DISCUSSION PAPER SERIES

No. 9417

MONOPOLISTIC COMPETITION AND OPTIMUM PRODUCT SELECTION: WHY AND HOW HETEROGENEITY MATTERS

Antonella Nocco, Gianmarco Ottaviano and Matteo Salto

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP9417.asp**

ISSN 0265-8003

MONOPOLISTIC COMPETITION AND OPTIMUM PRODUCT SELECTION: WHY AND HOW HETEROGENEITY MATTERS

Antonella Nocco, Università del Salento
Gianmarco Ottaviano, LSE, Bocconi University and CEPR
Matteo Salto, European Commission

Discussion Paper No. 9417
April 2013

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

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL TRADE AND REGIONAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

CEPR Discussion Paper No. 9417

April 2013

ABSTRACT

Monopolistic Competition and Optimum Product Selection: Why and how heterogeneity matters*

After some decades of relative oblivion, the interest in the optimality properties of monopolistic competition has recently re-emerged due to the availability of an appropriate and parsimonious framework to deal with firm heterogeneity. Within this framework we show that non-separable utility, variable demand elasticity and endogenous firm heterogeneity cause the market equilibrium to err in many ways, concerning the number of products, the size and the choice of producers, the overall size of the monopolistically competitive sector. More crucially with respect to the existing literature, we also show that the extent of the errors depends on the degree of firm heterogeneity. In particular, the inefficiency of the market equilibrium seems to be largest when selection among heterogenous firms is needed most, that is, when there are relatively many firms with low productivity and relatively few firms with high productivity.

JEL Classification: D4, D6, F1, L0 and L1
Keywords: heterogeneity, monopolistic competition, product diversity, selection and welfare

Antonella Nocco
University of Salento
Department of Management, Economics, Mathematics and Statistics
Ecotekne, via Monteroni
73100 Lecce
ITALY

Email: antonella.nocco@unisalento.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=159950

Gianmarco I.P. Ottaviano
London School of Economics
Department for Economics
Hougton Street
WC2A 2AE
UK

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

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=125330

Matteo Salto
European Commission
Rue de la Loi 200
B- 1049
BELGIUM

Email: matteo.salto@ec.europa.eu

*We thank Olivier Biau, Matthieu Lequien, Marc Melitz, John Morrow, Peter Neary, Mathieu Parenti, Avner Shaked, Jacques Thisse, Evgeny Zhelobodko as well as participants to presentations at the University of Utrecht and at the conference Industrial Organization and Spatial Economics , held in Saint-Petersburg in October 2012, for useful comments, discussions and suggestions. The views expressed here are those of the authors and do not represent in any manner the EU Commission.

Submitted 30 March 2013

1 Introduction

Do monopolistically competitive industries yield an optimal level of product diversity? As discussed by Neary (2004), this 'classic issue' in industrial organization motivated the canonical formalization of the Chamberlinian model (Chamberlin, 1933) as put forth by Spence (1976) and Dixit and Stiglitz (1977). These propose 'reduced form' models that "regard aggregate demands as if they result from the maximisation of a utility function defined directly over the quantities of goods, and the form of the utility function is intended to capture the desire for variety" (Dixit, 2004, p.125).[1] The classic issue can be itself split into four questions concerning the optimality of the market outcome (Stiglitz, 1975): Are there too few or too many products? Are the quantities of the products too small or too large? Are the products supplied by the right set of firms, or are there 'errors' in the choice of technique? Are monopolistically competitive industries too large or too small with respect to the rest of the economy?

The Chamberlinian model makes four basic assumptions (Bishop, 1967; Brakman and Heijdra, 2004): the number of sellers in a group of firms is sufficiently large so that each firm takes the behavior of other firms in the group as given; the group is well defined and small relative to the economy; products are physically similar but economically differentiated so that buyers have preferences for all types of products ('love for variety'); there is free entry. In this setup, optimality rests on how the market mechanism deals with the crucial tradeoff of 'efficiency versus diversity' (Kaldor, 1934).

As forcefully highlighted by Dixit and Stiglitz (1975), there are good reasons to doubt that the market will generally strike the right balance due to the public nature of diversity in the reduced form approach. As in these models the range of products enters utility as a direct argument in addition to the quantities consumed, the range itself becomes a public good whose social benefit is not fully reflected in private incentives. In the words of Spence (1976, pp. 230-231):

> "[T]here are conflicting forces at work with respect to the number or variety of products. Because of setup costs, revenues may fail to cover the costs of a socially desirable product. As a result, some products may be produced at a loss at an optimum. This is a force tending towards too few products. On the other hand, there are forces tending towards too many products. First, because firms hold back output and keep price above marginal cost, they leave more room for entry than would marginal cost pricing. Second, when a firms enter with a new product, it adds its own consumer and producer surplus to the total surplus, but it also cuts into the profits of the existing firms. If the cross elasticities of demand are high, the dominant effect may be the second one. In this case entry does not increase the size of the pie much; it just divides it into more pieces. Thus, in the presence of high cross elasticities of demand, there is a tendency toward too many products".

[1] 'Structural' models, instead, "give an explicit model of a consumer's choice where diversity plays a role; discrete choice from a collection of products differentiated by location in a characteristic space in the most common framework" (Dixit, 2004, p.125). See Anderson, de Palma and Thisse (1992) for microfoundations of the representative-consumer reduced form approach based on random-utility models of discrete choice.

As the issue of optimal product diversity does not admit a general settlement, explicit models with a detailed formulation of demand are used to isolate and analyze the four questions described above. The canonical choice is to model an economy consisting of two sectors. The first sector is monopolistically competitive and is the focus of the analysis. The second sector is perfectly competitive and represents the rest of the economy. Its purpose is to hold factor prices in check and to create the slack needed to answer the question whether the monopolistically competitive sector is too small or too big. This way the market is allowed to eventually misallocate resources not only within the monopolistically competitive sector but also between this sector and the rest of the economy.

The best known insights of the canonical model concern the special case in which the 'group utility' defined over differentiated products is separable across them, the demand of each product is CES and firms are homogeneous. In this case, the model shows that the first-best ('unconstrained') optimum calls for larger firms and more product variety than the market provides. From a normative perspective, however, this result is traditionally regarded of little practical relevance for policy intervention because implementing the unconstrained optimum requires the use of lump-sum instruments that are hardly available in reality. These are needed to subsidize the entry of firms that otherwise would not cover their setup ('entry') costs due to marginal cost pricing at the optimum. A lot of attention has, therefore, been devoted to the 'constrained' optimum in which the monopolistically competitive sector is financially self-sufficient. Under this constraint, the market is shown to provide the optimal number of products, the optimal size firms and hence the optimal size of the sector.

The robustness of these results has been investigated along several dimensions, with particular attention devoted to the impact of variable demand elasticity and firm heterogeneity. These extensions are already discussed by Stiglitz (1975), Spence (1976) and Dixit and Stiglitz (1977), who show that, when the elasticity of demand is allowed to vary, the market equilibrium ceases to be constrained optimal. In particular, products are too many (too few) and are supplied in too small (too large) quantities when the elasticity of 'product utility' is increasing (decreasing) in the quantity consumed. As for firm heterogeneity, Dixit and Stiglitz (1977) consider a variant of their model in which there are two groups of differentiated products that are perfect substitutes for each other with each group having CES sub-utility. Both fixed and marginal costs are allowed to differ between the two groups but not within them. Dixit and Stiglitz (1977) use this variant to show that the determination of the set of products to be supplied depends on a richer list of factors: fixed and marginal costs, the elasticity of the demand schedule, the level of the demand schedule and the cross-elasticities of demand. As a result, constrained optimality eventually applies only to a zero-measure set of parametrizations. A more exhaustive treatment of this issue can be found in Spence (1976) while Stiglitz (1975) reaches similar conclusions in a model of the capital market in which firms with heterogeneous costs issue securities whose returns are imperfectly correlated with each other.

After some decades of relative oblivion, interest in the optimality properties of monopolistic competition has recently re-emerged due to the 'heterogeneous firms revolution' in international trade theory (Melitz and Redding, 2012). This has been initiated by Melitz (2003), who shows that a Dixit-Stiglitz model with CES demand, endogenous firm heterogeneity and fixed export costs (but without the homogeneous good sector) predicts 'new' gains from trade liberalization

through the selection of the most efficient firms. Subsequent papers show that a similar result holds when demand exhibits variable elasticity, though fixed export costs are not necessarily needed for the result to materialize in this case (Melitz and Ottaviano, 2008; Behrens and Murata, 2012).[2]

The validity of these (among other) insights on international trade issues when alternative specifications of demands are allowed for is discussed by Zhelobodko, Kokovin, Parenti, and Thisse (2012). Using a framework with variable elasticity of substitution (VES), they show that CES is just a knife-edge case. While this finding is reminiscent of the conclusions by Stiglitz (1975), Spence (1976) and Dixit and Stiglitz (1977), Zhelobodko, Kokovin, Parenti, and Thisse (2012) do not discuss its implications for optimum product variety as those early contributors do. This is done, instead, by Dhingra and Morrow (2012) who fully characterize the optimality properties of a general demand system derived from separable 'group utility'. Their normative analysis thus complements the positive analysis of Zhelobodko, Kokovin, Parenti, and Thisse (2012), showing that, in the absence of the homogeneous sector, the market outcome achieves the (unconstrained) optimum under CES but not under VES. When a homogeneous sector is instead introduced, Melitz and Redding (2012) show that CES leads to constrained rather than unconstrained optimality due to the misallocation of resources between sectors. In other words, with CES firm heterogeneity does not change the welfare insights of the original Dixit-Stiglitz framework while things change in the case of VES.

