Monopolistic Competition and Optimum Product Selection

By ANTONELLA NOCCO, GIANMARCO I.P. OTTAVIANO AND MATTEO SALTO*

Since the initial contribution by Melitz (2003) monopolistically competitive models with heterogeneous firms have changed the way economists understand the patterns of international trade by bringing firms' decision making to the forefront. Less progress has been made on normative questions with respect to which the debate has started only very recently and has mainly centered on whether firm heterogeneity brings new gains from trade (Melitz and Redding, 2013) or simply a new channel through which old gains from trade materialize (Arkolakis et al, 2012a).

Despite additional work by Arkolakis et al (2012b), this debate has essentially focused on the Krugman (1980) version of the Dixit and Stiglitz (1977) model with CES ('constant elasticity of substitution') demand and no sector other than the monopolistically competitive one ('no outside good'), and on its extension with firm heterogeneity by Melitz (2003). In this 'non-nested' CES setup the market equilibrium and the social optimum coincide, drastically reducing the potential to deal with interesting normative issues related to the many distortions that were originally associated with monopolistic competition (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 optimality of the market equilibrium has also sidelined others issues that are of crucial interest for policy making: How can the optimum be decentralized when the market outcome is inefficient? What is the best one can do when some policy tools are unavailable?

Attention to a richer set of distortions has been brought into the debate by Dhingra and Morrow (2013), who analyse the optimality of the market equilibrium when the monopolistically competitive sector is 'non-nested', utility is additive separable and demand exhibits 'variable elasticity of substitution' (VES).[1] They show that the equilibrium and the optimum do not coincide and the increase in market size associated with a move from autarky to free trade reduces the gap between them. They do not provide, however, a discussion of decentralization. This is what the present paper does.

Separable utility is a convenient simplification that allows one to gain interesting insights by limiting the channels through which monopolistically competitive firms interact. In particular, as the marginal utility from a good's consumption is independent from the quantity consumed of other goods, firms interact only through the budget constraint in the presence of a variable marginal utility of income. This paper makes the alternative, equally restrictive assumption that utility is not separable but the marginal utility of income is constant. Specifically, the paper adopts the non-separable VES framework with linear demand by Melitz and Ottaviano (2008). This complements Dhingra and Morrow (2013).[2] It also allows for a welfare analysis based on total surplus, which represents a natural first step to discuss decentralization policies.

* Nocco: University of Salento, Department of Management, Economics, Mathematics and Statistics, Ecotekne, via Monteroni, 73100, Lecce, ITALY, antonella.nocco@unisalento.it. Ottaviano: London School of Economics, Department for Economics, Hougton Street, WC2A 2AE, UK, g.i.ottaviano@lse.ac.uk. Salto: European Commission, Rue de la Loi 200, B- 1049, BELGIUM, matteo.salto@ec.europa.eu.
We thank Paolo Bertoletti, Olivier Biau, Vincenzo Denicolò, Swati Dhingra, Sergey Kokovin, Matthieu Lequien, Marc Melitz, John Morrow, Peter Neary, Mathieu Parenti, Thomas Sampson, Jacques Thisse, Evgeny Zhelobodko as well as participants to various presentations for useful comments and suggestions. The views expressed here are those of the authors and do not represent in any manner the EU Commission.

[1] For a detailed discussion of the positive (but not the normative) implications of VES vs. CES demand, see Zhelobodko et al (2013) .

[2] According to the taxonomy of Dhingra and Morrow (2013), the separable counterpart of Melitz and Ottaviano (2008) involves private markups that are positively correlated with quantity and social markups that are higher at higher levels of quantity.

The rest of the paper is organized in six sections. Section I briefly presents the model by Melitz and Ottaviano (2008). Sections II and III respectively derive the market equilibrium and the first best optimum. Section IV compares the two outcomes. Section V discusses unconstrained and constrained decentralization. Section VI concludes summarizing the main results and discussing possible directions of future research.

I. The model

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 quasi-linear utility function given by

(1)

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

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. In the limit, as η goes to zero, utility becomes separable across the differentiated varieties. 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 outcomes.

