(p_M(\ell), p_S(\ell), w(\ell) + R(\ell)/L(\ell)\big) = \bar{u}, \quad \text{for all } \ell \in [0, 1], \tag{2}$$

where $\bar{u}$ is determined in equilibrium. In the numerical examples in the next section we use a CES utility function

$$U(c_M, c_S) = \big(h_M c_M^{\beta} + h_S c_S^{\beta}\big)^{1/\beta}$$

with $\beta < 1$.

3.2. Technology

Each location can produce in both sectors or specialize in one of them. The manufacturing sector is assumed to be more land intensive than the service sector and both sectors face knowledge spillovers. The inputs of production are land and labor. Production per unit of land in the manufacturing sector is given by

$$M\big(L_M(\ell)\big) = Z_M(\ell)L_M(\ell)^{\mu},$$

[17] As counties are the smallest geographic units in our data, we will relate a location in the model (a point in space) to a county in the data.

[18] Since the dynamics of the model enter only through the endogenous evolution of technology, we do not introduce time dependence in the notation until we describe the evolution of technology in Section 3.4.

and, similarly, in the service sector we have

$$S\big(L_S(\ell)\big) = Z_S(\ell)L_S(\ell)^{\sigma},$$

where $Z_i(\ell)$ is TFP in sector i and location ℓ and $L_i(\ell)$ is the amount of labor per unit of land used at location ℓ in sector i. We assume that $\mu < \sigma < 1$, reflecting the higher land intensity of manufacturing. As we will specify below (see Eq. (8)), TFP in each sector depends on the amount of labor employed in the same sector in surrounding locations. We assume that a firm takes $Z_i(\ell)$ as given, so that it does not take into account the effect of other producers on productivity: an externality. The problem of a firm in sector $i \in \{M, S\}$ at location ℓ is thus given by

$$\max p_i(\ell)Z_i(\ell)L_i(\ell)^{\iota} - w(\ell)L_i(\ell), \tag{3}$$

where $\iota \in \{\mu, \sigma\}$. The maximum per unit land rent that firms in sector i are willing to pay, the bid rent, is then given by

$$R_i(\ell) = p_i(\ell)Z_i(\ell)\hat{L}_i(\ell)^{\iota} - w(\ell)\hat{L}_i(\ell). \tag{4}$$

3.3. Land, goods, and labor markets

Because of trade, locations specialize in one industry, as in Rossi-Hansberg [30]. Goods are costly to transport. For simplicity we assume iceberg transportation costs that are identical in manufacturing and services. This is without loss of generality, given that the equilibrium depends only on the sum of transport costs in both industries.[19] If one unit of any of the goods is transported from ℓ to r, only $e^{-\kappa|\ell-r|}$ units of the good arrive in r. Since the technology to transport goods is freely available, the price of good i produced in location ℓ and consumed in location r has to satisfy $p_i(r) = e^{\kappa|\ell-r|}p_i(\ell)$.

Land is assigned to its highest value. Hence, land rents are such that $R(\ell) = \max\{R_M(\ell), R_S(\ell)\}$. Denote by $\theta_i(\ell)$ the fraction of land at location ℓ used in the production of good i. If $R(\ell) = R_i(\ell)$, then $\theta_i(\ell) > 0$. Of course, with complete specialization this condition becomes $\theta_i(\ell) = 1$.

In order to guarantee equilibrium in product markets, we need to take into account that some of the goods are lost in transportation. To do this, let $H_i(r)$ denote the stock of excess supply of product i between locations 0 and r. Define $H_i(r)$ by $H_i(0) = 0$ and by the differential equation

$$\frac{\partial H_i(r)}{\partial r} = \theta_i(r)x_i(r) - \hat{c}_i(r)\left(\sum_i \theta_i(r)\hat{L}_i(r)\right) - \kappa\big|H_i(r)\big|, \tag{5}$$

where $x_M(r) = M(\hat{L}_M(r))$ and $x_S(r) = S(\hat{L}_S(r))$ denote the equilibrium production of good i at location r per unit of land. That is, at each location we add to the stock of excess supply the amount of good i produced and we subtract the consumption of good i by all residents of r. We then need to adjust for the fact that if $H_i(r)$ is positive and we increase r, we have to ship the stock of excess supply a longer distance. This implies a cost in terms of goods and services given by κ. The equilibrium conditions in the goods markets are then $H_i(1) = 0$ for all i.

We impose trade balance location by location. The value of the goods shipped to location ℓ must thus be identical to the value of the goods shipped from location ℓ, so that

[19] To see this, suppose transportation costs are zero in services and prohibitive in manufacturing. Because we impose trade balance location by location, services would become as non-tradeable as manufactured goods, in spite of being freely transportable. The difference in transportation costs across sectors therefore plays no role in the current model.

$$p_M(\ell)H_M(\ell) + p_S(\ell)H_S(\ell) = 0 \quad \text{for all } \ell. \tag{6}$$

The trade balance condition says that the value of goods produced and consumed at ℓ is equal, once transport costs in terms of goods are covered.

