Regional specialisation:
from the geography of industries to the geography of jobs

Antoine Gervais

James R. Markusen

Anthony J. Venables

Our analysis begins with an empirical investigation of how employment concentration in industries and occupations across United States regions 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. Region-function-specific productivity differences 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.

Keywords: regional specialisation, regional trade, fragmentation, geographic concentration

JEL codes: F12, R11, R12, R13

Authors' affiliations

Antoine Gervais
Department of Economics
Université de Sherbrooke
Antoine.Gervais2@USherbrooke.ca

James R. Markusen
Department of Economics
University of Colorado
james.markusen@colorado.edu

Anthony J. Venables
The Productivity Institute, University of Manchester
& Monash University
anthony.venables@manchester.ac.uk

1. Introduction

From popular press reports to formal journal articles, much has been written about the changing nature of work 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 also 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 which 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 (discussed in the following section). We calculate how employment concentration in industries and occupations across US regions 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 specialisation by sector and increasing specialisation 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 a number of 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 various 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 a homogeneous primary factor - labor - and we refer to them as functions. The key feature of our approach is that regions differ in the relative productivity of labor 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.

A final ingredient in our approach is labor mobility between regions, a typical assumption in the urban/regional literature. This endogenizes nominal wages (as faced by producers). Migration equalizes real wages, but a larger population in a region, other things equal, implies higher land prices and (in an urban context) longer average commutes, thus creating regional variation in the cost of living and hence in nominal wages.

The model creates a distribution of fragmented and integrated sectors across industries and across regions and identifies the characteristics of industries that are fragmented versus integrated, and of the regions in which integrated sectors locate. Falling fragmentation costs are then the key experiment applied to the model. The central result is that, as these costs fall, regions become more specialized across functions and less specialized in sectors. At high fragmentation costs, general equilibrium resembles the archetype HO model of international trade theory. At low fragmentation costs, comparative advantage lies in functions, with employment in a region dispersed across many sectors.

Turning from regional specialization to sector and function concentration across regions, the model similarly predicts that sectors become less concentrated as some of their employment is spread across regions. But functions become more concentrated as employment in a function occurs in fewer regions. Here is a simple example to illustrate this result. 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.

Ideas in this paper are complementary to the influential paper by Duranton and Puga (2005, henceforth D&P) on sectoral and functional specialisation. By design, the present paper is tailored to be simpler than D&P in a number of respects, having perfect competition (rather than monopolistically competitive input sectors), and a given set of places (rather than endogenous city formation). D&P have a tight input-output structure of business services, headquarters, production plants, and intermediates to production, in contrast to our twofold classification of sectors and functions, with all sectors using a mix of functions in different proportions. The main advantage of our approach is that it gives a relatively flexible way of thinking about the interactions between the range of functions and range of sectors present in a region, as compared to the central proposition of D&P.[1]

Our focus on functions is also distinct from the literature on trade in tasks (for example Grossman and Rossi-Hansberg, 2008).[2] We think of there as being relatively few functions (law, engineering, accountancy) most of them used by many sectors, as compared to the task approach of many tasks, each specific to a single sector. Fundamentally, the task literature asks questions about international trade

[1] D&P proposition 1 states that (depending on parameters) either all firms are fully integrated and all cities fully specialized; or all firms are fragmented (multi-locational) with each city fully specialized in either headquarters and business services, or in one sector's production and intermediate suppliers.

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

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 specialisation.

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, Henderson and Thisse (2015)). Relevant work includes Audretsch and Feldman (1996), Berhens, Duranton and Robert-Nicoud (2014), Brackman and van Marrewijk (2013), Courant and Deardorff (1992), Davis and Dingel (2018), Fujita, Krugman and Venables (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 documents 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 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 to those of small cities in 1997.

The remainder of the paper is as follows. In section 2, we analyze the data using region-level information on production and employment by sector and occupation for US states for the period 1990-2019 for industries, 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 characterise 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 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 the regions' 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

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

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

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

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)). However, as explained by Ellison and Glaeser (1997, henceforth E&G), 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, E&G develop a more sophisticated index of concentration which, for sectors, is defined as

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

where $H_s = \sum_j z_{js}^2$ is the Herfindahl index of the sector's plant size distribution and z_{js} is the j^{th} plant's share of sectoral employment. E&G refer to G_s, defined in equation (1) above, as the "raw geographic concentration" of employment in a sector. The adjustments seen in equation (3), such as the subtraction of H_s, are corrections that account for the fact that G_s is expected to be larger in sectors consisting of fewer larger plants.[3]

To measure functional concentration, we use a modified version of the E&G index defined as

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

As for sectors, the index adjusts the raw measure of concentration 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 to 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.

The E&G indices of concentration defined in equations (3) and (4) have many useful properties.[4] First, they are easy to implement. Second, they are widely used which allows us to compare our results with previous studies. Third, they use employment shares, which implies that it does not confound features in time-series data such as the general decline in manufacturing.

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 to the national

[3] In practice (see results in appendix 2 of the paper), 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-industry comparisons. Nevertheless, we use EG_s as our benchmark measure.

[4] The motivation for the E&G 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 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 E&G and refer to those industries with *EG*s above 0.05 as being concentrated and to those with *EG*s below 0.02 as being dispersed.

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

$$D_r^{sector} = \sum_s (q_{rs} - q_s)^2, \qquad \text{where } q_{rs} = L_{sr} / \sum_s L_{sr}, \text{ and } q_s = L_s / \sum_s L_s, \tag{5}$$

$$D_r^{function} = \sum_f \left(q_{rf} - q_f\right)^2, \qquad \text{where } q_{rf} = L_{fr} / \sum_f L_{fr}, \text{ and } q_f = L_f / \sum_f L_f. \tag{6}$$

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 in 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 dataset that contains state-level data on 626 six-digit NAICS industries (our empirical measure of sectors) for years 1990 to 2019. Data sources and measurement issues are discussed in appendix 1 at the end of the paper. In our sample, about 41 percent 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 appendix 2 at the end of the paper.

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 Table A2.1 of appendix 2 of the paper). Taking into account 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 as

$$EG_\tau^{sector} = \sum_s m_{s\tau} \, EG_{s\tau}. \tag{7}$$

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).

In appendix Table A2.2 we report estimated time series trend in geographic concentration for each broad economic sectors. 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 previous studies, such as Dumais, Ellison, and Glaeser (2002) who study the geographic concentration of sectoral employment across US states from 1972 to 1997.[5]

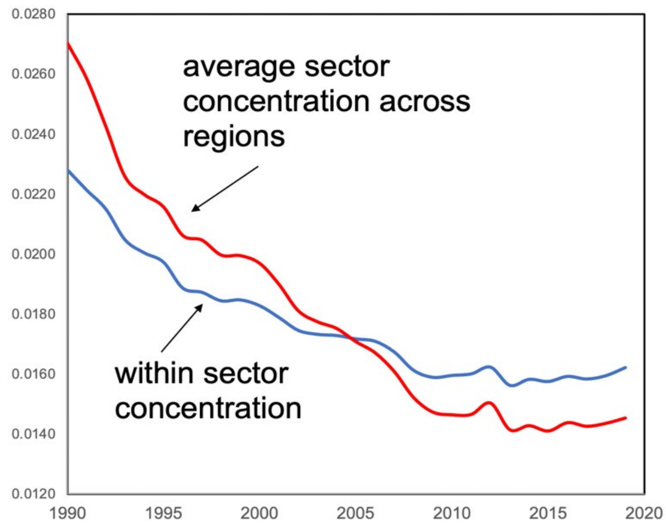


Figure 1: Geographic concentration of sectors over time

The results so far suggest that the average worker is employed in a more geographically dispersed sector in 2019 than he was in 1990. To gain additional insights, we decompose time series changes in the geographic concentration into two adjustments margins: within-industry changes in geographic concentration and across-industry reallocation of employment. While the theoretical model we develop in the next section produced across-sectors reallocation, 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 (7) as follows

$$EG_{\tau}^{sector} = \sum_s m_s \, EG_{s\tau} + \sum_s (m_{s\tau} - m_s) \, EG_{s\tau}, \tag{8}$$

[5] First, the two sets of estimates are of the same magnitude. They report a (simple) mean 0.034 for 1992. As reported in 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.

where $m_{s\tau}$ is industry-s's share of national employment in year τ and m_s is the industry'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-industry changes in concentration over time. The second term captures the remainder of the time series change.

The results from decomposition (8) 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 (8) 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 (8)). 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 (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 labor movement from less concentrated industries towards more concentrated industries, our results suggest that the decline in the within-industry 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 the 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. Data sources and measurement issues are discussed in appendix 1 at the end of the paper.

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). Taking into account 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 as

$$EG_\tau^{function} = \sum_f m_{f\tau} \, EG_{f\tau}. \tag{9}$$

The weighted average concentration increases by about 19% over the period (going from 0.0053 in 2000 to 0.0063 in 2019).

