Regional specialization: From the geography of industries to the geography of jobs

Antoine Gervais
Department of Economics, Université de Sherbrooke

James R. Markusen
Department of Economics, University of Colorado

Anthony J. Venables
Department of Economics, University of Oxford; Monash University

Abstract. Our analysis begins with an empirical investigation of how employment concentration in industries and occupations across regions of the United States has changed over time and how regional specialization has changed. Results show that industry concentration and specialization indices have fallen, while occupation concentration and specialization indices have risen. Using this background as motivation, we develop a model in which the comparative advantage of regions lies in their productivity of supplying functions such as law, finance, advertising and engineering, to multiple sectors. Productivity differences specific to region functions shape the location decisions of industries that use multiple functions and hence determine patterns of regional specialization both in functions and in sectors. A key parameter is the cost of sourcing functions from a different region (fragmentation costs), and we show that a fall in this cost mimics the data: sector concentration and regional specialization fall and function concentration and specialization rise. At high fragmentation costs, regional comparative advantage in sectors determines general equilibrium analogous to a Heckscher–Ohlin model (HO). At low fragmentation costs, comparative advantage in functions drives an equilibrium that has little resemblance to a HO world.

Résumé. Spécialisation régionale : de la géographie des industries à la géographie des emplois. Notre analyse commence par une enquête empirique de l'évolution au fil du temps de la concentration des emplois dans les industries et les professions dans l'ensemble des régions des États-Unis ainsi que de la modification de la spécialisation régionale. Les résultats démontrent que les indices de concentration et de spécialisation industrielles ont chuté, alors que les indices de concentration et de spécialisation des professions ont augmenté. À partir de ce constat, nous mettons au point un modèle où l'avantage comparatif des régions repose sur leur capacité à offrir des services comme le droit, les finances, la publicité et l'ingénierie à divers secteurs. Les différences de capacité des régions à offrir ces services façonnent les décisions relatives à l'emplacement des industries qui utilisent plusieurs services et déterminent donc les schémas de spécialisation régionale des services et des secteurs. Le coût lié à l'obtention de services dans une autre région (coûts de la fragmentation) et un paramètre clé, et nous démontrons qu'une chute de ce coût imite les données: la concentration sectorielle et la spécialisation régionale chutent alors que la concentration et la spécialisation des services augmentent. Lorsque le coût de la fragmentation est élevé, l'avantage comparatif régional dans certains secteurs détermine l'équilibre

---

Corresponding author: antoine.gervais2@usherbrooke.ca

We are grateful to editor, two referees and seminar participants for very helpful comments on earlier drafts. Any remaining errors are our own.

Canadian Journal of Economics / *Revue canadienne d'économique* 2024 57(4)
November 2024. / *Novembre 2024.*

24 / pp. 1236–1264 / DOI: 10.1111/caje.12747

général, comme dans un modèle Heckscher-Ohlin. Lorsque le coût de la fragmentation est faible, l'avantage comparatif des services stimule un équilibre qui ressemble peu au modèle de Heckscher-Ohlin.

JEL classification: F12, R11, R12 and R13

1. Introduction

From popular press reports to formal journal articles, much has been written about changing patterns of employment both within and across countries. A good deal of this effort focuses on the rise and fall of different sectors (industries), as changing technology, higher incomes, and foreign competition lead to a shift in production and demand across industries. There is also much interest in changing demand for different skills and occupations. These two are often closely linked: both within and across industry developments imply changes in the demand for different worker skills and occupations. The third phenomenon attracting attention is the change in the geography of production and jobs. Some regions grow and thrive, others stagnate or decline. This third phenomenon is linked to the first two, as growing areas are observed to specialize in the employment of workers needed in the expanding sectors, often drawing them from other regions.

The purpose of this paper is to develop both an empirical and a theoretical analysis that contributes to understanding the joint evolution of the spatial distribution of industries and occupations. Specifically, we analyze how industries (sectors) and occupations (functions) are becoming more regionally concentrated or more dispersed. Our empirics document these changes for US states over time. Our theory presents a model that mimics these changes and allows us to draw out further general equilibrium changes to many variables of interest.

We begin with an empirical investigation using US state-level data on sectoral and occupational employment. States are relatively large geographical units for the questions we are addressing, but because of data limitations, we believe it is better to operate at this level. We discuss the reasons in the next section, with alternative estimations using metropolitan statistical (MSA) data presented in the online appendix. We calculate how employment concentration in industries and occupations across US states has changed over time, and how regional specialization has changed. First, we find declining sectoral concentration and increasing occupational concentration over time. While the decline in sector concentration is perhaps widely acknowledged, we find that the decline has occurred within most sectors of activity and is driven principally by this within sector change rather than by compositional changes in the relative size of different sectors. Similarly, the spatial concentration of most occupational groups has increased over time. These results are important as they suggest that modelling should capture changes occurring within sectors and occupations. Second, regional specialization indices in sectors and occupations have the same properties as the concentration indices, showing decreasing specialization by sector and increasing specialization by occupation.

We develop a theoretical approach based on the core idea that regions' comparative advantages have evolved from being based on sectors, to being based on productivity differences in "functions" (occupations in the data). Our approach draws on elements from several literatures. In no particular order, these include international trade theory, new economic geography, multinational firms and outsourcing, and urban/regional economics. From each, we pick-and-chose certain features and discard others to try capture the correct combination of assumptions that seems consistent with the changing economic geography of industry, and occupational specialization and concentration within the country. We provide specific references to the literature below as we introduce the components of our model.

From international trade theory, we use the typical assumption that sectors (industries) differ in the intensity with which they use inputs. These inputs are produced by labour, and we refer to them as functions. The key feature of our approach is that regions differ in the relative productivity of labour in performing different functions. Crucially, regional comparative advantage, therefore, lies in region-function, not region-sector, productivity differentials, although in equilibrium these differentials will show up in patterns of both functional and sectoral specialization. What are the sources of region-function productivity differences? In developing the model, we start by assuming these are exogenous, as in Ricardian trade theory. Then, drawing on the new economic geography literature, we assume that the productivity advantages of a region may arise due to agglomeration economies (spillovers) where a larger set of workers specializing in the same function leads to higher productivity. This seems closely consistent with many of the examples in Moretti (2012). Regional productivity in functions such as software engineering, banking and finance, marketing, and biotechnology increases with the number of regional workers in those functions.

The extent to which productivity advantage in a function can be exploited by producers depends on the extent to which sectors can "fragment," performing different functions in different regions. We capture this by drawing on the literature on fragmentation, vertical multinational firms, and outsourcing. We assume that a sector in a region may draw all of its functions from within the region, or source them from other regions. While doing the latter brings the benefit of exploiting region−function-specific productivity and wage differentials, it incurs a fragmentation cost. When this cost is large, sectors are integrated, and each region contains multiple functions. With a lower fragmentation cost, sectors will outsource the region's comparative disadvantage functions thus leading to functional specialization.[1]

For physical products, the fragmentation cost is the added transport cost of bringing parts and components together from different locations rather than having assembly and all intermediate production in one location. For occupations (used in our empirics), think of complex and frequent back-and-forth interactions among designers, engineers, technicians, and managers. If they are working for a sector from different regions (e.g., remote work), there are time delays, incentive costs, and communication confusion costs relative to face-to-face working. Face-to-face is particularly important "where information is imperfect, rapidly changing, and not easily codified" (Storper and Venables 2004).

Final ingredients in the general equilibrium version of our model are: (i) labour mobility between regions and from a hinterland area and (ii) international trade with the Rest-of-World. These lead to added results due to falling fragmentation costs such as increases in the total regional population, and changes in national comparative advantage across sectors.

Central results are that, as fragmentation costs fall, regions become more specialized across functions and less specialized in sectors. At high fragmentation costs, functions and sectors are closely linked, so comparative advantage in functions is manifest in the location of sectors, as in standard international trade theory. At low fragmentation costs, this link is weakened, and regions come to specialize in functions. The model predicts that functions become more concentrated as employment in a function occurs in fewer regions, and that

1 We assume that goods/services are costlessly assembled from functions, and that the cost of shipping a final product produced by an integrated sector to the other region is normalized at zero. The fragmentation cost is a unit cost (in units of labour), not a fixed cost, although in section 3 (partial equilibrium) each firm produces only one unit so there is not difference between fixed and unit costs.

sectors become less regionally concentrated as some of their employment is spread across regions. These predictions are confirmed by our empirical work.

Intuition for these mechanisms is provided by a simple example. With high fragmentation costs, a region has lawyers, accountants, machinists, mechanics and many other occupations working in a small number of comparative advantage sectors, like the historic auto industry in Detroit. With lower fragmentation costs, a region has a smaller range of occupations working in a larger number of sectors. New York specializes in white-collar functions such as finance and marketing, but these individuals are working for many different sectors that employ functions drawn from many different places.[2]

Ideas in this paper are complementary to the influential paper by Duranton and Puga (2005). While the titles of their paper and ours are similar, the theories developed move in quite different directions. Duranton and Puga (2005) focuses on city formation and on the configuration of city types that can exist in equilibrium. The analysis is driven by two forms of agglomeration economies that differ for firm headquarters and plants. The technologies for headquarters and plants in all sectors are identical, and there are no ex ante differences in city characteristics such as comparative advantage in sectors or functions. Only two types of configurations can be equilibria: (i) all firms have an integrated form and all cities specialize by sector (high fragmentation costs) or (ii) some cities host only headquarters and the remaining cities host only production plants in a single sector. Specific industries, and whether or not they are integrated or fragmented, are not associated with specific cities, and a given city cannot have integrated production in some sectors along with fragmented production in others.

In our theory, sectors differ in function intensity and regions differ in function comparative or absolute advantage. Equilibrium characterizes which sectors are integrated, and for each integrated sector, in which region is it located; and similarly, which sectors are fragmented and, for each fragmented sector, which function is located in which region. We determine the characteristics of a sector that is integrated versus fragmented, such as the degree to which the sector's function intensity ratio is extreme or average.

The rich regional/international general equilibrium structure of our theory also allows for interesting further implications of falling fragmentation costs beyond those analyzed with Duranton and Puga's (2005) model. These include, first, how falling costs change the pattern of comparative advantage across sectors and how this in turn influences the composition of trade and trade balance with the outside world. Second, falling fragmentation costs lead to high regional productivity, leading to migration from the hinterland to the urbanized regions. In short, we believe that our model will be well suited for addressing a range of empirical issues involving changes across and over time in industries' spatial structure, regions' specialization, and national changes in comparative advantage and trade.

Our focus on functions is also distinct from the literature on trade in tasks (for example Grossman and Rossi-Hansberg 2008).[3] We think of as there being relatively few functions (law, engineering, accountancy), most of them used by many sectors, compared with the task

2 These examples also hint at why location-specific function productivity may depend on agglomeration. For exogenous location–function comparative advantage, we generally think of location-specific amenities or complementary inputs fixed in that location (e.g., natural resources). Explicitly modeling these is well beyond the scope of this paper.

3 The Grossmann and Rossi-Hansberg (2008, 2012), tasks are a narrow stage of production, like the earlier models of Feenstra and Hanson (1996) and Markusen (1989), while our concept is a broader professional concept. In Grossmann and Rossi-Hansberg (2008, 2012), each worker resides in one country and is either a low-skilled or high-skilled worker, and there is no

approach of many tasks, each specific to a single sector. Fundamentally, the task literature asks questions about international trade between countries with fixed factor endowments, and the effect of such trade on factor returns. International aspects of fragmentation are also addressed in the literatures on multinational firms (Markusen 1989, 2002) and on global value chains (Antràs and Chor 2021), although these literatures do not address our central question of the interplay between functional and sectoral specialization.

