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

Antoine Gervais

James R. Markusen

Anthony J. Venables

Abstract

We explore the idea that the comparative advantage of regions within a country is shaped by their productivity in supplying 'functions' such as law, finance, advertising and engineering, to multiple sectors. The paper addresses two questions. How do 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? How do changes in the ease with which industries can draw on functions produced in other regions affect these patterns of specialization? We derive theoretical answers to these questions in a model in which region-function specific productivity differentials may be exogenous or driven by agglomeration economies. The model's prediction that falling barriers to inter-regional trade in functions lead to increasing functional specialization and decreasing sectoral specialization is confirmed by empirical study of specialization of US states over a 20-30 year period.

Keywords: regional specialisation, regional trade, firm 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
tony.venables@economics.ox.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 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: rising service sectors and declining manufacturing 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. Moretti (2012) is a broad and sweeping study of these trends and how they are related.

This paper develops a theoretical approach and offers an empirical analysis of these issues, based on the core idea that regions' comparative advantages have evolved from being based on sectors, to being based on productivity differences in functions or occupations. 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.

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 primary factor – labor -- and we refer to them as functions, in the empirical section identifying them with occupations. 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-sectoral, 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. Drawing on the new economic geography literature, however, 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 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 so leading to functional specialization.

A final ingredient in our approach is labor mobility between regions, a typical assumption in the urban/regional literature. However, we draw on the urban-regional literatures to endogenise 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, these 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. The world less resembles the archetype model of international trade theory. 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. With high fragmentation costs, a region has lawyers, accountants, machinists, mechanics and many other occupations working in a small number of comparative-advantage sectors. With lower fragmentation costs, a region has a smaller range of occupations working in a large number of sectors. New York specializes in white-collar functions such as finance and marketing, but these individuals are working for many different sectors.

The final section of the paper is an empirical investigation using US state level data on sectoral and occupational (as a proxy for functional) employment. States are relatively large geographical units for the questions we are addressing, and the data limitations than require us to operate at this level are discussed in section 5. A further limitation is that fragmentation costs are not directly observed, and available proxies (e.g. travel costs for both personnel and physical products, internet applications from email to Skype and Zoom) do not provide either state- or sector-level variation. As a consequence, our empirical analysis instead examines how some of the key relationships obtained from the theory behave over a 20-30 period. Charnoz, Lelarge, and Tevien (2018), and Eckert, Ganapati, and Walsh (2020) present evidence that information, communications and technology costs (ICT) are decreasing over time.

This provide support for our suggestion that our empirical analysis over time is a rough proxy for falling fragmentation costs in the theory section (see also Eckert 2019).

Findings from the empirical work are consistent with predictions from the theory. First, we find declining sectoral concentration and increasing functional (occupational) concentration over time, and a large fraction of those changes is explained by within-sector and within-function changes in geographic concentration. Second, regional specialization indices in functions and sectors have the same properties as the concentration indices. Third finding, larger regions have lower levels of both sectoral and functional specialization.

Ideas in this paper are complementary to the influential paper by Duranton and Puga (D&P, 2005) on sectoral and functional specialisation. 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, head-quarters, 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. Our approach 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 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.

[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.

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), 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), Krugman (1991).

Empirical tools for measuring concentration and specialization are drawn from Krugman (1991), 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 detailed occupation categories, workers are increasingly concentrated. Our 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 develop a partial equilibrium model with two symmetric regions with exogenous Ricardian differences in productivity by function and region. In section 3, we endogenize productivity differences by adding external economies of scale in the form of spillovers. In section 4, we characterise the general equilibrium model and address these questions via simulation analysis. In section 5, we confront the main theoretical predictions with the data using region-level information on production and employment by sector and occupation for US states for the period 1990-2019.

2. Regions, sectors and functions

The ingredients of the model are locations, focussing in the theory on two regions; sectors, which we model as a continuum; two functions that are used as inputs to production each sector; and a single primary factor, labour, which is used to produce functions and is perfectly mobile between regions and

functions.[3] We build the model in stages. In this section and the next we focus on sectors and functions to draw out results on fragmentation and specialisation, whilst keeping the general equilibrium side of the model in the background; there is an outside good that we take as numeraire, and we make sufficient assumptions to ensure that the two regions are symmetric. In section 4, 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 (if any) 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 refer to as the function intensity of the sector.[4] These intensities vary with sector *s* but are the same in both regions; we assume that sectors can be ranked such that low *s* sectors are *A*-intensive and *B*-unintensive, i.e. $a'(s) < 0$ and $b'(s) > 0$.

Producers in each sector can source functions from either region, but if the two functions come from different regions then a per unit fragmentation cost *t* 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' (subscript *F*). The unit profits in sector *s* for each of the three production modes are therefore

$$\begin{aligned} \pi_1(s) &= p(s) - [a(s)\lambda_{A1} + b(s)\lambda_{B1}]w_1, \\ \pi_F(s) &= p(s) - [a(s)\lambda_{A1}w_1 + b(s)\lambda_{B1}w_2] - t, \\ \pi_2(s) &= p(s) - [a(s)\lambda_{A2} + b(s)\lambda_{B2}]w_2. \end{aligned} \tag{1}$$

[3] Thus, engineers can convert to lawyers. Comparative advantage comes from cross-region variation in the productivity of labour in producing functions. It would be possible to add a Heckscher-Ohlin flavour by assuming endowments of engineers and accountants, but this adds little to our basic story.

[4] $a(s)$ 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.

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 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}.

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 (1), these mode-boundaries are implicitly defined by

$$\pi_F(s_1) - \pi_1(s_1) = b(s_1)[\lambda_{B1}w_1 - \lambda_{B2}w_2] - t = 0, \qquad (2)$$

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

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. (1) and are given in appendix Table A.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, $L_r = \sum_f \int_s L_{fr}(s)ds$.

3. 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 (3.1) and then turning to economies of scale (3.2). Full general equilibrium is set out in section 4, while some material on asymmetric cases is found in Appendix 2. The empirical analysis is presented in section 5.

3.1 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. (2), and are implicitly given by

$$b(s_1)w\Delta\lambda = t, \qquad \text{and} \qquad a(s_2)w\Delta\lambda = t. \tag{3}$$

A simple case which we develop in detail takes 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.[5] Appendix Table A.2 gives employment levels by region, function, and sector, replicating Table A.1 with explicit expressions derived from this functional form. The unit profit functions of eqn. (1) become, $\pi_1(s) = p(s) - \{2\lambda + \Delta\lambda[1 - \gamma(1 - 2s)]\}w/2$, $\pi_F(s) = p(s) - \lambda w - t$, 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

$$\pi_1(s_1) = \pi_F(s_1): \quad s_1 = \frac{1}{2}\left[1 - \left(1 - \frac{2t}{w\Delta\lambda}\right)\frac{1}{\gamma}\right], \tag{4}$$

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

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, $\Delta\lambda$.

Integration to fragmentation: If $t = w\Delta\lambda/2$ then $s_1 = s_2 = 1/2$; i.e. half of sectors are integrated in 1, the other half integrated in 2, and no sectors are fragmented. We call this the critical value $t^* = w\Delta\lambda/2$ and note that there is no fragmentation for any values $t \geq t^*$. If $t < t^*$ then fragmented sectors emerge, first in sectors that have similar use of both functions, i.e. *s* in an interval around ½ and of width

[5] Thus, for all $s \in [0,1]$, $a(s), b(s) \geq 0$. The assumption is not necessary for our main results, see e.g. the proof of proposition 1 in appendix A1. Figure 1 has $\gamma = 1$, this being 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 sectors become fragmented at some positive value of *t*; if $\gamma > 1$ then extreme sectors use only one function.

$s_2 - s_1 = (1 - 2t/w\Delta\lambda)/\gamma$, wider the smaller is t, and the larger are productivity differences, $\Delta\lambda$. Intuitively, these are the sectors where both functions have a high share of costs (e.g. close to 50%), 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.

This and equations (4) are illustrated on figure 1, which has sectors on the vertical axis and fragmentation costs, t, on the horizontal. 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. Figure 1 is constructed with $\gamma = 1$ and $\Delta\lambda = 0.4$, and $w = 1$. The critical value t^* is proportional to $w\Delta\lambda$ and, for a given value of $t/w\Delta\lambda$ the range of fragmented sectors is larger the smaller is γ, the parameter that measures the range of function intensities.

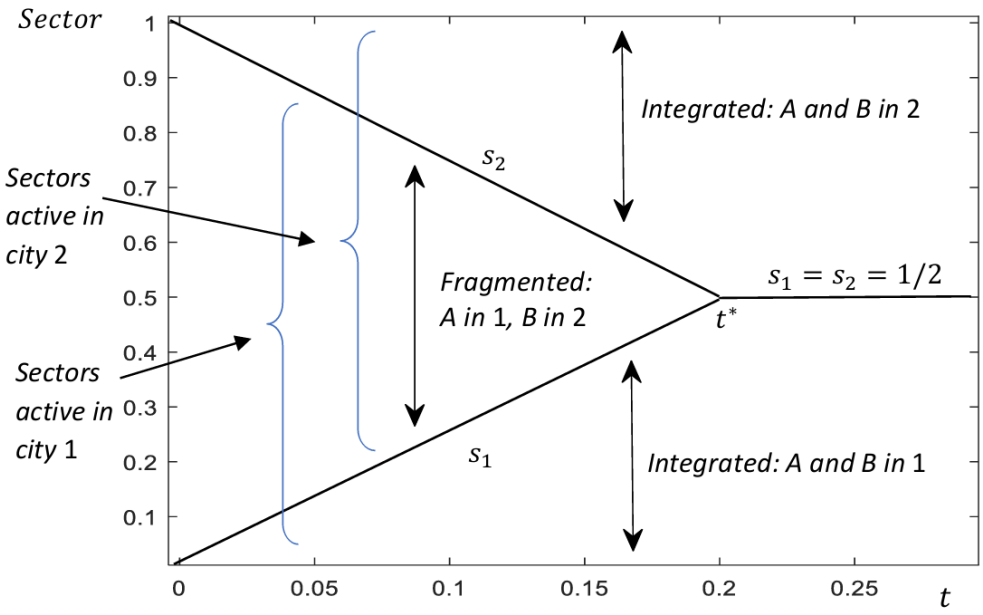


Figure 1: Modes of operation in each sector s.