In equilibrium labor markets clear. Given free mobility, we have to guarantee that the total amount of labor demanded in the economy is equal to the total supply $\bar{L}$. The labor market equilibrium condition is therefore

$$\int_0^1 \sum_i \theta_i(r)\hat{L}_i(r)\,dr = \bar{L}. \tag{7}$$

3.4. Evolution of technology

We still need to specify how TFP is determined in each industry. Since the dynamics come through the evolution of technology, we now introduce time dependence explicitly in the notation. We let

$$Z_i(\ell,t) = \max\begin{bmatrix} \rho\bar{Z}_i^{\max}(t-1) + (1-\rho)(\int_0^1 e^{-\delta_i|\ell-r|}\hat{L}_i(r,t)\theta_i(r,t)\,dr)^{\gamma_i}, \\ (\int_0^1 e^{-\delta_i|\ell-r|}\hat{L}_i(r,t)\theta_i(r,t)\,dr)^{\gamma_i} \end{bmatrix}, \tag{8}$$

where $\rho \in [0,1]$. If $\rho = 1$, then at time t all locations have access to the general level of technology $\bar{Z}_i^{\max}(t-1)$ established last period, as we describe below. If $\rho < 1$, the general level of technology still diffuses, but not perfectly. Locations can possibly improve upon this general level of technology, $\bar{Z}_i^{\max}(t-1)$, if they benefit from sufficiently large spillovers. These spillovers are a weighted average of employment at all locations, where the weights are a function of distance. We assume that $\gamma_i + \max[\mu,\sigma] < 1$. This guarantees that spillovers are a concave function of total population and so economic activity does not agglomerate in only one point. It also implies that very dense locations gain less from extra workers than less dense locations: a form of congestion.

Note that, given the general level of technology $\bar{Z}_i^{\max}(t-1)$, Eq. (8) *implicitly* determines technology at all locations, $Z_i(\cdot,t)$. The relationship is implicit because the optimal employment levels, $\hat{L}_i(\cdot,t)$, and the optimal land use patterns, $\theta_i(\cdot,t)$, depend themselves on the technology at all locations, $Z_i(\cdot,t)$. In other words, in equilibrium, the technology levels must generate employment levels and land-use patterns that generate those same technology levels.

We now describe how the general level of technology evolves over time. The best technology in a given period becomes the general technology in the next period. Thus, the general technology in period t and sector i, denoted by $\bar{Z}_i^{\max}(t-1)$, is the maximum level of technology across all locations in period $t-1$ and sector i:

$$\bar{Z}_i^{\max}(t-1) = \max_{\ell} Z_i(\ell,t-1). \tag{9}$$

As can be seen in (8), depending on ρ, spatial diffusion is perfect or imperfect, thus giving all locations complete or incomplete access to this general level of technology. The dynamics of the model come only through changes in the general levels of technology in manufacturing and services. An equilibrium of the dynamic economy is thus nothing other than a series of static equilibria (as the one described above), but with different levels of the general technology across periods. In other words, this is a two state variable spatial model, where the state in t is summarized by the general levels of technology in manufacturing, $\bar{Z}_M^{\max}(t-1)$, and services, $\bar{Z}_S^{\max}(t-1)$.

It is illustrative to study the case of $\rho = 1$. With perfect diffusion, after one period the better technology that some producers obtain from the benefits of agglomeration is in the public domain. That is, all producers in all locations have access to the best technology of the previous period. It is easy to show that in this case the economy will converge to a steady state in which all locations have the same technology. Given the specification of $Z_i(\ell, t)$ and $\rho = 1$, it is clear that $\bar{Z}_i^{\max}(t)$, $t = 0, 1, \ldots,$ is a weakly increasing sequence. Technology in both sectors is bounded by $\bar{L}^{\gamma}$, that is, $\lim_{t\to\infty} \bar{Z}_i^{\max}(t) \leqslant \bar{L}^{\gamma}$, all i. This, together with $\bar{Z}_i^{\max}(t)$ being weakly increasing for all i, implies that these sequences have to converge to a unique steady state, where

$$\bar{Z}_i^{\max}(t) = \bar{Z}_i^{\max}(t-1) \quad \text{all } i.$$

Since the model exhibits only transitional dynamics, growth rates of employment in all sectors in all locations become zero. (Note, however, that if $\bar{Z}_i^{\max}(t) = \bar{Z}_i^{\max}(t-1)$ in only one industry, there may be non-zero growth rates in all industries.) For the case of $\rho < 1$, the argument is more complex, since the sequence $\bar{Z}_i^{\max}(t)$, $t = 0, 1, \ldots,$ is still bounded but not necessarily increasing. A subsequence has to converge so there is either a steady state or a finite cycle. Numerically we find that the sequence converges for many values of ρ.

