NBER WORKING PAPER SERIES

URBAN SPECIALISATION; FROM SECTORAL TO FUNCTIONAL

Antoine Gervais
James R. Markusen
Anthony J. Venables

Working Paper 28352
http://www.nber.org/papers/w28352

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2021

Urban Specialisation; from Sectoral to Functional
Antoine Gervais, James R. Markusen, and Anthony J. Venables
NBER Working Paper No. 28352
January 2021
JEL No. F12,F23,R11,R12,R13

ABSTRACT

The comparative advantage of many cities is based on their efficiency in the production of 'functions', e.g., business services such as finance, law, engineering, or similar functions that are used by firms in a wide range of sectors. Firms that use these functions may choose to source them locally, or to purchase them from other cities. The former case gives rise to cities developing a pattern of sectoral specialization, and the latter a pattern of functional specialization. A two-city country trades with the larger world, and workers within the country are mobile between the two cities. Productivity in a given function varies across cities, giving rise to urban comparative advantage. This may be due to exogenous technological differences (Ricardian) or to city- and function-specific scale economies. Sectors differ in the intensity with which they use different functions, giving rise to a pattern of sectoral and functional specialisation. We generate a number of economic insights, and examine the model's predictions empirically over a 20-30-year period for US states. As geographic fragmentation costs fall, both our theory and empirical analysis show that sector concentration and regional specialization fall for sectors and rise for functions (occupations).

Antoine Gervais
Department of Economics
University of Sherbrooke
Canada
Antoine.Gervais2@usherbrooke.ca

James R. Markusen
Department of Economics
University of Colorado
Boulder, CO 80309-0256
and NBER
james.markusen@colorado.edu

Anthony J. Venables
Department of Economics
University of Oxford
Manor Road Building
Manor Road
Oxford OX1 3UQ
United Kingdom
and CEPR
tony.venables@economics.ox.ac.uk

1. Introduction

The production of final products and services typically requires numerous functions to be performed. Manufactured goods require engineering, finance and marketing; construction requires architects and lawyers, and so on. There may be spatial differences in the efficiency with which such functions can be supplied so, if the functions are not perfectly tradable, efficiency differences in functions will translate into a pattern of comparative advantage in final goods. This paper develops these ideas in the spatial context of cities or regions within a country, and investigates the impact of such differences for firm organisation, city specialisation, trade in goods, and for the associated gains from trade.

The concept of “function" is fuzzy, depending on how narrowly it is defined. A rather aggregate level is the distinction between headquarters and production, as developed in some of the literature on foreign direct investment (Markusen 2002) and work in the urban context (Duranton and Puga 2005, Rossi-Hansberg et al. 2009). Alternatively, functions could be identified with occupations. Indeed, a common statistical breakdown is to divide a firm’s workforce into production (or blue-collar) and non-production (or white-collar) workers, or into finer occupational definitions.

The concept we seek to capture in this paper is finer than HQ vs. production or blue-collar vs. white collar, and corresponds to functions such as engineering, finance, or law, and their specialisms. They have several characteristics. First, most functions are required in most sectors, though in different proportions, which we refer to as the function intensity of a sector. Second, many large cities appear to have developed strong functional specialisms. London and New York in business services: finance, but also legal and advertising; the San Francisco area in both hardware and software; Los Angeles in a range of media and creative sectors.

Third, labour productivity in a function differs by city, and we suggest that this may be the fundamental level at which city comparative advantage is based. Cities develop the skill set – through learning or the composition of its labour force – that comes to define what the city is good at. The fourth characteristic is that firms can purchase their functional needs in one place, or from several different ones. Many workers in London and New York, in functions such as finance, accounting, law, or advertising, provide services for firms in many sectors and in different places.

This concept of function is broader than that of a ‘task’, often thought of as a narrow stage of production and modelled as a continuum (Grossman and Rossi-Hansberg 2008, 2012, Autor 2013). While some cities specialise in quite narrow ‘tasks’, at least for large cities or regions, the broader functional or professional concept seems a better descriptor of their specialisms. Formally, the task models often go from a continuum of tasks to a single final good; in the analytical part of this paper we go from few functions to a continuum of final goods sectors.

In our model, spatial heterogeneity derives from the productivity of labour in performing functions. This translates, via the function intensity of sectors, into comparative advantage in goods. In the simplest cases in which all functions are perfectly tradable this is purely mechanical, adding an intermediate step from factor productivity to sectoral comparative advantage. However, the main direction we take in this paper is to vary the cost of sourcing functions from different cities. Firms then face a trade-off, between the efficiency with which functions operate in different cities and additional costs incurred if they source the functions they need from several different cities: we call these fragmentation costs, arising if e.g., engineering is purchased in one city, legal services in another, and so on.

The model shows how the interactions between fragmentation costs, the function intensity of different sectors, and efficiency differences between cities cause firms in some sectors to integrate production in one place, and in others to fragment it between cities. Firms' choices have implications for cities' production structures; to what extent are cities able to specialise in the functions in which they are most efficient, and how does this map into the sectoral specialisation of cities and countries?

In order to investigate these questions, we develop a model that has elements of economic geography, the literature on vertical multinationals, urban economics, and external economies of scale with some novel twists. A country contains two regions or cities, with identical workers who are mobile between jobs within and between cities. There are many final products (sectors) and just two functions, sectors requiring the functions in different proportions. There is free trade in final products, capturing the idea that the cities under study are embedded in an integrated market. However, sourcing functions from different places – i.e. splitting the production of a good between two locations – incurs a 'fragmentation cost'. This may be the cost of transporting 'functions' between cities, but is better thought of as coordination costs and the communication costs of maintaining links with suppliers in different cities.

The efficiency with which functions are produced is city specific, and we start with the simplest case in which these are exogenous Ricardian productivity differences. This provides a very clean example of how reducing fragmentation costs causes firms in some sectors to fragment (sourcing from both cities), and causes cities to move from sectoral towards functional specialisation. Sectors with extreme function intensities are more likely to contain integrated firms, concentrating production in the city with the advantage in the function in which they are intensive. Sectors which draw more equally on both functions will contain firms that are fragmented, performing each in the city with respective efficiency advantage.

The Ricardian model provides a simple introduction, but functional productivity differences are, we think, more likely to arise endogenously from learning and network formation amongst functionally specialist workers, and consequent increasing returns to scale. We therefore add agglomeration (localisation) economies to the model. These are specific to the function, not to the sector as in the standard model of Marshallian economies. It follows that there is a 'linkage' or complementarity between firms in different sectors. Although there are no direct technological spillovers between them, a range of sectors will benefit from an expansion and productivity increase in local production of a function which they use intensively. This model creates the possibility of multiple equilibria and discontinuous change; as fragmentation costs fall and firms fragment, functions are able to concentrate thereby raising productivity. Welfare gains from reductions in fragmentation costs can therefore be particularly large if they induce spatial reorganisation and the move from sectoral to functional specialisation.

The model generates a number of hypotheses about the effects of falling fragmentation costs. The principal implications are first that concentration of functions should rise and sectoral concentration should fall. Similarly, regional specialization in functions should rise and that of regional specialization in sectors should fall.

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. Our main 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 assertions that our empirical analysis over time is a rough proxy for falling fragmentation costs in the theory section.

Several findings emerge from our empirical analysis. First, we find declining sectoral concentration and increasing functional (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. A second result is that regional specialization indices in functions and sectors have the same properties as the concentration indices.   A third finding is that larger regions have lower levels of both sectoral and functional specialization.