The present paper goes back to the full set of classic questions laid down at the beginning of this introduction, with renewed emphasis on the question whether in the market equilibrium the products are supplied by the right set of firms. It does so in a Melitzian framework of endogenous firm heterogeneity with variable demand elasticity. Its aim is twofold. It shows that, with variable demand elasticity and endogenous firm heterogeneity, the market outcome errs with respect to the number of products, the size and the choice of producers, and the overall size of the monopolistically competitive sector. More crucially with respect to the existing literature, it also shows that the extent of the errors depends on the degree of firm heterogeneity.

None of the papers previously cited simultaneously addresses the four classic questions on the optimality of monopolistic competition in a framework with variable demand elasticity and endogenous firm heterogeneity. Moreover, none of them provides a systematic quantitative analysis of the impact of different degrees of firm heterogeneity on the extent of market inefficiencies. The discussion in Spence (1976) is systematic but qualitative, while Dixit and Stiglitz (1977) confine themselves to the special scenario discussed above. Dhingra and Morrow (2012) are closer to what the present paper tries to achieve but the focus of their comparative statics is on the parametrization of demand rather than on the parametrization of firm heterogeneity. In addition, not having the homogeneous good sector prevents them from discussing between-sector misallocation. Differently, Stiglitz (1975) presents comparative statics results on the heterogeneity parameters but his heterogeneity is not endogenous and his approach, based on a utility defined over alternative portfolios of assets, is quite distinct from the canonical model of monopolistic competition.

[2] See Arkolakis, Costinot and Rodriguez-Clare (2010) as well as Melitz and Redding (2013) for a discussion of the actual novelty of these findings.

Clearly, as pointed out by Stiglitz (1975) and others, without some appropriate parametrization of the problem, it would be hard to cut any new ground on the issues of interest. We rely on the specific parametrization of linear demand introduced by Ottaviano, Tabuchi and Thisse (2002) as applied to endogenous firm heterogeneity by Melitz and Ottaviano (2008). This parametrization is less general than the VES systems studied by Dhingra and Morrow (2012) and Zhelobodko, Kokovin, Parenti, and J. F. Thisse (2012) in terms of product utility but allows for cross-product effects that are absent in the former paper and only touched upon in the latter. For ease of exposition, in the main text we also focus on a specific but commonly used Pareto parametrization of firm heterogeneity, relegating the discussion of the validity of some key results in the case of a generic continuous parametrization to the appendix. There we also present the welfare analysis of the degenerate case in which firms are homogeneous as discussed by Ottaviano and Thisse (1999) for the same demand system.

The rest of the paper is organized in six sections. Section 2 briefly presents the model by Melitz and Ottaviano (2008). Sections 3 and 4 respectively derive and compare the market equilibrium and the (unconstrained) optimum. Section 5 investigates the impact of firm heterogeneity on the gap between the equilibrium and optimum outcomes. Section 6 discusses the constrained optimum. Section 7 concludes.

2 The model

Following Melitz and Ottaviano (2008), consider an economy populated by L consumers, each endowed with one unit of labor. Preferences are defined over a continuum of differentiated varieties indexed $i \in \Omega$, and a homogeneous good indexed 0. All consumers own the same initial endowment $\overline{q}_0$ of this good and share the same utility function given by

$$U = q_0^c + \alpha \int_{i\in\Omega} q_i^c di - \frac{1}{2}\gamma \int_{i\in\Omega} (q_i^c)^2 \, di - \frac{1}{2}\eta \left(\int_{i\in\Omega} q_i^c di \right)^2 \tag{1}$$

with positive demand parameters α, η and γ, the latter measuring the 'love for variety' and the others measuring the preference for the differentiated varieties with respect to the homogeneous good. The initial endowment $\overline{q}_0$ of the homogeneous good is assumed to be large enough for its consumption to be strictly positive at the market equilibrium and optimal solutions.

Labor is the only factor of production. It can be employed for the production of the homogeneous good under perfect competition and constant returns to scale with unit labor requirement equal to one. It can also be employed for the production of the differentiated varieties under monopolistic competition. The technology requires a preliminary R&D effort of $f > 0$ units of labor to design a new variety and its production process, which is also characterized by constant returns to scale. The R&D effort leads to the design of a new variety with certainty whereas the unit labor requirement c of the corresponding production process is uncertain, being randomly drawn from a continuous distribution with cumulative density

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

This corresponds to the empirically relevant case in which marginal productivity $1/c$ is Pareto distributed with shape parameter $k \geq 1$ over the support $[1/c_M, \infty)$. Hence, as k rises, density is skewed towards the upper bound of the support of G(c).[3] The R&D effort cannot be recovered and this gives rise to a sunk setup ('entry') cost.

3 Equilibrium and optimum

3.1 The market outcome

In the decentralized equilibrium consumers maximize utility under their budget constraints, firms maximize profits given their technological constraints, and markets clear. It is assumed that the labor market as well as the market of the homogeneous good are perfectly competitive. This good is chosen as numeraire, which then implies that the wage equals one. The market of differentiated varieties is, instead, monopolistically competitive with a one-to-one relation between firms and varieties.

The first order conditions for utility maximization give individual inverse demand for variety i as

$$p_i = \alpha - \gamma q_i^c - \eta Q^c \tag{3}$$

whenever $q_i^c > 0$, with $Q^c = \int_{i\in\Omega} q_i^c di$. Demand for consumed varieties can be derived from (3) as

$$q_i \equiv L q_i^c = \frac{\alpha L}{\eta N + \gamma} - \frac{L}{\gamma} p_i + \frac{\eta N}{\eta N + \gamma} \frac{L}{\gamma} \bar{p}, \ \forall i \in \Omega^* \tag{4}$$

where the set Ω^* is the largest subset of Ω such that demand is positive, N is the measure ('number') of varieties in Ω^* and $\bar{p} = (1/N) \int_{i\in\Omega^*} p_i di$ is their average price. Variety i belongs to this set when

$$p_i \leq \frac{1}{\eta N + \gamma} (\gamma\alpha + \eta N \bar{p}) \equiv p_{\max} \tag{5}$$

where $p_{\max} \leq \alpha$ represents the price at which demand for a variety is driven to zero.[4]

When a variety is produced by a firm with unit labor requirement c, the corresponding first order conditions for profit maximization are satisfied by an output level equal to

$$q^m(c) = \begin{cases} \frac{L}{2\gamma} (c^m - c) & \text{if } c \leq c^m = p_{\max} = \alpha - \frac{\eta}{L} Q^m \\ 0 & \text{if } c > c^m \end{cases} \tag{6}$$

where 'm' labels equilibrium variables and $Q^m = \int_0^{c^m} q^m(c) dG(c)$ is the total supply of differentiated varieties. Expression (6) defines a cutoff rule for survival:

[3] While the analysis in the main text rests on the Pareto distribution, several results have more general validity as discussed in Appendix A.

