DEMS WORKING PAPER SERIES

Insufficient Entry in Monopolistic Competition

Paolo Bertoletti Federico Etro

No. 543 – July 2024

Department of Economics, Management and Statistics
University of Milano – Bicocca
Piazza Ateneo Nuovo 1 – 2016 Milan, Italy
http://dems.unimib.it/

Insufficient Entry in Monopolistic Competition

Paolo Bertoletti and Federico Etro[1]

University of Milan-Bicocca and University of Florence

July 2024

Keywords: Entry, Monopolistic competition, Business stealing, Heterogeneous firms
JEL Codes: D11, D43, L11

Abstract

We study entry in markets with monopolistic competition under quasi-linear preferences, with homogeneous and heterogeneous firms. For common demand systems with a price aggregator that is a demand shifter, we show that entry tends to be insufficient: namely that, given market pricing, the business stealing effect of entry cannot dominate the consumer surplus effect. We then identify preferences that deliver efficient production and selection of firms (including the isoelastic demand case), confirming the insufficient entry result also compared to first-best allocations, and discuss a specification (which includes the Logit case) that also delivers efficient entry. Finally, we introduce more general quasi-linear preferences (nesting those of Spence, Melitz-Ottaviano and other cases) that generate flexible demand systems depending on a price aggregator. In this framework, we show that competitive effects of entry on prices actually strengthen the case for insufficient entry, and discuss conditions for its emergence.

[1] We are grateful to Alberto Bennardo, Gino Gancia, Alessandro Lizzeri, Salvatore Piccolo, Patrick Rey, Yaron Yehezkel, Kreso Zigic and seminar participants at the CRESSE Conference in Crete, the IO Conference at Passo del Tonale, CERGE in Prague, and the University of Milan-Bicocca for useful comments. *Correspondence.* Paolo Bertoletti: Dept. of Economics, Management and Statistics, University of Milan-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy. Phone: +39-02-64483105, email: *paolo.bertoletti@unimib.it.* Federico Etro: Florence School of Economics and Management, Via delle Pandette 32, Florence, 50127. Phone: +39-055-2759603, email: *federico.etro@unifi.it.*

1 Introduction

As well known, free entry is unlikely to deliver efficient allocations in imperfectly competitive markets. Entry decisions by firms do not completely internalize the positive impact on consumers generated by the introduction of new and differentiated goods: this *consumer surplus effect* pushes for the creation of an insufficient number of firms and limited competition. But the same entry decisions neglect the negative externalities generated on competitors whose sales decrease with the number of rivals: this *business stealing effect* pushes for the creation of an excessive number of firms and a waste of entry costs. In general, either effect could be dominant depending on the relevance of product differentiation, scale economies and competitive effects of entry on prices, as shown in a second-best environment with monopolistic competition pricing by Spence (1976), Mankiw and Whinston (1986) and Vives (1999).[2] Ambiguous results emerge also when the monopolistic competition equilibrium is compared to a first-best environment with marginal cost pricing, as shown with homogeneous firms by Spence (1976) and Dixit and Stiglitz (1977), and with heterogeneous firms by Nocco, Ottaviano and Salto (2014) and Dhingra and Morrow (2017, 2019), who also unveil market inefficiencies in the selection of firms.

In this work we argue that the case for insufficient entry in monopolistic competition is stronger than usually thought. We show that, for demand systems that are commonly used in partial equilibrium analysis of markets with product differentiation, monopolistic competition generates insufficient entry *à la* Mankiw and Whinston (1986), in the sense that the business stealing effect cannot dominate the consumer surplus effect at the equilibrium. We also show that in a more general environment where entry can generate competitive effects that reduce prices, these effects weaken business stealing and therefore push also in the direction of insufficient entry.

We study monopolistic competition for demand systems where an aggregator of prices is a demand shifter, as in common Logit and isoelastic demand models. The general case is obtained through quasi-linear preferences represented by an indirect utility that depends on an additive price aggregator. These preferences generate direct demands that satisfy the so-called *Independence of Irrelevant Alternatives* (IIA) property, by which the relative demand of two goods depends only on their two prices (see e.g. Anderson, de Palma and Thisse, 1992), and the demand elasticity is unaffected by entry: therefore we refer to these as quasi-linear IIA preferences. Special versions featuring a *Constant Relative Elasticity of Substitution and Surplus* (CRESS preferences) have been implicitly used in a variety of partial equilibrium settings of product differentiation, as in models with Logit and isoelastic demand systems. More general versions have been recently used by Nocke and Schutz (2018) to study multiproduct pricing,

[2] The analysis of imperfect competition with strategic interactions has instead emphasized excess entry results under Cournot competition with homogeneous goods (Mankiw and Whinston, 1986; Suzumura, 2012; Bisceglia *et al.*, 2023) and under Bertrand competition with spatial differentiation (Salop, 1979). The empirical evidence on total welfare effects of entry, however, remains inconclusive (Berry and Waldfogel, 1999).

by Anderson, Erkal and Piccinin (2020) to study aggregative games and commitments by market leaders, and by Etro (2023) to analyze pricing by sellers on platforms. Here we establish the properties of the monopolistic competitive equilibrium under IIA preferences, and compare it with constrained and unconstrained optimal allocations under homogeneous and heterogeneous firms.

In our setting, a larger market does not create competitive effects on prices and does not affect firm selection, as in closed-economy models *à la* Krugman (1980) and Melitz (2003), but can increase more or less than proportionally the mass of firms. As a consequence, individual spending in the differentiated goods can either increase or decrease, and the gains from variety associated with a larger market size can be either amplified or dampened. Our main findings concern the welfare impact of entry. We identify a preference specification for which the equilibrium is efficient in a first-best sense, and this includes a version of the Logit demand system.[3] Beyond this, we point out a general tendency toward insufficient entry of firms compared to the constrained optimal allocation. In particular, with homogenous firms, entry is always insufficient compared to a constrained optimal allocation where the planner controls entry taking monopolistic competition pricing as given. In the case of heterogeneous firms we confirm this result at the equilibrium when the planner controls the mass of firms (taking as given pricing and market selection) and also when the planner controls both the mass of firms and their selection (taking as given only pricing). The comparison to the first-best allocation is more complex, but in the case of CRESS preferences we show that production levels and firms' selection are efficient, and entry tends again to be insufficient, with full efficiency for the preferences specification that nests the Logit case.

To relate our results to instances of excess entry emerging in other frameworks, we finally introduce *quasi-linear preferences à la Gorman-Pollak*, which generate demand systems that depend on a price aggregator in a more general way.[4] This allows us to nest our baseline setting as well as the Spence (1976) model, the Melitz and Ottaviano (2008) model and other cases. A crucial difference is that these models can account for a *competitive effect*, whereby entry reduces the equilibrium markups (in spite of the absence of strategic interactions under monopolistic competition). In this more general setting a competitive effect actually strengthens the case for insufficient entry *à la* Mankiw and Whinston (1986), reducing the relative importance of business stealing. The simple reason is that, by adjusting prices downward after entry of rivals, the incumbent firms limit their losses of sales and profits. Accordingly, our results of insufficient entry are not driven by the absence of competitive effects, and can be actually amplified by them.[5] However, excess entry can still emerge

[3] Full efficiency applies to the Logit case without exogenous outside options, otherwise insufficient entry holds (as in Besanko, Perry and Spady, 1990).

[4] The original version of the Gorman-Pollak preferences without outside good (see Gorman, 1970 and Pollak, 1972) has been recently analyzed by Fally (2019, 2022), Bertoletti and Etro (2021) and Macedoni and Weinberger (2022).

[5] In addition, we also show that with heterogenous firms selection effects tend to reduce both the consumer surplus and the business stealing effects of entry, with an ambigous welfare

when entry increases the ratio of firms' revenue and consumers' surplus generated by each variety or there are strong anti-competitive effects. We provide conditions for insufficient entry to go through in this more general environment, and exemplify our results within the model of Melitz and Ottaviano (2008).

Our work is related to the vast body of literature on monopolistic competition equilibria and their comparison to the efficient allocations with heterogeneous firms and preferences with a variable demand elasticity.[6] In particular, Dhingra and Morrow (2019) have characterized the efficient allocation for the Dixit and Stiglitz (1977) model (with a given amount of resources allocated to the differentiated goods) and compared it to the monopolistic competition equilibrium with heterogeneous firms. They have shown that efficiency holds under CES (Constant Elasticity of Substitution) preferences and they have characterized the inefficiencies that emerge beyond the CES case in the allocation of production across firms and in their selection. Similar inefficiencies emerge with other preferences (Nocco, Salto and Ottaviano, 2014, 2017; Bertoletti, Etro and Simonovska, 2018; Bagwell and Lee, 2023). Here, following Spence (1976), we focus on an environment where total expenditure over the differentiated goods is endogenous, confirming related inefficiencies and identifying the class of CRESS preferences for which production and selection are efficient (and also entry can be efficient in a particular case).

Our emphasis on constrained optimality is relevant for policy applications to traditional markets, where insufficient entry can be avoided through appropriate subsidies to entry and production, while marginal cost pricing is hard to implement in practice. Moreover, vertical integration has been often regarded as beneficial in markets characterized by monopolistic competition downstream to contrast excess entry (Kuhn and Vives, 1999), which would not be the case in our environment. For other applications to industrial organization we should mention monopolistic competition among sellers on digital platforms: in such a context, insufficient entry of third-party sellers can be addressed directly by policies implemented by platforms or indirectly through platform regulation (Zennyo, 2022; Etro, 2023; Anderson and Bedre-Defolie, 2024). In general equilibrium applications with costly trade between countries, our framework produces beneficial selection effects of trade liberalization *à la* Melitz (2003) and Melitz and Ottaviano (2008) under a flexible demand system, and could be used for further explorations of industrial and trade policy.

The paper is organized as follows. Section 2 describes our demand setting. Section 3 analyzes the case of homogeneous firms and the comparative statics of the monopolistic competition equilibrium, deriving the main insufficient entry result. Section 4 extends the baseline analysis to heterogeneous firms. Section 5 introduces more general quasi-linear preferences to clarify the relation of the previous result with other models. Section 6 concludes. Appendix A analyses the general case and Appendices B and C provide further details.

impact.

[6] See Dhingra and Morrow (2017, 2019), Nocco, Ottaviano and Salto (2014, 2017, 2019), Bertoletti and Etro (2017, 2021), Simonovska (2015), Fally (2019, 2022), Anderson and de Palma (2020), Macedoni and Weinberger (2022), Bagwell and Lee (2023) and others.

2 The model

We study a model of monopolistic competition where the demand of differentiated goods is generated by L consumers with preferences represented by the following, quasi-linear indirect utility:

$$V = H\left(\int_{\omega\in\Omega} v(p(\omega))d\omega\right) + E. \tag{1}$$

Here E is the total individual expenditure allocated between a *numéraire* and the differentiated goods of type $\omega \in \Omega$, which are purchased at price $p(\omega)$ and generate an "incremental" surplus function $v(p(\omega))$ assumed positive, strictly decreasing and convex in the price, with $v(\infty) = 0$. The price aggregator:

$$A = \int_{\omega\in\Omega} v(p(\omega))d\omega \tag{2}$$

is additive across varieties, and changes in its value provide a sufficient statistic for changes in consumer welfare. The $H(A)$ transformation is increasing and concave in the aggregator, and we assume that it is a convex function of the underlying prices to satisfy the regularity conditions of an indirect utility function. We will discuss below a restriction that is necessary for this to be the case.

Assuming the suitable differentiability, the Roy's identity provides the demand of variety ω as:

$$q(\omega) = |v'(p(\omega))|\, H'(A), \tag{3}$$

which decreases in its own price $p(\omega)$ and in the price aggregator A, which is effectively a demand shifter for all goods. Therefore, a market providing more and/or cheaper varieties implies a lower demand for each individual product. Notice that the relative demand of two varieties $q(\omega)/q(\iota)$ depends only on their two prices $p(\omega)$ and $p(\iota)$, as in common aggregative models of product differentiation satisfying the IIA property (Anderson, Erkal and Piccinin, 2020), and that demand elasticity depends only on the own price.

The direct utility dual to (1) is implicitly defined by:

$$U = H(\bar{A}) + \int_{\omega\in\Omega} q(\omega)\, v'^{-1}\left(\frac{-q(\omega)}{H'(\bar{A})}\right) d\omega + Y, \tag{4}$$

where the quantity aggregator $\bar{A}$ satisfies $\bar{A} = \int_{\Omega} v[v'^{-1}(-q(\omega)/H'(\bar{A}))]d\omega$, and Y is consumption of the outside *numéraire*. We derive this expression in the Appendix and provide some explicit examples for relevant specifications. The corresponding inverse demand system depends on $\bar{A}$, which is in general not additive across varieties. Notice that the direct utility function (4) depends on more than one aggregator, and differs from the one used by Spence (1976),[7]

[7] As reviewed in Appendix B, the preferences of Spence (1976) can be represented by a direct utility $U = \Gamma(B) + Y$, where $B = \int_{\Omega} u(q(\omega))d\omega$ is a quantity aggregator that is additive across varieties, and Γ and u are increasing, concave functions with $u(0) = 0$. These preferences provide inverse demands that depend on the own quantity and the common aggregator B.

with a notable exception arising when an isoelastic transformation is coupled with a power surplus function. In Section 5 and Appendix A we will examine more general quasi-linear preferences, nesting our IIA ones and also those of Spence (1976), Melitz and Ottaviano (2008) and others.

We should remark that our framework differs from settings without an outside good, where the whole income is allocated across the differentiated products, as in common versions of the Dixit and Stiglitz (1977) model with direct additivity, or related models with indirect additivity (Bertoletti and Etro, 2017). In the current setting, any income increase is spent in the *numéraire* without impact on the demand for differentiated varieties, but the expenditure on these varieties $\Upsilon = \int_{\Omega} p(\omega)q(\omega)d\omega$ is endogenous and depends on prices and market dimension.

In particular, we study a (large) market where each variety is provided by a single firm with a common fixed cost of entry $F > 0$. The gross profits of an active firm producing variety ω with a marginal cost c and setting price $p(\omega)$ are:

$$\pi(\omega) = (p(\omega) - c)q(\omega)L, \tag{5}$$

where demand $q(\omega)$ is provided by (3) and the aggregator is given with respect to the individual choice of each firm under monopolistic competition.

In the next section we examine the case of homogeneous firms with a common marginal cost, as in early applications of monopolistic competition *à la* Spence (1976), Dixit and Stiglitz (1977), and Mankiw and Whinston (1986). Then we will extend our baseline analysis to the more general case of heterogeneous firms as in Melitz (2003), Melitz and Ottaviano (2008) and Dhingra and Morrow (2019), where the marginal cost of each firm is drawn from a continuous distribution upon the payment of the entry cost, and production can involve an additional fixed cost. In the remaining of this section we discuss further the nature of our preferences.

2.1 Properties of the demand systems

Substitutability among the differentiated goods depends on the shape of the incremental surplus function. However, substitutability between the differentiated goods and the outside commodity is also affected by the transformation of the price aggregator. It is convenient to provide some measures of these substitutabilities because they will play a central role in the analysis of equilibrium and optimal allocations.

We start by defining the following first-order and second-order elasticities of the surplus function $v(p)$:

$$\zeta(p) \equiv \frac{-v'(p)p}{v(p)} \quad \text{and} \quad \varepsilon(p) \equiv \frac{-v''(p)p}{v'(p)}.$$

The first is the direct elasticity of surplus with respect to the price. A higher value of $\zeta(p)$ implies that the introduction of a new good creates a higher revenue to its producer compared to the surplus generated for the consumers, which weakens the consumer surplus effect of entry. The second elasticity captures the elasticity of substitution between variety ω and any other variety, and determines the pricing rules of firms because it corresponds to the own price elasticity of demand:

$$\left|\frac{\partial \ln q(\omega)}{\partial \ln p(\omega)}\right| = \varepsilon\left(p(\omega)\right).$$

A higher value of $\varepsilon(p)$ implies that the introduction of goods takes place at lower prices and with a lower profitability. Accordingly, if we define the *relative elasticity of substitution and surplus* as:[8]

$$\eta(p) \equiv \frac{\varepsilon(p)}{\zeta(p)}, \tag{6}$$

a high value of it is associated to a strong consumer surplus effect relative to the business stealing effect. "Average" values of the surplus and demand elasticities across all varieties are defined as follows:

$$\overline{\zeta} \equiv \int_{\Omega} \zeta(p(\iota)) \frac{v(p(\iota))}{\int_{\Omega} v(p(\omega)) d\omega} d\iota > 0, \quad \overline{\varepsilon} \equiv \int_{\Omega} \varepsilon\left(p(\iota)\right) \frac{p(\iota)q(\iota)}{\int_{\Omega} p(\omega)q(\omega)d\omega} d\iota > 0.$$

We also define the following measure of curvature of the transformation function:

$$\rho\left(A\right) \equiv \frac{-H''\left(A\right)A}{H'\left(A\right)} > 0,$$

which captures the elasticity of demand with respect to the aggregator. A high value of $\rho(A)$ implies that the introduction of new goods reduces significantly the demand and thus the profits of the incumbent firms, and therefore it strengthens the relevance of the business stealing effect.

Finally, we can measure the substitutability between the differentiated varieties and the outside good. A natural approach is to look at how much the expenditure Υ over the differentiated products reacts to a proportional increase of all their prices, for a given set Ω of consumed varieties (Bertoletti, 2018). This can be measured by:

$$\Psi \equiv \left. \frac{d \ln\left\{\int_{\Omega} \lambda p\left(\omega\right)\left|v'\left(\lambda p\left(\omega\right)\right)\right| H'\left(\int_{\Omega} v\left(\lambda p\left(\iota\right)\right) d\iota\right) d\omega\right\}}{d \ln \lambda} \right|_{\lambda=1} = 1 - \overline{\varepsilon} + \rho\left(A\right)\overline{\zeta}.$$

The average "direct" impact of an hypothetical proportional price increase on spending in differentiated goods is captured by $1-\overline{\varepsilon}$ and depends therefore on the average demand elasticity. This is countered by an "indirect" effect due to the reduction of the price aggregator, whose average impact is given by $\rho(A)\overline{\zeta} > 0$. When $\Psi < 0$ overall spending in the differentiated goods decreases after a proportional price increase, and spending in the outside good rises. When $\Psi > 0$ spending in the differentiated goods increases, thereby

[8] The relative elasticity $\eta(p)$ is a measure of curvature of the incremental surplus, such that $v(p)$ is (locally) log-convex if and only if $\eta(p) \geq 1$.

reducing spending in the outside good (accordingly, Ψ is an *inverse* measure of outside substitutability). In the extreme case where $\Psi = 1$, namely when $\overline{\varepsilon} = \rho(A)\overline{\zeta}$, the *numéraire* cannot substitute for the differentiated products and their expenditure increases in the same proportion as the prices: in such a case the preferences are weakly convex and exhibit interesting welfare properties, as we will see.

Accordingly, in what follows we assume $\Psi \leqslant 1$, or equivalently:

$$\overline{\varepsilon}/\overline{\zeta} \geqslant \rho(A) \quad \text{for all prices and sets of goods } \Omega. \tag{7}$$

This requires that the expenditure in the differentiated goods does not increase more than proportionally following a proportional increase of all prices. A violation of this condition would imply an increase of the aggregate demand of differentiated products following a proportional price increase and is inconsistent with preference convexity.

2.2 Examples of IIA preferences

In this section we introduce some specifications of indirect utility functions that will be used to illustrate our results, and a family of preferences whose properties will play a relevant role in our analysis. In Appendix B we present the corresponding direct utility functions.

