

Gallo, Fredrik

Working Paper

Cournot Competition, Market Size Effects and Agglomeration

Working Paper, No. 2005:23

Provided in Cooperation with:

Department of Economics, School of Economics and Management, Lund University

Suggested Citation: Gallo, Fredrik (2005) : Cournot Competition, Market Size Effects and Agglomeration, Working Paper, No. 2005:23, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/259908>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Cournot Competition, Market Size Effects and Agglomeration*

Fredrik Gallo**

Department of Economics, Lund University

Abstract

We analyse a two-stage location-quantity game with many firms and two regions. We show that the firms will never agglomerate in the same location if transportation is costly between the regions. We also analyse the effects of differences in market size and economic integration on the allocation of industrial activity. For high levels of trade costs firms locate in different regions. Lowering the trade costs beyond a critical level triggers an agglomeration of industry in the larger region. This process of agglomeration is gradual in nature and trade costs have to be successively lowered for a full-scale agglomeration to take place.

Keywords: agglomeration, cross-hauling, market size effect, spatial Cournot competition

JEL Classification: D43, F12, L13, R3

* I am grateful to Fredrik Andersson, Fredrik CA Andersson, Richard E. Baldwin, Rikard Forslid, Joakim Gullstrand, Kristian Jönsson and Gianmarco Ottaviano for very helpful suggestions. The responsibility for any remaining errors and shortcomings rests with the author. Financial support from the *Tore Browaldh Foundation*, the *Foundation for Advancement of Economic Research at Lund University*, the *Crafoord Foundation* and the *Royal Swedish Academy of Sciences* is gratefully acknowledged.

** P.O. Box 7082, S-220 07 Lund, Sweden. e-mail: fredrik.gallo@nek.lu.se, fax no.: + 46 (0) 46 222 41 13.

1 Introduction and Previous Studies

Agglomeration externalities arising from the interaction of increasing returns, costly transportation of goods, and linkages between firms and consumers, have been thoroughly analysed in the *new economic geography* (NEG). There are two features common to almost every model in the field. The first is the extensive use of the Dixit-Stiglitz (1977) formalisation of monopolistic competition¹, and the second is the catastrophic nature of agglomeration. In this paper we present a model challenging these two features. While the Dixit-Stiglitz setting is chosen for its simplicity and analytical tractability in a general equilibrium framework, the approach imposes some strong assumptions on firm behaviour. Specifically, there is no strategic interaction whatsoever between firms; each and every one ignores any effects their own actions may have on other firms. The frequently used assumption of monopolistically competitive firms is a valid description of some real world industries. Other industries (like the petrochemical industry), however, have fewer firms and competition is clearly oligopolistic in nature. In its present state, the NEG has nothing to say about the location of such industries.

There are a few NEG models, which deviate from the standard set-up. Ottaviano et al. (2002) check the robustness of the core-periphery model's results by giving alternative specifications of consumer preferences and transportation costs. Strategic effects, however, are still ignored as the monopolistically competitive behaviour of firms is preserved. Ludema and Wooton (2000) develop a variant of Krugman's (1991) core-periphery model featuring Cournot competition and homogeneous goods. Their aim is to analyse the role of tax competition, not market structure *per se*, in the agglomeration process.

¹ See Fujita et al. (1999, ch. 4) for the basic framework.

Furthermore, it is the well-known backward and forward linkages that create a circular process of agglomeration of industrial activity.

In both these contributions agglomeration is catastrophic in nature, the second feature shared by most NEG models. The typical story goes as follows. Imagine a world consisting of some regions or countries, which are identical regarding factor endowments, preferences and technology. All barriers to trade, both formal and informal, are captured by a single measure, which initially is nearly prohibitive, forcing each region to be essentially self-sufficient. Industry is thus equally divided between the regions and regional income levels are the same. Then the standard experiment is to exogenously liberalise trade and analyse the effects on industry location. The initial symmetric equilibrium is stable until we reach a critical level of trade costs, called the *break point*, below which *all* of industry subject to agglomeration economies locates in one of the regions. The world hence spontaneously divides into an industrial centre having *all* the manufacturing activity affected by agglomeration forces, and a periphery having *none* of it. While the absence of firms acting strategically makes the NEG models less suitable to address industrial policy issues (Neary, 2001), the all-or-nothing character of agglomeration equilibria is unattractive from an empirical point of view.

The aim of this paper is to analyse how *strategic interaction* between firms, market size effects and economic integration can give rise to regional differences in industrial structure. In sharp contrast to the NEG's standard Dixit-Stiglitz (1977) set-up with differentiated goods and monopolistic competition, we employ a homogeneous-good Cournot oligopoly model. We deliberately rule out the self-reinforcing backward and forward linkages driving the results in the NEG models in order to isolate the pure effects of strategic interaction on firms' choice of location. It turns out that qualitatively different agglomeration

equilibria arise when firms' decisions to locate are driven entirely by strategic interaction, than when the standard linkages are operating. Specifically, agglomeration occurs gradually in the model.

We know of few NEG models sharing this feature. An extension of the vertical model in Fujita et al. (1999, ch. 14) is one example. If the competitive sector exhibits decreasing instead of constant returns to scale, then the wage paid to labour in that sector will rise when its employment falls. Since newly established firms in a region draw labour from the competitive sector this, in turn, will weaken the incentive for remaining labour to work in the monopolistically competitive sector and dampen the agglomeration tendency. Pflüger (2001) and Forslid and Ottaviano (2002) are two other models featuring gradual agglomeration. The former replaces the Cobb-Douglas upper-tier utility function with a quasi-linear utility function, removing income effects for the manufacturing sector, while the latter introduces regional size asymmetries. In both models the inter-regionally mobile factor enters only into the fixed cost. Together, these modifications result in smoother agglomeration processes. By analysing strategic interaction without linkages we add a different mechanism to the list.

Our approach in this paper is close in spirit to two lines of research. The first is the large body of literature on spatial competition descending from the pioneering work of Hotelling (1929). Until recently the conventional wisdom seemed to be that if firms compete in prices they tend to locate far from each other (d'Aspremont et al., 1979, d'Aspremont et al., 1983, Hamilton et al., 1989, Kats, 1995); if they compete in quantities they tend to agglomerate (Hamilton et al., 1989, Anderson and Neven, 1991). However, Pal (1998) shows that the results hinge on whether the economy has end points or not, or in economic terms, if firms face competition from all sides. If the economy is assumed to be

a circle then Bertrand and Cournot competition yield the same result (dispersion), if the economy is linear then Bertrand yields dispersion while Cournot gives rise to agglomeration. All of these studies are two-stage location games where either prices or quantities are strategic variables in the game's second stage. Amongst them only Hamilton et al. (1989) explicitly consider how different levels of transportation costs influence firms' profits and choice of location. However, their model is restricted to a duopoly and is of limited use when analysing how large-scale agglomerations are affected by economic integration.

In the spirit of these papers we analyse a two-stage location-quantity game albeit with discrete space. We then introduce differences in market size and analyse how economic integration influences the firms' choice of location. This brings us to the second strand of related research: Krugman and Venables (1990), Combes (1997) and Head et al. (2002), all of which focus on Cournot competition, economic integration and market size differences as determinants of industry location. Combes (1997) analyses a one-stage game where firms playing Cournot simultaneously decide in which of two regions to set up production and how much to supply to each region. Various regional asymmetries, including a difference in size, are introduced. The experiment in the paper is to allow free entry of *new* firms and analyse if the firms agglomerate in one of the regions, or if they locate symmetrically in both regions. Head et al. (2002) is an analysis of home market effects for *inter alia* a Cournot oligopoly with homogeneous goods. In their framework it is the tension between increasing returns and trade costs that is the focus of analysis in order to examine the robustness of Krugman's (1980) home market effect. Krugman and Venables (1990) also examine the interaction of scale economies, trade liberalisation and market size in determining firms' choice of location. However, neither of these papers scrutinises the nature of the agglomeration equilibria.

While Krugman and Venables (1990) assert that all of industry will eventually locate in the large region, they do not explicitly analyse intermediate agglomeration equilibria. As in the NEG models we analyse how economic integration affects the movement of *existing* firms *between* two regions, starting from a symmetric distribution of firms. This enables us to pinpoint the interplay of the level of trade costs, market size differences and the regional distribution of firms, in shaping agglomeration equilibria. In addition to analysing intermediate agglomeration equilibria we extend the analysis in Krugman and Venables (1990) in two ways. First, our analysis contains a more thorough examination of the welfare effects of trade liberalisation. Second, we introduce imperfect arbitrage possibilities for consumers.

In this paper we consider a simple model where firms, sharing the same technology and producing the same homogeneous good, play a two-stage game, simultaneously choosing a location in the first stage and the quantity to supply to each location in the second. In a benchmark case we show that Cournot competition yields spatial *dispersion* when the regions are identical and transportation of goods is costly. We then examine how economic integration and exogenous differences in market size affect the firms' choices of location. Starting from a sub-game perfect equilibrium with two firms located in separate regions of unequal size, we find that when transport costs fall below a critical level the firms agglomerate in the larger region. In the terminology of the NEG we call this symmetry-breaking level of trade costs the *break point*. We then extend the analysis to many firms to examine the agglomeration equilibria resulting from strategic interaction between firms. We find that they are strikingly different from the ones arising from the standard backward and forward linkages in the NEG as the model displays stable dispersed asymmetric equilibria. Agglomeration thus only occurs gradually and trade costs have to successively fall for a full-scale agglomeration to occur in the larger region. We

also find that the market provides an incentive to agglomerate “too early”. A social planner free to locate the firms as she wishes always prefers to move the firms at a level of trade costs lower than the break point. As a final extension we allow imperfect consumer arbitrage. If the cost of shipping goods for consumers is not too low, this is beneficial for the firms in the agglomerated equilibrium, increasing the break point.

The rest of this paper is organised as follows. Section 2 lays out the basic structure of the model. Section 3 analyses the location-quantity game with many firms when regions are of equal size. In section 4 we introduce differences in market size and examine the effects of economic integration on industrial structure and welfare. Section 5 analyses how the results are affected if the possibility of arbitrage is allowed. Finally, some tentative conclusions are drawn in section 6.