The questions we pose and the model we develop touch on many strands of international trade, economic geography, and urban economics. Sorting the considerable volume of literature related to our work into boxes is not an easy task: many papers overlap several categories, some papers are largely or entirely theoretical, some are solely empirical. Here we attempt to identify some rough groupings that form the background of our work. There are many papers on trade, the fragmentation of production, expansion of trade at the extensive margin, multinational firms, and trade costs between countries. These include Autor (2013), Duranton and Puga (2005), Grossman and Rossi-Hansberg (2008, 2012), Limao and Venables (2002), Markusen (2002, 2013), Markusen and Venables (2007).[1]

A general work on urban structure is Henderson and Thisse (2004). Some of our analysis here relates to a large literature on economic geography, agglomeration, and multiple equilibria. 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).

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 sweep of Moretti (2012.

In the next section, we develop a partial equilibrium model with two symmetric regions with exogenous Ricardian differences in functional productivity. 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 for US states for the period 1990-2019.

2. Cities, sectors and functions

The ingredients of the model are locations, focussing on two cities; 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 cities and functions.[2] 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

[1] Duranton and Puga (2005) is the closest to our theory model as the titles suggest. But the approach is quite different. In Duranton and Puga, there is no function intensity differences across sectors and cities do not have function comparative advantage as we do. There is a function (headquarters) which has agglomeration economies similar to our approach. But plants have agglomeration economies at the sector level, while ours are at the function level. Our model creates a distribution of fragmented and integrated firms across industries and across cities and identifies the characteristics of industries that are fragmented versus integrated and of the city in which integrated firms locate.

[2] Thus, engineers can convert to lawyers. Comparative advantage comes from efficiency differences for a given function between cities. It would be possible to add a Heckscher-Ohlin flavour by assuming endowments of engineers and accountants, but this seems to add little to our basic story. One way of interpreting this assumption is that labour is perfectly mobile internationally, although the national housing stock is fixed (so that engineers can be traded for accountants).

to ensure that the two cities 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 cities are indexed $r = 1, 2$, and the wage rate in city *r* is denoted w_r. The single factor of production, labour, is perfectly mobile between cities but, since the cost of living may vary across cities, so may the nominal wage rate. The two functions, labelled *f* = *A*, *B*, are produced by labour with productivity that varies by city and function; production of one unit of function *f* in city *r* requires $\lambda_{fr} > 0$ units of labour. Cities are labelled such that productivity differences (if any) give city 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 $s \in [0,1]$. Sector *s* contain $n(s)$ firms, each of which produces one unit of output which 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. These intensities vary with sector *s* but are the same in both cities; we rank sectors such that low *s* sectors are *A*-intensive, i.e. $a'(s) < 0$ and $b'(s) > 0$.

Firms in each sector can source functions from either city, but if the two functions come from different cities then a fragmentation cost *t* is incurred.[3] Each firm therefore operates in one of three modes, choosing to operate entirely in city 1, entirely in 2, or to purchase one function from city 1 and the other from city 2.[4] Firms that produce in a single city are 'integrated' and will be labelled by subscript 1, 2 according to city of operation; those operating in both are 'fragmented' (subscript *F*). The profits of a firm in sector *s* for each of the three production modes are therefore

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

Costs are that of labour in the functions used, times the city wage rate. Thus, a firm in sector *s* uses $a(s)$ units of function *A* and $b(s)$ units of *B*. The functions use labour, with input per unit output in city *r* given by λ_{fr}, *f* = *A*, *B*, and costed at 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 cities, urban comparative advantage is determined entirely by the efficiency with which cities use labour to produce functions, λ_{fr}.

[3] We think of functions as being produced within the organisational boundaries of each firm, although they could just as well be outsourced and purchased through an arms-length relationship.

[4] The assignment of which function to which city will become clear, and does not merit additional notation.

Firms' choice of mode partitions the continuum of sectors into three groups. First is a range of *s* in which firms are integrated and source both functions in city 1. Since we have labelled cities such that city 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 firms are fragmented, sourcing function *A* from city 1 and function *B* in city 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 firms are integrated and operate only in city 2, the city 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

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

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

Given the number of firms in each sector, $n(s)$, employment levels by function, city, 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 sum employment of each factor in each city over sectors (giving L_{fr}), employment in each sector embodied in functions (giving $L_r(s)$), and over both to give total employment in each city, L_r.

3. Sectoral and functional specialisation in symmetric equilibria

We start by analysing the way in which firms' mode 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 cities and sectors symmetrical, enabling us to derive a number of key results. We assume that the number of firms in each sector *s* is the same and constant, such that $n(s) = n$, and that wages are the same in both cities taking common value w. Labour productivity in functions is assumed to be symmetric across cities, which we capture by denoting the labour input coefficient in each city'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 replicates Table A.1 with explicit expressions derived from this functional form. The profit functions of eqn. (1) become, $\pi_1(s) = p(s) - \{2\lambda + \Delta\lambda[1 - \gamma(1 - 2s)]\}w/2$, $\pi_F(s) = p(s) - \lambda w - t$, $\pi_2(s) = p(s) - \{2\lambda + \Delta\lambda[1 + \gamma(1 - 2s)]\}w/2$, and give explicit expressions for the mode boundaries,

$$s_1 = \frac{1}{2}\left[1 - \left(1 - \frac{2t}{\mathrm{w}\Delta\lambda}\right)\frac{1}{\gamma}\right], \quad \text{and} \quad s_2 = \frac{1}{2}\left[1 + \left(1 - \frac{2t}{\mathrm{w}\Delta\lambda}\right)\frac{1}{\gamma}\right]. \tag{4}$$

These relationships capture the way in which the sourcing of functions by firms in each sector depends on fragmentation costs *t* relative to wages, the range of function intensities γ, and inter-city 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 firms emerge, first in sectors that have similar use of both functions, i.e. *s* in an interval around ½ and of width $s_2 - s_1 = (1 - 2t/w\Delta\lambda)/\gamma$, wider the smaller is *t*, and the larger are productivity differences, $\Delta\lambda$. Intuitively, these are the sectors where both functions have a high share of costs (e.g. close to 50%), so it is worthwhile incurring fixed cost *t* to source each from the lowest cost city. Sectors with more extreme function intensities remain integrated in the city 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 firms in city 1, the most *B*-intensive are integrated in city 2, and those with intermediate

[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.

function intensities are fragmented. 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 firms is larger the smaller is γ, the parameter that measures the range of function intensities.