Labor is the only productive factor. It can be employed in 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 to produce the differentiated varieties under monopolistic competition. A sunk labor requirement $f > 0$ is needed to design a new variety and its production process with unit labor requirement c 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 case in which marginal productivity $1/c$ is Pareto distributed over the support $[1/c_M, \infty)$. As $k \geq 1$ rises, density is skewed towards c_M.[3]

II. Equilibrium

The labor market and the market of the homogeneous good are assumed to be perfectly competitive. This good is chosen as numeraire, which then implies that also 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 $Q^c = \int_{i\in\Omega} q_i^c di$ over the set Ω of varieties in positive supply. When a variety is produced by a firm with unit labor requirement c, the profit maximizing output satisfies

$$q^m(c) = \begin{cases} \frac{L}{2\gamma}\left(c^m - c\right) & \text{if } c \leq c^m \\ 0 & \text{if } c > c^m \end{cases} \tag{4}$$

where 'm' labels equilibrium values and $c^m = \alpha - \eta Q^m/L$ with $Q^m = \int_0^{c^m} q^m(c) dG(c)$ is the endogenous cutoff for survival: only entrants that are productive enough ($c \leq c^m$) eventually produce. For them the price that corresponds to $q^m(c)$ is $p^m(c) = (c^m + c)/2$, implying markup $\mu^m(c) = (c^m - c)/2$ and maximized profit $\pi(c) = L\left(c^m - c\right)^2/4\gamma$. Due to free entry, an entrant's expected profit is exactly offset by the sunk entry cost so that $\int_0^{c^m} \pi(c) dG(c) = f$. Given (2), this 'free entry condition' determines the unique equilibrium cutoff marginal cost

$$c^m = \left[\frac{2\gamma(k+1)(k+2)\left(c_M\right)^k f}{L}\right]^{\frac{1}{k+2}} \tag{5}$$

[3] While our analysis rests on the Pareto distribution, several results have more general validity as discussed in Nocco et al (2013).

The mass ('number') of producers as a function of this cutoff can be found by rewriting the 'zero cutoff profit condition' $c^m = \alpha - \eta Q^m/L$ as

$$N^m = \frac{2\gamma\,(k+1)}{\eta}\frac{\alpha - c^m}{c^m} \tag{6}$$

with the corresponding equilibrium number of entrants given by $N_E^m = N^m/G(c^m) = N^m\,(c_M/c^m)^k$. Finally, by substituting (4) and (6) in (1), also the equilibrium welfare level can be expressed as a function of the cutoff

$$W^m = L + \overline{q}_0 L + \frac{L}{2\eta}\left(\alpha - c^m\right)\left(\alpha - \frac{k+1}{k+2}c^m\right) \tag{7}$$

where (5) has been used to replace f.

III. Unconstrained optimum

As quasi-linear utility 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 N_E and their output levels $q(c)$ so as to maximize individual utility (1) times L, subject to the resource constraint, the varieties' unit labor requirements and the stochastic 'variety generating technology' (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 its creation).

The tradeoffs the first best planner faces when firms are heterogeneous can be highlighted by writing social welfare as

$$\begin{aligned} W = \Big[L &+ \overline{q}_0 L \\ +N_E\left(\alpha\widehat{q} - \tfrac{1}{2}\tfrac{\gamma}{L}\widehat{q}^2 - \tfrac{1}{2}\tfrac{\eta}{L}N_E\widehat{q}^2 - \widehat{c}\widehat{q} - f\right)\Big]& \\ -\left[N_E\left(\tfrac{1}{2}\tfrac{\gamma}{L}\widehat{\sigma}_q^2 + \widehat{\sigma}_{cq}\right)\right]& \end{aligned} \tag{8}$$

