DISCUSSION PAPER SERIES

No. 9046

AGGLOMERATION, TRADE AND SELECTION

Gianmarco Ottaviano

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP9046.asp

ISSN 0265-8003

AGGLOMERATION, TRADE AND SELECTION

Gianmarco Ottaviano, London School of Economics and CEPR

Discussion Paper No. 9046

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 9046

ABSTRACT

Agglomeration, Trade and Selection*

This paper studies how firm heterogeneity in terms of productivity affects the balance between agglomeration and dispersion forces in the presence of pecuniary externalities through a selection model of monopolistic competition with endogenous markups. It shows that firm heterogeneity matters. However, whether it shifts the balance from agglomeration to dispersion or the other way round depends on its specific features along the two defining dimensions of diversity: 'richness' and 'evenness'. Accordingly, the role of firm heterogeneity in selection models of agglomeration can not be fully understood without paying due attention to various moments of the underlying firm productivity distribution.

JEL Classification: F12, R11 and R12
Keywords: agglomeration, economic geography, heterogeneity, selection and trade

Gianmarco Ottaviano
Department of Economics and Centre
for Economic Performance
London School of Economics and
Political Science
Houghton Street
London WC2A 2AE
United Kingdom

Email: g.i.ottaviano@lse.ac.uk

*The author would like to thank Kristian Behrens, Giordano Mion and Diego Puga for helpful interactions on the topic of this paper.

Submitted 22 June 2012

1 Introduction

Does firm heterogeneity matter at the aggregate level? Since the seminal paper by Melitz (2003) on the associated 'new' gains from trade, the lack of systematic answers to such a fundamental question has created a gap between micro and macro applications. While firm heterogeneity has become a cornerstone in much recent micro modeling in international trade and, increasingly, in regional and urban economics, its impact on aggregate analyses has been so far rather subdued. Indeed, in one of the very few papers in international economics trying to bridge the gap between micro and macro, Arkolakis, Costinot and Rodrigues-Clare (2010) conclude that firm heterogeneity does not really matter much for the aggregate gains from trade as only the first moment of the firm productivity distribution affects those gains. It is true that this result holds only under very restrictive assumptions that grant a perfect aggregation property of the Melitz model, thus looking very much like an 'impossibility theorem'. Still, the result is striking and defines a useful benchmark ("zero co-ordinate") for future research on the aggregate implications of firm heterogeneity.

Against this background, the aim of the present paper is to tackle the above general question from the specific viewpoint of regional and urban economics by focusing on the relevance of firm heterogeneity for one of its main issues: the emergence of spatial imbalances (Fujita and Thisse, 2002; Combes, Mayer and Thisse, 2008). In particular, the paper addresses the specific question whether firm heterogeneity affects the aggregate balance between agglomeration and dispersion forces in the presence of pecuniary externalities through a selection model of monopolistic competition with endogenous markups. This is achieved by introducing firm heterogeneity *à la* Melitz and Ottaviano (2008) in a core-periphery framework *à la* Ottaviano, Tabuchi and Thisse (2002) noting that the endogeneity of markups as derived in those papers represents a major deviation from the restrictive set-up of Arkolakis, Costinot and Rodrigues-Clare (2010). The present paper also builds on Ottaviano (2011) but with major departures. The model in Ottaviano (2011) is a dynamic model of capital accumulation with forward-looking agents in closed economy. Differently, this paper proposes a dynamic model of migration with short-sighted agents in open economy. As in Ottaviano, Tabuchi and Thisse (2002), the economy is 'open' in terms of both goods trade and factor mobility while 'short sight' (due to heavy time discounting) is assumed in order to remove the possibility of self-fulfilling equilibria. These would add an extra layer of complexity beyond the scope of the present paper.

In the proposed model there are two locations that are identical in terms of their exogenous attributes. There are two factors of production: high-skill labor and low-skill labor. The former is freely mobile whereas the second is spatially immobile and evenly distributed between locations. There are two sectors: a perfectly competitive sector employing only low-skill labor to produce a homogenous good under constant returns to scale; and a monopolistic competitive sector employing both high-skill and low-skill labor to produce varieties of a horizontally differentiated good. In this sector high-skill labor is hired to design blueprints for the production of varieties and low-skill labor to produce the varieties according to those blueprints. In each period, high-skill workers first decide in which location to reside, then the monopolistic competitive firms decide whether and where to enter the market by hiring them. Subsequently high-skill workers engage in research and development with uncertain outcome in terms of the productivities of their blueprints. Once these productivities are revealed, firms decide whether to use the corresponding blueprints for production or just leave the market without producing. At the end of period, blueprints fully depreciate becoming useless. This admittedly stark assumption is made to abstract from sorting and focus on selection.

In this framework, the effects of heterogeneity on the balance between agglomeration and dispersion forces depend on which dimension of heterogeneity is affected and how it is affected. In particular, defining heterogeneity as 'diversity', heterogeneity is considered along two dimensions: 'richness' measures the 'number' of alternative productivity levels that can be drawn; 'evenness' is defined as the similarity between the probabilities with which those alternative productivity levels are drawn (Maignan, Ottaviano, Pinelli and Rullani, 2003). It is shown that, when productivity draws follow a Pareto distribution, the effects of more heterogeneity differ depending on whether more heterogeneity is achieved through more richness (captured by the "scale parameter" of the Pareto distribution) or more evenness (captured by the "shape parameter" of the Pareto distribution). There are two orders of reasons for this. First, under the Pareto distribution assumption, more richness comes with a higher chance of low productivity draws whereas more evenness comes with a higher chance of high productivity draws. Second, under the Pareto distribution assumption, the elasticity of the success rate of entry to tougher competition is affected by evenness but not by richness.

In terms of findings, the proposed model exhibits all the key feature of the model by Ottaviano, Tabuchi and Thisse (2002) without urban costs and of similar models in the 'new economic geography' tradition (see, e.g., Baldwin, Forslid, Martin, Ottaviano, Robert-Nicoud, 2003). In particular, trade barriers regulate the balance between agglomeration forces (market-size and cost-of-living effects) and dispersion forces (competition effect): starting with high enough trade barriers, trade liberalization shifts the spatial equilibrium from dispersion to agglomeration. The proposed model, however, introduces firm selection as an additional force affecting the balance between agglomeration and dispersion.

A first implication of this additional force is that, differently from Ottaviano, Tabuchi and Thisse (2002), the emergence of agglomerated equilibria is not catastrophic with the spatial economy suddenly moving from dispersion to full agglomeration when trade barriers fall below a certain threshold. It is, instead, smooth: as trade barriers gradually fall, at some point the dispersed allocation loses stability to two stable equilibria with partial agglomeration evenly spaced around it. These are initially in a neighborhood of the dispersed allocation. Then, as trade barriers keep on falling, they gradually move away from dispersion until the economy hits full agglomeration. Hence, thanks to selection among heterogeneous firms, the model is able to generate the realistic feature of partial agglomeration as a stable equilibrium outcome provided that trade barriers are neither too high nor too low. In this equilibrium, the larger location exhibits more entrants, more sellers and thus more product variety, lower average cost, lower average price, lower average markup. As all these features imply higher consumer surplus, the engineers' indifference condition that sustains the equilibrium holds due lower expected profits driven by a lower success rate of entry that more than offsets a higher average profit from successful entry.

A second implication concerns the impact of heterogeneity on the balance between agglomeration and dispersion forces for given trade barriers. More (cost-increasing) richness shifts the balance in favor of agglomeration forces. This happens because selection in the larger location gets weaker as worse productivity draws become possible. The impact of more (cost-decreasing) evenness is more complex. When the initial distribution of productivity draws is already rather even, more evenness shifts the balance in favor of agglomeration forces. Vice versa, when the initial distribution of productivity draws is rather uneven, more evenness shifts the balance in favor of dispersion forces. The reason for this is that, when the initial evenness is low, more evenness has a weak positive impact on the average profit differential and a strong negative effect on the entry success rate differential between locations, thus fostering dispersion. Vice versa, when the initial evenness is already high, more evenness has a strong positive effect on the average profit differential and a weak negative effect on the entry success rate differential, thus fostering agglomeration. Such ambiguity of the impact of more evenness is due to the fact that evenness affects the elasticity of the success rate of entry to the toughness of competition. Differently, more richness does not affect that elasticity.

The punchline of the paper is that firm heterogeneity matters for the balance between agglomeration and dispersion forces. However, whether it shifts the balance from agglomeration to dispersion or the other way round depends on its specific features along both the richness and the evenness dimensions.