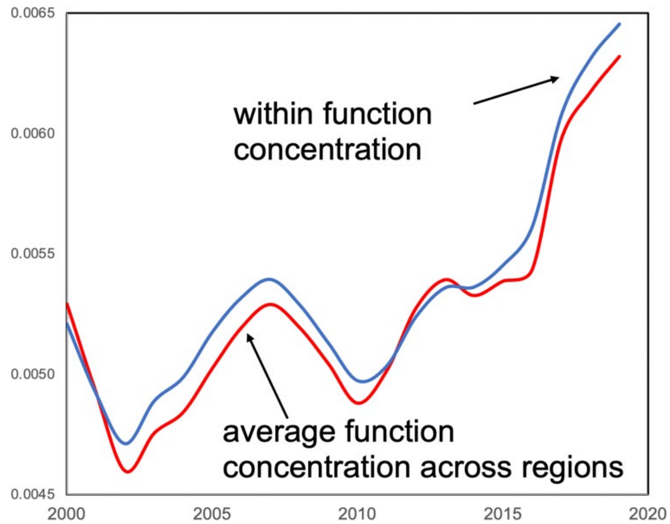


Figure 2: Geographic concentration of functions over time

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. Appendix Table A2.4 at the end of the paper sreport estimated time series trend in geographic concentration for each broad occupations categories. 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.

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 datasets described in the previous section to construct the two measures of regional specialization defined in equations (5) and (6) for each state-year in our datasets. In each case, we aggregate state-level measures using a weighted average, where the weights are the states' shares of national employment in the corresponding year

$$D_\tau^{sector} = \sum_r m_{r\tau} D_{r\tau}^{sector}, \qquad \text{and} \qquad D_\tau^{function} = \sum_r m_{r\tau} D_{r\tau}^{function}. \tag{10}$$

The main results are reported in Figure 3 (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 time. Conversely, the upward trend in the blue line indicates that states' distribution of employment across function is becoming increasingly uneven.

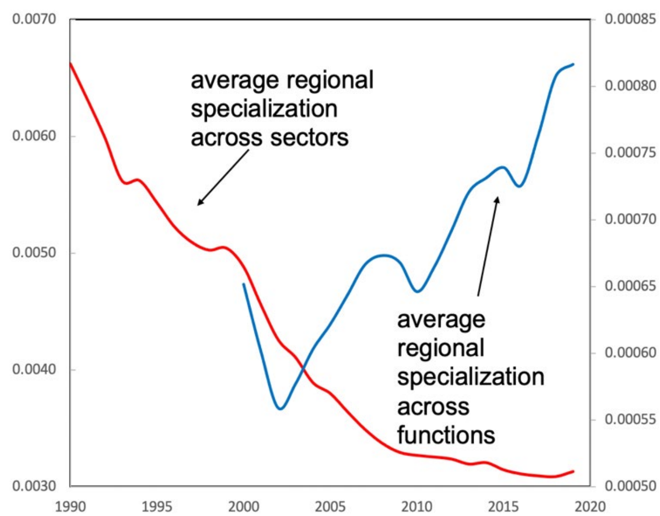


Figure 3: Regional specialization over time

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 regional specialisation, with decreased specialisation in sectors, but increased specialisation in functions. These findings suggests 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 specialisations which then shape the location of employment in a manner consistent with the data.

The ingredients of the model are locations, 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. [6] 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 $a(s)$ units of function A per unit output, and $b(s)$ units of function B, technical coefficients which we

[6] 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.

refer to as the function intensity of the sector.[7] 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-intensive, 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 is 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

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

Unit costs are those of the functions purchased, sector s using $a(s)$ units of function A and $b(s)$ units of B. 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 entirely 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. Since 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 by

---

[7] $a(s)$ 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 only operates in circumstances where there is sufficient spatial variation in productivity or wages, and sufficiently low costs of fragmentation.

$$\pi_F(s_1) - \pi_1(s_1) = b(s_1)[\lambda_{B1}w_1 - \lambda_{B2}w_2] - t = 0, \qquad (12)$$

$$\pi_F(s_2) - \pi_2(s_2) = a(s_2)[\lambda_{A2}w_2 - \lambda_{A1}w_1] - t = 0.$$

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 eqn. (11) and are given in appendix Table A3.1. 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 specialisation in symmetric equilibria

We start by analysing 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 (4.1) and then turning to economies of scale (4.2). Full general equilibrium is set out in section 5.

4.1 Functional productivity: Ricardian differences

Throughout this section, we make strong assumptions which make regions and sectors symmetrical, enabling us to derive key results on the location of sectors of 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 come from eqns. (12), and are implicitly given by

$$b(s_1)w\Delta\lambda = t, \qquad \text{and} \qquad a(s_2)w\Delta\lambda = t. \qquad (13)$$

A simple case is where the function intensity of sectors as 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.[8] Appendix Table A3.2 gives employment levels by region, function, and sector, replicating Table A3.1 with explicit expressions derived from this functional form. The unit profit functions of eqn. (11) become, $\pi_1(s) =$

[8] 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.

$p(s) - \{2\lambda + \Delta\lambda[1 - \gamma(1 - 2s)]\}w/2$, $\pi_F(s) = p(s) - \lambda w - t$, 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

$$\pi_1(s_1) = \pi_F(s_1): \quad s_1 = \frac{1}{2}\left[1 - \left(1 - \frac{2t}{\mathrm{w}\Delta\lambda}\right)\frac{1}{\gamma}\right], \tag{14}$$

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

These relationships capture the way in which the sourcing of functions in each sector depends on fragmentation costs *t* relative to wages, 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 $t^* = w\Delta\lambda/2$ then fragmentation becomes profitable, first in sectors that have similar use of both functions, i.e. *s* in an interval around ½ and of width $s_2 - s_1 = (1 - 2t/w\Delta\lambda)/\gamma$, wider the smaller is *t*, and the larger are productivity differences, $\Delta\lambda$. Intuitively, these are the sectors where both functions have a similar share of 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 specialisation 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 specialised. In simulation analysis of section 5.3 specialisation and concentration indices are calculated for the distribution of both sectoral and functional employment across regions in a full general equilibrium context.

4.2 Functional productivity: localisation economies

Ricardian efficiency differences provide the simplest model framework, but we think it unlikely that regional differences in the productivity of functions are principally due 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.

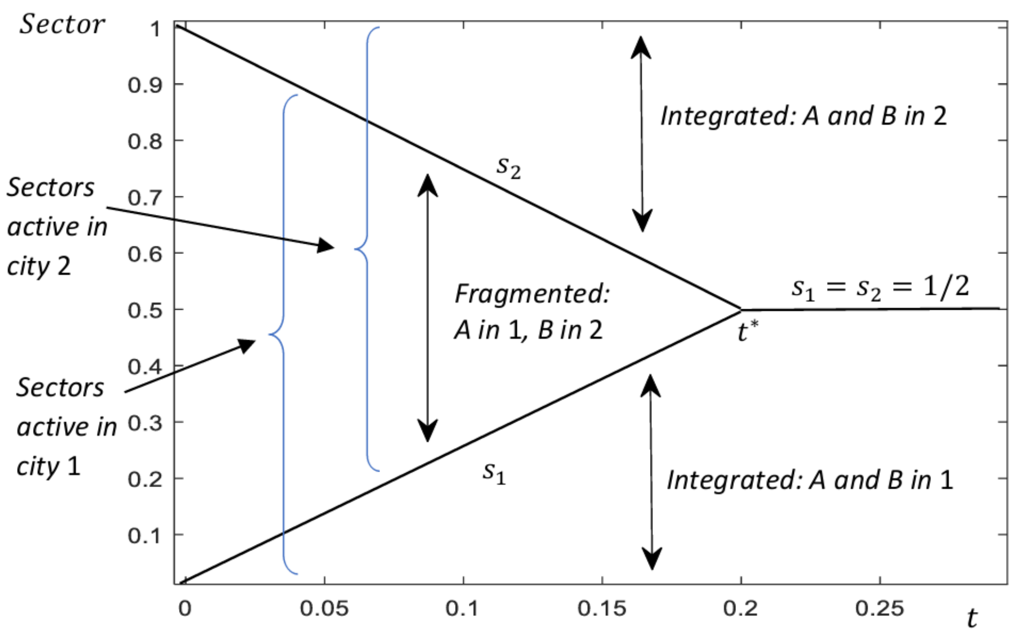


Figure 4: Modes of operation IN each sector *s*.

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, giving

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

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

Hence, productivity differentials are, using expressions from appendix Table A3.2, block IV,

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

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

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 eqn. (2) become,

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

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

To analyse these relationships, we focus on (16a) and (17a), the other pair, (16b) and (17b), being symmetric. Substituting (16a) in (17a) gives $\pi_F(s_1) - \pi_1(s_1)$ as a function of s_1. This expression and full analysis is given in appendix 3. Here we note the following facts and illustrate outcomes on Figure 5.

First, there is full integration if $\pi_F(s_1) \leq \pi_1(s_1)$ at s_1 = ½, and straightforward calculation gives value $t^{**} = [\Delta\lambda + n\sigma\gamma/4]w/2$ at which $\pi_F(s_1) = \pi_1(s_1)$. Evidently, this reduces to the Ricardian equivalent t^* if $\sigma = 0$, while $\sigma > 0$ implies a strictly higher critical point t^{**}. At higher values of t, $t \geq t^{**}$, 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 $(t^{**}, \tilde{t})$.[9] Integrated production is an equilibrium, because at this equilibrium productivity differences are small. But so too is a fragmented 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. Appendix 3 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 specialisation, and increases functional concentration and specialisation. 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, and 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.

