Functional specialization, sectoral specialization, and inter-city trade

James R. Markusen
Anthony J. Venables

Abstract

The comparative advantage of cities is based on their efficiency in the production of ‘functions’, e.g. business services such as finance, law, engineering, or other inputs that are used in a wide range of sectors. Efficiency differences may be Ricardian or based on increasing returns to scale. Firms in each sector use several of these functions and choose where to produce, and whether to produce in a single city (integrated) or pay a fragmentation cost for purchasing functions from different cities. What are the implications of this for the organization of firms, the specialization and income levels of cities, and the overall structure of production and trade in the economy? We develop a simple model combining elements of several literatures including economic geography, multinational firms, urban economics, and comparative advantage trade theory. Sectoral differences in function intensities and in fragmentation costs create a distribution of fragmented and integrated firms across sectors and across cities. Lower fragmentation costs cause cities to move from sectoral specialization to functional specialization. Outcomes depend on the function-intensity of sectors, the extent of returns to scale, and the level and sectoral variation of fragmentation costs. Lower fragmentation costs may increase inequalities between cities and lead to large shifts in the production structure (and hence the external trade) of the economy as a whole.

Keywords:

JEL classification: F1, F23, R3

Authors’ Addresses:

J.R. Markusen      A.J. Venables
Department of Economics  Department of Economics
University of Colorado  Manor Road
Boulder            Oxford
Colorado           OX1 3UQ
USA                UK

j.markusen@colorado.edu  tony.venables@economics.ox.ac.uk
1. **Introduction.**

The production of final products and services typically requires a number of functions to be performed. Manufactured goods require engineering, finance and marketing; construction requires architects and lawyers, and so on. There may be spatial (e.g. inter-urban) 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 the final goods that use these functions. This paper 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 HQ and production, as developed in some of the literature on foreign direct investment (Markusen 2002) and more recent work in the urban context (Duranton and Puga 2005, Rossi-Hansburg et al. 2009). A much finer level is that of a ‘task’, often thought of as a narrow stage of production and modelled as a continuum (Grossman and Rossi-Hansburg 2008, 2012, Autor 2013). We have in mind an intermediate level such as engineering (of distinct sorts), finance, legal services, advertising. The sort of function we seek to model has several properties. First, many functions are ubiquitous; virtually all final sectors require finance, law, and design, albeit in different proportions. Second, many large cities appear to have developed quite broad 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. Some cities specialise in quite narrow ‘tasks’ (Qiatou produces 200,000 km of zippers per year) but, at least for large cities, the broader functional concepts seem more relevant. Third, we think of functions as being associated with labour skills and firm capabilities, and suggest that these 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.

Differences in the efficiency with which cities or regions supply functions pose a number of questions. The first is; if firms in a particular sector use many functions, how do they access these functions? We develop a simple model to show how the interaction between fragmentation costs (the cost of dividing production between locations), 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. The second question concerns the implication of these choices 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 final product (sectoral) specialisation of cities?
Following from this, two further questions arise. As fragmentation costs and the associated production structure of cities change, what happens to relative city size and the real incomes of each city? Real income gains are particularly large with increasing returns, as fragmentation of firms allows cities to develop their functional specialisms. Changes in the production structure of each city may also change the production structure of the economy as a whole. How do changes in the costs of trading functions within an economy shape the external trade of the country?

In order to investigate these questions we develop a model that has elements of economic geography, the multinationals’ literature with vertical firms, urban economics, and external economies of scale with some novel twists. There are 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, each final product 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. Firms in each sector may perform both functions in one location, referred to as integrated firms, or one function in each location, referred to as fragmented firms. However, 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 there are Ricardian differences in the productivity of functions between cities. This provides a very clean example of how reducing fragmentation costs causes firms (in some sectors) to fragment, 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 comparative advantage is, we think, more likely to arise endogenously from cities’ acquired skills and consequent increasing returns to scale. Economies of scale are, we assume, external to the firm and sector, occurring at the city-function level. In this case city specialisation turns out to be a discontinuous function of fragmentation costs. This is because economies of scale large enough to overcome fragmentation costs are achieved only if a wide range of sectors fragment. Furthermore, there is a range of fragmentation costs at which there are multiple equilibria. If firms in all sectors are fully integrated neither city has a large enough comparative advantage to induce fragmentation; but if firms are fragmented then cities are functionally specialised, creating the scale and productivity differences that support
Welfare gains from reductions in fragmentation costs are then particularly large, providing they induce spatial reorganisation and the move from sectoral to functional specialisation.

The questions we pose and the model we develop touches on many strands of international and urban economics. The division of firms’ activities (at least, HQ and production) has been studied in the literature on foreign direct investment (see Markusen 2002). Perhaps closest in spirit to this paper is the urban model of Duranton and Puga (2005), the focus of which is precisely the move from sectoral to functional specialisation, although again in the context of the division of HQ and production.¹ The international trade literature has analysed trade in tasks, both in constant returns models (Grossman and Rossi-Hansburg 2008) and under increasing returns (Grossman and Rossi-Hansburg 2012). As well as its international focus, this literature works in a framework of many tasks and few final sectors; it does not capture the ubiquity feature which, we argued above, distinguishes functions from tasks. The present paper also draws on economic geography modelling (Fujita et al. 1999), particularly in its analysis of the multiplicity of equilibria occurring at intermediate levels of spatial frictions. Finally, there are literatures on the impact of internal geography on external trade. Uneven distribution of factors of production within a country is studied by Courant and Deardorff (1992) and following literature (Brakman and van Marrewijk 2013), and the physical geography of proximity to ports is studied by Limão and Venables (2002). In the present paper the internal geography arises from city variation in efficiency in the production of functions.

The contribution of the present paper is twofold. We provide a relatively simple and integrated framework to unify several themes in the literature, although some of the simplicity comes from the use of specific functional forms. We use the framework to show how lower fragmentation costs can (depending on technologies) bring about the transition from integrated to fragmented firms and sectoral to functional specialisation. It may cause divergence of city structure, size and income, and can cause large changes in the economies overall production structure and pattern of external trade.

2. The model

We build the model in stages, starting with a description of sectors, firms, and functions. In section 3 we use this to draw out results on fragmentation and specialisation, whilst keeping the general equilibrium side of the model in the background; we are able to do this by making sufficient assumptions to ensure that the two cities are symmetric, with the same wages.

¹ See Rossi-Hansburg et al. (2009) for intra-urban separation of HQ and production.
Section 4 then adds the general equilibrium side of the model enabling analysis of a richer set of possibilities.

There is a continuum of sectors indexed \( x \in [0, 1] \), each producing output with price \( p_x \). We assume that that there are just two functions, \( A \) and \( B \). Each sector contains a number of firms each of which employs 2 units of labour, of whom \( a(x) \) are employed on function \( A \) and \( 2 - a(x) \) on \( B \). Internal returns to scale are constant, so setting firm scale at 2 is without loss of generality. The share of each function varies across sectors, and we rank sectors such that low \( x \) sectors are \( A \)-intensive, \( a'(x) < 0 \).

There are two locations (cities) and the wage rate in city \( i \) is \( w_i, i = 1, 2 \). The prices of final output, \( p_x \), are the same in both. The productivity of labour in producing function \( j \) in city \( i \) is \( Q_{jk} \), \((j = A, B, k = 1, 2)\). There are no sectoral productivity differences between cities, so urban comparative advantage is determined entirely at the function, not the sector level. We look at cases where these productivity differences are Ricardian, and where they are endogenous due to increasing returns. Cities are labelled such that productivity differences (if any) give city 1 a comparative advantage in function \( A \) \((Q_{A1}/Q_{B1} \geq Q_{A2}/Q_{B2})\). Since low \( x \) sectors are \( A \)-intensive, city 1 will be attractive (other things equal) for firms in low \( x \) sectors, and city 2 attractive for high \( x \) sectors.