[4] Melitz and Ottaviano (2008) show that rewriting the indirect utility function in terms of average price and price variance reveals that it decreases with average prices $\bar{p}$, but rises with the variance of prices σ_p^2 (holding $\bar{p}$ constant), as consumers then re-optimize their purchases by shifting expenditures towards lower priced varieties as well as the numeraire good. Note also that the demand system exhibits 'love of variety': holding the distribution of prices constant (namely holding the mean $\bar{p}$ and variance σ_p^2 of prices constant), utility rises with product variety N.

only entrants that are productive enough ($c \leq c^m$) eventually produce. For them the price that corresponds to the profit-maximizing output $q^m(c)$ is $p^m(c) = (c^m + c)/2$, implying markup $\mu^m(c) = p^m(c) - c = (c^m - c)/2$ and maximized profit

$$\pi(c) = \frac{L}{4\gamma}\left(c^m - c\right)^2 \tag{7}$$

Due to free entry and exit, in equilibrium expected profit is exactly offset by the sunk entry cost

$$\int_0^{c^m} \pi(c) dG(c) = f$$

Given (2) and (7), this 'free entry condition' can be rewritten as

$$\left(\frac{c^m}{c_M}\right)^k \frac{L\left(c^m\right)^2}{2\gamma(k+1)(k+2)} = f \tag{8}$$

where, due to the law of large numbers, $G(c^m) = \left(c^m/c_M\right)^k$ is the *ex ante* probability that an entrant will produce as well as the *ex post* share of entrants that eventually produce while $L\left(c^m\right)^2/[2\gamma(k+1)(k+2)]$ is the *ex ante* expected profit conditional on producing as well as the *ex post* average profit of producers. Condition (8) can be solved for the unique equilibrium cutoff marginal cost

$$c^m = \left[\frac{2\gamma(k+1)(k+2)\left(c_M\right)^k f}{L}\right]^{\frac{1}{k+2}} \tag{9}$$

Finally, the number of producers can be determined as a function of c^m by observing that marginal firms with unit labor requirement $c = c^m$ make zero profit, i.e. $p(c^m) = c^m = p_{\max}$. Recalling (5), that implies the following 'zero cutoff profit condition'

$$c^m = \frac{1}{\eta N^m + \gamma}\left(\gamma\alpha + \eta N^m \bar{p}^m\right) \tag{10}$$

where, again due to the law of large numbers, $\bar{p}^m$ is the *ex ante* expected price conditional on producing as well as the *ex post* average price of producers: $\bar{p}^m = \int_0^{c^m} p(c) dG^m(c)$ with $G^m(c) = G(c)/G(c^m) = \left(c/c^m\right)^k$. The 'zero cutoff profit condition' can then be solved to obtain the equilibrium number of producers (and varieties) as a function of the equilibrium cutoff as

$$N^m = \frac{2\gamma(k+1)}{\eta}\frac{\alpha - c^m}{c^m} \tag{11}$$

with the corresponding equilibrium number of entrants given by $N_E^m = N^m/G(c^m) = N^m\left(c_M/c^m\right)^k$.

3.2 The optimal outcome

As the quasi-linearity of (1) implies transferable utility, social welfare may be expressed as the sum of all consumers' utilities. This implies that the first best ('unconstrained') planner chooses the number of varieties and their output levels so as to maximize the social welfare function given by individual utility (1)

times the number of consumers L, subject to the resource constraint, the varieties' production functions and the stochastic 'innovation production function' (i.e. the mechanism that determines each variety's unit labor requirement as a random draw from $G(c)$ after f units of labor have been allocated to R&D).

Specifically, given (1), the planner chooses the number N_E of R&D projects and the output levels of associated varieties so as to maximize social welfare

$$\begin{array}{c} W = q_0^c L + \alpha N_E \int_0^{c_M} \left[q^c(c)L\right] dG(c) - \frac{1}{2}\frac{\gamma}{L} N_E \int_0^{c_M} \left[q^c(c)L\right]^2 dG(c) \\ -\frac{1}{2}\frac{\eta}{L}\left[N_E \int_0^{c_M} \left[q^c(c)L\right] dG(c)\right]^2 \end{array} \tag{12}$$

with respect to q_0^c, $q^c(c)$ and N_E subject to the aggregate resource constraint

$$q_0^c L + fN_E + N_E \int_0^{c_M} cq^c(c)LdG(c) = L + \overline{q}_0 L \tag{13}$$

stating that the supply of the homogeneous good ($q_0^c L$), the supply of differentiated varieties ($N_E \int_0^{c_M} cq^c(c)LdG(c)$) and the R&D investment (fN_E) are costrained by the amount of available resources ($L + \overline{q}_0 L$).

After substituting (13) into (12), the planner's problem can be rewritten as the maximization of

$$\begin{array}{c} W = L + \overline{q}_0 L - fN_E + N_E \int_0^{c_M} (\alpha - c)\, q(c) dG(c) \\ -\frac{1}{2}\frac{\gamma}{L} N_E \int_0^{c_M} \left[q(c)\right]^2 dG(c) - \frac{1}{2}\frac{\eta}{L}\left[N_E \int_0^{c_M} q(c) dG(c)\right]^2 \end{array} \tag{14}$$

with respect to $q(c)$ and N_E. The corresponding first order conditions are then:

$$\frac{\partial W}{\partial q(c)} = \left[N_E (\alpha - c) - \frac{\gamma}{L} N_E q(c) - \frac{\eta}{L} (N_E)^2 \int_0^{c_M} q(c) dG(c)\right] dG(c) = 0 \quad \forall c \tag{15}$$

$$\begin{array}{c} \frac{\partial W}{\partial N_E} = -f + \int_0^{c_M} (\alpha - c)\, q(c) dG(c) - \frac{1}{2}\frac{\gamma}{L} \int_0^{c_M} \left[q(c)\right]^2 dG(c) \\ -\frac{\eta}{L} N_E \left[\int_0^{c_M} q(c) dG(c)\right]^2 = 0 \end{array} \tag{16}$$

Rearranging (15) shows that optimal output $q^o(c)$ has to satisfy

$$q(c) = \frac{L}{\gamma}(\alpha - c) - \frac{\eta}{\gamma} N_E \int_0^{c_M} q(c) dG(c) = \frac{L}{\gamma}(\alpha - c) - \frac{\eta}{\gamma} Q$$

with $Q \equiv L \int_{i \in \Omega} q_i^c di = N_E \int_0^{c_M} q(c) dG(c)$, i.e.

$$q^o(c) = \begin{cases} \frac{L}{\gamma}(c^o - c) & \text{if } c \leq c^o = \alpha - \frac{\eta}{L} Q^o \\ 0 & \text{if } c > c^o \end{cases} \tag{17}$$

where 'o' labels first best optimum variables and $Q^o = N_E^o \int_0^{c_M} q^o(c) dG(c)$ is the optimum total supply of differentiated varieties. Result (17) reveals that, just like the market, also the planner follows a cutoff rule allowing only for the production of varieties whose unit labor requirements are low enough: $q^o(c) \geq 0$ only for $c \leq c^o$. We can thus define the conditional distribution of unit input requirements for varieties that the planner actually produces as $G^o(c) = G(c)/G(c^o)$. The number N^o of those varieties thus satisfies $N^o = G(c^o) N_E^o$.

Expressions (17) and (3) can be used to show that the first best output levels would clear the market in the decentralized scenario only if each producer priced at its own marginal cost. To see this, note that (3) implies $q(c) =$

$[\alpha - p(c)]L/\gamma - \eta Q/\gamma$. Then, imposing $q(c) = q^o(c) = (c^o - c)\, L/\gamma$ and $Q = Q^o = (\alpha - c^o)\, L/\eta$ from (17) respectively on the left and on the right hand sides of $q(c) = [\alpha - p(c)]L/\gamma - \eta Q/\gamma$ gives $p(c) = c$.

Integrating (15) gives

$$Q^o = \frac{\gamma N^o}{\gamma + \eta N^o} \frac{L}{\gamma} (\alpha - \bar{c}^o)$$

where $\bar{c}^o = \int_0^{c^o} c dG^o(c)$. Substituting this result in $c^o = \alpha - \eta Q^o/L$ from (17) and solving for N^o gives a planner's cutoff condition analogous to the market 'zero cutoff profit condition' (10)

$$N^o = N_E^o G(c^o) = \frac{\gamma(k+1)}{\eta} \frac{\alpha - c^o}{c^o} \tag{18}$$

In order to find a second condition analogous to the market 'free entry condition' (8), we can substitute the optimal quantities from (17) as well as the optimal number of varieties (18) in the second condition in (16) to obtain

$$\left(\frac{c^o}{c_M}\right)^k \frac{L\,(c^o)^2}{\gamma(k+1)(k+2)} = f \tag{19}$$

so that the first best cutoff marginal cost evaluates to

$$c^o = \left[\frac{\gamma(k+1)(k+2)\,(c_M)^k\, f}{L}\right]^{\frac{1}{k+2}} \tag{20}$$

This then determines the first best number of varieties through (18). To sum up, (20) and (18) are the first best planner's analogues of expressions (9) and (11) derived for the market equilibrium.