As noted above, the questions we pose and the model we develop touch on many strands of international trade, economic geography, and urban economics. Some of our analysis builds on the large literature on economic geography, agglomeration, and multiple equilibria (see Henderson and Thisse 2004 and Duranton et al. 2015). Relevant work includes Audretsch and Feldman (1996), Berhens et al. (2014), Brackman and van Marrewijk (2013), Courant and Deardorff (1992), Davis and Dingel (2018), Fujita et al. (1999) and Krugman (1991b).

The empirical tools we use for measuring concentration and specialization are drawn from Krugman (1991b), Audretsch and Feldman (1996), and especially Ellison and Glaeser (1997). Evidence on urban specialization (sectoral and functional) includes Barbour and A. Markusen (2007), Duranton and Overman (2005), Ellison and Glaeser (1997), Gabe and Able (2012), Michaels, Rauch and Redding (2019), and the broad analysis of Moretti (2012). Our empirical results are also related to recent studies in the urban economics literature. For instance, Berry and Glaeser (2005), Moretti (2013) and Diamond (2016) all document skill divergence across cities. While these studies concentrate on dichotomous differences (i.e., skilled vs unskilled workers) across regions, our paper reports changes in concentration at a much more disaggregated level. We find that even within relatively detailed occupation categories, workers are increasingly concentrated. Our empirical results also complement previous works on functional specialization, including Duranton and Puga (2005). Using data from the Decennial Census of Population and Housing, they find that the ratio of managers to production worker is diverging across U.S. cities: ratios were similar across cities in 1977, but ratios for larger cities were significantly higher compared with those of small cities in 1997. Finally, our work is related to recent studies on the internal organization of firms such as Charnoz et al. (2018), Acosta and Lyngemark (2020), Gokan et al. (2019) and Acosta and Håkonsson Lyngemark (2021), which study the spatial distribution of firms' headquarter and production establishments.

The remainder of the paper is as follows. In section 2, we analyze the data using state-level information on production and employment by sector and occupation for the US for the period 1990–2019 for industries and 2000–2019 for occupations. In sections 3 and 4, we develop and provide analytical solutions for a partial equilibrium model with two symmetric regions. Section 3 assumes exogenous Ricardian differences in productivity by function and region. In section 4, we endogenize productivity differences by adding external economies of scale in the form of spillovers. In section 5, we characterize the general equilibrium model and address these questions numerically in a non-linear complementarity formulation. This also allows us to draw out further implications of fragmentation costs: relative and absolute employment levels and wage differentials across regions, relative output

endogenous switching of location or between high and low-skilled work. We assume workers can move between regions or from a hinterland to one of the regions, shaping the comparative advantage of each region. The ability to trade tasks in Grossmann and Rossi-Hansberg (2008, 2012) allows for some of the continuum of low-skilled tasks to be offshored for example, to a low-skilled-abundant country. But this cannot change the occupational structure and functional specialization of a region's workers, nor (with only two final goods) does it change the sectoral specialization of regions.

levels and prices across sectors, and net trade flows across sectors with the Rest-of-World. In section 6, we offer some concluding comments.

2. Concentration and specialization in the United States

In this section, we document time series changes in geographical concentration of sectors and functions and in regional specialization for the US. In section 2.1, we begin with a brief description of the methods we use to compute concentration and specialization indices. We then implement our measures in US data. In sections 2.2 and 2.3, we report declining sectoral concentration and increasing functional concentration over time, and that a large fraction of those changes is explained by within-sector and within-function changes in geographic concentration. Finally, in section 2.4, we report that states' sectoral specialization is decreasing over time, whereas their functional specialization is increasing. These empirical regularities help delineate the theoretical framework we develop in subsequent sections of the paper.

2.1. Measures of concentration and specialization

Using information on employment for each sector s and function f in each of r geographic areas, which we denote L_{sr} and L_{fr}, respectively, we can define the concentration of sector s as the sum over regions r of the share of sector s's national employment that is in region r minus region r's share of national employment, squared:

$$G_s = \sum_r (m_{sr} - m_r)^2, \quad \text{where } m_{sr} = L_{sr} / \sum_r L_{sr}, \text{ and } m_r = L_r / \sum_r L_r. \tag{1}$$

The concentration of function f employment across regions can be defined similarly as

$$G_f = \sum_r (m_{fr} - m_r)^2, \quad \text{where } m_{fr} = L_{fr} / \sum_r L_{fr}, \text{ and } m_r = L_r / \sum_r L_r. \tag{2}$$

Indices such as G_s and G_f are often used to measure agglomeration across regions (e.g., Krugman 1991b and Audretsch and Feldman 1996). An important limitation of these measures is that they could suggest high levels of concentration in sectors comprised of a few large plants located in a dispersed, random pattern. To control for this possibility, we adjust our indices following Ellison and Glaeser (1997) to obtain

$$EG_s = \frac{G_s / (1 - \sum_r m_r^2) - H_s}{1 - H_s}, \tag{3}$$

where $H_s = \sum_j z_{js}^2$ is the Herfindahl index of the sector's plant size distribution and z_{js} is the jth plant's share of sectoral employment.[4,5]

4 In practice (see results in online appendix 2), we find that changes in the value of EG_s over time are well approximated by changes in G_s. This happens because plant size distributions tend to change fairly slowly over time, so the correction is less important in cross-time comparisons (within a short time period) than in cross-sectors comparisons. Nevertheless, we use EG_s as our benchmark measure.

5 The motivation for the Ellison and Glaeser (1997) indices, defined in equations (3) and (4), is that it is an unbiased estimate of a sum of two parameters that reflect the strength of agglomeration forces (spillovers and unmeasured comparative advantage) in a model of

To measure functional concentration, we use a modified version of the Ellison and Glaeser (1997) index defined as

$$EG_f = \frac{G_f/\left(1-\sum_r m_f^2\right) - H_f}{1-H_f}. \quad (4)$$

As for sectors, the index adjusts G_f, defined in equation (2) above, to account for the fact that functions that are specific to a small number of plants will be more concentrated geographically compared with functions that are ubiquitous. Because we do not have information on plant-level employment by function, we cannot control directly for the dispersion of functions across plants. Instead, we use the Herfindahl index $H_f = \sum_s m_{fs}^2$, where m_{fs} is the share of employment in sector s performing function f. The intuition for the correction factor H_f, suggested by Gabe and Able (2010), is that when a function's employment is concentrated in a few sectors, the measured geographic concentration of the function should be higher all else equal.

We are interested in explaining aggregate time series patterns of concentration, so we combine sector-level indices into a single value by computing the employment-share weighted average over sectors for each year τ in our sample:

$$EG_\tau^{sector} = \sum_s m_{s\tau}\, EG_{s\tau}. \quad (5)$$

To gain additional insights, we can decompose time series changes in the geographic concentration into two adjustments margins: within-sector changes in geographic concentration and between-sector reallocation of employment. While the theoretical model we develop in the next section generates reallocation across sectors, we are particularly interested in explaining the within-sector component. For any given year τ, we can decompose the mean sectoral concentration defined in equation (5) as follows:

$$EG_\tau^{sector} = \sum_s m_s\, EG_{s\tau} + \sum_s (m_{s\tau} - m_s)\; EG_{s\tau}, \quad (6)$$

where $m_{s\tau}$ is sector s's share of national employment in year τ and m_s is the sector's share of employment in the sample (i.e., the mean over time of $m_{s\tau}$). The first term of the decomposition holds employment shares constant at the sample mean and provides information on the contribution of the within-sector changes in concentration over time. The second term captures the remainder of the time series change.

We can follow a similar process for functions. The employment-share weighted average over functions for each year, defined as

$$EG_\tau^{function} = \sum_f m_{f\tau}\, EG_{f\tau}, \quad (7)$$

can be decompose into a within-sector changes in geographic concentration and across-sector reallocation of employment as follows:

$$EG_\tau^{function} = \sum_f m_f\, EG_{f\tau} + \sum_f (m_{f\tau} - m_f)\; EG_{f\tau}. \quad (8)$$

location choice. At one extreme, the case of $EG = 0$, corresponds to a model in which location decisions are independent of region characteristics. In this case, the probability of choosing area r is m_r, the share of total employment in the region. At the other extreme, when $EG = 1$, region characteristics are so important that they completely overwhelm other factors, and the one region that offers the most favourable conditions will attract all the firms. In describing our results, we follow Ellison and Glaeser (1997) and refer to those industries with EGs above 0.05 as being concentrated and to those with EGs below 0.02 as being dispersed.

As explained in Aiginger and Rossi-Hansberg (2006), while regional specialization and geographic concentration are often considered almost identical economic phenomena (e.g., Krugman 1991a), they do not always develop in parallel. So, in addition to examining sector and function concentration, we also compute indices of regional specialization. Each region is compared with the national distribution of employment across sectors and functions via specialization indices, D_r^{sector} and $D_r^{function}$. Similar to our measures of concentration, the specialization of region r is defined as the sum over sectors (functions) of the square of the difference between the share of region r's employment in sector s (function f) and the share of national employment that is in sector s (function f) as follows:

$$D_r^{sector} = \sum\nolimits_s (q_{rs} - q_s)^2, \text{ where } q_{rs} = L_{sr}/\sum\nolimits_s L_{sr}, \text{ and } q_s = L_s/\sum\nolimits_s L_s, \tag{9a}$$

$$D_r^{function} = \sum\nolimits_f (q_{rf} - q_f)^2, \text{ where } q_{rf} = L_{fr}/\sum\nolimits_f L_{fr}, \text{ and } q_f = L_f/\sum\nolimits_f L_f. \tag{9b}$$

We aggregate region-level measures using a weighted average, where the weights are the regions' shares of national employment in the corresponding year:

$$D_\tau^{sector} = \sum\nolimits_r m_{r\tau} D_{r\tau}^{sector}, \quad \text{and} \quad D_\tau^{function} = \sum\nolimits_r m_{r\tau} D_{r\tau}^{function}. \tag{10}$$

This completes the description of the methods. In the next section, we implement the indices in US data.

2.2. Sectoral concentration

In this section, we study the time series of the geographic concentration of sectors. For this part of the empirical analysis, we use a balanced panel derived from the Bureau of Labor and Statistics' Quarterly Census of Employment and Wage data set that contains state-level data on 626 six-digit NAICS industries (our empirical measure of sectors) for years 1990 to 2019. As explained in online appendices 1 and 3, we believe that the publicly available state-level data is more reliable for our purpose compared with comparable MSA-level information. Nevertheless, we report MSA-level results in online appendix 3. Overall, the empirical analysis suggests that the main results are not affected by our choice of geography.

Data sources and measurement issues are discussed in online appendix 1. In our sample, about 41 % of the 18,780 observations are in manufacturing industries, the remainder of the observations are distributed across business services (23%), personal services (20%), and wholesale, retail and transportation (15%) industries. In the interest of space, we present only the most relevant empirical findings in the main text, additional results are presented in online appendix 2.

We compute the index of geographical concentration defined in equation (3) for each sector s and year τ in the data and denote it $EG_{s\tau}$. We find that for 362 of the 626 sectors (which together account for about 59% of US employment in our sample) the index of concentration is lower in 2019 than it was in 1990. The simple average of the concentration index over sectors decreases about 12% between 1990 and 2019 (going from 0.058 in 1990 to 0.051 in 2019, as reported in online appendix table A2.1). Considering the relative size of sectors strengthens the finding that sectoral concentration is declining on average. The red line in figure 1 depicts the employment-share weighted average over sectors for each year defined in equation (7). As seen in the figure, the mean sectoral concentration decreases by about 44% over the period (going from 0.027 in 1990 to 0.015 in 2019).