There are a few related models in the spatial economics literature. These differ among themselves in terms of whether agents' heterogeneity is assumed to be revealed before or after their location decisions. Sorting models study how *ex ante* heterogenous agents self-select into locations of different sizes (Nocke, 2006; Baldwin and Okubo, 2006; Davis, 2010; Okubo, Picard and Thisse, 2010; Okubo and Picard, 2011).[1] The present paper differs from these models in that it studies selection, where heterogeneity materializes *ex post* after agents have already committed to their locations and where agents self-select in whatever economic activities are available in those locations. In this respect, the most closely related models are the ones put forth by Behrens and Robert-Nicoud (2012) and Behrens, Duranton and Robert-Nicoud (2010). The former is a selection model that also builds on Melitz and Ottaviano (2008) where *ex ante* identical individuals decide whether or not to move from a common rural hin-

[1] While this and other papers focus on firm heterogeneity on the supply side in terms of productivity, the distinctive feature of Okubo and Picard (2011) is their study of heterogeneity on the demand side in terms of tastes.

terland to cities. Their heterogeneity is revealed after this decision has been made and the decision itself is assumed to be irreversible so as to rule out sorting. They show that larger market size increases productivity not only through a finer division of labour driven by pecuniary externalities (richer availability of intermediates) but also through a selection process, whereas higher productivity increases market size by providing incentives for rural-urban migration. Behrens, Duranton and Robert-Nicoud (2010) analyze both sorting and selection in a model in which agglomeration is driven by technological externalities. They distinguish between *ex ante* heterogenity ('talent'), known to agents before they decide where to locate, and *ex post* heterogeneity ('luck'), revealed to agents after their location decisions have been made. Agents choose locations based on their talent and occupations in the chosen locations based on luck too. More talented agents stand a better chance of finding more productive occupations in larger locations and this complementarity between talent and market size leads to the sorting of more talented agents into larger markets. Then, tougher selection in more talented locations implies more productive occupations. Higher productivity, in turn, complements the agglomeration benefits of larger locations so that more talented markets are larger in equilibrium. Unlike Behrens, Duranton and Robert-Nicoud (2010) but just like Behrens and Robert-Nicoud (2012), the model in the present paper dispenses with *ex ante* heterogeneity to focus on selection. However, differently from Behrens and Robert-Nicoud (2012), it allows for reversible location decisions between sites whose urban vs. rural status is endogenously determined as an outcome of those decisions.[2] Moreover, differently from Behrens, Duranton and Robert-Nicoud (2010), in the model of the present paper the emergence of agglomeration is driven by pecuniary rather than technological externalities. Finally, while in Behrens, Duranton and Robert-Nicoud (2010) constant markups imply that, after conditioning out sorting and agglomeration, selection becomes independent of market size, in the present paper tougher selection is associated with larger market size as in Behrens and Robert-Nicoud (2012).[3]

[2] Behrens, Mion, Murata and Südekum (2011) make a similar assumption in their study of spatial frictions within a multi-location framework that allows for the joint determination of locations' sizes, productivities, markups, wages, consumption diversity, and the number and size distribution of firms. Their demand system is, however, different from the one in the present paper. Also different is the fact that the currently available version of the model always features a unique spatial equilibrium. It is, therefore, silent on the effects of firm heterogeneity on the balance between agglomeration and dispersion forces, which is the focus of the present analysis. Nonetheless, a two-region version of their model would also exhibit multiple equilibria, with comparative statics results for k and c_M qualitatively similar to the ones highlighted in the present paper but with opposite comparative statics results for τ due to the presence of urban costs.

[3] Combes, Duranton, Gobillon, Puga and Roux (2011) also extend the model of Melitz and Ottaviano (2008) to allow for agglomeration economies driven by technological externalities. Their aim is to derive a parsimonious framework to test the relative importance of selection and agglomeration in determining the spatial distribution of firm productivities. Following Melitz and Ottaviano (2008), they do not allow for labor mobility across locations. Nonetheless, in a separate on-line appendix (http://diegopuga.org/papers/selectagg_webapp.pdf), they show how their model can be further extended to introduce worker mobility, consumption amenities, and urban crowding costs, without at the same time affecting its key equilibrium

The rest of the paper is organized in three sections. Section 2 develops the model in an isolated economy to investigate how heterogeneity affects the relation between market size and selection. Section 3 extends the model to a two-location spatial economy with high-skill labor mobility to examine how heterogeneity affects the balance between agglomeration and dispersion forces. Section 4 concludes.

2 The Isolated Economy

There are L identical 'workers', each supplying a unit of low-skill labor inelastically. Accordingly, L is both the number of workers and the economy endowment of low-skill labor. There are also M identical 'engineers', each supplying a unit of high-skill labor inelastically. Accordingly, M is both the number of engineers and the economy endowment of high-skill labor.

2.1 Preferences

Workers and engineers all share the same individual preferences captured by the following quasi-linear quadratic utility function defined over a homogenous good and a continuum of varieties of a horizontally differentiated good

$$U^c = q_0^c + \alpha \int_0^N q^c(\omega)d\omega - \frac{\gamma}{2}\int_0^N \left(q^c(\omega)\right)^2 d\omega - \frac{\eta}{2}\left(\int_0^N q^c(\omega)d\omega\right)^2 \tag{1}$$

with q_0^c, N and $q^c(\omega)$ respectively denoting the individual consumption level of the homogenous good, the measure ('number') of available varieties of the differentiated good and the individual consumption level of variety ω. Parameters are all positive with γ measuring product differentiation.

2.2 Technology

There are two sectors, one supplying the homogenous good and the other supplying the varieties of the differentiated good. The homogenous good is produced under perfect competition employing low-skill labor as the only input. Specifically, production of a unit of homogenous output requires a worker. Marginal cost pricing then implies that the price of the homogenous good equals the low-skill wage. By choosing this good as numeraire, also the low-skill wage is set to unity.

The differentiated varieties are supplied by monopolistic competitive firms employing both low- and high-skill labor. In particular, a firm enters the market by hiring f_E engineers to design the blueprint of a differentiated variety and the

equations on which their empirical analysis is based. In so doing, they restrict their attention to a situation in which there exists a unique stable spatial equilibrium with (asymmetric) dispersion. Whether heterogeneity fosters agglomeration or dispersion is, thus, beyond the scope of their paper.

corresponding production process, whose implementation then requires hiring a number of workers proportionate to the desired scale of production.

Developing the blueprint of a variety and its production process is an activity with uncertain outcome in terms of productive efficiency. Specifically, while engineers are always certain to design new varieties, the unit worker requirements of the corresponding production processes are determined by random draws from some distribution. The timing of events is as follows. Firms decide whether to enter or not. If they decide to enter, they have to competitively bid for the given stock of M engineers. Once these have been allocated to the $N_E = M/f_E$ winning bidders, each of these entrants is assigned its unit worker requirement c as a random draw from a common continuous differentiable distribution with c.d.f. $G(c)$ over the support $[0, c_M]$. Based on their draws, entrants then decide whether to produce or not. Letting ρ denote the (endogenous) share of entrants that decide to produce ('success rate of entry'), the mass ('number') of producers equals $N = \rho N_E$ as the number of producers equals the number of varieties available for consumption. Following Melitz and Ottaviano (2008), Behrens and Robert-Nicoud (2012) and Behrens, Duranton and Robert-Nicoud (2010), the marginal product of workers φ is assumed to follow a Pareto distribution with shape parameter $k \geq 1$ and support $[1/c_M, \infty]$. This implies that the c.d.f. of the unit worker requirement $c = 1/\varphi$ is

$$G(c) = \left(\frac{c}{c_M}\right)^k, \ c \in [0, c_M] \tag{2}$$

The two parameters in (2) regulate the 'heterogeneity', or 'diversity', of cost draws. This has two dimensions (Maignan, Ottaviano, Pinelli and Rullani, 2003). First, the scale parameter c_M quantifies 'richness', defined as the measure ('number') of different unit labor requirements that can be drawn. Larger c_M leads to a rise in heterogeneity along the richness dimension, and this is achieved by making it possible to draw also larger unit worker requirements than the original ones. Second, the shape parameter k is an inverse measure of 'evenness', defined as the similarity between the probabilities of those different draws to happen. When $k = 1$, the unit worker requirement distribution is uniform on $[0, c_M]$ with maximum evenness. As k increases, the unit worker requirement distribution becomes more concentrated at higher unit worker requirements close to c_M: evenness falls. As k goes to infinity, the distribution becomes degenerate at c_M: all draws deliver a unit worker requirement c_M with probability one. Hence, smaller k leads to a rise in heterogeneity along the evenness dimension, and this is achieved by making low unit worker requirements more likely without changing the unit worker requirements that are possible. Accordingly, more richness (larger c_M) comes with higher average unit worker requirement ('cost-increasing richness'), more evenness (smaller k) comes with lower average unit worker requirement ('cost-decreasing evenness').

To finance their entry, firms borrow from workers. The functioning of the capital market is modeled as in Ottaviano, Tabuchi and Thisse (2002), but modified to deal with *ex ante* uncertainty on cost draws in the wake of Melitz and Ottaviano (2008). Specifically, as in Ottaviano, Tabuchi and Thisse (2002), borrowing and lending transactions are closed within each period and there is no interest rate.[4] As in Melitz and Ottaviano (2008), workers are assumed to hold identical balanced portfolios across all entrants, so they do not face any risk. Due to free entry, in equilibrium firms have to be indifferent between entering or not. This implies that, through competitive bidding, the engineers' remunerations absorb all expected profits from entry. The law of large numbers then ensures that these *ex ante* expected profits exactly match the *ex post* average profits of producers (themselves equal to the *ex ante* expected profit conditional on producing) times the share of entrants that decide to produce (itself equal to the *ex ante* probability that an entrant becomes a producer). Engineers' remunerations are, therefore, the same *ex ante* and *ex post*, and both *ex ante* and *ex post* workers' earnings on their lending are driven to zero. The corresponding budget constraints are thus

$$q_0^c + \int_0^N p(\omega) q^c(\omega) d\omega = 1 + \overline{q}_0^c \tag{3}$$

for a worker and

$$q_0^c + \int_0^N p(\omega) q^c(\omega) d\omega = \rho \widetilde{\pi} / f_E + \overline{q}_0^c \tag{4}$$

for an engineer, with $p(\omega)$, $\widetilde{\pi}$ and $\overline{q}_0^c$ respectively denoting the price of variety ω, the average producer profit and an initial endowment of the numeraire good that is assumed to be the same for all individuals and large enough to ensure its positive consumption.

2.3 Consumption

Due to the quasi-linearity of (1), the fact that workers' and engineers' incomes differ has no bearing on their consumption of the differentiated varieties: income differences are entirely transmitted to different consumption levels of the homogenous good.[5] Thus, the maximization of (1) subject to the budget constraints (3) or (4) gives the same FOC with respect to $q^c(\omega)$ for all individuals, implying an inverse demand relation that is independent of income

$$p(\omega) = \alpha - \gamma q^c(\omega) - \eta Q^c \tag{5}$$

where

$$Q^c = \int_0^N q^c(\omega) d\omega$$

is total individual consumption of the differentiated varieties.