3.5. Definition of equilibrium

An equilibrium in this economy is a set of real functions $(c_i, \hat{L}_i, \theta_i, H_i, p_i, R_i, w, Z_i)$ of locations $\ell \in [0, 1]$ and time $t = 1, \ldots,$ and a function $\bar{Z}_i^{\max}$ of time, for $i \in \{M, S\}$, such that:

- Agents choose consumption, c_i, by solving the problem in (1).
- Agents locate optimally, so w, p_i, R_i and L_i satisfy (2).
- Firms maximize profits by choosing the number of workers per unit of land, $\hat{L}_i$, that solves (3), and by choosing the land bid rent, R_i, that solves (4).
- Land is assigned to its highest value, so if $\max\{R_M(\ell), R_S(\ell)\} = R_i(\ell)$, then $\theta_i(\ell) = 1$.
- Goods markets clear, so H_i is given by (5) and $H_i(1) = 0$.
- Trade is balanced location by location, so (6) is satisfied.
- The labor market clears, so θ_i and $\hat{L}_i$ satisfy (7).
- Technology Z_i satisfies the fixed point in (8) and the general technology, $\bar{Z}_i^{\max}$, evolves according to (9).

Lucas and Rossi-Hansberg [25] and Rossi-Hansberg [30] show that a static equilibrium of this economy exists for an arbitrary level of technology in both sectors. Hence, since the dynamics depend only on the level of technology in the different locations, an equilibrium of this model exists as well.

We think of a GPT as a new general production technology that induces technological innovation, spillovers and diffusion in an industry, following a period in which technological innovation had essentially come to a halt. In the wake of a GPT, the technology level is determined by the process described above. As the technology in an industry matures, spillovers are no longer useful and the technology in that industry reaches a steady state. Then a new GPT may emerge benefiting another industry. Thus, in the beginning of the 20th century manufacturing was the young sector and services the old one. In this context it is natural to think of the younger sector as having the higher level of technology, so that initially $\bar{Z}_S^{\max} < \bar{Z}_M^{\max}$. This reverses by the end of the century, as manufacturing is now old and services young.

4. Numerical results

In the following numerical simulations we use parameter values that maximize visibility and computational ease.[20] The qualitative results are robust to variations in most of these parameters. The resulting equilibrium has three regions of specialization and is symmetric. The areas at the boundaries of the interval [0, 1] specialize in manufacturing and the region in the middle in services. The size of these regions changes as technology evolves, but this pattern is constant across periods. As shown in Rossi-Hansberg [30], higher trade costs would result in more switches in land specialization.

We want to illustrate the implications of the model for sectoral growth rates and for the relative shares of manufacturing and services. Since the main general purpose innovation in manufacturing occurred at the beginning of the 20th century, much earlier than in services, we set initial productivity in manufacturing to be higher than in services.[21] Starting from those values, in an initial stage we let productivity in manufacturing evolve as implied by the model in Section 3.4, whereas we keep productivity in the service sector fixed. This reflects the fact that manufacturing is young and firms are adapting to the general purpose technology. Then, at a later stage, we introduce a new GPT that impacts the service sector. From that point onward we let productivity in both the manufacturing and the service sector evolve as described in Section 3.4.

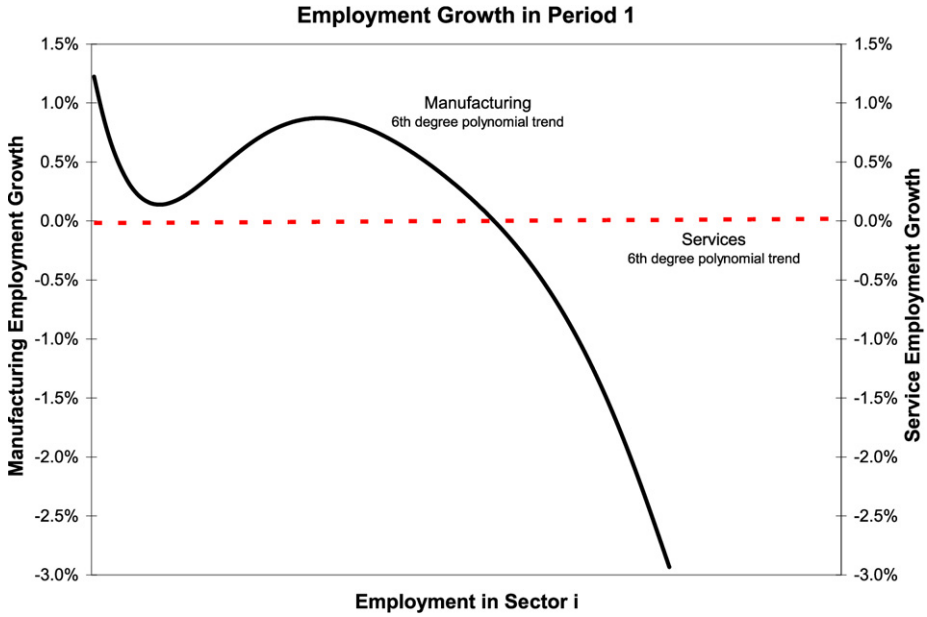
We present results for employment growth in both industries for three distinct time periods: before the GPT in services; just after the GPT in services; and some time later when technologies in both sectors are mature. The first period should illustrate one in which the GPT in manufacturing happened recently, but there is not much innovation in services (as in 1900–1920 in the data). The second period is one in which manufacturing is already a mature industry, and innovation is very active in the service industry (as in 1980–2000 in the data). The third period illustrates the case when both technologies are mature (the future, absent new GPTs).