Sectoral to functional specialisation: The preceding paragraph established where producers in each sector source their input of functions. The dual question is: what activities are present in which regions? The range of sectors with a presence in regions of each type increases as fragmentation costs fall below t^*, as more sectors become fragmented (see the curly brackets in figure 1). This implies a decline in the sectoral specialisation of regions. In the empirical section we will measure this by calculating

specialisation indices defined on the shares of each region in total employment of each sector, i.e. $m_{sr} \equiv L_{sr}/\Sigma_r L_{sr}$. While regions' specialisation in sectors is falling, their specialisation in functions is increasing; intuitively, as fragmentation costs fall so production of each function moves into the region according to comparative advantage. In later sections of the paper we compute functional specialisation indices, based on shares of each region in total employment of each function, $m_{fr} \equiv L_{fr}/\Sigma_r L_{fr}$. Pulling this together, we summarise results in the following proposition:

Proposition 1: In the symmetric model with $\gamma \leq 1$.

i) If fragmentation costs are high, i.e. $t \geq t^* = w\Delta\lambda/2$, then $s_1 = s_2 = 1/2$ and:
   a) Mode: All sectors are integrated.
   b) Sectors: Each region contains activity in half the sectors; each sector operates in a single region (region 1 for $s \leq 1/2$, and region 2 for $s > 1/2$).
   c) Functions: Both functions are present in each region.

ii) If fragmentation costs are low, $t < t^* = w\Delta\lambda/2$, then $s_2 - s_1 = (1 - 2t/w\Delta\lambda)/\gamma > 0$ and:
   a) Mode: Sectors with $s \in [s_2, s_1]$ are fragmented, operating in both regions; sectors with more extreme function intensities ($s < s_2,\ s > s_1$) are integrated, operating in a single region.
   b) Sectors: Each region contains activity in more than half the sectors. If $t \leq (1-\gamma)\, w\Delta\lambda/2$ then each region contains activity from all sectors.
   c) Functions: If $t \leq (1-\gamma)\, w\Delta\lambda/2$ then each region specialises in a single function, $L_{A1} = L_{B2} > 0,\ L_{A2} = L_{B1} = 0$, (complete functional concentration).

The implications of this proposition will be discussed further in section 4.3 where, in the context of the full general equilibrium model, specialisation and concentration indices are calculated for the distribution of both sectoral and functional employment across regions. They are central to the empirical work of section 5.

3.2 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 due to exogenous efficiency differences. We therefore explore an alternative mechanism in which there are agglomeration economies in production of functions, and it is these regionally focussed scale economies that drive the location of functions and thus of sectors.

Labour input coefficients are function and region specific, and are now assumed to be based on 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 λ + Δλ. 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

$$\lambda_{A1} = \lambda - \sigma_A X_{A1}, \qquad \lambda_{A2} = \lambda + \Delta\lambda - \sigma_A X_{A2}, \tag{5}$$

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

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

$$\lambda_{B1} - \lambda_{B2} = \Delta\lambda - \sigma_B n\left\{-\tfrac{1}{2} + s_1[1 - \gamma(1 - s_1)]\right\}, \tag{6a}$$

$$\lambda_{A2} - \lambda_{A1} = \Delta\lambda - \sigma_A n\left\{\tfrac{1}{2} - s_2[1 + \gamma(1 - s_2)]\right\}. \tag{6b}$$

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,

$$\pi_F(s_1) - \pi_1(s_1) = \{[1 - \gamma(1 - 2s_1)](\lambda_{B1} - \lambda_{B2})\}w/2 - t = 0, \tag{7a}$$

$$\pi_F(s_2) - \pi_2(s_2) = \{[1 + \gamma(1 - 2s_2)](\lambda_{A2} - \lambda_{A1})\}w/2 - t = 0. \tag{7b}$$

To analyse these relationships, we focus on (6a) and (7a), the other pair, (6b) and (7b), being symmetric. Substituting (6a) in (7a) gives $\pi_F(s_1) - \pi_1(s_1)$ as a function of s_1. The objective is to find sets of parameters at which different types of equilibria hold.

Notice first that there is full integration if $\pi_F(s_1) \leq \pi_1(s_1)$ at s_1 = ½. Straightforward calculation gives critical value $t^{**} = [\Delta\lambda + n\sigma\gamma/4]w/2$ at which $\pi_F(s_1) = \pi_1(s_1)$ evaluated at s_1 = ½. Evidently, this reduces to the Ricardian case 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. This is illustrated by the solid horizontal line on Figure 2.

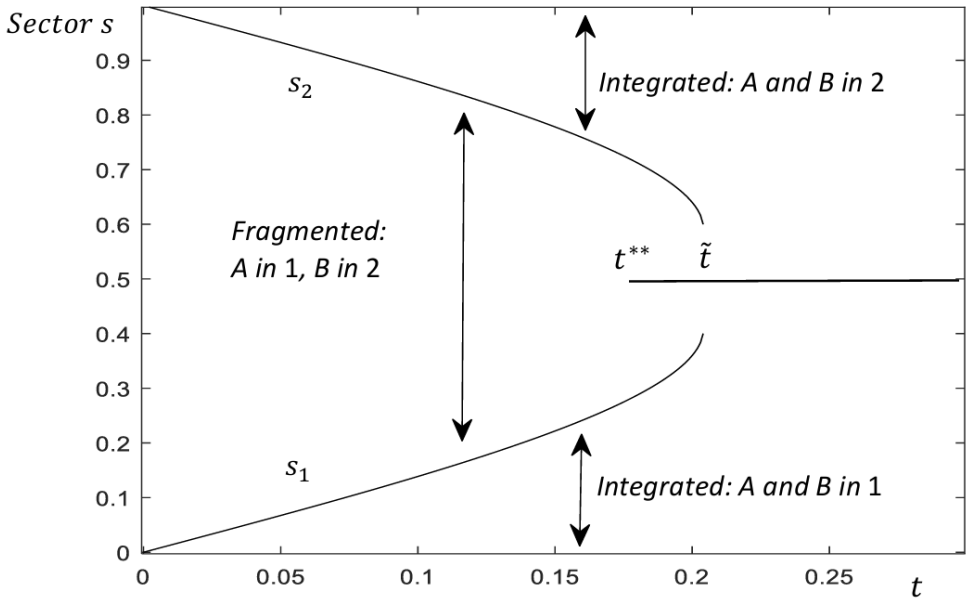


Figure 2: Modes of operation in each sector s: with increasing returns.

Figure 2 differs from Figure 1 in the non-linearity of the mode boundaries and, in particular, the overlap between these lines that occurs in the interval $(t^{**}, \tilde{t})$.[6] This is a region of multiple equilibria. 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.

Formally, this occurs because using (6a) in (7a) generates a cubic equation. Appendix 1 works this through in some detail, deriving the critical value $\tilde{t}$ below which fragmented production is an equilibrium. There is a positive interval $(t^{**}, \tilde{t})$ in which there are multiple equilibria if spillovers σ are large relative to any Ricardian productivity difference, $\Delta\Lambda$. To summarise:

Proposition 2: In the symmetric model with external economies of scale

i) If $t \geq t^{**} = [\Delta\Lambda + n\sigma\gamma/4]w/2$, there is an equilibrium in which all sectors are integrated.

ii) If $t < t^{**}$, there is a unique equilibrium, in which sectors $s \in [s_2, s_1]$ are fragmented.

[6] Figure 2 has the same parameters as Figure 1, except that $\Delta\lambda = 0$ and $\sigma_A = \sigma_B = 1.5$.

iii) There is a range of values of $t \in (t^{**}, \tilde{t})$ at which integration of all sectors and fragmentation of a range of sectors are both equilibria.

iv) Increasing returns ($\sigma > 0$) means that, should fragmentation occur, the range of sectors that are fragmented is wider, at each *t* and for each $\Delta\lambda$, than if $\sigma = 0$.

Parts (i) and (ii) of the proposition mean that the qualitative predictions concerning the effect of reductions in *t* on regions' sectoral diversification and functional specialisation are as in proposition 1; we use these predictions in the empirical section. Parts (iii) and (iv) are a consequence of the externality created by technological spillovers. An important difference is that the localisation economy operates at the functional rather than the sectoral level. Thus, while there are no direct technology spillovers between sectors, expansion in one sector will increase the quantity of functions supplied, this raising productivity in functions and reducing costs for other sectors, particularly those with similar function intensities. Linkages between sectors are created via the medium of localisation economies in functions.

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.

4. General Equilibrium

To this point we have assumed product prices are constant, wages are constant and equal in both regions, and the total output of each sector is fixed and the same in all sectors. We now relax these assumptions and develop the general equilibrium of the model. In section 4.1 we model regions as sets of urban areas, between which labour mobility equalises real, but not necessarily nominal, wages. In section 4.2 we look at product supply and demand, adding zero profit conditions and endogenizing prices. Section 4.3 sets out the full equilibrium structure of the production side as the basis for the empirical analysis of section 5. Section 4.4 considers asymmetric cases.