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

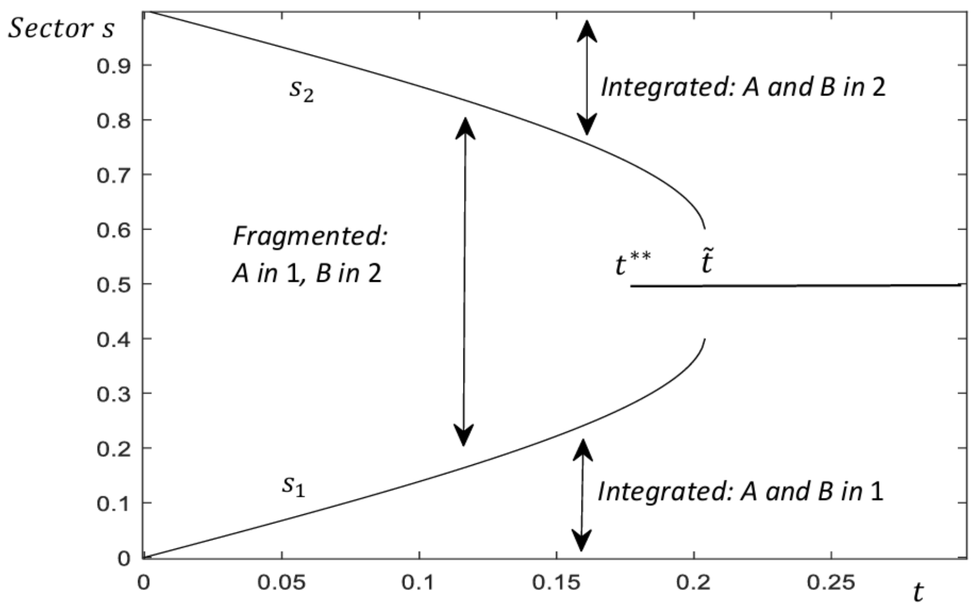


Figure 5: Modes of operation in each sector s: with increasing returns

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.

The following diagram 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) are 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.

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.

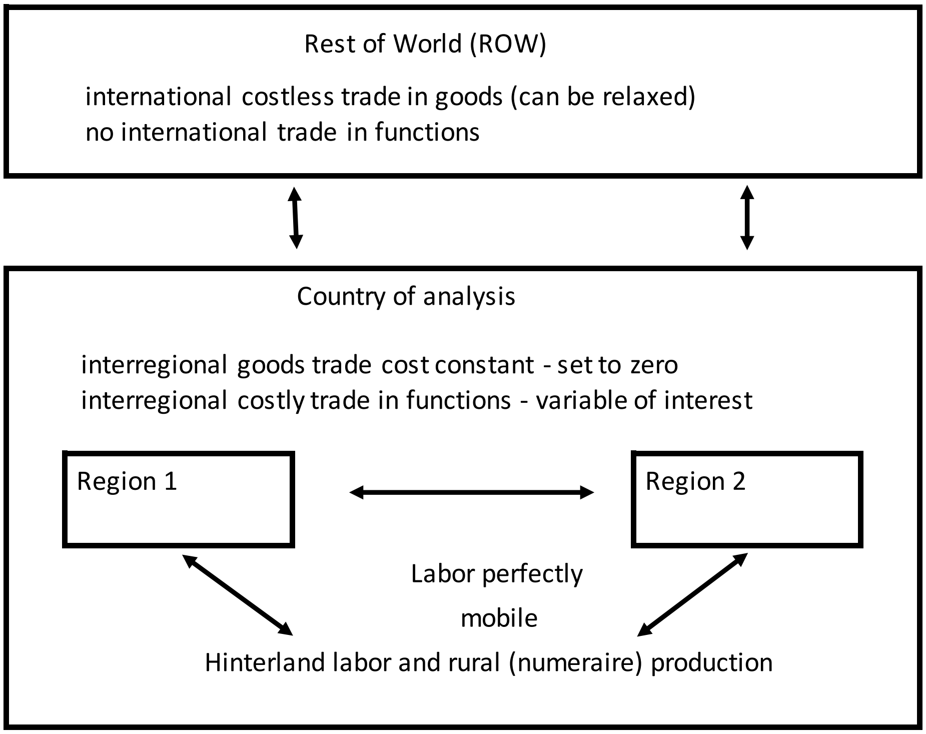


Figure 6: The spatial structure of the general-equilibrium model

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 which 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 micro-foundations 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 has to 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

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

These equations imply that, given the number of cities and commuting costs, regions with a larger population and labour force have to pay higher wages in order 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

$$H_r = c_r L_r^2/2K_r. \tag{19}$$

Thus, while workers' utility is equalised 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 good 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 appendix 2 of the paper.

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	labor demand or employment in region i
w_i	wages in region i

X_{ij}	output of function j in region i
λ_{ij}	labor 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. First, the supply-demand relationships for labor demand in the two regions are given as follows, where ⊥ denotes complementarity between the inequality and a variable. Labor 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 labor use, which divides it between the two regions, each using $t/2$ per F type firm.

$$L_1 \ \geq \textstyle\sum_s n_1(s)(a(s)\lambda_{A1} + b(s)\lambda_{B1}) + n_F(s)a(s)\lambda_{A1} + \ n_F(s)t/2 \qquad \perp \ L_1 \qquad (20)$$
$$L_2 \ \geq \textstyle\sum_s n_2(s)(a(s)\lambda_{A2} + b(s)\lambda_{B2}) + n_F(s)a(s)\lambda_{A2} + \ n_F(s)t/2 \qquad \perp \ L_2 \qquad (21)$$

Second, from eqn. (11) wages are given by:

$$(w_1 - w_0)K/c \geq L_1 \qquad \perp \qquad w_1 \qquad (22)$$
$$(w_2 - w_0)K/c \geq L_2 \qquad \perp \qquad w_2 \qquad (23)$$

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

$$X_{A1} \ \geq \textstyle\sum_s a(s)\big(n_1(s) + n_F(s)\big) \qquad \perp \qquad X_{A1} \qquad (24)$$
$$X_{A2} \ \geq \textstyle\sum_s a(s)n_2(s) \qquad \perp \qquad X_{A2} \qquad (25)$$
$$X_{B1} \ \geq \textstyle\sum_s b(s)n_1(s) \qquad \perp \qquad X_{B1} \qquad (26)$$
$$X_{B2} \ \geq \textstyle\sum_s b(s)\big(n_2(s) + n_F(s)\big) \qquad \perp \qquad X_{B2} \qquad (27)$$

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

$$\lambda_{A1} \ \geq \Lambda_{A1} - \sigma_A \mathrm{X}_{A1} \qquad \perp \qquad \lambda_{A1} \qquad (28)$$
$$\lambda_{A2} \ \geq \Lambda_{A2} - \sigma_A \mathrm{X}_{A2} \qquad \perp \qquad \lambda_{A2} \qquad (29)$$
$$\lambda_{B1} \ \geq \Lambda_{B1} - \sigma_B \mathrm{X}_{B1} \qquad \perp \qquad \lambda_{B1} \qquad (30)$$
$$\lambda_{B2} \ \geq \Lambda_{B2} - \sigma_B \mathrm{X}_{B2} \qquad \perp \qquad \lambda_{B2} \qquad (31)$$

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 region 1 and 2's urban labor: $t(w_1 + w_2)/2$ . .[10] Therefore

$$w_1(a(s)\lambda_{A1} + b(s)\lambda_{B1}) \geq p(s) \qquad \perp \qquad n_1(s) \qquad (32)$$

$$w_2(a(s)\lambda_{A2} + b(s)\lambda_{B2}) \geq p(s) \qquad \perp \qquad n_2(s) \qquad (33)$$

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

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

$$Q_d(s) \geq n_1(s) + n_2(s) + n_F(s) \qquad \perp \qquad Q_d(s) \qquad (35)$$

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 "short-hand" scaling parameters for domestic and foreign, that could depend on the relative market sizes for example (see appendix). θ_d and θ_f are the weights on the domestic and foreign varieties in the nest for each sector *s*.

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

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

$$Q_{fd}(s) = \frac{\alpha_d \theta_f \bar{p}^{-\epsilon}}{\theta_d\, p(s)^{1-\epsilon} + \theta_f \bar{p}^{1-\epsilon}} \qquad \perp \qquad Q_{fd}(s) \qquad (37)$$

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 7-11, present simulation results that develop economic implications of the model for the symmetric Ricardian case. Spillovers and asymmetric cases are found in an appendix. Figure 7 shows

[10] Note that this assumption makes (32)-(34) homogeneous of degree 0 in wages and prices.

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 1 earlier in the paper. 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.