Fig. 11 presents the results for manufacturing and service growth in these three time periods. (We present only three periods but we iterate several times between periods.) The solid curve shows the growth rate in manufacturing employment ordered by the size of manufacturing employment in that location. The dashed curve shows the same information for the service sector. The length of a given interval on the horizontal axis measures a set of locations specializing in a given industry.[22] Because of trade, all locations specialize, so the set of locations for which employment is positive differs across sectors. Of course, the sum of the lengths of the intervals for manufacturing and services adds up to one. In all cases we present a trend calculated by fit-

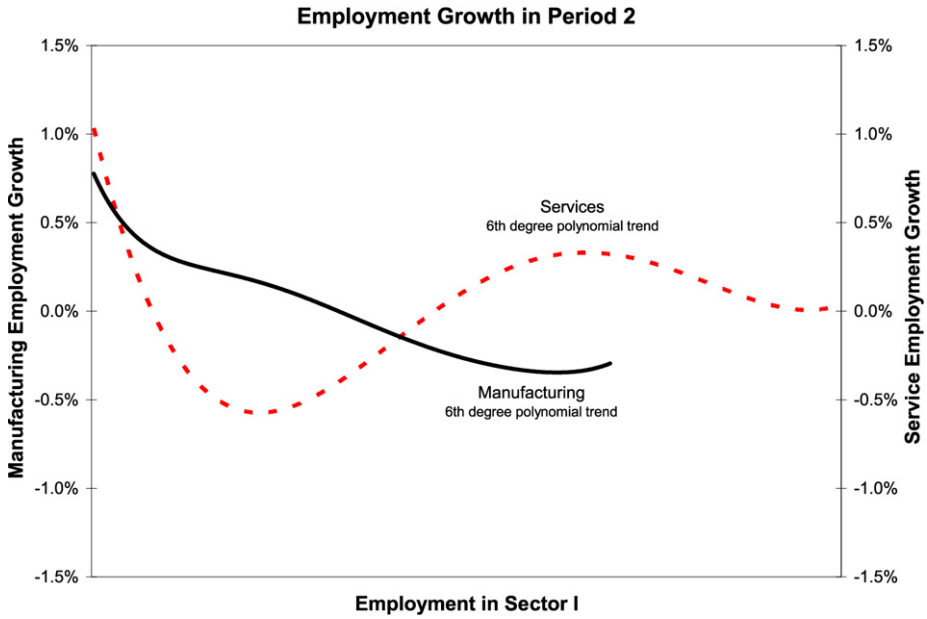
[20] We use $\beta = 0.5$, $\mu = 0.5$, $\sigma = 0.55$, $\gamma_i = 0.05$, $\delta_i = 10$, $\kappa = 0.005$ and $\rho = 0.95$. All other parameters are set to 1. For computational simplicity we fix $\bar{u} = 1$ and let total population size adjust, rather than fixing the population size and adjusting the utility levels. Doing this eliminates one of the iterations needed to compute an equilibrium and makes the computations feasible. Thus, one should interpret our economy as one with migration and population growth. We also distribute rents to absentee landlords instead of rebating them to the consumers of each location.

[21] We need to start the economy with an initial level of technology. Instead of assuming a constant level, in both sectors we use an initial productivity that is quadratic and symmetric around $\ell = 0.5$, where it obtains its maximum. We use a quadratic instead of a constant function to simplify the computations and avoid multiplicity of equilibria (e.g., manufacturing areas can be at the boundaries or at the center). The initial technologies are such that the quadratic functions have a minimum of 0.39 in manufacturing and 0.19 in services and a maximum of 0.4 in manufacturing and 0.2 in services.

[22] The horizontal axis in Fig. 11 exhibits the location index ordered by employment size, not employment size itself. The lengths of the intervals therefore measure the number of locations that specialize in each industry. In all industries and periods, the graphs using actual employment size in the industry exhibit the same patterns of scale dependence.