4.1 Region size, employment and wages

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

Labour is perfectly mobile, equating utilities across regions. To prevent corner solutions – such as all population ending up in one region -- we require some sort of diminishing returns to regional population, and this is achieved by supposing the existence of a fixed factor in each region. We take this

to be the number of urban areas, each of which is described by the standard urban model (the Alonso-Mills-Muth model, see for example Henderson and Thisse 2004). Thus, region *r* contains K_r cities, assumed to be identical. In each of these cities, workers face costs of commuting and land rent, costs 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

$$L_r = K_r(w_r - w_0)/c_r, \quad r = 1, 2. \tag{8}$$

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

$$H_r = c_r L_r^2/2K_r. \tag{9}$$

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 city model $K_r = 1$ or a state level model…

4.2 Sectoral price and output

We have to this point held the price and output of each sector constant. We now endogenize these variables by modelling demand for each sector's output and letting output adjust such that price equals unit cost.

Demands for final output comes from domestic spending and from exports. The domestic country is assumed small as an importer, and so foreign prices in all of the *s* sectors take exogenous value $\bar{p}$, common across all sectors. 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 outside good (numeraire) is additively separable with a constant marginal utility, implying that income does not appear in the demand functions for the *Q* goods (though we will introduce a demand shifter later). With these assumptions, demand for the output of each sector is

$$Q_d(s) = Q_{dd}(s) + Q_{df}(s) = \frac{\alpha\, \theta_d p(s)^{-\epsilon}}{\theta_d\, p(s)^{1-\epsilon} + \theta_f\, \bar{p}^{1-\epsilon}} + \frac{\bar{\alpha}\bar{\theta}_d p(s)^{-\epsilon}}{\bar{\theta}_d\, p(s)^{1-\epsilon} + \bar{\theta}_f \bar{p}^{1-\epsilon}}. \qquad (10)$$

Demand parameters are α, θ, and overbars are used to denote parameters in foreign. The utility functions and budget constraints that support these demand functions are given in appendix 2, and are used in some welfare calculations that follow.

In equilibrium price is less than or equal to unit cost in all sectors and regions (complementary slack inequalities since not all modes will be active in all sectors) so, from equations 1,

$$\begin{aligned} w_1[a(s)\lambda_{A1} + b(s)\lambda_{B1}] \geq p(s) \quad &\perp \quad n_1(s), \\ w_2[a(s)\lambda_{A2} + b(s)\lambda_{B2}] \geq p(s) \quad &\perp \quad n_2(s), \qquad (11) \\ w_1 a(s)\lambda_{A1} + w_2 b(s)\lambda_{B1} + t(w_1 + w_2)/2 \geq p(s) \quad &\perp \quad n_F(s).^{7} \end{aligned}$$

Total domestic supply in each sector *s* is may come from each mode, so

$$Q_d(s) = n_1(s) + n_F(s) + n_2(s). \qquad (12)$$

Prices and output adjust to clear markets, and ss they do so employment levels, wages, and the structure of economic activity in each region are determined. Sectors and modes may be active or non-active in each region, so the equilibrium can be thought of as a non-linear complementarity problem in which corner solutions are a crucial feature of the model. To explore this we use numerical techniques, and the full set of equations and inequalities used simulation are given in appendix 3. To implement this, we discretize the number of sectors: in the simulations to follow model development, there are 51 sectors (i.e., *s* = 1, 2, …, 51. an odd number allows for a middle sector). The total number of weak inequalities and non-negative unknowns is 318 (appendix 3).

4.3 Symmetric Ricardian and localization economies in general equilibrium

Figures 3 and 4, and appendix Figures A2 to A5 present simulation results that develop economic implications of the model. Figure 3 presents the symmetric Ricardian case, 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

[7] Fragmentation costs *t* are in units of labor; we arbitrarily assume that this is divided evenly between labour from each region, giving $t(w_1 + w_2)/2$ as fragmentation cost.

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.

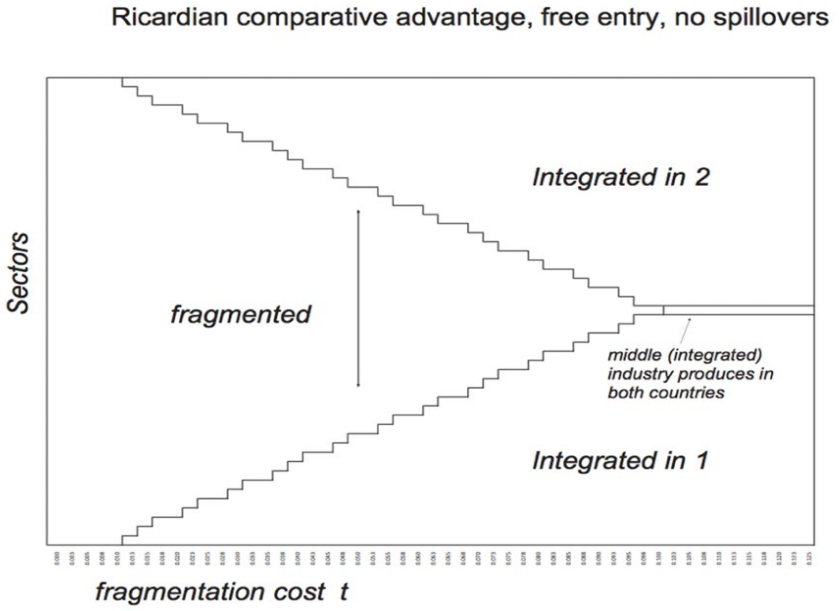


Figure 3: Symmetric Ricardian Case (fragmentation cost t on horizontal axes)

Figure 4 shows further results for the case in Figure 3 in four panels. The upper left panel gives Herfindahl employment concentration indices for sectors and functions across the two regions for each level of fragmentation costs. The concentration of sector s is the sum over regions r of the share of sector s's national employment that is in r minus region r's share of national employment, squared:[8]

$$G_s = \sum_r (m_{sr} - m_r)^2, \qquad m_{sr} = L_{sr} / \sum_r L_{sr} , \qquad m_r = L_r / \sum_r L_r, \tag{13}$$

The concentration of function f employment across regions is similarly defined.

$$G_f = \sum_r \left(m_{fr} - m_r\right)^2 \qquad m_{fr} = L_{fr} / \sum_r L_{fr} , \qquad m_r = L_r / \sum_r L_r. \tag{14}$$

These are then averaged over all sectors s and functions f to get the indices used in the upper left-hand panel of Figure 4. As fragmentation costs fall, the sectoral concentration index falls and the function

[8] Definitions of employment levels L_{sr}, L_{fr}, L_r are given in appendix tables A1 and A2. $L_r = \sum_s L_{sr} = \sum_f L_{fr}$, $L_s = \sum_r L_{sr}$, $L_f = \sum_r L_{fr}$.

concentration index rises. This is a central prediction of the model, which will be examined empirically in section 5 below.

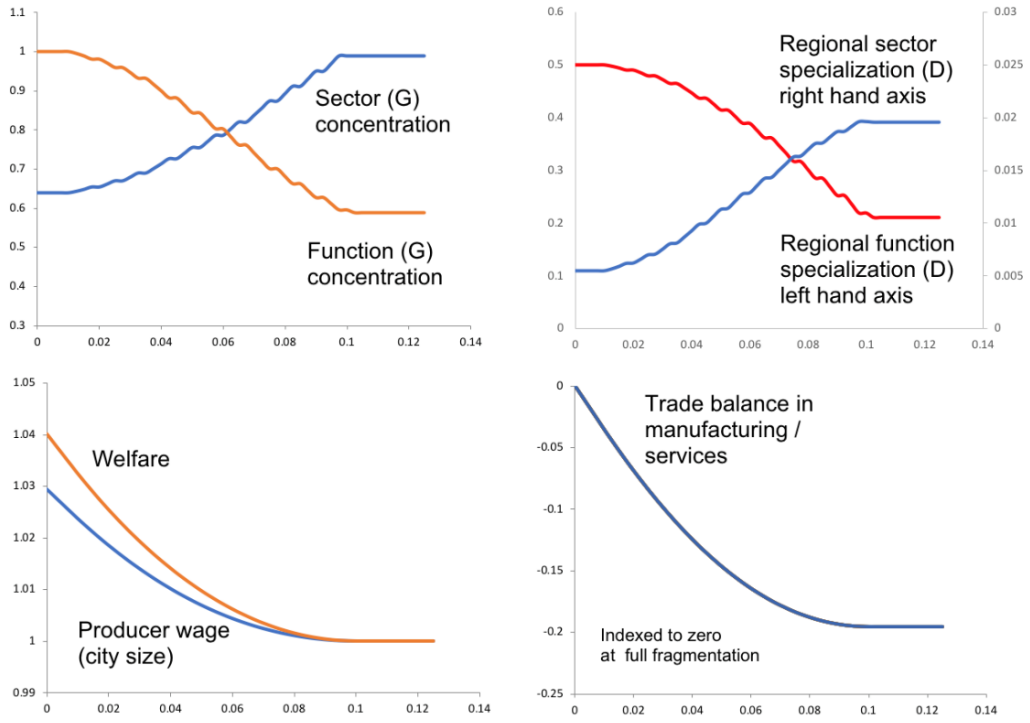


Figure 4: Symmetric Ricardian Case (fragmentation cost *t* on horizontal axes)

In addition to examining sector and function concentration theoretically here and empirically in section 5, we can compute indices of regional specialization. Each region is compared to the national distribution of employment across sectors and functions via specialization indices, $D_r^{sector}, D_r^{function}$. Similar to our measure 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

$$D_r^{sector} = \sum_s (q_{rs} - q_s)^2, \qquad q_{rs} = L_{sr} / \sum_s L_{sr}\,, \qquad q_s = L_s / \sum_s L_s, \tag{13a}$$

$$D_r^{function} = \sum_f \left(q_{rf} - q_f\right)^2, \quad q_{rf} = L_{fr} / \sum_f L_{fr}\,, \qquad q_f = L_f / \sum_f L_f\,. \tag{14a}$$

These are then averaged over all regions to get the indices used in the upper right-hand panel of Figure 4.