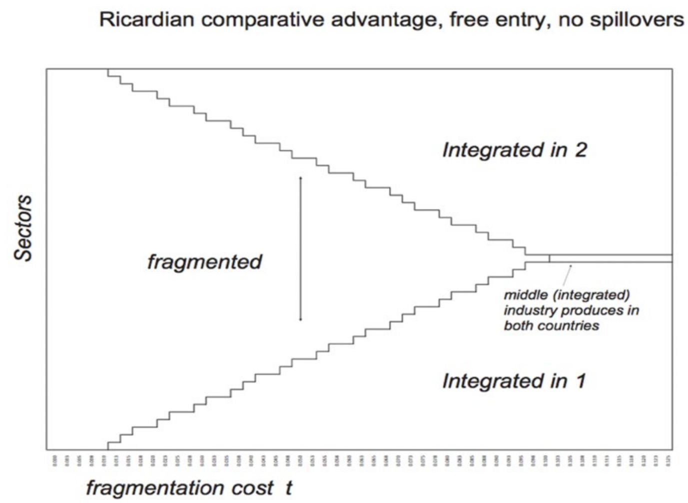
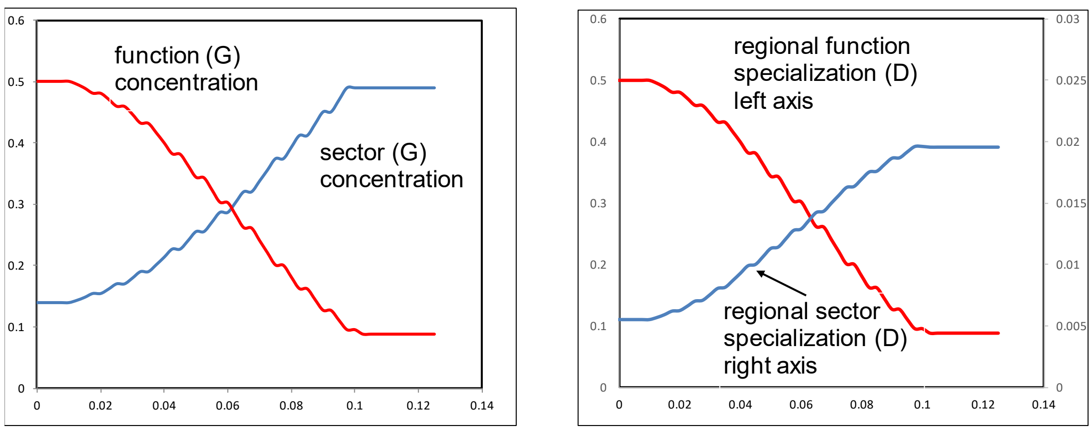


Figure 7: Symmetric Ricardian Case (fragmentation cost t on horizontal axes)

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 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 labor is reallocated to the two regions which due to 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.



Symmetric Ricardian Case   fragmentation cost t on horizontal axes

Figure 8: Concentration (D) and specialization (G) indices

The right panel of Figure 9 illustrates an effect which 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, than 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.

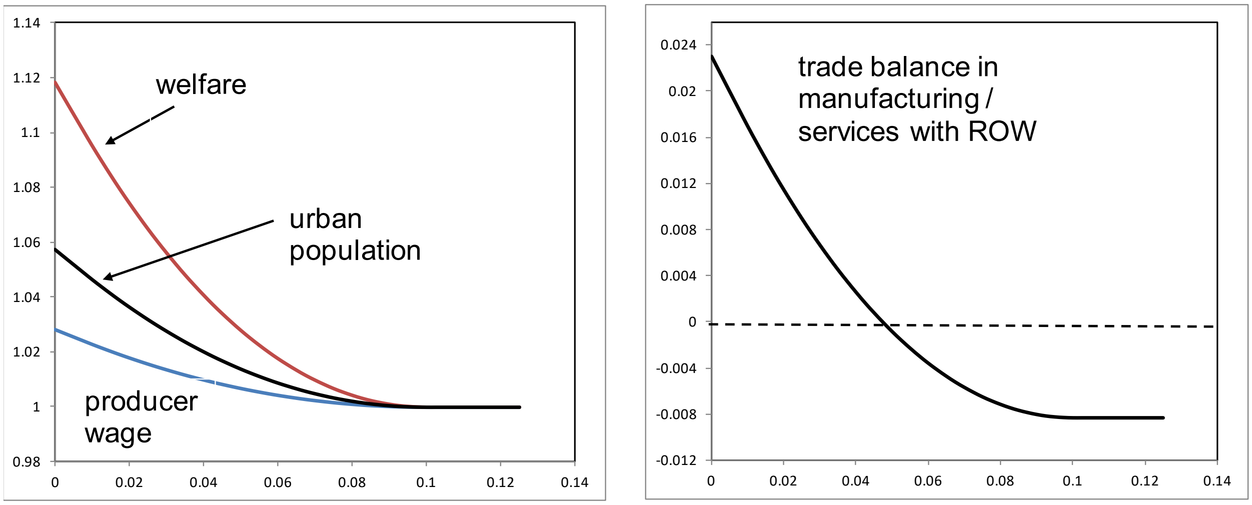


Figure 9 : Welfare, wages and trade balance

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 inter-sectoral 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.

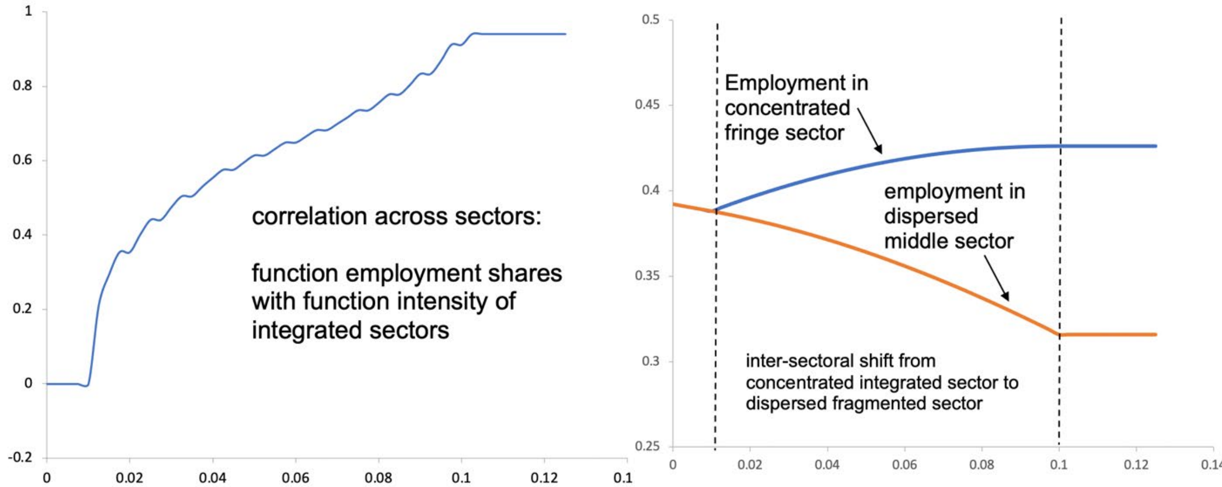
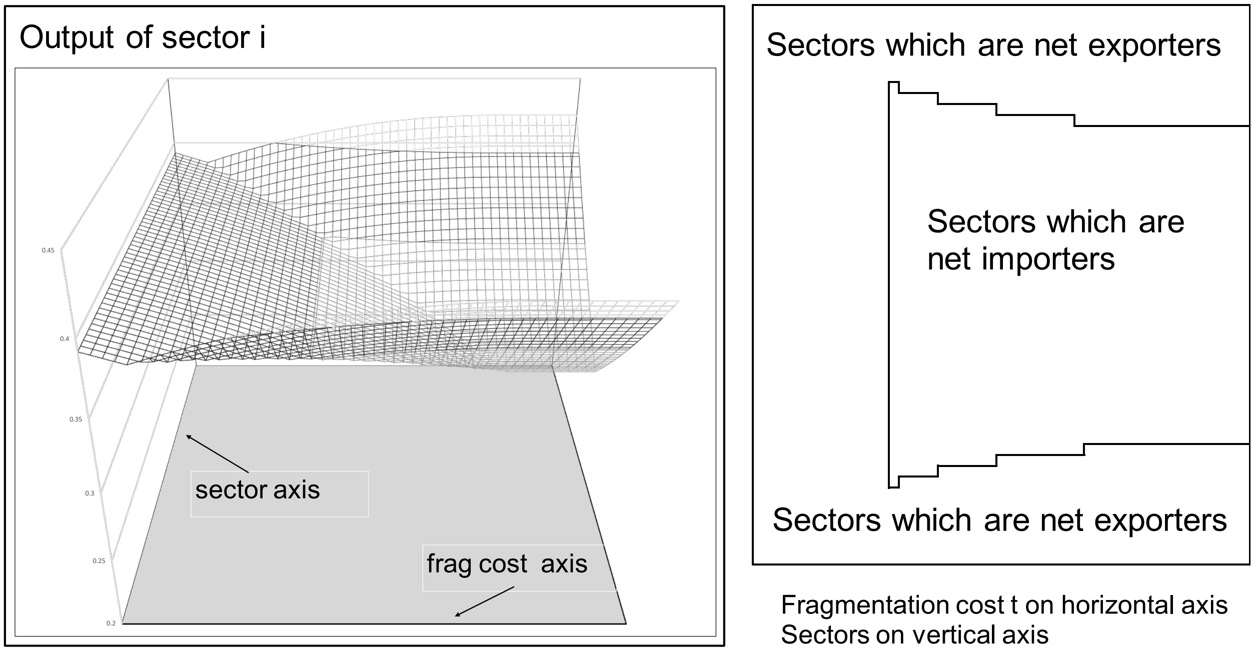


Figure 10: Intra and inter-sectoral shifts in technique and employment

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 middles sectors leads them to expand output and draw labor 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 inter-sectoral 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.[11]

Our final Figure 11 continue to illustrate the heterogeneous effects across sectors, with much the same intuition as the right-hand panel of Figure 10. The left-hand panel shows the effect of falling fragmentation costs on sector outputs. At high *t*, the smallest sectors are the middle sectors, due to their high costs. As *t* falls, the 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.