2 The Theoretical Setting

Our model is in fact one of reciprocal dumping à la Brander (1981) and Brander and Krugman (1983). The major difference is that we have a two-stage game where both quantities and locations are strategic variables to the firms. The world consists of two regions, $R = \{1, 2\}$, which are perceived to be spatially segmented by firms. The regions are of equal size in terms of population (we relax this assumption in section 4) and the consumers are assumed to be inter-regionally immobile. There is a set $N = \{1, \dots, n\}$ of identical firms playing a two-stage game where the first stage is simultaneously choosing in which region to locate production. In addition we define $N_1 = \{1, \dots, n_1\}$ and $N_2 = \{n_1 + 1, \dots, n_1 + n_2\}$ as the sets of firms located in regions 1 and 2. A firm can only *locate* in one of the regions, hence the sets N_1 and N_2 are disjoint,

$N_1 \cap N_2 = \emptyset$. In the game's second stage each firm observes the other firms' choices of location and given that information they simultaneously decide which quantities to supply to each region. The game is solved by backward induction. First each firm's choice of location is taken as given and the game's second stage is solved. We then analyse the choice of location.

All firms use the same technology to produce a homogeneous good and face identical marginal costs, which we denote $c > 0$ and assume to be constant. The *supply* of firm $i \in N$ in region $j \in R$ is denoted y_{ij} . The inverse demand function

in region $j \in R$ is $P_j = a - b \sum_{i=1}^n y_{ij}$, where $a, b > 0$ are constants and n is the total

number of firms in the world. Interregional trade incurs transportation costs whereas transportation *within* any of the regions is costless. Costs of transportation between the two regions take the iceberg form and are denoted t . For every unit a firm located in region k sells in location l , $t \geq 1$ units have to be produced and shipped. Consequently if y units are to arrive, then ty units have to be produced and shipped. By definition $t > 1$ if transportation is costly and $t = 1$ if it is free. Because only $\frac{1}{t}$ of the goods produced in region k arrives in location

l , production and supply are *not* identical if $k \neq l$, whereas they are equal when $k = l$. Throughout the paper we assume that $a > ct$, implying that $a > c$ since by definition $t \geq 1$. Finally, we let π_{ij} denote the profit of firm $i \in N$ when supplying y_{ij} units to region $j \in R$. Each firm's pay-off function in the game is given by its total profits, $\sum_{j \in R} \pi_{ij}$. Each of this paper's sections contains an

analysis that follows the standard analysis in the *NEG* closely. That is, given an initial symmetric distribution of firms, will the world divide into a core-periphery pattern with one of the regions hosting most of industry if trade is liberalised? We emphasise that we have no backward or forward linkages in the

model. All results are driven entirely by firms' spatial competition for market shares. In all sections, except section 5, it is assumed that there are no arbitrage possibilities for consumers. We first analyse the location-quantity game when regions are of equal size.

3 Many Firms and Equally Sized Regions

The profit of firm i from supplying region 1 is

$$\pi_{i1} = \left[a - b \left(y_{i1} + \sum_{k=1, k \neq i}^n y_{k1}^e \right) \right] y_{i1} - cy_{i1}, \quad i \in N_1, \text{ i.e. we assume firm } i \text{ is located in}$$

region 1 and does not pay any transportation costs when supplying the home market. Firm i 's expectation about firm k 's supply to region 1 is y_{k1}^e , $k \neq i$ and $k \in N_1 \cup N_2$. For a firm $j \in N_2$ the profit of *exporting* to market 1 is

$$\pi_{j1} = \left[a - b \left(y_{j1} + \sum_{k=1, k \neq j}^n y_{k1}^e \right) \right] y_{j1} - ct y_{j1}, \text{ where } y_{k1}^e \text{ is firm } j\text{'s belief about firm}$$

k 's supply to region 1, $k \neq j$ and $k \in N_1 \cup N_2$. While the domestic firms need not pay any costs of transportation when supplying the home market, firm j has to *produce* ty_{j1} units if it wants to *sell* y_{j1} units in market 1. This means that

when firms are located in different regions the exporting firms *de facto* face higher marginal costs (ct) than the domestic firms (c). Solving for the Cournot-Nash equilibrium quantities² yields firm i 's and j 's supply to region 1 as

$$y_{i1} = \frac{a - c + cn_2(t-1)}{b(1+n_1+n_2)} \quad \text{and} \quad y_{j1} = \frac{a - ct - cn_1(t-1)}{b(1+n_1+n_2)}.$$

With linear demand and costs functions profits are strictly concave in y_{i1} and y_{j1} respectively, and the first-order conditions are sufficient for maxima.³ Substituting into the

² See "Solving the two-region model with many firms" in the Appendix.

³ It can be verified that our model fulfils the conditions for local stability for an n firm oligopoly with asymmetric costs, see Dastidar (2000, p. 208 and proposition 2 on p. 211). For

expressions for profits gives $\pi_{i1} = \frac{1}{b} \left[\frac{a - c + cn_2(t-1)}{(1+n_1+n_2)} \right]^2$ and

$\pi_{j1} = \frac{1}{b} \left[\frac{a - ct - cn_1(t-1)}{(1+n_1+n_2)} \right]^2$. By symmetry the profit of exporting to region 2 for

firm i is $\pi_{i2} = \frac{1}{b} \left[\frac{a - ct - cn_2(t-1)}{(1+n_1+n_2)} \right]^2$ and the domestic profit for firm j equals

$\pi_{j2} = \frac{1}{b} \left[\frac{a - c + cn_1(t-1)}{(1+n_1+n_2)} \right]^2$. The total profits for firm $i \in N_1$ and $j \in N_2$ are then

equal to

$$(1) \quad \pi_i^T(n_1, n_2) = \frac{1}{b} \left[\frac{a - c + cn_2(t-1)}{(1+n_1+n_2)} \right]^2 + \frac{1}{b} \left[\frac{a - ct - cn_2(t-1)}{(1+n_1+n_2)} \right]^2$$

and

$$(2) \quad \pi_j^T(n_1, n_2) = \frac{1}{b} \left[\frac{a - c + cn_1(t-1)}{(1+n_1+n_2)} \right]^2 + \frac{1}{b} \left[\frac{a - ct - cn_1(t-1)}{(1+n_1+n_2)} \right]^2.$$

The first term in (1) and (2) is home market profits and the second is profits from exporting. If trade costs are low enough, then a firm exports to the other

region, $t < \frac{(a + cn_j)}{c(1+n_j)} \Rightarrow y_{ij} > 0, i \notin N_j, j = 1, 2$.⁴ So far we have solved the

game's second stage taking firms' locations as given. We now look at the choice of location. The first thing we note using (1) and (2) is $n_1 = n_2 \Rightarrow \pi_i^T(\cdot) = \pi_j^T(\cdot)$.

more on conditions ensuring existence, stability and uniqueness of Cournot equilibria, see Novshek (1985), Dixit (1986), Shapiro (1989) and Dastidar (2000).

⁴ Invasion takes place and there is cross-hauling, i.e. intra-industry trade in identical products, for the reasons given in Brander (1981) and Brander and Krugman (1983). Note that the prohibitive level is identical for the two regions only when firms are evenly distributed between them. Otherwise, they differ.

Whenever the firms are evenly distributed between the regions they all make the same profit. To see if this is an equilibrium in locations we look at a symmetric perturbation of the distribution of firms. Suppose a firm relocates from region 2 to region 1, what happens with firms' profits in the receiving region? Totally

differentiating (1), $d\pi_i^T(\cdot) = \frac{\partial \pi_i^T(\cdot)}{\partial n_1} dn_1 + \frac{\partial \pi_i^T(\cdot)}{\partial n_2} dn_2$, where $dn \equiv dn_1 = -dn_2$

yields⁵ $\frac{d\pi_i^T(\cdot)}{dn} = -\frac{2c^2(t-1)^2(1+2n_2)}{b(1+n_1+n_2)^2} < 0$ (since $n_2 \geq 0$). The profit of firm

$i \in N_1$ thus *strictly decreases* when a firm relocates from region 2 to that region.

In addition, the profits made by the remaining firms in region 2 *strictly increase* (see the Appendix). If we start from a symmetric distribution of firms with all firms making the same profit, then no firm has an incentive to relocate, since if it does its profits decrease, whereas profits go up in the region it left, creating an incentive to move back. By symmetry, if a firm moves from region 1 to region 2, profits for the remaining firms *increase*, whereas the profit of firm j in region 2 *decreases*. Because $\pi_i^T(\cdot) = \pi_j^T(\cdot)$ whenever $n_1 = n_2$ and the profit of any relocating firm decreases, we can conclude that any symmetric distribution of firms is an equilibrium. No firm wants to move to the other market since it will earn a smaller profit there.

What if the firms are initially unevenly spread between the regions and trade is costly? Due to the analysis above we know that each company in the region with more firms makes a smaller profit than the ones in the other region, which creates an incentive for firms to move. Even though profits decrease in the region with fewer firms when an additional firm locates production in it, it is straightforward to show (using (1) and (2)) that they are still greater than the

⁵ See the Appendix.

profits in the region with more firms as long as its number of firms is smaller. Any asymmetric distribution of firms is thus not an equilibrium as firms in the region with many firms have an incentive to move to the region with fewer firms. This is so until a symmetric equilibrium is reached and no further firm wants to relocate.

Finally, we also have $\pi_i^T(\cdot) = \pi_j^T(\cdot)$ and $\frac{d\pi_i^T(\cdot)}{dn} = 0$ without trade costs ($t = 1$).

When trade is free all firms have full access to both markets irrespective of where they locate rendering the choice of location pointless. Each firm is then indifferent about which region to locate production in; they can all agglomerate in either region or they can be distributed in any proportion between the two regions. However, provided that transport costs exist, firms will always spread evenly between the regions. This result is contrary to the conclusions reached in the NEG, where firms agglomerate in one region for an intermediate level of trade costs. The finding is also contrary to the results in Hamilton et al. (1989) and Anderson and Neven (1991), where Cournot competition yields spatial agglomeration. The reason is that firms face competition from all sides in our model, a crucial feature emphasised by Pal (1998).

4 Market Size Effects

In the previous section the only sub-game perfect equilibrium is a symmetric distribution of firms. This is no surprise as there were no explicit advantages of locating in a region with many competitors. The analysis in that section is included only as a benchmark, a point of reference, as the NEG also departs from regions that are identical in every respect. We now introduce a difference in market size. The effects of trade liberalisation on firms' choices of location are first traced out in the simplest possible setting (two firms only). We then

extend the analysis to many firms and we examine the welfare properties of the agglomerated equilibria.