The top left panel of Figure 4, giving the sector and function concentration indices and the top right panel giving the regional specialization indices in both sectors and functions, are qualitatively almost identical, though they differ some in scale (note the different scale on the right and left axis in the top left panel). This is largely due to the fact that this example has regions and sectors that are symmetric in size. This choice of example is deliberate, providing an intuitive base case which is examined empirically in section 5. We do not hypothesize that the concentration and specialization indices are qualitatively the same, only that the sector indices both fall with falling fragmentation costs and the function indices both rise with falling *t*. We have done many simulations with various asymmetries between regions and sectors, and these slope relations always hold for both concentration and specialization.

The bottom left panel of Figure 4 graphs the producer wage and welfare (recall all workers earn a wage net of commuting costs and land rent equal to w_0). Note from equation (8) that the producer wage is proportional to urban population or region size. The producer wage / region size curve shown in the bottom left of Figure 4 indicates that lowering fragmentation costs does have a significant effect on region size: increased outputs depress product prices some, and so from the free-entry conditions, producer wages (region populations) do not change much. The increase in welfare as fragmentation costs fall is larger. Part of potential welfare gains is dissipated by falling prices (worsening terms of trade with the outside world) due to the increased domestic productivity. Average prices $p(s)$ are 2.5% lower with full fragmentation than under fully integrated production. This fall in prices also holds down urbanization (producer wages and employment) as fragmentation costs fall. Nevertheless, falling fragmentation costs is analogous to an aggregate productivity improvement and raise welfare.

The bottom right panel of Figure 4 illustrates an effect which was not discussed in previous sections. The fall in fragmentation costs improves the competitiveness of the urban (manufacturing and services) sectors relative to the outside good. The vertical axis gives the trade balance (exports minus imports) of urban goods as a proportion all domestic urban goods production. This trade balance in urban sectors is normalized to zero at zero fragmentation costs. 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.

Turning to the spillovers case, Figure A2a shows results confirming those in Figure 2 earlier. There is a region of multiple equilibria: one in which all sectors are integrated and one in which some (middle) sectors are fragmented. Results corresponding to those in Figure 4 for the Ricardian case are qualitatively the same as for the Ricardian case, and thus we won't show them here.

One thing that is qualitatively different between the Ricardian and spillovers cases is the effect of increasing demand (increases in the alphas) in (10) on the equilibrium regime. In the Ricardian case in

which the λ's are constants, a symmetric situation ($w_1 = w_2$) means that the boundaries between the integrated and fragmented sectors do not depend on demand (also true in the partial-equilibrium case as seen in (4)). However, in (7) and here in (A14) - (A17) we see that increases in total market demand will affect the λ's and hence will affect regime boundaries in the spillovers case. Figure A2b shows the effect on the regime boundaries following a 50 percent increase in α_d and α_f. For middle levels of *t*, additional sectors will now fragment as shown, which implies increases function specialization and lower sectoral specialization for a given level of fragmentation costs.

4.4 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. This also seems closely consistent with the many examples given in Moretti (2012).

Figures A.3 and A.4 consider asymmetry between the sectors/regions in the Ricardian case. Figure A.3 assumes that region 1 has a comparative and absolute advantage in function *A*, while 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 A.3 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.

Two further results follow in the asymmetric Ricardian case. The right-hand panel of Figure A.3 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 A.4 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 empirics in section 5, we show that larger regions do have lower levels of both forms of concentration.

Finally, consider an asymmetric spillovers case, motivated by this idea that perhaps business service occupations are more characterized by spillovers than other functions. Figure A.5 shows a case where only function A has spillovers, but in both regions (in contrast to the Ricardian case where only λ_A is smaller in region 1 only). 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 A.5. 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.

An important point about Figure A.5 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. This also seems to fit much of the evidence and discussion in Moretti (2012).

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.

5. Sectoral and functional concentration in the US

The theoretical model provides a rich set of predictions that relate changes in fragmentation costs to changes in a region's sectoral and functional composition. In this section, we explore the empirical validity of three key predictions of the model using information on US employment. For empirical purposes, we interpret sectoral as industries, functional as occupational, and geographical as US states.

In section 5.1, we look at the spatial concentration of sectors and functions in order to test the hypotheses that as fragmentation costs fall sectoral concentration declines while occupational concentration rises. Fragmentation costs are not directly observed, and available proxies (e.g., travel costs, long-distance phone calls, or access to internet) do not provide either state- or sector-level variation.

Therefore, we simply assume fragmentation costs are falling over time, and use time as the proxy.[9] We find declining sectoral concentration and increasing occupational concentration over time. In line with the predictions of theory, a large fraction of those changes is explained by within-sector and within-function changes in geographic concentration.

In section 5.2, we explore time series changes in states' distributions of employment across sectors and across functions. The model predicts that, as fragmentation costs decrease, regions should experience decreasing sectoral specialization and increasing functional specialization. To test this hypothesis, we calculate our two measures of regional specialization defined in equations (13a) and (14a) for each state-year in the sample. As predicted by the model, we find that the states' sectoral specialization is decreasing overtime, whereas the functional specialization is increasing.

Finally, in section 5.3, we estimate the correlation between regional specialization and size (i.e., total employment in the region). The theoretical model predicts that larger regions have lower sectoral and functional specialization. In line with the prediction of model, we find a negative correlation between US states' size and measures of specialization for both sectors and functions.

5.1 Sectoral and functional concentration over time

In this section, we explore the first prediction of the model related to sectoral and functional geographic concentration. We begin by describing the main data sources. We then develop the method we use to measure the geographic concentration. Finally, we implement the index of concentration to study the time series changes in sectoral and functional concentration.

5.1.1 Data

To construct the indexes of concentration, such as those defined in equation (13), 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 theoretical model can accommodate different definition of sectors and regions. 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

[9] While this assumption is consistent with the general decrease in the cost of exchanging goods and services at a distance, it has two main drawbacks. First, it prevents us from exploiting across sector variation to identify the impact of changes in fragmentation costs. Second, it prevents us from making quantitative predictions regarding the impact of fragmentation costs on regional outcomes.

detailed industry.[10] 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 (15).

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).[11] 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 introduced in 1997 is revised multiple times, first in 2002, and subsequently in 2007, 2012, and 2017.[12] 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.[13] We also remove industries in the "Farming" (NAICS 11),

[10] Additional information on the QCEW is available online at https://www.bls.gov/cew/overview.htm.
[11] 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.
[12] For years 1990 to 1996, the QCEW is available on a NAICS basis even if the NAICS was introduced only in 1997.
[13] 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.

"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.[14] 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.[15] 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.

Together, the QCEW and the OES data allow us to construct the sectoral and functional concentration indices for each year in our sample.

5.1.2 The Ellison and Glaeser concentration index

Indices similar to G_s, defined in (13), are often used to measure agglomeration across regions (e.g., Krugman (1991) and Audretsch and Feldman (1996)). As explained by EG97, an important limitation of these measures is that they could suggest high levels of concentration in sectors comprised of a few large companies locate in a dispersed, random pattern. To control for this possibility, EG97 incorporate information about the size distribution of firms in the sector to construct the following index of concentration

[14] Additional information on the OES can be found online at https://www.bls.gov/oes/oes_emp.htm.
[15] 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.

$$EG_s = \frac{G_s/(1-\sum_r m_r^2) - H_s}{1 - H_s}, \tag{15}$$

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. EG97 refer to G_s (equation (13) above) as the "raw geographic concentration" of employment in a sector. The subtraction of H_s is a correction that accounts for the fact that the index G_s is expected to be larger in industries consisting of fewer larger plants if locations were chosen completely at random.[16]

The EG97 index of concentration defined in equation (15) has many useful properties.[17] First, it is easy to implement. Second, it is widely used which allows us to compare our results with previous studies. Third, it uses employment shares, which implies that it does not confound features in time-series data such as the general decline in manufacturing.

To measure functional concentration index, we use a modified version of the EG97 index defined as follows

$$EG_f = \frac{G_f/\left(1-\sum_r m_f^2\right) - H_f}{1 - H_f}. \tag{16}$$

As for sectors, we adjust our raw measure of concentration G_f, defined in (14), 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 occupations across plants. Instead, we use $H_f = \sum_s m_{fs}^2$, where m_{fs} is the share of employment in sector *s* performing function *f*.[18] The intuition for the correction factor H_f, suggested by Gabe and Able (2010), is that when a function's employment is

[16] In practice, changes in the value of the EG_s index over time are well approximated by changes in G_s. This happens because plant size distributions tend to change fairly slowly, 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.

[17] The motivation for the EG97 index defined in equation (15) 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 EG97 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.

[18] In the 2-function model of earlier sections, $m_{As} = a(s)/[a(s) + b(s)]$ if productivity λ_{fr} is the same for all *f, r*. If λ_{fr} varies then m_{As} is a mode weight average of these ratios adjusted by productivity factors λ_{fr}.

concentrated in a few industries, the measured geographic concentration of the function should be higher all else equal.