[11] The left-hand panel of Figure 10 is the total effect, the left-hand side of equation (8). We do not emphasize the inter-sectoral 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.



Fragmentation cost t on horizontal axis
Sectors on vertical axis

Fragmentation cost t on horizontal axis
Sectors on depth axis

Figure 11: Output and net sector exports

The right-and panel of Figure 11 presents the qualitative result on net trade balance in the *s* sectors with the rest-of-world (this figure depends on parameters values chose of course). At high *t*, the costly middle sectors are net importers, 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 exports at a sufficiently low value of t and the country is fully specialized in exporting the s goods and importing the hinterland good.

6. Conclusions

Our paper is motivated by what are widely seen as changes in scope of activities and occupations performed in our 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 thirty-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, result show that a concentration index for occupations and a regional specialization index for occupations have both rise over a twenty-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 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 both 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 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.

References

Aiginger, Karl and Esteban Rossi-Hansberg (2006), “Specialization and concentration: a note on theory and evidence” Empirica, 2006, vol. 33, issue 4, 255-266

Antràs, Pol and David Chor (2021), “Global Value Chains”, NBER working paper 28549, March 2021.

Audretsch, David B. and Maryann P. Feldman (1996), “R&D Spillovers and the Geography of Innovation and Production”, *American Economic Review* 86, 630-640.

Autor, David H. (2013), “The Task Approach to Labor Markets; an Overview”, *Journal for Labour Market Research* 46, 185-199, NBER working paper 18711.

Barbour, Elisa and Ann Markusen (2007), “Regional Occupational and Industrial Structure: Does one Imply the Other?”, *International Regional Science Review* 30, 92-90.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud (2014), “Productive Cities: Sorting, selection, and agglomeration”, *Journal of Political Economy* 122, 507-553.

Berry, Christopher R., and Edward L. Glaeser. "The divergence of human capital levels across cities." Papers in regional science 84, no. 3 (2005): 407-444.

Brakman, Steven, and Charles van Marrewijk (2013), “Lumpy Countries, Urbanization and Trade”, *Journal of International Economics*, 89, 252-261.

Charnoz, Pauline, Claire Lelarge and Corentin 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, 949-994.

Courant, Paul N. and Alan Deardorff (1992). ‘International Trade with Lumpy Countries’. *Journal of Political Economy* 100, 198–210

Davis, Donald R., and Johnathan Dingel (2018), “The Comparative Advantage of Cities”, *Journal of International Economics* 123, Article 103291.

Diamond, Rebecca. "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000." American Economic Review 106, no. 3 (2016): 479-524.

Dumais, Guy, Glenn Ellison, and Edward L. Glaeser (2002), “Geographic concentration as a dynamic process.” *Review of Economics and Statistics,* 84(2), 193-204.

Duranton, Giles, J. Vernon Henderson and William Strange (eds.) (2015), *Handbook of Regional and Urban Economics,* volume 4, Amsterdam North Holland.

Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4), 1077-1106.

Duranton, Giles, and Diego Puga (2005), “From sectoral to functional urban specialization”, *Journal of Urban Economics*, 57, 343-370.

Eckert, Fabian, Sharat Ganapati and Conor Walsh (2020), “Skilled Scalable Services: The New Urban Bias in Economic Growth”, CESifo Working Paper No 8705.

Eckert, Fabian. "Growing apart: Tradable services and the fragmentation of the US economy." mimeograph, Yale University (2019).

Eckert, Fabian, Teresa C. Fort, Peter K. Schott, and Natalie J. Yang (2021) *Imputing Missing Values in the US Census Bureau's County Business Patterns*. NBER working paper 28549.

Ellison, Glenn and Glaeser, Edward L. (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", *Journal of Political Economy* 105, 889-927.

Feenstra, Robert C, and Gordon H Hanson (1996), "Globalization, Outsourcing, and Wage Inequality", *American Economic Review* 86, 240-245.

Fujita, Masahisa, Paul Krugman and Anthony. J. Venables (1999), *The Spatial Economy: Cities, Regions and International Trade*, MIT press.

Gabe, Todd M. and Jaison R. Able (2012), "Specialized Knowledge and the Geography Concentration of Occupations", *Journal of Economic Geography* 12, 435-453.

Grossman, Gene. M. and Esteban Rossi-Hansberg (2008), "Trading Tasks: A Simple Theory of Off-shoring", *American Economic Review*, 98, 1978-1997.

Grossman, Gene M. and Esteban Rossi-Hansberg (2012), "Task Trade between Similar Countries", *Econometrica,* 80, 593-629.

Henderson, Vernon, and Jacques-Francois Thisse (eds.) (2004), *Handbook of Regional and Urban Economics,* volume 4, Amsterdam North Holland.

Krugman PR (1991a), *Geography and trade*. Leuven University Press, Leuven.

Krugman, Paul (1991b), "Increasing Returns and Economic Geography" *Journal of Political Economy* 99, 483-99.

Markusen, James R. (1989), "Trade in Producer Services and in Other Specialized Intermediate Inputs", *American Economic Review* 79, 85-95.

Markusen, James R. (2002), *Multinational Firms and the Theory of International Trade*, Cambridge: MIT Press.

Michaels, Guy, Ferdinand Rauch and Stephen Redding (2019), "Task Specialization in U.S. Cities from 1880-2000", *Journal of the European Economic Association* 17, 754-798.

Moretti, Enrico (2012), *The New Geography of Jobs*, Mariner Press.

Moretti, Enrico. "Real wage inequality." American Economic Journal: Applied Economics 5, no. 1 (2013): 65-103.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte and Raymond Owens III (2009), "Firm Fragmentation and Urban Patterns", *International Economic Review*, 50, 143-186.

Timmer, Marcel P, Sebastien Miroudot and Gaaitzen J. de Vries, (2019), "Functional Specialisation and Trade", *Journal of Economic Geography*, 19, 1-30.

Appendix 1: Data

To construct the indices of concentration and specialization, we need information on the geographic distribution of sectoral and functional economic activity, measured throughout by employment. The two sources from which we derive information are the BLS's Quarterly Census of Employment and Wage dataset (QCEW) and Occupational Employment Statistics (OES). We discuss each in turn.

The QCEW program publishes a quarterly count of employment reported by employers covering more than 95 percent of U.S. jobs available at the county, Metropolitan Statistical Area (MSA), state and national levels by detailed industry.[12] For the analysis, we use employment by six-digit North-American Industrial Classification System (NAICS) industries for each US state for the period 1990-2019. We supplement this data with sector-level information on employment by firm size class, also from the QCEW, to compute the Herfindahl index, H_s, defined in (3).

Using states as our unit of geography has three advantages. First, our results are comparable to previous studies on industry concentration such as Ellison, and Glaeser (1997, henceforth EG97) and Dumais, Ellison, and Glaeser (2002). Second, using states ensures a consistent geography over time. The delineations (i.e., the list of geographic components at a particular point in time) of states remains constant over our sample period. By contrast, between censuses, the delineations for MSAs are revised to reflect Census Bureau population estimates (even the number of counties changes over time).[13] Third, using states increases the reliability of our estimates. In accordance with the BLS Confidentiality policy, data reported under a promise of confidentiality are published in a way so as to protect the identifiable information of respondents. Obviously, the share of observations suppressed is inversely related to the size of regions. We note that totals at the industry level for the states and the nation include the undisclosed data suppressed within the detailed tables without revealing those data. In some case, missing or undisclosed values (at the states-level) create significant gaps in otherwise continuous levels of employment. We fill in the gaps in the data using linear interpolation. About 15 percent of the observations in our sample are imputed using this procedure.

A difficulty we face in developing our data is the frequent reclassification of sectors and functions over time. To minimize the impact of industry reclassification on our results, we restrict our attention to years 1990 to 2019. Information for years prior to 1990 is available only on a Standard Industrial Classification (SIC) basis. Over the period covered by our sample, the NAICS classification

[12] Additional information on the QCEW is available online at https://www.bls.gov/cew/overview.htm.