Let us begin with the *LogSumExp* (LSE) specification of indirect utility. Mathematically, the LSE function is a convex function defined as the logarithm of the sum of the exponential of the arguments. It combines the logarithmic transformation $H(A) = \log A$ with the exponential surplus functions $v(p) = e^{-\alpha p}$, where $\alpha > 0$. This implies a unitary elasticity of demand with respect to the aggregator and $\zeta(p) = \varepsilon(p) = \alpha p$, and results in the following indirect utility:

$$V = \log\left(\int_\Omega e^{-\alpha p(\omega)} d\omega\right) + E. \tag{8}$$

The LSE specification provides the demand:

$$q(\omega) = \frac{\alpha e^{-\alpha p(\omega)}}{\int_\Omega e^{-\alpha p(\iota)} d\iota},$$

which corresponds to a version of the multinomial Logit demand under discrete choices (see Anderson, de Palma and Thisse, 1992, Ch. 2) and implies a constant "aggregate" quantity, namely $\int_\Omega q(\omega)d\omega = \alpha$. The demand functions remain constant if all prices are changed by the same constant. This specification is of particular interest, because, as we will see, the monopolistic competition equilibrium implements the efficient allocation.[9]

[9] Many theoretical and empirical applications of the multinomial logit demand depart from the LSE specification by considering an additional option providing an exogenous level of surplus. For this reason Besanko, Perry and Spady (1990) find insufficient entry. The same would happen in the model of Nocke and Schutz (2018) where $H(A) = \log(A + K)$ for a positive constant K, which implies $\rho(A) < 1$.

For another example, let us consider the logarithmic transformation coupled with an isoelastic surplus function $v(p) = p^{1-\varepsilon}$ for a constant $\varepsilon > 1$. This implies $\zeta(p) = \varepsilon - 1$, $\varepsilon(p) = \varepsilon$, and the following indirect utility:

$$V = \log\left(\int_{\Omega} p(\omega)^{1-\varepsilon} d\omega\right) + E, \tag{9}$$

which provides an instance of the classic isoelastic demand

$$q(\omega) = \frac{(\varepsilon-1)p(\omega)^{-\varepsilon}}{\int_{\Omega} p(\iota)^{1-\varepsilon} d\iota},$$

as in common applications of monopolistic competition with CES preferences. Under this *log*-CES specification the demand functions are homogeneous of degree -1 with respect to prices, meaning that expenditure Υ remains constant after a proportional increase of all prices ($\Psi = 0$). This specification features a direct utility, derived in Appendix B, which depends on a single CES aggregator of quantities, and is therefore nested in the preferences adopted by Spence (1976).

CRESS preferences In the previous examples the surplus functions exhibit a constant ratio (6) of the elasticity of substitution and the elasticity of surplus, namely, $\eta = 1$ in the LSE specification and $\eta = \frac{\varepsilon}{\varepsilon-1} > 1$ in the isoelastic case. This property (a generalization of homogeneity) identifies the CRESS family of surplus functions, where the acronym stands for *Constant Relative Elasticity of Substitution and Surplus*. In particular, it can be shown that the ratio of the elasticity of substitution and the elasticity of the surplus is a constant η if and only if the surplus function belongs to the following family:

$$v(p) = \begin{cases} ae^{-\alpha p} & \text{if } \eta = 1 \text{ and } a > 0, \\ (a-\alpha p)^{\frac{1}{1-\eta}} & \text{if } \eta \in (0,1) \text{ and } a > \alpha p > 0, \\ (\alpha p - a)^{\frac{1}{1-\eta}} & \text{if } \eta \in (1,\infty) \text{ and } \alpha p > a \geqslant 0, \end{cases} \tag{10}$$

for $\alpha > 0$ (see Etro, 2021). We call *CRESS preferences* the quasi-linear preferences characterized by a surplus function which belongs to this family and *any* transformation $H(A)$ such that $\rho(A) \leqslant \eta$ to satisfy condition (7).[10] These preferences are of particular interest because, as we will see, in a monopolistic competitive equilibrium they generate efficient production levels as well as efficient firm selection.

Our last example is based on CRESS preferences under the restriction of an isoelastic transformation $H(A) = \frac{A^{1-\rho}}{1-\rho}$, where the parameter $\rho > 0$ determines the constant elasticity of demand with respect to the aggregator (clearly, the logarithmic transformation arises as a limit case for $\rho \to 1$). This provides the indirect utility:

$$V = \frac{\left(\int_{\Omega} v(p(\omega))d\omega\right)^{1-\rho}}{1-\rho} + E, \tag{11}$$

[10] Notice that a linear demand system arises for $\eta = 1/2$, implying $q(\omega) = 2\alpha\left[a - \alpha p(\omega)\right] H'(A)$.

where the surplus function belongs always to the CRESS family (10), generalizing our previous examples.[11]

3 Monopolistic competition with homogeneous firms

In this section we study monopolistic competition assuming that all firms have the same marginal cost, $c > 0$, and that the only fixed cost is the entry cost $F > 0$. We then compare the equilibrium with optimal allocations.

Each firm sets its price to maximize profits (5) taking as given the aggregator in the demand function (3). The equilibrium price p, which is the same for all firms, is implicitly given by:

$$p \equiv \frac{\varepsilon(p)\, c}{\varepsilon(p) - 1}. \tag{12}$$

To satisfy the first-order and second-order conditions for profit maximization, it is assumed that $\varepsilon(p) > 1$ and $2\varepsilon(p) \geq \phi(p)$, where $\phi(p) \equiv \frac{-v'''(p)p}{v''(p)}$. According to standard results, the equilibrium price increases in the marginal cost, with undershifting (overshifting) of cost changes whenever the demand elasticity $\varepsilon(p)$ is increasing (decreasing). Moreover, the price is independent from the price aggregator, and therefore from the mass of firms. In practice, this model does not generate competitive (or anti-competitive) effects of entry under monopolistic competition.[12] However, as we will show in Section 5, the neutrality of entry on prices is not crucial for our welfare results, in the sense that competitive effects actually strengthen the case for insufficient entry.

Notice that our condition (7) for the convexity of preferences implies that:

$$\eta(p) \geqslant \rho(A) \tag{13}$$

must hold for any price p and aggregator $A = nv(p)$, where n is the number of active firms. The value of $\eta(p)$ is also inversely related to profitability, since the equilibrium gross profits can be written as:

$$\pi = \frac{v(p)}{\eta(p)} H'(A) L,$$

where the price p is given by (12).

[11] The LSE specification (8) arises for $\rho \to 1 = \eta = a$. The other two cases of the CRESS family correspond to "translated power" surplus functions and imply $\varepsilon(p) = \frac{\alpha\eta p}{(1-\eta)(a-\alpha p)} = \eta\zeta(p)$. One case requires $a < \alpha p$ and generalizes the isoelastic demand specification, which emerges for $a = 0$ and $\eta = \frac{\varepsilon}{\varepsilon - 1}$. The remaining case requires $a > \alpha p$ and generalizes the linear demand specification.

[12] The neutrality of prices with respect to the aggregator is inherited from the indirect additivity of the price aggregator. It does not apply under the preferences adopted by Spence (1976), except for the case of CES preferences.

The concavity of the transformation $H(A)$ implies that the profit decreases with respect to n. Accordingly, in a free entry equilibrium the number n^e of firms satisfies the zero-profit condition:

$$n^e = \frac{H'(A^e)A^e L}{\eta(p)F}, \tag{14}$$

where $A^e = n^e v(p)$, and the individual consumption level is given by:

$$q^e = \frac{\varepsilon(p) - 1}{c}\frac{F}{L}.$$

As usual, n^e depends on the gross profitability of each firm compared to the fixed cost, and therefore increases with market size L, though, as we will discuss next, their relationship depends on the specification of preferences. Instead, the individual consumption is inversely proportional to the market size, so that the total production of each firm $q^e L$ is independent from it.

The monopolistic competition equilibrium characterized above shares some of the features of the classic Krugman (1980) model, based on CES preferences, but it also displays a few novel properties. We briefly explore them before moving to the welfare analysis. The market size L and the transformation embedded into preferences are neutral on prices and firm sizes, but affect the equilibrium number of firms n^e. The latter increases with the size of the market according to:

$$\frac{\partial \ln n^e}{\partial \ln L} = \frac{1}{\rho(A^e)}, \tag{15}$$

which is decreasing in the elasticity of demand with respect to the aggregator. This marks a difference compared to Krugman (1980) and its extension to indirectly additive preferences (Bertoletti and Etro, 2017), where a larger market size exerts a proportional impact on the mass of firms. When $\rho(A^e) < 1$ entry has a relatively small (negative) effect on the individual demand, and the profit opportunities created by a larger market induce a more than proportional entry. Instead, $\rho(A^e) > 1$ provides a substantial crowding out of entry on individual demand, which implies that a less than proportional increase of the mass of firms is enough to dissipate the profit opportunities emerging in a larger market. In case of a logarithmic transformation a larger market exerts a proportional impact on the equilibrium number of firms, as in the Krugman model.

It follows that the expansion of market size can also affect the endogenous expenditure $\Upsilon^e = pn^e q^e$ over the differentiated varieties in a flexible way. In particular, we can compute:

$$\frac{\partial \ln \Upsilon^e}{\partial \ln L} = \frac{1 - \rho(A^e)}{\rho(A^e)} \gtreqqless 0 \quad \text{iff } \rho(A^e) \lesseqqgtr 1. \tag{16}$$

Under the logarithmic transformation individual spending in differentiated varieties is constant and a larger market size exerts the same effects as in Krugman

(1980), where that spending is exogenous. However, when $\rho(A^e) < 1$ a larger market size fosters business creation so much that consumers increase their spending in the differentiated sector and reduce purchases of the outside good. In particular, when the market size doubles, each consumer reduces by half its spending in each single variety, but the mass of consumed goods more than doubles, leading to an increase in total individual spending and amplifying the gains from variety. Instead, when $\rho(A^e) > 1$ a larger market size generates limited equilibrium entry and induces consumers to spend less in differentiated goods and more in the *numéraire*, dampening the gains from variety.

We can also evaluate the impact of changes in the marginal cost (an inverse measure of productivity), which is null on the mass of firms in the Krugman model with CES preferences. Here, we can compute:

$$\frac{\partial \ln n^e}{\partial \ln c} = -\frac{\partial \ln p}{\partial \ln c}\frac{1}{\rho(A^e)}\left[(1-\rho(A^e))\zeta(p) + \frac{\partial \ln \eta(p)}{\partial \ln p}\right],$$

whose sign is in general ambiguous. The CRESS family of surplus functions (10) implies that $\eta(p)$ is constant and thus $\partial \ln n^e/\partial \ln c \lesseqgtr 0$ iff $\rho(A) \lesseqgtr 1$: under the logarithmic transformation a change in productivity is then neutral on entry as in Krugman (1980), otherwise a productivity growth can either induce business creation or business destruction. These comparative statics results are summarized as follows:

Proposition 1. *Monopolistic competition with IIA preferences and homogeneous firms generates prices and production levels that are independent from the market size, but its expansion increases more (less) than proportionally the mass of firms and increases (decreases) overall individual spending in the differentiated goods if* $\rho(A^e) < (>)1$. *Under CRESS preferences an increase of the marginal cost decreases (increases) the mass of firms if* $\rho(A^e) < (>)1$.

Our examples can be used to illustrate these findings. Standard effects of the market size emerge under the logarithmic transformation. In particular, with the LSE specification (8) we can derive the price $p = c + \frac{1}{\alpha}$, and the number of firms and consumption level:

$$n^e = \frac{L}{F} \quad \text{and} \quad q^e = \frac{\alpha F}{L}.$$

Instead, with the *log*-CES specification (9) we obtain the price $p = \frac{\varepsilon c}{\varepsilon - 1}$ and:

$$n^e = \frac{(\varepsilon-1)L}{\varepsilon F} \quad \text{and} \quad q^e = \frac{(\varepsilon-1)F}{cL}.$$

In both cases, the number of firms is proportional to the market size, while changes in the marginal costs are inconsequential on entry.

More flexible results emerge with the isoelastic transformation (11), even when the surplus function belongs to the CRESS family (10). In this case, for $\eta \neq 1$, we can compute the price $p = c\eta + \frac{a}{\alpha}(1-\eta)$, and the equilibrium number of firms and consumption as follows:

$$n^e = v(c)^{\frac{1-\rho}{\rho}} \left(\frac{L}{F}\right)^{\frac{1}{\rho}} \eta^{\frac{\eta-\rho}{\rho(1-\eta)}} \quad \text{and} \quad q^e = \frac{\alpha F}{(1-\eta)(a-\alpha c)L}. \tag{17}$$

Accordingly, the number of firms increases more than proportionally with the market size and it increases with a productivity growth if $\rho < 1$, with opposite results when $\rho > 1$.

The main objective of our analysis of the monopolistic competition equilibrium is to explore its welfare properties. However, it is worth mentioning that the model can be easily applied to a general equilibrium open economy framework with trade frictions in the spirit of Krugman (1980). The production of the outside good under constant returns to scale would fix wages. Introducing iceberg costs of trade, all firms would export, but they would sell products at a higher price in the foreign markets. The endogenous number of firms would depend on trade costs in a complex way. However, focusing on symmetric countries and CRESS preferences, it can be shown that trade liberalization (namely a reduction of the iceberg cost) improves welfare, with an ambiguous impact on entry: the number of firms is not affected under a logarithmic transformation, as in the Krugman model, but otherwise it can either increase or decrease depending on the shape of the transformation function. Therefore, our set-up introduces new features which could be explored in trade applications.

3.1 Welfare: second-best analysis

We now move to evaluate the welfare properties of the equilibrium. Following Spence (1976) and Mankiw and Whinston (1986), we start by assuming that the social planner can control the number of firms n but cannot affect prices, set under monopolistic competition according to (12).[13] Then, Marshallian welfare can be written as:

$$W(n) = H(A(n))L + n[\pi(n) - F] + EL,$$

where $A(n) = nv(p)$ and $\pi(n) = \frac{v(p)}{\eta(p)} H'(A(n))L$. The derivative of welfare with respect to the number of firms is:

$$W'(n) = H'(A(n))v(p)L + [\pi(n) - F] + \frac{v(p)}{\eta(p)} H''(A(n))A(n)L.$$

The three terms in this expression represent the three classic effects of the introduction of an additional firm:

[13] An alternative exercise involves a social planner controlling the price p but not entry. In such a case the planner maximizes the price aggregator A under the free entry constraint $(p-c)|v'(p)|H'(A)L = F$, whose solution is again the equilibrium price rule (12). As already noted by Besanko, Perry and Spady (1990) for a particular Logit model, the equilibrium price is accordingly second best optimal. This is not the case in the Spence model, except once again for CES preferences. We are grateful to Patrick Rey for suggesting this exercise.

1) a positive consumer surplus effect, through a rise of the value of the price aggregator for all consumers, $H'(A(n))v\,(p)\,L$;

2) an impact on producer surplus given by net profits $\pi(n) - F$, which is null in equilibrium;

3) a negative business stealing effect, $n\pi'(n)$, through the reduction of the profits of active firms due to the decrease of demand (following the price aggregator rise).

The previous derivative can be evaluated at the equilibrium $n = n^e$, where the profit effect is null due to free entry (namely, $\pi\,(n^e) = 0$):

$$W'\,(n^e) = H'(A^e)v\,(p)\,L\left[1 - \frac{\rho(A^e)}{\eta\,(p)}\right] \geqslant 0.$$

The non-negative sign derives from assumption (13) which is necessary to satisfy the convexity of preferences. In particular, whenever $\eta\,(p) > \rho\,(A^e)$ market entry is insufficient: in this case the business-stealing effect of additional entry is more than compensated by the direct positive impact on consumer surplus. Only in special cases where $\eta\,(p) = \rho\,(A^e)$ the equilibrium is optimal. This happens in particular when preferences are represented by the isoelastic transformation (11) and the surplus functions belong to the CRESS family (10) with $\eta = \rho$, as in the LSE case that generate a Logit demand system:[14] as we will discuss below, in this case the associated allocation is also efficient in an unconstrained sense. We summarize our main welfare result as follows:

Proposition 2. *Monopolistic competition with IIA preferences and homogeneous firms generates (weakly) insufficient entry of firms.*

Finally, we can provide an implicit expression for the constrained optimal mass of firms solving $W'\,(\widetilde{n}) = 0$ as follows:

$$\widetilde{n} = \frac{H'(\widetilde{A})\widetilde{A}L}{\eta(p)F}\left[1 + \eta(p) - \rho(\widetilde{A})\right],$$

where $\widetilde{A} = \widetilde{n}v(p)$. The comparison to (14) confirms that $n^e \leqslant \widetilde{n}$, and as a consequence in general there are too few firms, each one producing too much compared to the second best allocation (since $H'(\widetilde{A}) \leqslant H'\,(A^e)$).

The strength of the consumer surplus effect relative to the business stealing effect depends on the relative values of $\eta(p) = \varepsilon(p)/\zeta(p)$ and $\rho(A)$. Intuitively, a higher value of the demand elasticity $\varepsilon(p)$ weakens the business stealing effect compared to the consumer surplus effect because it reduces the prices and the profits at which goods are marketed, and a lower value of the surplus elasticity $\zeta(p)$ works in the same direction because entry creates relatively more surplus for consumers than revenues for firms. Crucially, under IIA preferences these

[14] In Appendix B we discuss the utility specifications that deliver an efficient equilibrium. They include the CES case with $v(p) = p^{\frac{1}{1-\eta}}$ and $\eta > 1$ for $H(A) = \frac{A^{1-\eta}}{1-\eta}$, and the linear demand case which arises with $v(p) = (a - \alpha p)^2$ for $H(A) = 2\sqrt{A}$.

effects are independent from entry, which does not affect demand elasticity and the allocation of surplus. Instead, a higher value of $\rho(A)$ strengthens the business stealing effect compared to the consumer surplus effect, because entry reduces more the demand and then the profits of all the firms.

In our setting entry has a direct impact on the aggregator, which is a measure of consumer surplus, and involves a large enough gain from variety and a sufficiently limited business stealing effect to generate a systematic tendency toward insufficient entry. This result is in contrast with previous findings emerging under the quasi-linear preferences introduced by Spence (1976) and used by Mankiw and Whinston (1986), Kuhn and Vives (1999) and others, for which entry could be either insufficient or excessive. While we postpone to Section 5 (and Appendix A) a detailed comparison, we anticipate here that the key issues are the impact of entry on the equilibrium prices and on the revenue/surplus ratio of each variety. As we will see, competitive effects that reduce prices upon entry weaken the business stealing effect and push for insufficient entry. However, entry can be excessive when it increases the revenue/surplus ratio of each variety and the associated bias is not compensated by sufficient competitive effects. In our setting there are no competitive effects and the revenue/surplus ratio is unaffected by entry, which ensures insufficient entry, but in a more general setting this is not necessarily the case. For instance, under the preference specification of Melitz and Ottaviano (2008) entry increases the revenue/surplus ratio of each variety creating the possibility of excess entry for low levels of product differentiation. Nevertheless, in Appendix A we show that the competitive effects can be strong enough to generate insufficient entry in a second-best sense also under the Melitz-Ottaviano specification.