4.1 An Illustrative Example: Two Firms

Suppose that one of the regions is larger than the other one in the sense that the number of individuals living there is greater. As before, all individuals in the world are identical so the demand intercepts in the two regions are the same. Summing individual demand horizontally at any given price then results in a more elastic market demand in the large region due to the larger number of individuals living there. More specifically, if we assume that region 1 is larger than region 2 then $P_1 = a - d(y_{11} + y_{21})$ and $P_2 = a - b(y_{12} + y_{22})$, where $d < b$ ⁶. Note that the difference in market size is entirely captured by the relationship between b and d ; the larger region 1 is relative to region 2, the smaller is $\frac{d}{b}$.

First, each firm's choice of location is taken as given and the game's second stage is solved. We then analyse the choice of location. Suppose the firms are located in different regions, firm 1 in region 1 and firm 2 in region 2.⁷ The firms' profits in region 1 are $\pi_{11} = [a - d(y_{11} + y_{21}^e)]y_{11} - cy_{11}$ and $\pi_{21} = [a - d(y_{11}^e + y_{21})]y_{21} - cy_{21}$, where e denotes each firm's expectation about the other firm's supply. In the Cournot-Nash equilibrium each firm's supply to region 1 is $y_{11} = \frac{(a - 2c + ct)}{3d}$ and $y_{21} = \frac{(a - 2ct + c)}{3d}$, respectively, and profits earned are $\pi_{11} = \frac{(a - 2c + ct)^2}{9d}$ and $\pi_{21} = \frac{(a - 2ct + c)^2}{9d}$. For region 2 we have

⁶ All the results in this section are qualitatively the same if we illustrate the difference in market size using different vertical intercepts (and allow for equal slopes) instead.

⁷ This choice is arbitrary, but harmless to the results due to the symmetry of the problem.

$$y_{12} = \frac{(a - 2ct + c)}{3b}, \quad y_{22} = \frac{(a - 2c + ct)}{3b}, \quad \pi_{12} = \frac{(a - 2ct + c)^2}{9b} \quad \text{and}$$

$\pi_{22} = \frac{(a - 2c + ct)^2}{9b}$. When the firms are located in separate regions the total profit for firm 1 is

$$(3) \quad \pi_1^{TD} = \frac{(a - 2c + ct)^2}{9d} + \frac{(a - 2ct + c)^2}{9b},$$

whereas for firm 2 it is

$$(4) \quad \pi_2^{TD} = \frac{(a - 2c + ct)^2}{9b} + \frac{(a - 2ct + c)^2}{9d}.$$

The first term on the right-hand side in (3) and (4) is home market profits made by each firm; the second term is export market profits. As in section 3 we assume that the level of trade costs is low enough to prevent domestic monopolies, $t < \frac{a+c}{2c}$, so both firms export to the other market when located in

separate regions. We next turn to the case when both firms are located in the same region, say region 1. The Cournot-Nash equilibrium is $y_{11} = y_{21} = \frac{(a-c)}{3d}$,

$$\pi_{11} = \pi_{21} = \frac{(a-c)^2}{9d}, \quad y_{12} = y_{22} = \frac{(a-ct)}{3b} \quad (\text{which are positive since } a > ct) \text{ and}$$

$$\pi_{12} = \pi_{22} = \frac{(a-ct)^2}{9b}. \text{ So, if both firms are located in the large region then the}$$

total profit of each firm becomes

$$(5) \quad \pi_i^{TA} = \frac{(a-c)^2}{9d} + \frac{(a-ct)^2}{9b}, \quad i = 1, 2.$$

If they both locate in the small region each firm's profit is

$$(6) \quad \pi_i^{TA} = \frac{(a-c)^2}{9b} + \frac{(a-ct)^2}{9d}, i = 1, 2.$$

Above, we have solved the game's second stage taking firms' locations as given. We now look at the choice of location by comparing the total profits earned by each firm in the different location scenarios. Three *a priori* obvious observations about location choice follow directly. First, using (3) and (4) it is straightforward to show that $\pi_1^{TD} > \pi_2^{TD}$ if $d < b$ (which holds since region 1 is defined to be the larger market) and $t > 1$ (trade is costly). When the firms are located in separate regions the one in the larger region always makes the greater profit. Second, the profit in equation (5) is greater than the profit in equation (6) when $t > 1$ and $d < b$. If both firms agglomerate in the same region they make greater profits when the agglomeration is in the larger region rather than in the smaller one. Third, the profit in equation (3) is larger than the profit in equation (6) if $\frac{b}{d} > \frac{a-ct}{a-c}$, which is true for all $t \geq 1$ when $d < b$. A firm always prefers locating in the larger market given that the other firm is located in the smaller market. A less obvious question is whether the firm located in the smaller region ever has an incentive to move to the larger region when the other firm is located there, i.e. is $\pi_2^{TA} > \pi_2^{TD}$ ever possible? Using (5) and (4) and solving for the value of trade costs, there is a market size effect, $\pi_2^{TA} > \pi_2^{TD}$, if

$$(7) \quad t < \frac{a}{c} - \frac{d}{b} \left(\frac{a-c}{c} \right).$$

In the terminology of the NEG we call this level of trade costs the *break point*, because if we start from a symmetric equilibrium, then lowering the trade costs below this level changes the equilibrium in locations from dispersion to agglomeration in the larger region. Combining (7) with the condition that trade costs are low enough to prevent domestic monopolies, we have *no* market size effect if $\frac{a}{c} - \frac{d}{b} \left(\frac{a-c}{c} \right) < t < \frac{a+c}{2c}$ as the firm located in the smaller region

makes a larger total profit staying in that region, $\pi_2^{TD} > \pi_2^{TA}$. That t indeed can lie in this interval follows from $\frac{a}{c} - \frac{d}{b} \left(\frac{a-c}{c} \right) < \frac{a+c}{2c} \Rightarrow d > \frac{b}{2}$. If market 1 is

not too large relative to market 2 there exists an interval of trade costs where the firm located in the smaller market wants to stay there. On the other hand, if $d \leq \frac{b}{2}$, then region 1 is so large that both firms want to locate there for all

$$t \in \left[1, \frac{a+c}{2c} \right).$$

Now suppose the two firms are located in different regions. Lowering the transportation costs so that inequality (7) holds induces the firm in the smaller region to relocate to the larger region. The reason is illustrated in Figure 1 below.⁸

⁸ The parameter values used for all the figures in the paper can be found in the Appendix.

Figure 1. Profits per firm and integration (two firms)

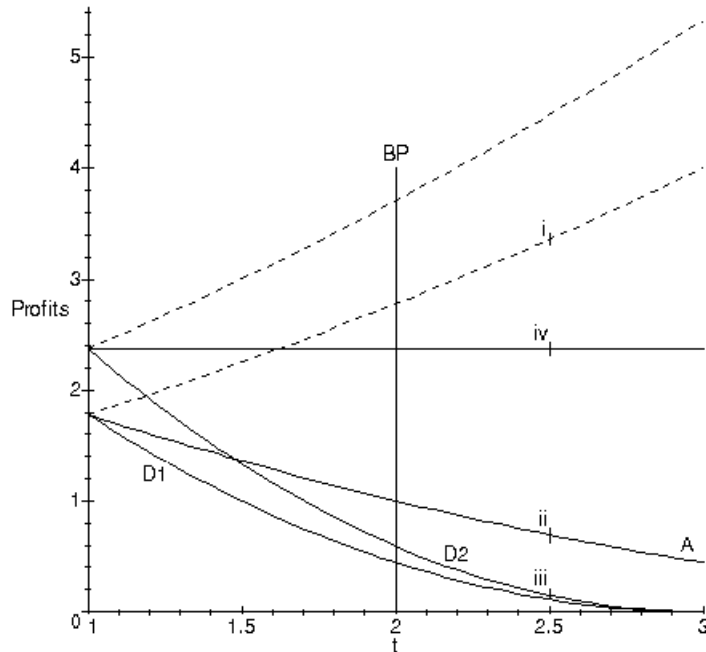


Figure 1 is a plot of the profits the two firms make in the two different location scenarios. The two dotted curves show the profit each firm earns in its home market when they are located in separate regions (given by the first terms on the right-hand sides of equations (3) and (4)). The horizontal line is the profit that each firm makes in the larger region if both firms agglomerate there (first term in equation (5)). Finally, the curves A , $D1$ and $D2$ are export market profits per firm if they agglomerate (A) or are dispersed ($D1$ for firm 1, $D2$ for firm 2). Suppose trade costs are $t = 2.5$ and consider firm 2's choice of location. The profit it earns in region 2 drops from i to ii , should it move to region 1, whereas the profit in region 1 rises from iii to iv . For high levels of trade costs (i.e. above the break point BP) the loss dominates the gain. At the break point these changes in profits balance, whereas the gain dominates the loss for levels lower than the break point. This is so because the firm's home market profit decreases as a result of firm 1's invading exports. For high levels of trade costs it is

sufficiently protected from firm 1’s exports and does not want to move. For lower trade costs, however, firm 2’s home market profit is further squeezed and it approaches the home market profit it would make if it relocated to region 1. In addition, each firm’s export profit, if they are agglomerated in region 1, is higher than firm 2’s export to the larger region ($A > D2$), fuelling the incentive to relocate. This outcome of agglomeration in the larger region is more likely the bigger the difference in size between the regions as illustrated in Figure 2 below, where the boundary BP is a plot of the break point against $\frac{d}{b}$ using (7).