[13] In a recent paper, Eckert et al. (2021) describe a method to impute missing employment to counties in the County Business Patterns. They provide a very detailed description of the types of issues researchers face when trying to construct longitudinal dataset. In particular, their analysis brings to light the fact that undisclosed information along with changes in geographic units and industrial classification present almost unsurmountable obstacles to the creation of long panels at detailed levels of geography.

introduced in 1997 is revised multiple times, first in 2002, and subsequently in 2007, 2012, and 2017.[14] We limit the sample to industries that we can track accurately across changes in classification. This reduces the size of the sample but ensures that our results are not driven by changes in the scope of our sample or changes in sector definitions.[15] We also remove industries in the "Farming" (NAICS 11), "Mining, Quarrying, and Oil and Gas Extraction" (NAICS 21), Utilities (NAICS 22), and "Public Administration" (NAICS92) sectors because the mapping from sectors to functions is too direct (i.e., "miners" work in "mining") so that the distinction between functional and sectoral specialisation is hard to establish.

The OES program is the only comprehensive source of regularly produced occupational employment and wage rate information for the U.S. economy.[16] It produces employment estimates annually for over 800 occupations. These estimates are available for the nation as a whole and for individual States; national occupational estimates for specific industries are also available. From the OES, we derive function-by-state data, specifically employment by six-digit Standard Occupational Classification (SOC) occupations by US states for the period 2000-2019. We also draw on national function-by-sector data from the OES to construct or to compute the Herfindahl index, H_f, defined in (16).

As was the case with the QCEW, we face data limitations. Beginning in year 2000, the OES survey began using the Office of Management and Budget (OMB) Standard Occupational Classification (SOC) system, which was revised in 2010 and in 2018. To limit the impact of reclassification, we exclude years prior to 2000.[17] For the analysis, we construct a longitudinal region-function datasets restricted to functions that we can defined consistency across changes in classification. We remove "Farming, Fishing, and Forestry Occupations" and occupations that contain the word "other" in their title. Finally, we fill in gaps in the data using interpolation. About 11 percent of the data in our sample is imputed.

---

[14] For years 1990 to 1996, the QCEW is available on a NAICS basis even if the NAICS was introduced only in 1997.

[15] We remove the "Other Services" (NAICS 81) sectors and industries that contain the word "other" in their title, because by their nature these categories are likely to vary from year to year.

[16] Additional information on the OES can be found online at https://www.bls.gov/oes/oes_emp.htm.

[17] Before 1997, data is available only at the national level. For years 1997, 1998 and 1999 the information on employment was collected under a OES proprietary occupational classification system.

Appendix 2: Additional empirical results

To get a sense of which component of the weighted mean drives the time series changes, Table A2.1 reports the simple means of the EG97 index, EG, the raw geographic concentration, G, and the correction factor, H. As seen in the table, the simple average decreases by about 14% over the period. The time series changes in raw concentration closely mimic those of the EG97 index. This happens because changes in the plant-level Herfindahl are an order of magnitude smaller compared to the raw geographic concentration index. Comparing the simple and the weighted mean reveals that large sectors tend to be more dispersed on average compared to smaller ones. The simple mean suggests that the average sector is geographically concentrated ($EG > 0.05$), whereas the weighted mean suggests that the average employee works in a geographically dispersed industry ($EG < 0.02$)

Table A2.1 Mean levels of sectoral concentration for selected years

	1990	1995	2000	2005	2010	2015	2019
Employment-year weighted mean (EG)	0.027	0.022	0.020	0.017	0.015	0.014	0.015
Employment weighted mean (EG)	0.023	0.020	0.018	0.017	0.016	0.016	0.016
Simple mean (EG)	0.058	0.054	0.052	0.047	0.046	0.049	0.051
Raw concentration (G)	0.063	0.058	0.056	0.053	0.053	0.055	0.057
Plant Herfindahl (H)	0.008	0.008	0.008	0.009	0.010	0.010	0.009

Notes: The table reports means (across 626 U.S. six-digit NAICS industries) of the Ellison-Glaeser (1997) index of geographic concentration and of two components, the raw geographic concentration and the Herfindahl measure of plant-level concentration.

Overall, changes in the weighted averages are useful indicators of the time series behavior of geographic concentration. However, to provide a more formal assessment of the time series trend in geographic concentration, we estimate regressions of the sectoral indices on a time trend controlling for sector-level factors using fixed effects

$$\ln EG_{s\tau} = \beta_s + \beta\, Trend_\tau + \varepsilon_{s\tau}\,. \tag{A2.1}$$

If concentration is declining over time the estimated time trend, β, will be negative.

The results from estimating equation (A2.1) by OLS are reported in Table A2.2. The first row reports the results for the full sample of 626 six-digit NAICS sectors. The point estimate is negative and statistically significant and suggests that the within-sector geographic concentration of employment is declining over time. To evaluate if the results are driven by a specific set of sectors, we estimate equation (A2.1) separately for each broad group: manufacturing, business services, personal services, and wholesale, retail and transportation. As reported in Table A2.2, every point estimate is negative and statistically significant.

Table A2.2 Time series trend of sectoral concentration

	Estimates	*Std. Err.*	R^2	*Nb. Obs.*
Full sample	-0.00028	0.00007	0.877	18,780
Manufacturing	-0.00029	0.00005	0.899	7,710
Business services	-0.00019	0.00007	0.745	4,320
Personal services	-0.00013	0.00005	0.849	3,810
Wholesale, retail and transportation	-0.00060	0.00010	0.856	2,910

Notes: This table reports OLS results from regressing indexes of concentration on a time trend. Standard errors are robust (and clustered by major sector for full sample). Every estimated coefficient is significant at the 1 percent level.

We produced similar results for occupation concentration. As seen in the table A2.3, the Herfindahl correction factor has little impact on the index because of its small magnitude, such that most of the changes in concentration over time is explained by the raw concentration index G_f, defined in equation (4). Comparing the simple and the weighted means reveals that occupations that represent a large shares of employment tend to be more dispersed on average compared to occupations that accounts for small shares.

Table A2.3 Mean levels of functional concentration for selected years

	2000	2005	2010	2015	2019
Employment-year weighted mean (EG)	0.0053	0.0050	0.0049	0.0054	0.0063
Employment weighted mean (EG)	0.0052	0.0052	0.0050	0.0055	0.0065
Simple mean (EG)	0.0223	0.0181	0.0181	0.0204	0.0246
Raw concentration (G)	0.0214	0.0174	0.0174	0.0199	0.0237
Plant Herfindahl (H)	0.0000	0.0001	0.0000	0.0004	0.0001

Notes: The table reports means (across 704 six-digit OCC occupations) of the Ellison-Glaeser (1997) index of geographic concentration and of two components, the raw geographic concentration and the Herfindahl measure of plant-level concentration.

As we did for the concentration of sectoral employment, we estimate OLS regressions of the form

$$\ln EG_{f\tau} = \beta_f + \beta\, Trend_\tau + \varepsilon_{f\tau} \tag{A2.2}$$

to estimate the time trend of geographic concentration. The results are reported in Table A2.4 for the full sample and by broad function categories defined in the OCC. As seen in the first row of the table, the time trend is positive and statistically significant in the full sample. This is not surprising given that the estimated beta is the slope of the fitted value through the solid line in Figure 2. The remaining rows of Table A2.4 show that 17 out of 21 estimated time trends are positive and 13 of those are statistically significant at conventional levels.

Table A2.4 Time series trend of functional concentration

	Estimates	*Std. Err.*	R^2	*Nb. Obs.*
Full sample	0.00019***	0.00007	0.870	14,080
Architecture and engineering	0.00018*	0.00010	0.934	640
Arts, design, entertainment, sports	0.00038**	0.00012	0.860	700
Building and ground cleaning and maintenance	-0.00011	0.00007	0.727	160
Business and financial	0.00010**	0.00004	0.903	540
Community and social services	0.00015	0.00019	0.683	260
Computer and mathematical	0.00038**	0.00016	0.776	260
Construction and Extraction Occupations	0.00013	0.00012	0.890	1,100
Education, trainning, and library	0.00015***	0.00005	0.537	1,120
Food preparation and serving	0.00085***	0.00024	0.916	300
Healthcare practitioner and technicians	-0.00024***	0.00004	0.570	920
Healthcare support	0.00039***	0.00010	0.613	280
Installation, Maintenance, and Repair	-0.00020*	0.00010	0.782	980
Legal	0.00011	0.00009	0.324	160
Life, physical, and social science	0.00001	0.00013	0.811	760
Management	0.00015***	0.00005	0.845	600
Office and Administrative Support	0.00014***	0.00004	0.846	1,040
Personal care and service	0.00017	0.00012	0.831	600
Production	0.00050***	0.00006	0.859	2,000
Protective services	-0.00020	0.00014	0.785	380
Sales	0.00049*	0.00027	0.594	400
Transportation and Material Moving	0.00044***	0.00011	0.886	880

Notes: This table reports OLS results from regressing indexes of concentration on a time trend. Standard errors are robust (and clustered by major sector for the full sample). The *, **, and *** indicate statistical significance at the 10, 5, and 1 percent level, respectively.