Firms can source functions from either city, but if the two functions come from different cities then a fragmentation cost is incurred. (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). Each firm can therefore operate in one of three modes, choosing to operate entirely in city 1, entirely in 2, or to locate one function in city 1 and the other in city 2. Firms that produce in a single city are ‘integrated’ (and will be labelled subscript \( i = 1, 2 \) according to city of operation); those operating in both are ‘fragmented’ (subscript \( F \)). Fragmented firms incur additional cost, \( t p_x \), of operating in two locations.

The profits of a firm in sector \( x \) for each of the three production modes are

\[
\pi_1(x) = p_x [a(x)Q_{A1} + (2 - a(x))Q_{B1}] - 2w_1,
\]

\[
\pi_F(x) = p_x [a(x)Q_{A1} + (2 - a(x))Q_{B2}] - a(x)w_1 - (2 - a(x))w_2 - t p_x,
\]

\[
\pi_2(x) = p_x [a(x)Q_{A2} + (2 - a(x))Q_{B2}] - 2w_2.
\]

\[\text{2 The assignment of which function to which city will become clear, and does not merit additional notation.}\]

\[\text{3 This is ‘iceberg’, so measured in terms of the price of output, } p_x \]
The term in square brackets is output. Thus, in sector \( x \) function \( A \) uses \( a(x) \) workers and function \( B \) uses \( 2 - a(x) \), with productivities given by the \( Q_{jk} \), depending on the city \( (k = 1, 2) \) in which the sector performs the function \( (j = A, B) \). We assume that outputs from functions \( A \) and \( B \) are additive, although a production function with a lower elasticity of substitution between functions is possible.\(^1\) Wage costs depend on where the functions are performed, and hence on the sectoral function intensity and the chosen mode.

The most profitable mode for firms in a particular sector depends on the sector’s function intensity, \( a(x) \), and sectors partition into three groups. First is a range of \( x \) in which firms are integrated and produce both functions in city 1; as we show below, these will be low \( x \) sectors, intensive in function \( A \). Second is a range of sectors in which firms are fragmented producing function \( A \) in city 1 and function \( B \) in city 2; if such sectors exist they will be those with intermediate values of \( x \) (i.e. using both functions in similar proportions). Third are high \( x \) (\( B \)-intensive) sectors in which firms are integrated and operate only in city 2.

The boundaries between these ranges are denoted \( x_1, x_2 \), and are the sectors for which different modes of operation are equi-profitable, i.e. \( \pi_1(x_1) = \pi_F(x_1), \pi_F(x_2) = \pi_2(x_2) \). Using (1), these mode-boundaries are implicitly defined by

\[
\begin{align*}
\pi_F(x_1) - \pi_1(x_1) &= (2 - a(x_1))\left[p_x(Q_{B2} - Q_{B1}) + w_1 - w_2\right] - tp_x = 0, \\
\pi_F(x_2) - \pi_2(x_2) &= a(x_2)\left[p_x(Q_{A1} - Q_{A2}) + w_2 - w_1\right] - tp_x = 0.
\end{align*}
\]

The relationship between sectors, functions, and chosen modes of production is illustrated on figure 1, where the horizontal axis is the range of sectors, \( x \in [0, 1] \), and the vertical is employment per firm, with line \( a(x) \) indicating the function intensity of each sector, so low \( x \) sectors are \( A \)-intensive.\(^5\) The inequalities at the bottom indicate the relative profitability of operating each mode, with mode-boundaries \( x_1, x_2 \), indicated by the vertical dashed lines.

The shaded area is total employment in city 1 if each sector contains just one firm; this consists of functions \( A \) and \( B \) for integrated low \( x \) firms, plus function \( A \) for fragmented firms.

This framework of sectors, functions, and firms provides the basis for investigating patterns of firm organisation and urban specialisation.

---

\(^1\) This is the economics department production function. You need someone in every field, but output is simply the sum of papers produced.

\(^5\) This is illustrated as \( a(x) = 1 + \gamma(1/2 - x) \), an example that will be developed more fully in the following sub-section: the equations in the figure give employment levels for this example and can be ignored for the moment.
We start by looking at how firms’ mode of operation and the consequent location of sectors and functions depend on technology and fragmentation costs. Throughout this section we make a number of strong assumptions which make cities and sectors symmetrical. We assume that the function intensity of sectors is linear in $x$ and symmetric (the middle sector, $x = \frac{1}{2}$, is equally intensive in $A$ and $B$) so, as in figure 1, $a(x) = 1 + \gamma(1/2 - x)$; parameter $\gamma$ measures the heterogeneity of function intensities across sectors. Productivity differences are described below, and will be constructed to be symmetrical (so city 1’s productivity advantage in $A$ will be equal to city 2’s advantage in $B$). Together with assumptions of symmetry of cities (developed explicitly in section 4), these conditions imply equality of wages in each city, $w_1 = w_2$.

3.1: Ricardian functional advantage.

Ricardian productivity differentials are captured by an exogenous element of $Q_{jk}$, denoted $q_{jk}$, $j = A, B, k = 1, 2$. We assume that city 1 has a productivity advantage in function $A$ and city 2 has an equal advantage in function $B$, so define $\Delta q \equiv q_{A1} - q_{A2} = q_{B2} - q_{B1} > 0$. This supports full symmetry of cities and functions so equilibrium will have $x_1 = 1 - x_2$, i.e. the mode-boundaries are equi-distant above and below the mid-sector, $x_2 - \frac{1}{2} = \frac{1}{2} - x_1$ (see figure 1). Using these facts explicit values for the mode-boundaries, $x_1, x_2$, come from eqns. (2),
\( x_1 = \frac{1}{2} - \frac{1}{\gamma} \left[ 1 - \frac{t}{\Delta q} \right], \quad x_2 = \frac{1}{2} + \frac{1}{\gamma} \left[ 1 - \frac{t}{\Delta q} \right]. \)  

(3)

It is now straightforward to show how both firm organisation and the pattern of trade and specialisation depend on fragmentation costs and functional comparative advantage. If \( t = \Delta q \) then \( x_1 = x_2 = \frac{1}{2} \); this is the highest value of \( t \) at which any sector fragments, and we denote it \( t^* \) and refer to it as the break-point. For \( t \geq t^* \) all firms are integrated and sectors are equally divided between cities; city 1 has sectors \( x < \frac{1}{2} \), i.e. sectors intensive in function \( A \), and city 2 has sectors \( x > \frac{1}{2} \). Fragmented firms come to exist if \( t < t^* \), and the range of industries that are fragmented is given by \( x_2 - x_1 = \frac{2}{\gamma} \left[ 1 - \frac{t}{\Delta q} \right] \), larger the smaller is \( t \). This is illustrated on figure 2, 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 function intensities are fragmented.

**Figure 2: From integration to fragmentation.**

The figure is constructed with \( \gamma = 2 \) and \( \Delta q = 0.2 \). The break-point value \( t^* \) is proportional to \( \Delta q \) and, for a given value of \( t/\Delta q \) the range of fragmented firms is larger the smaller is \( \gamma \), the parameter that measures the range of function intensities.\(^6\)

\(^6\) The figure has \( \gamma = 2 \), this being the special case in which all sectors are fragmented (\( x_1 = 0 \) and \( x_2 = \))
City specialisation can be analysed by looking at employment in each function/city pair. In general terms, employment levels are given by

\[ L_{AI} = \int_0^x [a(x)(n_1(x) + n_F(x))] \, dx, \quad L_{B1} = \int_0^x (2 - a(x))n_1(x) \, dx, \tag{4} \]

\[ L_{A2} = \int_0^x a(x)n_2(x) \, dx, \quad L_{B2} = \int_0^x [(2 - a(x))(n_2(x) + n_F(x))] \, dx. \]

In these expressions the number of firms of each mode in sector \( x \) is denoted \( n_i(x) \geq 0, \quad i = 1, 2, F \). However, as we have seen, only one mode is active in each range of \( x \), so \( n_1(x) > 0 \) only if \( x < x_1 \); \( n_F(x) > 0 \) only if \( x_1 < x < x_2 \); and \( n_2(x) > 0 \) only if \( x > x_2 \). Employment in each function in each city is given by integrating over sectors and total employment in each city is

\[ L_1 = L_{AI} + L_{B1}, \quad L_2 = L_{A2} + L_{B2}. \]

We suppose for the remainder of this section that the number of firms in each sector is fixed and equal (and set at unity), that is \( n_1(x) + n_F(x) + n_2(x) = 1 \) for all \( x \). (Results for free entry cases are contained in section 4). With this assumption employment in each function in each city is (from eqn. (4) or calculated directly from figure 1).

\[ L_{AI} = x_2 \{1 + (1-x_2)\gamma/2\}, \quad L_{B1} = x_1 \{1 - (1-x_1)\gamma/2\}, \quad L_{AI} + L_{B1} = 1 \tag{5} \]

\[ L_{A2} = (1-x_2) \{1-x_2\gamma/2\}, \quad L_{B2} = (1-x_1) \{1+x_1\gamma/2\}, \quad L_{A2} + L_{B2} = 1. \]

Notice that, with one firm in each sector, total employment is 2 and symmetry implies that employment in each city is 1.