Figure 2. The relationship between relative market size and the break point

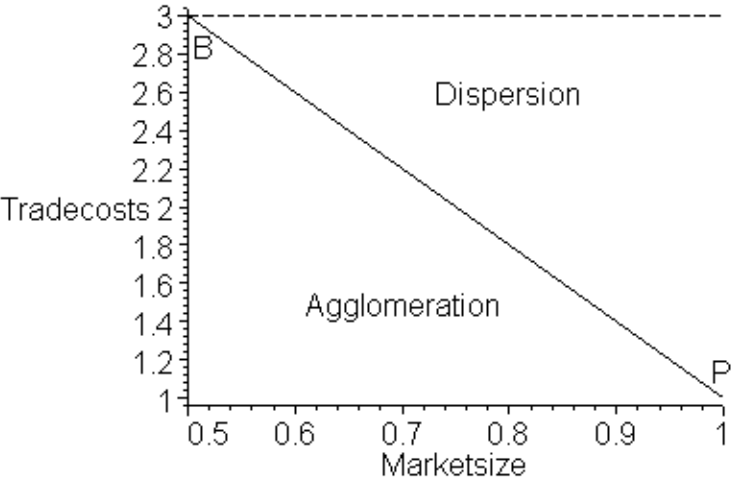


Figure 2 illustrates how a change in the ratio between d and b *ceteris paribus* affects the critical value of t . The dotted line is the level of trade costs that allows trade, $t = \frac{a + c}{2c}$. Note that the *lower* the value of $\frac{d}{b}$ the *bigger* is region 1 relative to region 2. Trade costs below BP trigger agglomeration in the large region for any given difference in market size, whereas values above it are high enough to preserve the dispersed equilibrium. From the figure it is clear that the

smaller the difference in market size, ($\frac{d}{b}$ rises), the less attractive it is to locate in the larger region and hence the lower the critical value of t triggering the market size effect. Conversely, the larger region 1 is relative to region 2, ($\frac{d}{b}$ falls), the stronger the incentive to move to that region and the higher the value of the break point.

4.2 Market Size Effects: the Case of Many Firms and Two Regions

We believe that the analysis in the previous section is interesting and extend it here to n firms. The total profits for firm $i \in N_1$ and $j \in N_2$ are then equal to domestic plus export market profits:

$$(8) \quad \pi_i^T(n_1, n_2) = \frac{1}{d} \left[\frac{a - c + cn_2(t-1)}{(1+n_1+n_2)} \right]^2 + \frac{1}{b} \left[\frac{a - ct - cn_2(t-1)}{(1+n_1+n_2)} \right]^2$$

$$(9) \quad \pi_j^T(n_1, n_2) = \frac{1}{b} \left[\frac{a - c + cn_1(t-1)}{(1+n_1+n_2)} \right]^2 + \frac{1}{d} \left[\frac{a - ct - cn_1(t-1)}{(1+n_1+n_2)} \right]^2.$$

We follow the NEG in supposing that the firms initially are evenly distributed between the regions. Denoting the number of firms in each region n_1^D and n_2^D we have $n_1^D = n_2^D$. Inserting in (8) and (9) it is straightforward to show that the firms in region 1 (the large region) make greater profits for all $t > 1$ than the firms in region 2 do. Consider now firm j in region 2. Will moving to region 1 be profitable? To answer this question we have to compare the profit it earns *after* the relocation with its *current* profit. As in section 3 the profits in the receiving (sending) region will decrease (increase). As argued above, however, the firms in the large region make greater profits when the distribution of firms

is symmetric, and they may continue to do so after the relocation. To analyse when this is the case we move from a situation with the number of firms being (n_1^D, n_2^D) and firm j 's profit given by (9), to a new one with $(n_1, n_2) = (n_1^D + dn_1, n_2^D + dn_2)$, where $dn_1 = -dn_2$ and the firm's profit is given by (8). Denote the new distribution of firms (n_1^A, n_2^A) . Inserting (n_1^A, n_2^A) in (8), (n_1^D, n_2^D) in (9) and solving for the level of trade costs⁹ we have $\pi_i^T(n_1^A, n_2^A) > \pi_j^T(n_1^D, n_2^D)$ if

$$(10) \quad t < \frac{[(2a - c)(d - b) + c(n_2^A - n_1^D)(d + b)]}{c[(n_2^A - n_1^D)(d + b) + (d - b)]}.$$

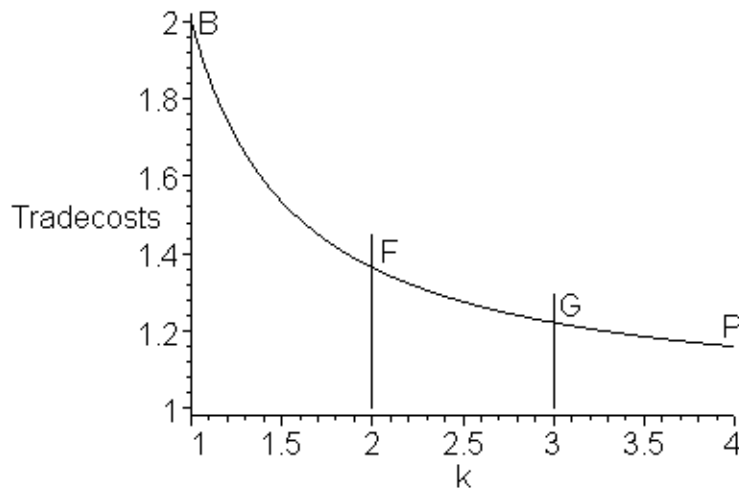
As in the previous section we call this level of trade costs the *break point*. If economic integration lowers trade costs below this level, firm j will find it profitable to move to the larger region. Suppose we start from *any* symmetric distribution of firms, (n_1^D, n_2^D) with $n_1^D = n_2^D$, and that one firm relocates to region 1. After the relocation we have $(n_1^A, n_2^A) = (n_1^D + 1, n_2^D - 1)$, hence $n_2^A - n_1^D = -1$. Inserting in (10) and simplifying yields $t < \frac{a}{c} - \frac{d}{b} \left(\frac{a - c}{c} \right)$, which we recognise from inequality (7) above. The analysis in the previous section with two firms is also valid with many firms. If trade costs are lower than the break point in (7), then *any* symmetric distribution of firms is not an equilibrium as a relocating firm will earn a larger profit. However, the break point given by

⁹ Because the total number of firms is constant it must be that the denominators in (8) and (9) are equal, i.e. $1 + n_1^D + n_2^D = 1 + n_1^A + n_2^A$. Also, after having derived (10) we must make sure that the break point is smaller than the prohibitive level of trade costs. Otherwise there are no exports and using (8) and (9) would be wrong. A condition for this can be found in the Appendix.

(10) is valid for any other allocation of firms between the regions. Inequality (10) also makes clear that the symmetry-breaking level of trade costs changes with the number of firms that has moved as illustrated numerically below.

Suppose we start from the symmetric equilibrium $(n_1^D, n_2^D) = (3, 3)$. If a firm relocates we have $(n_1^A, n_2^A) = (4, 2) \Rightarrow n_2^A - n_1^D = -1$. Inserting in (10) $\Rightarrow t < 2$ will induce the firm to relocate. Suppose that the *actual* level of trade costs is lower than this, say $t = 1.5$, and the firm moves. Now, what about the next firm, will it move too? If it chooses to move, the new configuration of firms is $(n_1^{A*}, n_2^{A*}) = (5, 1)$ and profits made under that configuration should be compared to current profits with $(n_1^A, n_2^A) = (4, 2)$. In this case $n_2^{A*} - n_1^A = -3$, inserting in (10) $\Rightarrow t < 1.36$ and the firm stays in region 2. Not until trade costs fall below this level will it move to region 1. Proceeding with this numerical example we find that $n_2^A - n_1^D$ in (10) evolves according to $n_2^A - n_1^D = 1 - 2k$, $k = 1, 2, 3 \dots$ where k stands for the k :th firm considering a relocation. Figure 3 illustrates how this difference in the number of firms affects the break point by inserting $1 - 2k = n_2^A - n_1^D$ in (10) and plotting it against k^{10} .

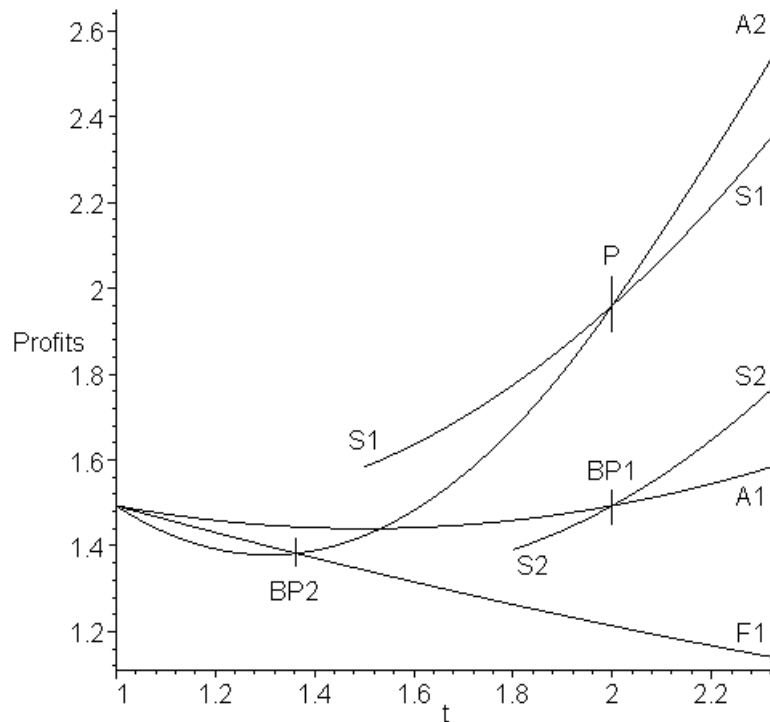
Figure 3. The break point for the k :th firm considering relocation



The boundary BP is the locus of break points, values below this boundary trigger agglomeration in the large region. For the first firm considering relocation, trade costs have to be lower than 2 to induce the firm to move. We see in Figure 3 that once the firm has moved to the larger region, the lower the break point has to be to induce the next firm to relocate. The break points for the second and third firms considering moving to region 1 are given by the intersection of the vertical lines and BP , at the points F and G . The break point in F is $t \approx 1.36$ and in G it is $t \approx 1.22$. For the next firm considering moving the trade costs have to be even lower for the relocation to be profitable. Figure 4 illustrates why this is so.

¹⁰ We ignore the integer problem and plot against a continuous k for expositional purposes.