Next, we evaluate the average time series changes in regional specialization using OLS regressions of the form

$$\ln D_{r\tau} = \beta_r + \beta \, Trend_\tau + \varepsilon_{r\tau}, \tag{A2.3}$$

where β_r represent region-level fixed effects. The results are reported in Table A2.5. As seen in the first row of the table, the time trend β is negative and statistically significant for the sectoral specialization, and positive and statistically significant for the functional specialization.

Table A2.5. Time series trend of regional specialization

	Estimates	*Std. Err.*	R^2	*Nb. Obs.*
Sectoral employment	-0.0203	0.0022	0.9342	1,500
Functional employment	0.0193	0.0028	0.8834	1,000

Notes: This table reports OLS results from regressing our measures of distance on a time trend. Robust standard errors are clustered by state. Estimated coefficients are significant at the 1 percent level.

Appendix 3: Theory

3.1 Analytical expression for employment

Table A3.1: 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 A3.2: 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*	
	$L_1 = L_{A1} + L_{B1} = \int_0^1 L_1(s)ds$	$L_2 = L_{A2} + L_{B2} = \int_0^1 L_2(s)ds$

3.2 Localisation economies : analysis

Using equation (16a) in (17a) gives the unit profit advantage from integration,

$$\Pi(s_1,t) \equiv \pi_1(s_1) - \pi_F(s_1) = t - [1-\gamma(1-2s_1)]\left(\Delta\Lambda - \sigma n\left\{-\frac{1}{2} + s_1[1-\gamma(1-s_1)]\right\}\right)w/2, \quad \text{(A3.1)}$$

There exists an integrated equilibrium if $t \geq t^{**}$, where t^{**} is the minimum value at which $\Pi(s_1 = 1/2, t) \geq 0$, and its value is (from inspection of A3.1), $t^{**} = [\Delta\Lambda + n\sigma\gamma/4]w/2$.

The function $\Pi(s_1,t)$ is cubic in s_1, and is illustrated in figure A1 over the interval $s_1 \in [0, 0.5]$, for three different values of *t*, higher values of *t* shifting the curve upwards. At the lowest value of *t* illustrated, integration is profitable for sector 1 at $s_1 \leq 0.22$. The middle curve is drawn for value t^{**}, i.e. is the value of *t* at which $\Pi(s_1 = 1/2\,, t^{**}) = 0$. There is an interval of values somewhat greater than t^{**} at which there are two values of s_1 at which $\Pi(s_1,t) = 0$, the lower one of which is stable, the upper unstable. The highest curve is the greatest value of *t* at which there is a fragmented equilibrium, this occurring at values $\{\tilde{s}_1, \tilde{t}, \}$. It is possible to derive the values $\{\tilde{s}_1, \tilde{t}\}$ from the pair of equations $\partial\Pi(\tilde{s}_1,\tilde{t})/\partial s_1 = 0,\ \Pi(\tilde{s}_1,\tilde{t}) = 0$. If $\Delta\Lambda = 0$, the value is, $\tilde{t} = n\sigma(1+\gamma^2)^{3/2}3^{1/2}w/(36\gamma)$. There is a positive interval $(t^{**}, \tilde{t})$ in which there are multiple equilibria if spillovers $n\sigma$ are large relative to Ricardian productivity difference, $\Delta\Lambda$.

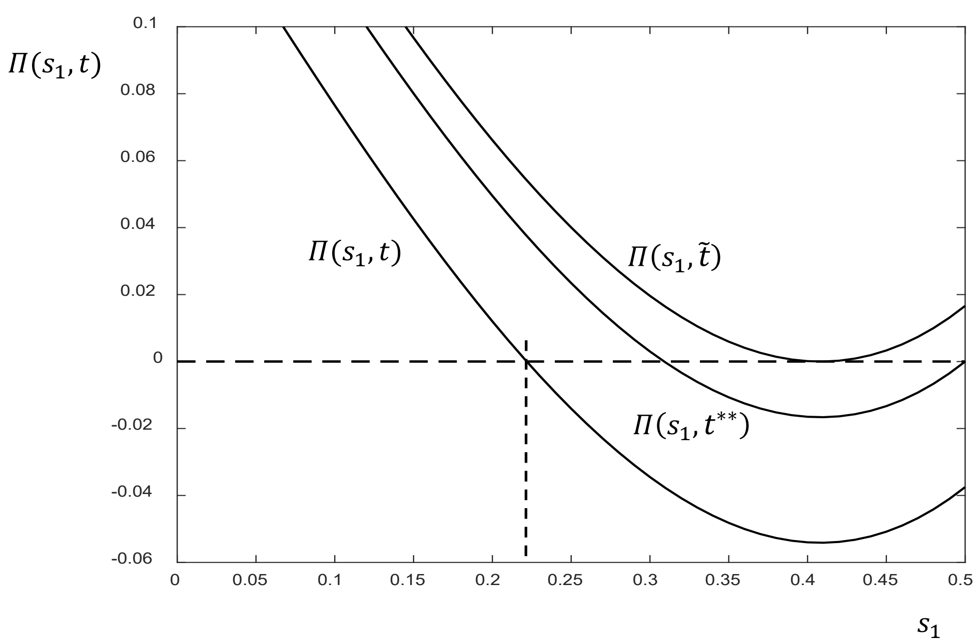


Figure A3.1: Expression A3.1 for different values

Appendix 4: Specification of utility and income

The specification of utility (welfare) is quite standard for trade models. The Q goods are a two-level CES nest. Domestic and foreign varieties for any z sector have an elasticity of substitution of $\varepsilon > 1$ whereas goods from different s sectors are Cobb-Douglas substitutes. R is the outside good, giving a standard quasi-linear utility function

$$U = \beta \ln \left\{ \prod_s \left[\theta_d \left(\frac{Q_{dd}(s)}{\theta_d} \right)^{\frac{\varepsilon-1}{\varepsilon}} + \theta_f \left(\frac{Q_{fd}(s)}{\theta_f} \right)^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}} \right\} + R \tag{A4.1}$$

where β is a scaling parameter. Income (Y) is given the sum of wages (net of commuting costs and rents = w_0) for all urban and outside workers ($\bar{L}$) plus land rents H_1 and H_2 from (19).

$$Y = w_0 \bar{L} + H_1 + H_2 \tag{A4.2}$$

The domestic economy's budget constraint is that Y is spent on R (used as numeraire) plus domestic and foreign urban goods.

$$Y = R + \sum_s p(s) Q_{dd}(s) + \sum_s \bar{p} Q_{fd}(s) \tag{A4.3}$$

(A4.3) can be substituted into (A4.1) to replace R.

$$U = \beta \ln \left\{ \prod_s \left[\theta_d \left(\frac{Q_{dd}(s)}{\theta_d} \right)^{\frac{\varepsilon-1}{\varepsilon}} + \theta_f \left(\frac{Q_{fd}(s)}{\theta_f} \right)^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}} \right\} + Y - \sum_s p(s) Q_{dd}(s) - \sum_s \bar{p} Q_{fd}(s) \tag{A4.4}$$

Maximization of (A4.4) with respect to the Q's (and equivalently for foreign) yields the demand functions in the body of the paper, which do not depend directly on Y as is the usual result in quasi-linear preferences. Domestic demand for domestic good s for example is:

$$Q_{dd}(s) = \alpha_d \theta_f p(s)^{-\epsilon} / \{\theta_d \, p(s)^{1-\epsilon} + \theta_f \bar{p}^{1-\epsilon}\} \tag{A4.5}$$

where α_d is a scaling parameter that is increasing in β (β_d which could differ from the foreign β_f). Suppose $\theta_d = \theta_f = 0.5$ and all $p(s) = \bar{p} = 1$. Then $\alpha = 2$ in the demand functions implies $\beta = 2^{1/\varepsilon}$ and $Q_{ij} = 1$. Parameters α_d and α_f in the demand functions in section 2 are increasing in the β of the domestic or foreign economy, and increases in the α's or β's can represent increases in or differences in market size.[18] Equations (36) and (37) are derived in a similar fashion.

[18] Our algebra indicates that the relationship between the β in (A4.1) and the α in the demand functions above are related by $\alpha = (\beta/2)^{\frac{\varepsilon}{1+2\varepsilon}}$. Because of the concavity of the log formulation of utility, β must more than double to double market demand (α) at constant prices.

Appendix 5: Asymmetric cases

Asymmetric cases are not just a theoretical curiosity, nor is the possibility of multiple equilibria. Several papers referenced above could be interpreted to suggest (translated into our framework) that some functions such as occupations in business services may be more subject to agglomeration economies that other functions. While agglomeration due to spillovers (as opposed to site-specific resources) is generally not explicitly investigated, evidence in Davis and Dingel (2018) and Eckert et. al. (2020) is important in this regard. Duranton and Puga (2005) assume that headquarter services across industries are subject to agglomeration economies while plants have agglomeration economies at the sector level. Theirs is quite a different model from ours as explained earlier, but it is consistent with an analogous view that spillovers may be more important in some functions than others.