Given the dependence of mode-boundaries \( \{x_1, x_2\} \) on \( t \), and the dependence of function/city employment on \( \{x_1, x_2\} \), we can now how reducing fragmentation costs changes the pattern of activity in the economy. Figure 3 has the same structure to figure 1, and we use it to consider the effect of a reduction in \( t \). As on figure 1, the shaded area is total employment in city 1 some of it employed in task \( A \), and some in task \( B \).

Sectoral diversification can be measured by the range of sectors in which each city is active. Thus, city 1 is active in sectors \( [0, x_2] \), and city 2 in \( [x_1, 1] \). At \( t > t^* \) all firms are integrated, \( x_1 = x_2 = \frac{1}{2} \), so each sector operates in just one city. Evidently, diversification increases (sectoral specialisation decreases) as \( t \) falls.

---

1) \( t = 0 \). If sectors are more similar in function intensity, \( \gamma < 2 \), then all sectors become fragmented at some positive value of \( t \); if \( \gamma > 2 \) then extreme sectors use only one function (see figure 1).

7 These equations hold only for \( \gamma \leq 2 \), which we assume from now on. If \( \gamma > 2 \) then some sectors use only one task and integrals would have to take this truncation into account.
Functional specialisation can be measured by the share of each city’s labour employed in each function, so city 1’s specialisation is simply $L_{A1}$. This takes value $\frac{1}{2} \left[ 1 + \frac{\gamma}{4} \right] \in [0.5, 0.75]$ if $x_2 = 1/2$ and is decreasing in $t$. In terms of figure 3, a reduction in $t$ reallocates city 1 labour as indicated by the arrow. City 1 labour that was being used in supplying task $B$ to integrated firms is now used to supply task $A$ to fragmented firms. City 2 is, of course, the mirror image. We summarise these results of in proposition 1.

Figure 3: From sectoral to functional specialisation.

Proposition 1:

i) If $t > t^* = \Delta q$ then $x_2 = x_1 = 1/2$:
   a) Mode: All firms are integrated.
   b) Sectoral specialisation: Each sector operates in a single city.
   c) Functional specialisation: Fraction $\frac{1}{2} \left[ 1 + \frac{\gamma}{4} \right] \in [0.5, 0.75]$ of each city’s labour force is employed in the function in which it has comparative advantage.

---

8 From eqn. (5) $\frac{dL_{A1}}{dt} = \frac{dx_2}{dt} \frac{dL_{A1}}{dx_2} = -\frac{1}{\gamma \Delta q} \left[ 1 + \frac{\gamma}{2} (1 - 2x_2) \right]$ which is negative for $x_2 < 1$ and $\gamma \leq 2$. 

9
ii) If \( t \leq t^* = \Delta q \) then \( x_2 - x_1 = \frac{2}{\gamma} \left[ 1 - \frac{t}{\Delta q} \right] \): reductions in \( t \) bring,

a) Mode: An increase in the range of sectors \( x_2 - x_1 \) in which firms are fragmented.

b) Sectoral specialisation: An increase in the range of sectors, \([0, x_2]\) to which city 1 contributes at least one function (similarly \([x_1, 1]\) for city 2).

c) Functional specialisation: An increase in the share of each city’s labour force employed in the function in which the city has comparative advantage.

### 3.2. External economies of scale.

Ricardian efficiency differences might be due differences in cities’ history or physical geography but, critically, are exogenous. We now turn to the case in which productivity is endogenous, determined by the scale of activity of each function in each city. Given the substantial evidence base on the presence of urban agglomeration economies this case is empirically relevant. It is also more complex although, since economies of scale are assumed to be external to the firm, we can keep the description of firms simple, as above.

Labour productivity is function and city specific, based on an exogenous Ricardian component, \( q_{A1}, q_{B1}, q_{A2}, q_{B2} \), and an endogenous part deriving from spillovers in the same function and city. These spillovers are denoted, \( s_{A1}, s_{B1}, s_{A2}, s_{B2} \), with parameter \( \sigma_A, \sigma_B \) measuring the impact of spillovers in each function. These two determinants of productivity are additive, so output per worker is:

\[
Q_{A1} = q_{A1} + \sigma_A s_{A1}, \quad Q_{A2} = q_{A2} + \sigma_A s_{A2},
\]
\[
Q_{B1} = q_{B1} + \sigma_B s_{B1}, \quad Q_{B2} = q_{B2} + \sigma_B s_{B2}.
\]

We assume that the spillovers generated by each function in each city are equal to employment in the function/city pair, so \( s_{jk} = L_{jk} \rightarrow j = A, B, \ k = 1, 2 \). Hence, productivity differentials are (using eqns. (5) and continuing to assume a single firm in each sector)

\[
Q_{A1} - Q_{A2} = q_{A1} - q_{A2} + \sigma_A [2x_2 - 1 + x_2 (1 - x_2)\gamma],
\]
\[
Q_{B2} - Q_{B1} = q_{B2} - q_{B1} + \sigma_B [1 - 2x_1 + x_1 (1 - x_1)\gamma].
\]

Thus, if \( x_2 \) is large a wide range of sectors undertake function \( A \) in city 1 (some with integrated firms, some fragmented), this raising city 1’s productivity in function \( A \). If these spillovers are equally powerful in both functions, \( \sigma = \sigma_A = \sigma_B > 0 \), and there is full symmetry so \( x_1 = 1 - x_2 \) and \( w_1 = w_2 \), then equations (2) which define the mode-boundaries become,
\[
\pi_F(x_i) - \pi_I(x_i) = p_i \left[ 1 + \gamma (x_i - 1/2) \right] \left[ \Delta q + \sigma (1 - 2x_i + x_i(1 - x_i)\gamma) \right] - tp_x = 0, \quad (8)
\]
\[
\pi_F(x_2) - \pi_I(x_2) = p_x \left[ 1 + \gamma (1/2 - x_2) \right] \left[ \Delta q + \sigma (2x_2 - 1 + x_2(1 - x_2)\gamma) \right] - tp_x = 0.
\]

These are a pair of cubic equations, but the break-point \( t^* \), i.e. the highest value of \( t \) at which integration becomes profitable, can be easily found. Fragmentation occurs first in the sectors with the most diversified function intensities, \( x = 1/2 \), so evaluating (8) at \( x_1 = x_2 = 1/2 \), \( t^* \) is given by
\[
\pi_F(1/2) - \pi_I(1/2) = p_x \left[ \Delta q + \sigma \gamma / 4 - t^* \right] = 0. \quad (9)
\]

This reduces to the Ricardian case if \( \sigma = 0 \), while \( \sigma > 0 \) implies a strictly higher break point \( t^* \).