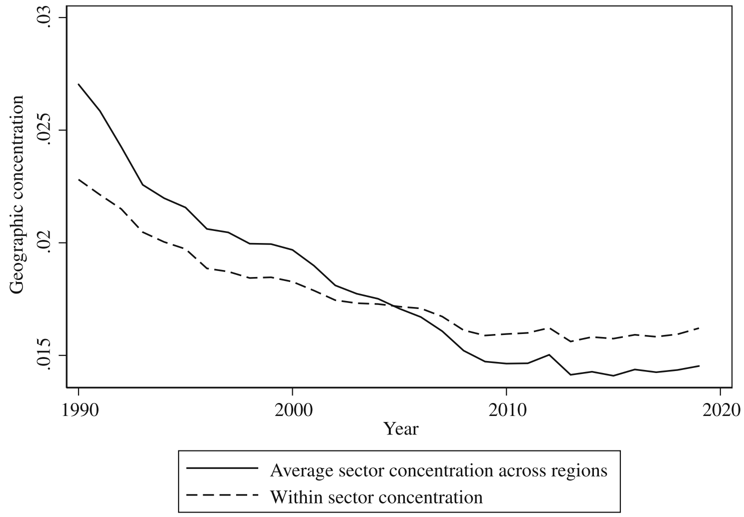

FIGURE 1 Geographic concentration of sectors over time

In online appendix table A2.2, we report estimated time series trend in geographic concentration for each broad economic sector. All estimates are negative and statistically significant, but the estimated decline in concentration is more important on average in wholesale, retail and transportation, and manufacturing sectors, then it was in business services and personal services sectors. The finding that sectoral concentration is decreasing over time is in line with the results of Dumais et al. (2002), who study the geographic concentration of sectoral employment across US states from 1972 to 1997.[6] Crafts and Klein (2021) also report a steady decline in the spatial concentration of US manufacturing industries between years 1880 and 2007. Similar results are found in countries. Dauth et al. (2016) report declining concentration in Germany between 1980 and 2010, while Barrios et al. (2005) for Ireland and Portugal between 1985 and 1998.

The results so far suggest that the average worker is employed in a more geographically dispersed sector in 2019 than was the case in 1990. To gain additional insights, we decompose time series changes in the geographic concentration into within-sector changes in geographic concentration and across-sector reallocation of employment. The results from decomposition (6) are depicted in figure 1. The blue line represents the within-sector component of the decomposition (i.e., the term $\sum_s m_s \, EG_{s\tau}$). The second term on the right-hand side of equation (6) is represented implicitly by the difference between the blue line and the red line (recall that the red line represents the overall change in concentration, EG_τ^{sector}, on the left-hand side of equation (6)). As seen in the figure, the rate of decline in concentration is lower when considering only the within-sector changes in concentration. We estimate that the within-sector component decreases by about 30% over the sample period

6 First, the two sets of estimates are of the same magnitude. They report a (simple) mean 0.034 for 1992. As reported in online appendix 2, our corresponding estimate is 0.056. The fact that our sectors are more concentrated on average can be explained by differences in scope and aggregation levels for sectors across studies. We include services and manufacturing sectors, whereas they focus on manufacturing, and we use six-digit NAICS industries as our definition of sectors, whereas they use three-digit NAICS. Second, they also find a decline in geographical concentration of sectors using US data. Both the simple and the employment weighted means of their index declines by more than 10% between 1972 and 1992.

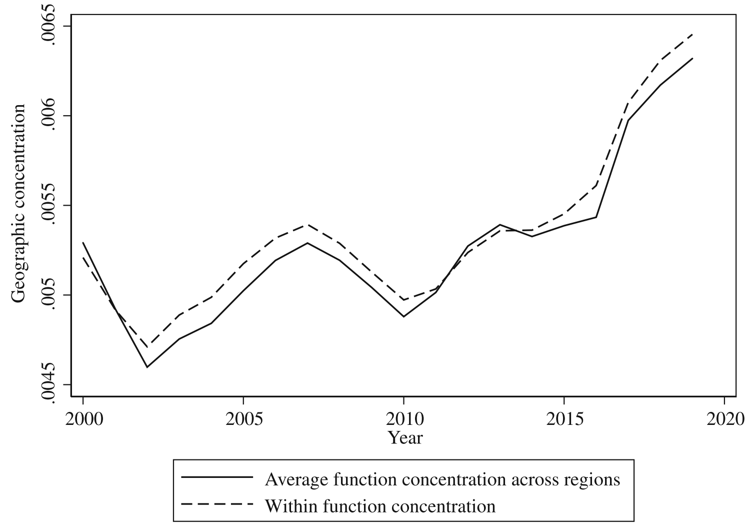

FIGURE 2 Geographic concentration of functions over time

(going from 0.023 in 1990 to 0.016 in 2019), whereas the average concentration decreases by 44% as reported earlier. While part of the observed decrease in sectoral concentration is due to labour movement from less concentrated industries towards more concentrated industries, our results suggest that the decline in the within-sector component of geographic concentration represents the majority of the time series change in geographic concentration.

2.3. Functional concentration

In this section, we study the times series properties of the geographic concentration of functional employment. For this part of he empirical analysis, we use a balanced panel that contains state-level data on 704 six-digit SOC occupations derived from the BLS's Occupational and Employment Statistics for years 2000 to 2019.[7] Data sources and measurement issues are discussed in online appendix 1. As a robustness check, we present results using broad two-digit function in online appendix 4. The main results are robust to this change in definition.

We begin by computing the index of geographical concentration defined in equation (4) for each function f and year τ in the data, $EG_{f\tau}$. We find that for 378 of the 704 functions (which together account for 55% of US employment in our sample), the difference between the 2019 and 2000 measures of concentration is positive, which implies an increase in the geographic concentration of functions over time. The simple average of the concentration index over functions increases about 10% between 1990 and 2019 (going from 0.0223 in 2000 to 0.0246 in 2019). Considering the relative size of functions increases the estimated increase in concentration. The red line in figure 2 depicts the employment-share weighted average over functions for each year, defined in equation (7). The weighted average concentration increases by about 19% over the period (going from 0.0053 in 2000 to 0.0063 in 2019).

7 We believe that the publicly available state-level data are more reliable for our purpose compared with comparable MSA-level information. MSA-level results are reported in online appendix 3. The main results are not affected by our choice of geography.

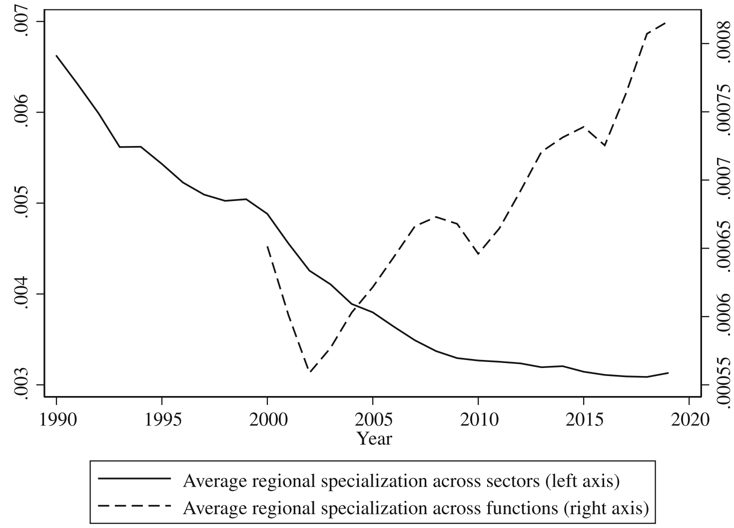

FIGURE 3 Regional specialization over time

Results from the decomposition defined in equation (8), but applied to functions, are also illustrated in figure 2. The blue line represents the within-sector component of the decomposition (defined as $\sum_f m_f \; EG_{f\tau}$). The figure clearly shows that there is an increase in the geographic concentration of functions over time even when holding the employment weights constant. As was the case for sectors, most of the time series change in functional concentrations is explained by the within-function component. Online appendix table A2.4 reports estimated time series trend in geographic concentration for each broad occupation category. The results show that 17 out of 21 estimated time trends are positive and 13 of those are statistically significant at conventional levels.

The empirical results in this section complement those of previous studies, such as Berry and Glaeser (2005), Duranton and Puga (2005), Moretti (2013) and Diamond (2016), that document divergence in the skill level of US cities over time. We provide evidence that functional concentration holds even within disaggregated definitions of occupations. To the best of our knowledge, the extant literature does not contain other times series estimates of the geographic concentration of functional employment to which we can compare our results.[8]

2.4. Regional specialization over time

We now turn to the sectoral and functional structure of regional employment. We use the region–sector and the region–function data sets described in the previous section to construct the two measures of regional specialization defined in equations (9a) and (9b) for each state–year in our data sets. In each case, we aggregate state-level measures using the weighted averages defined in equation (10).

The main results are reported in figure 3 (online appendix 2 presents additional results). The red line and the blue line depict respectively the time series of regional specialization in sectors (D_τ^{sector}) and functions ($D_\tau^{function}$). The decreasing trend in the red line indicates that the states' employment is becoming more evenly distributed across sectors over

8 Gabe and Abel (2012) report cross-sectional estimates for geographic concentration of occupations across US metropolitan areas using individual-level data from the 1% Public Use Microdata Sample (PUMS) of the 2008 American Community Survey. Gabe and Abel (2016) and Behrens and Guillain (2017) study the determinants of coagglomeration of occupations.

time. Conversely, the upward trend in the blue line indicates that states' distribution of employment across function is becoming increasingly uneven.

This completes the empirical section of the paper. To summarize, the empirical results suggest that the average US worker is employed in a more geographically disperse sector in 2019 than in 1990 but performs a function that is more geographically concentrated in 2019 than in 2000. We also find that, over time, states' employment is becoming more evenly distributed across sectors but increasingly unevenly distributed across functions.

3. Regions, sectors and functions

The empirical findings reported in section 2 indicate the spatial deconcentration of a majority of sectors of activity, while at the same time the sectoral concentration of almost all occupations—or "functions"—increased. The pattern is mirrored in spatial specialization, with decreased specialization in sectors, but increased specialization in functions. These findings suggest that, during the period of study, the latent comparative advantage of places in particular functions became more influential in shaping the location of employment. This in turn requires that firms became able to spatially fragment, performing different functions in different places. In the remainder of this paper, we set out a minimal model in which falling costs of fragmentation (due perhaps to technical progress in communication) enable regions to develop functional specializations that then shape the location of employment in a manner consistent with the data.

The ingredients of the model are regions, functions and sectors. For simplicity, we focus on just two regions and two functions. These functions can be used by multiple sectors, which we represent as a continuum. The functions use a single primary factor, labour, and are used as inputs to production of final output in each sector.[9] To capture the regional aspect of the model, we assume that labour is perfectly mobile, but its nominal wage may vary between places as the cost of living depends on employment in each place, as in the standard urban model. In our base model, comparative advantage is driven by region–function-specific Ricardian productivity differences, and we extend the model to endogenize these productivity differences through agglomeration effects. In this section and the next, we keep the general equilibrium side of the model in the background and make sufficient assumptions to ensure that the two regions are symmetric. In section 5, we fully specify the general equilibrium side of the model, enabling analysis of a richer set of possibilities.