5.1.3 Sectoral concentration

As explained in sections 3 and 4 above, the theoretical model predicts that a decrease in fragmentation costs leads to lower sectoral concentration. To test this prediction, we explore the time-series in the geographic concentration index defined in equation (15). For this part of the empirical analysis, we use a balanced panel that contains state-level data on 626 six-digit NAICS industries across all sectors of the economy for years 1990 to 2019. About 41 percent of the 18,780 observations are in the manufacturing sector, the remainder of the observations are distributed across industries in the business services (23%), personal services (20%), and wholesale, retail and transportation (15%).

Time series changes in the geographic concentration of sector employment can be decomposed into two adjustments margins, within-sector changes in geographic concentration and across-sector reallocation of employment. We are mostly interested in quantifying the contribution of the first margin because the theoretical model's predictions are related to within-sector changes in employment concentration. For any given year τ, the mean sectoral concentration can be decomposed as follows

$$EG_{\tau}^{Sector} = \sum_s m_{s\tau}\, EG_{st} = \sum_s m_s\, EG_{s\tau} + \sum_s (m_{s\tau} - m_s)\, EG_{s\tau}, \qquad (17)$$

where $m_{s\tau}$ is sector-s's share of national employment in year τ and m_s is the sector's share of employment in the sample (i.e., the mean over time of $m_{s\tau}$). The first equality follows by definition of a weighted average. The second equality decomposes time series changes into two components. 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.

We report the results from decomposition (17) in Figure 5. The solid line depicts the weighted average EG_{τ}^{Sector}. It clearly shows the steady decline in the weighted mean geographic concentration of sector employment. The dashed line depicts the within-industry component of the decomposition, i.e., the term $\sum_s m_s\, EG_{s\tau}$ in equation (17). The figure makes clear that even when holding the employment weights constant, the mean geographic concentration of sectors declines steadily over time.

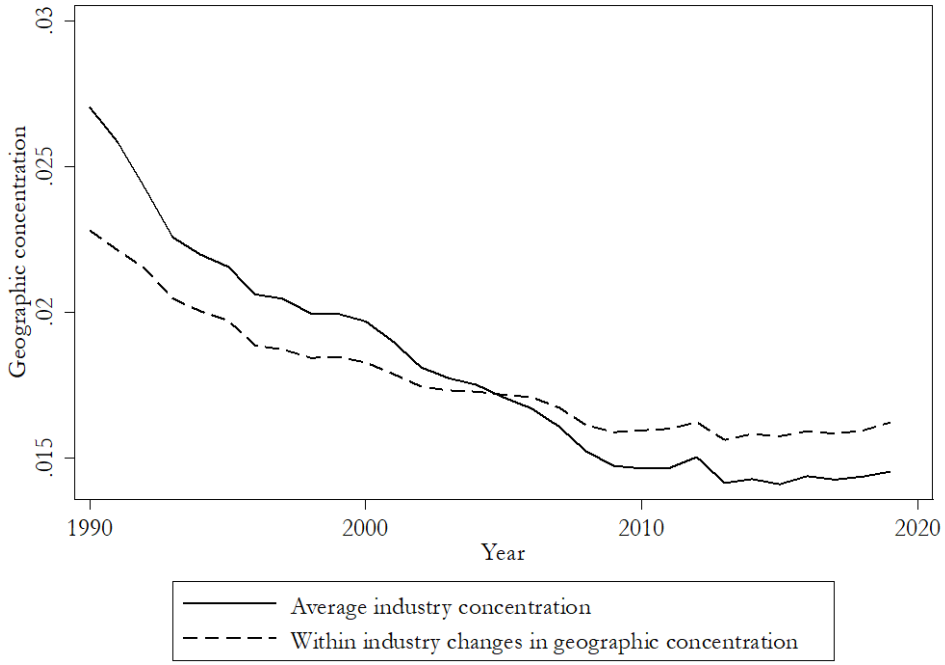


Figure 5: Geographic concentration of sectors over time

As illustrated in Figure 5, the rate of decay is lower when considering only the within-sector changes in concentration. This happens because part of the observed decrease in sectoral concentration is due to labor movement from less concentrated industries towards more concentrated industries. As seen from the first line table 1, the mean sectoral concentration decreases by about 44% over the period (going from 0.027 in 1990 to 0.015 in 2019), while the within-sector component decreases by about 30% (going from 0.023 in 1990 to 0.016 in 2019) as shown in the second line. So, the decline in the within-industry component of geographic concentration is large in absolute term and represents the majority of the time series change in geographic concentration. Overall, the results presented in Table 1 suggest that the average worker is employed in a more geographically dispersed sector in 2019 than he was in 1990. Given our assumption on the evolution of fragmentation costs, the decrease in the sectoral concentration over time observed in the data is consistent with the predictions of the theoretical model.

Table 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 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.

To get a sense of which component of the weighted mean drives the time series changes, Table 1 also 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 ($EG < 0.02$).

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 EG97 indices on a time trend controlling for sector-level factors using fixed effects

$$\ln EG_{s\tau} = \beta_s + \beta\, Trend_\tau + \varepsilon_{s\tau}\,. \tag{18}$$

Under the assumption that fragmentation costs are decreasing over time, the theoretical model predicts that the trend, β, should be negative.

The results from estimating equation (18) by OLS are reported in Table 2. The first row reports the results for the full sample of 626 six-digit NAICS sectors. As predicted, 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 (18) separately for each broad group: manufacturing, business services, personal services, and wholesale, retail and transportation. As reported in Table 2, every point estimate is negative and statistically

significant. Overall, the results presented so far, support the prediction that the geographic concentration of sectoral employment is declining over time.

Table 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.

The results presented in this section share many similarities with the findings of Dumais, Ellison, and Glaeser (2002) who study the geographic concentration of sectoral employment across US states from 1972 to 1997. First, the two sets of estimates are of the same magnitude. They report a (simple) mean 0.034 for 1992. Our corresponding estimate is 0.056 (not in Table 1). 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.

5.1.4 Functional concentration

In this section, we use the decomposition in equation (17)—defined over functional shares instead of sectoral shares— to 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 across all sectors of the economy for years 2000 to 2019.

The results are depicted in Figure 6. The solid represents the employment-year weighted mean concentration, while the dashed line depicts the within-function component of the weighted average. The figure clearly shows that there is an increase in the geographic concentration of functions over time, even when holding the employment weights constant. Our empirical results complement those of previous studies, such as Berry and Glaeser 2005, Duranton and Puga (2005), Moretti (2013) and Diamond (2016), that documents divergence in the skill-level of U.S. cities over time. We provide evidence that functional concentration holds even within disaggregated definitions of occupations.

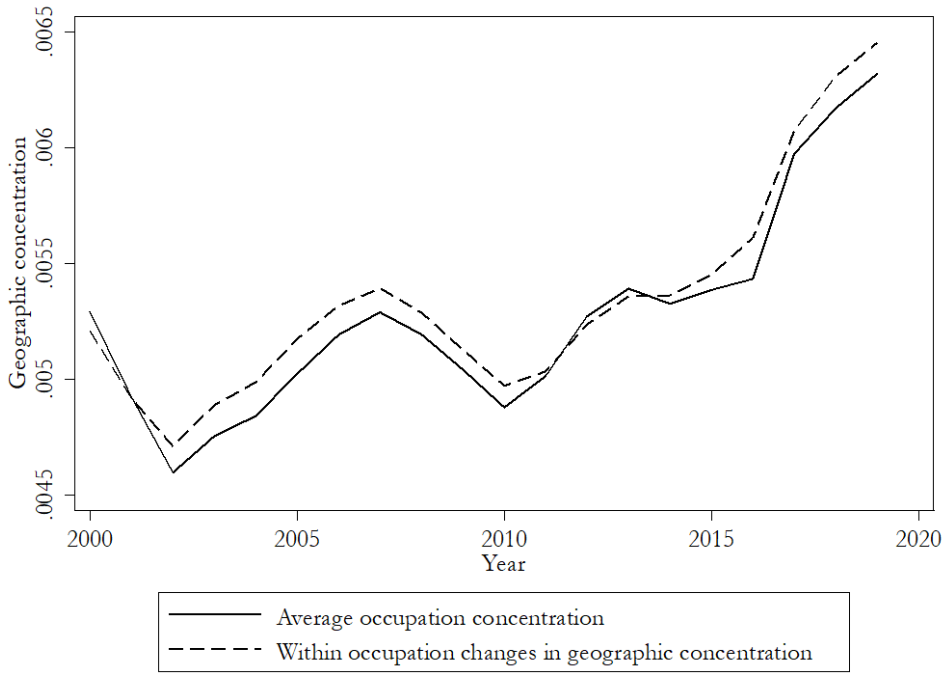


Figure 6: Geographic concentration of functions over time

Results from the decomposition (17), applied to functions, are reported in Table 3 for selected years. As seen in the table, 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 (14). 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 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

$$\ln EG_{f\tau} = \beta_f + \beta\, Trend_\tau + \varepsilon_{f\tau} \tag{19}$$

to estimate the time trend of geographic concentration. Under the assumption that fragmentation costs are decreasing over time, the theoretical model predicts that the trend, β, should be positive. The results are reported in Table 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 6. The remaining rows of Table 4 show that 17 out of 21 estimated time trends are positive and 13 of those are statistically significant at conventional levels.

Table 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.

Overall, the results presented in Figure 6 and Tables 3 and 4 provide empirical support to the predictions of the theoretical model. As explained in sections 3 and 4 above, as fragmentation costs fall, more sectors fragment such that regions move from sectoral to functional specialization. Under our assumption, this implies that function concentration should increase over time.