[4] This rules out consumption smoothing and is consistent with the myopic adjustment process (29) that will be assumed for migration in the spatial economy as a result of heavy time discounting. A discussion of how the capital market would work were consumption smoothing allowed for can be found in Ottaviano (2011).

[5] For a discussion of income effects in an urban context with selection among heterogeneous firms, see Behrens, Mion, Murata and Südekum (2011).

Individual consumption can be obtained as follows. First, integrate (5) across products and solve for

$$Q^c = \frac{N\alpha - P}{\gamma + \eta N} \tag{6}$$

with $P = \int_0^N p(\omega)d\omega$. Hence, $\alpha > \widetilde{p} = P/N$ has to hold if any consumption has to take place at all ($Q^c > 0$). In other words the average price $\widetilde{p}$ must not be too high. Then, substituting (6) in (5) yields

$$q^c(\omega) = \frac{1}{\gamma}\left(\frac{\alpha\gamma + \eta N\widetilde{p}}{\gamma + \eta N} - p(\omega)\right)$$

Accordingly, varieties priced above the choke price

$$p^* \equiv \frac{\alpha\gamma + \eta N\widetilde{p}}{\gamma + \eta N} \tag{7}$$

are not bought ($q^c(\omega) = 0$) while individual inverse demand of any variety with price below the choke price can be written as

$$p(\omega) = p^* - \gamma q^c(\omega)$$

with corresponding total demand and total inverse demand respectively equal to

$$q(\omega) = q^c(\omega)\,(L+M) = \frac{p^* - p(\omega)}{\gamma}\,(L+M)$$

$$p(\omega) = p^* - \frac{\gamma}{L+M}q(\omega) \tag{8}$$

after aggregating across all workers and engineers.

The associated price elasticity of demand is

$$\left|\frac{dq(\omega)}{dp(\omega)}\frac{p(\omega)}{q(\omega)}\right| = \left(\frac{p^*}{p(\omega)} - 1\right)^{-1} = \left(\frac{p^*}{p^* - \frac{\gamma}{L+M}q(\omega)} - 1\right)^{-1} \tag{9}$$

This is an increasing function of own price $p(\omega)$ and a decreasing function of the choke price p^*. It is also an increasing function of the number of consumers $L + M$ as well as a decreasing function of the quantity demanded $q(\omega)$ and the extent of product differentiation γ. Note that the impact of changing p^* is weaker for higher $p(\omega)$. In turn, given (7), the choke price p^* is a decreasing function of the number of producers N as well as a decreasing function of their average price $\widetilde{p}$. Hence, any increase (decrease) in the number of producers as well as any decrease (increase) in their average price leads to a rise (fall) in the elasticity of demand. This makes competition tougher (softer) for all firms but disproportionately so for low price firms.

2.4 Production

Profit maximization by monopolistic competitive firms requires marginal revenue to match marginal cost. Given total inverse demand (8), the FOC for profit maximization by a firm with unit worker requirement c requires output to be

$$q(c) = \frac{L+M}{2\gamma}(p^* - c)$$

This uniquely identifies a cutoff unit worker requirement or, equivalently given unit wage, a cutoff marginal cost

$$c^* = p^* \tag{10}$$

such that $q(c^*) = 0$ and only firms whose unit worker requirement satisfies $c \leq c^*$ end up producing. Conditional on this cutoff, the unit worker requirement distribution of producers is $G^*(c) = G(c)/G(c^*) = (c/c^*)^k$ so that

$$\begin{array}{ll} \rho = \left(\frac{c^*}{c_M}\right)^k & N = \left(\frac{c^*}{c_M}\right)^k M/f_E \\ \widetilde{c} = \frac{k}{k+1}c^* & \widetilde{\sigma}^2 = \frac{k}{(k+1)^2(k+2)}(c^*)^2 \end{array} \tag{11}$$

Expression (10) can be used to rewrite firm output as

$$q(c) = \frac{L+M}{2\gamma}(c^* - c) \tag{12}$$

which can be plugged into total inverse demand (8) to obtain the corresponding price, markup, revenue and profit as functions of own and cutoff marginal costs:

$$\begin{array}{ll} p(c) = \frac{1}{2}(c^* + c) & \mu(c) = \frac{1}{2}(c^* - c) \\ r(c) = \frac{L+M}{4\gamma}\left[(c^*)^2 - c^2\right] & \pi(c) = \frac{L+M}{4\gamma}(c^* - c)^2 \end{array} \tag{13}$$

As more productive firms have lower marginal cost c, they are bigger in terms of both output and revenues. They also quote lower prices but have higher markups. As higher markups are associated with larger output, more productive firms generate more profits. Moreover, a lower cutoff c^* reduces the price, the output, the revenue and the profit of all firms. By increasing the elasticity of demand, it also reduces the markup, which makes c^* an inverse measure of the toughness of competition.

Based on (13) and (2), average price, average markup and average output evaluate to

$$\begin{array}{l} \widetilde{p} = \frac{2k+1}{2(k+1)}c^* \\ \widetilde{\mu} = \frac{c^*}{2(k+1)} \\ \widetilde{q} = \frac{L+M}{2\gamma(k+1)}c^* \end{array} \tag{14}$$

where $\widetilde{c}$ labels the average unit worker requirement, i.e. the mean unit worker requirement calculated for the conditional distribution $G^*(c)$ as only firms with

$c \leq c^*$ produce.[6] Analogously, average profit evaluates to

$$\widetilde{\pi} = \int_o^{c^*} \pi(c) dG^*(c) = \frac{L+M}{2\gamma\,(k+2)\,(k+1)} (c^*)^2 \tag{15}$$

The 'free entry condition', entailing that the engineers' remunerations absorb all expected profits from entry, can then be stated as

$$\rho \widetilde{\pi} = w f_E \tag{16}$$

where w is the high-skill wage.

Finally, (10), (11) and (7) imply the 'zero cutoff profit condition'

$$N = \frac{2\gamma(k+1)}{\eta} \frac{\alpha - c^*}{c^*} \tag{17}$$

which shows that $N > 0$ requires $\alpha > c^*$. All the rest given, a larger number of producers (larger N) is associated with tougher competition (lower c^*).

2.5 Equilibrium

The equilibrium of the closed economy is fully characterize by five equations in the following five unknowns: ρ, $\widetilde{\pi}$, w, N and c^*. The five equations are: (15), (16), (17), $\rho = G(c^*)$ and $N = \rho M / f_E$. These last two equations can be used together with (15) to substitute $\widetilde{\pi}$, ρ and N out of (16) and (17), reducing the characterization of the equilibrium to the solution of a system of two equations in the two remaining unknowns c^* and w:

$$\left(\frac{c^*}{c_M} \right)^k M / f_E = \frac{2\gamma(k+1)}{\eta} \frac{\alpha - c^*}{c^*} \tag{18}$$

$$w = \frac{L+M}{2\gamma\,(c_M)^k\, f_E} \frac{(c^*)^{k+2}}{(k+2)\,(k+1)} \tag{19}$$

There exists a unique value of c^* solving the 'zero cutoff profit condition' (18): its left-hand side is increasing whereas its right-hand side is decreasing in c^*. Some but not all entrants decide to produce (hence, there is 'selection') as long

[6] Results (13) and (14) concern additive markups, for firm c and on average respectively. The corresponding expressions for multiplicative markups are $p(c)/c = (c^* + c)/(2c)$ for firm c and $\int_0^{c^*} [p(c)/c]\, dG^*(c) = (2k-1)/[2(k-1)]$ on average. The average multiplicative markup is thus independent from the cutoff. This is explained by two opposing effects. On the one hand, tougher competition (lower c^*) forces high cost firms with low markups to exit, which on its own would increase the average markup. On the other hand, tougher competition reduces the markups of all surviving firms, which on its own would decrease the average markup. Under the Pareto assumption, the two effects exactly offset each other.

as that unique value of c^* falls in the interval $[0, c_M]$. A sufficient condition for this to happen is:

$$M > \frac{2\gamma(k+1)f_E}{\eta}\frac{\alpha - c_M}{c_M} \tag{20}$$

The argument behind (20) goes as follows. The left-hand side of (18) evaluates to 0 at $c^* = 0$ and increases in c^* for $c^* > 0$. Its right-hand side goes to infinity – and is thus larger that the left-hand side – in a neighborhood of $c^* = 0$ and decreases in c^* for $c^* > 0$. Accordingly, as (20) implies that at c_M the left-hand side is larger than the right-hand side, the two sides must achieve the same value for some value of c^* between 0 and c_M. Hence, the Pareto assumption and (20) together ensure the existence and uniqueness of an equilibrium cutoff marginal cost with selection.