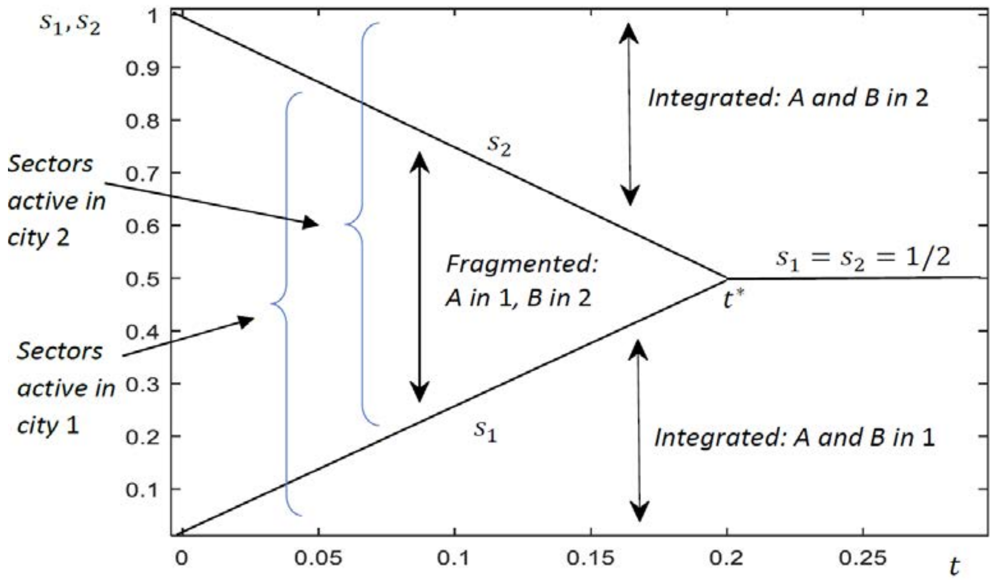


Figure 1: Sectoral mode of operation

Sectoral to functional specialisation: The preceding paragraph established where firms in each sector source their input of functions. The dual question is: what activities take place in which cities? As fragmentation costs fall below t^* so some sectors remain integrated in a single city, but others ($s \in (s_1, s_2)$ fragment, so there is a decline in the average sectoral specialisation of cities. In Figure 1 the curly brackets indicate the range of sectors with a presence in each city and evidently, once $t < t^*$, further reductions in *t* increase this range. In the empirical section we will measure this by calculating specialisation indices defined on the shares of each city's total employment in each sector, i.e. $m_{sr} = L_{sr}/\Sigma_r L_{sr}$. While cities' specialisation in sectors is falling, their specialisation in functions is increasing. Employment of each function in city *r* is L_{Ar}, L_{Br} (appendix Table A.2). Intuitively, production of each function moves into the city according to comparative advantage. In later sections of the paper we compute functional specialisation indices, based on shares of each city's total employment in each function, $m_{fr} = L_{fr}/\Sigma_r L_{fr}$. 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 city contains activity in half the sectors; each sector operates in a single city (city 1 for $s \leq 1/2$, and city 2 for $s > 1/2$).

c) Functions: If every sector uses both functions ($1 \geq \gamma$), then every function is present in each city.

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 $s \,\epsilon\, [\mathrm{s}_2, \mathrm{s}_1]$ are fragmented, operating in both cities; sectors with more extreme function intensities ($s < s_2,\ s > s_1$) are integrated, operating in a single city.

b) Sectors: Each city contains activity in more than half the sectors. If $t \leq (1-\gamma)\, w\Delta\lambda/2$ then each city contains activity from all sectors.

c) Functions: If $t \leq (1-\gamma)\, w\Delta\lambda/2$ then each city specialises in a single function, $L_{\mathrm{A1}} = L_{\mathrm{B2}} > 0,\ L_{\mathrm{A2}} = L_{\mathrm{B1}} = 0$, (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 (concentration) indices are calculated for the distribution of both sectoral and functional employment across cities. 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 differences the productivity of functions varies because of exogenous efficiency differences. There is considerable evidence of agglomeration economies in this and other sectors, so we develop a variant of the model in which these economies of scale drive productivity.

Labour input coefficients are function and city specific, and are now assumed to be based on an endogenous part deriving from productivity spillovers in the same function and city, as well as a possible Ricardian component. The Ricardian component is as before, taking values λ and $\lambda + \Delta\lambda$. Productivity spillovers generated by each function in each city are equal to output in the function-city 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

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

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

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

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

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

Thus, if s_2 is large a relatively small range of sectors undertake function *A* in city 2, thereby reducing city 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 cities 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 $\pi_F(s_1) \le \pi_1(s_1)$ 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*, $t \ge t^{**}$, there is an equilibrium with fully integrated production. This is illustrated by the solid horizontal line on Figure 2.

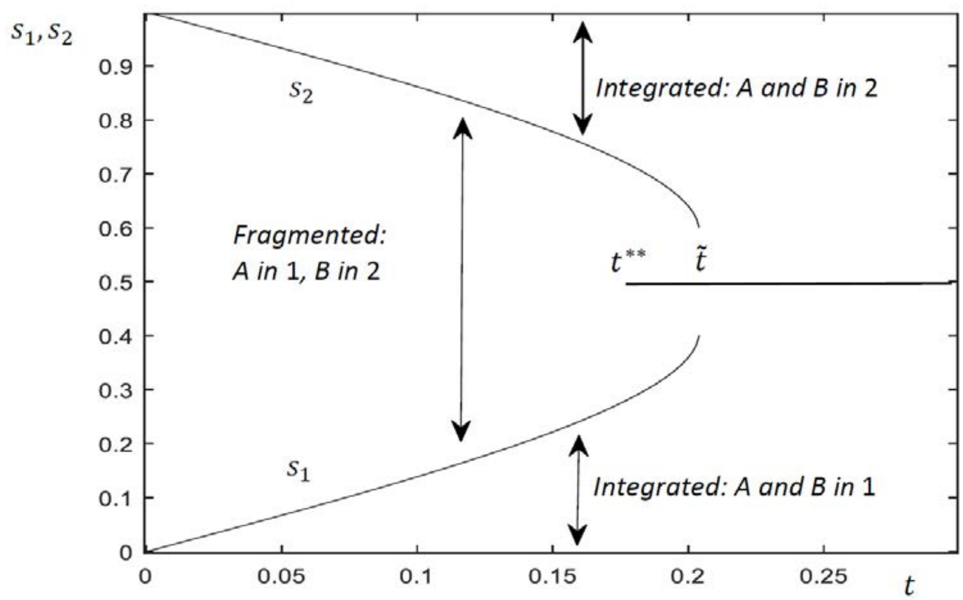


Figure 2: Sectoral mode of operation with increasing returns

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 city 1, and B in city 2; the presence of increasing returns means that the productivity differential is now large, justifying firms' 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 $n\sigma$ 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 $s \, \epsilon \, [s_2, s_1]$ are fragmented.

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 cities' 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 and, 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.

---

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

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 (number of firms) in each sector.

4. General Equilibrium

To this point, we have assumed product prices are constant, a fixed and equal number of firms in all sectors, and that wages are constant and the same in both cities. We now relax these assumptions and develop the general equilibrium of the model. In section 4.1 we model urban structure and the labour market, endogenizing wages; then in section 4.2 we look at product supply and demand, adding free entry of firms 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 City size, employment and wages

In addition to the sectors and functions modelled above we now add a hinterland region producing an 'outside good' which we use as numeraire. This good is produced using labour alone, at constant productivity giving fixed wage w_0. This and all other final goods are perfectly freely traded.

Labour is perfectly mobile, equating utilities across cities and the outside region. To give determinate city sizes and wages we use the standard urban model (the Alonso-Mills-Muth model, see for example Henderson and Thisse 2004). City workers face urban costs of commuting and land rent, costs which depend on city size. It follows that the cost of living may vary across locations, so labour mobility implies that equilibrium wages in each city, $w_1,\ w_2$, may differ from w_0 and from each other. The micro-foundations of this are that each household occupies one unit of land, all urban jobs are in the city centre (CBD), and commuting costs are c_r per unit distance. Urban costs at distance z from the CBD consist of 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, so real wages are equalised when $w_r - c_r z - h_r(z) = w_0$ for all r and at all occupied distances z. There are K spokes from the CBD, along which people live and commute, so population is $L_r = K z_r^*$, where z_r^* is the edge of the city (length of each spoke). At the city edge land rent is zero, so $w_0 = w_r - c_r z_r^* = w_r - c_r L_r / K$ giving the city-size equations

$$L_1 = (w_1 - w_0)K/c_1, \quad \text{and} \quad L_2 = (w_2 - w_0)K/c_2. \tag{8}$$

It should be noted that L_r denotes both the number of residents and the number of workers in the city. These equations simply say that larger cities have to pay higher wages in order to cover the commuting costs and rents incurred by workers. Finally, we note that rent in each city can be expressed as, $h_r(z) =$

$w_r - w_0 - c_r z = c_r(L_r/K - z)$, so integrating over z and adding over all spokes, total rent in a city of size L_r is

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

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.