4 Equilibrium vs. optimum

There are two dimensions along which the efficiency of the market outcome can be evaluated: the number of varieties actually produced N^m and the (conditional) cost distribution of the firms producing them as dictated by the cutoff c^m. In turn, the cost distribution determines the efficiency of the corresponding distributions of firm sizes and prices.[5]

The tradeoffs the first best planner faces when firms are heterogeneous can be highlighted by rewriting the first best objective (12) in terms of means and variances of the distribution $G(c)$ as follows

$$W = \left[L + \overline{q}_0 L + N_E \left(\alpha \widehat{q} - \tfrac{1}{2}\tfrac{\gamma}{L}\widehat{q}^2 - \tfrac{1}{2}\tfrac{\eta}{L} N_E \widehat{q}^2 - \widehat{cq} - f\right)\right] - \left[N_E \left(\tfrac{1}{2}\tfrac{\gamma}{L}\widehat{\sigma}_q^2 + \widehat{\sigma}_{cq}\right)\right] \tag{21}$$

where $\widehat{c} = \int_0^{c_M} c dG(c)$ is the unconditional mean unit labor requirement, $\widehat{q} = \int_0^{c_M} q(c) dG(c)$ and $\widehat{\sigma}_q^2 = \left\{\int_0^{c_M} [q(c)]^2\, dG(c) - \widehat{q}^2\right\}$ are the unconditional mean

[5] Dhingra and Morrow (2012) provide a detailed discussion of these issues that emphasizes the role of alternative parametrizations of demand when utility is separable. The bias in market allocations by demand characteristics is summarized in their Table 2. If we also assumed separability (by imposing $\eta = 0$), our demand system would be compatible with the parametrizations classified in the upper right hand corner of that table.

and variance of quantities, and $\widehat{\sigma}_{cq} = \left\{\int_0^{c_M} cq(c)dG(c) - \widehat{c}\widehat{q}\right\}$ is the convariance between quantities and unit input requirements.[6] The first bracketed term on the right hand side of (21) corresponds to the planner's objective when marginal costs are homogeneous. Here the tradeoffs are in terms of: (a) average quantity vs. average marginal cost; (b) number of varieties vs. fixed costs. The second bracketed term has to be considered when unit labor requirements are heterogeneous. It shows that, due to love of variety, consumers dislike a consumption bundle in which the quantity consumed varies across varieties. Formally, they dislike a consumption bundle with large deviations from the average (large $\widehat{\sigma}_q$), the more so the stronger the love of variety (larger γ). On the other hand, there is a penalty in offering a basket of varieties with small deviations around the average as higher productivity could be achieved by assigning little production to varieties with high marginal costs ($\widehat{\sigma}_{cq} < 0$).

4.1 Selection

Comparing the equilibrium cutoff with the optimal one is straightforward. Specifically, comparing expressions (9) with (20) reveals that $c^m = 2^{1/(k+2)}c^o$, which implies $c^o < c^m$. Accordingly, varieties with $c \in [c^o, c^m]$ should not be supplied. We thus have:

Proposition 1 ***(Selection)*** *Firm selection in the market equilibrium is weaker than optimal.*

The intuition behind this proposition can be gauged by recalling that, as discussed in Section 3.1, in the market equilibrium the markup of a firm with marginal cost c equals $\mu^m(c) = p^m(c) - c = (c^m - c)/2$. Accordingly, consumption is inefficiently biased against the differentiated varieties and in favor of the numeraire good as the prices of the former are inefficiently high.

Differences in the strength of selection map into aggregate performance. In particular, defining aggregate productivity $\bar{\Phi}$ as average output per worker weighted by firm size, expressions (2), (6) and (17) imply

$$\bar{\Phi}^j \equiv \frac{\int_0^{c^j} q(c)dG(c)}{\int_0^{c^j} cq(c)dG(c)} = \frac{k+2}{k}\frac{1}{c^j}$$

with $j \in \{m, o\}$ Hence, the cutoff ranking $c^o < c^m$ maps into the productivity ranking $\bar{\Phi}^o > \bar{\Phi}^m$, with $\bar{\Phi}^o = 2^{1/(k+2)}\bar{\Phi}^m$, giving rise to the following result:

Corollary 2 ***(Average productivity)*** *Aggregate productivity in the market equilibrium is lower than optimal.*

4.2 Firm size

Proposition 1 has also implications in terms of optimality of the firm size distribution. To see this, one can use (6) and (17) to rewrite output levels as

$$q^m(c) = \frac{L}{2\gamma}(c^m - c) \quad \text{and} \quad q^o(c) = \frac{L}{\gamma}(c^o - c)$$

[6] With homogeneous unit labor requirements we would have $\widehat{\sigma}_q = \widehat{\sigma}_{cq} = 0$ and the planner's objective boils down to the one in Ottaviano and Thisse (1999). See Appendix B for further details.

Since $c^m = 2^{1/(k+2)} c^o$ implies $c^o < c^m$, it is readily seen that $q^m(c) > q^o(c)$ if and only if $c > \left(2 - 2^{1/(k+2)}\right) c^o$, which falls in the relevant interval $[0, c^o]$ given that $0 < \left(2 - \sqrt[3]{2}\right) < \left(2 - 2^{1/(k+2)}\right) < 1$. Hence, with respect to the optimum, the market equilibrium undersupplies varieties with marginal cost $c \in [0, \left(2 - 2^{1/(k+2)}\right) c^o)$ and oversupplies varieties with marginal cost $c \in (\left(2 - 2^{1/(k+2)}\right) c^o, c^m]$. Hence, we have:

Corollary 3 ***(Within-sector misallocation)*** *The market equilibrium oversupplies high cost varieties and undersupplies low cost ones with respect to the optimum.*

In other words, misallocation materializes as a lack of market concentration: in the market equilibrium there are relatively too many small firms and relatively too few large firms with respect to the optimum. The intuition behind this corollary can be explained as follows. The markup $\mu^m(c) = (c^m - c)/2$ is a decreasing function of c. This implies that more productive firms do not pass on their entire cost advantage to consumers as they absorb part of it in the markup. As a result, the price ratio of less to more productive firms is smaller than their cost ratio and thus the quantities sold by less productive firms are too large from an efficiency point of view relative to those sold by more productive firms.

Turning to average firm size $\overline{q}$, given (2), expressions (6) and (17) together with expressions (9) and (20) imply

$$\begin{aligned} \overline{q}^m &= \int_0^{c^m} q^m(c) dG^m(c) = \frac{L}{2\gamma} \frac{1}{k+1} c^m \\ \overline{q}^o &= \int_0^{c^o} q^o(c) dG^o(c) = \frac{L}{\gamma} \frac{1}{k+1} c^o = 2^{\frac{k+1}{k+2}} \overline{q}^m \end{aligned} \tag{22}$$

so that the cutoff ranking $c^o < c^m$ dictates the average output ranking $\overline{q}^m < \overline{q}^o$. Accordingly, we can write:

Corollary 4 ***(Average firm size)*** *In the market equilibrium firms are on average smaller than optimal.*

The intuition behind this corollary follows from the discussion of the previous one: a lower cutoff with markup pricing makes firms on average larger in the optimum than in the market equilibrium.

Finally, given (11), (18) and (22), the total output of the differentiated varieties evaluates to $N^m \overline{q}^m = (L/\eta)(\alpha - c^m)$ and $N^o \overline{q}^o = (L/\eta)(\alpha - c^o)$ at the market equilibrium and at the optimum respectively. Hence, $c^o < c^m$ implies $N^o \overline{q}^o > N^m \overline{q}^m$ and we have:

Corollary 5 ***(Between-sector misallocation)*** *In the market equilibrium the total supply of differentiated varieties is smaller than optimal.*

4.3 Product variety and entry

The equilibrium is suboptimal also when it comes to the number of varieties supplied. However, given (11) and (18), the ranking of cutoffs $c^o < c^m$ does not

allow to rank N^m and N^o unambiguously. In particular, since $c^m = 2^{1/(k+2)}c^o$, we have $N^m > N^o$ as long as

$$\alpha > \alpha_1 \equiv \frac{c^o}{2^{\frac{k+1}{k+2}} - 1} = \frac{1}{2^{\frac{k+1}{k+2}} - 1} \left[\frac{\gamma(k+1)(k+2)\left(c_M\right)^k f}{L} \right]^{\frac{1}{k+2}} \tag{23}$$

which is the case when α as well as L are large and when γ, f as well as c_M are small. Hence, we can state the following result:

Corollary 6 ***(Product variety)*** *Product variety is richer (poorer) in the market equilibrium than in the optimum when varieties are close (far) substitutes, the sunk entry cost is small (large), market size is large (small) and the difference between the highest and the lowest possible cost draws is small (large).*

This corollary has an interesting implication for the impact of larger market size, driven for example by the integration of previously autarkic national markets. In this scenario, it could well be that each national market on its own is small enough to entail $\alpha < \alpha_1$ whereas the internationally integrated market is large enough to entail $\alpha > \alpha_1$. Then, according to the corollary, market integration would cause the transition from a situation in which product variety is inefficiently poor ($N^m < N^o$) to a situation in which it becomes inefficiently rich ($N^m > N^o$).