The two regions are indexed $r = 1, 2$, and the wage rate in region r is denoted w_r. The single factor of production, labour, is perfectly mobile between regions but, since the cost of living may vary across regions, so may the nominal wage. The two functions, labelled $f = A,\ B$, are produced by labour with productivity that varies by region and function; production of one unit of function f in region r requires $\lambda_{fr} > 0$ units of labour. Regions are labelled such that productivity differences give region 1 a comparative advantage in function A, i.e., $\lambda_{A1}/\lambda_{B1} \leq \lambda_{A2}/\lambda_{B2}$.

There is a continuum of sectors, indexed $s \in [0, 1]$. Production occurs with constant returns to scale and perfect competition, and the output of sector s is denoted $n(s)$. This is freely traded at price $p(s)$. A unit of sector s output requires inputs of the two functions, and no other inputs. Sector s uses $a(s) > 0$ units of function A per unit output and $b(s) > 0$

9 Thus, workers can choose to become e.g., engineers or lawyers. Comparative advantage comes from cross-region variation in the productivity of labour in these functions. It would be possible to add a Heckscher–Ohlin flavour by assuming fixed endowments of engineers and lawyers, but this is inconsistent with the long-run perspective of the model.

units of function B, technical coefficients that we refer to as the function intensity of the sector.[10] These intensities vary with sector s but are the same in both regions; we assume that sectors can be ranked such that low s sectors are A-intensive and B-unintensive, i.e., $a'(s) < 0$ and $b'(s) > 0$.

Producers in each sector can source functions from either region, but if the two functions come from different regions, then a per unit fragmentation cost $t\overline{w}$ is incurred, where $\overline{w}$ is some average of wages in each region, depending on where these costs are incurred. Producers in each sector therefore operate in one of three modes, choosing to operate entirely in region 1, entirely in 2, or to purchase one function from region 1 and the other from region 2. Producers in a single region are "integrated" and will be labelled by subscript 1, 2 according to region of operation; those operating in both are "fragmented" and will be labelled by subscript F. The unit profits in sector s for each of the three production modes are, therefore,

$$\begin{aligned}\pi_1(s) &= p(s) - [a(s)\lambda_{A1} + b(s)\lambda_{B1}]\, w_1,\\ \pi_F(s) &= p(s) - [a(s)\lambda_{A1}w_1 + b(s)\lambda_{B2}w_2] - t\overline{w},\\ \pi_2(s) &= p(s) - [a(s)\lambda_{A2} + b(s)\lambda_{B2}]\, w_2.\end{aligned} \tag{11}$$

Costs are those of the functions purchased, sector s using $a(s)$ units of function A and $b(s)$ units of B per unit output. The functions use labour, with region r productivity λ_{fr}, $f = A,\ B$, and are costed at the region's wage w_r, $r = 1,\ 2$. Since the technology with which functions are combined into final goods $(a(s),\ b(s))$ is the same in both regions, urban comparative advantage is determined by the efficiency with which regions use labour to produce functions, λ_{fr}.

The endogenous choice of mode partitions the continuum of sectors into three groups. First is a range of s in which production is integrated, sourcing both functions in region 1. Because we have labelled regions such that region 1 has a comparative advantage in function A and ranked sectors such that low s sectors are A-intensive, it follows that these will be low s sectors. Second is a range of sectors in which production is fragmented, sourcing function A from region 1 and function B in region 2; if this range exists it will contain sectors with intermediate values of s (i.e., using both functions in similar proportions). Third are high s (B-intensive) sectors in which production is integrated in region 2, the region with comparative advantage in function B.

The boundaries between these ranges are denoted s_1, s_2 and are the sectors for which different modes of operation are equi-profitable, i.e., $\pi_1(s_1) = \pi_F(s_1)$, and $\pi_2(s_2) = \pi_F(s_2)$. Using (11), these mode-boundaries are implicitly defined, for interior solutions ($0 < s_1$, $s_2 < 1$), by

$$\begin{aligned}\pi_F(s_1) - \pi_1(s_1) &= b(s_1)\,[\lambda_{B1}w_1 - \lambda_{B2}w_2] - t\overline{w} = 0,\\ \pi_F(s_2) - \pi_2(s_2) &= a(s_2)\,[\lambda_{A2}w_2 - \lambda_{A1}w_1] - t\overline{w} = 0.\end{aligned} \tag{12}$$

Given technologies and wage rates, there are sectors operating in each mode if both these equations have solutions s_1, s_2 lying in the range (0,1). There are no sectors

10 $a(s)$ and $b(s)$ can be thought of as rows of a matrix mapping sectors to functions, as in Timmer et al. (2019). We show how the mapping operates only in circumstances where there is sufficient spatial variation in productivity or wages, and sufficiently low costs of fragmentation.

integrated in region 1 if $\pi_F(s)$ or $\pi_2(s) > \pi_1(s)$ for all s in this range, and similarly for other modes.[11]

For a given level of output each sector, $n(s)$, the levels of employment by function, region, and sector, denoted $L_{fr}(s)$, follow directly from equation (11) and are given in appendix table A1. The lower rows of the table give employment by function in each region, $L_{fr} = \int_s L_{fr}(s)\,ds$, employment by sector in each region, $L_r(s) = \sum_f L_{fr}(s)$, and total employment in each region, $L_r = \sum_f \int_s L_{fr}(s)\,ds$.

4. Sectoral and functional specialization in symmetric equilibria

We start by analyzing the way in which modes of operation and the consequent location of sectors and functions depend on technology and fragmentation costs, looking first at the case where efficiency differences are exogenous (section 4.1) and then turning to economies of scale (section 4.2). Full general equilibrium is set out in section 5.

4.1. Functional productivity: Ricardian differences

Throughout this section, we make strong assumptions that make regions and sectors symmetrical, enabling us to derive key results on the location of sectors and functions. We assume that output in each sector s is the same and constant, $n(s) = n$. Wages are the same in both regions taking common value w. Labour productivity in functions is assumed to be symmetric across regions, which we capture by denoting the labour input coefficient in each region's high productivity function as $\lambda \equiv \lambda_{A1} = \lambda_{B2}$, and that of the lower productivity function $\lambda_{A2} = \lambda_{B1} = \lambda + \Delta\lambda$, with $\Delta\lambda > 0$. Values for the mode-boundaries s_1, s_2 come from equations (12) and are given by

$$b\left(s_1\right)\Delta\lambda = t, \quad \text{and} \quad a\left(s_2\right)\Delta\lambda = t. \tag{13}$$

A simple case is where the function intensity of sectors is linear in s, taking the form $a(s) = [1 + \gamma(1 - 2s)]/2$ and $b(s) = [1 - \gamma(1 - 2s)]/2$ with $1 \geq \gamma > 0$. This is symmetric, with middle sector, $s = 1/2$, equally intensive in A and B. The parameter γ measures the heterogeneity of function intensities across sectors and $1 \geq \gamma$ means that both functions are used in all sectors.[12] Appendix table A2 gives employment levels by region, function and sector, replicating appendix table A1 with explicit expressions derived from this functional form. The unit profit functions of equation (11) become $\pi_1(s) = p(s) - \{2\lambda + \Delta\lambda[1 - \gamma(1 - 2s)]\}w/2$, $\pi_F(s) = p(s) - \lambda w - tw$ and $\pi_2(s) = p(s) - \{2\lambda + \Delta\lambda[1 + \gamma(1 - 2s)]\}w/2$, from which explicit expressions for the mode boundaries are

$$\pi_1\left(s_1\right) = \pi_F\left(s_1\right): \quad s_1 = \frac{1}{2}\left[1 - \left(1 - \frac{2t}{\Delta\lambda}\right)\frac{1}{\gamma}\right]$$
$$\pi_2\left(s_2\right) = \pi_F\left(s_2\right): \quad s_2 = \frac{1}{2}\left[1 + \left(1 - \frac{2t}{\Delta\lambda}\right)\frac{1}{\gamma}\right]. \tag{14}$$

11 There are no sectors integrated in region 2 if $\pi_F(s)$ or $\pi_1(s) > \pi_2(s)$ for all s and no fragmented sectors if $\pi_1(s)$ or $\pi_2(s) > \pi_F(s)$ for all s. These boundary cases will be addressed in section 5, which also endogenizes wages in each region.

12 Thus, for all $s \in [0, 1]$, $a(s), b(s) \geq 0$. Figure 4 has $\Delta\lambda = 0.4$, $w = 1$, and $\gamma = 1$. This value of γ is the special case in which all sectors become fragmented ($s_1 = 0$ and $s_2 = 1$) at $t = 0$. If sectors are more similar in function intensity, $\gamma < 1$, then all become fragmented at some positive value of t; if $\gamma > 1$ then extreme sectors use only one function.

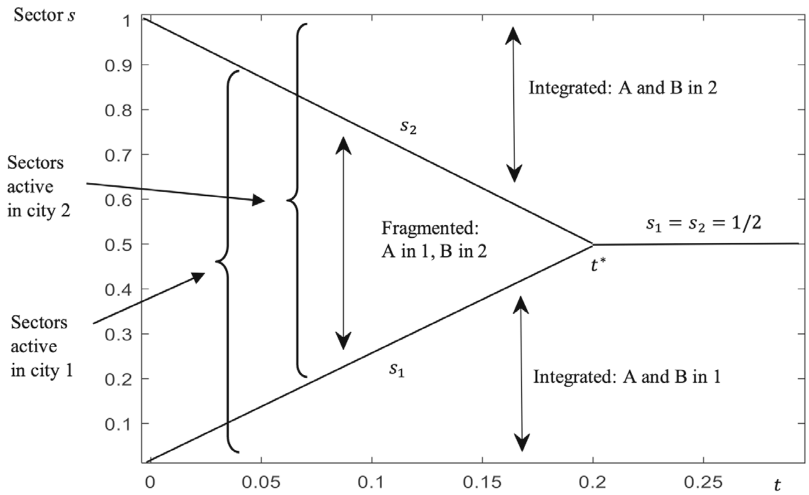

FIGURE 4 Modes of operation in each sector s

These relationships capture the way in which the sourcing of functions in each sector depends on fragmentation costs t, the range of function intensities γ, and inter-regional differences in relative labour productivity, $\Delta\lambda$.

Sectoral mode choice is illustrated on figure 4, which has sectors on the vertical axis and fragmentation costs, t, on the horizontal. If t is high, then all sectors are integrated, with an equal proportion of sectors in each region. If t falls below value $t^* = \Delta\lambda/2$, then fragmentation becomes profitable, first in sectors that have similar use of both functions, i.e., s in an interval around $1/2$ and of width $s_2 - s_1 = (1 - 2t/\Delta\lambda)/\gamma$, which is wider the smaller is t and the larger are productivity differences, $\Delta\lambda$.[13] Intuitively, these are the sectors where both functions have a similar share in costs, so it is worthwhile incurring cost t to source each from the lowest cost region. Sectors with more extreme function intensities remain integrated in the region where the function with highest cost share is relatively cheap. Thus, at $t < t^*$, the most A-intensive sectors operate with integrated production in region 1, the most B-intensive are integrated in region 2 and those with intermediate function intensities are fragmented, locating their functions according to inter-region differences in the productivity of labour in each function.