The monotonicity properties of the left- and right-hand sides of (18) also imply that the comparative statics properties of the equilibrium cutoff are readily assessed: a larger endowment of engineers (larger M), a smaller entry costs (smaller f_E), weaker product differentiation (smaller γ), weaker preference for the differentiated varieties with respect to the homogenous good (smaller α and larger η) all lead to a smaller c^* and, therefore, to tougher competition. More workers have, instead, no bearing on competition. As for the impact of heterogeneity, if c_M increases, c^* has also to increase to satisfy (18) and, given that the right-hand side of this equation also adjusts, the increase in c^* is less than proportionate to the increase in c_M. Product variety falls accordingly. Lower k raises $(c^*/c_M)^k$ on the left-hand side of (18) given $c^*/c_M < 1$ under (20), and reduces $(k+1)$ on its right-hand side so that c^* has to fall to keep (18) satisfied. Hence, when heterogeneity grows because some additional bad draws become possible, selection gets weaker. Vice versa, when heterogeneity grows because the probability of the already existing good draws increases, selection gets tougher. Formally:

$$\frac{dc^*}{dc_M} > 0,\ \frac{dc^*}{dk} > 0$$

Given the equilibrium cutoff, the 'free entry condition'(19) then uniquely determines the equilibrium high-skill wage w as an increasing function of c^*:[7]

$$w = \frac{L+M}{2\gamma\,(c_M)^k\, f_E}\frac{(c^*)^{k+2}}{(k+2)\,(k+1)}$$

[7] Note that the logical sequence of solving for the equilibrium in the present model is opposite to the one required to solve the model by Melitz and Ottaviano (2008). There the equilibrium cutoff marginal cost is uniquely determined by the 'free entry condition' and the 'zero cutoff profit condition' then determines the equilibrium number of producers given the equilibrium cutoff. The reason for this sequence reversal is that in Melitz and Ottaviano (2008) there is only one type of labor employed in both sectors. This has two implications. First, as the unit input requirement and the sunk cost of entry are paid in the same labor units, the wage drops out of the free entry condition allowing for a closed form solution for the cutoff. Second, the number of entrants is endogenous, being determined by the allocation of the unique type of labor between sectors rather than by the exogenous endowment of a second factor.

Therefore, for any given c^*, a larger number of consumers (larger L or larger M), weaker product differentiation (smaller γ) and lower entry costs (smaller f_E) increase the engineers' remuneration w. Three parameters, however, have an ambiguous overall effect on the high-skill wage once their parallel impact on c^* is factored in. Larger M leads, on the one hand, to tougher competition and hence lower expected profit but, on the other hand, to a larger market, hence to larger expected firm size and larger expected profit. Smaller f_E leads, on the one hand, to tougher competition and hence lower expected profit but, on the other hand, to fewer engineers sharing those expected profits. Smaller γ leads, on the one hand, to tougher competition and thus lower expected profit but, on the other hand, to larger expected firm size and thus larger expected profits.

Once the equilibrium cutoff marginal cost and the corresponding number of firms $N = G(c^*)M/f_E$ are determined, welfare can be evaluated by noticing that the consumption choice that maximizes utility (1) yields the following indirect utility function:

$$V^c = I^c + \frac{1}{2\eta}\left(\alpha - c^*\right)\left(\alpha - \frac{k+1}{k+2}c^*\right) \tag{21}$$

where I^c is consumer income (equal to either 1 for workers or w for engineers) and $V^c - I^c$ is consumer surplus. To ensure positive demand levels for the numeraire, one has to assume that $I^c > N\int_0^N p\left(\omega\right) q^c(\omega) dG^*(\omega)$. Tougher competition (smaller c^*) is good for consumer surplus $V^c - I^c$, neutral for workers' income ($I^c = 1$) and bad for engineers' income ($I^c = w$). It is good for consumer surplus because selection leads to lower prices (due to both lower marginal costs and smaller markups) and richer product variety. Vice versa, it is bad for engineers' income due to two concurring events. First, tougher competition reduces the fraction of entrants that eventually produce. Second, the average profit of these producers falls.

2.6 Market size, heterogeneity and competition

How do market size and heterogeneity interact in determining the intensity of competition? Consider the effect of increasing market size by raising the number of engineers as these will be the mobile factor in open economy. As already argued, larger M leads to smaller c^*. The question then is how heterogeneity affects this negative relation. Implicit differentiation of (19) yields

$$\frac{\partial c^*}{\partial M} = -\frac{\eta}{2\gamma(k+1)\left(c_M\right)^k f_E}\frac{\left(c^*\right)^{k+2}}{k\left(\alpha - c^*\right) + \alpha} < 0$$

$$\frac{\partial^2 c^*}{\partial M \partial c_M} = \frac{\alpha\eta k}{2\gamma(k+1) f_E \left(c_M\right)^{k+1}}\frac{\left(c^*\right)^{k+1}}{\left(k\left(\alpha - c^*\right) + \alpha\right)^2} > 0$$

$$\frac{\partial c^*}{\partial M \partial k} = \frac{\left[\left(\alpha\left(k+2\right) - \left(k+1\right)c^*\right) - \alpha\left(k+1\right)\ln\left(\frac{c^*}{c_M}\right)\right]\eta\left(c^*\right)^{k+2}}{\left(c_M\right)^k 2\gamma f_E\left(k+1\right)^2\left((k+1)\alpha - kc^*\right)^2} > 0$$

The first expression above confirms that larger market size (larger M) makes competition tougher (smaller c^*). Heterogeneity affects the strength of this effect. According to the second expression, more cost-increasing richness (larger c_M) weakens the impact of larger market size on competition. The opposite holds for more cost-decreasing evenness (smaller k) as the third expressions shows that it dampens the competition enhancing drive of larger market size. Hence, heterogeneity fosters the tougher competition of larger markets only if it generates a better breed of firms.

3 The Spatial Economy

Consider now a spatial economy consisting of two locations, H and F. Each location is endowed with $L/2$ identical workers. Workers are geographically immobile and, as before, each worker supplies a unit of low-skill labor inelastically so that $L/2$ is both the number of workers and the endowment of low-skill labor in each location. The economy is also endowed with M identical engineers each supplying a unit of high-skill labor inelastically so that M is both the number of engineers and the economy endowment of high-skill labor. Differently from workers, engineers are geographically mobile and choose to reside in the location that offers them higher utility. The shares of engineers residing in locations H and F are called $s^H = \lambda$ and $s^F = 1 - \lambda$ respectively, with $\lambda \in [0, 1]$.

3.1 Preferences and technology

Workers and engineers share the same preferences in both locations, defined over a homogenous good and a continuum of differentiated varieties as captured by the utility function (1). They also have the same exogenous endowment $\overline{q}_0^c$ of the homogenous good. As before, this good is produced under perfect competition and constant returns to scale. Moreover, it is freely traded between locations. As a worker is needed to produce a unit of the homogenous good, choosing this good as numeraire implies that both its price and workers' wage are equalized to one across locations. The varieties of the differentiated good are produced by monopolistic competitive firms and their trade between locations is hampered by iceberg costs: $\tau > 1$ units have to be shipped for a unit to reach destination. Production faces similar technological constraints as in the isolated economy. Specifically, firms choose which location to enter by hiring f_E local engineers in order to develop a differentiated variety and the corresponding production process. The supply of the differentiated variety then requires hiring a number of workers proportionate to the desired scale of production. Accordingly, the number of entrants in the two locations is dictated by the spatial allocation of engineers: $N_E^H = s^H M/f_E$ and $N_E^F = s^F M/f_E$.

In a period the timing of events is as follows. Firms decide whether and where to enter taking the residential choices of engineers as given. If they decide to enter a location, they have to competitively bid for the stock of local engineers. Once these have been allocated to the winning bidders, in both locations each entrant is assigned its unit worker requirement c as a random draw from the common distribution $G(c)$ with support $[0, c_M]$. This distribution is the same across locations and is specified by (2). Based on their draws, entrants then decide whether to produce or not. In so doing, they are bound to produce in the location they have entered. This assumption is made to avoid sorting and focus on selection instead.

To finance their entry in a given location, firms can borrow from local workers only. These are assumed to hold riskless balanced portfolios across all entrants in their own location. Due to free entry and competitive bidding, in equilibrium firms have to be indifferent between entering or not. This implies that in each location the local engineers' remunerations absorb all expected profits from entry in that location. Accordingly, the *ex ante* expected profits from entry in location H (F) exactly match the *ex post* average profits of local producers $\widetilde{\pi}^H$ $(\widetilde{\pi}^F)$ times the share of local entrants that decide to produce ρ_D^H (ρ_D^F). The high-skill wage in H (F) therefore equals $w^H = \rho_D^H \widetilde{\pi}^H / f_E$ $(w^F = \rho_D^F \widetilde{\pi}^F / f_E)$ both *ex ante* and *ex post*, and both *ex ante* and *ex post* workers' earnings on their lending are driven to zero.

3.2 Consumption and production

Following Ottaviano, Tabuchi and Thisse (2002), the characterization of the equilibrium of the spatial economy proceeds in two steps. The first characterizes the indirect utility of engineers as a function of their spatial distribution λ. The second step endogenously determines the equilibrium spatial allocation of engineers as the outcome of their utility maximizing decisions.

Differently from the isolated economy, in the spatial economy the number of producers in and the number of sellers to each location may differ from one another. This is due to the presence of trade costs that allow only firms with low enough marginal costs to export. For parsimony, focus on location H knowing that analogous expressions apply to location F. Let p^H denote the price threshold for positive demand in location H. Then (7) implies

$$p^H = \frac{\alpha\gamma + \eta N^H \widetilde{p}^H}{\gamma + \eta N^H} \tag{22}$$

where N^H is the total number of firms selling to location H, consisting of domestic producers and foreign exporters, and $\widetilde{p}^H$ is the average price in location H across domestic producers and foreign exporters.