3.2 Welfare: first-best analysis

We conclude this section by examining the first-best allocation that selects both p^* and n^* to maximize welfare. As discussed in Appendix C.1, the first best arises under marginal cost pricing $p^* = c$, so that the social planner problem simplifies to the selection of the number of firms that solves:

$$\max_{n^*} \{H(n^* v(c))L - n^* F\}.$$

The FOC then provides:

$$n^* = \frac{H'(A^*)A^* L}{F}, \tag{18}$$

where $A^* = n^* v(c)$ is the optimal value of the aggregator. Here n^* increases with respect to market size L, while the optimal consumption level:

$$q^* = \frac{\zeta(c)\, F}{cL}$$

is inversely proportional to it, as in equilibrium. Due to the different pricing, it is only in special cases that we can compare these efficient values to the equilibrium ones.[15]

However, when the surplus function belongs to the CRESS family (10) we can provide the following clear-cut result (see Appendix C.1 for the proof):

Proposition 3. *Under CRESS preferences, monopolistic competition with homogeneous firms generates efficient firm sizes, and (weakly) insufficient entry (with full efficiency if* ρ *is constant and equal to* η*).*

The result can be verified through our main example of an isoelastic transformation (11) combined with surplus functions of the CRESS family (10). The equilibrium allocation was derived above. Now we can compute the first best allocation, where the quantity is $q^* = q^e$ and the number of firms is given by:

$$n^* = v\left(c\right)^{\frac{1-\rho}{\rho}} \left(\frac{L}{F}\right)^{\frac{1}{\rho}}.$$

The comparison to the equilibrium value in (17) confirms that the latter is insufficient (namely $n^* > n^e$) for any $\eta > \rho$. Only when $\eta = \rho$ the equilibrium allocation is overall efficient ($n^* = n^e$), as in the LSE case. This efficiency result may be surprising because the equilibrium allocation involves positive markups for the differentiated goods and provides a smaller consumer welfare than the first-best allocation ($A^* > A^e$). The peculiarity of the specifications with $\eta = \rho$ is that the lower consumer surplus in equilibrium exactly matches the profit losses in the first-best allocations, so that total welfare is the same. In other words, the monopolistic competition prices implement at the equilibrium the efficient quantities and create the efficient number of firms.[16]

Comparisons between monopolistic competition equilibria and first-best allocations have been analyzed in a variety of works with homogeneous firms. The results are typically ambiguous due to the different pricing, both in the presence of an outside good and without (Spence, 1976; Dixit and Stiglitz, 1977; Bertoletti and Etro, 2016). In our environment, the ambiguity disappears at least for the CRESS preferences, where entry tends to be insufficient also compared to the efficient allocation.

4 Monopolistic competition with heterogeneous firms

We now move to study monopolistic competition when firms are *ex-post* heterogeneous. As in Melitz (2003) we assume that each firm draws its marginal cost c from a continuous distribution $G(c)$ with support $[0, \overline{c}]$, and is active after paying a fixed cost of production f, setting its price to maximize profits (taking as given the aggregator).

[15] For example, with a logarithmic transformation we immediately get $n^* \geqslant n^e$ for any surplus function, since condition (13) requires $\eta\left(p\right) \geqslant 1$. It follows, trivially, that $A^* > A^e$.

[16] Strictly speaking, in these very special cases first-best prices are multiple, and they need not to be equal to the marginal costs.

An active firm with marginal cost c sets a price according to the pricing rule:

$$p(c) \equiv \frac{\varepsilon(p(c))c}{\varepsilon(p(c)) - 1}, \tag{19}$$

which now depends on the marginal cost and is still independent from the mass of firms and the market size. More efficient firms set lower prices, and also get higher (lower) markups if the demand elasticity $\varepsilon(p)$ is increasing (decreasing) in the price.

The quantity demanded by each consumer of a good produced at marginal cost c is thus given by:

$$q(c) = |v'(p(c))|\, H'(A),$$

which is a function not only of the price $p(c)$, but also of the price aggregator A, that is affected by the measure and selection of firms. The equilibrium profits of an active firm with marginal cost c are then given by:

$$\pi(c) = \frac{v(p(c))}{\eta(p(c))} H'(A)L - f, \tag{20}$$

and are decreasing in the marginal cost by the Envelope theorem, which implies $\pi'(c) = -q\,(c)\,L < 0$.

Accordingly, there is a cutoff firm with marginal cost $\hat{c}$ such that $\pi(\hat{c}) = 0$, assuming that $\hat{c} < \bar{c}$. Given this, the value of the aggregator can be written as:

$$A = N \int_0^{\hat{c}} v(p(c))dG(c), \tag{21}$$

where N is the mass of created firms and $G(c)$ is the *ex-ante* distribution of marginal costs. Then, the measure of active firms and consumed products is given by $n = G(\hat{c})N$. The free entry condition requires that the expected profit is equal to the entry cost, namely $\int_0^{\hat{c}} \pi(c)dG(c) = F$, which implicitly pins down the measure of firms.

This allows us to express the zero-profit and free entry conditions as follows:

$$\frac{v(p(\hat{c}^e))}{\eta(p(\hat{c}^e))} H'(A^e)L = f, \tag{22}$$

$$\left[\int_0^{\hat{c}^e} \frac{v\,(p(c))}{\eta\,(p(c))} dG(c)\right] H'(A^e)L = F + G(\hat{c}^e)f, \tag{23}$$

where $A^e = N^e \int_0^{\hat{c}^e} v(p(c))dG(c)$. The equilibrium system (22)-(23) determines $(\hat{c}^e, N^e)$. Equivalently, it determines $(\hat{c}^e, A^e)$, and therefore it also pins down consumer welfare through the equilibrium value of the aggregator.[17] Assuming $f > 0$ and combining the two equations we obtain the condition:

$$\int_0^{\hat{c}^e} \frac{v(p(c))}{\eta(p(c))} \frac{\eta\,(p(\hat{c}^e))}{v(p(\hat{c}^e))} dG(c) = \frac{F}{f} + G(\hat{c}^e),$$

[17] In the limit case where f vanishes but demand exhibits a finite choke price, the latter determines the cutoff through the condition $v'(\hat{c}^e) = 0$.

which alone identifies the equilibrium cutoff value $\hat{c}^e$. The left hand side is a marginal rate of substitution between N and $\hat{c}$ in terms of gross profitability and the right hand side displays the marginal rate of transformation in terms of fixed costs. The fact that this condition does not depend on market size implies a dichotomy between selection and entry that also holds in the Melitz (2003) model.[18]

To express the equilibrium conditions in a convenient way for our welfare comparisons, it is useful to define the weighted harmonic mean value of $\eta(p(c))$ across active firms as:

$$\overline{\eta}(\hat{c}) = \left[\int_0^{\hat{c}} \frac{1}{\eta(p(c))} \frac{v(p(c))}{\int_0^{\hat{c}} v(p(c))dG(c)} dG(c)\right]^{-1}. \tag{24}$$

Remembering condition (13), in this environment the convexity of preferences requires:

$$\overline{\eta}\left(\hat{c}\right) \geq \rho\left(A\right), \tag{25}$$

which will be crucial for our welfare analysis. After some manipulations, we can express the system (22)-(23) as:

$$\int_0^{\hat{c}^e} \frac{v(p(c))}{v(p(\hat{c}^e))} dG(c) = \left[\frac{F}{f} + G(\hat{c}^e)\right] \frac{\overline{\eta}(\hat{c}^e)}{\eta\left(p(\hat{c}^e)\right)} \tag{26}$$

and

$$N^e = \frac{H'(A^e)A^e L}{\overline{\eta}(\hat{c}^e)[F + G(\hat{c}^e)f]}. \tag{27}$$

Condition (26) determines the equilibrium cutoff $\hat{c}^e$. Now the left hand side displays the ratio of average and marginal surplus at the equilibrium prices, which is a marginal rate of substitution between N and $\hat{c}$ in terms of consumer welfare, and is increasing in the cutoff. The right hand side displays the corresponding marginal rate of transformation, corrected for the ratio between mean and marginal values of the relative elasticity of substitution and surplus, which accounts for gross profitability. The equilibrium value of the cutoff is an increasing function of the fixed cost of entry, as well as of the mean value of the relative elasticity. Condition (27) determines residually the mass of firms created as a function of the equilibrium cutoff $\hat{c}^e$. When the mean $\overline{\eta}(\hat{c}^e)$ is higher (profitability is lower on average), the market tends to create fewer firms. The same happens when the fixed costs increase.

An expansion of the market size L is neutral on the price rules (19) and does not affect selection, since condition (26) is independent from it. However, it affects the measure of entrant firms and therefore the mass of consumed varieties. It is straightforward to verify that:

$$\frac{\partial \ln N^e}{\partial \ln L} = \frac{1}{\rho\left(A^e\right)},$$

[18] The neutrality of the cutoff with respect to market size is inherited from the indirect additivity of the price aggregator (see Bertoletti, Etro and Simonovska, 2018), and in general it does not hold under the preferences adopted by Spence (1976) or Dixit and Stiglitz (1977).

analogously to what happens under homogeneous firms. Similarly, the impact on the overall expenditure in the differentiated varieties, $\Upsilon^e = N^e \int_0^{\hat{c}^e} p(c)q(c)dG(c)$, is the same as under homogeneous firms, namely given by (16). Thus, an increase in the market size expands the mass of consumed varieties more or less than proportionally depending on whether $\rho(A^e)$ is smaller or larger than unity in equilibrium, and spending in differentiated varieties can either amplify or dampen the gains from variety in larger markets. The case of a logarithmic transformation, instead, preserves the properties of the Melitz (2003) model with respect to this dimension.

Finally, the equilibrium consumption of a variety produced with marginal cost c is given by:

$$q^e(c) = |v'(p(c))| \frac{\eta(p(\hat{c}^e))f}{v(p(\hat{c}^e))L}, \tag{28}$$

which is inversely proportional to the market size, while the equilibrium firm size does not depend on it. Also the transformation embedded into preferences is neutral on quantities and selection, and only affects the measures of entrant and active firms.

The main findings concerning the comparative statics of our setting are summarized as follows:

Proposition 4. *Monopolistic competition with IIA preferences and heterogeneous firms generates prices, production levels and selection of firms that are independent from the market size, but its expansion increases more (less) than proportionally the mass of firms and increases (decreases) overall individual spending in the differentiated goods if* $\rho(A^e) < (>)1$.

We can illustrate these results through our usual examples.[19] The LSE specification (8) provides the following equilibrium measure of entrant firms and consumption level:

$$N^e = \frac{L}{F + G(\hat{c}^e)f} \quad \text{and} \quad q^e(c) = \frac{\alpha f}{L} e^{\alpha(\hat{c}^e - c)}.$$

The *log*-CES specification (9) provides:

$$N^e = \frac{(\varepsilon - 1)L}{\varepsilon[F + G(\hat{c}^e)f]} \quad \text{and} \quad q^e(c) = (\varepsilon - 1)\frac{f}{L}\frac{(\hat{c}^e)^{\varepsilon-1}}{c^\varepsilon}.$$

These are cases where an expansion of market size has a proportional impact on the measure of differentiated goods without affecting overall spending Υ^e.

In case of the isoelastic transformation (11) with surplus functions that belong to the CRESS family (10) for $\eta \neq 1$, we can recover the following mass of firms and consumption levels:

$$N^e = v(\hat{c}^e)^{\frac{1-\rho}{\rho}} \left(\frac{L}{f}\right)^{\frac{1}{\rho}} \frac{\eta^{\frac{\eta-\rho}{\rho(1-\eta)}} f}{F + G(\hat{c}^e)f} \quad \text{and} \quad q^e(c) = \frac{\alpha}{|\eta - 1|}\frac{f}{L}\frac{v(c)^\eta}{v(\hat{c}^e)}. \tag{29}$$

[19] In all of these examples the cutoff satisfies $\int_0^{\hat{c}^e} \frac{v(c)}{v(\hat{c}^e)} dG(c) = \frac{F}{f} + G(\hat{c}^e)$, independently from the market size.

This immediately shows that the impact of market size on the measure of consumed goods can be either more or less than proportional.

As in the case of homogenous firms, our model can be extended to a general equilibrium open economy setting to study costly trade in the spirit of Melitz (2003) and Melitz and Ottaviano (2008).[20] In presence of iceberg costs of trade only the most efficient firms would be able to cover the fixed costs of export and thus engage in international trade. The price aggregator would then reflect the surplus from both domestic and imported goods, and free entry would determine the mass of firms created in each country. The cutoff costs for domestic activation and for export would be independent from their size, extending the dichotomy between selection and entry that holds in a closed economy. Under CRESS preferences it is also straightforward to verify that a trade liberalization would induce selection effects and improve welfare, while its impact on entry would remain ambiguous.

In the remaining of this section we examine the welfare properties of the monopolistic equilibrium with heterogeneous firms, to verify whether also in this case the market allocation involves insufficient entry and what are the additional implications for the selection of firms. We start by exploring what happens when the social planner controls the measure of firms, but cannot affect pricing and the decision of becoming active. In principle, this could be implemented by subsidizing or taxing entry (affecting the fixed cost of entry). We then move to the case where the social planner can also control selection, potentially through subsidies or taxes on actual production (or affecting the fixed cost of production). Finally, we consider the case where the social planner controls entry, activation and pricing of all firms. It is natural to refer to these cases respectively as third-, second- and first-best analysis.

4.1 Welfare: third-best analysis

Suppose that the social planner chooses the mass of firms N to maximize the sum of consumer surplus and expected profits of the firms under the constraint that firms decide on pricing and activation. This amounts to the problem:

$$\max_N W(N) = H(A)\,L + N\left[\int_0^{\hat{c}} \pi(c)dG(c) - F\right] + EL,$$

where $\pi(c)$ is given by (20), for a pricing rule (19) under constraints (21) and (22), which determine the indirect impact of N through A and $\hat{c}$. The planner takes into account the equilibrium reactions of the value of the aggregator and of the cutoff for activation to changes in the mass of firms (remember that the pricing rules are independent from the latter in our setting).

Differentiating (21) and (22) we get, in matrix form:

$$\mathbf{D}\begin{bmatrix} \frac{dA}{dN} \\ \frac{d\hat{c}}{dN} \end{bmatrix} = \begin{bmatrix} \frac{A}{N} \\ 0 \end{bmatrix}, \text{ with } \mathbf{D} = \begin{bmatrix} 1 & -Nv\left(p\left(\hat{c}\right)\right)g\left(\hat{c}\right) \\ -\rho\left(A\right)\frac{f}{A} & \pi'\left(\hat{c}\right) \end{bmatrix},$$

[20] For an analysis of new trade models with demand systems depending on a single price aggregator see Alfaro (2022).

where the determinant of the matrix $\mathbf{D}$ is given by:

$$\det\{\mathbf{D}\} = \pi'(\widehat{c}) - \frac{\rho(A)}{A} N v(p(\widehat{c})) g(\widehat{c}) f < 0.$$

Thus, by Cramer's rule we obtain the following effects of changes in the measure of firms respectively on the value of the aggregator and the cutoff:

$$\frac{dA}{dN} = \frac{A}{N}\frac{\pi'(\widehat{c})}{\det\{\mathbf{D}\}} > 0 \quad \text{and} \quad \frac{d\widehat{c}}{dN} = \frac{A}{N}\frac{\rho(A) f}{\det\{\mathbf{D}\}} < 0.$$

The overall impact of N on the aggregator depends on $d\hat{c}/dN$ and thus on the transformation function and on the cost distribution, but its sign is necessarily positive: as one would expect, the creation of more firms enhances consumer welfare. A rise of N decreases $\hat{c}$ by reducing demand and then necessarily the profit of each active firm, which creates a selection effect. The size of this effect depends on the elasticity of demand with respect to the aggregator, given by $\rho(A)$: the larger this elasticity the larger the selection effect induced by the creation of more firms.

The welfare impact of entry is thus given by:[21]

$$W'(N) = H'(A) L \frac{dA}{dN} + \left[\int_0^{\hat{c}} \pi(c) dG(c) - F\right] - \frac{\rho(A)}{\overline{\eta}(\hat{c})} H'(A) L \frac{dA}{dN}, \tag{30}$$

which exhibits the three welfare effects familiar from the case of homogeneous firms:

1) a positive consumer surplus effect, through a rise of the value of the price aggregator;

2) an impact on producer surplus given by the expected profits;

3) a negative business stealing effect, through the reduction of the profits of active firms.

Evaluating $W'(N)$ at the equilibrium value N^e, such that the expected profit is null, we obtain:

$$W'(N^e) = H'(A^e)\left[1 - \frac{\rho(A^e)}{\overline{\eta}(\hat{c}^e)}\right]\frac{dA}{dN} L \gtrless 0, \tag{31}$$

whose sign relies on condition (25).[22] In particular, if $\overline{\eta}(\hat{c}^e) > \rho(A^e)$ the market equilibrium is creating an insufficient mass of entrant firms, and a constrained social planner would like to promote entry through subsidies. Only in the special case where $\overline{\eta}(\hat{c}^e) = \rho(A^e)$ the equilibrium is constrained optimal. In particular, this happens in the LSE case generating a Logit demand system and, more

[21] The welfare direct impact of N through $\hat{c}$ is null by condition (22).

[22] The third-best mass of firms satisfies $W'(N^c) = 0$ and it is implicitly given by:

$$N^c = \frac{H'(A^c) A^c L}{\overline{\eta}(\hat{c}^c)[F + G(\widehat{c}^c) f]}\left\{1 + [\overline{\eta}(\hat{c}^c) - \rho(A^c)]\frac{\pi'(\widehat{c}^c)}{\det\{\mathbf{D}\}}\right\},$$

where $\pi(c)$ is given by (20), and A^c and $\widehat{c}^c$ are defined by (21) and (22).

generally, when preferences are represented by the isoelastic transformation in (11) and the surplus functions belong to the CRESS family (10) with $\eta = \rho$.

In summary, these results confirm that in our setting the business-stealing effect cannot dominate the consumer surplus effect, even under heterogeneous firms and endogenous market selection:

Proposition 5. *Monopolistic competition with IIA preferences and heterogeneous firms generates (weakly) insufficient entry of firms.*

This result implies that a social planner would like to expand the mass of firms at the equilibrium, even at the cost of reducing their sizes and the fraction of active firms according to the selection provided by the market. Next, we will verify what happens when the social planner can avoid this constraint and can control firm activation.

4.2 Welfare: second-best analysis

We now consider what happens when the social planner cannot affect equilibrium pricing, given by (19), but can control both the measure of entrant firms and also their activation, possibly through appropriate taxes or subsidies. In this case the planner chooses $(\widehat{c}, N)$, and therefore A, to solve the second-best problem:

$$\max_{\widehat{c},N} \widetilde{W}(\widehat{c}, N) = H(A)L + N\left[\int_0^{\hat{c}} \pi(c)dG(c) - F\right] + EL,$$

where $\pi(c)$ is given by (20), under the pricing rule (19) and the constraint (21). Differentiating the latter we get the following effects of the decision variables on the aggregator:

$$\frac{\partial A}{\partial \widehat{c}} = Nv\left(p\left(\widehat{c}\right)\right)g\left(\widehat{c}\right) > 0 \quad \text{and} \quad \frac{\partial A}{\partial N} = \frac{A}{N} > 0. \tag{32}$$

The welfare derivatives are then given by:

$$\frac{\partial \widetilde{W}}{\partial \widehat{c}} = H'(A)L\frac{\partial A}{\partial \widehat{c}} + N\pi(\widehat{c})g(\widehat{c}) + N\int_0^{\hat{c}} \frac{\partial \pi(c)}{\partial A}dG(c)\frac{\partial A}{\partial \widehat{c}}, \tag{33}$$

$$\frac{\partial \widetilde{W}}{\partial N} = H'(A)L\frac{\partial A}{\partial N} + \int_0^{\hat{c}} \pi(c)dG(c) - F + N\int_0^{\hat{c}} \frac{\partial \pi(c)}{\partial A}dG(c)\frac{\partial A}{\partial N} \tag{34}$$

Conditions (33)-(34) show that a rise of the cutoff and the creation of new firms have once again three welfare effects. Beyond the consumer surplus and business stealing effects due to induced changes of the aggregator, there are the direct impacts on producer surplus of the cutoff and of the mass of entrant firms, which are given respectively by the profit $\pi(\widehat{c})$ of the marginal firms multiplied by their measure $Ng(\widehat{c})$, and by the expected profit, both of which are null at the equilibrium.