As fragmentation costs fall so more sectors become fragmented. This means that the number of sectors with a presence in each region increases, and hence both the regional concentration of sectors and the sectoral specialization of regions decline. At the same time activity in each region becomes more skewed toward the function in which it has comparative advantage, so each function becomes more regionally concentrated, and each region more functionally specialized. In simulation analysis of section 5.3, specialization and concentration indices are calculated for the distribution of both sectoral and functional employment across regions in a full general equilibrium context.

13 With fragmented production, costs are $\lambda w + tw$, so the ratio of fragmentation costs to production cost t/λ. The critical value is $t^* = \Delta\lambda/2$ so that, at this point fragmentation costs as a share of production cost are half the proportionate productivity difference $t^*/\lambda = (\Delta\lambda/\lambda)/2$.

4.2. Functional productivity: Localization economies

Ricardian efficiency differences provide the simplest model framework, but we think it unlikely that regional differences in the productivity of functions are due principally to exogenous efficiency differences. A further mechanism is the presence of function and location specific agglomeration economies, creating endogenous variation in the productivity of labour across functions and regions. We model this by assuming that labour input coefficients λ_{fr} now contain an endogenous part deriving from productivity spillovers in the same function and region, as well as a possible Ricardian component. The Ricardian component is as before, taking values λ and $\lambda + \Delta\lambda$. Productivity spillovers generated by each function in each region are equal to output in the function–region pair, $X_{fr} = L_{fr}/\lambda_{fr},\ f = A, B,$ $r = 1, 2$ with parameters σ_A and σ_B measuring the impact of spillovers on productivity. The Ricardian and endogenous components of labour input coefficients are additive, so

$$\lambda_{A1} = \lambda - \sigma_A X_{A1}, \qquad \lambda_{A2} = \lambda + \Delta\lambda - \sigma_A X_{A2}, \tag{15}$$

$$\lambda_{B1} = \lambda + \Delta\lambda - \sigma_B X_{B1}, \quad \lambda_{B2} = \lambda - \sigma_B X_{B2}.$$

Using expressions for employment, and hence output and spillovers, from appendix table A2 block IV, productivity differentials are

$$\lambda_{B1} - \lambda_{B2} = \Delta\lambda + \sigma_B n \left\{ \frac{1}{2} - s_1 \left[1 - \gamma (1 - s_1) \right] \right\}, \tag{16a}$$

$$\lambda_{A2} - \lambda_{A1} = \Delta\lambda + \sigma_A n \left\{ -\frac{1}{2} + s_2 \left[1 + \gamma (1 - s_2) \right] \right\}. \tag{16b}$$

Thus, if s_2 is large a relatively small range of sectors undertake function A in region 2, thereby reducing region 2's productivity in A, i.e., raising $\lambda_{A2} - \lambda_{A1}$. If these spillovers are equally powerful in both functions ($\sigma \equiv \sigma_A = \sigma_B > 0$) and wages are the same in both regions, then the mode-boundaries defined in equation (12) become

$$\pi_F(s_1) - \pi_1(s_1) = \left[1 - \gamma (1 - 2s_1) \right] (\lambda_{B1} - \lambda_{B2}) \, w/2 - tw = 0, \tag{17a}$$

$$\pi_F(s_2) - \pi_2(s_2) = \left[1 + \gamma (1 - 2s_2) \right] (\lambda_{A2} - \lambda_{A1}) \, w/2 - tw = 0. \tag{17b}$$

To analyze these relationships, we focus on equations (16a) and (17a), the other pair, equations (16b) and (17b), being symmetric. Substituting equations (16a) in (17a) gives $\pi_F(s_1) - \pi_1(s_1)$ as a function of s_1. Full analysis is given in the appendix at the end of the paper, and here we note the following facts and illustrate outcomes on figure 5.

First, there is full integration if $\pi_F(s_1) \le \pi_1(s_1)$ at $s_1 = 1/2$, and straightforward calculation gives value $t^{**} = [\Delta\lambda + n\sigma\gamma/4]/2$ at which $\pi_F(s_1) = \pi_1(s_1)$. This reduces to the Ricardian equivalent t^* if $\sigma = 0$, while $\sigma > 0$ implies a strictly higher critical point t^{**}, as expected given the additional source of productivity differences. At higher values of t, $t \ge t^{**}$, there is an equilibrium with fully integrated production, illustrated by the solid horizontal line on figure 5.

Second, the expression for $\pi_F(s_1) - \pi_1(s_1)$ is cubic in s_1 (see (16a) and (17a)), and this generates curvature of the mode boundaries and a range of multiple equilibria. In figure 5, this multiplicity occurs in the interval $(t^{**}, \tilde{t})$.[14] Integrated production is an equilibrium, because at this equilibrium productivity differences are small. But so too is a fragmented

14 Figure 5 has the same parameters as figure 4, except that $\Delta\lambda = 0$ and $\sigma_A = \sigma_B = 1.5$.

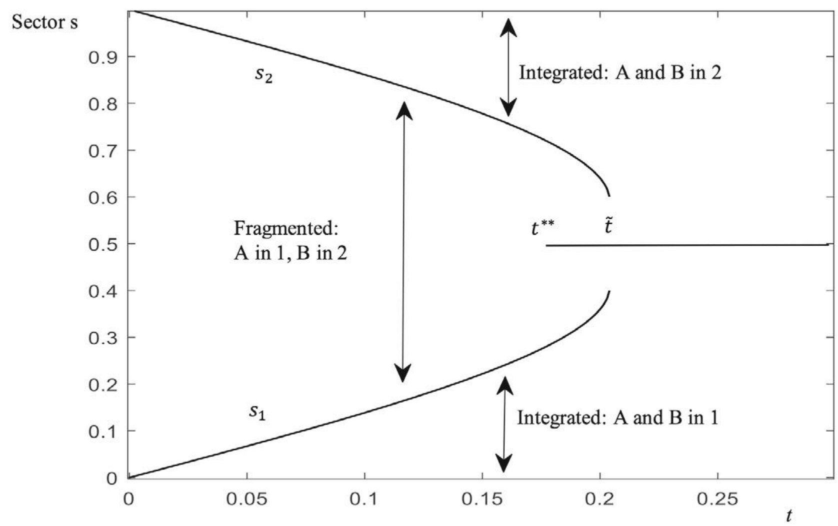

FIGURE 5 Modes of operation in each sector s: With increasing returns

equilibrium; at such an equilibrium production of function A is relatively concentrated in region 1, and B in region 2; the presence of increasing returns means that the productivity differential is now large, justifying sectors' choices to fragment production. Online appendix 5 works this through in detail, deriving the critical value $\tilde{t}$ below which fragmented production is an equilibrium and establishing that multiple equilibria arise if spillovers σ are large relative to any Ricardian productivity difference, $\Delta\lambda$.

Third, the qualitative effects of reducing fragmentation costs are as in the preceding case. Fragmentation reduces sectoral concentration and specialization and increases functional concentration and specialization. Importantly, these results do not depend on arbitrary Ricardian differences, but can also arise if places are ex ante identical and technology has location–function-specific agglomeration economies. Fragmentation allows these agglomeration forces to operate, thereby concentrating the location of functions and allowing the process we see in the data to operate. These arguments set out the driving mechanisms that we want to explore, and we now move to place them in a general equilibrium setting, endogenizing wages and the scale of activity (total output) in each sector.

5. General equilibrium

To this point, we have assumed product prices are constant, wages are constant and equal in both regions, the total output of each sector is fixed and the same in all sectors, and there is no interaction with the Rest-of-World. We now relax these assumptions and develop the general equilibrium of the model. The model does not admit analytical solutions, so the results are derived from numerical simulation.

Figure 6 illustrates the spatial structure of the general equilibrium model. The country of analysis consists of two regions surrounded by a hinterland. The regions can draw labour from the hinterland. Workers are homogeneous and move freely to equalize real wages. Inter-regional cost of trading goods is constant, and we set this at zero. The inter-regional cost of trading functions (fragmentation costs) is the variable of interest. Final goods, which are costlessly assembled from functions can be traded at fixed world prices with the Rest-of-World, but there is no international trade in functions.

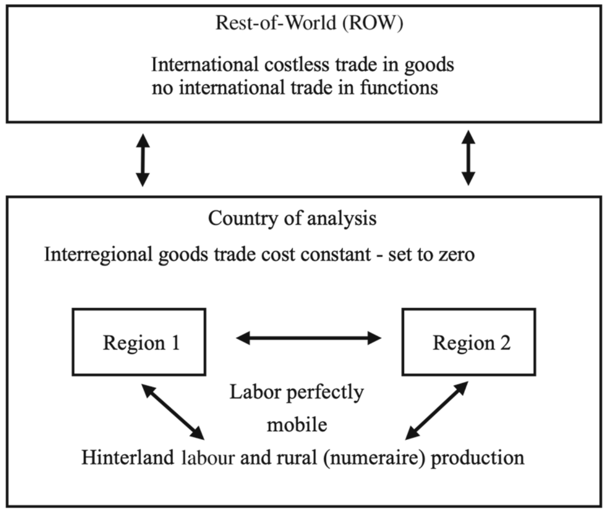

FIGURE 6 The spatial structure of the general equilibrium model

5.1. Region size, employment and wages

In addition to the sectors and functions modelled above, we now add an "outside good," which we use as numeraire. This good is produced in a hinterland region, using labour alone at constant productivity giving fixed hinterland wage w_0. The hinterland produces no other goods or functions, and this and all other final goods are perfectly freely traded.

Labour is perfectly mobile, equating utilities across regions. To prevent corner solutions—such as all population ending up in one region—we require some sort of diminishing returns to regional population, and this is achieved by supposing the existence of a fixed factor in each region. We take this to be the number of urban areas, each of which is described by the standard urban model (the Alonso–Mills–Muth model; see, for example, Henderson and Thisse 2004).

Thus, region r contains K_r cities, assumed to be identical. In each of these cities, workers face costs of commuting and land rent, costs that depend on city population. Since the cost of living may vary across regions, labour mobility is consistent with equilibrium nominal wages in each region, w_1, w_2, differing from w_0 and from each other. The microfoundations of the simplest possible urban model are that each urban household occupies one unit of land, all urban jobs are in the city centre and commuting costs are c_r per unit distance. A worker living at distance z from the centre must pay commuting costs $c_r z$, plus rent at distance z from the centre, denoted $h_r(z)$. Workers choose residential location within and between cities and regions, and real wages are equalised when $w_r - c_r z - h_r(z) = w_0$ for all r and at all occupied distances z. People in each city live and commute along a spoke from the centre, so city population is z_r^*, where z_r^* is the edge of the city (length of the spoke). At the city edge land rent is zero, so $z_r^* = (w_r - w_0)/c_r$. The total urban population living in region r cities is $K_r z_r^*$, so the relationship between the region r wage and its total urban population, $L_r = K_r z_r^*$, is

$$L_r = K_r (w_r - w_0)/c_r, r = 1, 2. \tag{18}$$

These equations imply that, given the number of cities and commuting costs, regions with a larger population and labour force must pay higher wages to cover the commuting costs and rents incurred by workers. Note that rent in each city can be expressed as,

$h_r(z) = w_r - w_0 - c_r z = c_r(L_r/K_r - z)$, so integrating over z and adding over all cities, total rent in a region of size L_r is

$$H_r = c_r L_r^2 / 2K_r. \quad (19)$$

Thus, while workers' utility is equalized across all locations, the productivity gap associated with $w_1, w_2 > w_0$ is partly dissipated in commuting costs, with the rest going to recipients of land rents. This is general enough to be a model of a single city, $(K_r = 1)$, or a model of a state containing multiple cities.