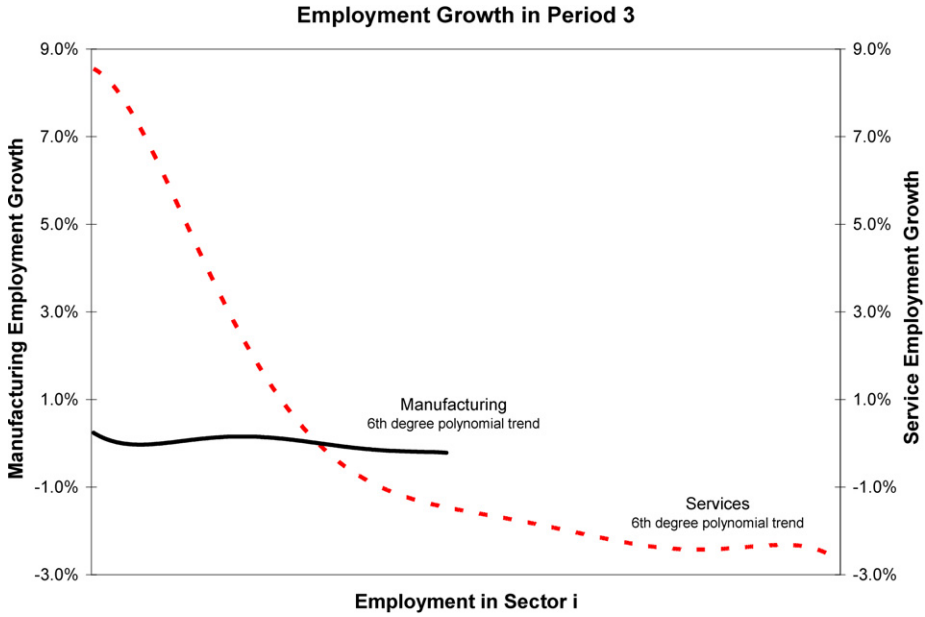
(a)



(b)

Fig. 11.

ting a polynomial of degree six. The polynomial of degree six allows us to smooth employment growth rates in a sector and present just the scale dependence implied by the model, without the noise that results from having several distinct growth rates associated with a particular employ-



(c)

Fig. 11. (*continued*)

ment size. This issue is present only in the manufacturing industry in periods 2 and 3, since this industry is losing employment in these periods in locations that have the same level of employment as locations that keep producing manufactured goods. In all other graphs this issue is not present, but we apply the same smoothing for comparison purposes. This smoothing also eliminates growth rates equal to infinity or minus one in locations that switch specialization patterns. All growth rates in Fig. 11 are relative to the mean growth rate in the period, so we present growth rates after subtracting the average growth rate in the sector and period.

It is clear in Fig. 11a for period 1 that the model produces the non-monotonic (S-shaped) pattern observed in the data for manufacturing in 1900–1920. Locations with little or no manufacturing employment exhibit convergence, with smaller locations experiencing faster growth. These are locations, specialized in services and isolated from existing manufacturing clusters, that switch to manufacturing production. Since they do not benefit from spillovers from surrounding locations, the standard convergence forces are at work. Locations that already have some manufacturing employment exhibit divergence, with smaller locations growing slower than larger ones. Because of agglomeration forces, the larger ones tend to have larger surrounding locations and benefit from greater spillovers, the driving force for the concentration of employment. Once locations become large enough, the pattern switches around once again, and convergence is back: congestion dominates spillovers from surrounding locations, and employment disperses.

According to our simulations, during period 1 employment growth in the service sector is essentially zero (although actual growth rates are not exactly zero). Of course, this is the result of our assumption that technology in the service sector is fixed in the first period. Unfortunately there are no data on employment growth in services at the county level at the beginning of the 20th century that are able to discipline our theoretical exercise for services in the first period.

The service sector loses some locations to the manufacturing sector, though. This is natural, since manufacturing technology is improving and diffusing.

Just before the second period, the technology in the service industry starts evolving as described in Section 3.4. This can be observed in Fig. 11b. This reflects the impact of a new GPT. By now manufacturing has become a mature industry, although its technology is still diffusing. Its employment is becoming more disperse in order to take advantage of low land rents. In contrast, the service industry now exhibits the S-shaped pattern or non-monotonicity previously observed in the manufacturing sector. This seems to be the period that best approximates the data for the U.S. between 1970–2000.

In the third period, both industries are mature and technology is relatively uniform, although diffusion continues, particularly in the now old (but relatively younger) service sector. Both industries are still taking advantage of low land rents in certain areas that used to have inferior technologies. Employment is becoming more dispersed. Growth rates in both industries decline with the size of employment in the corresponding sector. If we let technology diffuse perfectly, eventually growth rates become zero and the economy stops growing. This cycle is almost complete in the manufacturing sector, as can be observed in Fig. 11c.

The focus in Fig. 11 is on the shape, and not the level, of the curves. We detrend all growth rates as we do in the data. The reason is that we can arbitrarily change the level by changing the initial technologies and the utility level $\bar{u}$. However, changing these parameters proportionally does not change the qualitative features of these curves. Note that the difference in growth rates across locations is smaller in the model than in the data presented in Section 2. Of course, this depends on the definition of a period. If we group several periods, we can obtain differences of the same magnitude.