4.2 Sectoral output and the number of firms

The price of output of sector s is $p(s)$ and, since each firm produces one unit of output, total supply of good s is simply $n_1(s) + n_F(s) + n_2(s)$. We have to this point held the price and number of firms constant. We now endogenize these variables by modelling demand for each sector's output and letting the number of firms adjust until profits in each sector are zero.

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

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

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.

Domestic supply response comes from free entry of firms. The zero profit conditions are complementary slack inequalities since not all modes will be active in all sectors, and we state them (from equations 1) as,

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

Recalling that firms operate at unit scale, total domestic supply in each sector s is

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

Output prices and numbers of firms adjust to clear markets. As they do so employment levels, wages, and the structure of economic activity in each city will also change. Firm types may be active or non-active in each city, 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 spillovers cases 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 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. With all firms integrated, the middle sector (there is an odd number of sectors) is produced in both countries.

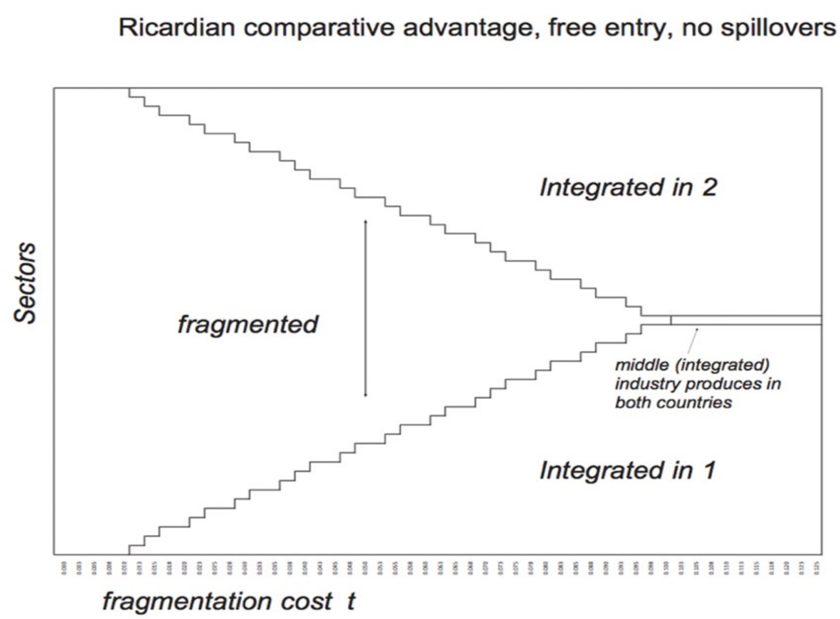


Figure 3: Symmetric Ricardian Case (fragmentation cost *t* on horizontal axes)

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 cities for each level of fragmentation costs. The concentration of sector *s* is the sum over cities *r* of the share of sector *s*'s national employment that is in *r* minus city *r*'s share of national employment, squared:[7]

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

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

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

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

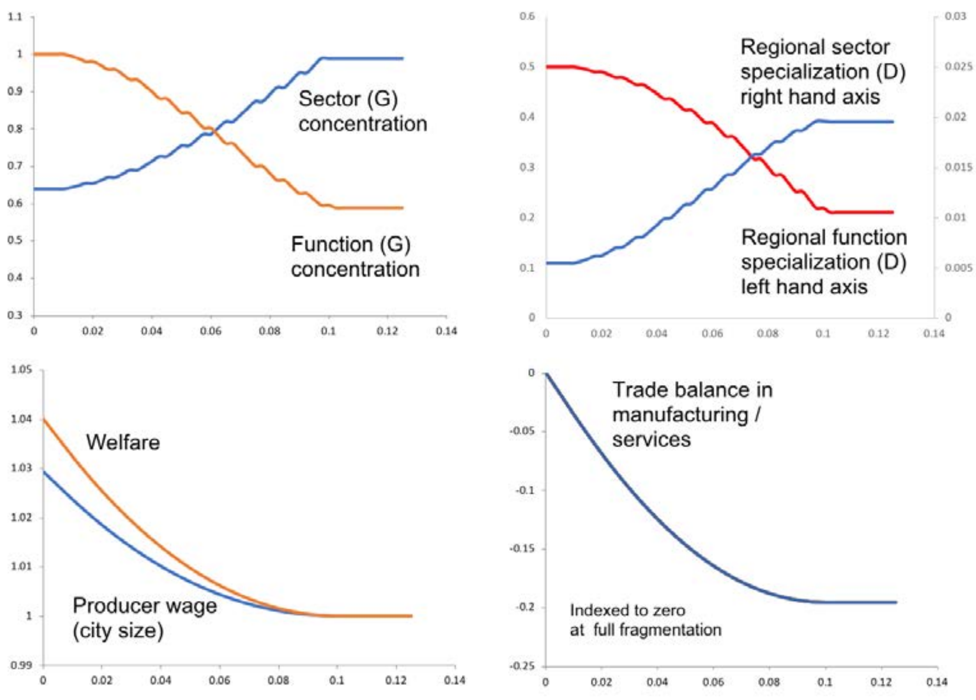


Figure 4: Symmetric Ricardian Case (fragmentation cost *t* on horizontal axes)

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

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 a specialization index D. 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 city size. The producer wage / city size curve shown in the bottom left of Figure 4 indicates that a lowering a fragmentation costs does not have a big effect on city size: increased outputs depress product prices some and so from the free-entry conditions, producer wages (city populations) don't 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

Figures A3 and A4 consider asymmetry between the sectors/cities in the Ricardian case. Figure A3 assumes that city 1 has a comparative and absolute advantage in function A, while city 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 city 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 cities. Then if sector *s* = 0.5 is just breaking even in city 2, there would be positive profits for sector *s* in city 1.

Two further results follow in the asymmetric Ricardian case. The right-hand panel of Figure A3 shows the employment levels in the two cities. Intuitively, the city with the absolute advantage (city 1) will be larger for all levels of fragmentation costs, but this difference shrinks as these costs fall. Figure A4 shows the function and sector concentration indices for the same asymmetric Ricardian case. The more productive city 1 will have lower concentration for both sectors and functions. The intuitive follows from the previous paragraph: city 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 cities do have lower levels of both forms of concentration.

Figure A5 shows a similar result for the spillovers case: here only function A has spillovers, but in both cities (in contrast to the Ricardian case where only λ_A is smaller in city 1 only). In equilibrium however, the spillovers case is similar: city 1 will have a comparative and an endogenous absolute advantage in function A, while city 2 has a comparative but not absolute advantage in function B.

These results show up as differences in city size/employment (which in turn translate into producer wages), shown in the right-hand panel of Figure A5. The city 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 city sizes (employment) were the same, then producer wages would be the same, in which case there must be positive profit opportunities in city 1 and/or losses incurred in city 2.

The convergence in city 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 city 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 city 1 workers in the A function means that less workers are required to produce those tasks at given output prices and hence city 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.[8] We

[8] 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

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 method we use to measure the geographic concentration. We then describe the main data sources. Finally, we implement the index of concentration to study the time series changes in sectoral and functional concentration.