5.2. Production and demand

Sectors are perfectly competitive and produce goods by costlessly assembling them from functions. Sector outputs and prices are endogenous, and the number of sectors s becomes a discrete (and exogenous) parameter. The domestic country is assumed small as an importer and so all foreign prices for the s sectors are given by an exogenous value, $\bar{p}$, common across all sectors.

The agricultural good R is treated as a numeraire. It is additively separable with a constant marginal utility, and hence income does not appear in the demand functions for the Q goods (though we will introduce a demand shifter later). Demand comes from domestic and foreign sales, respectively $Q_{dd}(s), Q_{df}(s)$ for sector s, and domestic and foreign goods are CES substitutes in each market with an elasticity of substitution $\varepsilon > 1$. Sectoral composites (domestic and foreign varieties) are Cobb–Douglas substitutes. The utility function and budget constraint that produces the demand functions are given in online appendix 6.

5.3. General equilibrium as a non-linear complementarity problem

Here we give the specification for the model with agglomeration economies, which has more equations and unknowns than the Ricardian model. The latter is simpler because the λs are exogenous, and the model nests the Ricardian model as a special case with the σ parameters equal to zero.

Non-negative variables:

L_i	Labour demand or employment in region i
w_i	Wages in region i
X_{ij}	Output of function j in region i
λ_{ij}	Labour requirements in function j in region j
$Q_d(s)$	Total output of sector s (all firm types)
$Q_{fd}(s)$	Domestic demand for foreign goods
$n_k(s)$	Output of type $k = 1, 2, F$ in sector s
$p(s)$	Price of (domestic) good s

With the dimension of s equal to 51, the model has 318 non-negative variables complementary to 318 weak inequalities. A strict inequality corresponds to a zero value for the complementary variable.[15] First, the supply–demand relationships for labour demand in the

15 The solution method for non-linear complementarity problems follows exactly from the Karush–Kuhn–Tucker theorem. Slack variables are added to each weak inequality to make them equations. Then, for each added slack variable, there is an added equation that requires that the product of the slack variable and the complementary variable (identified in each weak inequality (20) to (37)) be zero. The system of weak inequalities is thus converted to a (larger) set of equations, and the model is solved by an iterative procedure akin to a sophisticated Newton method.

two regions are given as follows, where $\perp$ denotes complementarity between the inequality and a variable. Labour is used in variables costs for all firm types in all sectors, plus used in fragmentation costs for fragmented sectors. We use a simple formulation of the fragmentation labour use, which divides it between the two regions, each using $t/2$ per F-type firm:

$$L_1 \geq \sum_s n_1(s)\left(a(s)\lambda_{A1} + b(s)\lambda_{B1}\right) + n_F(s)a(s)\lambda_{A1} + n_F(s)t/2 \perp L_1 \tag{20}$$

$$L_2 \geq \sum_s n_2(s)\left(a(s)\lambda_{A2} + b(s)\lambda_{B2}\right) + n_F(s)a(s)\lambda_{A2} + n_F(s)t/2 \perp L_2 \tag{21}$$

Second, from equation (11), wages are given by

$$\left(w_1 - w_0\right) K/c \geq L_1 \quad \perp \quad w_1 \tag{22}$$

$$\left(w_2 - w_0\right) K/c \geq L_2 \quad \perp \quad w_2 \tag{23}$$

Third, output levels of the two functions in the two regions are given by

$$X_{A1} \geq \sum\nolimits_s a(s)\left(n_1(s) + n_F(s)\right) \quad \perp \quad X_{A1} \tag{24}$$

$$X_{A2} \geq \sum\nolimits_s a(s)n_2(s) \quad \perp \quad X_{A2} \tag{25}$$

$$X_{B1} \geq \sum\nolimits_s b(s)n_1(s) \quad \perp \quad X_{B1} \tag{26}$$

$$X_{B2} \geq \sum\nolimits_s b(s)\left(n_2(s) + n_F(s)\right) \quad \perp \quad X_{B2} \tag{27}$$

Fourth, the labour input coefficients (inverse productivity) are given by

$$\lambda_{A1} \geq \Lambda_{A1} - \sigma_A \mathrm{X}_{A1} \quad \perp \quad \lambda_{A1} \tag{28}$$

$$\lambda_{A2} \geq \Lambda_{A2} - \sigma_A \mathrm{X}_{A2} \quad \perp \quad \lambda_{A2} \tag{29}$$

$$\lambda_{B1} \geq \Lambda_{B1} - \sigma_B \mathrm{X}_{B1} \quad \perp \quad \lambda_{B1} \tag{30}$$

$$\lambda_{B2} \geq \Lambda_{B2} - \sigma_B \mathrm{X}_{B2} \quad \perp \quad \lambda_{B2} \tag{31}$$

The volume of output in each sector is complementary to a zero-profit condition, that unit cost is greater than or equal to price. Fragmentation costs are incurred with a half unit each of urban labour of regions 1 and 2: $t\left(w_1 + w_2\right)/2$.[16] Therefore,

$$w_1\left(a(s)\lambda_{A1} + b(s)\lambda_{B1}\right) \geq p(s) \quad \perp \quad n_1(s) \tag{32}$$

$$w_2\left(a(s)\lambda_{A2} + b(s)\lambda_{B2}\right) \geq p(s) \quad \perp \quad n_2(s) \tag{33}$$

$$w_1 a(s)\lambda_{A1} + w_2 b(s)\lambda_{B1} + t\left(w_1 + w_2\right)/2 \geq p(s) \quad \perp \quad n_F(s) \tag{34}$$

Total output of good s is given by the sum the outputs across firm types:

$$Q_d(s) \geq n_1(s) + n_2(s) + n_F(s) \quad \perp \quad Q_d(s) \tag{35}$$

16 Note that this assumption makes (32) to (34) homogeneous of degree 0 in wages and prices.

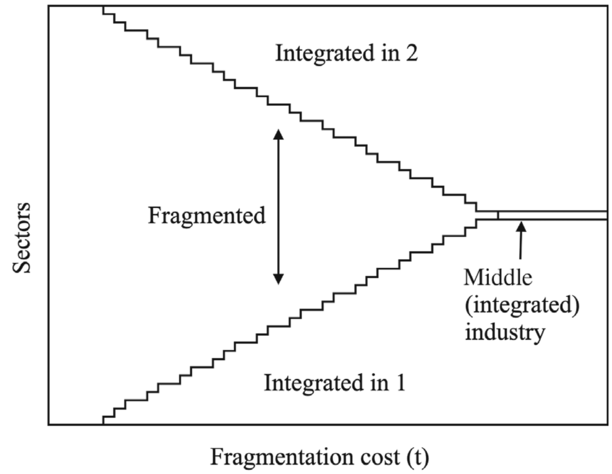

FIGURE 7 Symmetric Ricardian case (fragmentation cost t on horizontal axes)

The final element is to specify the demand size of the model, which links outputs, prices and the external foreign market. The market clearing equation for the domestic good s is that supply equal the sum of domestic and foreign demand. α_d and α_f are "shorthand" scaling parameters for domestic and foreign, that could depend on the relative market sizes for example (see online appendix). θ_d and θ_f are the weights on the domestic and foreign varieties in the nest for each sector s:

$$Q_d(s) = Q_{dd}(s) + Q_{fd}(s) = \frac{\alpha_d \theta_d p(s)^{-\epsilon}}{\theta_d p(s)^{1-\epsilon} + \theta_f \overline{p}^{1-\epsilon}} + \frac{\alpha_f \theta_d p(s)^{-\epsilon}}{\theta_f p(s)^{1-\epsilon} + \theta_f \overline{p}^{1-\epsilon}}, \perp p(s). \tag{36}$$

Domestic demand for foreign goods is not needed to solve the core model, but it is needed for welfare calculations after solution. These are given by

$$Q_{fd}(s) = \frac{\alpha_d \theta_f \overline{p}^{-\epsilon}}{\theta_d p(s)^{1-\epsilon} + \theta_f \overline{p}^{1-\epsilon}} \quad \perp Q_{fd}(s). \tag{37}$$

As noted above, the core model is then 318 weak inequalities complementary with 318 non-negative unknowns.

5.4. Symmetric Ricardian and localization economies in general equilibrium

Figures 7 to 11 present simulation results that develop economic implications of the model for the symmetric Ricardian case. Spillovers and asymmetric cases are found in online appendix 7. Figure 7 shows results with fragmentation costs t on the horizontal axis. Each column of the figure is a solution to the model for that value of t, as will be the case in the following figures (the jagged line is a consequence of the discreteness of sectors). The results naturally qualitatively resemble figure 4. At high t, all production is integrated in either one country or the other—except for the middle sector (there is an odd number of sectors), where integrated sectors produce in both countries.

Figures 8 shows further results for the case in figure 7 in two panels. The left panel of figure 8 show the G concentration indices for sectors and functions as defined in equation (1) above. The right-hand panel shows the D specialization indices for sectors and functions as defined in equation (2). Falling fragmentation costs lower both sector concentration and regional sector specialization and raise both regional function concentration and regional function specialization. Falling fragmentation costs thus mimic the empirical trends in these indices that we documented in section 2. The G and D indices seem to be conveying the