Output growth follows very similar patterns given that employment growth is solely the product of productivity growth in a particular sector and location. So it is interesting to look at the evolution of productivity at these three points in time. The following two figures show the level of productivity in manufacturing and services across the interval [0, 1]. Since both sectors exhibit substantial productivity growth between the first and second period, we use different axes for the first period (left axis) and for the second and third periods (right axis).

Fig. 12 shows how technology diffuses in the model. In the first period the productivity distribution in manufacturing is mostly the outcome of spillovers.[23] In the second period, diffusion is the dominating force, except for the few locations exhibiting a bump. In those areas spillovers continue to prevail because of the high concentration of manufacturing employment in the surroundings. In particular, the productivity that results from agglomeration effects is higher than the productivity these regions had access to through technological diffusion. By the third period this more advanced technology has smoothly diffused to the other regions, as everyone has access (up to ρ) to the best technology used last period. Fig. 13 shows the evolution of TFP in the service sector. In this sector diffusion is strong enough so that technology increases substantially in all locations. By the third period technology is relatively uniform across locations.[24]

[23] Note that because of specialization not all locations actually have manufacturing employment. In particular, the region in the middle is specialized in services. In that region manufacturing productivity represents what productivity would be were the region to produce manufactured goods.

[24] As in the data, the model yields a share of employment in services that is larger than in manufacturing and increases as the service sector starts innovating and these innovations diffuse in space. However, given our CES assumption, the model also implies a reduction in the relative price of services, which may be counterfactual (Buera and Kaboski [6]).

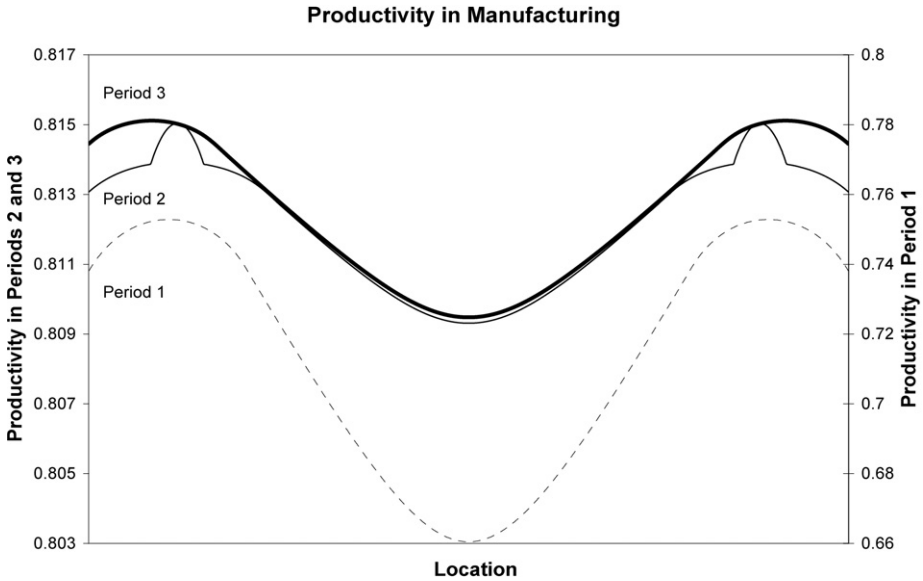


Fig. 12.

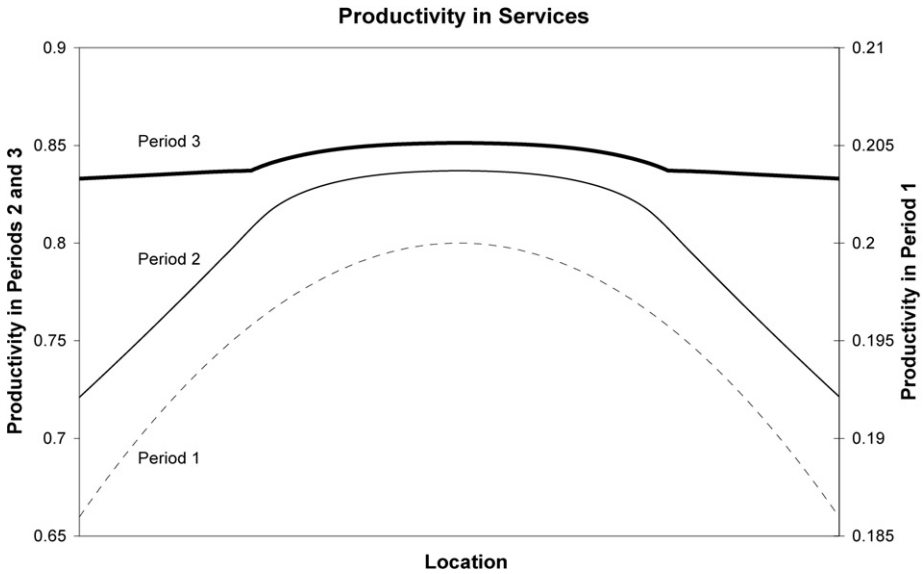


Fig. 13.