5.1.1 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)). 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, Ellison, and Glaeser (1997, henceforth EG97) incorporate information about the size distribution of firms in the sector to construct the following index of concentration

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

impact of changes in fragmentation costs. Second, it prevents us from making quantitative predictions regarding the impact of fragmentation costs on regional outcomes.

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.[9]

The EG97 index of concentration defined in equation (15) has many useful properties.[10] 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

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

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*.[11] The intuition for the correction factor H_f, suggested by Gabe and Able (2010), is that when a function's employment is concentrated in a few industries, the measured geographic concentration of the function should be higher all else equal.

---

[9] 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.

[10] 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.

[11] 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}.

5.1.2 Data

To construct the EG97 indexes of concentration, we need information on the geographic distribution of sectoral and functional economic activity, measured throughout by employment. The two main sources from which we derive information are the BLS's Quarterly Census of Employment and Wage dataset (QCEW) and Occupational Employment Statistics (OES).

The QCEW provides sector-by-state level data, specifically 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). 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). Together, these data allow us to construct the sectoral and functional concentration indices for each year in our sample. Additional information on the datasets is provided in Appendix 4 at the end of the paper.

A difficulty we face in developing our data is the frequent reclassification of sectors and functions over time. Over the period covered by our sample, the 1997 NAICS classification used in the QCEW is revised multiple times, first in 2002, and subsequently in 2007, 2012, and 2017. Similarly, the original SOC classification introduced in 2000 was revised in 2010 and in 2018. For the analysis, we construct longitudinal region-sector and region-function datasets restricted to sectors and functions 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 and function definitions.

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.

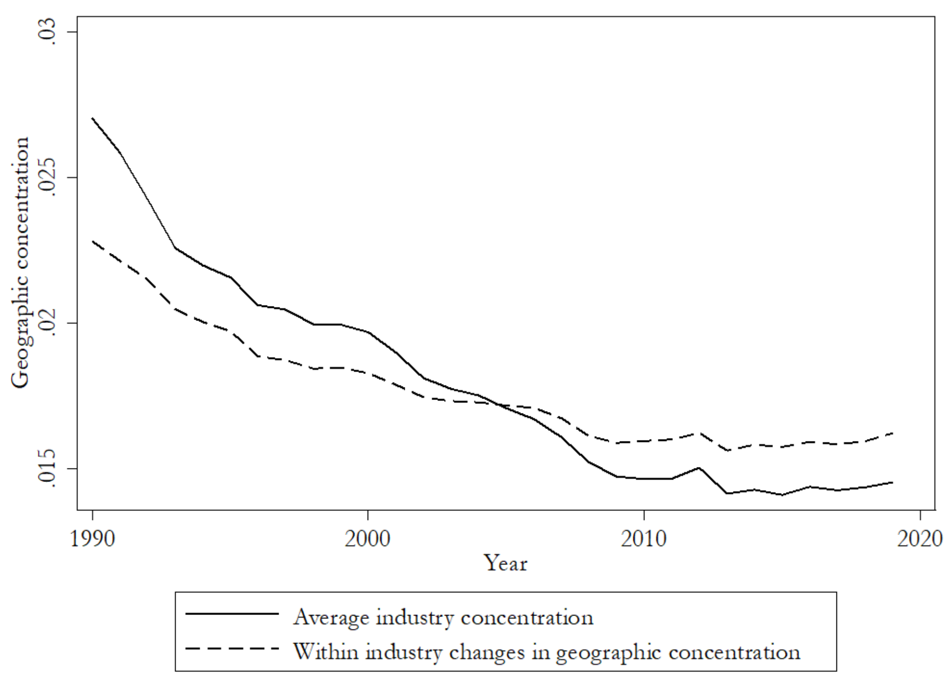


Figure 5: Geographic concentration of sectors over time

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 U.S. six-digit NAICS industries) of the Ellison-Glaeser (1997) index of geographic concentration and of two components, the raw geographic concentration and the Herfindahl measure of plant-level concentration.

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 groups: 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. Dev.*	R^2	*Nb. Obs.*
Full sample	-0.00028	0.00003	0.877	18,780
Manufacturing	-0.00029	0.00004	0.899	7,710
Business services	-0.00019	0.00006	0.745	4,320
Personal services	-0.00013	0.00006	0.849	3,810
Wholesale, retail and transportation	-0.00060	0.00008	0.856	2,910

Notes: This table reports OLS results from regressing indexes of concentration on a time trend. 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 function, even when holding the employment weights constant.

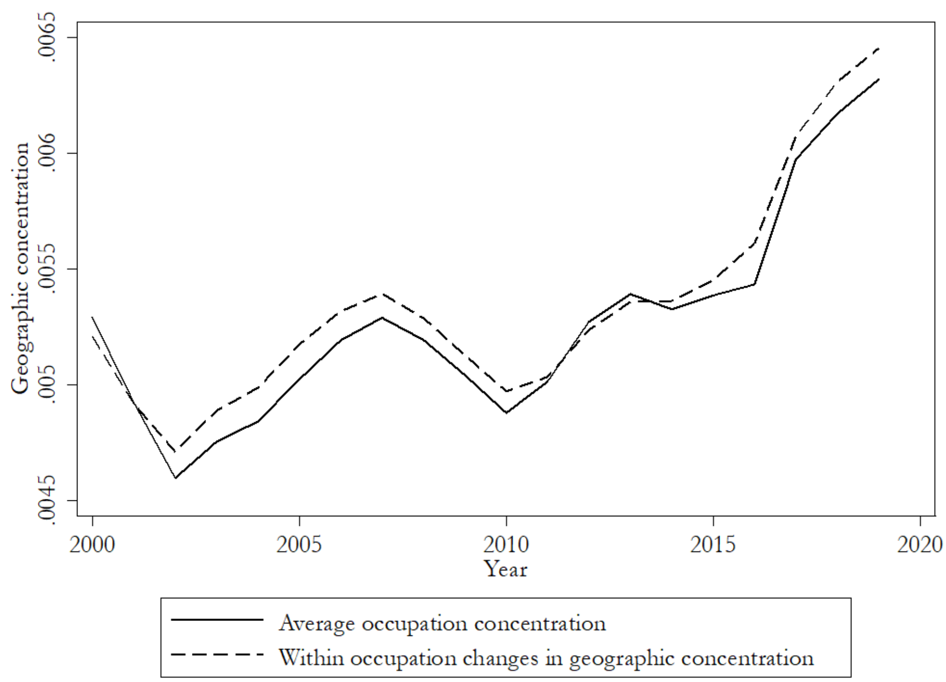


Figure 6: Geographic concentration of functions over time

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

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

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 12 of those are statistically significant at conventional levels.