Figure 4. Total profits per firm and integration (many firms)



In Figure 4 the curves $S1$ and $S2$ show the *total* profit per firm in each region at the symmetric distribution of firms, $(n_1, n_2) = (2, 2)$. Similarly, the curves $A1$ and $A2$ illustrate the total profit per firm at the asymmetric distribution $(n_1, n_2) = (3, 1)$. With two firms in each region the prohibitive level of trade costs is $t = 7/3$. After one firm has moved to region 1, the new prohibitive level for the firm remaining in region 2 is $t = 2$, illustrated by P . The part of $A2$ to the right of P is thus home market profits only. To the left of P profits of exporting are added. Finally, $F1$ is the total profit per firm when all four firms are agglomerated in region 1, $(n_1, n_2) = (4, 0)$. Starting from the symmetric equilibrium and lowering the trade costs below the prohibitive level induces invasion, reducing profits ($S1$ and $S2$ fall). As before, for high levels of trade costs the firms in the smaller region still make higher profits than if one of them should relocate to region 1 ($S2 > A1$). When we reach the first break point

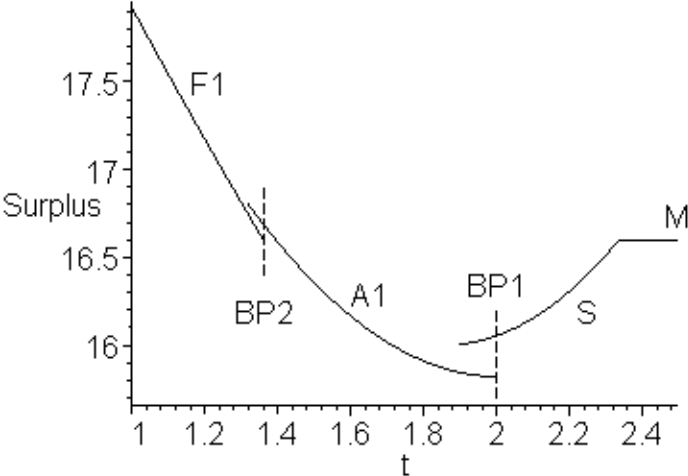
($BP1$), however, it becomes profitable for one of the firms to move to the large region ($A1 > S2$). Note that the profit of the remaining firm is larger than the profit of each firm in the core ($A2 > A1$) for high t , the reason being the now larger share of its home market, reducing its incentive to move. Furthermore, at high t it is sufficiently protected from invading exports from firms located in region 1. Also, because there are now more firms in the large region competition will be fiercer there, and the gain in market share smaller, if it moves. Indeed, this asymmetric distribution of firms is an equilibrium for all trade costs $t \in [BP2, BP1]$, as the remaining firm will not find it profitable to move until we pass the second break point $BP2$ (when $F1 > A2$). A large-scale agglomeration in region 1 is only possible if trade costs successively fall. Agglomeration equilibria due to pure strategic interaction between firms are thus different from the ones arising from backward and forward linkages. In the latter case passing the break point triggers a large-scale agglomeration in one region as *all* of industry relocates, whereas here the break point itself changes with agglomeration. Agglomeration of industrial activity driven by strategic interaction is thus not a catastrophic “all-or-nothing” process as it is in most of the NEG literature.

4.3 Welfare Analysis

Recently the welfare properties of agglomeration in NEG models have received interest (see chapter 11 in Baldwin et al. 2002 and Ottaviano et al., 2002). We next analyse how integration between two regions of unequal size affects global welfare. The welfare consequences of the market outcome when agglomeration is driven entirely by strategic interaction between firms are investigated and compared to the ones resulting from the outcome a social planner would achieve. We start from a high level of trade costs so firms are evenly distributed between the two regions. Each region’s welfare is calculated as the sum of

consumer and producer surplus; since the analysis is one of partial equilibrium and we do not model incomes, this is a natural benchmark specification. Figure 5 below shows how the total surplus is affected by trade liberalisation and compares the different location scenarios from Figure 4 above.

Figure 5. Global surplus changes



Without fixed costs the producer surplus for a firm is equal to its profits. The total surplus in region i under scenario j is then equal to the sum of consumer surplus and profits made in the region under that scenario: $TS_i^j = CS_i^j + \Pi_i^j$, $i = 1, 2$ and $j \in \{F1, A1, S, M\}$. The horizontal line segment M is welfare under autarky, which we use as a benchmark. The curve S is welfare in the symmetric equilibrium with trade. Finally, $A1$ and $F1$ display welfare in the agglomerated equilibria with trade ($A1$ for the configuration $(n_1, n_2) = (3, 1)$ and $F1$ when $(n_1, n_2) = (4, 0)$). The vertical dotted lines show the two break points. We see in Figure 5 that integration initially lowers global welfare compared to autarky. The reasons are the same as in Brander and Krugman (1983). For high values of trade costs (i.e. near the prohibitive level), the negative effects on welfare

stemming from trade diversion (domestic production being replaced by high cost imports) and the costly transportation of identical goods dominate the positive effects of firms' eroding monopoly power. As trade costs are gradually lowered, the negative effects become less important and the pro-competitive effect starts to dominate, increasing welfare (not shown in the figure for expositional clarity). In Brander and Krugman (1983) the end of the story is that successively falling trade costs increase welfare in both regions, which are both better off with low trade costs compared to autarky. In our setting, however, the regions are of different size and lowering trade costs to the first break point, $BP1$, triggers a relocation of one firm to the larger region. At the break point welfare jumps down and is initially lower than that in the symmetric equilibrium. To the left of the first break point global surplus is given by AI . Liberalising trade is now unambiguously beneficial and total welfare surpasses that of the other two scenarios (M and S) as trade costs are lowered even further.

When we reach the second break point, $BP2$, the whole industry agglomerates in the larger region and welfare again jumps down. Note that we have intentionally extended the S and AI curves somewhat to the left of the two break points, even though they are not part of the set of location equilibria. The reason is to illustrate what the global surplus would have been had the configuration of firms not changed. From Figure 5 it is clear that, from a social planner's point of view, the firms agglomerate "too early". The level of trade costs where a social planner would choose to move the first firm is given by the intersection of the curves S and AI . For the second firm it is given by the intersection of the AI and FI curves (the two intersections are not shown in the figure). That the planner's choice of trade costs is lower than the corresponding break points is a general feature of the model as shown below.

Suppose we have an arbitrary equilibrium with the number of firms in each region being n_1^A and n_2^A . Consider a social planner who thinks about relocating a firm from the small to the large region so that the new configuration of firms becomes n_1^{A*} and n_2^{A*} . She will only do so when the total surplus under the new configuration of firms exceeds the total surplus associated with the existing one. The planner thus compares welfare under the new distribution of firms,

$$TS_k \equiv n_1^{A*} \pi_i^T(n_1^{A*}, n_2^{A*}) + \frac{d(Y_1^*)^2}{2} + n_2^{A*} \pi_j^T(n_1^{A*}, n_2^{A*}) + \frac{b(Y_2^*)^2}{2},$$

with welfare under the old one, $TS_{k-1} \equiv n_1^A \pi_i^T(n_1^A, n_2^A) + \frac{d(Y_1)^2}{2} + n_2^A \pi_j^T(n_1^A, n_2^A) + \frac{b(Y_2)^2}{2}$, where

$\pi_i^T(\cdot)$ and $\pi_j^T(\cdot)$ are from equations (8) and (9) and Y_i is total quantity supplied to region i by the appropriate configuration of firms.

Next we note two relationships among the distributions of firms. First, the number of firms after the relocation is $n_1^{A*} = n_1^A + 1$ and $n_2^{A*} = n_2^A - 1$. Second, the relationship between the number of firms in the arbitrary equilibrium is $n_1^A = n_2^A + 2(k-1)$, where $k = 1, 2, 3, \dots$ stands for the k :th firm the planner considers relocating. If k is unity the comparison is between a symmetric equilibrium and the first agglomerated equilibrium, if k equals two the comparison is between the first and second agglomerated equilibria and so on. Making use of these two observations (i.e. inserting them in the expressions for TS_k and TS_{k-1}) we have $TS_k \geq TS_{k-1}$ if

(11)

$$t \leq \frac{4a(b-d)(k+n_2^A) + c[b(8k(k-1)+1+n_2^A(8k-6)) + d(n_2^A(8k-2)+1+4k(2k-1))]}{c[d(8k(k-1)+1+n_2^A(8k-6)) + b(n_2^A(8k-2)+1+4k(2k-1))]}$$

, where n_2^A is the number of firms in region 2 in our arbitrary equilibrium *before* the planner moves the k :th firm. Using (11) we can calculate the value of t when the planner chooses to move the first firm. Inserting $k = 1$ and $n_2^A = 2$ we obtain $t \approx 1.58$, which is less than the first break point ($t = 2$). For the second firm we have $k = 2$ and $n_2^A = 1$, and so $t \approx 1.2$, which is less than the second break point ($t \approx 1.36$). What we want to show is that the level of t in (11), which we call t^{planner} , is less than the break point in (10). We have $t^{\text{break}} - t^{\text{planner}} > 0$ if

$$\frac{(b-d)(d+b)(2k-1)(a-c)(2k-1+2n_2^A)}{c[d(k-1)+bk]\left(b\left[2n_2^A(4k-1)+4k(2k-1)+1\right]+d\left[2n_2^A(4k-3)+8k(k-1)+1\right]\right)} > 0$$

, which is true for all $k \geq 1$ as $b > d$ and $a > c$ by assumption. The planner's choice of t is always lower than the market's break point. The reason is that when a firm decides to move it takes only the effect on its own profits into account. But the relocation also has effects on consumer surplus and other firms' profits in both regions, something the planner takes into account. To disentangle all these effects we next decompose the welfare change. We begin by analysing the change in consumer surplus.

Differentiating each region's consumer surplus totally we obtain

$$dCS_1 = \frac{dCS_1}{dY_1} \frac{\partial Y_1}{\partial n_1} dn_1 + \frac{dCS_1}{dY_1} \frac{\partial Y_1}{\partial n_2} dn_2 \quad \text{and} \quad dCS_2 = \frac{dCS_2}{dY_2} \frac{\partial Y_2}{\partial n_1} dn_1 + \frac{dCS_2}{dY_2} \frac{\partial Y_2}{\partial n_2} dn_2,$$

where $dn \equiv dn_1 = -dn_2$ and all the derivatives can be found in the Appendix.