Let $p_D^H(c)$ and $q_D^H(c)$ represent the domestic levels of the profit maximizing price and quantity sold by a firm producing in location H with cost c. This firm may also decide to produce some output $q_X^H(c)$ to be exported at a delivered price $p_X^H(c)$. The two local markets in H and F are assumed to be segmented, which implies that firms independently maximize the profits earned from domestic and export sales. Then, by analogy with the closed economy, the profit maximizing prices and quantities are

$$
\begin{aligned}
p_D^H(c) &= \frac{1}{2}\left(c_D^H + c\right), \qquad & q_D^H(c) &= \frac{L/2 + s^H M}{2\gamma}\left(c_D^H - c\right), \\
p_X^H(c) &= \frac{\tau}{2}(c_X^H + c), \qquad & q_X^H(c) &= \frac{L/2 + s^F M}{2\gamma}\tau\left(c_X^H - c\right),
\end{aligned}
\tag{23}
$$

where $c_D^H = p^H$ and $c_X^H = p^F/\tau$ are the marginal cost cutoffs above which producers located in H are unable to sell in their domestic and export markets respectively. As trade barriers make it harder for exporters to break even relative to domestic producers, entrants with marginal cost c between 0 and c_X^H serve both the domestic and export markets. Entrants with marginal cost c between c_X^H and c_D^H serve only the domestic market. Entrants with marginal cost c above c_D^H are not able to serve any market and exit. Prices and output levels (23) then yield the following maximized profits from the domestic and export markets:

$$
\begin{aligned}
\pi_D^H(c) &= \frac{L/2 + s^H M}{4\gamma}\left(c_D^H - c\right)^2 \text{ for } c \in [0, c_D^H] \\
\pi_X^H(c) &= \frac{L/2 + s^H M}{4\gamma}\left(c_D^F - \tau c\right)^2 \text{ for } c \in [0, c_D^H/\tau]
\end{aligned}
\tag{24}
$$

where the second expression exploits the fact that $c_D^H = p^H$ and $c_X^H = p^F/\tau$ imply $c_X^H = c_D^F/\tau$. As in the isolated economy, the cutoffs measure the toughness of competition and summarize all the effects of market conditions relevant for firm performance in the corresponding locations.

Expected profits from entry in location H evaluate to $\rho_D^H \widetilde{\pi}^H = \rho_D^H \widetilde{\pi}_D^H + \rho_X^H \widetilde{\pi}_X^H$ with $\rho_D^H = G(c_D^H)$ and $\rho_X^H = G(c_X^H) = G(c_D^F/\tau)$. Expressions (24) and the distributional assumption (2) then yield

$$
\begin{aligned}
w^H &= \rho_D^H \widetilde{\pi}^H = \frac{\left(L/2 + s^H M\right)\left(c_D^H\right)^{k+2} + \left(L/2 + s^F M\right)\tau^{-k}\left(c_D^F\right)^{k+2}}{2\gamma(k+1)(k+2)\left(c_M\right)^k f_E} \qquad (25) \\
w^F &= \rho_D^F \widetilde{\pi}^F = \frac{\left(L/2 + s^F M\right)\left(c_D^F\right)^{k+2} + \left(L/2 + s^H M\right)\tau^{-k}\left(c_D^H\right)^{k+2}}{2\gamma(k+1)(k+2)\left(c_M\right)^k f_E}
\end{aligned}
$$

with the latter being the analogous expression holding for location F.

The corresponding indirect utilities evaluate to

$$
\begin{aligned}
V^H &= 1 + \frac{1}{2\eta}\left(\alpha - c_D^H\right)\left(\alpha - \frac{k+1}{k+2}c_D^H\right) \\
V^F &= 1 + \frac{1}{2\eta}\left(\alpha - c_D^F\right)\left(\alpha - \frac{k+1}{k+2}c_D^F\right)
\end{aligned}
$$

for workers and

$$
\begin{aligned}
V_E^H &= w^H + \frac{1}{2\eta}\left(\alpha - c_D^H\right)\left(\alpha - \frac{k+1}{k+2}c_D^H\right) \\
V_E^F &= w^F + \frac{1}{2\eta}\left(\alpha - c_D^F\right)\left(\alpha - \frac{k+1}{k+2}c_D^F\right)
\end{aligned}
\tag{26}
$$

for engineers. All these expressions clearly subsume the ones of the isolated economy as a special case when trade vanishes ($\tau \to \infty$) for identical stocks of workers and engineers in a location.

Given that $s^H = \lambda$ and $s^F = 1 - \lambda$, the foregoing expressions show that indirect utilities (26) are determined by three endogenous variables: the two cutoffs c_D^H and c_D^F, and the geographical distribution of engineers λ. However, c_D^H and c_D^F are themselves determined by λ. To see this, consider two sets of relations. First, given positive masses of entrants N_E^H and N_E^F in both locations, there are $G(c_D^H)N_E^H$ and $G(c_D^F)N_E^F$ domestic producers as well as $G(c_X^H)N_E^H$ and $G(c_X^F)N_E^F$ exporters satisfying $G(c_D^H)N_E^H + G(c_X^F)N_E^F = N^H$ and $G(c_D^F)N_E^F + G(c_X^H)N_E^H = N^F$. Recalling (2), $c_X^H = c_D^F/\tau$ and $c_X^F = c_D^H/\tau$ as well as $N_E^H = s^H M/f_E$ and $N_E^F = s^F M/f_E$, those relations can be rewritten as

$$
\begin{aligned}
N^H &= \left(\frac{c_D^H}{c_M}\right)^k s^H M/f_E + \left(\frac{c_D^H}{c_M}\right)^k \tau^{-k} s^F M/f_E \\
N^F &= \left(\frac{c_D^F}{c_M}\right)^k s^F M/f_E + \left(\frac{c_D^F}{c_M}\right)^k \tau^{-k} s^H M/f_E
\end{aligned} \tag{27}
$$

Moreover, (22) and (2) also imply

$$
\begin{aligned}
N^H &= \frac{2\gamma(k+1)}{\eta}\frac{\alpha - c_D^H}{c_D^H} \\
N^F &= \frac{2\gamma(k+1)}{\eta}\frac{\alpha - c_D^F}{c_D^F}
\end{aligned} \tag{28}
$$

For any given spatial allocation of engineers, equations (27) and (28) define a system of four equations in four unknowns (c_D^H, c_D^F, N^H, N^F) that implicitly determines them as functions of λ only. Just like in the isolated economy, uniqueness of these mappings is granted by the opposite monotonicity properties of the relations between the number of sellers and the marginal cost cutoffs embedded in (27) and (28) respectively. The system is unfortunately not amenable to analytical solutions. Exploiting uniqueness, it can nonetheless be solved numerically, finding the values of c_D^H and c_D^F that correspond to all possible values of λ, thus allowing for a numerical characterization of engineers' indirect utilities (26) as functions of their spatial allocation λ: $V_E^H(\lambda)$ and $V_E^F(\lambda)$.

3.3 Spatial equilibrium

The previous section has argued that two equilibrium marginal cost cutoffs c_D^H and c_D^F are uniquely associated with any given distribution of engineers λ. In turn, these cutoffs determine the expected profits from entry and, therefore, engineers' incomes. They also determine their consumer surpluses and, thus, their indirect utilities in the two locations. The next step is to determine the equilibrium spatial allocation of engineers based on the fact that their location choices must be utility maximizing.

Following the definition in Ottaviano, Tabuchi and Thisse (2002), the distribution $\lambda \in [0,1]$ is a *spatial equilibrium* if and only if no engineer can achieve higher indirect utility by changing location. As V_E^H and V_E^F are both functions of λ, the indirect utility differential between locations H and F is also a function of λ:

$$\Delta V(\lambda) \equiv V_E^H(\lambda) - V_E^F(\lambda)$$

A spatial equilibrium then arises at $\lambda \in (0,1)$ when

$$\Delta V(\lambda) = 0$$

or at $\lambda = 0$ when $\Delta V(0) \leq 0$, or at $\lambda = 1$ when $\Delta V(1) \geq 0$.

In order to study the stability of a spatial equilibrium, assume that local markets adjust instantaneously when some engineers move from one region to the other. Assume also a myopic adjustment process such that the driving force in the migration process is engineers' current indirect utility differential:

$$\dot{\lambda} \equiv d\lambda/dt = \begin{cases} \Delta V(\lambda) & \text{if } 0 < \lambda < 1 \\ \min\{0, \Delta V(\lambda)\} & \text{if } \lambda = 1 \\ \max\{0, \Delta V(\lambda)\} & \text{if } \lambda = 0 \end{cases} \tag{29}$$

when t is time. Crucially, assume finally that every period firms have first to bid for engineers and then to draw their unit worker requirements. This implies that draws are not carried over from one period to the next. The reason for this assumption is to focus on spatial selection while avoiding the different issue of spatial sorting.

Clearly, a spatial equilibrium implies $\dot{\lambda} = 0$. If $\Delta V(\lambda)$ is positive, some engineers will move from F to H; if it is negative, some will move in the opposite direction. A spatial equilibrium is *stable* for (29) if, for any marginal deviation from the equilibrium, the migration process brings the spatial allocation of engineers back to the original one. Therefore, the agglomerated configuration ($\lambda = 0$ or $\lambda = 1$) is always stable when it is an equilibrium while the dispersed configuration ($\lambda \in (0,1)$) is stable if and only if the slope of $\Delta V(\lambda)$ is nonpositive in a neighborhood of this point.

3.4 Core and periphery

The forces at work in the model are the ones highlighted by Ottaviano, Tabuchi and Thisse (2002), here enriched by firm selection *à la* Melitz and Ottaviano (2008). First, the immobility of workers is a dispersion force as long as there is trade between the two locations. The agglomeration force finds its origin in a demand effect generated by the preference for variety. If a larger number of firms are located, say, in location H, there are two effects at work. First, less varieties are imported. Second, the average price of all varieties sold in H is lower due not only to lower markups but also to lower marginal costs thanks to selection. Both effects, in turn, induce some consumers to migrate toward region H. The resulting increase in the number of consumers creates a larger local demand for the differentiated good, which locally raises expected profits (hence, engineers' wages) promoting more entry and this maps into a larger number of local producers. Differently from Ottaviano, Tabuchi and Thisse (2002), the increase in the number of producers is less than proportionate to the increase in the number of entrants and this is due to tougher selection in the expanding market.