Figures A5.1 and A5.2 consider asymmetry between the sectors/regions in the Ricardian case. Figure A5.21 assumes that region 1 has a comparative advantage in function *A* but an absolute advantage in both *A* and *B*. Region 2 has a comparative advantage in function *B*, but no absolute advantage. For intermediate or high levels of fragmentation costs, the result in Figure A5.1 is that region 1 will have a larger range of integrated industries. The intuition follows from a simple argument by contradiction. Consider high fragmentation costs such that all sectors are integrated. Suppose that the solution was symmetric across regions. Then if sector s = 0.5 is just breaking even in region 2, there would be positive profits for sector s in region 1.

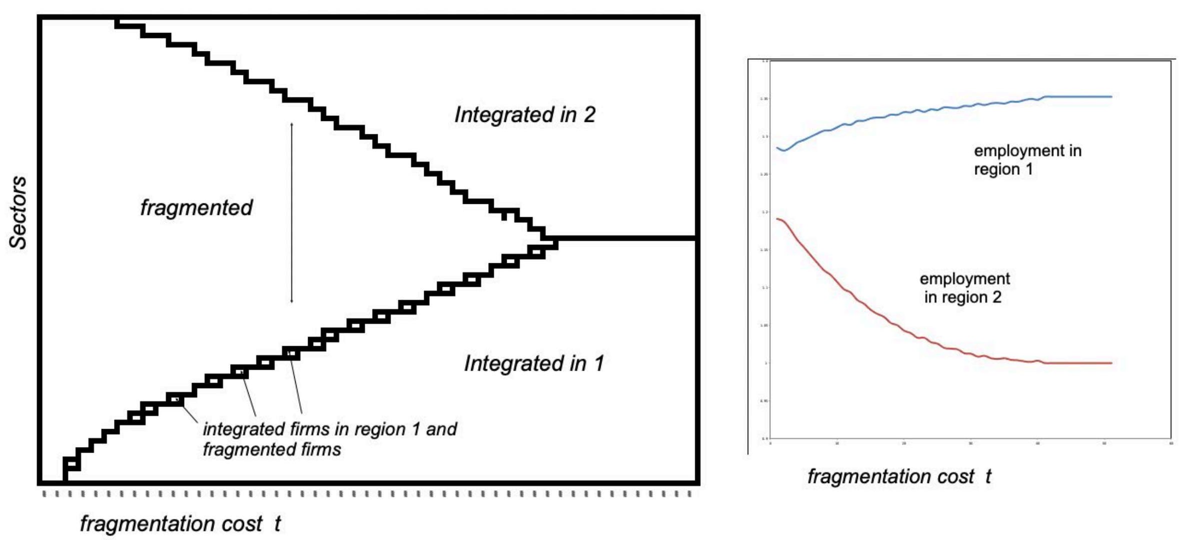


Figure A5.1: Asymmetric Ricardian Case

Region 1: comparative and absolute advantage in function *A*

Two further results follow in the asymmetric Ricardian case. The right-hand panel of Figure A5.1 shows the employment levels in the two regions. Intuitively, the region with the absolute advantage (region 1) will be larger for all levels of fragmentation costs, but this difference shrinks as these costs fall. Figure A5.2 shows the function and sector concentration indices for the same asymmetric Ricardian case. The more productive region 1 will have lower concentration for both sectors and functions. The intuitive follows from the previous paragraph: region 1 will have more integrated industries. But the difference disappears as fragmentation costs go to zero. In our empirical results, we show that larger regions do have lower levels of both forms of concentration.

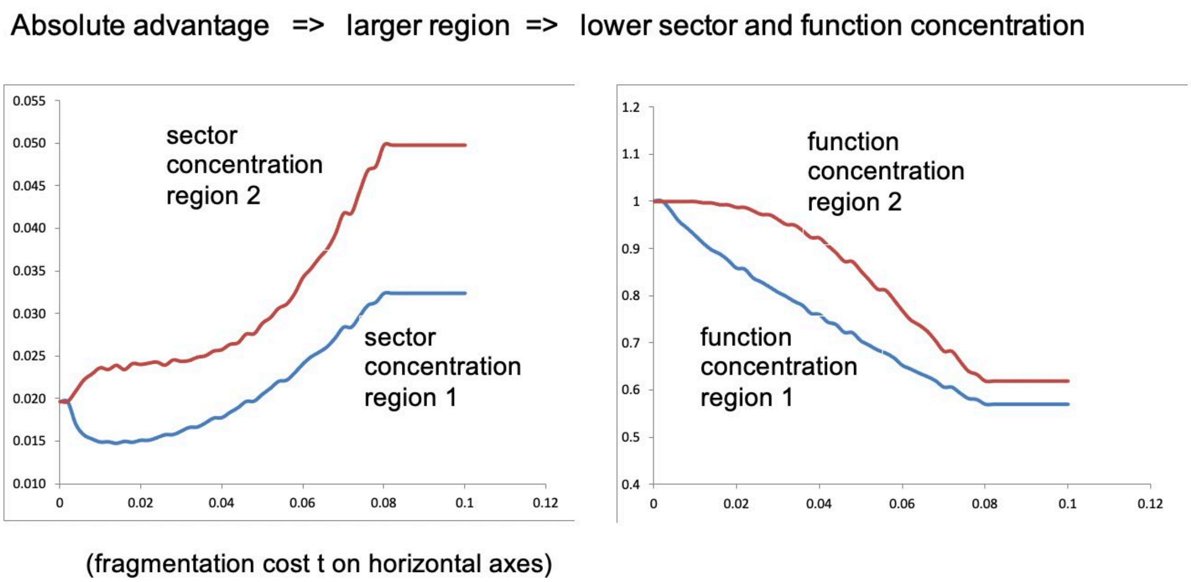


Figure A5.2: Asymmetric Ricardian Case

Region 1: comparative and absolute advantage in function *A*.

Finally, consider an asymmetric spillovers case, motivated by this idea that perhaps business service occupations are more characterized by spillovers than other functions. Figure A5.3 shows a case where only function A has spillovers, but in both regions. In equilibrium however, the spillovers case is similar: region 1 will have a comparative and an endogenous absolute advantage in function *A*, while region 2 has a comparative but not absolute advantage in function *B*.

These results show up as differences in region size/employment (which in turn translate into producer wages), shown in the right-hand panel of Figure A5.3. The region size difference is large when all industries are integrated and small when all are fragmented (though largest in the middle for the spillovers case). Again, the intuition follows from a simple argument by contradiction. If region sizes (employment) were the same, then producer wages would be the same, in which case there must be positive profit opportunities in region 1 and/or losses incurred in region 2.

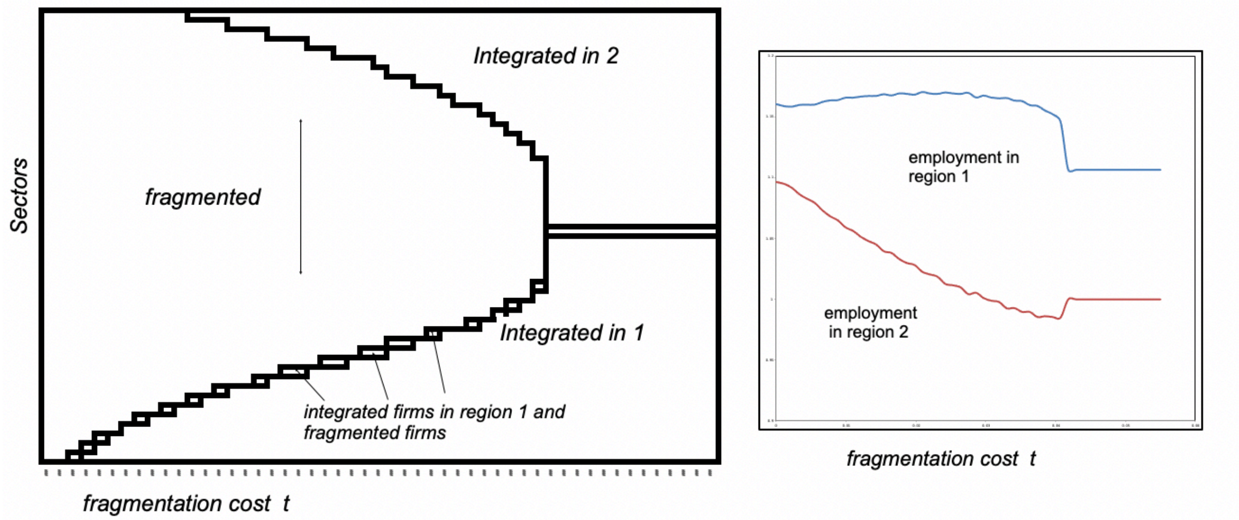


Figure A5.3: Asymmetric Spillovers Case; spillovers in function *A* only

An important point about Figure A5.3 is that it illustrates the possibility that regional fortunes may diverge over some range of falling fragmentation costs. The region with the comparative advantage in the function characterized by spillovers grows and the other region can actually shrink.

The convergence in region sizes as fragmentation costs become small seems to be in large part a terms-of-trade effect: as fragmentation costs fall, the relative prices of goods with low sector indices (located in region 1) fall a lot more in general equilibrium than the prices of the high index goods. An alternative way to think about this is that the high productivity of region 1 workers in the *A* function means that less workers are required to produce those tasks at given output prices and hence region 1's employment falls some in response to that increased productivity.