Accordingly, the second-best inefficiency of the equilibrium depends on the comparison of the consumer surplus effect and the business stealing effect. For a given value of A and of either $dA/d\widehat{c}$ or dA/dN, the latter is captured by:

$$N\int_0^{\hat{c}}\frac{\partial\pi(c)}{\partial A}dG(c)=-\frac{\rho\left(A\right)}{\overline{\eta}\left(\hat{c}\right)}H'\left(A\right)L. \tag{35}$$

Thus, the indirect impact on profits of a greater $\widehat{c}$ or N is larger (in absolute value) the larger is $\rho\left(A\right)$ and the smaller is $\overline{\eta}\left(\hat{c}\right)$. As a result, the net welfare impact of $\widehat{c}$ and N through A, that is the net impact of consumer surplus and business stealing effects, depends once again on the sign of $\left[1-\rho\left(A\right)/\overline{\eta}\left(\hat{c}\right)\right]H'\left(A\right)L$, and therefore it is non-negative due to condition (25).

Thus, when evaluated at a market equilibrium both $\partial\widetilde{W}/\partial\widehat{c}$ and $\partial\widetilde{W}/\partial N$ are non-negative. In particular, if $\overline{\eta}\left(\hat{c}^e\right)>\rho\left(A^e\right)$ the planner can achieve a *local* welfare improvement by subsidizing both entry and activation. We summarize our last finding as follows:

Proposition 6. *At a monopolistic competition equilibrium with IIA preferences and heterogeneous firms welfare can be (weakly) increased by expanding locally both the mass of firms and the fraction of active firms.*

This confirms and extends our previous welfare results: the social planner has a local incentive to promote both entry and activation of firms (thereby reducing the sizes of active firms). However, this does not mean that at the second-best allocation the social planner necessarily expands the mass of active firms using both tools. For instance, it may be optimal to create more firms than in the monopolistic competition equilibrium, but ultimately activate a smaller fraction of them. To verify this, we now characterize the second-best optimum.

Setting (33)-(34) equal to zero we obtain second-best values $\widetilde{N}$ and $\widetilde{c}$ that satisfy the system:

$$\int_0^{\widetilde{c}}\frac{v(p(c))}{v\left(p\left(\widetilde{c}\right)\right)}dG(c)=\left[\frac{F}{f}+G\left(\widetilde{c}\right)\right]\frac{\eta\left(p\left(\widetilde{c}\right)\right)\left[\overline{\eta}\left(\widetilde{c}\right)-\rho(\widetilde{A})\right]+\overline{\eta}\left(\widetilde{c}\right)}{\eta\left(p\left(\widetilde{c}\right)\right)\left[\overline{\eta}\left(\widetilde{c}\right)-\rho(\widetilde{A})+1\right]}, \tag{36}$$

$$\widetilde{N}=\frac{H'(\widetilde{A})\widetilde{A}L}{\overline{\eta}\left(\widetilde{c}\right)\left[F+G(\widetilde{c})f\right]}\left[1+\overline{\eta}\left(\widetilde{c}\right)-\rho(\widetilde{A})\right]. \tag{37}$$

for $f>0$, with a lower production for the active firms compared to the equilibrium (see Appendix C.2). These expressions differ from the equilibrium analogues (26) and (27) for additional terms on the right hand side, depending on the mean and marginal values of the relative elasticity of demand and surplus and on the elasticity of demand to the aggregator. In general, the market size L affects both $\widetilde{N}$ and $\widetilde{c}$, so this allocation loses the dichotomy that we found in a market equilibrium where the cutoff was independent from the market size and from the transformation embedded into preferences.

Nevertheless, condition (36) can be compared with the corresponding equilibrium condition (26). The left hand sides present the same ratio of expected and marginal surplus functions, which are increasing in the cutoff. The right hand sides differ by the last term. However, it can be easily verified that:

$$\frac{\overline{\eta}\left(\widetilde{c}\right)}{\eta\left(p\left(\widetilde{c}\right)\right)} < \frac{\eta\left(p\left(\widetilde{c}\right)\right)\left[\overline{\eta}\left(\widetilde{c}\right) - \rho(\widetilde{A})\right] + \overline{\eta}\left(\widetilde{c}\right)}{\eta\left(p\left(\widetilde{c}\right)\right)\left[\overline{\eta}\left(\widetilde{c}\right) - \rho(\widetilde{A}) + 1\right]}$$

corresponds to $\overline{\eta}\left(\widetilde{c}\right) < \eta\left(p\left(\widetilde{c}\right)\right)$, which holds when $\eta\left(p\right)$ is monotonic increasing. It follows that:

$$\widehat{c}^e \lesseqqgtr \widetilde{c} \text{ if (everywhere) } \eta'\left(p\right) \gtreqqless 0.$$

Accordingly, in the case of the CRESS family (10), where $\eta\left(p\right)$ is constant, the market provides the second-best selection of firms. Moreover, in that case the comparison of (37) with (27) implies that $\widetilde{N} \geqslant N^e$ always, confirming the insufficient entry result.[23] Instead, when $\eta\left(p\right)$ is not constant, we cannot draw general conclusions. We summarize these findings as follows:

Proposition 7. *Monopolistic competition with IIA preferences and heterogeneous firms generates excessive (insufficient) selection if $\eta\left(p\right)$ is monotonic increasing (decreasing). Under CRESS preferences, monopolistic competition delivers optimal selection, with insufficient entry whenever $\eta > \rho(A^e)$ and optimal entry if $\eta = \rho(A^e)$.*

This result is reminiscent of findings by Dhingra and Morrow (2019) in an environment with a given amount of resources allocated to differentiated goods, where the equilibrium can generate either excessive or insufficient selection and production of the active firms. However, in our setting with an endogenous resource allocation the equilibrium is associated to a (weakly) excessive production of the active firms.

4.3 Welfare: first-best analysis

In the first best the social planner maximizes Marshallian welfare by choosing the measure N^* of entrants, the threshold for active firms $\hat{c}^*$ and the price schedule $p^*(c)$ under a non-binding resource constraint. The solution (see Appendix C.3) can be obtained by marginal cost pricing, namely by $p^*(c) = c$. Accordingly, the planner problem can be written as:

$$\max_{N^*,\hat{c}^*} H\left(N^*\int_0^{\hat{c}^*} v(c)dG(c)\right)L - N^*\left[F + G(\hat{c}^*)f\right],$$

Assuming $f > 0$, the FOCs for this problem provide a cutoff and a mass of firms satisfying:

$$\int_0^{\hat{c}^*}\frac{v(c)}{v(\hat{c}^*)}dG(c) = \frac{F}{f} + G(\hat{c}^*), \tag{38}$$

[23] Note that insufficient equilibrium entry in a constrained sense arises also in the limit case of $f = 0$ whenever the cutoff is fixed by an exogenous choke price.

$$N^* = \frac{H'(A^*)A^*L}{F + G(\hat{c}^*)f}, \tag{39}$$

where $A^* = N^* \int_0^{\hat{c}^*} v(c)dG(c)$. Condition (38) defines the efficient cutoff by equating the marginal rate of substitution between N^* and $\hat{c}^*$ in terms of consumer surplus on the left hand side to the corresponding marginal rate of transformation in terms of fixed costs on the right hand side. Condition (39) determines residually the efficient mass of firms. The efficient consumption level can be derived for each product as follows:

$$q^*(c) = \frac{|v'(c)|\, f}{v(\hat{c}^*)L}. \tag{40}$$

A comparison to the equilibrium system (26)-(27)-(28) provides only partial insights, given the different pricing. The equilibrium production of the active firms and their selection could be biased in different directions. Nevertheless, both firms' size and selection are efficient when the surplus function belongs to the CRESS family. The reason is that in such a case the "virtual social surplus" $v(p(c))/\eta(p(c))$, which determines the market cutoff, is proportional to the actual social surplus $v(c)$ considered by the planner.[24]

Moreover, the same argument used with homogeneous goods implies that under CRESS preferences entry is weakly insufficient. For instance, in our example with an isoelastic transformation (11) the first-best measure of entrant firms is given by:

$$N^* = v(\hat{c})^{\frac{1-\rho}{\rho}} \left(\frac{L}{f}\right)^{\frac{1}{\rho}} \frac{f}{F + G(\hat{c})f},$$

whose comparison to the equilibrium value in (29) implies immediately $N^* \gtrless N^e$, with efficiency for $\eta = \rho$, as in the LSE case. We summarize these findings as follows:

Proposition 8. *Under CRESS preferences, monopolistic competition with heterogeneous firms generates efficient selection and firm sizes, and (weakly) insufficient entry (with full efficiency if ρ is constant and equal to η).*

A variety of works have made comparisons between monopolistic competition equilibria and first-best allocations under alternative settings. In the absence of an outside good (i.e. with exogenous resources allocated to the differentiated goods), full efficiency with heterogeneous firms emerges under CES preferences (Dhingra and Morrow, 2019) and also for the wider class of so-called *implicit* CES preferences (Bertoletti and Etro, 2021): this is essentially a consequence of the common markup that they imply on all goods. With other separable preferences (Dhingra and Morrow, 2019; Bertoletti and Etro, 2021) both entry and selection can be either excessive or insufficient. Further results can be

[24] It can be proved (see Appendix C.3) that $\hat{c}^* \lesseqgtr \hat{c}^e$ if everywhere $\Phi'(c) \gtrless 0$, where $\Phi(c) \equiv [v(p(c))/\eta(p(c))]\,/v(c)$. In these cases $v(p(c))/\eta(p(c))$ diverges in a predictable way from $v(c)$ and, of course, $\Phi(c)$ is a constant under CRESS preferences. These results resonate well with those obtained by Dhingra and Morrow (2019) in a different setting.

derived in special cases: for instance, particular preference specifications combined with a Pareto distribution and without fixed costs have been shown to generate efficient entry and insufficient selection: this is the case of the translated power preferences of Bertoletti, Etro and Simonovska (2018) and of the quadratic preferences of Bagwell and Lee (2023).[25]

In the presence of an outside good, we are only aware of the welfare analysis by Nocco, Ottaviano and Salto (2014), based on the quasi-linear quadratic preferences of Melitz and Ottaviano (2008), again with a Pareto distribution and without fixed costs. This framework delivers insufficient selection in equilibrium compared to the first best, and ambiguous results on entry, depending on the substitutability among varieties. Not surprisingly, more general preference and cost conditions are consistent with an even wider range of outcomes, but in our setting CRESS preferences deliver efficient selection and (weakly) insufficient entry also compared to first-best allocations.

5 General preferences and competitive effects

Our results on entry have been derived for demand systems depending on a price aggregator that represents a demand shifter. In this baseline setting there are no competitive effects of entry, since under monopolistic competition entry does not affect demand elasticity and thus markups. Alternative demand systems, as those popularized by Spence (1976) and Melitz and Ottaviano (2008), can generate competitive effects and also, in particular cases, instances of excess entry *à la* Mankiw and Whinston (1986). This leads us to consider more general demand systems and clarify the role played by the competitive effects and the conditions under which insufficient entry emerges.

With this goal in mind, we introduce a quasi-linear version of the Gorman-Pollak (GP) preferences,[26] which deliver a large class of demand systems depending on a price aggregator in a flexible way. They can be represented, under suitable technical conditions, by the following indirect utility:

$$V = \int_{\omega \in \Omega} s(p(\omega), M) d\omega - \theta(M) + E, \tag{41}$$

where both the incremental surplus produced by a variety, $s(p(\omega), M)$, and the function $\theta(M)$ depend on a price aggregator M, which is implicitly defined by:

$$\int_{\omega \in \Omega} s_M(p(\omega), M) d\omega \equiv \theta'(M), \tag{42}$$

with subscripts denoting partial derivatives.

[25] Bagwell and Lee (2023) offer also an exhaustive discussion of why marginal cost pricing is part of a global optimum in these kinds of problems.

[26] The original version of GP preferences, which does not consider an outside good, has been recently explored by Fally (2019, 2022) and Bertoletti and Etro (2021) under monopolistic competition with exogenous resources allocated to the differentiated goods. Applications can be found in Lashkaripour (2020), Macedoni and Weinberger (2022), and elsewhere.

In Appendix A we derive the direct utility corresponding to (41) and characterize the properties of the aggregator M, which is not necessarily additive across varieties and it is conveniently defined here as increasing with prices and decreasing with entry.[27] By its definition, the aggregator M is such that its impact on the indirect utility is null, therefore by Roy's identity demand $q(p(\omega), M) = |s_p(p(\omega), M)|$ depends on this single price aggregator in a flexible way (i.e. not in a multiplicative way as a pure demand shifter). As shown in Appendix A, different assumptions on the functional form for the incremental surplus allow us to nest both the IIA preferences (1) and those of Spence (1976),[28] as well as other preferences generating demand systems with a single aggregator, as the Melitz and Ottaviano (2008) specification.

To analyze the case for insufficient entry *à la* Mankiw and Whinston (1986), we extend our analysis with homogenous firms to GP preferences. Here all firms set a common price p that can change with the number of firms, with the possibility of either competitive or anti-competitive effects. Both the equilibrium values of the aggregator, implicitly defined by $ns_M(p, M) = \theta'(M)$, and of the individual quantity $q(p, M)$ depend on the number of firms. We are interested in the impact of entry on welfare, which is given by the sum of aggregate consumer surplus and total profits:

$$W(n) = [ns(p, M) - \theta(M)]\, L + n\, [(p - c)q(p, M)L - F] + EL.$$

Using the Roy's identity and the neutrality of the aggregator on consumer welfare, the impact of entry at the free entry equilibrium can be written as follows:

$$W'(n^e) = s(p, M)L + n^e(p - c)\frac{\partial q(p, M)}{\partial M}\frac{\partial M}{\partial n}L + n^e(p - c)\frac{dq(p, M)}{dp}\frac{dp}{dn}L.$$

The first two terms are the same encountered in our baseline model: a positive consumer surplus effect provided by the incremental surplus of a new variety and a negative producer surplus effect which depends on the impact of entry on demand through the aggregator. The third term is novel, and captures the impact of entry on demand and profits through prices. The crucial point is that the sign of the new term is constrained by the convexity of preferences, which requires that the aggregate demand of differentiated products is non-increasing in their common price (namely, $\frac{dq(p,M)}{dp} \leqslant 0$). As a consequence, whenever entry exerts a competitive effect that reduces prices (i.e., whenever $\frac{dp}{dn} < 0$), which is often regarded as the plausible scenario (Krugman, 1979), this competitive effect works to reduce the size of the overall business stealing effect, strengthening the case for insufficient entry. The simple intuition is that the downward adjustment of the incumbents' prices which follows entry does limit their loss of sales. Remarkably, this is a general insight: business stealing

[27] Notice that the aggregator A of the baseline model was defined as decreasing with prices and increasing with entry.

[28] In particular, the IIA preferences emerge when $s(p, M) = Mv(p)$ and the Spence preferences when $s(p, M) = Mv(p/M)$.

is weakened by any competitive effects for the same reason under more general quasi-linear preferences.[29]

We conclude that our results of insufficient entry under IIA preferences are not driven by the absence of competitive effects, and can be actually amplified by those arising under other preferences. Nevertheless, instances of excess entry could emerge when the consumer surplus effect is relatively weak (for example when products are highly substitutable) or there is a sufficiently strong anti-competitive effect (namely, when entry increases prices). In Appendix A.1 we show that insufficient entry emerges under GP preferences if and only if, at the market equilibrium:

$$\frac{\varepsilon_p}{\zeta_p} - \rho + \left(\varepsilon_p - \frac{\varepsilon_M}{\zeta_M}\zeta_p\rho\right)\sigma > 0, \tag{43}$$

where ζ_p and ε_p are the elasticities of surplus and demand with respect to the price, ζ_M and ε_M those with respect to the aggregator, ρ captures the elasticity of demand with respect to entry through the aggregator, and $\sigma = -\frac{d\ln p}{d\ln n}$ measures the strength of the competitive effects. To satisfy the convexity of preferences the term in parenthesis in (43) cannot be negative, therefore the presence of competitive effects makes insufficient entry more likely.

We can also isolate a sufficient condition for the emergence of insufficient entry under GP preferences: assuming that there are no anti-competitive effects (namely, $\sigma \geqslant 0$), entry is insufficient whenever it increases (or leaves unchanged) the ratio between surplus and revenues (namely, when ζ_p is non-decreasing in the aggregator). This condition is always satisfied under IIA preferences, where this ratio does not change with entry. And it can be satisfied under the Spence preferences (for which we also identify a weaker condition). However, it is a sufficient but not necessary condition: for instance, it does not hold for the Melitz-Ottaviano preference specification, but we show that in this case the competitive effect can be so strong that entry can still be insufficient when combined with a high level of product differentiation.

In Appendix A.2 we finally extend the analysis to the case of heterogeneous firms, focusing on whether entry is excessive or insufficient at the equilibrium whenever firms decide on pricing and activation. In spite of substantial complications, the insights are similar to the previous ones, and in particular the competitive effect works to reduce the business stealing component of welfare changes. In addition, we show that any selection effect which reduces firm activation tends to decrease both the consumer surplus effect (by eliminating the surplus contribution of the cutoff firms) and the business stealing effect (by increasing the aggregator). We conclude this section by illustrating these results here through an example based on the model of Melitz and Ottaviano (2008) without fixed costs of production and with a Pareto distribution of marginal costs.

[29] Notice that a case of demand systems depending on multiple aggregators emerges when M is a vector of price aggregators defined by a system of conditions analogous to (42).

The Melitz-Ottaviano preference specification can be represented by the following indirect utility:

$$V = N \int_0^{\hat{c}} \frac{(M - p(c))^2}{2\gamma} dG(c) - \theta(M) + E \quad \text{with } \theta(M) = \frac{\alpha M}{\eta} - \frac{M^2}{2\eta}, \quad (44)$$

where the parameters $\alpha, \gamma, \eta > 0$ are the same as in Melitz and Ottaviano (2008) and by (42) the price aggregator is implicitly defined by:

$$N \int_0^{\widehat{c}} \frac{M - p(c)}{\gamma} dG(c) = \theta'(M).$$

As in our baseline framework, $p(c)$ is the price set by an active firm with marginal cost c, the cutoff $\hat{c}$ represents the threshold for activation and N is the measure of entrant firms.

The quadratic surplus function $s(p, M) = \frac{(M-p)^2}{2\gamma}$ implies the linear demand $q(p, M) = \frac{M-p}{\gamma}$, where γ parametrizes the level of product differentiation. The active firms use the familiar price rule $p = \frac{c+M}{2}$. Under the assumption that there are no fixed costs of production ($f = 0$), all firms with a positive demand at a price equal to their marginal cost are active, so that $\hat{c} = M$. The definition of the aggregator then provides an inverse relation between the measure of entrants and the cutoff. Under a Pareto distribution of marginal cost $G(c) = (c/\overline{c})^{\kappa}$ with $\kappa \geqslant 1$, this relation be expressed as:[30]

$$N = \frac{2\overline{c}^{\kappa} (\kappa + 1) \gamma(\alpha - \hat{c})}{\eta \hat{c}^{\kappa+1}}.$$

Intuitively, entry induces a selection effect and also reduces the prices of the active firms. The last relation defines the cutoff $\hat{c} = \hat{c}(N)$ as a decreasing function of the mass of firms (as long as it is below α).