Some possibilities are illustrated on figure 4. The inner pair of lines gives values \( \{x_1, x_2\} \) with increasing returns but no Ricardian productivity differences \( (\gamma = 2, \sigma = 0.5, \Delta q = 0) \). The break point is at \( t = t^* = \Delta q + \sigma \gamma / 4 \) \( (= 0.25) \), at which point there is discontinuous jump, with a finite range of sectors becoming fragmented. The reason is that there is an interval of \( t \) above \( t^* \) in which there are multiple equilibria, one with all sectors integrated, and another with a range of sectors fragmented. (In the terminology of Fujita et al. (1999) there is a ‘sustain point’ to the right of the break point). The outer line gives \( \{x_2, x_1\} \) when there are increasing returns plus the same Ricardian productivity differences as illustrated in figure 2, \( (\gamma = 2, \sigma = 0.5, \Delta q = 0.2) \). The additional source of productivity difference means that \( t^* \) is higher, although in this case there is no discontinuity. Notice that this case has the same parameter values as figure 2, except for the addition of increasing returns, so it is apparent that, at all values of \( t \) below the break point, a wider range of sectors is fragmented.

For our simple example, the possibilities in figure 4 can be established analytically and are summarised in proposition 2.
Proposition 2:

i) Fragmentation of some range of industries, $x_2 - x_1 > 0$, occurs for $t \leq t^* = \Delta q + \sigma \gamma / 4$.

ii) Increasing returns ($\sigma > 0$) means that the range of industries that are fragmented is wider, at each $t$ and for each $\Delta q$, than if $\sigma = 0$.

iii) If $\gamma \Delta q + \sigma \gamma^2 / 4 - 2\sigma < 0$, then there is a discontinuity at $t = t^*$.

Part (ii) comes from noting that the Ricardian mode-boundary lies strictly inside the set of fragmented industries with $\sigma > 0$; formally, $\pi_F(x_i) - \pi_i(x_i) > 0$ at the Ricardian solution, where $x_1$ is given by eqn. (3). Part (iii) comes from the gradient of $\pi_F(x_i) - \pi_i(x_i) > 0$ in the neighbourhood of $t = t^*$ and $x_1 = \frac{1}{2}$. This is $\left. \frac{d[\pi_F(x_i) - \pi_i(x_i)]}{dx_i} \right|_{x_i = 1/2} = p_i \left[ \gamma \Delta q + \sigma \gamma^2 / 4 - 2\sigma \right]$.

If this is negative then $\pi_F(x_i) - \pi_i(x_i) > 0$ for some interval of $x_1 < \frac{1}{2}$; there is therefore an interval in which fragmentation of some range of industries – if it occurs – is an equilibrium. If there are no Ricardian comparative advantage, $\Delta q = 0$, then this is necessarily the case.

A further comment concerns the welfare effects of reductions in fragmentation costs. These come through three mechanisms. The first is the direct effect of the change in fragmentation costs – lowering them for fragmented firms, but also inducing more sectors to fragment. The second is the allocative effect of sourcing functions in line with efficiency differences. The
third are extra efficiency gains coming from increasing returns to scale. In the Ricardian case just the first two operate, and the net effect is to raise welfare. With increasing returns there is the additional benefit of higher productivity as cities develop their functional specialism.

4. Wages, prices, and industry scale

To this point we have assumed that product prices are constant, that there is a fixed and equal number of firms in all sectors, and that there is sufficient symmetry for wages to be the same in both cities. We now lay out the general equilibrium framework needed to move beyond these cases, starting with the standard urban model to give the relationship between urban employment and wages.

4.1 City structure and wages.

We use the simplest form of the standard urban model (the Alonso-Mills-Muth model, Henderson and Thisse 2004) in which there is perfect labour mobility between cities and other areas. Non-urban locations have exogenously given wage \( w_0 \) (fixed by constant returns production of a perfectly tradable good which we use as numeraire) and land rent of zero (a normalization). Urban rent and commuting costs mean that the cost of living may vary across locations, in which case labour mobility implies that the equilibrium wages paid by producers in each city, \( w_1, w_2, \) may differ from \( w_0 \) and from each other. Within each city a household occupies one unit of land and the rent in city \( i \) at distance \( z \) from the centre is \( r_i(z) \). All urban jobs are in the city centre (CBD), and commuting costs are \( c_i \) per unit distance. Workers choose residential location within and between cities so (since final goods prices are the same everywhere), real wages are equalised when \( \frac{w_i - c_i z - r_i(z)}{z} = w_0 \) for all \( i, z \).

In a linear city in which there are \( K \) spokes from the CBD, along which people live and commuting takes place, population is \( L_i = K z_i^* \), where \( z_i^* \) is the edge of the city (length of each spoke). At the city edge land rent must be zero, so \( w_0 = w_i - c_i z_i^* = w_i - c_i L_i / K \) giving the city-size equations

\[
L_1 = (w_1 - w_0) K / c_1, \quad L_2 = (w_2 - w_0) K / c_2.
\]

These equations simply say that larger cities have to pay higher wages in order to cover the commuting costs and rents incurred by workers. It should be noted that \( L_i \) denotes both the number of workers in the tradable activities that we study and the number of workers resident in the city. This may include workers in non-tradable sectors; if the number of these workers is in fixed proportion to \( L_i \) then they are subsumed in the constant \( c_i \).
This framework implies a simple welfare indicator. Prices of final goods are fixed as are real wages (at \( w_0 \)). Welfare changes therefore accrue only through changes in profits or land rents.\(^9\) The rent function in city \( i \) is \( r_i(z) = w_i - w_0 - c_i z = c_i \left( \frac{L_i}{K} - z \right) \). Integrating over \( z \) and adding over all spokes, total rent in a city of size \( L_i \) is
\[
R_i = c_i L_i^2 / 2K .
\]

4.2: Free entry, industry scale, and goods prices.

The final building block is to move from having a fixed number of firms in each sector to an endogenous number, determined by free entry. Profit levels are given by eqns. (1), so the free entry conditions are, for firms in each sector \( x \) and for each of the production modes,
\[
\pi_1(x) = p_x \left[ a(x)Q_{A1} + (2 - a(x))Q_{B1} \right] - 2w_1 \leq 0, \quad n_1(x) \geq 0,
\]
\[
\pi_F(x) = p_x \left[ a(x)Q_{A1} + (2 - a(x))Q_{B2} \right] - a(x)w_1 - (2 - a(x))w_2 - tp_x \leq 0, \quad n_F(x) \geq 0
\]
\[
\pi_2(x) = p_x \left[ a(x)Q_{A2} + (2 - a(x))Q_{B2} \right] - 2w_2 \leq 0, \quad n_2(x) \geq 0 .
\]

We proceed in two steps. First, suppose that the economy is a price-taker, so \( p_x \) is exogenous for all \( x \). This will generally involve the economy specialising in just two goods, since the endogenous variables in the definitions of profits are simply the two wages, \( w_1 \) and \( w_2 \).\(^{10}\) We therefore look first at a case in which we assume that there are just two final products. Second, we endogenise prices by adding a downward sloping demand curve for each product. It is then possible that the full range of sectors operates, and we explore this largely by numerical means in the final sections.

The 2x2 Ricardian case.

There are just two final sectors (or products), i.e. two values of \( x \) with associated \( A \)-intensities \( \alpha_1 \) and \( \alpha_2 \). We suppose the these function intensities are symmetric, so \( \alpha_1 + \alpha_2 = 2 \), and sector 1 is function \( A \)-intensive, \( 2 > \alpha_1 > 1 > \alpha_2 > 0 \).\(^{11}\) The prices of the two sectors are fixed

---

\(^9\) We do not specify where the owners of these rents live: since final expenditure goes on the perfectly tradable final product the location of spending has no bearing on the equilibrium.

\(^{10}\) With non-constant returns \( Q_{jk} \) are also endogenous. Diminishing returns would give non-specialisation but increasing returns, as assumed here, reinforce specialisation.

\(^{11}\) With \( a(x) = 1 + \gamma(1/2 - x) \) the corresponding values of \( x \) solve \( \alpha_1 = 1 + \gamma(1/2 - x) \) and \( \alpha_2 = 1 + \gamma(1/2 - x) \).
and equal, denoted \( p \). However, we allow for the possibility that two sectors have different fragmentation costs, \( t_1 \) and \( t_2 \).