Turning to entry, the equilibrium number of entrants is given by

$$N_E^j = \frac{N^j}{G(c^j)} = N^j \left(\frac{c_M}{c^j} \right)^k \tag{24}$$

with $j \in \{m, o\}$. Then, together with (11) and (18) as well as (9) and (20), expression (24) can be used to show that $c^m = 2^{1/(k+2)}c^o$ imply $N_E^m > N_E^o$ as long as

$$\alpha > \alpha_2 \equiv \frac{2^{2/(k+2)} - 1}{2^{1/(k+2)} - 1} c^o = \frac{2^{2/(k+2)} - 1}{2^{1/(k+2)} - 1} \left[\frac{\gamma(k+1)(k+2)\left(c_M\right)^k f}{L} \right]^{\frac{1}{k+2}} \tag{25}$$

which is the case when α as well as L are large and γ, f as well as c_M are small. This leads to:

Corollary 7 ***(Entry)*** *More (fewer) firms enter in the market equilibrium than in the optimum if varieties are close (far) substitutes, the sunk entry cost is small (large), market size is large (small) and the difference between the highest and the lowest possible cost draws is small (large).*

As larger market size reduces α_2, it causes the transition from a situation in which the resources devoted to develop new varieties are inefficiently small ($N_E^m < N_E^o$) to a situation in which they are inefficiently large ($N_E^m > N_E^o$).

Given that (23) and (25) imply $\alpha_1 < \alpha_2$, corollaries 6 and 7 together imply that the market provides too little entry with too little variety for $\alpha < \alpha_1$ and too much entry with too much variety for $\alpha > \alpha_2$. For $\alpha_1 < \alpha < \alpha_2$ it provides, instead, too much variety and too little entry.

5 The impact of firm heterogeneity

We now turn to the relation between the degree of heterogeneity and the extent of the market inefficiency. The key question here is whether or not the inefficiency of the market equilibrium is largest when selection is needed most, that is, when there are a lot of low productivity firms and few high productivity ones.

As discussed by Ottaviano (2012), the scale and shape parameters of the Pareto distribution (2) regulate the 'heterogeneity' of cost draws along two dimensions: 'richness' and 'evenness' (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 labor 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 labor requirement distribution is uniform on $[0, c_M]$ with maximum evenness. As k increases, the unit labor requirement distribution becomes more concentrated at higher unit labor requirements close to c_M: evenness falls. As k goes to infinity, the distribution becomes degenerate at c_M: all draws deliver a unit labor 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 labor requirements more likely without changing the unit labor requirements that are possible. Accordingly, more richness (larger c_M) comes with higher average unit labor requirement ('cost-increasing richness'), more evenness (smaller k) comes with lower average unit labor requirement ('cost-decreasing evenness').

Given the cutoff expressions (9) and (20), more heterogeneity has different impacts on selection depending on whether it comes through more richness or evenness. To see this, rewrite (9) and (20) as:

$$\begin{aligned} \left(\frac{c^m}{c_M}\right)^k \left[\frac{L}{4\gamma}\frac{2\left(c^m\right)^2}{(k+2)(k+1)}\right] &= f \\ \left(\frac{c^o}{c_M}\right)^k \left[\frac{L}{2\gamma}\frac{2\left(c^m\right)^2}{(k+2)(k+1)}\right] &= f \end{aligned}$$

where $\left(c^m/c_M\right)^k$ and $\left(c^o/c_M\right)^k$ are the shares of viable varieties and the bracketed terms are average firm profit for the market equilibrium and average surplus per variety for the optimum respectively. For any given cutoffs, more cost-increasing richness (larger c_M) decreases the left hand sides of both expressions through its depressing effect on the share of viable varieties. As the right hand sides are constant, (9) and (20) can keep on holding only if the cutoffs rise. Differently, for any given cutoffs (smaller than c_M), more cost-decreasing evenness (smaller k) increases the left hand sides of both expressions through its enhancing effect on both the share of viable varieties and average profit or surplus. Again, as the right hand sides are constant, (9) and (20) can keep on holding only if the cutoffs fall. Hence, while more cost-increasing richness makes selection softer, more cost-decreasing evenness makes it tougher.

When we focus on the percentage deviation of the market equilibrium from

the optimum, only the change in evenness matters for several outcomes. Specifically, given $c^m = 2^{1/(k+2)}c^o$, more eveness (smaller k) leads to a larger percentage gap in the cutoffs between the market equilibrium and the optimum ($(c^m - c^o)/c^o$ rises) whereas more richness is immaterial. Hence, we have:

Proposition 8 ***(Heterogeneity and selection)*** *More cost-decreasing evenness increases the percentage gap in the cutoffs between the market equilibrium and the optimum. Cost-increasing richness has no impact on this gap.*

As in the case of Proposition 1, Proposition 8 gives rise to a series of parallel corollaries. First, given $\bar{\Phi}^o = 2^{1/(k+2)}\bar{\Phi}^m$, smaller k increases the percentage aggregate productivity gap between the market equilibrium and the optimum ($\left(\bar{\Phi}^o - \bar{\Phi}^m\right)/\bar{\Phi}^o$ rises). We can therefore state:

Corollary 9 ***(Heterogeneity and productivity)*** *More cost-decreasing evenness increases the percentage gap in the aggregate productivity between the market equilibrium and the optimum. Cost-increasing richness has no impact on this gap.*

Second, recall that, with respect to the optimum, the market equilibrium undersupplies varieties with marginal cost $c \in [0, \left(2 - 2^{1/(k+2)}\right)c^o)$ and oversupplies varieties with marginal cost $c \in (\left(2 - 2^{1/(k+2)}\right)c^o, c^m]$. When k falls $\left(2 - 2^{1/(k+2)}\right)c^o/c^m$ also falls whereas it does not change when c_M changes. This leads to:

Corollary 10 ***(Heterogeneity and within-sector misallocation)*** *More cost-decreasing evenness makes the overprovision of varieties relatively more likely than its underprovision in the market equilibrium. Cost-increasing richness has no impact on this.*

Third, given that by (22) $\overline{q}^o = 2^{\frac{k+1}{k+2}}\overline{q}^m$, smaller k decreases the percentage average size gap between the optimum and the market equilibrium ($(\overline{q}^o - \overline{q}^m)/\overline{q}^o$ falls). Hence, we have:

Corollary 11 ***(Heterogeneity and average firm size)*** *More cost-decreasing evenness decreases the percentage gap in average firm size between the market equilibrium and the optimum. Cost-increasing richness has no impact on this gap.*

Fourth, expressions $N^m\overline{q}^m = (L/\eta)(\alpha - c^m)$ and $N^o\overline{q}^o = (L/\eta)(\alpha - c^o)$ allow us to write

$$\frac{N^o\overline{q}^o - N^m\overline{q}^m}{N^o\overline{q}^o} = \frac{c^m - c^o}{c^o}\frac{c^o}{\alpha - c^o}$$

By Proposition 8 changes in c_M have no impact on $(c^m - c^o)/c^o$ whereas, by (20), larger c_M leads to larger c^o and therefore larger $c^o/(\alpha - c^o)$. Accordingly, larger c_M implies larger $(N^o\overline{q}^o - N^m\overline{q}^m)/N^o\overline{q}^o$. Proposition 8 also states that smaller k leads to larger $(c^m - c^o)/c^o$ whereas, by (20), smaller k leads to smaller c^o and therefore smaller $c^o/(\alpha - c^o)$. As the latter effect is strong when c^o is far from α and this is the case for small k, falling k decreases $(N^o\overline{q}^o - N^m\overline{q}^m)/N^o\overline{q}^o$ when k is initially small and increases it when k is initially large. We can then write:

Corollary 12 ***(Heterogeneity and between-sector misallocation)*** *More cost-increasing richness increases the percentage gap in the total output of the differentiated varieties between the market equilibrium and the optimum. More cost-decreasing evenness increases the percentage gap if evenness is initially low and decreases it if evenness is initially high.*

Fifth, given again Proposition 8, expressions (11) and (18) with the associated condition (23) lead to:

Corollary 13 ***(Heterogeneity and product variety)*** *Less cost-decreasing evenness and more cost-increasing richness makes the underprovision of variety relatively more likely than its overprovision in the market equilibrium.*

Analogously, given (24) and the associated condition (25), we can write:

Corollary 14 ***(Heterogeneity and entry)*** *Less cost-decreasing evenness and more cost-increasing richness makes the lack of entry relatively more likely than excess entry in the market equilibrium.*

In the limit, when k goes to infinity, the Pareto distribution convergences to a Dirac distribution with all density concentrated at c_M. In this case without heterogeneity, in which the number of entrants and the number of producers coincide, the market always yields too much entry and too much variety with respect to the optimum (Ottaviano and Thisse, 1999).[7]

Finally, we can look at the relation between heterogeneity and welfare. The welfare level attained in the market equilibrium can be expressed as a function of a corresponding cutoff through the following substitutions in the planner's objective (14):expression (11) can be used together with $N_E = N\left(c_M/c^m\right)^k$ to substitute for N_E; expression (8) can be used to substitute for f; expression (6) can be used to substitute for $q(c)$. The result is:

$$W^m = L + \overline{q}_0 L + \frac{L}{2\eta}\left(\alpha - c^m\right)\left(\alpha - \frac{k+1}{k+2}c^m\right) \tag{26}$$

Analogously, the welfare level attained in the optimum can be expressed as a function of a corresponding cutoff through the following substitutions in the planner's objective (14): expression (18) can be used together with $N_E = N\left(c_M/c^m\right)^k$ to substitute for N_E; expression (19) can be used to substitute for f; expression (17) can be used to substitute for $q(c)$. This gives:

$$W^o = L + \overline{q}_0 L + \frac{L}{2\eta}\left(\alpha - c^o\right)^2 \tag{27}$$

Given $c^m = 2^{1/(k+2)}c^o$, it is readily verified that we have $W^m < W^o$, as to be expected. Comparing (26) and (27) also reveals:[8]

Corollary 15 ***(Heterogeneity and welfare)*** *More cost-decreasing evenness and less cost-increasing richness reduce the percentage gap in welfare between the market equilibrium and the optimum.*

In other words, from a welfare point of view, the inefficiency of the market equilibrium is largest when selection is needed most: a lot of low productivity firms and few high productivity ones.