Under free entry, the expected profits match the fixed costs of entry F, and we can compute the equilibrium cutoff as:

$$\hat{c}^e = \left[\frac{2(\kappa + 2)(\kappa + 1)\overline{c}^{\kappa} \gamma F}{L} \right]^{\frac{1}{\kappa+2}}, \quad (45)$$

and recover the equilibrium measure of entrants N^e of Melitz and Ottaviano (2008) from the condition above.

Let us now consider total welfare as a function of N:

$$\begin{aligned} W(N) &= \left[N \int_0^{\hat{c}(N)} \frac{(M - p(c))^2}{2\gamma} dG(c) - \theta(M) \right] L + \\ &\quad + N \left[\int_0^{\hat{c}(N)} \frac{(p(c) - c)(M - p(c))L}{\gamma} dG(c) - F \right] + EL \end{aligned}$$

[30] See Appendix A.2 for the case of a general cost distribution.

where $p(c) = \frac{c+M}{2}$ and $M = \hat{c}(N)$. Clearly, the first term represents the aggregate consumer surplus and the second term the aggregate expected profits. The mass of firms has direct effects on both terms, but the second is null at the free entry equilibrium. In addition, entry exerts indirect welfare impacts through the cutoff, the aggregator and pricing. In this model the first impact is null, because both the incremental surplus provided by the cutoff firms and their demand are zero in this case. The indirect impact on consumer surplus through the aggregator is null due to the welfare neutrality of the latter. The indirect impact on aggregate profits through the aggregator represents the business stealing effect, which is partially compensated by the competitive effect on pricing. Finally, price changes for given quantities are neutral on welfare by the Roy's identity.

Formally, we obtain:

$$W'(N^e) = \int_0^{\hat{c}^e} \frac{(\hat{c}^e - c)^2}{8\gamma} dG(c)L + N^e L \left[\int_0^{\hat{c}^e} \frac{(\hat{c}^e - c)}{2\gamma} dG(c) \right] \hat{c}'(N^e).$$

and, under the Pareto distribution, we can directly compute the welfare impact of entry. This is negative for $\hat{c}^e \in \left(0, \frac{\alpha}{2}\right)$, but positive, implying insufficient entry, if:

$$\hat{c}^e \in \left(\frac{\alpha}{2}, \alpha\right). \tag{46}$$

Since the equilibrium cutoff (45) is increasing in γ, insufficient entry requires a sufficiently high product differentiation. Intuitively, in this case the consumer surplus effect and the competitive effect are strong enough relatively to the business stealing effect to induce insufficient entry in equilibrium.[31]

6 Conclusion

In this work we have argued that the case for insufficient entry in monopolistic competition is stronger than usually thought. Following the partial equilibrium approach of Spence (1976) we have adopted quasi-linear preferences that generate demand systems depending on a price aggregator. Our main concern has been whether entry is excessive or insufficient, taking as given market pricing in a second-best sense *à la* Mankiw and Whinston (1986). The comparison is usually regarded as ambiguous due to the presence of consumer surplus and business stealing effects. The former pushes for insufficient entry of new varieties, and the latter for excess entry, with complications due to the effects of entry on prices (see Vives, 1999). Also the comparison with first-best allocations is generally ambiguous, as in models with homogeneous firms (Dixit and Stiglitz,

[31] This third-best result would emerge also with a second-best approach, since the welfare derivative with respect to the cutoff is null (in fact, also a social planner would set the cutoff at the choke price M). While this entry analysis *à la* Mankiw and Whinston (1986) has not been explored with the Melitz-Ottaviano model, we refer to Nocco *et al.* (2014) for the comparison between equilibrium and first-best allocations.

1977) and with heterogeneous firms (Dhingra and Morrow, 2019), where also market selection can be inefficient.

We have explored the welfare impact of entry starting from a framework where a price aggregator is a demand shifter, as in common models with Logit and isoelastic demands and more general models of product differentiation (Nocke and Schutz, 2018; Anderson, Erkal and Piccinin, 2020). Their characteristic feature is that, since entry does not affect demand elasticity, there are no effects of entry on prices under monopolistic competition. Within this framework, we found that entry is (weakly) insufficient *à la* Mankiw and Whinston (1986), in the sense that the business stealing effect cannot dominate the consumer surplus effect of entry. We have also analyzed CRESS preferences that encompass the cases of Logit, CES and other demand systems. Within these preferences the market provides the first-best production of each good and efficient selection under heterogeneous firms, and entry is (weakly) insufficient also compared to the first-best. However, we have shown that a special preference specification which includes the Logit case leads to fully efficient allocations.

We have then extended the analysis to a novel and more general type of quasi-linear preferences that nests our baseline setting, the one of Spence (1976) and others with demand systems depending on a price aggregator in a flexible way, as in Melitz and Ottaviano (2008). These quasi-linear Gorman-Pollak preferences account for competitive and selection effects of entry. Our main finding is that the competitive effects push for insufficient entry, for the simple reason that price reductions due to entry limit the strength of business stealing. We have finally derived conditions for insufficient entry in this more general environment.

Future research may evaluate the welfare properties of monopolistic competition with endogenous spending in differentiated goods beyond the restriction of quasi-linearity. Our basic framework could be exploited to study strategic interactions among a discrete number of firms, which may strengthen business stealing effects and change the welfare assessment of entry. In fact, Logit and CES demand systems have been often used to study Bertrand competition, and IIA preferences generate aggregative games that have been already used to study multiproduct pricing, commitments by market leaders,[32] and pricing by sellers on platforms, and could be applied further. Finally, our framework could inspire new macroeconomic and trade applications due to the flexible impact that population growth exerts on investments in new goods and trade liberalization exerts on costly trade.

References

Alfaro, Martin, 2022, The Microeconomics of New Trade Models, *Canadian Journal of Economics*, 55, 3, 1539-65.

Anderson, Simon and Özlem Bedre-Defolie, 2024, Hybrid platform model: monopolistic competition and a dominant firm, *RAND Journal of Economics.*

Anderson Simon and André de Palma, 2020, Decoupling the CES Distribution

[32] This is the environment where strategies and commitments by market leaders are neutral on consumer welfare under free entry of price-setting followers.

Circle with Quality and Beyond: Equilibrium Distributions and the CES-Logit Nexus, *Economic Journal*, 130, 628, 911-936.

Anderson Simon, André de Palma and Jacques-François Thisse, 1992, *Discrete Choice Theory of Product Differentiation*, MIT Press: Cambridge.

Anderson, Simon, Nisvan Erkal and Daniel Piccinin, 2020, Aggregative games and oligopoly theory: short-run and long-run analysis, *RAND Journal of Economics*, 51, 2, 470-495.

Bagwell, Kyle and Seung Hoon Lee, 2023, Monopolistic Competition and Efficiency under Firm Heterogeneity and Non-additive Preferences, *American Economic Journal: Microeconomics*, 15, 4, 208-67.

Berry, Steve and Joe Waldfogel, 1999, Free Entry and Social Inefficiency in Radio Broadcasting, *RAND Journal of Economics*, 30, 3, 397-420.

Bertoletti, Paolo, 2018, A Note on Ramsey Pricing and the Structure of Preferences, *Journal of Mathematical Economics*, 76, 45-51.

Bertoletti, Paolo and Federico Etro, 2016, Preferences, Entry and Market Structure, *RAND Journal of Economics*, 47, 4, 792-821.

Bertoletti, Paolo and Federico Etro, 2017, Monopolistic Competition when Income Matters, *Economic Journal*, 127, 603, 1217-43.

Bertoletti, Paolo and Federico Etro, 2021, Monopolistic competition with Generalized Additively Separable preferences, *Oxford Economic Papers*, 73, 2, 927-52.

Bertoletti, Paolo, Federico Etro and Ina Simonovska, 2018, International Trade with Indirect Additivity, *American Economic Journal: Microeconomics*, 10, 2, 1-57.

Besanko, David, Martin Perry and Richard Spady, 1990, The Logit Model of Monopolistic Competition: Brand diversity, *Journal of Industrial Economics*, 38, 4, 397-415.

Bisceglia, Michele, Jorge Padilla, John Perkins and Salvatore Piccolo, 2023, On Excessive Entry in Bayes-Cournot Oligopoly, *RAND Journal of Economics*.

Dhingra, Swathi and John Morrow, 2017, Efficiency in Large Markets with Firm Heterogeneity, *Research in Economics*, 71, 4, 718-728.

Dhingra, Swathi and John Morrow, 2019, Monopolistic Competition and Optimum Product Diversity Under Firm Heterogeneity, *Journal of Political Economy*, 127, 196-232.

Dixit, Avinash and Joseph Stiglitz, 1977, Monopolistic Competition and Optimum Product Diversity, *American Economic Review*, 67, 297-308.

Etro, Federico, 2021, Product Selection in Online Marketplaces, *Journal of Economics and Management Strategies*, 30, 3, 614-37.

Etro, Federico, 2023, Hybrid Marketplaces with Free Entry of Sellers, *Review of Industrial Organization*, 62, 2, 119-148.

Fally, Thibault, 2019, Generalized Separability and the Gains from Trade, *Economics Letters*, 178 (C), 8-12.

Fally, Thibault, 2022, Generalized Separability and Integrability: Consumer Demand with a Price Aggregator, *Journal of Economic Theory*, 203, 105471.

Gorman, William Moore, 1970, Conditions for Generalized Additive Separability, in *Collected Works of W.M. Gorman: Separability and Aggregation*, 1,

186-207, Oxford University Press.
Krugman, Paul, 1979, Increasing Returns, Monopolistic Competition and International Trade, *Journal of International Economics*, 9, 469-79.
Krugman, Paul, 1980, Scale Economies, Product Differentiation, and the Pattern of Trade, *American Economic Review*, 70, 950-9.
Krugman, Paul, Maurice Obstfeld and Marc Melitz, 2022, *International Economics: Theory and Policy*, 12th Edition, Pearson.
Kuhn, Kai-Uwe and Xavier Vives, 1999, Excess Entry, Vertical Integration, and Welfare, *RAND Journal of Economics*, 30, 4, 575-603.
Lashkaripour, Ahmad, 2020, Within-industry Specialization and Global Market Power, *American Economic Journal: Microeconomics*, 12, 1, 75-124.
Macedoni, Luca and Ariel Weinberger, 2022, Quality Heterogeneity and Misallocation: The welfare benefits of raising your standards, *Journal of International Economics*, 134, 103544.
Mankiw, Gregory and Michael Whinston, 1986, Free Entry and Social Inefficiency, *Rand Journal of Economics*, 17, 1, 48-58.
Melitz, Marc, 2003, The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity, *Econometrica*, 71, 6, 1695-725.
Melitz, Marc and Gianmarco Ottaviano, 2008, Market Size, Trade, and Productivity, *Review of Economic Studies*, 75, 1, 295-316.
Mrázová, Monika and Peter Neary, 2019, Selection Effects with Heterogeneous Firms, *Journal of the European Economic Association*, 17, 4, 1294–334.
Nocco, Antonella, Gianmarco Ottaviano and Matteo Salto, 2014, Monopolistic Competition and Optimum Product Selection, *American Economic Review*, 104, 5, 304-309.
Nocco, Antonella, Gianmarco Ottaviano and Matteo Salto, 2017, Monopolistic Competition and Optimum Product Selection: Why and how heterogeneity matters, *Research in Economics*, 71, 4, 704-717.
Nocco, Antonella, Gianmarco Ottaviano and Matteo Salto, 2019, Geography, Competition, and Optimal Multilateral Trade Policy, *Journal of International Economics*, 120, 145-161.
Nocke, Volker and Nicolas Schutz, 2018, Multiproduct-firm Oligopoly: An aggregative games approach, *Econometrica*, 86, 2, 523-557.
Pollak, Robert, 1972, Generalized Separability, *Econometrica*, 40, 3, 431-53.
Salop, Steven, 1979, Monopolistic Competition with Outside Goods, *Bell Journal of Economics*, 10, 1, 141-156.
Simonovska, Ina, 2015, Income Differences and Prices of Tradables: Insights from an online retailer, *Review of Economic Studies*, 82, 4, 1612-1656.
Spence, Michael, 1976, Production Selection, Fixed Costs, and Monopolistic Competition, *Review of Economic Studies*, 43, 2, 217-35.
Suzumura, Kotaro, 2012, Excess Entry Theorems after 25 Years, *Japanese Economic Review*, 63, 152-170.
Vives, Xavier, 1999, *Oligopoly Pricing: old ideas and new tools*, MIT Press.
Zennyo, Yusuke, 2022, Platform Encroachment and Own-content Bias, *Journal of Industrial Economics*, 70, 3, 684-710.

Appendix

A. Quasi-linear preferences *à la* Gorman-Pollak

In this Appendix we extend our analysis to more general demand systems characterized by a price aggregator, nesting our baseline case as well as other well-known cases, as those emerging in Spence (1976) and Melitz and Ottaviano (2008). We do so by introducing *quasi-linear preferences à la Gorman-Pollak*, building on contributions by Fally (2019, 2022) and Bertoletti and Etro (2021) originally developed in an environment without an outside commodity due to Gorman (1970) and Pollak (1972). The quasi-linear version with a *numéraire* has special characteristics which allows us to generalize and clarify the conditions for insufficient entry.

Consider preferences that can be represented by the following quasi-linear indirect utility:

$$V = \int_{\omega\in\Omega} s(p(\omega), M)d\omega - \theta(M) + E, \tag{47}$$

where the incremental surplus $s(p, M) > 0$ of a commodity depends on its price p and a price aggregator M, which is implicitly defined by:

$$\int_{\omega\in\Omega} s_M(p(\omega), M)d\omega = \theta'(M) \tag{48}$$

(we denote partial derivatives with subscripts). Let us assume, under the suitable differentiability, that $s_M > 0$, $s_p < 0$, $s_{pp} > 0$, $s_{Mp} < 0 \leqslant s_{MM}$, and that the function $\theta(M)$ is increasing and concave, with curvature captured by $\beta(M) = \frac{-\theta''(M)M}{\theta'(M)} > 0$. Notice that (47)-(48) imply that the direct impact of changes in the aggregator on consumer welfare is null ($\partial V/\partial M = 0$), and that our other assumptions imply that in this setting M increases with the prices of goods and decreases with entry. The latter features are just a matter of convenience, and while they are the opposite of what happens in our baseline model to the price aggregator A, as we show below the ultimate impacts of prices and entry on demand through the aggregator are nevertheless consistent.

The Roy's identity provides the following demand function for each variety:

$$q(p(\omega), M) = |s_p(p(\omega), M)|,$$

which is decreasing in its own price and increasing in the price aggregator in a flexible way. Therefore the price aggregator is not necessarily a demand shifter as it was under IIA preferences. We define the following, positive elasticities of surplus and demand with respect to price and aggregator:

$$\begin{aligned} \zeta_p(p, M) &= \frac{|s_p(p, M)|\, p}{s(p, M)} \quad \text{and} \quad \zeta_M(p, M) = \frac{s_M(p, M)M}{s(p, M)}, \\ \varepsilon_p(p, M) &= \frac{s_{pp}(p, M)p}{|s_p(p, M)|} \quad \text{and} \quad \varepsilon_M(p, M) = \frac{s_{pM}(p, M)M}{s_p(p, M)}. \end{aligned}$$

It is easily verified that:

$$\frac{\partial \ln \zeta_p(p,M)}{\partial \ln M} = \varepsilon_M(p,M) - \zeta_M(p,M), \tag{49}$$

implying that the price elasticity of surplus (locally) increases with respect to the aggregator M if and only if demand is more sensitive than surplus to its changes. Since $\zeta_p(p,M)$ represents the ratio of revenues and incremental surplus generated by a good, entry reduces the aggregator and this ratio if $\varepsilon_M(p,M) > \zeta_M(p,M)$. In other words, under this condition entry increases surplus compared to revenues. This was not possible under IIA preferences where entry was not affecting the ratio of revenues and incremental surplus.

We can also obtain a direct utility representing preferences (47)-(48). In particular, the direct demand can be inverted with respect to p to obtain a function $p = g\left(q, M\right)$. Given this, we can use (48) to define the *quantity* aggregator $\bar{M}$ by:

$$\int_{\omega\in\Omega} s_M(g(q(\omega),\bar{M}),\bar{M})d\omega = \theta'(\bar{M}).$$

The direct utility U dual to (47) then satisfies:

$$\begin{aligned}
U &= \underset{\{p(\omega)\}}{Min}\left\{\int_{\omega\in\Omega} s\left(p(\omega),M\right)d\omega - \theta(M) + \int_{\omega\in\Omega} p\left(\omega\right)q\left(\omega\right)d\omega\right\} + Y \\
&= \int_{\omega\in\Omega}\left[s(g(q(\omega),\bar{M}),\bar{M}) + g(q(\omega),\bar{M})q\left(\omega\right)\right]d\omega - \theta(\bar{M}) + Y \\
&= \int_{\omega\in\Omega} z(q(\omega),\bar{M})d\omega - \theta(\bar{M}) + Y,
\end{aligned}$$

where Y is the quantity of the *numéraire* and

$$z(q,\bar{M}) = s(g(q,\bar{M}),\bar{M}) + g(q,\bar{M})q$$

is the "sub-utility" (incremental surplus) of q when the quantity aggregator is $\bar{M}$. Since we can compute:

$$z_{\bar{M}}(q,\bar{M}) = s_M(g(q,\bar{M}),\bar{M}),$$

the quantity aggregator $\bar{M}$ also satisfies the condition:

$$\int_{\omega\in\Omega} z_{\bar{M}}(g(q(\omega),\bar{M}),\bar{M})d\omega = \theta'(\bar{M}),$$

and the inverse demand of commodity ω can be computed as:

$$p(\omega) = z_q(q(\omega),\bar{M}).$$

We now show how preferences represented by (47)-(48) nest our and other well-known preferences.

IIA preferences The quasi-linear GP preferences nest our baseline preferences represented by the indirect utility (1). This happens when:

$$s(p, M) = Mv(p) \tag{50}$$

because the Roy's identity provides the direct demand $q(p, M) = M\,|v'(p)|$, which is multiplicative in the price aggregator. Notice that in such a case M is just a demand shifter. Accordingly, (48) implies:

$$\int_{\omega\in\Omega} v(p(\omega))d\omega = \theta'(M),$$

and the price aggregator is additive across varieties. Replacing it within the indirect utility, we can express the latter as:

$$V = M\theta'(M) - \theta(M) + E,$$

which corresponds to our baseline case under suitable assumptions on the function $\theta(M)$.[33] In this case the elasticities $\zeta_p = \zeta(p)$, $\varepsilon_p = \varepsilon(p)$, $\zeta_M = \varepsilon_M = 1$ do not depend on the aggregator, and in particular $\partial \ln \zeta_p / \partial \ln M = 0$.

Spence preferences The quasi-linear GP preferences also nest the preferences considered by Spence (1976) and represented by a direct utility:

$$U = \Gamma\left(\int_{\omega\in\Omega} u(q(\omega))d\omega\right) + Y \tag{51}$$

where the quantity aggregator:

$$B = \int_{\omega\in\Omega} u(q(\omega))d\omega$$

is additive across varieties, and $\Gamma(B)$ and $u(q)$ are increasing and concave functions with $u(0) = 0$. These preferences provide an inverse demand $p(\omega) = u'(q(\omega))\Gamma'(B)$ that decreases in the common quantity aggregator.