The results above were computed for a particular set of parameter values. The general pattern of scale dependence is, however, robust to many of these parameters and, in general, depends more on the relative values we choose in manufacturing versus services rather than on their level. Fig. 14 presents results for the growth rate in service employment in period 2 for different values of σ. The figure shows that the S-shaped pattern becomes more pronounced the larger the value of σ. The results for different manufacturing employment shares, μ, behave similarly.

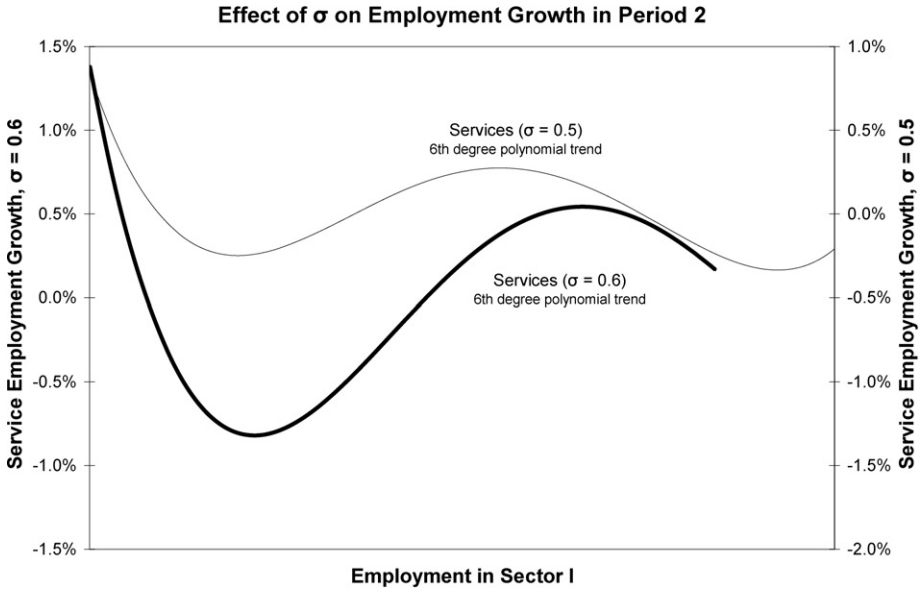


Fig. 14.

The larger μ, the more pronounced the S-shaped pattern observed in manufacturing employment growth in the first period.

One potential concern with the results above is the initial productivity function we have chosen in both sectors. As mentioned, we start both sectors with a quadratic productivity function. The shape of these curves does affect the results. However, as we make the curvature smaller, so the initial variation in productivities across locations decreases, we obtain the same pattern of scale dependence but with amplified S-shaped patterns.

5. Conclusions

We have documented a new fact about the evolution of employment across sectors and industries. The spatial evolution across regions seems to be related to industry age. At a minimum, it is clear that the scale dependence in employment growth is different in the service and manufacturing sectors and that the scale dependence in manufacturing at the turn of the 20th century resembles the one in services in the last couple of decades of the 20th century.

Our theory suggests that this distinct evolution in the manufacturing and service sectors may be related to the age of an industry as measured by the time since the last GPT innovation had a major impact on that sector. Young industries innovate and benefit from spatial knowledge spillovers. This, together with technological diffusion and trade, leads to changes in spatial specialization patterns, consistent with the observed S-shaped scale dependence in employment growth in young industries. Mature sectors disperse as technology diffuses further and firms move to locations where land rents are low. Importantly, the data show similar patterns of scale dependence for productivity. This is consistent with our theory where the driving force behind the observed employment patterns are technological.

One caveat to our findings is that we did not document employment growth in the service industry in the first two decades of the 20th century, a period during which we would argue the service industry was old. Data limitations prevented us from doing so.

The theory presented endogenizes technological growth across regions by making it a function of the level of employment in nearby locations. However, the evolution of these technologies and the technological spillovers themselves are modeled only in reduced form. Modeling explicitly the adoption decisions of firms would, of course, lead to a richer spatial theory of endogenous growth.

Acknowledgments

We thank Kunal Dasgupta for his excellent research assistance and Tom Holmes, Robert Lucas, Jordan Rappaport, Paul Romer, Giorgio Topa, and several seminar participants for useful comments. Desmet acknowledges the financial support of the Fundación Ramón Areces, and Rossi-Hansberg acknowledges the support of the Sloan Foundation.