where $\widehat{c} = \int_0^{c_M} c dG(c)$ is the mean unit labor requirement, $\widehat{q} = \int_0^{c_M} q(c)dG(c)$ and $\widehat{\sigma}_q^2 = \left\{\int_0^{c_M}\left[q(c)\right]^2 dG(c) - \widehat{q}^2\right\}$ are the and variance of quantities, and $\widehat{\sigma}_{cq} = \left\{\int_0^{c_M} cq(c)dG(c) - \widehat{c}\widehat{q}\right\}$ is the covariance between quantities and unit input requirements, all calculated for the unconditional distribution $G(c)$. The first bracketed term in (8) 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. sunk 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 productive efficiency could be improved by assigning little production to varieties with high marginal costs ($\widehat{\sigma}_{cq} < 0$). The fact that η appears in the first but not in the second bracketed terms suggests that the degree of non-separability should not affect firm selection. We will see below that this is indeed the case.

The first order condition with respect to the output of a variety with unit input requirement c is satisfied by the optimal level

$$q^o(c) = \begin{cases} \frac{L}{\gamma}\,(c^o - c) & \text{if } c \le c^o \\ 0 & \text{if } c > c^o \end{cases} \tag{9}$$

where 'o' labels optimum values and $c^o = \alpha - \eta Q^o/L$ with $Q^o = N_E^o\int_0^{c_M} q^o(c)dG(c)$ is the planner's cutoff such that $q^o(c) \ge 0$ only for $c \le c^o$. The conditional distribution of unit input requirements for varieties that the planner actually supplies is $G^o(c) = G(c)/G(c^o)$, so their number N^o satisfies $N^o = G(c^o)N_E^o$. Given (3), first best output (9) would clear the market in the decentralized scenario only if each producer priced at its own marginal cost $p^o(c) = c$. Solving $c^o = \alpha - \eta Q^o/L$ for N^o gives the planner's cutoff condition analogous to the market 'zero cutoff profit condition' (6)

$$N^o = N_E^o G(c^o) = \frac{\gamma\,(k+1)}{\eta}\frac{\alpha - c^o}{c^o} \tag{10}$$

Together with (10), the first order condition with respect to the number of varieties gives the analogue of the market 'free entry condition'. This can be solved for the optimal cutoff

$$c^o = \left[\frac{\gamma\,(k+1)(k+2)\,(c_M)^k\,f}{L}\right]^{\frac{1}{k+2}} \tag{11}$$

which then determines the first best number of varieties through (10). This number is a decreasing function of η: the higher the degree of non-separability (larger η), the less product variety is provided. Lastly, (9) and (10) can be plugged into (1) to express the welfare level at the first best optimum as a function of the cutoff

$$W^o = L + \overline{q}_0 L + \frac{L}{2\eta}\left(\alpha - c^o\right)^2 \tag{12}$$

where (11) has been used to substitute for f.

IV. Equilibrium vs. unconstrained optimum

The efficiency of the market outcome can be evaluated along several dimensions: the number of entrants N^E, the number of varieties produced N^m, the cost distribution of producers and their output levels as dictated by the cutoff c^m.

Comparing (5) with (11) 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. Intuitively, as discussed in Section II, in the market equilibrium the markup of a firm with marginal cost c equals $\mu^m(c) = (c^m - c)/2$, so more productive firms absorb half of their cost advantage into fatter markups making it inefficiently easy for less productive firms to survive. Hence, we have:

PROPOSITION 1: ***(Selection)*** *In the market equilibrium firm selection is weaker than optimal.*

This is reflected in the firm size distribution. Comparing (4) and (9) shows that $q^m(c) > q^o(c)$ if and only if $c > [2^{(k+1)/(k+2)} - 1]c^m$, which falls in the relevant interval $[0, c^m]$. Hence, the market equilibrium oversupplies high cost varieties with $c \in ([2^{(k+1)/(k+2)} - 1]c^m, c^m]$ and undersupplies low cost ones with $c \in [0, [2^{(k+1)/(k+2)} - 1]c^m)$. Given that the price ratio of less to more productive firms is smaller than their cost ratio, the corresponding quantity ratio is inefficiently large and 'within-sector misallocation' materializes as a lack of market concentration. Given (4) and (9), markup pricing also implies that in the market equilibrium firms are on average smaller than optimal as $c^o < c^m$ dictates $\overline{q}^m < \overline{q}^o$. We can thus state:

COROLLARY 1: ***(Quantity)*** *In the market equilibrium: (i) average firm size is smaller than optimal; (ii) low cost firms are smaller and high cost firms are larger than optimal.*

Turning to the number of varieties supplied, N^m and N^o cannot be ranked unambiguously. In particular, since $c^m = 2^{1/(k+2)}c^o$, we have $N^m > N^o$ as long as $\alpha > \alpha_1$ with

$$\alpha_1 \equiv \frac{c^o}{2^{(k+1)/(k+2)} - 1}$$

This is the case when α as well as L are large and when γ, f as well as c_M are small. The gap between N^m and N^o is a decreasing function of η: the higher the degree of non-separability (larger η), the smaller the market inefficiency in terms of product variety. We can thus state:

COROLLARY 2: ***(Variety)*** *In the market equilibrium product variety is richer (poorer) than optimal 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 proposition has an interesting implication for the impact of larger market size driven by the integration of previously autarkic national markets as in Dhingra and Morrow (2013). 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 could cause the transition from a situation in which product variety is inefficiently poor* ($N^m < N^o$) *to a situation in which it is inefficiently rich* ($N^m > N^o$). An analogous result holds for the number of varieties created. The only difference lies in the associated threshold for α, which evaluates to

$$\alpha_2 \equiv \frac{2^{2/(k+2)} - 1}{2^{1/(k+2)} - 1}c^o$$

Given $\alpha_1 < \alpha_2$, 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 little entry and too much variety. Differently, the total outputs of the differentiated varieties at the equilibrium and at the optimum can be unambiguouly ranked as $N^o\overline{q}^o > N^m\overline{q}^m$: in the market equilibrium the total output of the differentiated varieties is smaller than optimal. Markup pricing for the differentiated varieties implies that consumption is inefficiently biased towards the numeraire good leading to 'between-sector misallocation'. Unsurprisingly, also the levels of welfare (7) and (12) can be unambiguously ranked as $c^m = 2^{1/(k+2)}c^o$ implies $W^m < W^o$.

V. Decentralization

The first best optimum can be decentralized through a firm-specific per-unit production subsidy $s^o(c) = c^o - c$ accompanied by a lump-sum entry tax $T^o = f$ per entrant and a lump-sum tax $N_E^o T^o$ on consumers. The production subsidy is decreasing in the marginal cost, being zero for firms with $c = c^o$, negative ('tax') for high cost firms with $c \in (c_o, c_M]$ and positive for low cost firms with $c \in [0, c_o)$. Given the optimal cutoff c^o, the production subsidy delivers the optimal number of varieties N^o and the optimal output levels $q^o(c)$ with associated marginal cost prices $p^o(c)$. Given optimal output levels and prices, the entry tax ensures that c^o indeed solves the corresponding 'free entry condition' $\int_0^{c^o} \left[p^o(c) + s^o(c) - c\right] q^o(c) dG(c) - T^o = f$. As only entrants with $c \le c^o$ eventually produce, (11) implies that the total production subsidy is twice the entry tax revenues: $N^o \int_0^{c^o} s^o(c) q^o(c) dG^o(c) = 2N_E^o T^o$. A total lump-sum tax $N_E^o T^o$ on consumers is thus needed to finance the deficit: production subsidies are convered by lump-sum transfers whose burden is equally shared between entrants and consumers.

When differentiated subsidies or lump-sum instruments for firms or consumers are not available, the first best optimum cannot be decentralizes. Different second best scenarios arise depending on the specific policy tools that are available. Traditional analyses without firm heterogeneity have focused on the implications of the lack of lump-sum instruments for firms (e.g. Dixit and Stiglitz, 1977). Here we highlight the novel issues that arise with heterogeneous firms when not only lump-sum instruments for firms but also differentiated subsidies are not available.