To nest this case we employ the following incremental surplus function:

$$s(p, M) = Mv\left(\frac{p}{M}\right), \tag{52}$$

so that the Roy's identity provides the direct demand $q(p, M) = |v'(p/M)|$, which can be inverted with respect to p to obtain $p\,(q, M) = g(q)M$, where $g(q) \equiv v'^{-1}(-q)$ with $g'(q) = -1/v''(g(q)) < 0$.

Since in this case $s_M(p, M) = v(p/M) - (p/M)v'(p/M) > 0$, we can use (48) to define the (additive) quantity aggregator $\bar{M}$:

$$\int_{\omega\in\Omega} u(q(\omega))d\omega = \theta'(\bar{M}),$$

[33] For instance, $\theta(M) = \frac{\rho}{\rho-1}M^{\frac{\rho-1}{\rho}}$ provides $H(A) = \frac{A^{1-\rho}}{1-\rho}$ in our baseline notation, with $M = H'(A)$.

where

$$u(q) \equiv v(g(q)) - g(q)v'(g(q)), \quad \text{with } u'(q) = g(q) > 0 \text{ and } u''(q) = g'(q) < 0,$$

is the positive, increasing and concave sub-utility function of the quantity of each variety.

Then the direct utility can be expressed as:

$$U = \bar{M}\theta'(\bar{M}) - \theta(\bar{M}) + Y,$$

which corresponds to the one in Spence (1976) under suitable assumptions on the function $\theta(\bar{M})$. In this case the elasticities $\zeta_p = \zeta_M - 1 = \zeta(p/M)$ and $\varepsilon_p = \varepsilon_M = \varepsilon(p/M)$ depend on the ratio p/M, and are related to the more familiar quantity elasticities used in Spence (1976), Vives (1999) and elsewhere by:

$$\zeta\left(g(q)\right) = \frac{qu'(q)}{u(q) - qu'(q)} = \frac{\varphi(q)}{1-\varphi(q)} \quad \text{and} \quad \varepsilon\left(g(q)\right) = \frac{-g(q)}{qg'(q)} = \frac{-u'(q)}{qu''(q)} = \frac{1}{\epsilon(q)},$$

where $\varphi(q) = \frac{u'(q)q}{u(q)}$ and $\epsilon(q) = \frac{u''(q)q}{u'(q)}$ are the first-order and second-order elasticities of $u(q)$. Notice that:

$$\frac{\partial \ln \zeta(p/M)}{\partial \ln M} = \varepsilon(p/M) - 1 - \zeta(p/M) = -\frac{\partial \ln \zeta(p/M)}{\partial \ln p}$$

is positive when $\varepsilon > 1 + \zeta$, i.e. when $\varphi'(q) > 0$.

Melitz-Ottaviano preference specification The quasi-linear GP preferences nest also the Melitz and Ottaviano (2008) specification of the direct utility:

$$U = \int_{\omega\in\Omega} \left(\alpha q(\omega) - \gamma \frac{q(\omega)^2}{2}\right) d\omega - \frac{\eta Q^2}{2} + Y, \tag{53}$$

where $Q = \int_{\omega\in\Omega} q(\omega)d\omega$ is total quantity and we used the original parametrization of the authors with α, γ, $\eta > 0$. This happens in our framework when:

$$s(p(\omega), M) = \frac{\left(M - p\left(\omega\right)\right)^2}{2\gamma}, \tag{54}$$

and

$$\theta(M) = \frac{\alpha M}{\eta} - \frac{M^2}{2\eta}.$$

Then, the Roy's identity provides the linear demand $q\left(p, M\right) = (M - p)/\gamma$, while from (48) the aggregator can be explicitly derived as the choke price:

$$M = \frac{\alpha\gamma + \eta \int_{\omega\in\Omega} p(\omega)d\omega}{\gamma + \eta n},$$

which can be also expressed in terms of the average price $\frac{\int_{\omega\in\Omega} p(\omega)d\omega}{n}$. By inverting the direct demand with respect to the price, we obtain the quantity aggregator:

$$\bar{M} = \alpha - \eta Q,$$

and the direct utility function:

$$U = \int_{\omega\in\Omega} \left(\bar{M} q(\omega) - \gamma \frac{q(\omega)^2}{2} \right) d\omega - \theta(\bar{M}) + Y$$

which is equal to (53) up to a constant, and delivers the inverse demand $p(\omega) = \bar{M} - \eta Q = \alpha - \gamma q(\omega) - \eta Q$. The relevant elasticities are $\zeta_p = \frac{2p}{M-p} = 2\varepsilon_p$ and $\zeta_M = \frac{2M}{M-p} = 2\varepsilon_M$, which decrease with respect to the price aggregator.

A.1. Monopolistic competition with homogeneous firms

We now consider the equilibrium of monopolistic competition when all firms have a common marginal cost c, with profits given by:

$$\pi(p, M) = (p - c)q(p, M)L - F,$$

where by Roy's identity:

$$q(p, M) = |s_p(p, M)| . \tag{55}$$

The monopolistic competition price p satisfies:

$$p = \frac{\varepsilon_p(p, M)c}{\varepsilon_p(p, M) - 1} \tag{56}$$

under the assumption $\varepsilon_p(p, M) > 1$, where from (48) the value of the aggregator is given by:

$$n s_M(p, M) = \theta'(M) \tag{57}$$

as a function of the common price and the number of firms. These conditions provide the equilibrium price $p = p(n)$ and the aggregator $M = M(n)$ in function of the number of firms.

In a free entry equilibrium the zero-profit condition implies that the equilibrium number of firms n^e satisfies:

$$n^e = \frac{\zeta_p(p^e, M^e)\, \theta'(M^e) M^e L}{\varepsilon_p(p^e, M^e)\, \zeta_M(p^e, M^e)\, F},$$

where (p^e, M^e) are implicitly determined by (56)-(57) for the given value of n^e.

In a constrained optimum, the number of firms is selected to maximize welfare, which is given by the sum of aggregate consumer surplus and total profits:

$$W(n) = [ns(p, M) - \theta(M)]\, L + n\, [(p - c)q(p, M)L - F] + EL.$$

By using (55) and (57) the welfare impact of entry can be written as:

$$W'(n) = s(p,M)L + \pi\,(p,M) + n(p-c)\frac{\partial q(p,M)}{\partial M}\frac{\partial M}{\partial n}L + n(p-c)\frac{dq(p,M)}{dp}p'\,(n)\,L.$$

It includes four terms:

1) a positive consumer surplus effect;

2) a producer surplus effect, which is null in equilibrium;

3) a negative business stealing effect through the direct impact of n on the aggregator M, for which entry reduces the demand of each incumbent firm at given prices, and then its profit;

4) a "competitive" effect through the impact of n on price p, which affects (directly and indirectly, through its impact on the aggregator) the demand, generating an additional impact on profits.

The last term was absent under our baseline IIA preferences since in such a case there were no effects on prices. In general, it has an unambiguous sign because the convexity of preferences requires that the demand of differentiated products should be non-increasing in their common price ($\frac{dq(p,M)}{dp} \leqslant 0$). As a consequence, whenever entry exerts a competitive effect reducing prices (namely, $p'\,(n) \leq 0$), it must be that this works (if any) to limit the size of the business stealing effect, thereby strengthening any tendency toward insufficient entry. This shows that our case for insufficient entry does *not* actually rely on the absence of competitive effects.

More formally, differentiating the equilibrium conditions, the convexity of preferences requires:

$$\frac{d\ln q(p,M)}{d\ln p} = -\varepsilon_p + \frac{\zeta_p \varepsilon_M}{\zeta_M}\rho \leqslant 0, \tag{58}$$

where $\rho(p,M) = \left|\frac{\partial \ln q}{\partial \ln M}\frac{\partial \ln M}{\partial \ln n}\right|$ is the elasticity of quantity with respect to the number of firms at given prices. In particular, differentiating (57) we have:

$$\rho(p,M) = \frac{\varepsilon_M}{\beta(M) + \chi(p,M)} \quad \text{with } \chi(p,M) = \frac{s_{MM}(p,M)M}{s_M(p,M)},$$

and (58) follows from the fact that:

$$\frac{\partial \ln M}{\partial \ln p} = \frac{\zeta_p}{\zeta_M}\rho \quad \text{and} \quad \frac{\partial \ln M}{\partial \ln n} = \frac{-1}{\beta(M) + \chi(p,M)}.$$

Moreover, we can derive the impact of entry on equilibrium prices from the system (56)-(57). Differentiating it, we obtain in matrix notation:

$$\begin{bmatrix} \varepsilon_p(p,M) + (p-c)\frac{\partial \varepsilon_p(p,M)}{\partial p} - 1 & (p-c)\frac{\partial \varepsilon_p(p,M)}{\partial M} \\ ns_{Mp}\,(p,M) & ns_{MM}\,(p,M) - \theta''(M) \end{bmatrix}\begin{bmatrix} p'\,(n) \\ M'\,(n) \end{bmatrix}$$
$$= \begin{bmatrix} 0 \\ -s_M\,(p,M) \end{bmatrix}.$$

Thus, by Cramer's rule:

$$\begin{aligned} p'(n) &= \frac{s_M(p,M)\,\frac{p}{M}\frac{\partial \ln \varepsilon_p(p,M)}{\partial \ln M}}{\Delta}, \\ M'(n) &= \frac{-s_M(p,M)\left[\varepsilon_p(p,M)-1+\frac{\partial \ln \varepsilon_p(p,M)}{\partial \ln p}\right]}{\Delta}, \end{aligned}$$

where

$$\begin{aligned} \Delta &= \left[\varepsilon_p(p,M)-1+\frac{\partial \ln \varepsilon_p(p,M)}{\partial \ln p}\right]\frac{ns_M(p,M)}{M}\left[\beta(M)+\chi(p,M)\right] \\ &\quad -\frac{\partial \ln \varepsilon_p(p,M)}{\partial \ln M}\frac{ns_{Mp}(p,M)\,p}{M}. \end{aligned}$$

This shows that, in general, we cannot determine the sign of the impact of entry on prices. However, it is convenient to focus on the most appealing scenario with competitive effects, namely $p'(n) \leq 0$. In particular, by defining the elasticity of the equilibrium prices with respect to the mass of firms as $\sigma \equiv -\frac{d\ln p}{d\ln n}$, we can express the welfare impact of entry in equilibrium ($\pi(p^e, M^e) = 0$) as proportional to total revenues:

$$W'(n^e) = \left[\frac{1}{\zeta_p} - \frac{\rho}{\varepsilon_p} + \left(1 - \frac{\varepsilon_M}{\zeta_M}\frac{\zeta_p\rho}{\varepsilon_p}\right)\sigma\right]\Bigg|_{n^e} q^e p^e L. \tag{59}$$

which provides the general condition for insufficient entry (43) in the text.[34]

By exploiting (58), in the scenario in which entry weakly decreases prices ($\sigma \geqslant 0$), a *sufficient* condition for *insufficient* entry (namely, $W'(n^e) \geq 0$) is given by:

$$\varepsilon_M \geqslant \zeta_M, \tag{60}$$

because this insures that $\frac{\varepsilon_p}{\zeta_p} \geqslant \rho$ and then that the overall expression (59) cannot be negative.

This sufficient condition requires that changes in the aggregator have not a stronger impact on surplus than on demand. From (49), this is equivalent to the condition that, for given prices, entry (weakly) increases the ratio between surplus and revenues generated by goods. Only when entry distorts enough the ratio of surplus and revenues in favour of the firms (or exerts anti-competitive effects), instances of excess entry can emerge. Of course, (60) is just a sufficient but not necessary condition for entry to be insufficient. We illustrate this fact by discussing our three main examples, which present a case where the sufficient condition is always satisfied, one where it can be satisfied, and one where it is never satisfied, but still entry can be insufficient due to strong enough competitive effects.

[34] We can also derive the second-best optimal number of firms $\widetilde{n}$ as satisfying:

$$\widetilde{n} = \frac{\zeta_p \theta'(\tilde{M})\tilde{M}L}{\varepsilon_p \zeta_M F}\left[1 - \rho + \frac{\varepsilon_p}{\zeta_p} + \sigma\left(\varepsilon_p - \frac{\varepsilon_M}{\zeta_M}\zeta_p\rho\right)\right],$$

where $(\tilde{p}, \tilde{M})$ are implicitly determined by (56)-(57) for a given value of $\tilde{n}$.

IIA preferences To start with, under our IIA baseline preferences $\varepsilon_M = \zeta_M$ and $\sigma = 0$, so that (60) is always satisfied: in particular, the convexity condition (58) becomes $\varepsilon/\zeta \geqslant \rho$ and thus (59) cannot be negative.

Spence preferences Under the preferences used by Spence (1976) the convexity condition (58) is always satisfied,[35] but the sign of (59) is ambiguous. Assuming competitive effects ($\sigma \geqslant 0$), (60) becomes $\varepsilon(p/M) \geqslant \zeta(p/M) + 1$, which implies that the price elasticity of the surplus function $\zeta(p/M)$ is (everywhere) either constant or decreasing. However, insufficient entry arises also under an increasing price elasticity of surplus as long as:[36]

$$\beta(M^e) \geqslant \frac{\zeta(p^e/M^e)}{1+\zeta(p^e/M^e)} \frac{d\ln\zeta(p^e/M^e)}{d\ln(p^e/M^e)}.$$

To the contrary, excess entry can only emerge when the surplus elasticity is increasing enough compared to the curvature measure $\beta(M^e)$.

Similarly, starting from the direct utility (51), one can show that insufficient entry emerges whenever the elasticity $\varphi(q)$ of the direct sub-utility $u(q)$ is (everywhere) increasing or not too much decreasing.[37] Excess entry emerges only when at the free-entry equilibrium $\varphi(q)$ is decreasing enough for a given curvature of the transformation $\Gamma(B)$ measured by $\xi(B) \equiv -\frac{\Gamma''(B)B}{\Gamma'(B)}$. This was indeed the case in the example of excess entry discussed by Mankiw and Whinston (1986), where $u(q) = aq - bq^2/2$ and $\xi = 2$.

Melitz-Ottaviano preference specification Under the Melitz-Ottaviano specification again (58) always holds,[38] and moreover (60) is never satisfied ($\zeta_M = 2\varepsilon_M$), suggesting that only the strength of the competitive effects can induce insufficient entry. To verify this we recall the surplus and demand elasticities for this case and compute the additional elasticities $\beta = \frac{M}{\alpha - M}$, $\chi = \frac{M}{M-p}$ and:

$$\rho(p, M) = \frac{\alpha - M}{\alpha - p}.$$

[35] By using $\chi = \frac{\varepsilon\zeta}{1+\zeta}$ one can verify that:

$$\frac{d\ln q(p,M)}{d\ln p} = -\varepsilon + \frac{\zeta\varepsilon}{1+\zeta}\frac{\varepsilon}{\beta+\chi} = \frac{-\varepsilon\beta}{\beta+\chi} < 0.$$

[36] From (59), a sufficient condition for insufficient entry under competitive effects is $\beta \geqslant \frac{\zeta(1+\zeta-\varepsilon)}{1+\zeta}$, that can be expressed as in the text after computing the elasticity of the surplus elasticity.

[37] It may be useful to point out that the related discussion of Spence (1976, p. 232) erroneously attributes his Fig. 6 to the case of a decreasing $\varphi(q)$.

[38] One can verify that:

$$\frac{d\ln q(p,M)}{d\ln p} = \varepsilon_p\left(\frac{\varepsilon_M}{\varepsilon_M+\beta} - 1\right) < 0.$$

Moreover, we employ the equilibrium price $p(n) = c + \frac{\gamma(\alpha-c)}{2\gamma+\eta n}$ emerging under homogeneous firms as a function of the number of firms. Then, we can compute:

$$\sigma(n) = \frac{n\eta\gamma(\alpha-c)}{(2\gamma+\eta n)\left[\gamma(\alpha+c)+\eta nc\right]} > 0.$$

The condition for insufficient entry becomes:

$$\left[\frac{1}{\zeta_p} - \frac{\rho}{\varepsilon_p} + \left(1 - \frac{\varepsilon_M}{\zeta_M}\frac{\zeta_p\rho}{\varepsilon_p}\right)\sigma\right]\Bigg|_{n^e} = \frac{M^e - p^e}{(\alpha - p^e)}\left[\frac{2M^e - p^e - \alpha}{2p^e} + \sigma(n^e)\right] \geq 0,$$

which holds if and only if:

$$2\gamma^2 - (n^e\eta)^2 + n^e\eta\gamma \geq 0 \iff n^e \leq \frac{2\gamma}{\eta},$$

that is satisfied for limited enough entry. The same result can be directly obtained by the derivative of welfare evaluated at the equilibrium number of firms:

$$W'(n^e) = \frac{\gamma(\alpha-c)^2 L}{2(2\gamma+\eta n^e)^3}(2\gamma - n^e\eta).$$

Since the Melitz-Ottaviano model with homogeneous firms has an equilibrium number of firms given by:

$$n^e = \left(\frac{\alpha-c}{\eta}\right)\sqrt{\frac{\gamma L}{F}} - \frac{2\gamma}{\eta},$$

insufficient entry occurs in equilibrium whenever:

$$\alpha \leqslant c + 4\sqrt{\frac{\gamma F}{L}},$$

This requires a strong enough competitive effect or a high product differentiation (high γ) that generates a large enough consumer surplus effect. Notice that the condition differs from the one of Nocco *et al.* (2017) because they did not consider the second best allocation analyzed here.

A.2. Monopolistic competition with heterogeneous firms

We finally sketch the analysis of the case where firms are heterogeneous in marginal costs. An active firm with marginal cost c sets a price according to the pricing rule:

$$p(c, M) = \frac{\varepsilon_p(p(c,M),M)c}{\varepsilon_p(p(c,M),M) - 1}, \tag{61}$$

generating the individual demand $q(c, M) = |s_p(p(c, M), M)|$, where the price aggregator M satisfies:

$$N\int_0^{\hat{c}} s_M(p(c,M),M)dG(c) = \theta'(M). \tag{62}$$

Its gross profits are:

$$\pi(c) = (p(c,M) - c)q(c,M)L - f = \frac{s(p(c,M),M)}{\eta(p(c,M),M)}L - f,$$

where

$$\eta(p(c,M),M) \equiv \frac{\varepsilon_p(p(c,M),M)}{\zeta_p(p(c,M),M)}.$$

The equilibrium $(N^e, M^e, \hat{c}^e)$ satisfies simultaneously the definition of the aggregator (62), the zero-profit condition:

$$\frac{s(p(\hat{c},M),M)}{\eta(p(\hat{c},M),M)}L = f, \tag{63}$$

and the free entry condition:

$$\int_0^{\hat{c}} \frac{s(p(c,M),M)}{s(p(\hat{c},M),M)}dG(c) = \left[\frac{F}{f} + G(\hat{c})\right] \frac{\overline{\eta}(\hat{c},M)}{\eta\left(p(\hat{c},M),M\right)},$$

where:

$$\overline{\eta}(\hat{c},M) = \left[\int_0^{\hat{c}} \frac{1}{\eta(p(c,M),M)} \frac{s(p(c,M),M)}{\int_0^{\hat{c}} s(p(c,M),M)dG(c)}dG(c)\right]^{-1}.$$

Notice that (63) together with (61) defines the equilibrium value of the cutoff $\hat{c}\,(M)$ as a function of the aggregator.