The free entry conditions (12) apply for each sector in the continuum \( x \in [0, 1] \). Appendix 1 gives the corresponding conditions when these are evaluated for just two sectors. If fragmentation costs are the same in both sectors then the model is, once again, symmetric. The level of fragmentation costs at which sectors change mode depends on their function intensity (as in figure 2), and \( t^* = (2 - \alpha_i)\Delta q = \alpha_2 \Delta q \). At \( t > t^* \) firms in both sectors are integrated and cities are sectorally specialised. At \( t < t^* \) firms are fragmented and cities are functionally specialised. This is as before although now, with free entry, lower \( t \) reduces firms’ costs, attracting entry and increasing urban employment, nominal wages, and land rents.

To investigate asymmetries, suppose that fragmentation costs in the two sectors are different, with \( t_1 < t_2 \). There are now three regimes. If both sectors’ fragmentation costs are greater than \( t^* \) (regime I) then all firms are integrated and entry and exit occurs until wages are such that they make zero profits. These wage rates are given by appendix equations A1 and A4 reported in eqns. (13) below (where \( Q_+ = Q_{a1} = Q_{a2} \)). If \( t_1 < t^* \) and \( t_2 > t^* \) then firms in sector 1 are fragmented while those in sector 2 remain integrated. Wages then solve appendix eqns. A2 and A4, solutions of which are given in eqns. (14). This is regime II, and its lower edge is a value \( t^* < t^* \). The final regime, III, is that in which fragmentation costs in both sectors are less than \( t^* \), so all firms are fragmented. Wages are given by A2 and A3, and reported in eqns. (15) below.

\[
\text{I: } w_i = p\left[Q_+ - \frac{(2 - \alpha_i)\Delta q}{2}\right] = w_2 = p\left[Q_+ - \alpha_2 \Delta q / 2\right],
\]

\[
\text{II: } w_i = p\left[Q_+ + \frac{1}{\alpha_i}\left(\frac{(2 - \alpha_i)\alpha_2 \Delta q}{2} - t_1\right)\right], \quad w_2 = p\left[Q_+ - \alpha_2 \Delta q / 2\right]
\]

\[
\text{III: } w_i = p\left[Q_+ - \left\{\frac{t_1 (2 - \alpha_2) - t_2 (2 - \alpha_1)}{2(\alpha_1 - \alpha_2)}\right\}\right], \quad w_2 = p\left[Q_+ - \frac{t_2 \alpha_1 - t_1 \alpha_2}{2(\alpha_1 - \alpha_2)}\right]
\]

Several points come from these wage equations. In regime III in which both sectors are fragmented, a reduction in fragmentation costs in sector \( i \) raises wages in the city with comparative advantage in the function intensive in sector \( i \), but reduces wages in the other city (eqn. 15). Thus, a reduction in \( t_1 \) raises the return to function \( A \) (intensive in 1) and

\[\text{II: } w_i = p\left[Q_+ + \frac{1}{\alpha_i}\left(\frac{(2 - \alpha_i)\alpha_2 \Delta q}{2} - t_1\right)\right], \quad w_2 = p\left[Q_+ - \alpha_2 \Delta q / 2\right]
\]

\[\text{III: } w_i = p\left[Q_+ - \left\{\frac{t_1 (2 - \alpha_2) - t_2 (2 - \alpha_1)}{2(\alpha_1 - \alpha_2)}\right\}\right], \quad w_2 = p\left[Q_+ - \frac{t_2 \alpha_1 - t_1 \alpha_2}{2(\alpha_1 - \alpha_2)}\right]
\]

Several points come from these wage equations. In regime III in which both sectors are fragmented, a reduction in fragmentation costs in sector \( i \) raises wages in the city with comparative advantage in the function intensive in sector \( i \), but reduces wages in the other city (eqn. 15). Thus, a reduction in \( t_1 \) raises the return to function \( A \) (intensive in 1) and

\[\text{II: } w_i = p\left[Q_+ + \frac{1}{\alpha_i}\left(\frac{(2 - \alpha_i)\alpha_2 \Delta q}{2} - t_1\right)\right], \quad w_2 = p\left[Q_+ - \alpha_2 \Delta q / 2\right]
\]

\[\text{III: } w_i = p\left[Q_+ - \left\{\frac{t_1 (2 - \alpha_2) - t_2 (2 - \alpha_1)}{2(\alpha_1 - \alpha_2)}\right\}\right], \quad w_2 = p\left[Q_+ - \frac{t_2 \alpha_1 - t_1 \alpha_2}{2(\alpha_1 - \alpha_2)}\right]
\]

\[\text{II: } w_i = p\left[Q_+ + \frac{1}{\alpha_i}\left(\frac{(2 - \alpha_i)\alpha_2 \Delta q}{2} - t_1\right)\right], \quad w_2 = p\left[Q_+ - \alpha_2 \Delta q / 2\right]
\]

\[\text{III: } w_i = p\left[Q_+ - \left\{\frac{t_1 (2 - \alpha_2) - t_2 (2 - \alpha_1)}{2(\alpha_1 - \alpha_2)}\right\}\right], \quad w_2 = p\left[Q_+ - \frac{t_2 \alpha_1 - t_1 \alpha_2}{2(\alpha_1 - \alpha_2)}\right]
\]

\[\text{II: } w_i = p\left[Q_+ + \frac{1}{\alpha_i}\left(\frac{(2 - \alpha_i)\alpha_2 \Delta q}{2} - t_1\right)\right], \quad w_2 = p\left[Q_+ - \alpha_2 \Delta q / 2\right]
\]

\[\text{III: } w_i = p\left[Q_+ - \left\{\frac{t_1 (2 - \alpha_2) - t_2 (2 - \alpha_1)}{2(\alpha_1 - \alpha_2)}\right\}\right], \quad w_2 = p\left[Q_+ - \frac{t_2 \alpha_1 - t_1 \alpha_2}{2(\alpha_1 - \alpha_2)}\right]
\]

\[\text{II: } w_i = p\left[Q_+ + \frac{1}{\alpha_i}\left(\frac{(2 - \alpha_i)\alpha_2 \Delta q}{2} - t_1\right)\right], \quad w_2 = p\left[Q_+ - \alpha_2 \Delta q / 2\right]
\]

\[\text{III: } w_i = p\left[Q_+ - \left\{\frac{t_1 (2 - \alpha_2) - t_2 (2 - \alpha_1)}{2(\alpha_1 - \alpha_2)}\right\}\right], \quad w_2 = p\left[Q_+ - \frac{t_2 \alpha_1 - t_1 \alpha_2}{2(\alpha_1 - \alpha_2)}\right]
\]

\[\text{II: } w_i = p\left[Q_+ + \frac{1}{\alpha_i}\left(\frac{(2 - \alpha_i)\alpha_2 \Delta q}{2} - t_1\right)\right], \quad w_2 = p\left[Q_+ - \alpha_2 \Delta q / 2\right]
\]

\[\text{III: } w_i = p\left[Q_+ - \left\{\frac{t_1 (2 - \alpha_2) - t_2 (2 - \alpha_1)}{2(\alpha_1 - \alpha_2)}\right\}\right], \quad w_2 = p\left[Q_+ - \frac{t_2 \alpha_1 - t_1 \alpha_2}{2(\alpha_1 - \alpha_2)}\right]
\]

\[\text{II: } w_i = p\left[Q_+ + \frac{1}{\alpha_i}\left(\frac{(2 - \alpha_i)\alpha_2 \Delta q}{2} - t_1\right)\right], \quad w_2 = p\left[Q_+ - \alpha_2 \Delta q / 2\right]
\]

\[\text{III: } w_i = p\left[Q_+ - \left\{\frac{t_1 (2 - \alpha_2) - t_2 (2 - \alpha_1)}{2(\alpha_1 - \alpha_2)}\right\}\right], \quad w_2 = p\left[Q_+ - \frac{t_2 \alpha_1 - t_1 \alpha_2}{2(\alpha_1 - \alpha_2)}\right]
\]

\[\text{II: } w_i = p\left[Q_+ + \frac{1}{\alpha_i}\left(\frac{(2 - \alpha_i)\alpha_2 \Delta q}{2} - t_1\right)\right], \quad w_2 = p\left[Q_+ - \alpha_2 \Delta q / 2\right]
\]

\[\text{III: } w_i = p\left[Q_+ - \left\{\frac{t_1 (2 - \alpha_2) - t_2 (2 - \alpha_1)}{2(\alpha_1 - \alpha_2)}\right\}\right], \quad w_2 = p\left[Q_+ - \frac{t_2 \alpha_1 - t_1 \alpha_2}{2(\alpha_1 - \alpha_2)}\right]
\]
hence the wage in city 1. But, given the price and fragmentation cost of good 2, the higher cost of function $A$ must be matched by a cost reduction elsewhere; this means a reduction in wages in city 2, and thence a reduction in the return to function $B$.\footnote{The ‘paradoxical’ effect of outsourcing on wages in Grossman and Rossi-Hansberg (2008) is an example of this result.}