[7] See Appendix B.

[8] See Appendix C for a proof.

6 Constrained optimum

The unconstrained optimum discussed so far has been traditionally regarded of little practical relevance from a normative point of view. Its implementation requires the use of lump-sum instruments to subsidize the entry of firms that otherwise would not cover their entry costs due to marginal cost pricing at the unconstrained optimum. As these instruments are considered hardly available in reality, it is interesting to look at the 'constrained' optimum, in which the differentiated sector has to be financially self-sufficient.

The constrained planner maximizes (14) with respect to N_E subject to two constraints: profit maximizing output (6) and the 'free entry condition' (8). These impose the planner the market cutoff (9). Substituting (6) and (8) in (14) allows us to rewrite the constrained problem as the maximization of

$$W = L + \overline{q}_0 L + \frac{2\alpha\left(k+2\right) - \left(2k+3\right)c^m}{2c^m} f N_E - \frac{\eta\left(k+2\right)\left(c^m\right)^k}{4\gamma\left(k+1\right)\left(c_M\right)^k} f\left(N_E\right)^2 \quad (28)$$

with respect to N_E. Then, using $N_E = N\left(c_M/c^m\right)^k$ to substitute for N_E in the first order condition of the planner's problem yields

$$N^s = \frac{2\gamma\left(k+1\right)}{\eta} \frac{\alpha - \frac{2k+3}{2(k+2)}c^m}{c^m} \quad (29)$$

Comparing this expression with (11) reveals that product variety is richer in the constrained optimum than in the market equilibrium.

Expression (11) can be used together with $N_E = N\left(c_M/c^m\right)^k$ to substitute for N_E while (8) can be used to substitute for f in the planner's objective. The result expresses welfare in the constrained optimum as a function of the market cutoff

$$W^s = L + \overline{q}_0 L + \frac{L}{2\eta}\left[\alpha - \frac{2k+3}{2(k+2)}c^m\right]^2$$

This is smaller than W^o but larger than W^m. In particular, we have

$$W^s - W^m = \frac{L}{8\eta}\left(\frac{c^m}{k+2}\right)^2$$

Given (9), less cost-increasing richness (smaller c_M) reduces the percentage gap between the market outcome and the constrained optimum. The same happens in the case of more cost-decreasing evenness (smaller k) when initial evenness is high. Differently, when initial evenness is low, more cost-decreasing evenness raises the gap.[9] As in the case of the unconstrained optimum, the inefficiency of the market equilibrium is largest when selection is needed most.

7 Conclusion

After some decades of relative oblivion, the interest in the optimality properties of monopolistic competition has recently re-emerged due to the 'heterogeneous firms revolution' in international trade theory initiated by Melitz (2003). The

[9] See Appendix C for a proof.

availability of an appropriate and parsimonious framework to deal with firm heterogeneity allows to bring back into the normative debate the full set of questions the canonical formalization of the Chamberlinian model by Spence (1976) and Dixit and Stiglitz (1977) was designed to answer. In particular, it provides a useful analytical tool to address the question whether in the market equilibrium the products are supplied by the right set of firms, or there are rather 'errors' in the choice of technique.

We have contributed to this debate by showing that in a model with non-separable utility, variable demand elasticity and endogenous firm heterogeneity, the market outcome errs in many ways: with respect to the number of products, the size and the choice of producers, the overall size of the monopolistically competitive sector. More crucially with respect to the existing literature, we have also shown that the extent of the errors depends on the degree of firm heterogeneity. In particular, we have found that the inefficiency of the market equilibrium seems to be largest when selection is needed most, that is, when there are relatively many firms with low productivity and relatively few firms with high productivity. This holds from the viewpoints of both unconstrained and constrained efficiency.

These insights have been obtained for a parametrization of demand that is admittedly specific but still non-separable and more flexible than the CES. It would be important to understand how general they are by checking their validity under alternative non-separable parametrizations with variable demand elasticity, such as the one proposed by Behrens and Murata (2007). This is left to future research.

References

[1] Anderson S., A. de Palma and J.-F. Thisse (1992) *Discrete Choice Theory and Product Differentiation* (Cambride MA: MIT Press).

[2] Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012) New trade models, same old gains?, *American Economic Review* 102, 94-130.

[3] Bagnoli, M. and T. Bergstrom, (2005) Log-concave probability and its applications, *Economic Theory* 26, 445-469.

[4] Behrens K. and Y. Murata (2007) General equilibrium models of monopolistic competition: A new approach, *Journal of Economic Theory* 136, 776-787.

[5] Behrens K. and Y. Murata (2012) Trade, competition, and efficiency, *Journal of International Economics* 87, 1-17.

[6] Bishop R. (1967) Monopolistic competition and welfare economics, in Kuenne R. (ed.), *Monopolistic Competition Theory: Studies in Impact: Essays in Honor of Edward H. Chamberlin* (New York: John Wiley).

[7] Brakman and Heijdra (2004) Introduction, in Brakman R. and B. Heijdra (eds.), *The Monopolistic Competition Revolution in Retrospect* (Cambridge: Cambridge University Press).

[8] Chamberlin E. (1933) *The Theory of Monopolistic Competition* (Cambride MA: Harvard University Press).

[9] Dhingra S. and J. Morrow (2012) Monopolistic competition and optimum product diversity under firm heterogeneity, LSE Department of Economics, mimeo: `http://www.sdhingra.com/selection3rdGain.pdf`

[10] Dixit A. (2004) Some reflections on theories and applications of monopolistic competition, in Brakman R. and B. Heijdra (eds.), *The Monopolistic Competition Revolution in Retrospect* (Cambridge: Cambridge University Press).

[11] Dixit A. and J. Stiglitz (1975) Monopolistic competition and optimum product diversity, The Warwick Economics Research Series 64, University of Warwick, Department of Economics.

[12] Dixit A. and J. Stiglitz (1977) Monopolistic competition and optimum product diversity, *American Economic Review* 67, 297-308.

[13] Kaldor N. (1935).Market imperfection and excess capacity, *Economica* 2, 33-50.

[14] Maignan C., G. Ottaviano, D. Pinelli and F. Rullani (2003) Bio-ecological diversity vs. socio-economic diversity: A comparison of existing measures, FEEM Working Paper 58.2003.

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

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

[17] Melitz M. and S. Redding (2012) Heterogeneous firms and trade, NBER Working Paper 18652, forthcoming in Gopinath G., G. Grossman and K. Rogoff (eds.) *Handbook of International Economics* (Amsterdam: North Holland).

[18] Melitz M. and S. Redding (2013) Firm Heterogeneity and Aggregate Welfare, NBER Working Paper 18919.

[19] Neary P. (2004) Monopolistic competition and international trade theory, in Brakman R. and B. Heijdra (eds.), *The Monopolistic Competition Revolution in Retrospect* (Cambridge: Cambridge University Press).

[20] Ottaviano G. (2012) Agglomeration, trade and selection, CEPR Discussion Paper 9046, *Regional Science and Urban Economics*, forthcoming.

[21] Ottaviano G. and J.-F. Thisse (1999) Monopolistic competition, multiproduct firms and optimum product diversity, CEPR Discussion Paper 2151.

[22] Ottaviano G., T. Tabuchi and J.-F. Thisse (2002) Agglomeration and trade revisited, *International Economic Review* 43, 409-436.

[23] Spence M. (1976) Product selection, fixed costs and monopolistic competition, *Review of Economic Studies* 43, 217-235.

[24] Stiglitz (1975) Monopolistic competition and the capital market, in Brakman R. and B. Heijdra (eds.), *The Monopolistic Competition Revolution in Retrospect* (Cambridge: Cambridge University Press).

[25] Zhelobodko E., S. Kokovin, M. Parenti and J.-F. Thisse (2010) Monopolistic competition: Beyond the CES, CEPR Discussion Paper 7947, *Econometrica*, forthcoming.