The constrained (third-best) optimal measure of firms for given price and activation rules maximizes total welfare under the constraints (61)-(63), which define $M = M\,(N)$, $\widehat{c}\,(N) \equiv \widehat{c}\,(M\,(N))$ and $p\,(c,N) \equiv p\,(c,M(N))$. By total differentiation we can compute $M'\,(N)$, $\widehat{c}'\,(N) = \widehat{c}'\,(M\,(N))\,M'\,(N)$ and $\frac{\partial p(c,N)}{\partial N} = \frac{\partial p(c,M(N))}{\partial M}M'\,(N)$.

Then the social optimum selects N to maximize welfare:

$$W\,(N) = NL\int_0^{\hat{c}(N)} s(p(c,N),M\,(N))dG - \theta(M\,(N))L + N\left[\int_0^{\hat{c}(N)} \pi(c)dG - F\right] + EL,$$

where $\pi(c) = (p(c,N) - c)q(c,N)L - f$ and $q(c,N) \equiv |s_p(p(c,N),M\,(N))|$.

Given this, the derivative of welfare with respect to N evaluated at the free entry equilibrium can be computed as follows:

$$W'(N^e) = N^e L s(p(\hat{c}\,(N^e)\,,N^e),M\,(N^e))g\,(\hat{c}\,(N^e))\,\hat{c}'\,(N^e) \tag{64}$$

$$+L\int_0^{\hat{c}(N^e)} \left[s(p(c,N^e),M\,(N^e)) + N^e(p(c,N^e) - c)\frac{dq(c,N^e)}{dN}\right]dG(c),$$

where

$$\frac{dq(c,N^e)}{dN} = \frac{\partial q(c,N^e)}{\partial p\,(c,N)}\frac{\partial p\,(c,N^e)}{\partial N} + \frac{\partial q(c,N^e)}{\partial M}M'\,(N^e)\,.$$

The first term of (64), which has a sign that depends on the sign of $\hat{c}'(N^e)$, accounts for the impact of entry on consumer welfare through the cutoff. This is null if the surplus provided by the varieties of the cutoff firms vanishes (as in the absence of fixed costs of production) and otherwise is negative if entry reduces the cutoff. Therefore selection effects tend to reduce the consumer surplus effect.

The direct impact of the cutoff on producer surplus is null by (63), and the business stealing effect ($\frac{dq(c,N^e)}{dN} < 0$) works by reducing $q(c, N^e)$ through a reduction of the aggregator ($M'(N^e) < 0$), which would be countered by any competitive effect of entry ($\frac{\partial p(c,N^e)}{\partial N} < 0$).

By using (62) to define $M = \underline{M}(N, \hat{c}, p(c, M))$ we can also disentangle the impact of entry on demand through the aggregator:

$$\frac{\partial \ln \underline{M}}{\partial \ln N} = -\frac{1}{\overline{\chi}(M) + \beta(M)} < 0,$$

where $\chi(p(c, M), M) = \frac{s_{MM}(p(c,M),M)M}{s_M(p(c,M),M)} > 0$ and $\overline{\chi}(M)$ is the average:

$$\overline{\chi}(M) = \int_0^{\hat{c}} \chi(p(c), M) \frac{s_M(p(c), M)}{\int_0^{\hat{c}} s_M(p(c), M) dG(c)} dG(c).$$

Moreover, we have:

$$\frac{\partial \ln \underline{M}}{\partial \ln \hat{c}} = -\frac{\frac{s_M(p(\hat{c},M),M)}{\overline{s_M}(M)} g(\hat{c})\hat{c}}{\overline{\chi}(M) + \beta(M)} \leqslant 0,$$

where $\overline{s_M}(M) = \int_0^{\hat{c}} s_M(p(c, M), M) dG(c)$, and

$$\frac{\partial \ln \underline{M}}{\partial \ln p(c, M)} = -\frac{\frac{\overline{s_{Mp}}(M)}{\overline{s_M}(M)} p(c, M)}{\overline{\chi}(M) + \beta(M)} > 0,$$

where $\overline{s_{Mp}}(M) = \int_0^{\hat{c}} s_{Mp}(p(c, M), M) dG(c) < 0$.

Accordingly we can finally rewrite the welfare derivative as

$$W'(N^e) = N^e L s(p(\hat{c}(N^e), N^e), M(N^e)) g(\hat{c}(N^e)) \hat{c}'(N^e) \tag{65}$$

$$+L \int_0^{\hat{c}(N^e)} \left\{ \begin{array}{c} s(p(c, N^e), M(N^e)) + \\ +(p(c, N^e) - c) q(c, N^e) \left[\begin{array}{c} \frac{\partial \ln q(c,N^e)}{\partial \ln M} T(N^e) \\ + \frac{d \ln q(c,N^e)}{d \ln p(c,N)} \frac{\partial \ln p(c,N^e)}{\partial \ln N} \end{array} \right] \end{array} \right\} dG(c),$$

where

$$T(N^e) \equiv \frac{\partial \ln \underline{M}(N^e)}{\partial \ln N} + \frac{\partial \ln \underline{M}(N^e)}{\partial \ln \hat{c}} \frac{\hat{c}'(N^e) N^e}{\hat{c}(N^e)}$$

and

$$\frac{d \ln q(c, N^e)}{d \ln p(c, N)} = \frac{\partial \ln q(c, N^e)}{\partial \ln p(c, N)} + \frac{\partial \ln q(c, N^e)}{\partial \ln M} \frac{\partial \ln \underline{M}(N^e)}{\partial \ln p(c, N)} < 0.$$

The last expressions confirm that, as in the case of homogenous firms, the competitive effects (here captured by $\frac{\partial \ln p(c,N^e)}{\partial \ln N} < 0$) would work (either directly or through the aggregator) to reduce the business stealing effect generated by entry through its direct impact on the aggregator ($\frac{\partial \ln \underline{M}(N^e)}{\partial \ln N} < 0$). In addition, the second term in the definition of $T(N^e)$ shows that the business stealing effect would also be countered by any selection effect of entry (namely $\hat{c}'(N^e) < 0$), as soon as $\frac{\partial \ln \underline{M}(N^e)}{\partial \ln \hat{c}} \neq 0$. Therefore selection effects tend to reduce also the business stealing effect.

In general, the evaluation of the sign of (65) is a rather complex matter (for which in principle one should also take into account the implications of the needed price convexity of the indirect utility function representing the quasi-linear GP preferences), and it seems a formidable task to assess the overall role of the selection effects (since they tend to reduce both consumer surplus and business stealing effects). However, in a framework without fixed cost of activation as in the model of Melitz and Ottaviano (2008), both the surplus provided by the cutoff firms and the elasticity of this surplus with respect to M can be null at the equilibrium, making null the direct and indirect welfare impact of entry through the cutoff $\hat{c}$. Accordingly, only the competitive effects would remain at work in the case of a marginal increase of entry. For this reason, a planner controlling not only entry but also the activation threshold (in a second-best analysis) would not marginally change it starting at the monopolistic competition equilibrium.

Insufficient entry in the Melitz-Ottaviano model We verify the conditions under which entry can be insufficient or excessive in the model of Melitz and Ottaviano (2008) taking as given the pricing rules of the firms *à la* Mankiw and Whinston (1986). As far as we know, such a second best analysis has not been explored in the literature (while comparisons between equilibrium and first best allocations are explored in Nocco *et al.*, 2014).

We consider heterogeneous firms with a generic cost distribution $G(c)$, generalizing the case of a Pareto distribution presented in Section 5 of the text. In our framework the indirect utility for the Melitz-Ottaviano model can be written as:

$$V = N \int_0^{\widehat{c}} \frac{(M - p(c))^2}{2\gamma} dG(c) - \theta(M) + E,$$

where $\theta(M) = \frac{\alpha M}{\eta} - \frac{M^2}{2\eta}$ and the aggregator M satisfies:

$$N \int_0^{\widehat{c}} \frac{M - p(c)}{\gamma} dG(c) = \frac{\alpha - M}{2\eta}. \tag{66}$$

In fact, each firm with a marginal cost c has demand and gross profits given respectively by:

$$q = \frac{M - p}{\gamma} \text{ and } \pi = \frac{(p - c)(M - p)L}{\gamma}$$

and there are no fixed costs of production, namely $f = 0$.

Since the pricing rule is given by:

$$p\left(c, M\right) = \frac{c + M}{2}, \tag{67}$$

the gross profit expression can be written as:

$$\pi\left(c, M\right) = \frac{(M - c)^2 L}{4\gamma}.$$

Accordingly, the zero profit of the cutoff firm (namely, $\pi\left(\hat{c}, M\right) = 0$) implies

$$\hat{c} = M, \tag{68}$$

and the expressions for the pricing rule, the gross profit and the aggregator can be re-written as

$$p\left(c\right) = \frac{c + \hat{c}}{2}, \quad \pi\left(c\right) = \frac{(\hat{c} - c)^2 L}{4\gamma}, \text{ and } N \int_0^{\hat{c}} \frac{\hat{c} - c}{2\gamma} dG\left(c\right) = \frac{\alpha - \hat{c}}{2\eta}.$$

Then, the free-entry condition:

$$\int_0^{\hat{c}^e} \frac{(\hat{c}^e - c)^2}{4\gamma} L dG(c) = F \tag{69}$$

identifies the equilibrium value of the cutoff $\hat{c}^e$ (alternatively, of the aggregator M^e). Integrating twice by parts delivers:

$$\int_0^{\hat{c}^e} \breve{G}\left(c\right) dc = \frac{2\gamma F}{L}, \tag{70}$$

where we defined $\breve{G}(c) \equiv \int_0^c G(x)dx$.

By using (66) we also get:

$$N \int_0^{\widehat{c}} \left[\hat{c} - p\left(c\right)\right] dG\left(c\right) = \gamma \frac{\alpha - \hat{c}}{\eta},$$

or, after integrating by parts,

$$N = 2\gamma \frac{\alpha - \hat{c}}{\eta \breve{G}\left(\hat{c}\right)}. \tag{71}$$

Given the constraints (67)-(68), the formula (71) implicitly determines the value of the cutoff (alternatively, the aggregator) as a function of the number of entrant firms $\hat{c}\left(N\right)$ (alternatively, $M\left(N\right)$). And we can compute its derivative by total differentiation:

$$\hat{c}'\left(N\right) = -\frac{\eta \breve{G}\left(\hat{c}\right)}{N\eta G\left(\hat{c}\right) + 2\gamma} < 0. \tag{72}$$

Let us now consider the welfare impact of entry. The planner controls N under the constraints (66)-(68). We can then write welfare as

$$\begin{aligned} W &= \left[N \int_0^{\hat{c}(N)} \frac{\left(\hat{c}\left(N\right) - c\right)^2}{8\gamma} dG(c) - \theta(\hat{c}\left(N\right)) \right] L + EL \\ & \quad + N \left[\int_0^{\hat{c}(N)} \frac{(\hat{c}\left(N\right) - c)^2 L}{4\gamma} dG(c) - F \right]. \end{aligned}$$

Thus, by using (66) and $\pi\left(\hat{c}\left(N^e\right)\right) = 0$, we have:

$$\begin{aligned} W'\left(N^e\right) &= L \int_0^{\hat{c}^e} \frac{\left(\hat{c}^e - c\right)^2}{8\gamma} dG(c) \\ & \quad + N^e L \left[\int_0^{\hat{c}^e} \frac{\left(\hat{c}^e - c\right)}{2\gamma} dG(c) \right] \hat{c}'\left(N^e\right). \end{aligned}$$

By integrating by parts the second term and using (69),(71) and (72) we obtain:

$$W'\left(N^e\right) = \frac{F}{2} + N^e L \frac{\breve{G}\left(\hat{c}^e\right)}{2\gamma} \hat{c}'\left(N^e\right) \geqslant 0,$$

which implies insufficient entry if and only if:

$$\frac{2\gamma F}{L} \breve{G}\left(\hat{c}^e\right) + \left[\frac{2\gamma F}{L} G\left(\hat{c}^e\right) - \breve{G}\left(\hat{c}^e\right)^2 \right] \left(\alpha - \hat{c}^e\right) \geqslant 0, \tag{73}$$

where it can be proven that the term within the square bracket is negative.

In the case of a Pareto distribution, namely $G\left(c\right) = \left(c/\overline{c}\right)^\kappa$ with $\kappa \geqslant 1$, we can compute:

$$\breve{G}\left(c\right) = \frac{c^{\kappa+1}}{\overline{c}^\kappa \left(\kappa + 1\right)} \text{ and } \int_0^{\hat{c}} \breve{G}\left(c\right) dc = \frac{\hat{c}^{\kappa+2}}{\overline{c}^\kappa \left(\kappa + 1\right)\left(\kappa + 2\right)},$$

Then, using (70) and (73) we obtain that there is insufficient entry if and only if:

$$\alpha \leqslant 2\hat{c}^e$$

as reported in the main text.

B. Direct utility for IIA and CRESS preferences

In this Appendix we derive the direct utility (4) for the general IIA preferences and for the CRESS specifications used in the text. According to standard duality results, the direct utility dual to (1) is given by:

$$U = \underset{\{p(\omega)\}}{Min} \left\{ H \left(\int_{\omega \in \Omega} v\left(p\left(\omega\right)\right) d\omega \right) + \int_{\omega \in \Omega} p\left(\omega\right) q\left(\omega\right) d\omega \right\} + Y, \tag{74}$$

where Y is the amount of the *numéraire*. The FOCs for this problem require that:

$$-H'(A)\,v'(p(\omega)) = q(\omega) \text{ for any } \omega.$$

They can be rewritten as:

$$p(\omega) = v'^{-1}\left(\frac{-q(\omega)}{H'(\bar{A})}\right), \tag{75}$$

where from (2) the quantity aggregator $\bar{A}$ is implicitly defined by:

$$\bar{A} = \int_{\omega\in\Omega} v\left(v'^{-1}\left(\frac{-q(\omega)}{H'(\bar{A})}\right)\right) d\omega \tag{76}$$

and is not necessarily additive across varieties. Accordingly, the direct utility can be expressed as in the text:

$$U = H(\bar{A}) + \int_{\omega\in\Omega} q(\omega)\,v'^{-1}\left(\frac{-q(\omega)}{H'(\bar{A})}\right) d\omega + Y.$$

We now consider this direct utility in the case of surplus functions belonging to the CRESS family (10) when the transformation $H(A)$ is either isoelastic as in (11) or logarithmic.

1) When $H(A) = \frac{A^{1-\rho}}{1-\rho}$ and $v(p) = (\alpha p - a)^{\frac{1}{1-\eta}}$ for $\alpha p > a \geqslant 0$, $\eta \geqslant \rho \neq 1$ and $\eta > 1$, we can compute (75) and (76) as:

$$p(\omega) = \frac{a}{\alpha} + \frac{(\eta-1)^{\frac{1-\eta}{\eta}}\, q(\omega)^{\frac{1-\eta}{\eta}}\, \bar{A}^{\frac{\rho(1-\eta)}{\eta}}}{\alpha^{\frac{1}{\eta}}}, \tag{77}$$

and

$$\bar{A} = \left(\frac{\eta-1}{\alpha}\right)^{\frac{1}{\eta}} \bar{A}^{\frac{\rho}{\eta}} \int_{\omega\in\Omega} q(\omega)^{\frac{1}{\eta}} d\omega.$$

If $\eta > \rho$, the last expression defines $\bar{A}$ and can be solved explicitly as a transformation of a CES aggregator:

$$\bar{A} \equiv \left(\frac{\eta-1}{\alpha}\right)^{\frac{1}{\eta-\rho}} \left[\int_{\omega\in\Omega} q(\omega)^{\frac{1}{\eta}} d\omega\right]^{\frac{\eta}{\eta-\rho}}. \tag{78}$$

Using this, the direct utility can be computed as:

$$U = \frac{\eta-\rho}{\eta-1}\frac{\bar{A}^{1-\rho}}{1-\rho} + \frac{a}{\alpha}Q + Y,$$

where

$$Q \equiv \int_{\omega\in\Omega} q(\omega) d\omega$$

is the "aggregate" quantity (intuitively, when $\rho = 0$ the *numéraire* and the CES aggregator are perfect substitutes). The direct utility is a concave function of quantities (its first term is an increasing concave transformation of a concave function), and in general depends on two distinct, additive aggregators, $\int_\omega q(\omega)^{\frac{1}{\eta}} d\omega$ and $\int_\omega q(\omega) d\omega$, departing from the preferences of Spence (1976). Notice, however, that the case of a power surplus function emerging when $a = 0$ has a single aggregator and it is nested within the Spence model.[39]

When $\eta = \rho > 1$ the indirect utility is:

$$V = \frac{\left[\int_{\omega\in\Omega} (\alpha p(\omega) - a)^{\frac{1}{1-\eta}} d\omega\right]^{1-\eta}}{1-\eta} + E,$$

providing a specification that leads to full efficiency of the market equilibrium (see the main text). Then, the expression (76) becomes the constraint:

$$\frac{\alpha}{\eta - 1} = \left[\int_{\omega\in\Omega} q(\omega)^{\frac{1}{\eta}} d\omega\right]^{\eta}, \tag{79}$$

and the arbitrary quantity $\bar{A}$ can be normalized to unit. Accordingly, the direct utility is given (up to a constant) by:

$$U = \frac{a}{\alpha} Q + Y$$

when the constraint (79) is satisfied, and $U = -\infty$ otherwise (intuitively, there is no substitutability between the *numéraire* and the CES aggregator).

2) When $H(A) = \frac{A^{1-\rho}}{1-\rho}$ and $v(p) = (a - \alpha p)^{\frac{1}{1-\eta}}$ for $a > \alpha p > 0$ and $0 < \rho \leqslant \eta < 1$, we can similarly express (75) and (76) as:

$$p(\omega) = \frac{a}{\alpha} - \left(\frac{1-\eta}{\alpha}\right)^{\frac{1-\eta}{\eta}} \frac{\bar{A}^{\frac{\rho(1-\eta)}{\eta}}}{\alpha} q(\omega)^{\frac{1-\eta}{\eta}},$$

and

$$\bar{A} = \left(\frac{1-\eta}{\alpha}\right)^{\frac{1}{\eta}} \bar{A}^{\frac{\rho}{\eta}} \int_{\omega\in\Omega} q(\omega)^{\frac{1}{\eta}} d\omega.$$

For $\eta > \rho$, the last expression defines $\bar{A}$ and can be solved explicitly as follows:

$$\bar{A} \equiv \left(\frac{1-\eta}{\alpha}\right)^{\frac{1}{\eta-\rho}} \left[\int_{\omega\in\Omega} q(\omega)^{\frac{1}{\eta}} d\omega\right]^{\frac{\eta}{\eta-\rho}}$$

Then, the direct utility can be derived as:

$$U = \frac{a}{\alpha} Q - \frac{(\eta-\rho)\bar{A}^{1-\rho}}{(1-\rho)(1-\eta)} + Y.$$

[39] This implies $u(q(\omega)) = q(\omega)^{\frac{1}{\eta}}$ and $\Gamma(B) = \frac{\eta-\rho}{(\eta-1)(1-\rho)} B^{\eta\frac{1-\rho}{\eta-\rho}}$ according to the notation of (51) in Appendix A.

Again the utility is a concave function of quantities (its second term is a decreasing concave transformation of a convex function), and depends on two distinct, additive aggregators.