The hallmark of Ottaviano, Tabuchi and Thisse (2002) and similar models in the 'new economic geography' tradition is the fact that trade barriers regulate the balance between agglomeration and dispersion forces: starting with high enough trade barriers, trade liberalization breeds agglomeration (see, e.g., Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud, 2003). Figure 1 shows that this is true also in the present model. As in Ottaviano, Tabuchi and Thisse (2002), with high trade barriers the dispersed allocation ($\lambda = 1/2$) is the only spatial equilibrium. As trade barriers fall (smaller τ), this allocation becomes unstable and agglomeration in either location emerges as the only spatial equilibrium. However, differently from Ottaviano, Tabuchi and Thisse (2002) the emergence of the agglomerated equilibrium is not catastrophic driving the economy from dispersion ($\lambda = 1/2$) straight to full agglomeration ($\lambda = 0$ or $\lambda = 1$). As trade barriers gradually fall, at some point the dispersed allocation loses stability to two stable equilibria with partial agglomeration ($0 < \lambda < 1/2$ or $1/2 < \lambda < 1$) evenly spaced around it. These are initially in a neighborhood of the dispersed allocation. Then, as trade barriers keep on falling, they gradually move away from dispersion until the economy hits full agglomeration.

Hence, thanks to selection among heterogeneous firms, the model is able to generate the realistic feature of partial agglomeration as a stable equilibrium outcome provided that trade barriers are neither too high nor too low. In this equilibrium, the larger location exhibits more entrants, more sellers and thus more product variety, lower average cost, lower average price, lower average markup. As all these features imply higher consumer surplus, indifference $\Delta V(\lambda) = 0$ is sustained by lower expected profits (smaller w) due to a lower success rate of entry (smaller ρ_D) that more than offset any higher average profit from successful entry (larger $\widetilde{\pi}$).

Note that introducing urban costs as in Ottaviano, Tabuchi and Thisse (2002) would generate redispersion as trade barriers get very low.[8] In the partial agglomeration equilibrium, it would also lead not only to higher average profits but also to higher expected profits in the larger location in order to compensate for its higher urban costs.

3.5 Agglomeration, dispersion and heterogeneity

What is the impact of heterogeneity on core-periphery patterns? The analysis of the isolated economy has shown that more cost-increasing richness (larger c_M) weakens the impact of larger market size on selection. The opposite holds

[8] This would be consistent with the predictions of a two-region model *à la* Behrens, Mion, Murata and Südekum (2011).

for more cost-decreasing evenness (smaller k). Therefore, heterogeneity must affect the balance between dispersion and agglomeration forces in the spatial economy, and the effect of heterogeneity must be different depending on whether it increases or decreases the chances of high unit worker requirement draws.

Figure 2 shows the effects of more cost-increasing richness (larger c_M) on the patterns depicted in Figure 1. High, mid and low heterogeneity corresponds to large, intermediate and small c_M respectively. In this case more heterogeneity shifts the balance in favor of agglomeration forces. This happens because selection in the larger location gets weaker as worse unit worker requirement draws become possible.

Figures 3 and 4 show the effects of more cost-decreasing evenness (smaller k). High, mid and low heterogeneity corresponds to small, intermediate and large k respectively. When the initial distribution of unit worker requirement draws is already rather even (Figure 3), additional evenness shifts the balance in favor of agglomeration forces. Vice versa, when the initial distribution of unit worker requirement draws is rather uneven (Figure 4), additional evenness shifts the balance in favor of dispersion forces. The reason for this is that, when initial evenness is low (large k), more evenness (smaller k) has a weak positive effect on the average profit differential ($\widetilde{\pi}^H$ vs. $\widetilde{\pi}^F$) and a strong negative effect on the success rate differential (ρ_D^H vs. ρ_D^F), thus fostering dispersion. Vice versa, when the initial evenness is already high, more evenness has a strong positive effect on the average profit differential and a weak negative effect on the success rate differential, thus fostering agglomeration. This does not happen in the case of more richness as larger c_M does not affect the elasticity of the success rate of entry to competition given that $d \ln \rho_D^H / d \ln c_D^H = k c_D^H$ and $d \ln \rho_D^F / d \ln c_D^F = k c_D^F$.

Differently from Ottaviano, Tabuchi and Thisse (2002) and related models with homogenous firms, the present model does not allow for the derivation of a closed form 'break point', which is typically useful in order to better understand numerical findings such as those reported in the figures (Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud, 2003). The reason for such impossibility is that the value of the cutoff $c_D^H = c_D^F = c^*$ in the symmetric equilibrium $\lambda = 1/2$ is itself an implicit function of trade costs. To see this, note that the symmetric equilibrium cutoff satisfies the 'zero cutoff profit condition'

$$\frac{1+\tau^{-k}}{2} \left(\frac{c^*}{c_M} \right)^k M/f_E = \frac{2\gamma(k+1)}{\eta} \frac{\alpha - c^*}{c^*} \tag{30}$$

where $\rho^* = (c^*/c_M)^k$ is the corresponding symmetric success rate of entry. Condition (30) subsumes the isolated economy 'zero cutoff profit condition' (18) as a special case when trade vanishes ($\tau \to \infty$) for identical stocks of workers and engineers in a location. This implies that the open economy cutoff qualitatively exhibits the same comparative statics properties with respect to heterogeneity as the closed economy one:

$$\frac{dc^*}{dc_M} > 0, \ \frac{dc^*}{dk} > 0 \tag{31}$$

On top of these common properties, (30) also implies that lower trade costs lead to tougher competition

$$\frac{dc^*}{d\tau} > 0 \tag{32}$$

as in Melitz and Ottaviano (2008).

While it is impossible to derive a closed form 'break point', intuition can still be gauged by discussing how the key equations determining engineers' indirect utility differential between locations behave for small deviations from the symmetric equilibrium. In particular, following Behrens and Robert-Nicoud (2012), consider a linearization of the equilibrium conditions (25), (26), (27) and (28) around symmetry, recalling $s^H = \lambda$ and $s^F = 1 - \lambda$. The indirect utility differential of engineers is determined by two components: their income differential and their consumer surplus differential. Using (25), the income differential can be linearized around symmetry to obtain

$$\frac{w^H - w^F}{w^*} = \left(\frac{1-\tau^{-k}}{1+\tau^{-k}}\right)\left(4\frac{M}{L+M}\frac{\lambda - 1/2}{1/2} + \frac{\rho^H - \rho^F}{\rho^*} + 2\frac{c_D^H - c^F}{c^*}\right) \tag{33}$$

On the right-hand side, inside the second parentheses, the first term captures the part of differential income that is driven by differential market size: all else equal, profits are higher in the larger location because this hosts a larger number of consumers ('backward linkage'; Fujita, Krugman and Venables, 1999). This is an agglomeration force and its strength depends on the share of mobile consumers $M/(L+M)$. The second and third terms reveal, instead, the presence of dispersion forces. These stem from the fact that in the larger location competition is tougher (the cutoff is lower), which reduces the success rate of entry (second term) as well as the average profits of successful entrants (third term). From the viewpoint of engineers' income, selection therefore acts a dispersion force.

To see that larger market size actually implies tougher competition (lower cutoff) and thus a lower success rate of entry, one has to remember that, on the one hand, by definition $\rho^H = \left(c^H/c_M\right)^k$ and $\rho^F = \left(c^F/c_M\right)^k$ so that linearization around symmetry yields

$$\frac{\rho^H - \rho^F}{\rho^*} = k\frac{c_D^H - c_D^F}{c^*} \tag{34}$$

which shows that a lower success rate is associated with a lower cutoff. On the other hand, linearizing (27) and (28) around symmetry gives

$$\frac{c_D^H - c_D^F}{c^*} = -2\frac{1-\tau^{-k}}{1+\tau^{-k}}\frac{1}{k+\frac{\alpha}{\alpha - c^*}}\frac{\lambda - 1/2}{1/2} \tag{35}$$

which shows that the cutoff is indeed lower in the larger location. Using these results, engineers' differential income (33) can be rewritten as

$$\frac{w^H - w^F}{w^*} = 4\left(\frac{1-\tau^{-k}}{1+\tau^{-k}}\right)\frac{M}{L+M}\frac{\lambda - 1/2}{1/2} - 2\left(\frac{1-\tau^{-k}}{1+\tau^{-k}}\right)^2\frac{k+2}{k+\frac{\alpha}{\alpha - c^*}}\frac{\lambda - 1/2}{1/2} \tag{36}$$

where the first and second terms on the right-hand side capture agglomeration and dispersion forces respectively. Note that, while for the agglomeration force driven by the 'backward linkage' the location of engineers matters as a share of the overall number of consumers, for the dispersion forces driven by competition and selection it matters as a share of the overall number of sellers.

Expression (36) reveals that lower trade costs τ weaken both agglomeration and dispersion forces until the income differential vanishes with free trade ($\tau = 1$). However, lower trade costs weaken agglomeration forces less than dispersion forces so that, if only income mattered, engineers would agglomerate in a single location provided trade costs were low enough (but still not negligible). This is due both to the explicit impact of lower τ on the ratio $\left(1-\tau^{-k}\right)/\left(1+\tau^{-k}\right)$ and to its implicit impact (32) on the cutoff c^*. The explicit impact disproportionately weakens the dispersion forces with respect to the agglomeration forces due to the squared power in the second term on the right hand side of (36). The implicit impact also weakens the dispersion forces but without affecting the agglomeration forces as the cutoff does not appear in the first term on the right-hand side of (36).