5.2 Regional specialization over time

In this section, we explore the sectoral and functional structure of regional employment. Under our assumption about the time series evolution of fragmentation costs, we expect to find a decrease in sector specialization and an increase in functional specialization.

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 (13a) and (14a) for each region-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. The results are reported in Figure 7. The decreasing trend observed in panel (a) indicates that the states' employment is becoming more evenly distributed across sectors over time. Conversely, panel (b) shows that states' distribution of employment across function is becoming increasingly uneven. As predicted by the theoretical model, these results indicate that states are becoming less specialized in terms of sectoral employment, but more specialized in terms of functional employment.

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

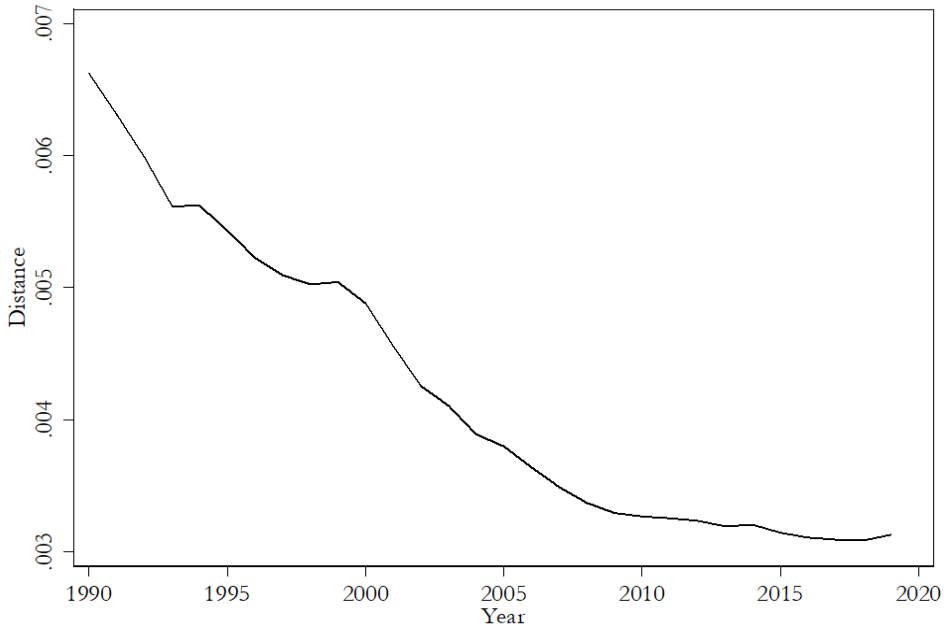
$$\ln D_{r\tau} = \beta_r + \beta\, Trend_\tau + \varepsilon_{r\tau}, \tag{20}$$

where β_r represent region-level fixed effects. The results are reported in Table 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. Overall, the results provide empirical support to the predictions of the theoretical model.

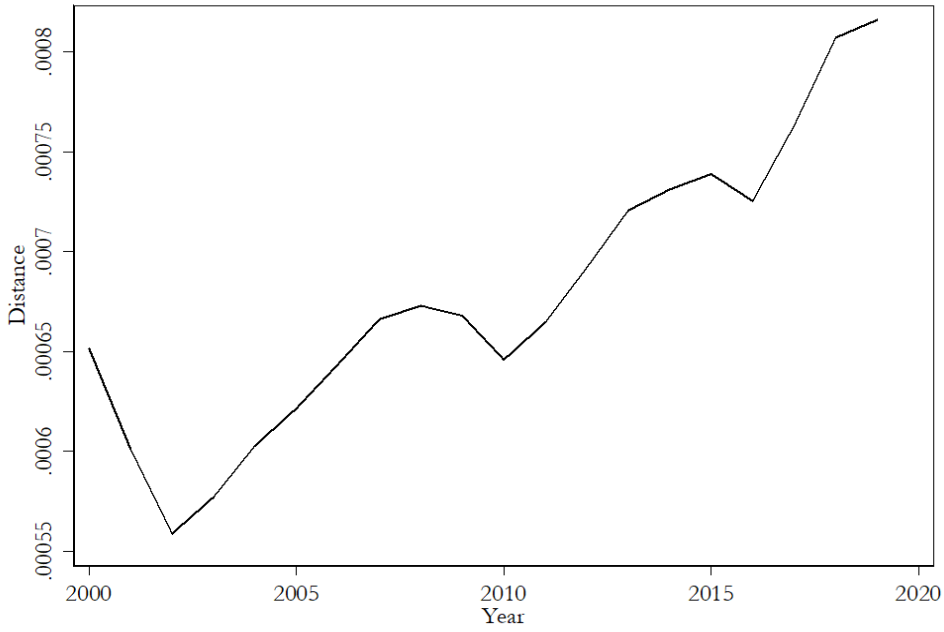
Table 5. Time series trend of regional specialization

	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.



(a) sectors



(b) Functions

Figure 7: Regional specialization over time

5.3 Region size

As explained in sections 3 and 4 above, the theoretical model predicts that larger regions have lower industrial and functional employment concentration. To test this prediction, we use thee indices of regional specialization for sectors and functions. The first set of measures are the indices D, defined in (13a) and (14a). The second set are Herfindahl-Hirschman indices (HHI) defined, respectively, over sectoral and functional employment for each region-year in the sample as

$$HHI_{r\tau}^{sector} = \sum_s q_{rs\tau}^2, \qquad \text{and} \qquad HHI_{r\tau}^{function} = \sum_f q_{rf\tau}^2. \tag{21}$$

These measures, which are commonly used in the literature, are similar to our index *D* but without the deviation from national employment. The third set of measures are Krugman (1991) indices of regional specialization defined as

$$K_{r\tau}^{sector} = \sum_s |q_{rs\tau} - \bar{q}_{rs\tau}|, \qquad \text{and} \qquad K_{r\tau}^{function} = \sum_s \left|q_{rf\tau} - \bar{q}_{rf\tau}\right|, \tag{22}$$

where $\bar{q}_{rs\tau}$ and $\bar{q}_{rf\tau}$ denote the region's average share of employment in a sector and a function in year τ, respectively. By definition, high values of the specialization indices imply that regional employment is concentrated among a small number of sectors or functions. In our sample, the correlation between the two indices is 0.50 for sectors and 0.85 for functions.

We test for the negative association between regional specialization and size by estimating regressions of the form

$$Spec_{r\tau} = \beta_\tau + \beta \log emp_{r\tau} + \varepsilon_{r\tau}, \tag{23}$$

where $Spec_{r\tau}$ represents one of the three specialization indices (D, HHI, or K), β_τ denotes year fixed effects, $emp_{r\tau}$ is the state's employment, and $\varepsilon_{r\tau}$ is a residual term that capture the impact of exogenous factors that affect regional specialization and are not included in the model.

We report results from estimating (23) by OLS in Table 6. Panels A and B report, respectively, results for sectoral specialization and functional specialization. The first, second and third line present, respectively, the results using the D indices defined in (13a) and (14a), the HHI index of specialization defined in equation (21) and the Krugman specialization index defined in equation (22). To obtain more meaningful magnitudes for the point estimates, we report so-called "beta coefficients" (defined as the usual OLS point estimates multiplied by the ratio of the independent and dependent variables' standard

deviation) which gives the number of standard deviations in the dependent variable associated with a one standard deviation change in the independent variable.

As seen in the table, the point estimates vary across measures of specialization but, in all cases, the partial correlation between the measures and regional employment is negative and statistically significant at conventional levels as expected. These results indicate that region size is a strong predictor of the cross-sectional variation in both sectoral and functional specialization.

Table 6. Regional size and specialization

	Estimates	*Std. Err.*	R^2
Panel A: Sectoral specialization			
D	-0.595***	0.077	0.412
HHI	-0.490***	0.071	0.297
Krugman	-0.800***	0.064	0.735
Panel B: Functional specialization			
D	-0.642***	0.095	0.458
HHI	-0.173*	0.103	0.237
Krugman	-0.171*	0.102	0.228

Notes: This table reports OLS results from regressing measures of regional specialization on region employment. The sample in Panel A contains 1,500 State-year observations over period 1990 to 2019. The sample for Panel B contains 1,000 State-year observations over period 2000 to 2019. Standard errors are clustered by major sector. The *, **, and *** indicate statistical significance at the 10, 5, and 1 percent level, respectively.

6. Conclusions

Our paper is motivated by what are widely seen as changes in the nature of work and changes in scope of activities 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.

The paper draws 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, synonymous with occupations in the empirical analysis. 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.

This basic model result is in turn used to draw out qualitative and quantitative predictions about a range of issues including how concentration indices for sectors and function behave, welfare effects and a country's trade position with the outside world. Asymmetric cases capture some interesting outcomes, including a possibility that falling fragmentation costs lead to a divergence in city size over some range of these costs.

We do not have good measures of these fragmentation costs and existing proxies do not provide either state or sector level variation. But we are able to measure key relationships over a twenty-year period for functions, thirty years for sectors. We find that over time our measure of sectoral concentration within regions has steadily decreased and functional concentration has increased. We show that these adjustments are not just due to employment shifting from concentrated sectors to dispersed sectors; e.g., it is not due to employment shifting from geographically concentrated manufacturing to dispersed services. Our effect holds just as strongly within sectors.

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. Empirically, they also have the property that regional sectoral specialization is falling over time and regional functional specialization is rising, though the former has a slight u-shaped feature at the end of the time period.

Finally, we find that larger regions are less specialized in both sectors and functions. All three results are consistent with the model and with fragmentation costs that are falling over time. We argue, though only informally, that our theoretical and empirical findings seem consistent with a wide range of results in several related literatures.