8 Appendix A - General distribution

The analysis in the main text is based on the assumption that the distribution $G(c)$ from which entrants draw their unit labor requirements is a Pareto distribution. In this appendix we show that some key results do not depend on such assumption.

8.1 Market outcome

Instead of the Pareto distribution, consider a generic $G(c)$ such that dG is positive in $[0, c_M]$. The cutoff rule is

$$q^m(c) = \begin{cases} \frac{L}{2\gamma}(c^m - c) & c \leq c^m = \alpha - \frac{\eta}{L}Q^m \\ 0 & c > c^m \end{cases} \tag{30}$$

with $Q^m \equiv N_E^m \int_0^{c_M} q^m(c) dG(c) = N^m \int_0^{c^m} q^m(c) dG^m(c)$, $N^m = N_E^m G(c^m)$ and $G^m(c) = G(c)/G(c^m)$. The 'free entry condition' becomes

$$\frac{1}{4}\int_0^{c^m} (c^m - c)^2 \, dG(c) = \frac{\gamma f}{L} \tag{31}$$

In turn, the 'zero cutoff profit condition' becomes

$$N^m = \frac{2\gamma}{\eta} \frac{\alpha - c^m}{c^m - \bar{c}^m} \tag{32}$$

with $\bar{c}^m = \left[\int_0^{c^m} c dG^m(c)\right]$. The number of entrants is then given by $N_E^m = N^m / G(c^m)$.

8.2 Unconstrained optimum

The planner maximizes

$$\begin{array}{c} W = L + \overline{q}_0 L - f N_E + N_E \int_0^{c_M} (\alpha - c) \, q(c) dG(c) \\ -\frac{1}{2}\frac{\gamma}{L} N_E \int_0^{c_M} [q(c)]^2 \, dG(c) - \frac{1}{2}\frac{\eta}{L} \left[N_E \int_0^{c_M} q(c) dG(c)\right]^2 \end{array} \tag{33}$$

The two first order conditions are

$$\begin{array}{c} \frac{\partial U}{\partial q(c)} = \left[N_E (\alpha - c) - \frac{\gamma}{L} N_E q(c) - \frac{\eta}{L} (N_E)^2 \int_0^{c_M} q(c) dG(c)\right] dG(c) = 0 \quad \nabla c \\ \frac{\partial U}{\partial N_E} = -f + \int_0^{c_M} (\alpha - c) \, q(c) dG(c) - \frac{1}{2}\frac{\gamma}{L} \int_0^{c_M} [q(c)]^2 \, dG(c) \\ -\frac{\eta}{L} N_E \left[\int_0^{c_M} q(c) dG(c)\right]^2 = 0 \end{array}$$

As utility can take only positive values, it must be $N_E^o > 0$ at the maximum. Rearranging the former first order condition gives

$$q^o(c) = \frac{L}{\gamma}(\alpha - c) - \frac{\eta}{\gamma} N_E^o \int_0^{c_M} q^o(c) dG(c) = \frac{L}{\gamma}(\alpha - c) - \frac{\eta}{\gamma} Q^o \qquad (34)$$

with $Q^o \equiv N_E^o \int_0^{c_M} q^o(c) dG(c) = N^o \int_0^{c^o} q^o(c) dG^o(c)$, $N^o = N_E^o G(c^o)$ and $G^o(c) = G(c)/G(c^o)$. Equation (34) and the constraint $q^o(c) \geq 0$ imply that the same cutoff rule for the planner as in the main text:

$$q^o(c) = \begin{cases} \frac{L}{\gamma}(c^o - c) & c \leq c^o = \alpha - \frac{\eta}{L} Q^o \\ 0 & c > c^o \end{cases} \qquad (35)$$

Integrating (34) across c to obtain Q^o, plugging the result in $c^o = \alpha - \eta Q^o / L$ and solving for N^o gives the planner's cutoff condition analogous to (32)

$$N^o = \frac{\gamma}{\eta} \frac{\alpha - c_D^o}{c_D^o - \overline{c}^o} \qquad (36)$$

with $\overline{c}^o = \left[\int_0^{c^o} c dG^o(c) \right]$ and $G^o(c) = G(c)/G(c^o)$. Then, substituting (36) and (35) in the second first order condition yields

$$\frac{1}{2} \int_0^{c^o} (c^o - c)^2 dG(c) = \frac{\gamma f}{L} \qquad (37)$$

8.3 Constrained optimum

When the differentiated sector has to be financially self-sufficient, the constrained planner cannot affect the profit maximizing choices of firms in terms of quantities and prices but it can affect the number of firms that operate in the economy. Hence, the planner follows the same free entry condition (31) and thus chooses the same cutoff as the market

$$\frac{1}{4} \int_0^{c^m} (c^m - c)^2 dG(c) = \frac{\gamma f}{L} \qquad (38)$$

As to the number of entrants, the planner maximize utility in (33) with respect to N_E subject to the market quantities

$$q^m(c) = \begin{cases} \frac{L}{2\gamma}(c^m - c) & c \leq c^m = \alpha - \frac{\eta}{L} Q^m \\ 0 & c > c^m \end{cases} \qquad (39)$$

The first order condition of the planner's problem is

$$-f + \int_0^{c_M} (\alpha - c) q^s(c) dG(c) - \frac{1}{2} \frac{\gamma}{L} \int_0^{c_M} [q^s(c)]^2 dG(c) - \frac{\eta}{L} N_E^s \left[\int_0^{c_M} q^s(c) dG(c) \right]^2 = 0$$

Substituting (38) and (39) then gives

$$N^s = N^m + \frac{\gamma}{2\eta} \left(1 + \frac{\overline{\sigma}_c^2}{(c^m - \overline{c})^2} \right) \qquad (40)$$

where $\overline{c} = \int_0^{c^m} c dG^m(c)$ is the conditional mean and $\sigma_c^2 = \int_0^{c^m} (c - \overline{c})^2 dG^m(c)$ is the conditional variance of the unit labor requirement in the market equilibrium.

8.4 Firm selection and product variety

Equations (31) and (38) are identical. They have the same right hand side as (37) and left hand sides that differ only up to a positive multiplicative constant that is larger for the unconstrained planner. Given that $\int_0^x (x-c)^2 \, dG(c)$ is an increasing function of x, the ranking of the multiplicative constants implies a reverse ranking of cutoffs $c^o < c^s = c^m$. This generalizes Proposition 1 in the text.

Turning to the number of varieties supplied at the different outcomes, the cutoff conditions (32) and (40) readily establish that the constrained planner provides richer product variety than the market equilibrium. On the other hand, (32) and (36) are functions of the cutoffs that differ from one another only up to a positive multiplicative constant: $N^m(x) \equiv N$ and $N^o(x) = N(x)/2$. However, the sign of the derivative $N'(x)$ depends on the properties of $G(c)$:

$$\text{sign}\left(N'(c)\right) = \text{sign}\left\{-\left[x - \bar{c}(x)\right] - \left[(\alpha - x)\left(1 - \bar{c}'(x)\right)\right]\right\}$$

where $\bar{c}(x) = \left[\int_0^x c dG(c)\right] / G(x)$ and $\bar{c}'(x)$ is its derivative. The sign is ambiguous because for a generic distribution function the derivative $\bar{c}'(x)$ of the conditional mean based on right truncation can be larger than 1. Hence, $N(x)$ need not be decreasing everywhere, even if it equals 0 at $x = \alpha$ and diverges to $+\infty$ when x goes to 0. We can, nonetheless, state a sufficient condition for $\bar{c}'(x) < 1$, and, therefore, for $N'(x) < 0$. The condition is that $G(c)$ is log-concave (see Lemma 1 in Bagnoli and Bergstrom, 2005).[10] As this is only a sufficient condition, there exists a larger family of functions than the log-concave ones that guarantee $N'(x) < 0$. This ensures that, within each outcome (whether market equilibrium or unconstrained optimum), a lower cutoff is associated with richer product variety. It does not allow, however, to unambiguously rank the unconstrained planner and the market equilibrium in terms of product variety as in the Pareto case discussed in the main text.

As a final comment, it should be noted that no result is, instead, available concerning the implications of different degrees of firm heterogeneity for the efficiency gap of the market equilibrium in the case of a generic $G(c)$ as, differently from the Pareto case, in the generic case the unconditional distribution generally puts little structure on the conditional (truncated) distribution.