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

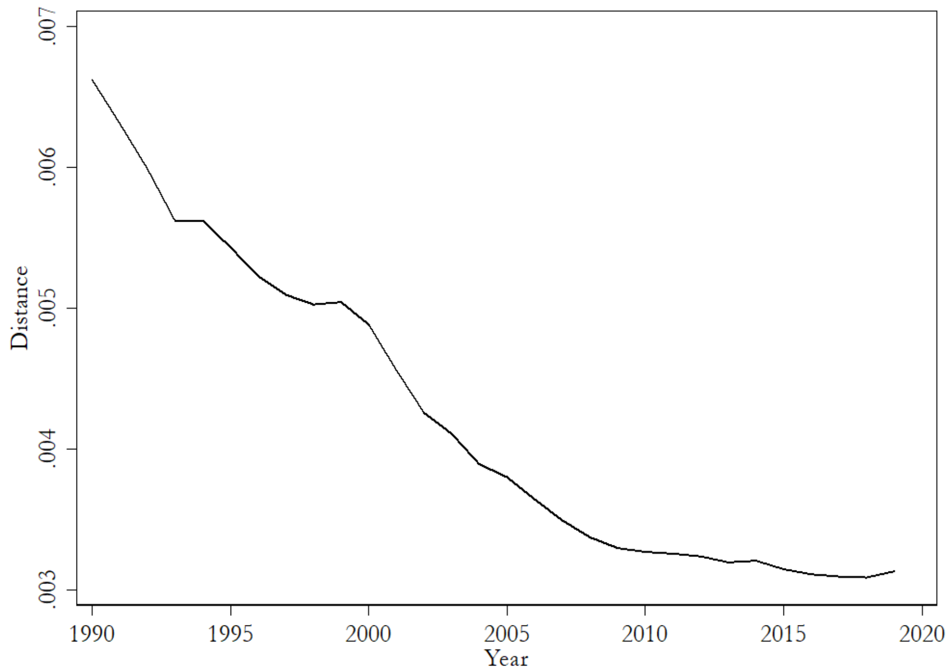
In this section, we explore the sectoral and functional structure of regional employment. Under our assumption about the time series evolution of fragmentation costs, we expect to find a decrease in sector specialization and an increase in functional specialization.

Table 4. Time series trend of functional concentration

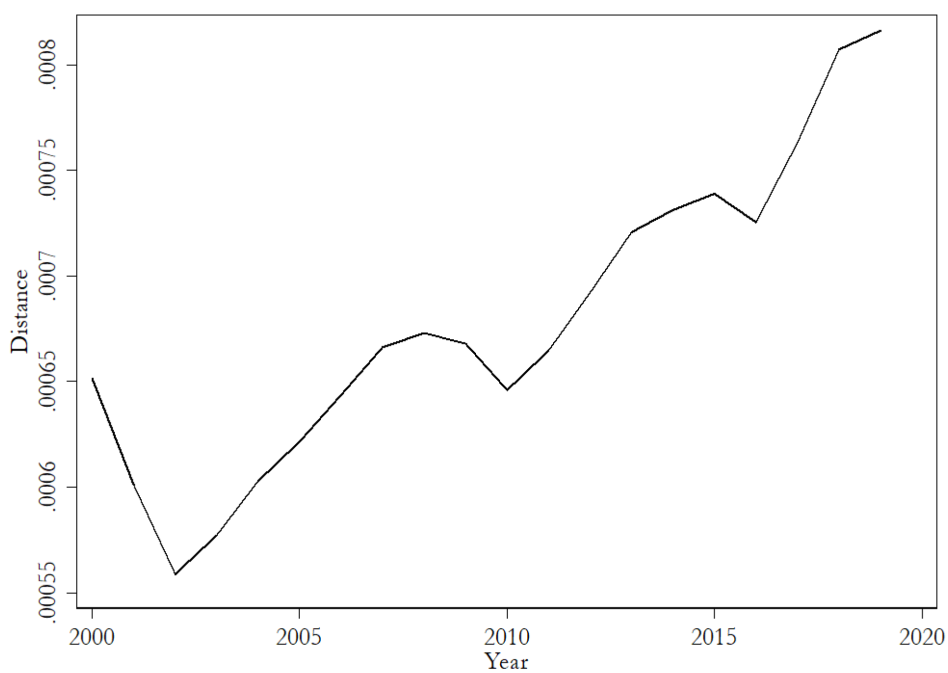
	Estimates	*Std. Dev.*	R^2	*Nb. Obs.*
Full sample	0.00019***	0.00002	0.870	14,080
Architecture and engineering	0.00018*	0.00008	0.934	640
Arts, design, entertainment, sports	0.00038***	0.00010	0.860	700
Building and ground cleaning and maintenance	-0.00011*	0.00005	0.727	160
Business and fincancial	0.00010**	0.00004	0.903	540
Community and social services	0.00015	0.00013	0.683	260
Computer and mathematical	0.00038**	0.00012	0.776	260
Construction and Extraction Occupations	0.00013	0.00010	0.890	1,100
Education, trainning, and library	0.00015***	0.00004	0.537	1,120
Food preparation and serving	0.00085***	0.00024	0.916	300
Healthcare practitioner and technicians	-0.00024***	0.00003	0.570	920
Healthcare support	0.00039***	0.00008	0.613	280
Installation, Maintenance, and Repair	-0.00020**	0.00006	0.782	980
Legal	0.00011	0.00007	0.324	160
Life, physical, and social science	0.00001	0.00011	0.811	760
Management	0.00015***	0.00003	0.845	600
Office and Administrative Support	0.00014***	0.00003	0.846	1,040
Personal care and service	0.00017	0.00010	0.831	600
Production	0.00050***	0.00005	0.859	2,000
Protective services	-0.00020	0.00010	0.785	380
Sales	0.00049*	0.00019	0.594	400
Transportation and Material Moving	0.00044***	0.00009	0.886	880

Notes: This table reports OLS results from regressing indexes of concentration on a time trend. The *, **, and *** indicate statistical significance at the 10, 5, and 1 percent level, respectively.

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.



(a) sectors



(b) Functions

Figure 7: Regional specialization over time

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. Dev.*	R^2	*Nb. Obs.*
Sectoral employment	-0.0203	0.0006	0.9342	1,500
Functional employment	0.0193	0.0011	0.8834	1,000

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

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

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

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 as expected. These results indicate that region size is a strong predictor of the cross-sectional variation in both sectoral and functional specialization.

6. Conclusions

Our paper is motivated by what is 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.

Table 6. Regional size and specialization

	Estimates	*Std. Dev.*	R^2
Panel A: Sectoral specialization			
D	-0.595	(0.020)	0.412
HHI	-0.490	(0.022)	0.297
Krugman	-0.800	(0.014)	0.735
Panel B: Functional specialization			
D	-0.642	(0.024)	0.458
HHI	-0.173	(0.028)	0.237
Krugman	-0.171	(0.028)	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. Every estimated coefficient is statistically significant at the 1 percent level.

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 city may produce with only locally sourced functions or may draw functions from other locations. 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 city's employment is concentrated in certain sectors, with each sector's employees performing many different functions. At low fragmentation costs, a city's employment is concentrated in certain functions, with employees in a certain function doing work for many different sectors. Instead of a city having production workers, managers, lawyers and accountants working in one sector, a lawyer or accountant does work for many different sectors, often at a distance.

This basic model result is in turn used to draw out a number of 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, and the correlation across regions between their degree of sectoral specialization with their degree of functions specialization.

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 cities 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.

References

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.

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.

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, 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 Ppaer No 8705.

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

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.

Limao, Nuno, and Anthony J. Venables (2002), "Geographical Disadvantage: a Heckscher-Ohlin-von Thünen Model of International Specialization", *Journal of International Economics*, 58, 239-263.

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

Markusen, James R. (2013), "Expansion of Trade at the Extensive Margin: A General Gains-from-Trade Result and Illustrative Examples", *Journal of International Economics* 88, 262-270.

Markusen, James R. and Anthony J. Venables (2007), "Interacting Factor Endowments and Trade Costs: A Multi-Country, Multi-Good Approach to Trade Theory", *Journal of International Economics* 73, 333-354.

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.