When the per-unit production subsidy cannot be differentiated across firms, marginal cost pricing cannot be enforced. With a common subsidy s, profit becomes $\pi(c) = \left[p(c) + s\right] q(c) - cq(c)$. This is equivalent to a situation in which the marginal costs of all firms are reduced by the same amount. Given that also their equilibrium prices are reduced by the same amount, it can be shown that output levels and maximized profits are the same as in the market equilibrium without the subsidy. Thus, in the absence of lump-sum tools for firms, the corresponding free entry condition is also the same, $\int_0^{c^m} \frac{L}{4\gamma} (c^m - c)^2 dG(c) = f$, resulting in the same cutoff marginal cost c^m. As the cost distribution of producers and the quantity each of them supplies cannot be affected, the constrained planner is left facing a tradeoff between the number of varieties and the supply of the homogeneous good.[4] The optimal balance is struck when the number of producers equals

$$N^s = \frac{2\gamma\ (k+1)}{\eta} \frac{\alpha - \frac{2k+3}{2(k+2)} c^m}{c^m} \tag{13}$$

which can be implement through a common per-unit production subsidy $s^s = c^m/[2(k+2)]$ financed through a lump-sum tax on consumers. Comparing (13) with (6) reveals that product variety is richer in the constrained optimum than in the market equilibrium. The ensuing welfare gain is

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

which is a decreasing function of η: the higher the degree of non-separability (larger η), the smaller the inefficiency of the market equilibrium with respect to the second best.

VI. Conclusion

Variable demand elasticity and endogenous firm heterogeneity enrich the dimensions along which the market equilibrium of monopolitically competitive models can err with respect to the social optimum. This raises new questions on the social optimality of market outcomes and the decentralization of optimum outcomes through policy instruments. These questions have been so far neglected with the excep-

[4] Of course, if other sets of policy instruments were available, the constrained planner would face different tradeoffs.

tion of Dhingra and Morrow (2013), who address optimality under additive separable utility but do not discuss decentralization.

Under admittedly restrictive assumptions on the functional form of non-separable utility and on the parametrization of firm heterogeneity, we have complemented the study of optimality in Dhingra and Morrow (2013) and provided novel insights on unconstrained and constrained decentralization of optimal outcomes. In our framework, non-separability turns out to be relevant only for product variety, with respect to which stronger non-separability leads to smaller market inefficiency. Relative to the uncostrained optimum, in the market equilibrium firm selection is too weak, average firm size is too small, low cost firms are too small and high cost firms are too large. Moreover, product variety is too rich (poor) 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 marginal cost realizations is small (large). We have also shown than the unconstrained optimum can be decentralized through differentiated production subsidies across heterogeneous producers financed through lump-sum taxes equally shared by entrants and consumers. When production subsidies cannot be differentiated and lump-sum transfers from entrants are not viable, the constrained optimum can be decentralized through a common production subsidy financed by a lump-sum tax on consumers.

These results may extend beyond the proposed framework under alternative demand systems with non-separable utility and variable demand elasticity, such as those proposed by Bertoletti (2006), Behrens and Murata (2007) and Arkolakis et al (2012b). Checking whether this is indeed the case is left to future research.

REFERENCES

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

[2] Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodriguez-Clare (2012b) The Elusive Pro-Competitive Effects of Trade, Yale University, mimeo.

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

[4] Bertoletti (2006) Logarithmic quasi-homothetic preferences, *Economics Letters* 90, 433–439.

[5] Dhingra S. and J. Morrow (2013) Monopolistic competition and optimum product diversity under firm heterogeneity, LSE, mimeo.

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

[7] Nocco A., G. Ottaviano and M. Salto (2013) Monopolistic Competition and Optimum Product Selection: Why and how heterogeneity matters, CEPR Discussion Paper 9417.

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

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

[10] 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).

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

[12] Zhelobodko E., S. Kokovin, M. Parenti and J.-F. Thisse (2012) Monopolistic competition: Beyond the constant elasticity of substitution, *Econometrica* 80, 2765-2784.