Consumer surplus in region 1 (region 2) increases (decreases) provided that total supply in the region increases (decreases) as a result of the relocation. Inserting the derivatives and simplifying it can be verified that this is the case if $c(t-1) > 0$, which always holds when trade is costly. Consumers in the large region gain, whereas consumers in the small region lose. The total change is

$$\frac{dTCS}{dn} = \frac{dCS_1}{dn} + \frac{dCS_2}{dn} = \frac{c(t-1)}{(1+n_1+n_2)}(Y_1 - Y_2), \quad \text{where we have used the}$$

derivatives in the Appendix, so $\frac{dTCS}{dn} > 0$ if $Y_1 > Y_2$. In the Appendix we show that the last inequality always holds. We can conclude that the net effect on consumer surplus is always positive. We next turn to profits.

Let $\pi_1^T = n_1 \pi_i^T(n_1, n_2)$ denote total profits in region 1, where $\pi_i^T(n_1, n_2)$ is profit per firm in region 1 (from equation 8). We know that the total number of firms is constant, $n_1 + n_2 = k$, so $n_2 = k - n_1$ and $\pi_1^T = n_1 \pi_i^T(n_1, k - n_1)$. The total effect on profits in region 1 when a firm relocates to that region is

$$\frac{d\pi_1^T}{dn_1} = \pi_i^T(\cdot) + n_1 \frac{d\pi_i^T(\cdot)}{dn_1}, \text{ where } \frac{d\pi_i^T(\cdot)}{dn_1} = \frac{-2c(t-1)}{(1+n_1+n_2)}(y_{i1} - y_{i2}).$$

The latter is negative if the quantity supplied to the home market is larger than the quantity supplied to the export market, $y_{i1} > y_{i2}$. This is always the case as region 1 is bigger and marginal costs of supplying region 2 are higher than when supplying region 1. We have that $y_{i1} > y_{i2}$ if $t > \frac{ad + dc n_2 - ab + bc + bc n_2}{bc n_2 + cd + cd n_2}$, where the

right-hand side can be shown to be less than unity. The inequality thus always holds and so $\frac{d\pi_i^T(\cdot)}{dn_1} < 0$. To summarise the total effect on profits in region 1:

there is one more firm earning a profit in region 1, increasing total profits by $\pi_i^T(\cdot)$, but the profit per firm is lower, reducing total profits by $n_1 \frac{d\pi_i^T(\cdot)}{dn_1}$.

For region 2 we have $\pi_2^T = (k - n_1) \pi_j^T(n_1, k - n_1)$ and $\frac{d\pi_2^T}{dn_1} = -\pi_j^T(\cdot) + n_2 \frac{d\pi_j^T(\cdot)}{dn_1}$,

where $\frac{d\pi_j^T(\cdot)}{dn_1} = \frac{2c(t-1)}{(1+n_1+n_2)}(y_{j2} - y_{j1})$. One firm less decreases the total profits

by $\pi_j^T(\cdot)$, but the profit per firm increases (provided that

$t > \frac{ab + bcn_1 - ad + cd + cdn_1}{bcn_1 + bc + cdn_1}$), which is represented by the second term. The

total effect on profits is $\frac{d\pi_1^T}{dn_1} + \frac{d\pi_2^T}{dn_1} = \pi_i^T(\cdot) - \pi_j^T(\cdot) - \frac{2c(t-1)}{(1+n_1+n_2)}(Y_1 - Y_2)$. The

first two terms on the right-hand side are what govern the individual firm's decision to relocate and the difference is positive when (10) holds. The third term is the total effect on profits for the rest of the industry. It is ignored by the moving firm and can be thought of as a negative externality. The firm considers only its own change in profit when moving, but the move has effects on overall profits of the industry. This is something that the social planner takes into account. Note that the sign of the last term is governed by the same inequality as the net effect on consumer surplus. In the Appendix we showed that $Y_1 > Y_2$ always holds, hence it is negative. Furthermore, it is twice as big as the positive net effect on consumer surplus, which is thus eliminated. As in Ottaviano et al. (2002) the market generates excess agglomeration, even though we work with a different model featuring a more gradual agglomeration mechanism. In our setting this is true for all trade costs, whereas over-agglomeration only happens for intermediate values of trade costs in Ottaviano et al. (2002).

5 An Extension: Allowing Consumer Arbitrage

So far we have ruled out the possibility of arbitrage. As a robustness check we allow it in this section (see Maskus, 2000, for a recent survey of parallel imports). To keep things simple we return to the two-firm case in section 4.1, where firms segmented markets perfectly, and analyse how the possibility of arbitrage influences the result. In that section we showed that there exist only two sub-game perfect equilibria. Either the firms are dispersed (high trade costs) or they are agglomerated in the large region (for trade costs below the break point). The first thing we note is that the regions have the same price level when

the two firms are located in different regions, $P_j = \frac{a+c+ct}{3}$, $j \in R$. Without a difference in prices between the markets there are no arbitrage possibilities. When both firms are located in the large region, however, prices are $P_1 = \frac{a+2c}{3} < P_2 = \frac{a+2ct}{3}$ whenever $t > 1$. With a price differential “entrepreneurs” will buy the goods in market 1 and sell them to market 2. We now allow such arbitrage possibilities, but we assume there is a cost $s > 1$ (also of the iceberg type) for “entrepreneurs” to ship goods from market 1 to market 2. We do not impose any restriction on s ; it may be equal to t or differ from it (in either direction). To exclude arbitrage and still segment markets as much as possible, the duopolists face the restriction $P_2 \leq sP_1$. The Lagrangean of firm i at the agglomerated equilibrium is

$$L = \left[a - d(y_{i1} + y_{j1}^e) \right] y_{i1} - cy_{i1} + \left[a - b(y_{i2} + y_{j2}^e) \right] y_{i2} - ct y_{i2} + \lambda [sP_1 - P_2]$$
 and the

first-order conditions are $\frac{\partial L}{\partial y_{i1}} = a - 2dy_{i1} - dy_{j1}^e - c - \lambda sd = 0$ and

$$\frac{\partial L}{\partial y_{i2}} = a - 2by_{i2} - by_{j2}^e - ct + \lambda b = 0.$$
 Imposing symmetry in the first-order

conditions ($y_{i1} = y_{j1}^e$ and $y_{i2} = y_{j2}^e$) and solving for quantities give

$$y_{i1} = \frac{a-c-\lambda sd}{3d} \quad \text{and} \quad y_{i2} = \frac{a-ct+\lambda b}{3b}.$$
 The price in each region becomes

$$P_1 = \frac{a+2c+2\lambda sd}{3} \quad \text{and} \quad P_2 = \frac{a+2ct-2\lambda b}{3}.$$
 The complementary slackness

condition is $\lambda \geq 0$ ($= 0$ if $P_2 < sP_1$). If the cost of shipping goods for “entrepreneurs” is very high, then the constraint is not binding, λ equals zero and we are back in the analysis in section 4.1.

For lower s , arbitrage possibilities will arise. If the firms want to eliminate them, the constraint will bind and λ will be positive. We see that the two firms raise

the price of region 1 by cutting back on home market production, whereas they lower the price in region 2 by increasing their export market production. The difference in prices which, if left unchecked, creates arbitrage possibilities is thus eliminated by an appropriate adjustment of quantities. The result is a price differential exactly equal to the transport cost of “entrepreneurs”. We now want to analyse how the threat of parallel imports affects profits and the location-quantity game. Profit per firm if the constraint is binding is

$$(12) \quad \pi_i = \frac{(a-c+2\lambda sd)(a-c-\lambda sd)}{9d} + \frac{(a-ct-2\lambda b)(a-ct+\lambda b)}{9b}, \quad i = 1, 2.$$

Compared to the profit per firm without consumer arbitrage,

$$\pi_i^{TA} = \frac{(a-c)^2}{9d} + \frac{(a-ct)^2}{9b} \quad (\text{from equation 5}),$$

we see that the home (export) market profit increases (decreases). The net effect depends on the level of s as follows. The profit in (12) can be shown to be greater than π_i^{TA} provided that

$$\frac{a(s-1)+c(t-s)}{2(s^2d+b)} > \lambda > 0 \quad (\text{call the left-hand side } \bar{\lambda}).$$

We can solve for the *actual* value of λ by using the binding constraint and the expressions for prices above,

$$\text{which yields } \lambda = \frac{a(1-s)+2c(t-s)}{2(s^2d+b)}. \quad \text{For } \lambda \text{ to be positive we need } \frac{a+2ct}{a+2c} > s.$$

We have $\bar{\lambda} > \lambda$ if $s > \frac{2a+ct}{2a+c}$ and $\lambda > \bar{\lambda}$ if $\frac{2a+ct}{2a+c} > s$. The firms thus benefit

from the threat of arbitrage if $\frac{a+2ct}{a+2c} > s > \frac{2a+ct}{2a+c}$; if the last inequality is

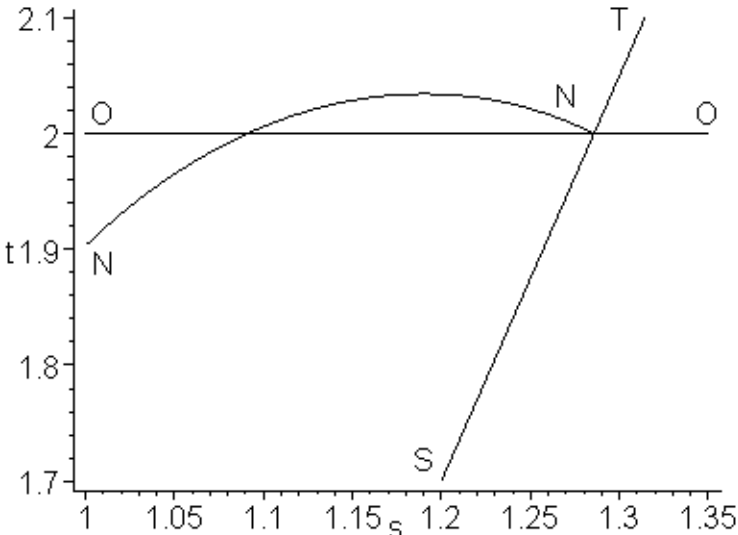
reversed they lose. We are now ready to investigate how the firms' choice of location is affected.