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

Appendix 1: Section 3 theory

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

	City 1	City 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/city (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/city (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 city*		
	$L_1 = L_{A1} + L_{B1} = \int_0^1 L_1(s)ds$	$L_2 = L_{A2} + L_{B2} = \int_0^1 L_2(s)ds$

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

	City 1	City 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/city (all sectors)*	
Function A	$L_{A1} = \lambda_{A1} s_2[1 + \gamma(1 - s_2)]n/2$	$L_{A2} = \lambda_{A2}\,(1 - s_2)(1 - \gamma s_2)n/2$
Function B	$L_{B1} = \lambda_{B1} s_1[1 - \gamma(1 - s_1)]n/2$	$L_{B2} = \lambda_{B2}(1 - s_1)(1 + \gamma s_1)n/2$
V	L_{sr}: *Employment in each sector/city (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 city*	
	$L_1 = L_{A1} + L_{B1} = \int_0^1 L_1(s)ds$	$L_2 = L_{A2} + L_{B2} = \int_0^1 L_2(s)ds$

Profit functions are:

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

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

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 $\gamma < 1$ 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 $\gamma > 1$ 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 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, $\Delta\Lambda$.

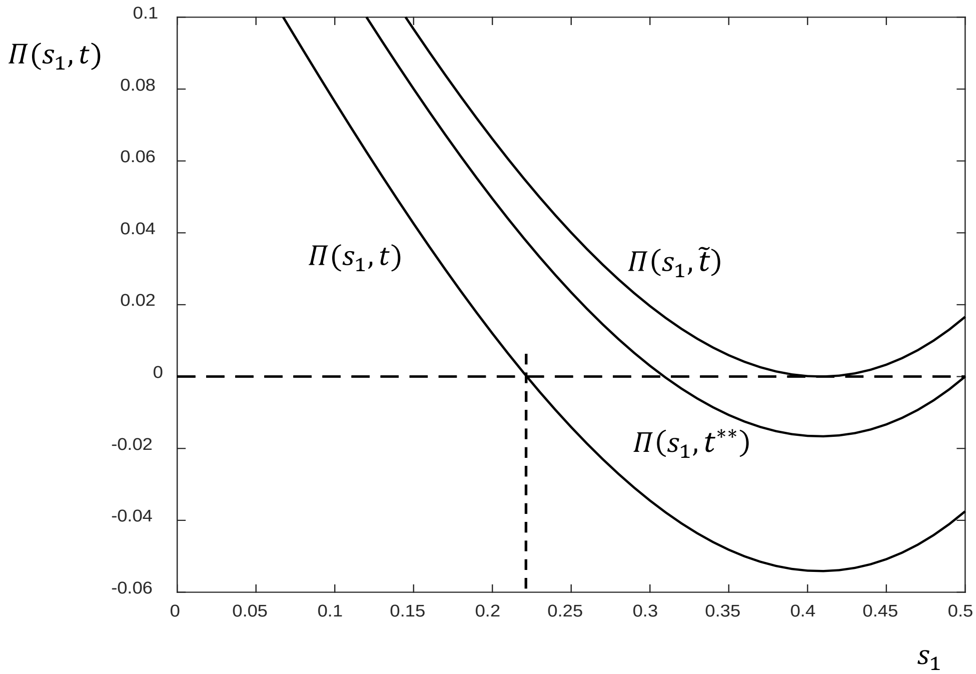


Figure A1: Expression A1 for different values of *t*

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 spend 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.[12]

[12]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.

Appendix 3: General equilibrium as a non-linear complementarity problem

Here we give the specification for the spillovers' model, which has more equations and unknowns than the Ricardian model. The latter is simpler because the lambdas are exogenous.

Non-negative variables:

L_i	labor demand or employment in city i
w_i	wages in city i
X_{ij}	output of function j in city i
λ_{ij}	labor requirements in function j in city j
$Q_d(s)$	total output of sector z (all firm types)
$Q_{fd}(s)$	domestic demand for foreign goods
$n_k(s)$	number of firms 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 cities are given as follows, where ⊥ denotes complementarity between the inequality and a variable. Labor is used in variables costs for all firm types in all sectors, plus used in fragmentation costs for fragmented sectors. We use a simple formulation of the fragmentation labor use, which divides it between the two cities, 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 cities are given by:

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

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

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

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

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

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

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

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

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

The number of active firms of each type in each sector is complementary to a zero-profit condition, that unit cost is greater than or equal to price. Fragmentation costs are: $t(w_1 + w_2)/2$.[13] Therefore