References

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, Paul (1991), "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: Section 3 theory

Table A1: Employment by function *f* = A, B, in sector *s* and region *r* = 1, 2.

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

Table A.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	$L_{\mathrm{A1}} = \lambda_{\mathrm{A1}} s_2[1 + \gamma(1 - s_2)]n/2$	$L_{A2} = \lambda_{A2}\,(1 - s_2)(1 - \gamma s_2)n/2$
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*	
	$L_1 = L_{A1} + L_{B1} = \int_0^1 L_1(s)ds$	$L_2 = L_{A2} + L_{B2} = \int_0^1 L_2(s)ds$

Unit profit functions are:

$\pi_1(s) = p(s) - \{2\lambda + \Delta\lambda[1 - \gamma(1 - 2s)]\}w/2,$ $\pi_F(s) = p(s) - \lambda w - T,$

$\pi_2(s) = p(s) - \{2\lambda + \Delta\lambda[1 + \gamma(1 - 2s)]\}w/2.$

Proposition 1:

iib) $s_2 = \frac{1}{2}\left[1 + \frac{1}{\gamma}\left(1 - \frac{2t}{\Delta\lambda}\right)\right]$, decreasing in *t*. $s_2 = 1$ when $t = (1-\gamma)\,\Delta\lambda/2$

iic) $L_{A2} = \lambda_{A2}\,(1 - s_2)(1 - \gamma s_2)n/2$. If γ < 1 then $L_{A2} = 0$ at $s_2 = 1$, i.e. $t = (1-\gamma)\,\Delta\lambda/2$. All sectors use all functions and all sectors are fragmented at this value of *t*.

If γ > 1 then $L_{A2} = 0$ at $\gamma s_2 = 1$, i.e. $t = (\gamma - 1)\,\Delta\lambda/2$. Some sectors use only one function: all sectors that use both functions are fragmented at this value of *t*.

Section 3.2: localisation economies

Using equation (6a) in (7a) gives the unit profit advantage from integration,

$$\Pi(s_1, t) \equiv \pi_1(s_1) - \pi_F(s_1) = t - [1 - \gamma(1 - 2s_1)]\left(\Delta\Lambda - \sigma n\left\{-\frac{1}{2} + s_1[1 - \gamma(1 - s_1)]\right\}\right)w/2, \quad \text{(A1)}$$

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 A1), $t^{**} = [\Delta\Lambda + n\sigma\gamma/4]w/2$.

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, $\Delta\Lambda$.

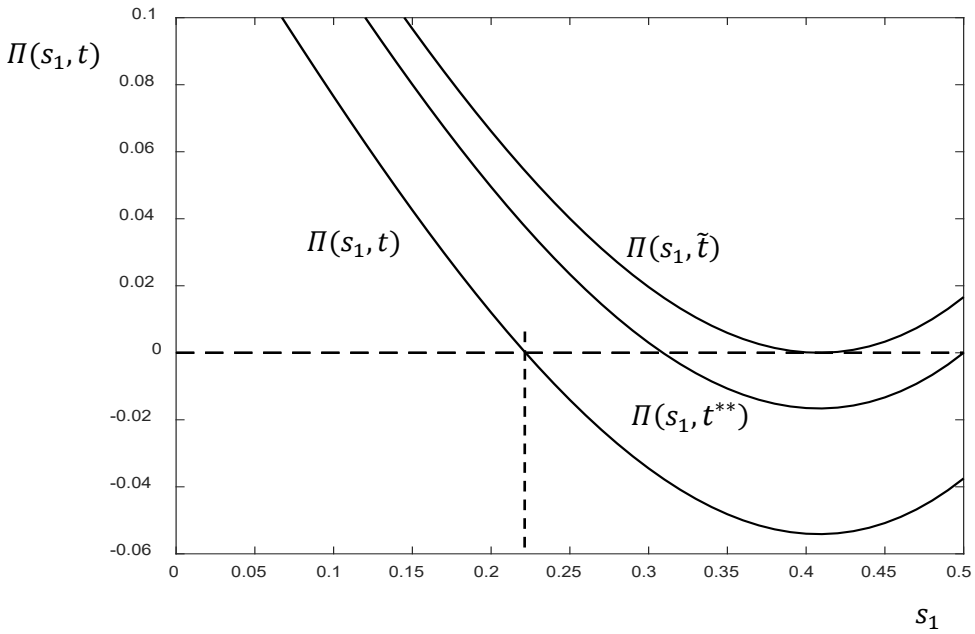


Figure A1: Expression A1 for different values of t

Appendix 2: 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

$$U = \beta \ln\left\{\prod_s \left[\theta_d \left(\frac{Q_{dd}(s)}{\theta_d}\right)^{\frac{\varepsilon-1}{\varepsilon}} + \theta_f \left(\frac{Q_{fd}(s)}{\theta_f}\right)^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}}\right\} + R \qquad \text{(A1)}$$

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 (12).

$$Y = w_0\bar{L} + H_1 + H_2 \qquad \text{(A2)}$$

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

$$Y = R + \sum_s p(s)Q_{dd}(s) + \sum_s \bar{p}Q_{fd}(s) \qquad \text{(A3)}$$

(A3) can be substituted into (A1) to replace R.

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

Maximization of (A4) 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:

$$Q_{dd}(s) = \alpha_d \theta_f p(s)^{-\epsilon} / \{\theta_d\, p(s)^{1-\epsilon} + \theta_f \bar{p}^{1-\epsilon}\} \qquad \text{(A5)}$$

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 $p(s) = \overline{p} = 1$. 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.[19]

Appendix 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.

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 z (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 z

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 $\perp$ 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

[19]Our algebra indicates that the relationship between the β in (A1) 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.

formulation of the fragmentation labor use, which divides it between the two regions, each using $t/2$ per F type firm.

$$L_1 \geq \sum_s n_1(s)(a(s)\lambda_{A1} + b(s)\lambda_{B1}) + n_F(s)a(s)\lambda_{A1} + n_F(s)t/2 \quad \perp L_1 \quad \text{(A6)}$$

$$L_2 \geq \sum_s n_2(s)(a(s)\lambda_{A2} + b(s)\lambda_{B2}) + n_F(s)a(s)\lambda_{A2} + n_F(s)t/2 \quad \perp L_2 \quad \text{(A7)}$$

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

$$(w_1 - w_0)K/c \geq L_1 \quad \perp \quad w_1 \quad \text{(A8)}$$

$$(w_2 - w_0)K/c \geq L_2 \quad \perp \quad w_2 \quad \text{(A9)}$$

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

$$X_{A1} \geq \sum_s a(s)\big(n_1(s) + n_F(s)\big) \quad \perp \quad X_{A1} \quad \text{(A10)}$$

$$X_{A2} \geq \sum_s a(s)n_2(s) \quad \perp \quad X_{A2} \quad \text{(A11)}$$

$$X_{B1} \geq \sum_s b(s)n_1(s) \quad \perp \quad X_{B1} \quad \text{(A12)}$$

$$X_{B2} \geq \sum_s b(s)\big(n_2(s) + n_F(s)\big) \quad \perp \quad X_{B2} \quad \text{(A13)}$$

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

$$\lambda_{A1} \geq \Lambda_{A1} - \sigma_A \mathrm{X}_{A1} \quad \perp \quad \lambda_{A1} \quad \text{(A14)}$$

$$\lambda_{A2} \geq \Lambda_{A2} - \sigma_A \mathrm{X}_{A2} \quad \perp \quad \lambda_{A2} \quad \text{(A15)}$$

$$\lambda_{B1} \geq \Lambda_{B1} - \sigma_B \mathrm{X}_{B1} \quad \perp \quad \lambda_{B1} \quad \text{(A16)}$$

$$\lambda_{B2} \geq \Lambda_{B2} - \sigma_B \mathrm{X}_{B2} \quad \perp \quad \lambda_{B2} \quad \text{(A17)}$$

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: $t(w_1 + w_2)/2$.[20] Therefore

$$w_1(a(s)\lambda_{A1} + b(s)\lambda_{B1}) \geq p(s) \quad \perp \quad n_1(s) \quad \text{(A18)}$$

$$w_2(a(s)\lambda_{A2} + b(s)\lambda_{B2}) \geq p(s) \quad \perp \quad n_2(s) \quad \text{(A19)}$$

$$w_1 a(s)\lambda_{A1} + w_2 b(s)\lambda_{B1} + t(w_1 + w_2)/2 \geq p(s) \quad \perp \quad n_F(s) \quad \text{(A20)}$$

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

$$Q_d(s) \geq n_1(s) + n_2(s) + n_F(s) \quad \perp \quad Q_d(s) \quad \text{(A21)}$$

[20] Note that all inequalities are homogeneous of degree 1 in wages and prices.

The final element is to specify the demand size of the model, which links outputs, prices, and the external foreign market. The domestic country is assumed small as an importer, and so all foreign prices for the z sectors are given by an exogenous value, common across all sectors. Domestic and foreign goods within a sector are CES substitutes with an elasticity of substitution $\varepsilon > 1$. Sectoral composites (domestic and foreign varieties) are Cobb-Douglas substitutes. The outside 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).