As in section 4.1 we want to see if there is a level of trade costs such that the sub-game perfect equilibrium of the location-quantity game changes from dispersion to agglomeration in the large region. There is a home market effect if the profit in (12) is greater than the profit in (4):

$$\frac{(a - c + 2\lambda sd)(a - c - \lambda sd)}{9d} + \frac{(a - ct - 2\lambda b)(a - ct + \lambda b)}{9b} - \frac{(a - 2c + ct)^2}{9b} - \frac{(a - 2ct + c)^2}{9d} > 0.$$

Since λ contains t the left-hand side is a polynomial of degree 2 in t . Inserting λ , simplifying and setting the result to zero, we arrive at an expression of the type $\alpha_2 t^2 + \alpha_1 t + \alpha_0 = 0$, where the coefficients are given in the Appendix. In the Appendix we also show that $\alpha_2 < 0$, $\alpha_1 > 0$ and $\alpha_0 < 0$. So far we have not been able to establish analytically when the quadratic equation has two real roots. Instead we have solved it numerically with Maple. Since the coefficient of t^2 is negative ($\alpha_2 < 0$) we know that the difference in profits is everywhere positive between the roots, i.e. the profit per firm is greater in the agglomerated equilibrium than the profit the firm in the small region makes in the dispersed equilibrium. Figure 6 illustrates the numerical solution.

Figure 6. The break point when consumer arbitrage is allowed.



The horizontal line OO is the break point without arbitrage possibilities (from section 4.1). The straight line ST shows the relationship between t and s for λ to be zero and is given by $\frac{a+2ct}{a+2c} = s$. For any given t , s has to be lower than the value read off the x-axis if the constraint is to be binding (in which case $\lambda > 0$). Any combination of t and s to the right of ST results in a non-binding constraint ($\lambda = 0$) and the location-quantity game is the same as the one we analysed in section 4.1. The curve NN is the larger of the two real roots that solves $\alpha_2 t^2 + \alpha_1 t + \alpha_0 = 0$; we used Maple to calculate it for various values of s . We only report roots to the left of ST due to the reason given above. The smaller root is always to the right of ST under our choice of parameter values.

Figure 6 is best understood as follows. Whenever the constraint is *not* binding, the break point is given by OO . Values of t above OO preserve the dispersed equilibrium and values below it result in agglomeration in the large region. Allowing arbitrage affects the break point, which is now given by NN . If the threat of arbitrage increases firms' profits in the agglomerated equilibrium (i.e. if $s > \frac{2a+ct}{2a+c}$), the firm in the small region will want to move sooner. In this case the break point is higher than without arbitrage possibilities. On the other hand, if profits are reduced ($\frac{2a+ct}{2a+c} > s$) it will move later (the break point is lower). The reason the firms may benefit from the threat of arbitrage is that it forces them to cut production levels in the large market (and increase them in the small one) to eliminate the price difference. It would seem that the possibility of arbitrage helps the firms to be more "monopoly-like" in the big region, increasing their profit, provided that trade costs for "entrepreneurs" are not too low.

6 Concluding Remarks

The important role played by backward and forward linkages in creating agglomerations of economic activity has been thoroughly analysed in the *new economic geography* (NEG). Most of this literature has relied on the Dixit-Stiglitz (1977) formalisation of monopolistic competition, whereas the implications of other forms of market structure for the allocation of industrial activity have largely been neglected. Furthermore, a general result (there are a few exceptions) is that agglomeration is catastrophic in nature. A marginal change in the level of trade costs either has no effect whatsoever on the allocation of industry, or it triggers a complete relocation of the whole industry to one single region. There are no intermediate cases; a region ends up with either all or nothing of the industry subject to agglomeration externalities once a critical level of trade costs is reached. Even though the use of the Dixit-Stiglitz set-up is a deliberate modelling choice due to its analytical simplicity, it still leaves the NEG empty-handed regarding the location of industries with an oligopolistic market structure. And the binary outcome of most NEG models fits poorly with casual empiricism.

The aim of this paper is to analyse how Cournot competition and *strategic interaction* between firms can give rise to differences in industrial structure when trade is liberalised. In a two-stage location-quantity game we first analyse a benchmark case with many firms and two identical regions. We show that the firms will *never* agglomerate in the same location if transportation is costly between regions. More importantly, we then analyse the effects of differences in market size and economic integration on the allocation of industrial activity. Assuming two regions of unequal size we find that spatial dispersion is a unique sub-game perfect equilibrium for high levels of trade costs. Lowering the costs of transportation beyond a critical value (the *break point*) triggers a relocation of

a firm from the small to the large region, giving rise to an agglomeration in the large market. It should be noted that the break point depends on the distribution of firms between the regions and that trade has to be liberalised even further for the next firm to move. A full-scale agglomeration of industry can only result from successively falling trade costs. The model thus displays stable dispersed asymmetric equilibria for intermediate levels of trade costs, something that is widely observed in the real world, but almost never in NEG models.

Krugman and Venables (1990) point out the gradual nature of agglomeration in settings such as ours, but include no explicit analysis of the matter. We also dig deeper into the welfare analysis. We show that the market provides an incentive for firms to agglomerate "too early". This is because they only take the effect on own profits into account, whereas both consumer surplus and other firms' profits are affected as well when a firm relocates. As a final extension we allow consumer arbitrage. Interestingly, this may affect the break point in an unexpected way. As firms strive to eliminate the price difference that gives rise to arbitrage possibilities, they cut back on production in the large market and increase production in the small one. The net effect on profits is positive if costs of shipping goods for consumers are not too low. The firms will then agglomerate in the large region at a higher level of trade costs compared to when there is no possibility of arbitrage.

References

- Anderson, S. P. and D. J. Neven (1991): "Cournot Competition Yields Spatial Agglomeration", *International Economic Review* **32**, 793-808.
- Baldwin, R. E., R. Forslid, P. Martin, G. I. P. Ottaviano and F. Robert-Nicoud (2002): *Economic Geography and Public Policy*, manuscript available at <http://heiwww.unige.ch/~baldwin/>
- Brander, J. (1981): "Intra-Industry Trade In Identical Commodities", *Journal of International Economics* **11**, 1-14.
- Brander, J. and P. Krugman (1983): "A Reciprocal Dumping Model of International Trade", *Journal of International Economics* **15**, 313-321.
- Combes, P.-Ph. (1997): "Industrial Agglomeration under Cournot Competition", *Annales d'Économie et de Statistique* **45**, 161-182.
- D'Aspremont, C., J. J. Gabszewicz and J.-F. Thisse (1979): "On Hotelling's Stability In Competition", *Econometrica* **47**, 1145-1150.
- D'Aspremont, C., J. J. Gabszewicz and J.-F. Thisse (1983): "Product Differences and Prices", *Economics Letters* **11**, 19-23.
- Dastidar, K. G. (2000): "Is a Unique Cournot Equilibrium Locally Stable?", *Games and Economic Behavior* **32**, 206-218.
- Dixit, A. (1986): "Comparative Statics for Oligopoly", *International Economic Review* **27**, 107-122.
- Dixit, A. and J. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity", *American Economic Review* **67**, 297-308.
- Forslid, R. and G. I. P. Ottaviano (2002): "An Analytically Solvable Core-Periphery Model", forthcoming in *Journal of Economic Geography*.
- Fujita, M., P. Krugman and A. J. Venables (1999): *The Spatial Economy. Cities, Regions, and International Trade*, MIT Press.
- Hamilton, J. H., J.-F. Thisse and A. Weskamp (1989): "Spatial Discrimination. Bertrand vs. Cournot in a Model of Location Choice", *Regional Science and Urban Economics* **19**, 87-102.

Head, K., T. Mayer and J. Ries (2002): "On the Pervasiveness of Home Market Effects", *Economica* **69**, 371-390.

Hotelling, H. (1929): "Stability In Competition", *Economic Journal* **39**, 41-57.

Kats, A. (1995): "More on Hotelling's stability in competition", *International Journal of Industrial Organization* **13**, 89-93.

Krugman, P. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade", *American Economic Review* **70**, 950-959.

Krugman, P. (1991): "Increasing Returns and Economic Geography", *Journal of Political Economy* **99**, 483-499.

Krugman, P. and A. J. Venables (1990): "Integration and the competitiveness of peripheral industry", in Bliss, C. and J. B. de Macedo (eds.): *Unity with diversity in the European economy*, Cambridge University Press.

Ludema, R. D. and I. Wooton (2000): "Economic geography and the fiscal effects of regional integration", *Journal of International Economics* **52**, 331-357.

Maskus, K. E. (2000): "Parallel Imports", *World Economy* **29**, 1269-1284.

Neary, P. (2001): "Of hype and hyperbolas: Introducing the new economic geography", *Journal of Economic Literature* **39**, 536-561.

Novshek, W. (1985): "On the Existence of Cournot Equilibrium", *Review of Economic Studies* **52**, 85-98.

Ottaviano, G. I. P., T. Tabuchi and J.-F. Thisse (2002): "Agglomeration and Trade Revisited", *International Economic Review* **43**, 409-435.

Pal, D. (1998): "Does Cournot competition yield spatial agglomeration?", *Economics Letters* **60**, 49-53.

Pflüger, M. (2001): "A Simple, Analytically Solvable Chamberlinian Agglomeration Model", IZA Discussion Paper No. 359.

Shapiro, C. (1989): "Theories of Oligopoly Behavior", in Schmalensee, R. and R. Willig (eds.): *Handbook of Industrial Organization*, Volume I. North-Holland.