References

[1] A. Atkeson, P.J. Kehoe, Modeling the transition to a new economy: Lessons from two technological revolutions, Amer. Econ. Rev. 97 (2007) 64–88.
[2] E.J. Bartelsman, J.J. Beaulieu, A consistent accounting of U.S. productivity growth, unpublished manuscript, 2004.
[3] S. Basu, J. Fernald, ICT as a general-purpose technology: Evidence from U.S. industry data, unpublished manuscript, 2006.
[4] S. Basu, J.G. Fernald, N. Oulton, S. Srinivasan, The case of the missing productivity growth: Or, does information technology explain why productivity accelerated in the United States but not the United Kingdom? Discussion Paper Number 2021, Harvard Institute of Economic Research, 2003.
[5] T.F. Bresnahan, M. Trajtenberg, General purpose technologies: Engines of growth? J. Econometrics 65 (1995) 83–108.
[6] Buera, F. and J. Kaboski, The rise of the service economy, Northwestern University Working Paper, 2006.
[7] P. Caselli, F. Paternò, ICT accumulation and productivity growth in the United States: An analysis based on industry data, Temi di Discussione 419, Banco d'Italia, 2001.
[8] H. Chun, J.-W. Kim, J. Lee, R. Morck, Information technology, creative destruction, and firm-specific volatility, unpublished manuscript, 2005.
[9] P.A. David, The dynamo and the computer: An historical perspective on the modern productivity paradox, Amer. Econ. Rev. 80 (1990) 355–361.
[10] P.A. David, G. Wright, General purpose technologies and surges in productivity: Historical reflections on the future of the ICT revolution, in: P.A. David, M. Thomas (Eds.), The Economic Future in Historical Perspective, Oxford University Press, 2003.
[11] K. Desmet, M. Fafchamps, Employment concentration across U.S. counties, Reg. Sci. Urban Econ. 36 (2006) 482–509.
[12] G. Duranton, D. Puga, Nursery cities: Urban diversity, process innovation, and the life cycle of products, Amer. Econ. Rev. 91 (2001) 1454–1477.
[13] G. Duranton, Urban evolutions: The fast, the slow, and the still, Amer. Econ. Rev. 97 (2007) 197–221.
[14] C. Forman, A. Goldfarb, S. Greenstein, Geographic location and the diffusion of Internet technology, Electron. Commerce Res. Appl. 4 (2005) 1–13.
[15] R.L. Forstall, Population of the states and counties of the United States: 1790 to 1990, U.S. Bureau of the Census, Washington, DC, 1996.
[16] X. Gabaix, Zipf's law for cities: An explanation, Quart. J. Econ. 114 (1999) 739–767.
[17] Z. Griliches, Hybrid corn: An exploration in the economics of technological change, Econometrica 25 (1957) 501–522.
[18] C. Gust, J. Marquez, International comparisons of productivity growth: The role of information technology and regulatory practices, International Finance Discussion Paper Number 727, Board of Governors of the Federal Reserve System, 2002.

[19] J.V. Henderson, Urbanization and growth, Chapter 24, in: Handbook of Economic Growth, vol. 1, Part B, Elsevier, 2005, pp. 1543–1591.
[20] J.V. Henderson, A. Venables, The dynamics of city formation: Finance and government, Brown University Working Paper, 2005.
[21] B. Hobijn, B. Jovanovic, The information-technology revolution and the stock market: Evidence, Amer. Econ. Rev. 91 (2001) 1203–1220.
[22] T. Holmes, Step-by-step migrations, Rev. Econ. Dynam. 7 (2004) 52–68.
[23] B. Jovanovic, P.L. Rousseau, General purpose technologies, NBER Working Paper #11093, 2005.
[24] J. Kendrick, Productivity Trends in the United States, Princeton University Press, Princeton, 1961.
[25] R.E. Lucas Jr., E. Rossi-Hansberg, On the internal structure of cities, Econometrica 70 (2002) 1445–1476.
[26] R.E. Lucas Jr., Life earnings and rural-urban migration, J. Polit. Economy 112 (2004) S29–S59.
[27] R.H. McGuckin, K.J. Stiroh, Computers and productivity: Are aggregation effects important? Econ. Inquiry 40 (2002) 42–59.
[28] N. Rosenberg, M. Trajtenberg, A general-purpose technology at work: The Corliss steam engine in the late-nineteenth-century United States, J. Econ. Hist. 64 (2004) 1–39.
[29] E. Rossi-Hansberg, Cities under stress, J. Monet. Econ. 51 (2004) 903–937.
[30] E. Rossi-Hansberg, A spatial theory of trade, Amer. Econ. Rev. 95 (2005) 1464–1491.
[31] E. Rossi-Hansberg, M.L.J. Wright, Urban structure and growth, Rev. Econ. Stud. 74 (2007) 596–624.
[32] J.E. Triplett, B.P. Bosworth, "Baumol's Disease" has been cured: IT and multifactor productivity in U.S. services industries, The Brookings Institution, Washington, DC, 2002.
[33] B. van Ark, J. Melka, N. Mulder, M. Timmer, G. Ypma, ICT investment and growth accounts for the European Union, 1980–2000, Final Report on ICT and Growth Accounting for the DG Economics and Finance of the European Commission, Brussels, 2002.