Figure 5 illustrates a case in which the two fragmentation costs vary together, but with $t_2$ twice the level of $t_1$. The horizontal axis gives the path of $t_1$ and, as discussed above, there are three regimes. The path of wages in each city comes directly from zero profit conditions and hence eqns. (13) – (15), but intuition comes from looking at these in conjunction with the varying output levels of each sector $Y_1$ and $Y_2$ (panel a).

**Figure 5: The structure of production and division of gains.**

**a) Output of each sector**

**b) Wages in each city**

At high $t_1$, in regime I, firms in both sectors are integrated, the sectors produce the same levels of output and the cities are the same size. Entering regime II, sector 1 fragments production, with function $A$ remaining in city 1 and function $B$ moving to city 2. This can only be accommodated in city 2 by a large reduction in the (integrated) output of sector 2, $Y_2$: at the same time, since sector 1 now draws on labour in both cities, there is a large increase in output $Y_1$. The direct effect of reductions in $t$ in regime II is to reduce costs in sector 1, leaving sector 2 unchanged. Wages in city 1 therefore increase, as do employment and land rents. These processes start to reverse when sector 2 fragments; the sector now uses city 1 labour to supply function $A$, so city 2 is able to accommodate an increase in output $Y_2$, and
contracts correspondingly. Since full symmetry must be restored at $t_1 = t_2 = 0$, there is convergence of city employment, wages, and sectoral output levels.

Two main messages come from this. The first is the uneven impact of falling fragmentation costs on the two cities: the city with comparative advantage in the function intensive in the first sector to fragment expands. Recall that nominal wages increase linearly with city size (eqn. 10) and land rents increase with the square of city size (eqn. 11). Since profits are zero, real wages are fixed at $w_0$ by the outside sector, and output prices are all constant, all the benefits of city 1 growth accrues to its landlords.

The second is the uneven impact on the output of the two sectors. The overall production structure of the economy (and hence its external trade pattern) moves dramatically in response to changes in internal communication costs that enable one sector to fully utilize the benefits of fragmenting production in line with urban comparative advantage.

**Many products and endogenous prices.**

With more sectors than wage rates (equal to the number of cities), non-specialisation will occur only if sectoral output prices adjust to maintain zero profits in all active sectors. Without this the economy would, in general, operate just two sectors, as in the preceding subsection. We now suppose that, in each sector, the country under study produces a good differentiated from those in the rest of the world, so faces a downward sloping demand curve and hence endogenous prices.

To capture this, recall that the numeraire good is produced in the hinterland, freely traded, and production of this good fixes the hinterland wage, $w_0$. We assume that there are quasi-linear preferences, with the hinterland good being the residual (linear) product. The sectors and cities we are studying produce goods at prices $p_x$, and these goods may also be produced in the rest of the world at fixed price $\overline{p}$. Within sector preferences between these sources of supply are given by an Armington aggregator, $G_x = \left[ \theta p_x^{1-\varepsilon} + (1 - \theta) \overline{p}^{1-\varepsilon} \right]^{\varepsilon/(1-\varepsilon)}$. Preferences across sectors are Cobb-Douglas, so the price index for all sectors is $P = \prod_{x \in \{0,1\}} G_x$. The indirect utility function is

$$V = M - C \ln(P) = M - C \sum_{x \in \{0,1\}} G_x$$

(16)

where $M$ is total income in the economy and $C$ is a constant. Demand is therefore

$$D(x) = C \theta p_x^{-\varepsilon} G_x^{\varepsilon-1}$$

(17)
The price of each sector is solved by equating this with supply,

\[
Y(x) = n(x)[a(x)Q_{A1} + (2-a(x))Q_{B1}] + n_F(x)[a(x)Q_{A1} + (2-a(x))Q_{B2}] + n_2(x)[a(x)Q_{A2} + (2-a(x))Q_{B2}],
\]

(18)

Free entry implies that the number of firms adjusts such that, for each sector \( x \), the active mode has \( \pi_i(x) = 0 \), and \( \pi_i(x) \leq 0 \) for inactive modes, \( i = 1, 2, F \). As the number of firms changes in each sector so does employment, city size, rents and wages.

The first point to note is that endogenising the number of firms, wages and prices has no effect whatsoever on the fragmentation story given in earlier sections. With our assumptions (in particular those supporting symmetry) the results in propositions 1 and 2 contain no price or wage terms.

**Figure 6a: Output change:**

**Figure 6b: City 1 employment shares**

\( \gamma = 2; \sigma_A = 0; \sigma_A = 0.0; v_0 = 1; v = 0.5; q_{A1} = 1.1; q_{B1} = 0.9; q_{B2} = 0.9; q_{B2} = 1.1; \quad \varepsilon = 5. \)

Second, the pattern of supply and prices across sectors changes as fragmentation costs fall. This is illustrated in figure 6.\(^{14}\) Figure 6a illustrates the proportionate change in supply from each sector at three values of \( t \). The highest value is where just a small intermediate range of sectors has fragmented, this causing entry and expanded production in these sectors. As \( t \) falls further the profits of fragmented firms increase, this causing a wider range of sectors to

\(^{14}\) Parameter \( C \) is calibrated such that all sectors are the same size at high value of \( t \) (i.e. when all integrated).
fragment, and causing further entry in fragmented sectors, as indicated by the two higher
curves. Notice that entry increases city employment and wages, and it is this cost increase
that causes the small reductions in output for sectors with integrated firms.

Figure 6b gives the shares of city 1 employment in each sector for the three values of t (so
areas under each curve integrate to unity). Without fragmentation all employment is in A-
intensive sectors, with $x \leq 0.5$. Fragmentation means that city 1 workers start to provide
function A to a wider range of sectors, so both cities become less sectorally specialised.

Further (preliminary) investigation of possibilities is contained in the second appendix to the
paper.

A final comment on welfare. There are now two components of welfare change, rent in each
of the cities and the change in consumer surplus from price index changes, $-C \ln(P)$. Notice
that we have not specified who receives these changes in consumer surplus; in general this is
a combination of landlords, workers in each city and the hinterland, or foreigners. While
rents unambiguously increase as $t$ falls price index effects can be ambiguous. In particular,
gains will be mitigated by a terms loss as domestically produced output is exported more
cheaply.

5: Concluding comments

There is abundant evidence that cities have high productivity, generally attributed to
agglomeration economies deriving from scale and density of economic activity. It is
empirically difficult to disentangle whether these benefits are city-wide, or are generated by
the clustering of particular sectors. This paper investigates the implications of productivity
differences arising neither within sectors nor within cities as a whole, but instead within
‘functions’, thought of as business services or other inputs that are widely used in production
in many sectors.

The paper shows how better communication between cities – be it transport or electronic –
can enable the economy to develop and exploit these productivity differentials. This has
consequences for firm organisation, city specialisation, the size and income levels of cities,
and the overall structure of the economy’s production and external trade. The results
presented in the paper are work in progress, and much remains to be done to better
understand the rich array of insights created by this way of thinking.
References:


Forslid, R. and T. Okubo (2010), “Spatial Relocation with Heterogeneous firms and Heterogeneous Sectors”, CEPR working paper 8117


Appendix 1: The 2x2 Ricardian case.