Appendix

Solving the two-region model with many firms

Firm $i \in N_1$ maximises its *home* market profits,

$$\pi_{i1} = \left[a - b \left(y_{i1} + \sum_{k=1, k \neq i}^{n_1} y_{k1}^e + \sum_{k=1}^{n_2} y_{k1}^e \right) \right] y_{i1} - c y_{i1}. \quad \text{The first-order condition is}$$

$$\frac{\partial \pi_{i1}}{\partial y_{i1}} = 0 \Rightarrow a - 2b y_{i1} - b \sum_{k=1, k \neq i}^{n_1} y_{k1}^e - b \sum_{k=1}^{n_2} y_{k1}^e - c = 0. \quad \text{Total supply in region 1}$$

equals $Y_1 = \sum_{k=1}^{n_1} y_{k1} + \sum_{k=1}^{n_2} y_{k1}$, where the first sum is total supply by domestic firms

and the second is region 2's exports to region 1. Because all domestic firms are symmetric and produce the same amount of output we have

$$\sum_{k=1}^{n_1} y_{k1} = y_{11} + y_{21} + \dots + y_{i1} + \dots + y_{n_11} = n_1 y_{i1} \Rightarrow \sum_{k=1, k \neq i}^{n_1} y_{k1}^e = (n_1 - 1) y_{i1}. \quad \text{Inserting in the}$$

first-order condition above and solving for y_{i1} :

$$(A.1) \quad y_{i1} = \frac{a - c - b \sum_{k=1}^{n_2} y_{k1}^e}{b(1 + n_1)}.$$

Similarly, firm $j \in N_2$ maximises its *export* market profits

$$\pi_{j1} = \left[a - b \left(\sum_{k=1}^{n_1} y_{k1}^e + y_{j1} + \sum_{k=1, k \neq j}^{n_2} y_{k1}^e \right) \right] y_{j1} - c y_{j1}. \quad \text{The first-order condition is}$$

$$\frac{\partial \pi_{j1}}{\partial y_{j1}} = 0 \Rightarrow a - b \sum_{k=1}^{n_1} y_{k1}^e - 2b y_{j1} - b \sum_{k=1, k \neq j}^{n_2} y_{k1}^e - c = 0. \quad \text{Using the fact that all}$$

exporting firms are identical the first-order condition equals

$$(A.2) \quad y_{j1} = \frac{a - ct - b \sum_{k=1}^{n_1} y_{k1}^e}{b(1+n_2)}.$$

(A.1) and (A.2) form a system of $n_1 + n_2$ equations in $n_1 + n_2$ unknowns. Because all firms located in the *same* region are symmetric and each firm's belief about every other firm's output level is confirmed in equilibrium, this system reduces to two equations in two unknowns:

$$(A.3) \quad \begin{cases} y_{i1} = \frac{a - c - bn_2 y_{j1}}{b(1+n_1)} \\ y_{j1} = \frac{a - ct - bn_1 y_{i1}}{b(1+n_2)} \end{cases}$$

Solving for y_{i1} , y_{j1} and substituting back into the expressions for profits yield equations (1) and (2) in the text.

Is a symmetric distribution of firms an equilibrium when regions are of equal size?

$$\text{Total profits for firm } i \in N_1 \text{ is } \pi_i^T(\cdot) = \frac{[a - c + cn_2(t-1)]^2 + [a - ct - cn_2(t-1)]^2}{b(1+n_1+n_2)^2}.$$

How does a change in the number of firms located in region 1, affect profits in that region? Because the total number of firms is constant, the only way the number of firms in region 1 can increase is if a firm from region 2 relocates to region 1. The effects on profits for firms located in region 1 is obtained by

$$\text{totally differentiating } \pi_i^T, \quad d\pi_i^T = \frac{\partial \pi_i^T}{\partial n_1} dn_1 + \frac{\partial \pi_i^T}{\partial n_2} dn_2, \quad \text{where } dn \equiv dn_1 = -dn_2,$$

$$\frac{\partial \pi_i^T}{\partial n_1} = \frac{-2}{(1+n_1+n_2)} \left(\frac{[a-c+cn_2(t-1)]^2 + [a-ct-cn_2(t-1)]^2}{b(1+n_1+n_2)^2} \right), \text{ where the term}$$

within parenthesis is $\pi_i^T(\cdot)$, and

$$\frac{\partial \pi_i^T}{\partial n_2} = \frac{-2}{(1+n_1+n_2)} \left(\frac{[a-c+cn_2(t-1)]^2 + [a-ct-cn_2(t-1)]^2}{b(1+n_1+n_2)^2} \right) +$$

$$\frac{2c(t-1)}{(1+n_1+n_2)} \left(\frac{a-c+cn_2(t-1)}{b(1+n_1+n_2)} - \frac{[a-ct-cn_2(t-1)]}{b(1+n_1+n_2)} \right), \text{ where the second factor is}$$

$$y_{i1} - y_{i2}. \text{ Then } \frac{d\pi_i^T}{dn} = \frac{\partial \pi_i^T}{\partial n_1} - \frac{\partial \pi_i^T}{\partial n_2} = -\frac{2c(t-1)}{(1+n_1+n_2)} (y_{i1} - y_{i2}), \text{ which is negative if}$$

$y_{i1} > y_{i2}$. Inserting the expressions for the quantities and simplifying we have

$$\text{that } \frac{d\pi_i^T}{dn} = -\frac{2c^2(t-1)^2(1+2n_2)}{b(1+n_1+n_2)^2} < 0 \text{ since } n_2 \geq 0. \text{ Whenever a firm relocates}$$

from region 2 to region 1, each firm's profit in region 1 *decreases*. Analysing how profits in region 2 are affected by the relocation, we get

$$\frac{d\pi_j^T}{dn} = \frac{2c^2(t-1)^2(1+2n_1)}{b(1+n_1+n_2)^2} > 0. \text{ Profits for the firms remaining in region 2}$$

increase.

Parameter values used for the figures in the paper

Figures 1-6 all use the same parameter values: $a = 10$, $b = 4$, $c = 2$, and $d = 3$.

A condition ensuring that the break point is less or equal to the prohibitive level of trade costs

Starting from *any* symmetric equilibrium, $n_1^D = n_2^D$, denote the prohibitive level of trade costs facing firms in the small region $t^{pr} = \frac{a + c(n_1^D + k - 1)}{c(1 + n_1^D + k - 1)}$, where $k = 1, 2, \dots$ stands for the k :th firm considering a move to region 1. Setting $n_2^A - n_1^D = 1 - 2k$ in (10) in the text, and naming the right-hand side t^{bp} , we have $t^{bp} \leq t^{pr}$ if $n_1^D \leq \frac{d(2k - 1)}{(b - d)}$. In order to be sure that the break point is less than or equal to the prohibitive level, and hence that the profits in equations (8) and (9) are well defined, this inequality has to be satisfied.

The derivatives used in the analysis of consumer surplus

Consumer surplus is $CS_1 = \frac{d(Y_1)^2}{2}$ in region 1 and $CS_2 = \frac{b(Y_2)^2}{2}$ in region 2.

The derivatives with respect to Y_1 and Y_2 are dY_1 and bY_2 , respectively. Total

supply in region 1 is $Y_1 = n_1 \left(\frac{a - c + cn_2(t - 1)}{d(1 + n_1 + n_2)} \right) + n_2 \left(\frac{a - ct - cn_1(t - 1)}{d(1 + n_1 + n_2)} \right)$. Then

$$\frac{\partial Y_1}{\partial n_1} = \frac{a - c + cn_2(t - 1)}{d(1 + n_1 + n_2)} - n_1 \frac{[a - c + cn_2(t - 1)]}{d(1 + n_1 + n_2)^2} - n_2 \left(\frac{c(t - 1)(1 + n_1 + n_2) + [a - ct - cn_1(t - 1)]}{d(1 + n_1 + n_2)^2} \right) \text{ and}$$

$$\frac{\partial Y_1}{\partial n_2} = n_1 \left(\frac{c(t-1)(1+n_1+n_2) - [a-c+cn_2(t-1)]}{d(1+n_1+n_2)^2} \right) + \frac{a-ct-cn_1(t-1)}{d(1+n_1+n_2)} -$$

$$n_2 \frac{[a-ct-cn_1(t-1)]}{d(1+n_1+n_2)^2}. \text{ Thus } \frac{\partial Y_1}{\partial n_1} - \frac{\partial Y_1}{\partial n_2} = \frac{c(t-1)}{d(1+n_1+n_2)} \text{ and}$$

$$\frac{dCS_1}{dn_1} = \frac{dCS_1}{dY_1} \left(\frac{\partial Y_1}{\partial n_1} - \frac{\partial Y_1}{\partial n_2} \right) = \frac{Y_1 c(t-1)}{(1+n_1+n_2)}. \text{ Proceeding in the same way for region}$$

$$2 \text{ we have } \frac{dCS_2}{dn_1} = \frac{dCS_2}{dY_2} \left(\frac{\partial Y_2}{\partial n_1} - \frac{\partial Y_2}{\partial n_2} \right) = -\frac{Y_2 c(t-1)}{(1+n_1+n_2)}. \text{ Then the total effect on}$$

consumer surplus is as given in the text.

Showing that the total effect on consumer surplus always is positive

Inserting the expressions for each region's total supply and solving for t there are two cases. Either i) $t > \frac{a(n_1+n_2)(b-d) - c(bn_1 - dn_2)}{c(bn_2 - dn_1)}$ (if $bn_2 - dn_1 < 0$) or

ii) $t < \frac{a(n_1+n_2)(b-d) - c(bn_1 - dn_2)}{c(bn_2 - dn_1)}$ (if $bn_2 - dn_1 > 0$). In the former case the

right-hand side can be shown to be less than unity: the exact condition is $(a-c)(b-d)(n_1+n_2) > 0$, which always holds as $a > c$ and $b > d$ by assumption.

Since $t \geq 1$ by definition the inequality i) always holds and the net effect on consumer surplus is positive. In the latter case the right-hand side can be shown

to be greater than the strictest of the prohibitive levels, $\frac{a+cn_1}{c(1+n_1)}$. The exact

condition is $(a-c)[bn_1 - dn_2 + (n_1^2 + n_1n_2)(b-d)] > 0$, which holds as $a > c$, $b > d$ and $n_1 \geq n_2$. The inequality ii) thus always holds implying that $Y_1 > Y_2$ and

$$\frac{dTCS}{dn} > 0.$$

The coefficients of the polynomial of degree 2

$$\text{We have } \alpha_2 = \frac{-2bdc^2}{2(s^2d + b)} - 4bc^2 < 0,$$

$$\alpha_1 = \frac{[5ac(s-1) + 4c^2s]bd + 2(s^2d + b)[4ac(b-d) + 4c^2(b+d)]}{2(s^2d + b)} > 0, \text{ and}$$

$$\alpha_0 = \frac{5acsbd(1-s) - 8c(s^2d + b)[a(b-d) + cd] - 2bd[a^2(s-1)^2 + s^2c^2]}{2(s^2d + b)} < 0, \text{ as}$$

$s \geq 1$ and $b > d$.