Heterogeneity affects (36) in different ways depending on whether richness or evenness change. By (31), more cost-increasing richness (larger c_M) increases the cutoff c^*, thus weakening the dispersion forces captured by the second term on the right hand side of (36). This happens because an increase in c^* relaxes competition and selection. More cost-decreasing evenness (smaller k) works, instead, along three channels. First, it reduces $\left(1-\tau^{-k}\right)/\left(1+\tau^{-k}\right)$, thus weakening dispersion more than agglomeration forces. Second, it decreases c^*, thus strengthening dispersion forces. Third, lower k has also an impact of ambiguous sign on $(k+2)/[k+\alpha/(\alpha-c^*)]$. As discussed below, these concurrent effects explain why lower k fosters or hampers agglomeration depending on initial parameter values.

Turning to consumer surplus, linearizing (26) around symmetry yields

$$\frac{S_E^H - S_E^F}{S_E^*} = -\left(\frac{c^*}{\alpha - c^*} + \frac{\frac{k+1}{k+2}c^*}{\alpha - \frac{k+1}{k+2}c^*}\right)\frac{c_D^H - c_D^F}{c^*}$$

which shows that differential surplus arises from the differential toughness of competition ($c_D^F - c_D^H$) as tougher competition is associated with more sellers (varieties) and lower prices. By (35), competition is tougher in the larger location. Accordingly, differential surplus can be rewritten as a function of differential market size:

$$\frac{S_E^F - S_E^F}{S_E^*} = 2\frac{1-\tau^{-k}}{1+\tau^{-k}}\frac{1}{k+\frac{\alpha}{\alpha-c^*}}\left(\frac{c^*}{\alpha - c^*} + \frac{\frac{k+1}{k+2}c^*}{\alpha - \frac{k+1}{k+2}c^*}\right)\frac{\lambda - 1/2}{1/2} \tag{37}$$

This shows that the larger location enjoys higher consumer surplus, the more so the higher trade costs are, due to both the positive explicit impact of these costs through larger $\left(1-\tau^{-k}\right)/\left(1+\tau^{-k}\right)$ and their positive implicit impact through larger c^*. From the point of view of engineers' consumption, the access to a richer set of cheaper varieties offered by the larger market always works as an agglomeration force ('forward linkage'; Fujita, Krugman and Venables, 1999). Hence, in line with Behrens and Robert-Nicoud (2012), selection acts as an agglomeration force on the demand side and as a dispersion force on the supply side.

Heterogeneity affects differential surplus (37) in two ways. First, by (31), more cost-increasing richness (larger c_M) increases the cutoff c^*, thus strengthening the agglomeration force on the demand side. Second, more cost-decreasing evenness (smaller k) reduces both $\left(1-\tau^{-k}\right)/\left(1+\tau^{-k}\right)$ and c^*, thus weakening that agglomeration force. Differently from the supply side, smaller k has no opposing effects on the demand side. Hence, its ambivalent impact on the balance between agglomeration and dispersion forces depicted in Figures 3 and 4 is entirely explained by the opposing effects on the supply side. In particular, as already pointed out, when initial evenness is already low, the additional effect of even more evenness is weakly positive on the average profit differential but strongly negative on the success rate differential, which fosters dispersion. Vice versa, when initial evenness is already high, the effect of more evenness is strongly positive on the average profit differential and weakly negative on the success rate differential, which fosters agglomeration.

Finally, linearization around symmetry is also useful to gain intuition on the reasons why, differently from Ottaviano, Tabuchi and Thisse (2002), the present model features stable asymmetric interior equilibria. In particular, the linearization of (27) yields

$$\frac{N^H-N^F}{N^*}=2\frac{1-\tau^{-k}}{1+\tau^{-k}}\frac{\lambda-1/2}{1/2}+\frac{\rho^H-\rho^F}{\rho^*} \tag{38}$$

which shows that the differential number of sellers is determined by a term proportional to differential market size $(\lambda-1/2)$ as well as by the differential success rate of entry $(\rho^H-\rho^F)$. Expressions (34) and (35) then allow to rewrite (38) as

$$\frac{N^H-N^F}{N^*}=2\frac{1-\tau^{-k}}{1+\tau^{-k}}\frac{\frac{\alpha}{\alpha-c^*}}{k+\frac{\alpha}{\alpha-c^*}}\frac{\lambda-1/2}{1/2} \tag{39}$$

This can be compared with the analogous expression in Ottaviano, Tabuchi and Thisse (2002)

$$\frac{N^H-N^F}{N^*}=2\frac{\lambda-1/2}{1/2}$$

which is due to the fact that not only all firms are identical and all patents are implemented but also restrictions on parameters are imposed to ensure that all firms sell in both locations. In (39) the result that in the present model not all patents are implemented is captured by the extra factor $[\alpha/\left(\alpha-c^*\right)]/[k+\alpha/\left(\alpha-c^*\right)]<1$. The fact that not all firms export is captured by the extra factor $\left(1-\tau^{-k}\right)/\left(1+\tau^{-k}\right)$. Both extra factors dampen the agglomeration force on the demand side as well as agglomeration and, to a lesser extent, dispersion forces on the supply side, which explains why asymmetric interior stable equilibria are feasible in the present model while they are not in Ottaviano, Tabuchi and Thisse (2002).

4 Conclusion

This paper has investigated how firm heterogeneity affects the balance between agglomeration and dispersion forces in the presence of pecuniary externalities in a model of monopolistic competition with endogenous markups. In so doing, it has proposed a model that exhibits all the key features of the model by Ottaviano, Tabuchi and Thisse (2002) and similar models in the 'new economic geography' tradition. In particular, trade barriers regulate the balance between agglomeration forces (market-size and cost-of-living effects) and dispersion forces (competition effect): starting with high enough trade barriers, trade liberalization shifts the spatial equilibrium from dispersion to agglomeration.

In the proposed model, however, firm selection acts as an additional force affecting the balance between agglomeration and dispersion. A first implication of this additional force is that, differently from Ottaviano, Tabuchi and Thisse (2002), the emergence of agglomeration is not catastrophic, which is more realistic. A second implication is that firm heterogeneity matters for the balance between agglomeration and dispersion forces. However, whether it shifts the balance from agglomeration to dispersion forces or the other way round depends on its specific features along the two defining dimensions of richness and evenness. This result shows that firm heterogeneity matters for aggregate outcomes: its role in spatial models with selection can not be fully understood without paying due attention to various moments of the underlying firm productivity distributions. A similar conclusion is reached by Okubo and Picard (2011) for a sorting model with taste heterogeneity.

References

[1] Baldwin R., R. Forslid, P. Martin, G. Ottaviano and F. Robert-Nicoud (2003) *Economic Geography and Public Policy* (Princeton: Princeton University Press).

[2] Baldwin R. and T. Okubo (2006) Heterogeneous firms, agglomeration and economic geography: Spatial selection and sorting, *Journal of Economic Geography* 6, 323-346.

[3] Behrens K. and F. Robert-Nicoud (2012) Survival of the fittest in cities: Agglomeration, selection and polarisation. CIRPÉE Discussion Paper n. 09-19. Revised February 2012.

[4] Behrens K., Duranton G. and F. Robert-Nicoud (2010) Productive cities: Sorting, selection and agglomeration, CEPR Discussion Paper n. 7922.

[5] Behrens K., Mion G., Murata Y. and J. Südekum (2011) Spatial frictions, CEPR Discussion Paper n. 8572.

[6] Combes P.-P., T. Mayer and J.-F. Thisse (2008) *Economic Geography* (Princeton: Princeton University Press).

[7] Combes P.-P., Duranton G., Gobillon L., Puga D. and S. Roux (2011) The productivity advantages of large cities: Distinguishing agglomeration from firm selection, CEPR discussion paper 7191, March 2009. Revised September 2011. Forthcoming: *Econometrica*.

[8] Davis D. (2010) A spatial knowledge economy, Columbia University, mimeo.

[9] Fujita M. and J.-F. Thisse (2002) *Economics of Agglomeration* (Cambridge: Cambridge University Press).

[10] Fujita M., P. Krugman and A. Venables (1999) The Spatial Economy. Cities, Regions and International Trade (Cambridge MA: MIT Press).

[11] Maignan C., Ottaviano G., Pinelli D. and F. Rullani (2003) Bio-ecological diversity vs. socio-economic diversity: A comparison of existing measures, FEEM Nota di Lavoro n. 12.2003.

[12] Melitz, M. (2003) The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica* 71, 1695-1725.

[13] Melitz M. and G. Ottaviano (2008) Market size, trade and productivity, *Review of Economic Studies* 75, 295-316.

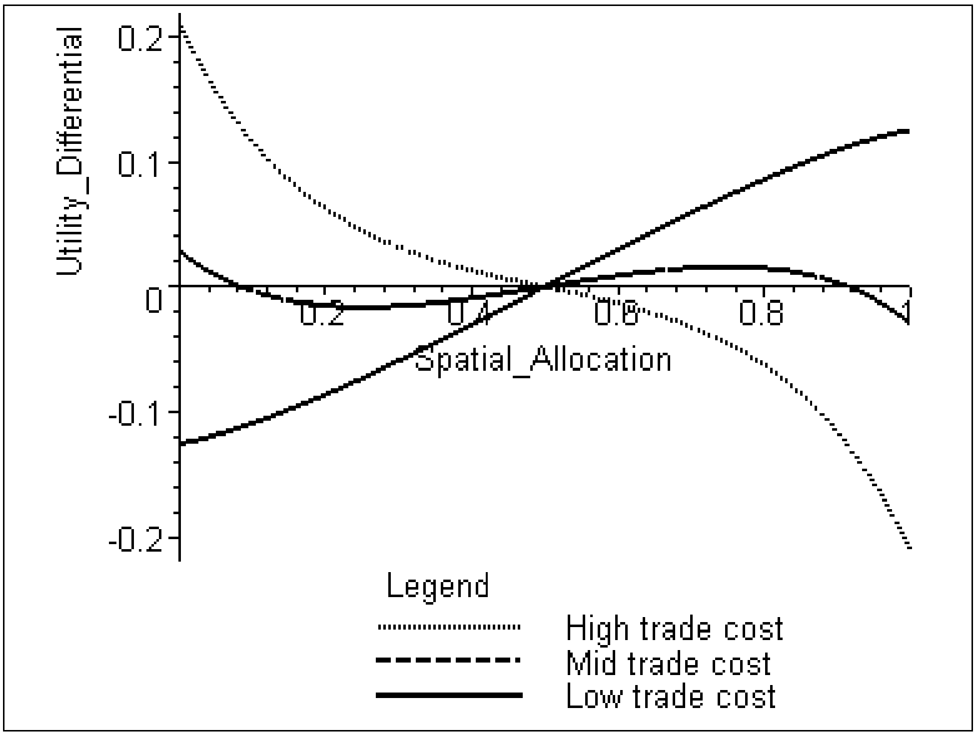
[14] Nocke V. (2006) A gap for me: Entrepreneurs and entry, *Journal of the European Economic Association* 4, 929-956.