The market clearing equation for the domestic good z 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 *z*.

$$Q_d(s) = Q_{dd}(s) + Q_{fd}(s) = \frac{\alpha_d \theta_d p(s)^{-\epsilon}}{\theta_d\, p(s)^{1-\epsilon} + \theta_f \bar{p}^{1-\epsilon}} + \frac{\alpha_f \theta_d p(s)^{-\epsilon}}{\theta_f\, p(s)^{1-\epsilon} + \theta_f \bar{p}^{1-\epsilon}} \qquad \perp \qquad p(s) \qquad \text{(A22)}$$

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

$$Q_{fd}(s) = \frac{\alpha_d \theta_f \bar{p}^{-\epsilon}}{\theta_d\, p(s)^{1-\epsilon} + \theta_f \bar{p}^{1-\epsilon}} \qquad \perp \qquad Q_{fd}(z) \qquad \text{(A23)}$$

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

Figure A2a: Symmetric Spillovers Case
Spillovers, free entry, no Ricardian comparative advantage
Base Case: region of multiple equilbria

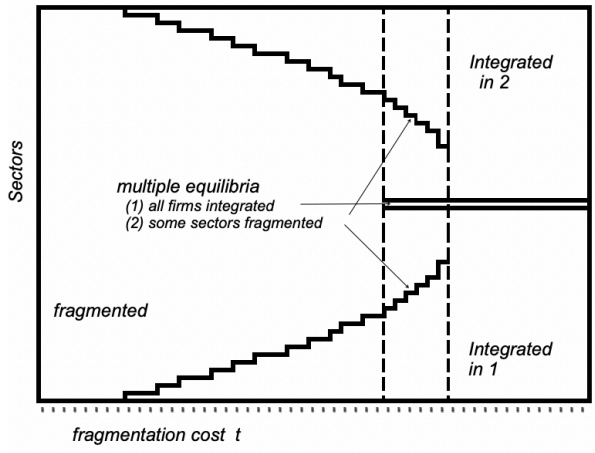


Figure A2b: Symmetric Spillovers Case
Spillovers, free entry, no Ricardian comparative advantage
Base case, + 50% increase in market size

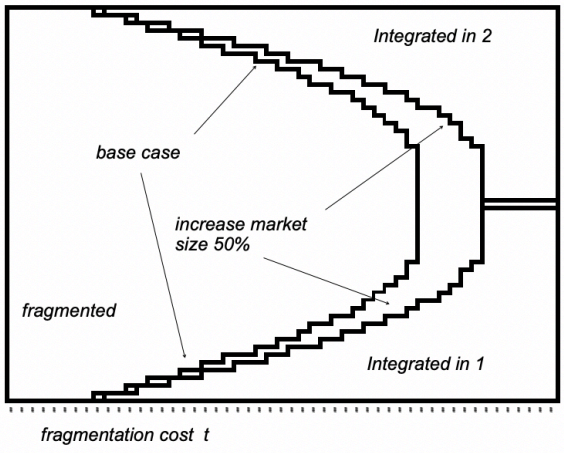


Figure A.2: Symmetric Spillovers Case

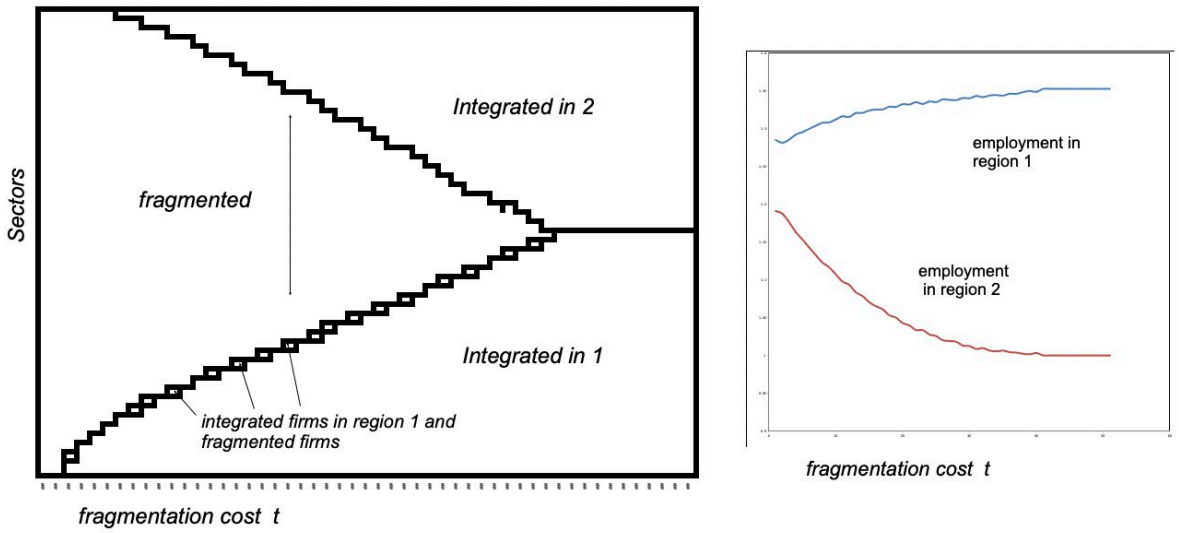


Figure A.3: Asymmetric Ricardian Case

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

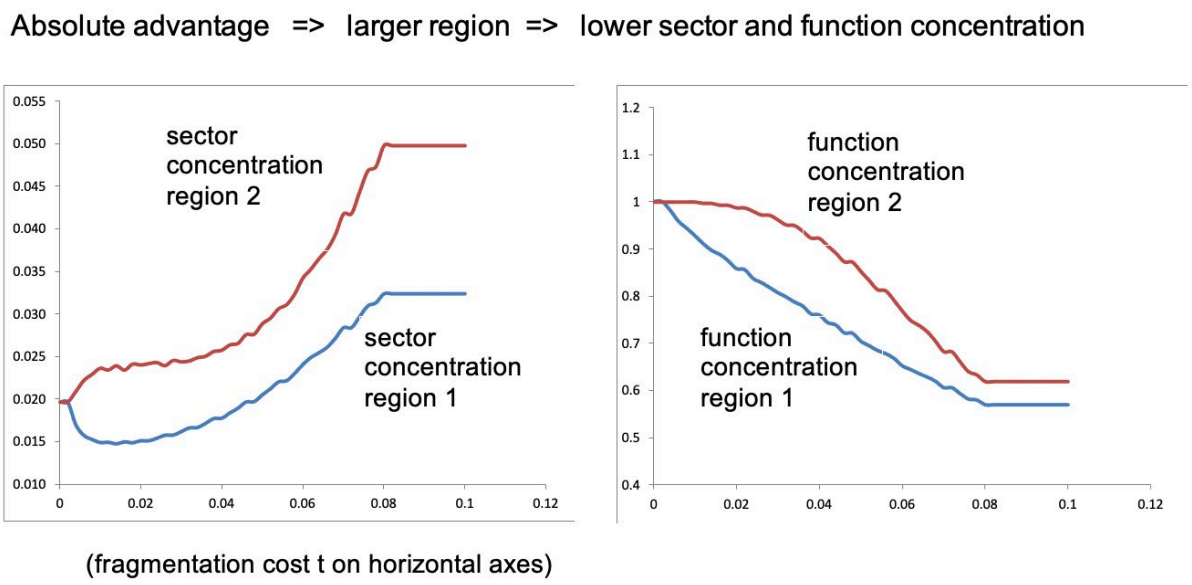


Figure A.4: Asymmetric Ricardian Case

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

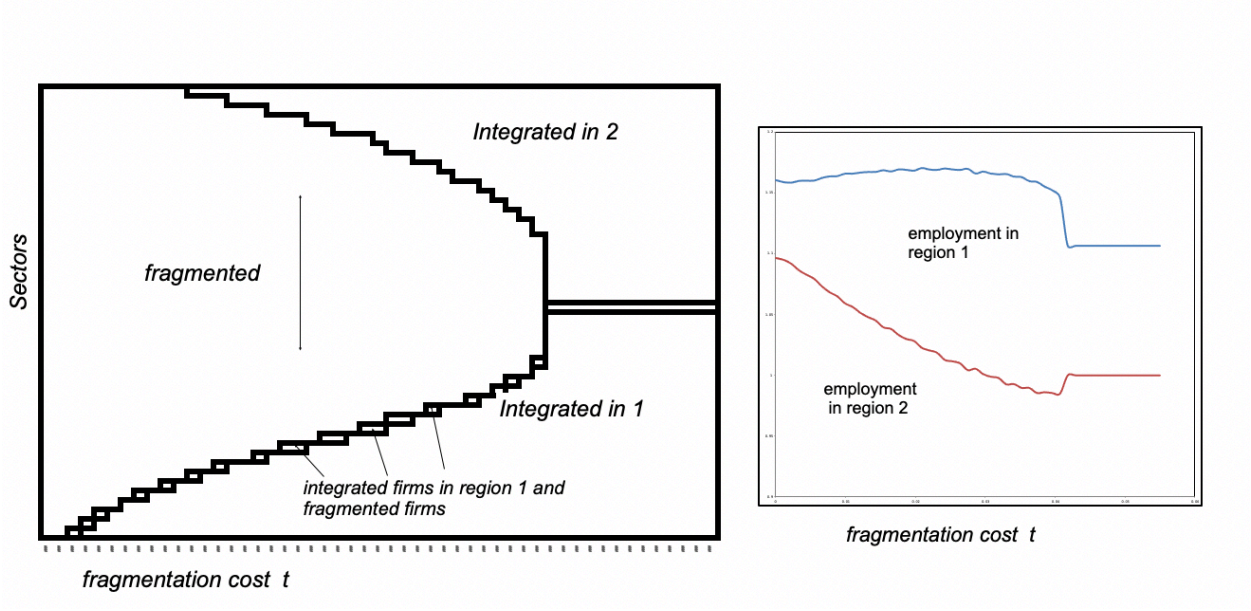


Figure A.5: Asymmetric Spillovers Case; spillovers in function *A* only