The two products are labelled by subscripts 1, 2, and their mode of operation by superscripts \( I \), \( F \). Free entry conditions are:

\[
\begin{align*}
\pi_1^I &= p[2Q_{A1} + (2 - \alpha_1)\Delta q] - 2w_1 \leq 0, \quad n_1^I \geq 0 \\
\pi_1^F &= p[2Q_{B2} - t_1] - \alpha_1 w_1 - (2 - \alpha_1)w_2 \leq 0, \quad n_1^F \geq 0 \\
\pi_2^I &= p[2Q_{A1} - t_2] - \alpha_2 w_1 - (2 - \alpha_2)w_2 \leq 0, \quad n_2^I \geq 0 \\
\pi_2^F &= p[2Q_{B2} - \alpha_2 \Delta q] - 2w_2 \leq 0, \quad n_2^F \geq 0
\end{align*}
\]

Employment levels are linked to wages by eqns. (10), and employment depends on the number of active firms according to:

\[
L_i = 2n_i^I + n_i^F \alpha_i + n_2^F \alpha_2, \quad L_2 = 2n_1^I + n_1^F (2 - \alpha_i) + n_2^F (2 - \alpha_2)
\]

Regime 1 has A1 and A4 holding with equality, so \( n_1^I > 0, n_2^I > 0, n_1^F = 0, n_2^F = 0 \).

Regime 2 has A2 and A4 holding with equality, so \( n_1^F > 0, n_2^F > 0, n_1^I = 0, n_2^I = 0 \).

Regime 3 has A2 and A3 holding with equality, so \( n_1^F > 0, n_2^F > 0, n_1^I = 0, n_2^I = 0 \).

Figure 5 is constructed with \( \alpha_1 = 1.25; \alpha_2 = 0.75; \Delta q = 0.2; t_1 = t; t_2 = 2t; \)
Appendix: General-equilibrium simulations with endogenous prices and free entry

In this section, we examine some properties of the model in more detail through the use of simulations. The model is formulated as a nonlinear complementarity problem, where weak inequalities are each matched with a non-negative variable. If one of the inequalities is strict in equilibrium (e.g., profits for a (potential) integrated firm entrant are negative), then the associated complementary variable is zero (the number of integrated firms is zero). This requires a discrete number of industries/sectors, so we fairly arbitrarily chose 51 sectors: an odd number gives a middle industry. These industries differ in relative function intensities in the same way they do in the continuous case.

With 51 industries, the model consists of 268 weak inequalities in 268 non-negative unknowns. This is coded in GAMS and the code is available from Markusen. Using set I to index the 51 industries, the variables and inequalities are:

**NON-NEGATIVE VARIABLES**
- $L_I, L_2$: labor demand or employment in city $i$
- $W_I, W_2$: wages in city $i$
- $SA_I, SB_1, SA_2, SB_2$: employment in function $i$ in city $j$
- $QA_I, QB_1, QA_2, QB_2$: labor productivity in function $i$ in city $j$
- $Q(I)$: total output of sector $i$ (all firm types)
- $N_I(I), N_2(I), N_3(I)$: number of firms of type $j = 1,2,3$ in sector $i$
- $PRICE(I)$: price of good $i$

**WEAK INEQUALITIES**
- $LF_I, LF_2$: employment equations
- $WF_I, WF_2$: wage equations
- $SFA_I, SFB_1, SFA_2, SFB_2$: equations for employment in function $i$ city $j$
- $QFA_I, QFB_1, QFA_2, QFB_2$: equations for productivity in function $i$ in city $j$
- $QF(I)$: equations for total output of sector $i$
- $NF_I(I), NF_2(I), NF_3(I)$: zero-profit conditions for firms of type $j = 1,2,3$
- $PFPRICE(I)$: demand equations

where firm type 3 is fragmented firms. The complementarity matching is that each row of inequalities is matched with the same row of unknowns.

In the following simulations, we use the same parameter values as those used in the earlier simulations except where noted. The model is solved repeated over a set of fragmentation costs $t$. Varying the number of firms will require a somewhat different set of $t$ values to see all the possible regimes, but this is not a substantive difference.

Figure A1 shows the symmetric Ricardian case using the same parameters as Figure 2. Except for the discreteness in Figure A1, the figures are qualitatively the same. One other small difference is that, with a discrete, odd number of industries, the middle sector will have firms in both cities when all firms are integrated on the right-hand side of the diagram. Two panels are shown at the bottom. The left-hand panel show producer wages and welfare as defined earlier in the paper. There is actually a small region of negative wage and “welfare” effects as fragmentation costs fall from their prohibitive level (remember that “welfare” here...
Well not quite. If there are a discrete and odd number of industries, then the middle sector may be producing in both countries. Then the sectoral specialization index will be a little less than one with integrated firms only. This is the case in the symmetric outcomes shown attached. This effect goes to does not included any measure of consumer surplus from falling prices!). This is due to a negative terms-of-trade effect due to the increased productivity of the fragmenting firms. The country has monopoly power in trade that is not being internalized. But further falls in $t$ lead to welfare gains, most of which are in the form of land rents.

The right-hand panel at the bottom of Figure A1 is an attempt to quantify the degree to which cities are specialized in sectors versus functions. We have not found an entirely satisfactory way of doing this, and are exploring alternatives. For now, the method is as follows, using the same notation developed earlier in the paper.

Let $i =$ country $\{1, 2\}$, $x =$ sector $\{1, n\}$, $k =$ function $\{A, B\}$

Let $s_{xi} =$ share of sector $j$ output located in country $i$

Let $r_{ki} =$ share of function $k$ employment located in country $i$

Let $\sigma_x =$ a Herfindal type index of specialization by sector

Let $\sigma_f =$ a Herfindal type index of specialization by function

Specifically, let $s_{x1}$ be the share of sector $x$’s output that is produced in country 1 with a corresponding equation for $s_{x2}$

$$s_{xl} = \frac{(n_1(x)(a(x)Q_{al} + (2 - a(x))Q_{bl}) + n_3(x)(a(x)Q_{al}))}{Q(x)}$$

Let $\sigma_x$ be $(1/n)$ times the sum of squared shares

$$\sigma_x = \frac{1}{n} \left[ \sum_x s_{xl}^2 + \sum_x s_{x2}^2 \right]$$

This index would take a value of 0.5 when fifty percent of each sector’s output is produced in country $i$ (all firms fragmented), if all the $a(x)$ and $Q$’s are the same. But these outputs are skewed toward region $i$’s industries of comparative advantage, so $\sigma_x$ generally takes on a value greater than 0.5 even when all firms are fragmented and all sector outputs are the same (e.g., region 1 will have output share of two industries 0.75 and 0.25, not 0.5 and 0.5).

In a symmetric high-$t$ outcome with only integrated firms and no sector producing in both regions, this index will take on a value of 1 (half the industries have a share of one, half a share of zero: $1 = (1/n)*((n/2 + n/2)$). If city 2 is depopulated so that all production is by city 1 firms, 2

---

1Well not quite. If there are a discrete and odd number of industries, then the middle sector may be producing in both countries. Then the sectoral specialization index will be a little less than one with integrated firms only. This is the case in the symmetric outcomes shown attached. This effect goes to
then the index takes on a maximum value of 1 (1 = (1/n)*(n + 0)).