[15] Okubo T. and P. Picard (2011) Firms location under demand heterogeneity, CREA Discussion Paper Series n. 11-07.

[16] Okubo T., P. Picard and J.-F. Thisse (2010) The spatial selection of heterogeneous firms, *Journal of International Economics* 82, 230-237.

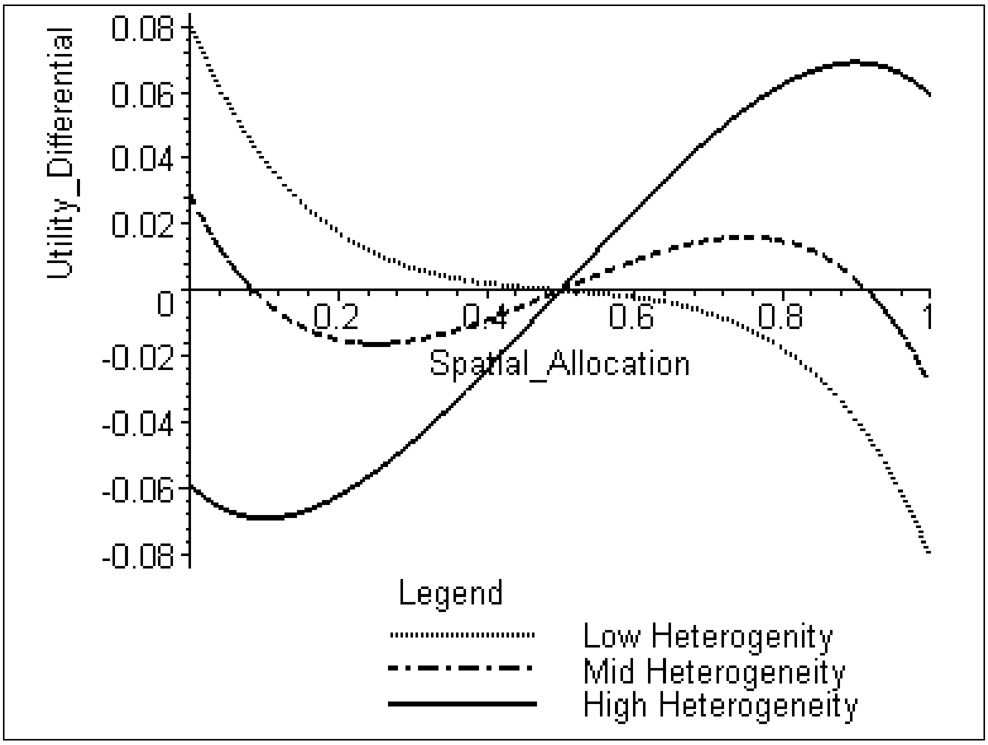
[17] Ottaviano G. (2011) Firm heterogeneity, endogenous entry, and the business cycle, NBER Working Paper n. 17433.

[18] Ottaviano G., T. Tabuchi and J.-F.Thisse (2002) Agglomeration and trade revisited, *International Economic Review* 43, 409-435.



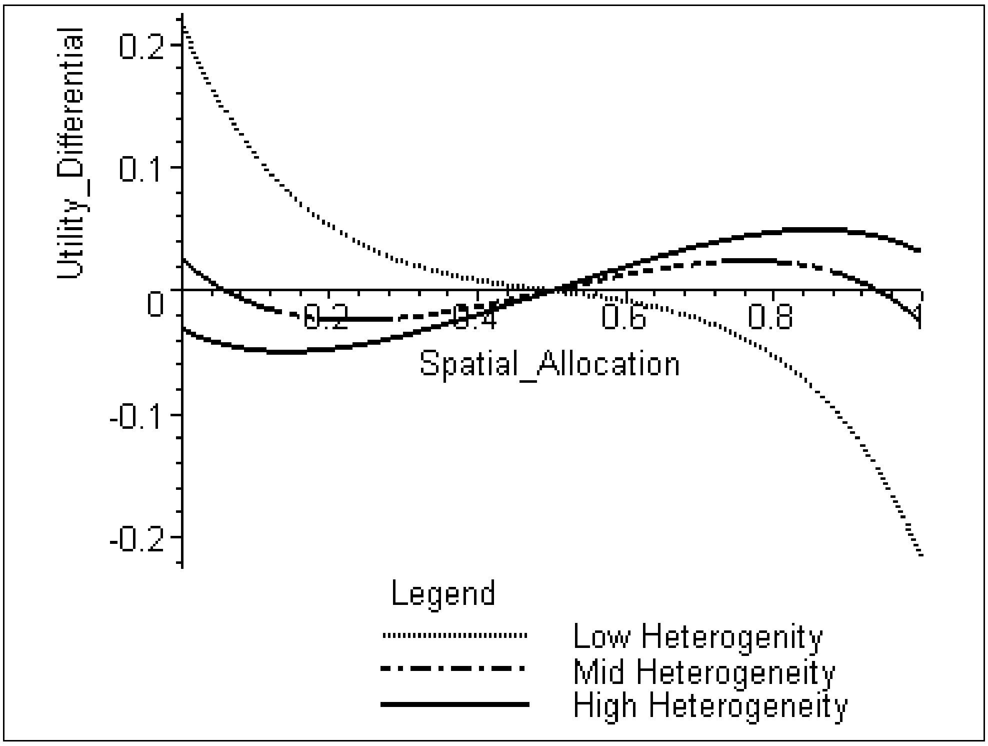
Parameters: α=10; γ=2; f_E=1; k=1; η=10; L=50; M=20; c_M=15; τ=2.4, 2.7, 3.0.

Figure 1 – Core-periphery patterns with stable partial agglomeration



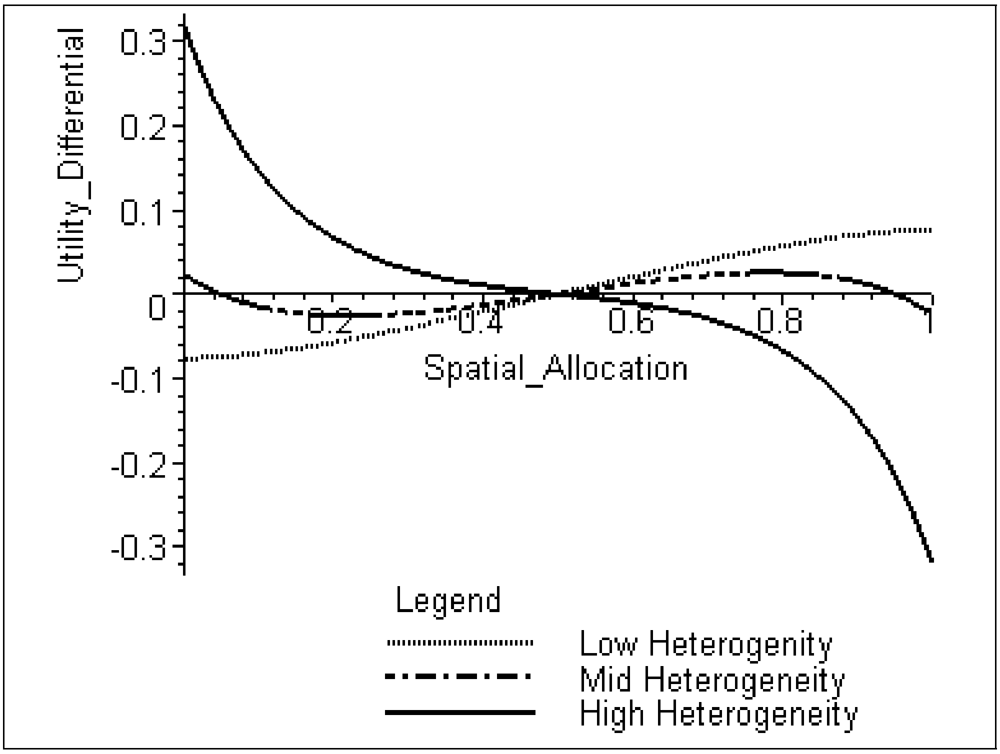
Parameters: α=10; γ=2; f_E=1; k=1; η=10; L=50; M=20; c_M=5, 15, 30; τ=2.7.

Figure 2 – More cost-increasing richness (larger c_M) fosters agglomeration



Parameters: α=10; γ=2; f_E=1; k=1, 1.05, 1.2; η=10; L=50; M=20; c_M=25; τ=2.7.

Figure 3 – More cost-decreasing evenness (lower *k*) fosters agglomeration if initial evenness is large enough



Parameters: α=10; γ=2; f_E=1; k=3, 3.55, 4; η=10; L=50; M=20; c_M=25; τ=2.7.

Figure 4 – More cost-decreasing evenness (lower *k*) fosters dispersion if initial evenness is small enough