9 Appendix B - Homogeneous firms

For parsimony, let us focus on the unconstrained optimum and the market equilibrium. The constrained optimum can be analyzed analogously. To connect to the previous analysis, rewrite (31) and (37) respectively as

$$\frac{1}{2} G(c^o) \left\{ \left[c^o - \bar{c}(c^o)\right]^2 + \sigma_c^2(c^o) \right\} = \frac{\gamma f}{L} \tag{41}$$

[10] Most of the most commonly used distribution functions are log-concave: Uniform, Normal, Exponential, Logistic, Extreme Value, Laplace (Double Exponential), Power Function, ($c \geq 1$), Weibull ($c \geq 1$), Gamma ($c \geq 1$), Chi-Squared ($c \geq 2$), Chi ($c \geq 1$), Beta ($a \geq 1$, e). Note also that Theorem 9 in Bagnoli and Bergstrom (2005) shows that, if a probability distribution has a log-concave (log-convex) density function (cumulative distribution function), then any truncation of this probability distribution will also have a log-concave (log-convex) density function (cumulative distribution function). Thus, $N'(x) < 0$ under all commonly used distributions.

and

$$\frac{1}{4}G(c^m)\left\{[c^m-\bar{c}(c^m)]^2+\sigma_c^2(c^m)\right\}=\frac{\gamma f}{L} \tag{42}$$

where we have used the following expression for the conditional variance of the marginal cost distribution $\sigma_c^2(x)=\left[\int_0^x c^2 dG(c)\right]/G(x)-\bar{c}(x)^2$. Without heterogeneity, there is no variance in costs ($\sigma_c^2=0$), all entrants produce ($G(c^m)=1$), and c is exogenous and common to all firms with $c=\bar{c}$.

Given (41), the first best planner's solution for the cutoff is

$$c^o=\bar{c}+\sqrt{\frac{2\gamma f}{L}} \tag{43}$$

This determines the willingness to pay of consumers for any variety. As $c^o>\bar{c}$, all entrants produce. How much they produce can be determined by noticing that expression (35) implies $q^o(\bar{c})=(L/\gamma)(c^o-\bar{c})$ so that we can rewrite (43) to obtain firm output as

$$q^o=\sqrt{\frac{2fL}{\gamma}} \tag{44}$$

Then, by using (36) and (43), we find that the number of varieties supplied is

$$N^o=\frac{(\alpha-\bar{c})\sqrt{\frac{\gamma L}{2f}}-\gamma}{\eta} \tag{45}$$

Turning to the market equilibrium, expression (42) implies that, with no heterogeneity, the cutoff evaluates to

$$c^m=\bar{c}+2\sqrt{\frac{\gamma f}{L}} \tag{46}$$

Again, the willingness to pay c^m is larger than the common marginal cost $\bar{c}$, so all entrants produce. Furthermore, expression (30) implies $q^m(\bar{c})=(L/2\gamma)(c^m-\bar{c})$, which can be used together with (46) to find firm output

$$q^m=\sqrt{\frac{fL}{\gamma}} \tag{47}$$

Furthermore, (32) and (46) imply that the number of firms producing in the market economy is

$$N^m=\frac{(\alpha-\bar{c})\sqrt{\frac{\gamma L}{f}}-2\gamma}{\eta} \tag{48}$$

It is readily verified from (44) and (47) that we always have $q^o>q^m$. Moreover, from (45) and (48), we we also have $N^m>N^o$ if and only if

$$\alpha>\bar{c}+\frac{\sqrt{2}}{\sqrt{2}-1}\sqrt{\frac{\gamma f}{L}} \tag{49}$$

which shows that each firm is smaller in the market equilibrium than in the unconstrained optimum, and the market tends to overprovide variety when varieties are close substitutes (γ small) and when the fixed cost f is low compared to market size as measured by α and L. These results concur with those in Ottaviano and Thisse (1999), taking into account that they assume $L=1$.

10 Appendix C - Heterogeneity and inefficiency

10.1 Equilibrium vs. optimum

Given $c^m = 2^{1/(k+2)}c^o$ as well as (27) and (26), the percentage gap in welfare between the optimum and the market equilibrium can be written as

$$\frac{W^o - W^m}{W^o} = \frac{ac^m}{1 + \overline{q}_0 + \left[\alpha - \left(\frac{1}{2}\right)^{1/(k+2)} c^m\right]^2}$$

where $a \equiv b\alpha - dc^m$, $b \equiv (2k+3)/(k+2) - 2^{-(k+3)/(k+2)}$ and $d \equiv (k+1)/(k+2) - 2^{-2/(k+2)}$. Given that $\alpha/c^m > 1 > d/b$, it is readily verified that $a > 0$ and, consequently, $W^o > W^m$ as to be expected. Derivation with respect to c_M then implies

$$\frac{\partial\left(\frac{W^o - W^m}{W^o}\right)}{\partial c_M} = \frac{W^o\left(b\alpha - 2dc^m\right) + \left(\frac{1}{2}\right)^{1/(k+2)} 2ac^m\left(\alpha - c^o\right)}{(W^o)^2}\frac{\partial c^m}{\partial c_M}$$

which is positive given that: $\partial c^m/\partial c_M > 0$; $W^o > 0$; $(\alpha - c^o) > 0$; $a > 0$ and $b\alpha - 2dc^m > 0$ as $\alpha/c^m > 1 > 2d/b$ holds. The signs of these expressions, together with

$$\partial c^m/\partial k = c^m\left\{\ln\left(c_M/c^m\right)/(k+2) + (2k+3)/\left[(k+2)^2(k+1)\right]\right\} > 0$$

as $c_M > c^m$, and

$$\alpha\frac{\partial b}{\partial k} - c^m\frac{\partial d}{\partial k} > 0$$

as $\alpha/c^m > 1 > 2\left(2^{-2/(k+2)}\ln 2 - 2^{-1}\right)/\left[2^{-(k+3)/(k+2)}\ln 2 - 1\right]$, also ensure that

$$\frac{\partial\left(\frac{W^o - W^m}{W^o}\right)}{\partial k} = \frac{W^o\left[\left(\alpha\frac{\partial b}{\partial k} - c^m\frac{\partial d}{\partial k}\right)c^m + \left(b\alpha - 2dc^m\right)\frac{\partial c^m}{\partial k}\right] + 2ac^m\left(\alpha - c^o\right)\left[\left(\frac{1}{2}\right)^{1/(k+2)}\frac{\partial c^m}{\partial k} + \frac{\left(\frac{1}{2}\right)^{\frac{1}{k+2}+2}\ln 2}{\left(\frac{1}{2}k+1\right)^2}c^m\right]}{(W^o)^2}$$

is positive.

10.2 Equilibrium vs. constrained optimum

The percentage gap in welfare between the constrained optimum and the market equilibrium is given by

$$\frac{W^s - W^m}{W^s} = \frac{\frac{1}{8\eta}\left(\frac{c^m}{k+2}\right)^2}{1 + \overline{q}_0 + \frac{1}{2\eta}\left(\alpha - \frac{2k+3}{2(k+2)}c^m\right)^2}$$

and it is readily verified that

$$\frac{\partial\left(\frac{W^s - W^m}{W^s}\right)}{\partial c_M} = \frac{W^s + \frac{(2k+3)c^m}{4(k+2)\eta}\left[\alpha - \frac{2k+3}{2(k+2)}c^m\right]}{4\eta(k+2)^2(W^s)^2}c^m\frac{\partial c^m}{\partial c_M}$$

is positive as $W^s > 0$, $\partial c^m / \partial c_M > 0$ and $\alpha - (2k+3)/[2(k+2)]\, c^m > 0$, which is required to have $N^s > 0$.

Derivation with respect to k then yields

$$\frac{\partial \left(\frac{W^s - W^m}{W^s}\right)}{\partial k} = \frac{(c^m)^2 \left\{(k+1)\ln\left[\frac{(c_M)^2 L}{2\gamma(k+1)(k+2)f}\right] - \left(k + k^2 - 1\right)\right\}}{4\eta\,(k+1)\,(k+2)^4\, W^s}$$

which is positive (negative) for $k < k^*$ ($k > k^*$) when $\ln\left[(c_M)^2 L/(12\gamma f)\right] > 1/2$, where $k^* > 1$ is the value of k that solves:[11]

$$\ln\left\{(c_M)^2 L/[2\gamma(k+1)(k+2)f]\right\} = \left(k - 1 + k^2\right)/(k+1)$$

Otherwise, when

$$\ln\left[(c_M)^2 L/(12\gamma f)\right] < 1/2, \partial\left[(W^s - W^m)/W^s\right]/\partial k < 0$$

holds.

[11] The term $(c_M)^2 L/[2\gamma(k+1)(k+2)f]$ is larger than 1 to ensure $c^m < c_M$. Then the left hand side of the equation, $\ln\left\{(c_M)^2 L/[2\gamma(k+1)(k+2)f]\right\}$, is a positive and decreasing function of k. The right hand side, $\left(k - 1 + k^2\right)/(k+1)$, is instead a positive and increasing function of k, attaining value $1/2$ at $k = 1$. Hence, the two functions cross only once at $k = k^* > 1$ if $\ln\left[(c_M)^2 L/(12\gamma f)\right] > 1/2$.