The function specialization index is straightforward. Let \( s_{ki} \) be the total amount of function \( k \) that is performed in region \( i \). Then \( r_{ki} \) is the share of function \( k \) performed in region \( i \):

\[
r_{ki} = \frac{s_{ki}}{s_{k1} + s_{k2}}
\]

The index is the sum of squared shares, given by

\[
\sigma_f = \frac{1}{2} \left[ \sum_k r_{k1}^2 + \sum_k r_{k2}^2 \right] \quad k = A, B
\]

If all firms are fragmented, then function \( A \) is only produced in region 1, with that region having zero production of function \( B \), so the index is 1 (1 = (1/2)*(1+0+0+1)). In a symmetric equilibrium with all firms integrated, this index would have a value of 1/2 if all the \( a(x) \)'s and sector outputs were equal. But with comparative advantage, the shares of functions are skewed toward the region’s comparative advantage, so the index will exceed 0.5 but be less than one (e.g., region 1 will function shares of the two functions 0.75 and 0.25, not 0.5 and 0.5). In an asymmetric equilibrium such as discussed in section 4, the more productive region could have share of both function significantly in excess of one-half. As in the case of the sector specialization index, if one region is depopulated than this index will take on a value of 1 (1 = (1/2)*(1+1+0+0)).

The outcome for the symmetric Ricarian case is shown in the lower right-hand panel of Figure A1. HFUN denotes the function specialization index and HSEC the sector specialization index. We see that functional specialization increases as \( t \) fall, while sectoral specialization decreases. A similar result is also found in Duranton and Puga (2005), though there the switch in specialization is abrupt and complete.

To this point in the paper, we have not discussed external trade with the rest-of-world and what effect that has on city and industry structure. Pursuing logic from the multinational enterprise literature, we might conjecture that falling trade costs with the rest of world (or rising demand there) would encourage firms to bear the cost of fragmentation since the payoff is larger. It turns out that this depends in large part on nature of the fragmentation costs. In working on the paper, we used both ad valorem and specific fragmentation costs. The former is equivalent to iceberg costs: the fragmentation cost is in units of the good itself. The latter is equivalent to assuming that the cost is incurred in units of the numeraire good. Neither alternative seems more persuasive than the other based on economic intuition.
In the analytical section of the paper, it turns out that an ad valorem cost is simpler and yielded much cleaner and simpler analytical solutions. So that is used there. Under this assumption, an increase in world demand due to falling trade costs for goods or a rise in rest-of-world incomes has no effect on fragmentation. If the world price of a sector’s good rises, its fragmentation cost rises in the same proportion. There is no incentive to switch from integrated to fragmented production. We do not find this intuitively appealing.

The situation is quite different if the fragmentation cost is in units of the numeraire good. Then the effect of a fall in external trade costs is shown in Figure A2. We first solve the model using the ad valorem formulation of Figure A1 in order to provide a clean comparison. Then falling trade costs to the rest-of-world are assumed to lead to the same quantity demanded from each sector when the domestic price is 25 percent higher. Fragmentation costs are held constant at their initial level however, equivalent to assuming that they are incurred in units of the numeraire good. Figure A2 gives a result that seems intuitive from the point of view of the multinational enterprise literature: the higher demand from the rest-of-world encourages (marginal) firms to incur the fixed fragmentation cost: the area of fragmented firms becomes larger. Cities become more function-specialized when world demand grows.

Figure A3 presents results when comparative advantage is neutralized but there are strong spillovers in both functions: a figure of $\sigma = 2$ is used as in the section 4 and Figure 4. The top panel is quite similar to Figure 4, except that trade costs need to be re-scaled with free entry. We see that there is a more abrupt transformation to fragmented sectors when trade costs fall. This is the self-reinforcing effect of fragmentation. When firms in the middle sector find it advantageous to fragment, this raises function productivity in function A in city 1 and function B in city 2. That means that firms in the sectors on either side of the middle sector will also now find it advantageous to fragment and so forth.

Aside from this, the big difference between Figure A3 and Figure A1 is in the left-hand lower panel. With such strong agglomeration economies, fragmentation and its effect on productivity is so strong that there is a big negative effect on prices. This is sufficiently big that we again have a case of “immiserizing growth” for range over which fragmentation costs fall from their prohibitive level. We strongly emphasize again that this measure of “welfare” is wages and rents only, and does not include consumer surplus. The right-hand lower panel of Figure A3 is qualitatively similar to that of Figure A1.

In order to emphasize that this “welfare” result is not general, we present the same analysis in Figure A4, with sigma lowered to $\sigma = 1$. The result in the upper panel of Figure A4 is quite similar to that in Figure A3, except the $t$ axis is re-scaled to show all regions. Fragmentation first occurs at a much lower cost $t$ in Figure A4 with a lower spillover coefficient than in Figure A3. The lower-left panel of Figure A4 shows that this eliminates the “immiserization” present with the very strong spillover coefficient in Figure A3.

Figure A5 brings this section to a close by considering an asymmetric situation, in parallel with Figure 5 of section 4, in which only function A has spillovers. We continue to use
$\sigma = 1$ in this Figure. Now we see that, for intermediate costs $t$, there are more integrated industries in city 1 than in city 2. When all industries are integrated at high $t$, most of the industries are located in city 1, where function A is more productive than in city 2, and function B is equally productive in both cities. Therefore, the middle sector cannot be indifferent to its location, and will choose city 1.

This added concentration of activity in city 1 means that producer wages cannot be the same in the two cities, as shown in the left-hand lower panel of Figure A5. This gap does decrease as fragmentation costs fall, but it is not eliminated. At the left-hand edge where $t = 0$, there is a lot more output in the A-intensive industries and hence a lot more demand for labor for function A in city 1 and demand for labor for function B in city 2. That is, even with all firms fragmented, output and employment are skewed toward A-intensive industries. The lower right-hand panel of Figure A5 shows qualitatively similar results for the two specialization indices as in the other cases.
Figure A1: Symmetric Ricardian Case

Ricardian comparative advantage, free entry, no spillovers

\[
\begin{align*}
\text{GAMMA} &= 2; \quad \text{V0} = 1; \quad V = 0.5; \quad \text{Armington elasticity of substitution} = 5 \\
\text{SIGMA} &= 0; \quad \text{SIGMB} = 0; \\
\text{qa1} &= 1.1; \quad \text{qa2} = 0.9; \quad \text{qb1} = 0.9; \quad \text{qb2} = 1.1;
\end{align*}
\]

Integrated in 1

Integrated in 2

middle (integrated) industry produces in both countries

welfare

producer wages

Herfindal-type indices of functional (HFUN) and sectoral specialization (HSEC)
Figure A2: Symmetric Ricardian Case: lower trade costs in goods

Ricardian comparative advantage, free entry, no spillovers

Same parameters as Figure A1
Parable for falling trade costs in goods to rest of world; fragmentation costs constant
Same world demand with domestic export prices 25% higher
Figure A3: Symmetric spillovers case, high sigmas

No Ricardian comparative advantage, free entry, spillovers

\[
\begin{align*}
\Gamma &= 2; \quad V_0 = 1; \quad V = 0.5; \quad \text{Armington elasticity of substitution} = 5 \\
\Sigma &= 2; \quad \Sigma_B = 2; \\
q_a &= 1; \quad q_b = 1; \quad q_{b1} = 1; \quad q_{b2} = 1;
\end{align*}
\]
Figure A4: Symmetric spillovers case: lower sigmas

No Ricardian comparative advantage, free entry, spillovers

\[ \text{GAMMA} = 2; \quad \text{V0} = 1; \quad \text{V} = 0.5; \quad \text{Armington elasticity of substitution} = 5 \]
\[ \text{SIGMA} = 1; \quad \text{SIGMB} = 1; \]
\[ qa1 = 1; \quad qa2 = 1; \quad qb1 = 1; \quad qb2 = 1; \]

Herfindal-type indices of functional (HFUN) and sectoral specialization (HSEC)
Figure A5: Asymmetric spillovers: agglomeration economics in function A only

No Ricardian comparative advantage, free entry, spillovers in function A

\[ \text{GAMMA} = 2; \quad \text{V0} = 1; \quad \text{V} = 0.5; \quad \text{Armington elasticity of substitution} = 5 \]
\[ \text{SIGMA} = 1; \quad \text{SIGMB} = 0; \]
\[ qa1 = 1; \quad qa2 = 1; \quad qb1 = 1; \quad qb2 = 1; \]