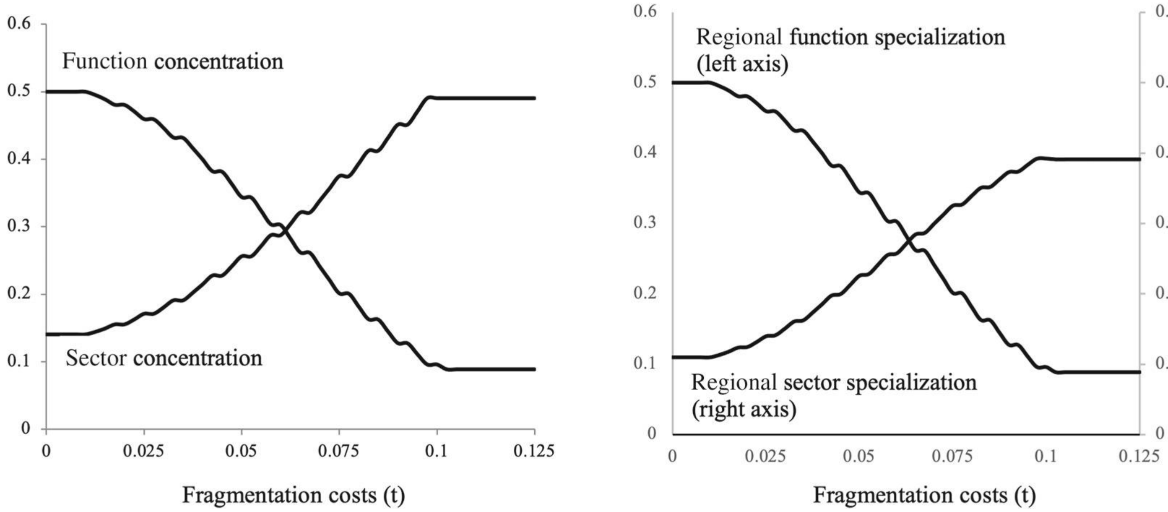

FIGURE 8 Concentration (D) and specialization (G) indices

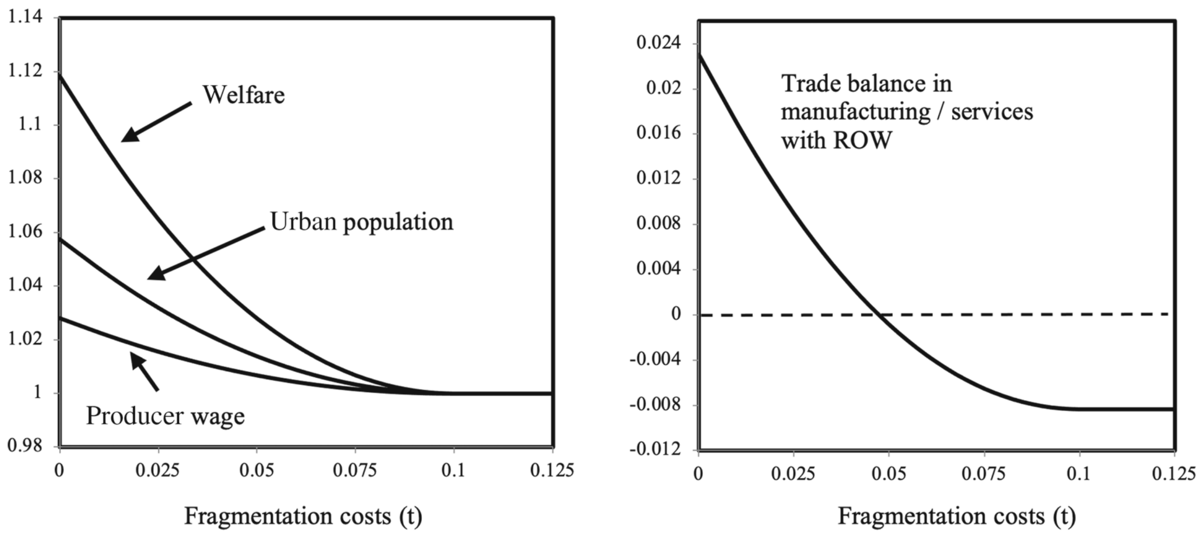

FIGURE 9 Welfare, wages and trade balance

same information in figure 3, but this is not a general result. It is due to the several symmetry assumptions in the model. If the regions are of quite different size (e.g., one region has absolute advantage in both functions), then sectors may all be concentrated in the larger region, but that region will also have a low sector specialization index. But the key qualitative properties about falling sector concentration/specialization and rising function concentration/specialization will remain as shown.

Figure 9 presents further economic implications of falling fragmentation costs. The left panel of figure 9 graphs the welfare, the "urban" population of both regions combined, and the producer (urban) wage (recall all workers earn a wage net of commuting costs and land rent equal to w_0). Note from equations (18) and (19) that the produce wage rises proportionally more slowly than the population. The increase in welfare as fragmentation costs fall is larger. The intuition behind all of these results is that falling fragmentation costs are analogous to an aggregate productivity improvement for the economy. More output can be produced for a lower cost in terms of the hinterland good, and so labour is reallocated

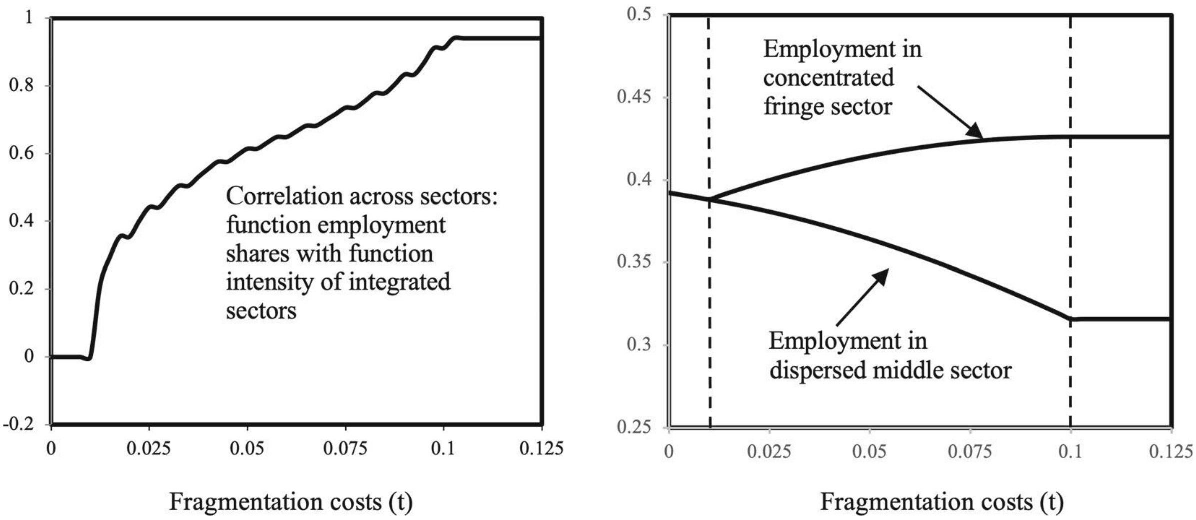

FIGURE 10 Intra and intersectoral shifts in technique and employment

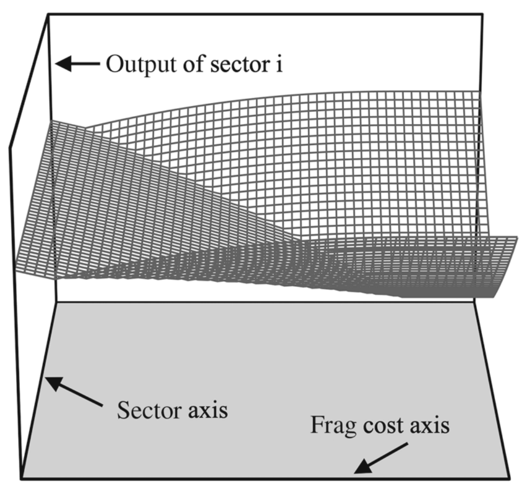

FIGURE 11 Sector outputs and comparative advantage

to the two regions, which, because of the congestion effect, raises the urban wage. Welfare rises as the sector goods become cheaper, so the real wage, w_0, divided by a price index for the s sectors rises.

The right panel of figure 9 illustrates an effect that was not discussed in previous sections. The fall in fragmentation costs improves the competitiveness of the s (manufacturing and services) sectors relative to the hinterland good. With Rest-of-World prices for their varieties of these sectors held constant, net exports increase with trade balance by increased net imports of the hinterland good. The vertical axis gives the trade balance (exports minus imports) of urban goods as a proportion all domestic urban goods production. The trade balance with the rest of the world is negatively related to fragmentation costs. Ease of internal transport and communications is a source of comparative advantage.

Figures 8 and 9 examine aggregate results from the simulation. Now we turn to analyzing economic implications of fragmentation at a more disaggregated level across sectors in the remaining figures. The left panel of figure 10 is trying to quantify an idea expressed earlier in the paper, that fragmentation makes the regions look less like classic Heckscher–Ohlin economies. The correlation shown in the left panel is as follows. Take function A in region 1. Form a vector across the s sectors for the share of employment of function A in total employment of each sector. We could refer to this as the technology observed in each sector. A second vector is the share of function A in total employment for an integrated sector. The calculate the correlation between the two vectors, with the result shown in the left-hand panel of figure 10.

When fragmentation costs are high, the technology employed in each sector of a region is identical to production technology of an integrated sector. Production specialization between the regions looks very Heckscher–Ohlin, with each region specializing in the sectors using intensively their comparative advantage function (as opposed to using intensively their abundant factor in HO). As fragmentation costs fall, actual employment in sectors in a region spreads out across the sectors, to the point where all employment shares for function A in region 1 equal one, and the correlation goes to zero. We could say that, when fragmentation costs are high, comparative advantage is found or observed in sectors (even though that is indirectly derived as in HO), while when fragmentation costs are low, comparative advantage lies directly in functions.

As noted earlier, we are primarily interested in the within-sector and occupation changes in concentration, since the results shown in figures 1 and 2 (the decomposition in equation (8)) indicate that the within effect is generally dominate, especially for occupations. But the general equilibrium model of this section does feature an intersectoral reallocation as well. Specifically, there is a reallocation of employment from concentrated sectors that are still not fragmented to dispersed sectors that are already fragmented as t continues to fall. The right panel of figure 10 shows the effect of employment reallocation from the integrated "fringe" sector 51 with the most extreme function intensity spread toward the fragmented middle sector 26, which uses both functions in equal shares. Between the dashed vertical lines, neither sector changes its status yet employment shifts.

The intuition is fairly straightforward. At high fragmentation costs, the middle sector is the most penalized, having to draw half its inputs from an expensive local source. Conversely, that sector benefits the most as the cost t falls. Falling t does not direct affect the fringe sectors until they fragment, but the falling t does indirectly affect the fringe sector. The lower costs of production in the middle sectors leads them to expand output and draw labour into the region centers. This raises the urban wage for the fringe sectors, leading to lower outputs and higher prices. In general equilibrium, falling fragmentation costs also lead to an intersectoral shift of employment from concentrated to dispersed sectors. This effect reinforces the within-sector effect of equation (8) such that the theory also produces a total effect curve (left-hand side of (8)) that is steeper that the within effect alone.[17]

Our final figure 11 continues to illustrate the heterogeneous effects across sectors, with much the same intuition as the right-hand panel of figure 10. Figure 11 shows the effect of falling fragmentation costs on sector outputs. At high t, the smallest sectors are the middle sectors, due to drawing one function from a costly local source. As t falls,

17 The left-hand panel of figure 10 is the total effect, the left-hand side of equation (8). We do not emphasize the intersectoral and occupation shifts in this paper, in part because there are many other candidates that contribute to this effect, such as technical change and import competition in manufacturing, and rising incomes that shift demand from goods to (likely) less-concentrated services.

the fragmenting middle sectors get a productivity boost and increase outputs, while the fringe sectors are harmed a little by rising urban wages. At zero t, all sectors produce the same output.

These sectoral differences in figure 11 map directly into national comparative advantage and net exports by sector (not shown). At high t, the costly middle sectors are net importers in this example, while the fringe sectors are net exports. As t falls, some fringe sectors actual switch signs due to the rising urban wage. But all sectors become net exporters at a sufficiently low value of t and the country is fully specialized in exporting the s goods and importing the hinterland good (for the parameters chosen in this example),

6. Conclusions

Our paper is motivated by what are widely seen as changes in the range of activities and occupations performed in urban areas. Our approach is necessarily circumscribed by the requirements of formal theory and data analysis, but many of the ideas here are consistent with the broad analysis and vision of Moretti (2012) for example.

We begin with an empirical exercise on US State data on employment by sector–occupation–state. Results show that a concentration index for industries and a regional specialization index in industries have both fallen over a 30-year period. Importantly, most of the fall is within industries and so the decrease is not primarily explained by employment moving from concentrated to less concentrated sectors. Second, results show that a concentration index for occupations and a regional specialization index for occupations have both risen over a 20-year period. As with (but opposite to) the indices for industries, this is not due to employment shifting from less concentrated to more concentrated occupations but occurs within occupations.

Using these results as motivation, we construct a model that can capture the features of the data. The key and novel aspect of the model is that regions have comparative advantage in functions (occupations in the data) rather than sectors. This comparative advantage may be Ricardian (exogenous) or due to agglomeration economies (arising endogenously between places that are ex ante identical). We draw on concepts and analyses from a number of fields of study including international trade, multinational corporations, urban economics and economic geography. Industries (sectors) produce with a range of functions. A sector in a region may produce with only locally sourced functions or may draw functions from other locations, the latter referred to as fragmentation. Our model creates a distribution of fragmented and integrated production across industries and across regions and identifies the characteristics of industries that are fragmented versus integrated, and of the regions in which integrated production occurs.