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

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

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

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

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

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

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

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

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

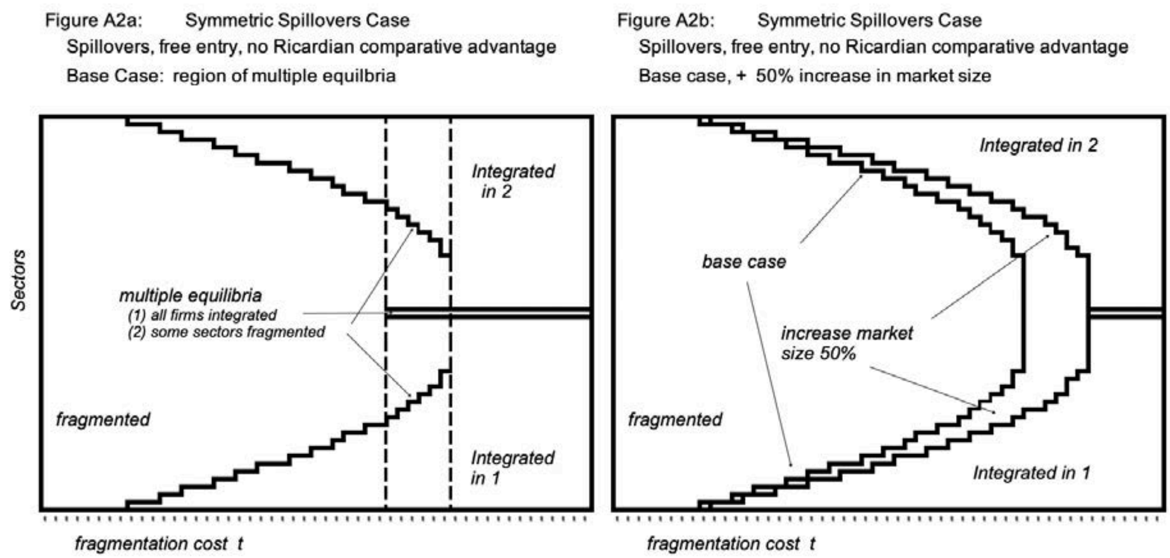


Figure A.2: Symmetric Spillovers Case

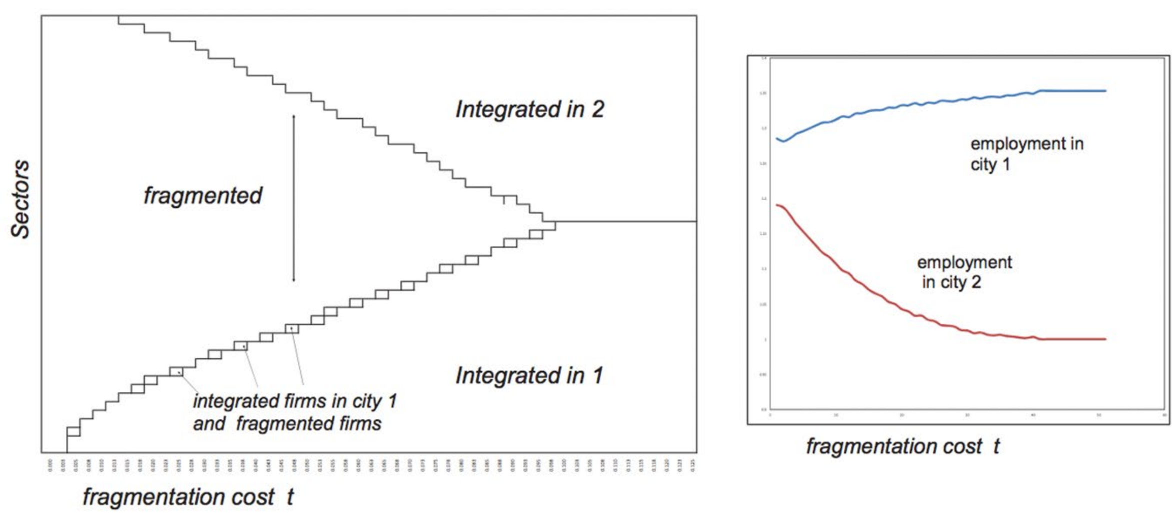


Figure A.3: Asymmetric Ricardian Case

City 1: comparative and absolute advantage in function A

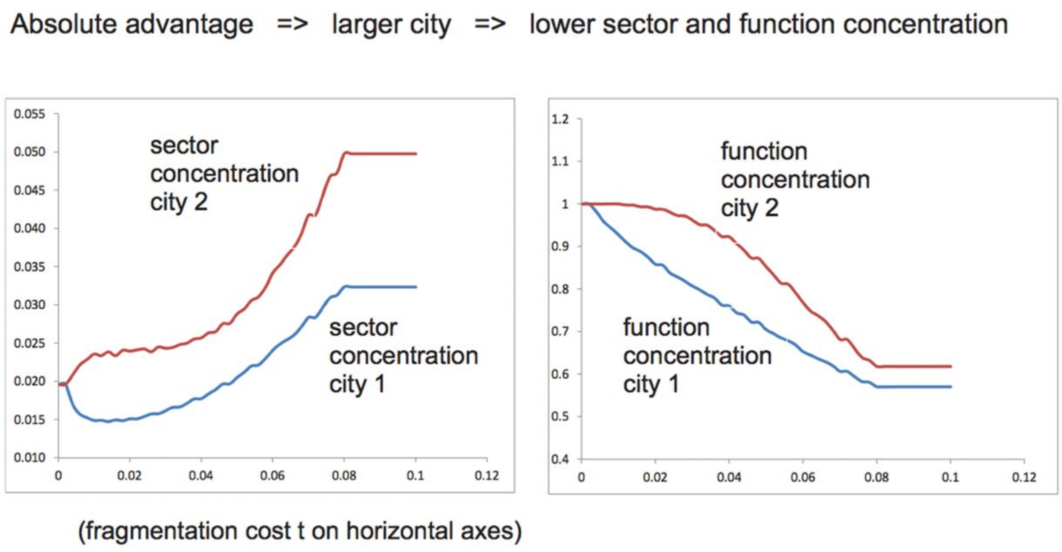


Figure A.4: Asymmetric Ricardian Case

City 1: comparative and absolute advantage in function A.

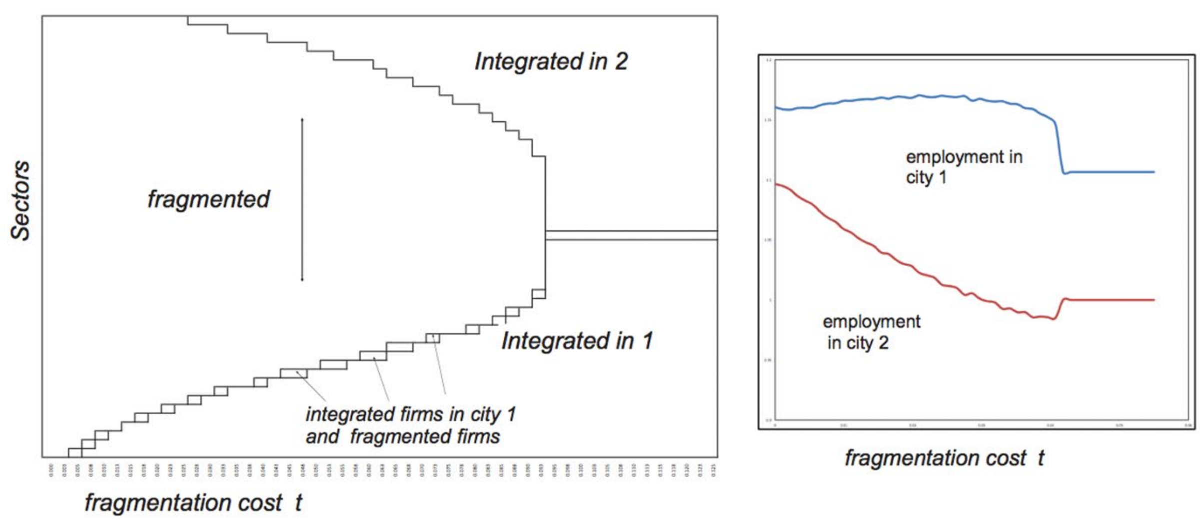


Figure A.5: Asymmetric Spillovers Case; spillovers in function A only

Appendix 4: Data

In this appendix, we provide information on the main dataset used in the empirical analysis. There are two main datasets, both from the Bureau of Labor Statistics: (i) the Quarterly Census of Employment and Wages (QCEW); (ii) the Occupational Employment Statistics (OES). We describe the main properties of each dataset in turn.

Quarterly Census of Employment and Wage

The Quarterly Census of Employment and Wages (QCEW) program publishes a quarterly count of employment reported by employers covering more than 95 percent of U.S. jobs available at the county, Metropolitan Statistical Area (MSA), state and national levels by detailed industry. Additional information on the QCEW is available online at https://www.bls.gov/cew/overview.htm.[14]

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. BLS withholds the publication of UI-covered employment and wage data for any industry level when necessary to protect the identity of employers. Totals at the industry level for the states and the nation include the undisclosed data suppressed within the detailed tables without revealing those data. Therefore, to limit the number of missing values, we use States as our geography instead of Metropolitan Statistical Area (MAS) or counties.

At the State level, the QCEW program publishes employment down to the 6-digit NAICS industry level. It also produces data on establishments and employment stratified by size of establishment for the first quarter of each year. In some case, imputed values create significant gaps in otherwise continuous levels of employment. We fill in the gaps in the data using linear interpolation (and extrapolation). About 15 percent of the observations are imputed.

For the analysis, we restrict our attention to years 1990 to 2019 (all the years using NAICS) to minimize the impact of industry reclassification on our results. We also remove industries in the "Farming" (NAICS 11), "Mining, Quarrying, and Oil and Gas Extraction" (NAICS 21), Utilities (NAICS 22), "Other Services" (NAICS 81) and "Public Administration" (NAICS92) sectors, as well as industries that contain the word "other" in their names.

[14] For our purpose, the QECW data is preferable to similar data from the Census' County Business Pattern data because it contains much less top coded entries. The correlation between the two dataset is 0.8. The "low" correlation is likely due to the imputation required to use CBP. While CBP has more data, imputation introduces a lot of noise.

Occupational Employment Statistics

The Bureau of Labor Statistics' Occupational Employment Statistics (OES) program is the only comprehensive source of regularly produced occupational employment and wage rate information for the U.S. economy. 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. Additional information on the OES can be found online at https://www.bls.gov/oes/oes_emp.htm.

Beginning in year 2000, the OES survey began using the Office of Management and Budget (OMB) Standard Occupational Classification (SOC) system. For that reason, we limit the analysis to years 2000 to 2019. As was the case for the QCEW, we fill in gaps in the data using interpolation (about 11 percent of the data is imputed) and drop occupations that contain "other" in their title.