A key variable in our theory is a cost of geographically separating the sourcing of function inputs into a sector, referred to as the fragmentation cost. Our principal result is that, at high fragmentation costs, a region's employment is concentrated in certain sectors, with each sector's employees performing many different functions. At low fragmentation costs, a region's employment switches to being concentrated in certain functions, with employees in a particular function doing work for many different sectors. Instead of a region having a range of production workers, managers, lawyers and accountants working in a few sectors, it comes to have a smaller range of functions, for example lawyers or accountants, working for many different sectors, often at a distance.

Second, we use the same data to calculate measures of regional specialization, more in line with a traditional international trade approach. With the confines of our theory model, these measures of regional specialization in sectors and functions should be qualitatively similar to the concentration measures and indeed they are in our simulations.

Thus, falling fragmentation costs in the model mimic the changes observed in the data over time.

The final section of the paper extracts added economic insights from the general equilibrium simulation model. Falling fragmentation costs are analogous to a productivity improvement, so at the national level, the falling cost leads to higher welfare, urban population, producer (urban) wages and an improved trade balance in urban goods and services with the Rest-of-World. But this hides considerable heterogeneity across sectors. Sectors that require large proportions of both functions benefit the most from falling fragmentation costs and may change from being net importing sectors to net exporters

Appendix

TABLE A1
Employment by function f = A, B, in sector s and region r = 1, 2

	Region 1	Region 2
	Integrated in 1: $0 < s < s_1$	
Function A	$L_{A1}(s) = n(s)a(s)\lambda_{A1}$	$L_{A2}(s) = 0$
Function B	$L_{B1}(s) = n(s)b(s)\lambda_{B1}$	$L_{B2}(s) = 0$
	Fragmented: $s_1 < s < s_2$	
Function A	$L_{A1}(s) = n(s)a(s)\lambda_{A1}$	$L_{A2}(s) = 0$
Function B	$L_{B1}(s) = 0$	$L_{B2}(s) = n(s)b(s)\lambda_{B2}$
	Integrated in 2: $s_2 < s < 1$	
Function A	$L_{A1}(s) = 0$	$L_{A2}(s) = n(s)a(s)\lambda_{A2}$
Function B	$L_{B1}(s) = 0$	$L_{B2}(s) = n(s)b(s)\lambda_{B2}$
	L_{fr}: Employment in each function/region (all sectors)	
Function A	$L_{A1} = \int_0^{s_2} L_{A1}(s)ds$	$L_{A2} = \int_{s_2}^1 L_{A2}(s)ds$
Function B	$L_{B1} = \int_0^{s_1} L_{B1}(s)ds$	$L_{B2} = \int_{s_1}^1 L_{B1}(s)ds$
	L_{sr}: Employment in each sector/region (all functions)	
	$L_{s1} = \Sigma_{f=A,B} L_{f1}(s)$	$L_{s2} = \Sigma_{f=A,B} L_{f2}(s)$
	L_r :Total employment in each region	
	$L_1 = L_{A1} + L_{B1} = \int_0^1 L_1(s)ds$	$L_2 = L_{A2} + L_{B2} = \int_0^1 L_2(s)ds$

TABLE A2
Employment by function f = A, B, in sector s and region r = 1, 2

	Region 1	Region 2
I	Integrated in 1: $0 < s < s_1$	
Function A	$L_{A1}(s) = n\lambda_{A1}[1+\gamma(1-2s)]/2$	$L_{A2}(s) = 0$
Function B	$L_{B1}(s) = n\lambda_{B1}[1-\gamma(1-2s)]/2$	$L_{B2}(s) = 0$
II	Fragmented: $s_1 < s < s_2$	
Function A	$L_{A1}(s) = n\lambda_{A1}[1+\gamma(1-2s)]/2$	$L_{A2}(s) = 0$
Function B	$L_{B1}(s) = 0$	$L_{B2}(s) = n\lambda_{B2}[1-\gamma(1-2s)]/2$
III	Integrated in 2: $s_2 < s < 1$	
Function A	$L_{A1}(s) = 0$	$L_{A2}(s) = n\lambda_{A2}[1+\gamma(1-2s)]/2$
Function B	$L_{B1}(s) = 0$	$L_{B2}(s) = n\lambda_{B2}[1-\gamma(1-2s)]/2$
IV	L_{fr}: Employment in each function/region (all sectors)	
Function A	$L_{A1} = \lambda_{A1}s_2[1+\gamma(1-s_2)]n/2$	$L_{A2} = \lambda_{A2}(1-s_2)(1-\gamma s_2)n/2$
Function B	$L_{B1} = \lambda_{B1}s_1[1-\gamma(1-s_1)]n/2$	$L_{B2} = \lambda_{B2}(1-s_1)(1+\gamma s_1)n/2$
V	L_{sr}: Employment in each sector/region (all functions)	
	$L_{s1} = \Sigma_{f=A,B} L_{f1}(s)$	$L_{s2} = \Sigma_{f=A,B} L_{f2}(s)$
VI	L_r : Total employment in each region	
	$L_1 = L_{A1} + L_{B1} = \int_0^1 L_1(s)ds$	$L_2 = L_{A2} + L_{B2} = \int_0^1 L_2(s)ds$

Supporting information

Supplementary material accompanies this article. The data and code that support the findings of this study are available in the Canadian Journal of Economics Dataverse at `https://doi.org/10.5683/SP3/4URVZP`.

References

Acosta, C., and D. H. Lyngemark (2021) "The internal spatial organization of firms: Evidence from Denmark," *Journal of Urban Economics* 124, 103366

——— (2020) "Spatial wage differentials, geographic frictions and the organization of labor within firms." Available at SSRN: `https://ssrn.com/abstract=3664082` or `https://doi.org/10.2139/ssrn.3664082`

Aiginger, K., and E. Rossi-Hansberg (2006) "Specialization and concentration: A note on theory and evidence" *Empirica* 33(4), 255–66

Antràs, P., and D. Chor (2021) "Global value chains," NBER working paper no. 28549

Audretsch, D.B., and M. P. Feldman (1996) "R&D spillovers and the geography of innovation and production," *American Economic Review* 86, 630–40

Autor, D.H. (2013) "The 'task approach' to labor markets: An overview," *Journal for Labor Market Research* 46, 185–99

Barbour, E., and A. Markusen (2007) "Regional occupational and industrial structure: Does one imply the other?," *International Regional Science Review* 30, 92–90

Barrios S., L. Bertinelli, E. Strobl, A.-C. Teixeira (2005) “The dynamics of agglomeration: Evidence from Ireland and Portugal,” *Journal of Urban Economics* 57, 170–88
Behrens, K., G. Duranton, and F. Robert-Nicoud (2014) “Productive cities: Sorting, selection, and agglomeration,” *Journal of Political Economy* 122, 507–53
Behrens, K., and R. Guillain (2017) “The determinants of coagglomeration: Evidence from functional employment patterns,” CEPR discussion paper no. 11884
Berry, C.R., and E. L. Glaeser (2005) “The divergence of human capital levels across cities,” *Papers in Regional Science* 84(3), 407–44
Brakman, S., and C. van Marrewijk (2013) “Lumpy countries, urbanization and trade,” *Journal of International Economics* 89, 252–61
Charnoz, P., C. Lelarge, and C. Trevien (2018) “Communication costs and the internal organization of multi-plant businesses: Evidence from the impact of the French high-speed rail,” *Economic Journal* 128, 949–94
Courant, P.N., and A. Deardorff (1992) “International trade with lumpy countries,” *Journal of Political Economy* 100, 198–210
Crafts, N., and A. Klein (2021) “Spatial concentration of manufacturing industries in the United States: Re-examination of long-run trends,” European Review of Economic History 25(2), 223–46
Dauth, W., M. Fuchs, and A. Otto (2018) “Long-run processes of geographical concentration and dispersion: Evidence from Germany,” *Papers in Regional Science* 97(3), 569–93
Davis, D.R., and J. Dingel (2018) “The comparative advantage of Ccities,” *Journal of International Economics* 123, 103291
Diamond, R. (2016) “The determinants and welfare implications of US workers’ diverging location choices by skill: 1980–2000,” *American Economic Review* 106(3), 479–524
Dumais, G., G. Ellison, and E. L. Glaeser (2002) “Geographic concentration as a dynamic process,” *Review of Economics and Statistics* 84(2), 193–204
Duranton, G., J. V. Henderson, and W. Strange (eds.) (2015) *Handbook of Regional and Urban Economics*, vol. 5. Amsterdam: North Holland
Duranton, G., and H. G. Overman (2005) “Testing for localization using micro-geographic data,” *Review of Economic Studies* 72(4), 1077–106
Duranton, G., and D. Puga (2005) “From sectoral to functional urban specialization,” *Journal of Urban Economics* 57, 343–70
Eckert, F. (2019) “Growing apart: Tradable services and the fragmentation of the US economy,” mimeo, Yale University
Eckert, F., T. C. Fort, P. K. Schott, and N. J. Yang (2021) “Imputing missing values in the US Census Bureau’s county business patterns,” NBER working paper no. 28549
Eckert, F., S. Ganapati, and C. Walsh (2020) “Skilled scalable services: The new urban bias in economic growth,” CESifo working paper no. 8705
Ellison, G., and E. L. Glaeser (1997) “Geographic concentration in US manufacturing industries: A dartboard approach,” *Journal of Political Economy* 105, 889–927
Feenstra, R.C, and G. H. Hanson (1996) “Globalization, outsourcing, and wage inequality,” *American Economic Review* 86, 240–45
Fujita, M., P. Krugman, and A. J. Venables (1999) *The Spatial Economy: Cities, Regions and International Trade*. Cambridge, MA: The MIT Press
Gabe, T.M., and J. R. Able (2012) “Specialized knowledge and the geography concentration of occupations,” *Journal of Economic Geography* 12, 435–53
——— (2016) “Shared knowledge and the coagglomeration of occupations,” *Regional Studies* 50(8), 1360–73
Grossman, G.M., and E. Rossi-Hansberg (2008) “Trading tasks: A simple theory of off-shoring,” *American Economic Review* 98, 1978–97
——— (2012)“Task trade between similar countries,” *Econometrica* 80, 593–629
Henderson, V., and J.-F. Thisse (eds.) (2004) *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: North Holland
Krugman, P. (1991a) *Geography and Trade*. Leuven: Leuven University Press

——— (1991b) "Increasing returns and economic geography" *Journal of Political Economy* 99, 483–99

Markusen, J.R. (1989) "Trade in producer services and in other specialized intermediate inputs," *American Economic Review* 79, 85–95

——— (2002) *Multinational Firms and the Theory of International Trade*. Cambridge, MA: The MIT Press

Michaels, G., F. Rauch, and S, Redding (2019) "Task specialization in US cities from 1880–2000," *Journal of the European Economic Association* 17, 754–98

Moretti, E. (2012) *The New Geography of Jobs*. Boston: Mariner Press

——— (2013) "Real wage inequality," *American Economic Journal: Applied Economics* 5(1), 65–103

Rossi-Hansberg, E., P.-D. Sarte, and R. Owens, III (2009) "Firm fragmentation and urban patterns," *International Economic Review* 50, 143–86

Storper, M., and A. J. Venables (2004) "Buss: Face-to-face contact and the urban economy," *Journal of Economic Geography* 4, 351–70

Timmer, M. P, S. Miroudot, and G. J. de Vries (2019) "Functional specialisation and trade," *Journal of Economic Geography* 19, 1–30