When $\eta = \rho \in (0,1)$ the indirect utility is:

$$V = \frac{\left[\int_{\omega\in\Omega} (a - \alpha p(\omega))^{\frac{1}{1-\eta}} \, d\omega\right]^{1-\eta}}{1-\eta} + E,$$

and again it leads to full efficiency of the market equilibrium (see the main text). Then, from (76) we obtain the constraint:

$$\frac{\alpha}{1-\eta} = \left[\int_{\omega\in\Omega} q(\omega)^{\frac{1}{\eta}} \, d\omega\right]^{\eta}, \tag{80}$$

and the direct utility (up to a constant) becomes:

$$U = \frac{a}{\alpha} Q + Y$$

when the constraint (80) is satisfied, and $U = -\infty$ otherwise.

3) When $H(A) = \log A$ and $v(p) = (\alpha p - a)^{\frac{1}{1-\eta}}$ for $\alpha p > a \geqslant 0$ and $\eta > 1$, we obtain $p(\omega)$ and A as given respectively by (77) and (78) with $\rho = 1$, and the direct utility can be computed as:

$$U = \frac{\eta}{\eta - 1} \log\left\{\int_{\omega\in\Omega} q(\omega)^{\frac{1}{\eta}} \, d\omega\right\} + \frac{a}{\alpha} Q + Y.$$

Only in the case of a power surplus function ($\eta = \frac{\varepsilon}{\varepsilon-1}$, $a = 0$) the utility has a single aggregator and it is nested within the model of Spence (1976).[40]

4) When $H(A) = \frac{A^{1-\rho}}{1-\rho}$ and $v(p) = e^{-\alpha p}$, where $\alpha > 0$ and $0 < \rho < \eta = 1$, we can express (75) and (76) as:

$$p(\omega) = -\frac{1}{\alpha} \ln \frac{q(\omega)}{\alpha \bar{A}^{-\rho}},$$

and:

$$\bar{A} = \frac{1}{\alpha \bar{A}^{-\rho}} \int_{\omega\in\Omega} q(\omega) \, d\omega.$$

The last expression can be solved explicitly for the aggregator as follows:

$$\bar{A} \equiv \alpha^{\frac{1}{\rho-1}} \left[\int_{\omega\in\Omega} q(\omega) d\omega\right]^{\frac{1}{1-\rho}}.$$

[40] This implies $u(q(\omega)) = q(\omega)^{\frac{\varepsilon-1}{\varepsilon}}$ and $\Gamma(B) = \varepsilon \ln B$ according to the notation of (51) in Appendix A.

Then, the direct utility can be derived as:

$$U = -\frac{1}{\alpha}\int_{\omega\in\Omega} q(\omega)\ln q(\omega)\,d\omega - \frac{\rho Q\ln Q}{\alpha(1-\rho)} + Y,$$

which is a concave function of quantities (being the addition of two decreasing concave transformation of two convex functions), and depends on the two distinct additive aggregators, namely total quantity $Q = \int_{\omega\in\Omega} q(\omega)\,d\omega$ and the "entropy index" $\int_{\omega\in\Omega} q(\omega)\ln q(\omega)\,d\omega$.

The LSE case where $\eta = \rho = 1$ has been already explored by Anderson, de Palma and Thisse (1992, Ch. 3). In our notation, the direct utility dual to (8) is given by:

$$U = -\frac{1}{\alpha}\int_{\omega\in\Omega} q(\omega)\log\left\{\frac{q(\omega)}{\alpha}\right\}d\omega + Y$$

when $Q = \alpha$, and $U = -\infty$ otherwise.

C: Proofs

C.1: Proof of Proposition 3

We consider the first best analysis under IIA preferences and homogeneous firms. In general, Marshallian welfare can be written as:

$$W^* = H(A^*)L + n^*\left[(p^*-c)\left|v'(p^*)\right|H'(A^*)L - F\right] + EL,$$

where $A^* = n^* v(p^*)$. This is maximized by marginal cost pricing $p^* = c$ and then becomes:

$$W^* = H(n^* v(c))L - n^* F + EL.$$

Maximizing the latter with respect to n^*, we immediately obtain a consumption level q^* and a number of firms n^* satisfying:

$$q^* = \frac{\zeta(c)}{c}\frac{F}{L} \quad \text{and} \quad n^* = \frac{H'(A^*)A^*L}{F}.$$

We now prove that, under CRESS preferences, monopolistic competition generates efficient firm sizes, and (weakly) insufficient entry. The fact that the equilibrium consumption is equal to the efficient one for the available goods can be directly verified by using the surplus function (10) to compute:

$$\zeta(c) = \varepsilon(p) - 1 = \frac{\alpha c}{(1-\eta)(a-\alpha c)},$$

implying $q^* = q^e$.

To prove that under CRESS preferences, with a general transformation $H(A)$, entry is (weakly) insufficient, we compute the equilibrium surplus as a constant fraction of the efficient one:

$$v(p(c)) = \eta^{\frac{1}{1-\eta}} v(c), \quad \text{where } \eta^{\frac{1}{1-\eta}} \in (0,1)$$

Given this, the equilibrium number of firms (14) and the efficient number of firms (18) satisfy:

$$H'\left(n^{*}v\left(c\right)\right)=\frac{F}{Lv\left(c\right)}=\Phi H'\left(\Phi\eta n^{e}v\left(c\right)\right),$$

where $\Phi\equiv\eta^{\frac{\eta}{1-\eta}}\in(0,1)$. Thus, a sufficient condition for (weakly) insufficient entry is $H'\left(nv\left(c\right)\right)\geqslant\Phi H'(\Phi\eta nv\left(c\right))$ for any $n>0$ and $v\left(c\right)>0$. Now, let us define the auxiliary function $t\left(x\right)=x^{\eta}H'\left(xnv\left(c\right)\right)>0$ for $x>0$, which satisfies $t(\eta^{\frac{1}{1-\eta}})=\Phi H'(\Phi\eta nv\left(c\right))$ and $t\left(1\right)=H'\left(nv\left(c\right)\right)$. Since its derivative is:

$$\begin{aligned}t'\left(x\right) &= \eta\frac{t\left(x\right)}{x}+x^{\eta}nv\left(c\right)H''\left(xnv\left(c\right)\right)\\ &= \frac{t\left(x\right)}{x}\left[\eta-\rho\left(xnv\left(c\right)\right)\right]\geqslant 0\end{aligned}$$

under our assumption (13), the previous sufficient condition always holds. This completes the proof.

C.2: Heterogenous firms: the second best

In this section we derive the conditions (36) and (37) for the second best with heterogeneous firms. By assuming $\overline{\eta}\left(\widetilde{c}\right)>\rho(\widetilde{A})$, where $\widetilde{A}=\widetilde{N}\int_{0}^{\widetilde{c}}v(p(c))dG(c)$, and using (32) and (35) we can write $\frac{\partial\widetilde{W}}{\partial\widetilde{c}}=\frac{\partial\widetilde{W}}{\partial N}=0$ as:

$$\begin{aligned}H'(\widetilde{A})L\frac{\partial A}{\partial\widetilde{c}}+\widetilde{N}\pi(\widetilde{c})g\left(\widetilde{c}\right) &= \frac{\rho(\widetilde{A})}{\overline{\eta}\left(\widetilde{c}\right)}H'(\widetilde{A})L\frac{\partial A}{\partial\widetilde{c}},\\ H'(\widetilde{A})L\frac{\partial A}{\partial N}+\left[\int_{0}^{\widetilde{c}}\pi(c)dG(c)-F\right] &= \frac{\rho(\widetilde{A})}{\overline{\eta}\left(\widetilde{c}\right)}H'(\widetilde{A})L\frac{\partial A}{\partial N},\end{aligned}$$

or:

$$\begin{aligned}\left[1-\frac{\rho(\widetilde{A})}{\overline{\eta}\left(\widetilde{c}\right)}\right]H'(\widetilde{A})L\frac{\partial A}{\partial\widetilde{c}} &= -\widetilde{N}\pi(\widetilde{c})g\left(\widetilde{c}\right),\\ \left[1-\frac{\rho(\widetilde{A})}{\overline{\eta}\left(\widetilde{c}\right)}\right]H'(\widetilde{A})L\frac{\partial A}{\partial N} &= -\left[\int_{0}^{\widetilde{c}}\pi(c)dG(c)-F\right],\end{aligned}$$

which imply $\pi(\widetilde{c})<0$ and $\int_{0}^{\widetilde{c}}\pi(c)dG(c)-F<0$. The fact that the second best involves negative expected profits implies that consumer welfare has to increase with respect to the equilibrium, namely $\widetilde{A}>A^{e}$. This, in turn implies that the equilibrium quantity of each good $q(c)=\left|v'(p(c))\right|H'(A^{e})$ is above the second best level $q(c)=\left|v'(p(c))\right|H'(\widetilde{A})$ for any active firm.

The last equation can be rewritten, by using (20), (32) and (35), as follows:

$$\left[1-\frac{\rho(\widetilde{A})}{\overline{\eta}\left(\widetilde{c}\right)}\right]H'(\widetilde{A})L\frac{\widetilde{A}}{\widetilde{N}}=-\left[H'(\widetilde{A})L\int_{0}^{\widetilde{c}}\frac{v(p\left(c\right))}{\eta(p\left(c\right))}dG(c)-G\left(\widetilde{c}\right)f-F\right],$$

We can rearrange this equation as:

$$H'(\widetilde{A})L\frac{\widetilde{A}}{\widetilde{N}}\left\{1-\frac{\rho(\widetilde{A})}{\overline{\eta}\left(\tilde{c}\right)}+\int_0^{\widetilde{c}}\frac{1}{\eta(p\left(c\right))}\frac{v(p\left(c\right))}{\int_0^{\widetilde{c}}v(p\left(c\right))dG(c)}dG(c)\right\}=G\left(\widetilde{c}\right)f+F,$$

or

$$H'(\widetilde{A})L\frac{\widetilde{A}}{\widetilde{N}}\left\{\frac{\overline{\eta}\left(\hat{c}\right)-\rho(\widetilde{A})+1}{\overline{\eta}\left(\hat{c}\right)}\right\}=G\left(\widetilde{c}\right)f+F,$$

and eventually as in (37).

Dividing the two previous optimality conditions we get:

$$\frac{\int_0^{\widetilde{c}}v(p\left(c\right))dG(c)}{v\left(p\left(\widetilde{c}\right)\right)}=\frac{\int_0^{\widetilde{c}}\pi(c)dG(c)-F}{\pi(\widetilde{c})},$$

or:

$$\frac{\int_0^{\widetilde{c}}v(p\left(c\right))dG(c)}{v\left(p\left(\widetilde{c}\right)\right)}\left[H'(\widetilde{A})L\frac{v\left(p\left(\widetilde{c}\right)\right)}{\eta\left(p\left(\widetilde{c}\right)\right)}-f\right]=H'(\widetilde{A})L\int_0^{\widetilde{c}}\frac{v\left(p\left(c\right)\right)}{\eta\left(p\left(c\right)\right)}dG(c)-G\left(\widetilde{c}\right)f-F,$$

or:

$$\int_0^{\widetilde{c}}\frac{v(p\left(c\right))}{v\left(p\left(\widetilde{c}\right)\right)}dG(c)-\frac{H'(\widetilde{A})L}{\widetilde{N}F}\left[\frac{1}{\overline{\eta}\left(\tilde{c}\right)}-\frac{1}{\eta(p(\widetilde{c}))}\right]=\frac{F}{f}+G\left(\widetilde{c}\right),$$

which can be written as in (36) after using (37).

C.3: Heterogenous firms: the first best

In this section we consider the first best analysis with heterogeneous firms, in which the social planner maximizes Marshallian welfare choosing the measure N^* of entrant firms, the threshold for active firms $\hat{c}^*$ and the price schedule $p^*(c)$ under a resource constraint.[41] The first-best allocation determines the price aggregator:

$$A^*=N^*\int_0^{\hat{c}^*}v(p^*(c))dG(c). \tag{81}$$

Assuming that expenditure E is large enough to allow us to ignore the resource constraint (so that the consumption of the *numéraire* is positive), the planner's problem can be stated as:

$$\max_{p^*(c),\ \hat{c}^*,\ N^*}\ W^*=H\left(A^*\right)L+EL$$

$$+N^*\left[H'(A^*)L\int_0^{\hat{c}^*}\left(p^*(c)-c\right)\left|v'(p^*(c))\right|dG(c)-G(\hat{c}^*)f-F\right].$$

[41] The resource constraint is:

$$N^*\left[\int_0^{\hat{c}^*}c\left|v'(p^*(c))\right|H'(A^*)LdG(c)+F+G(\hat{c}^*)f\right]\leq EL,$$

and it is satisfied if E is large enough.

Point-wise maximization of W^* with respect to $p^*(c)$ shows that, not surprisingly, marginal cost pricing:

$$p^*(c) = c \quad \text{for } c \in [0, \hat{c}^*]$$

delivers the first best. Accordingly, the previous program can be rewritten as:

$$\max_{N^*,\hat{c}^*} W^* = H\left(N^* \int_0^{\hat{c}^*} v(c)dG(c)\right) L + EL - N^* \left[G(\hat{c}^*)f + F\right].$$

The FOCs then give (assuming $\hat{c}^* < \bar{c}$):

$$v(\hat{c}^*)H'(A^*)L = f, \tag{82}$$

$$H'(A^*)L \int_0^{\hat{c}^*} v(c)dG(c) = G(\hat{c}^*)f + F. \tag{83}$$

Condition (82) says that the contribution to consumer surplus of the optimal cutoff firm is equal to its fixed cost of activation, while condition (83) shows that the expected contribution of the marginal entry is equal to its expected fixed costs. Together, for $f > 0$, they imply that the efficient cutoff is determined by:

$$\int_0^{\hat{c}^*} \frac{v(c)}{v(\hat{c}^*)} dG(c) = \frac{F}{f} + G(\hat{c}^*).$$

The efficient mass of firms N^* increases with respect to L and decreases with respect to f according to:

$$N^* = \frac{H'(A^*)A^*L}{F + G(\hat{c}^*)f}.$$

From (82) we obtain that the efficient individual consumption of a variety is given by (40) in the text. Notice that these results parallel those of the equilibrium analysis: in particular, a similar "dichotomy" arises, according to which market size and the transformation embedded into preferences do affect neither $\hat{c}^*$ (efficient firm selection) nor $q^*(c)\,L$ (efficient firm size).

Welfare comparison We finally compare the equilibrium values to the efficient ones but, not surprisingly, this comparison is in general made difficult by the rather different pricing, as it is also the case in models with exogenous resources allocated to the differentiated goods (see Dhingra and Morrow, 2019). We can illustrate this difficulty by referring to the simple case of the logarithmic transformation $H(A) = \log A$, for which we immediately get:

$$N^* = \frac{L}{F + G(\hat{c}^*)f} \quad \text{and} \quad A^* = v(\hat{c}^*)\frac{L}{f}. \tag{84}$$

Since in this case preference convexity requires $\eta(p) \geq 1$, a comparison with (27) may suggest at first sight that the equilibrium would deliver an insufficient measure of entrant firms and a smaller consumer surplus, as it actually does in the case of homogeneous firms. However, to ensure that these results always hold (whatever L, F, f and G), we also need that $\hat{c}^* \leq \hat{c}^e$, namely, that the efficient firm selection is not looser than the equilibrium one. It turns out that this is not necessarily the case, and accordingly that the equilibrium and optimal cutoffs are different and either one could be larger.

In fact, comparing the integrand on the left hand side of equilibrium and optimality conditions for the cutoffs shows that the market chooses $\hat{c}$ by considering $\widetilde{v}(c)/\widetilde{v}(\widehat{c})$, where $\widetilde{v}(c) = \frac{v(p(c))}{\eta(p(c))}$ is a sort of "virtual social surplus". Thus the market equilibrium distorts the surplus $v(p)$ by evaluating it at $p(c) > c$ and dividing it by $\eta(p(c))$. The relative pattern of $\widetilde{v}(c)/\widetilde{v}(\widehat{c})$ versus $v(c)/v(\widehat{c})$ is far from obvious, since it depends on the curvature features of $v(p)$. However, by defining $\Phi(c) = \frac{h(c)}{\eta(p(c))}$, where $h(c) = \frac{v(p(c))}{v(c)} < 1$, one can prove the following result:

A sufficient condition for $\hat{c}^* < (>)\ \hat{c}^e$ is that $\Phi(c)$ is monotonic increasing (decreasing). Accordingly, with a decreasing (increasing) $\eta(p)$ a sufficient condition for $\hat{c}^* < (>)\ \hat{c}^e$ is that $h(c)$ is increasing (decreasing).

The intuition for these results is that when $\Phi(c)$ is monotonic the social surplus $v(c)$ and the virtual social surplus $\widetilde{v}(c)$ provided by the variety produced with marginal cost c diverge (as a function of the marginal cost) in a predictable way, leading market selection to be tighter (looser) than socially efficient. However, the sufficient conditions above are rather involved: while in general $\varepsilon'(p)$ and $\zeta'(p)$ need not to agree in sign, they tend to agree when $\varepsilon(p)$ is monotonic,[42] and in such a case *a fortiori* we cannot predict the sign of $\eta'(p)$.[43] Moreover, $h'(c) > (<)\, 0$ is equivalent to $\zeta(p(c))\frac{d\ln p(c)}{d\ln c} < (>)\, \zeta(c)$, and we have $d\ln p(c)/d\ln c \lesseqgtr 1$ if $\varepsilon'(p) \gtreqless 0$. Accordingly, in general the equilibrium and efficient cutoffs do not coincide. Notice in particular that it might well be the case that $\hat{c}^* > \hat{c}^e$, implying that a planner would like to activate firms which cannot survive at the market equilibrium.

A simple case arises if the surplus function belongs to the CRESS family (10). In such a case we have that also $h(c)$ is constant and thus $\Phi(c)$ is constant too (and smaller than 1).[44] Accordingly, the virtual social surplus is proportional to the actual social surplus, and as a result the equilibrium cutoff is equal to the efficient one. In addition, a comparison of (28) and (40) reveals that also the

[42] One can prove that when $|v'(p)|$ is so-called (see Mrázová and Neary, 2019) *super*(*sub*)-convex, meaning $\frac{d^2\ln|v'(p)|}{d(\ln p)^2} = \frac{d\{-\varepsilon(p)\}}{d\ln p} > (<)\, 0$, then under some technical conditions also $v(p)$ is *super*(*sub*)-convex, meaning $\frac{d^2\ln v(p)}{d(\ln p)^2} = \frac{d\{-\zeta(p)\}}{d\ln p} > (<)\, 0$.

[43] But notice that $\eta'(p) \geq 0$ is equivalent to $\phi(p) + \zeta(p) \leq 2\varepsilon(p)$, which would be satisfied under log-convexity of v if demand were locally concave (i.e., if $\phi(p) \leq 0$).

[44] Computation shows that in this case $h(c) = \eta^{\frac{1}{1-\eta}} = \Phi^{\frac{1}{\eta}}$.

equilibrium individual consumption $q^e(c)$ is at its first-best level $q^*(c)$.[45] Since in this case $\hat{c}^* = \hat{c}^e$, applying to the comparison of the first term in (84) with (27) the same argument used in Appendix C.1 for the case of homogeneous firms shows that there is in general (weakly) insufficient entry, namely $N^* \geqslant N^e$. However, computation shows that when the isoelastic transformation (11) is combined with a CRESS surplus function (10) in such a way that $\rho = \eta$ then a fully efficient allocation arises (as in the LSE case).

Finally, (82) can be rewritten as:

$$A^* = H'^{-1}\left(\frac{f}{v(\hat{c}^*)L}\right),$$

so that a comparison with (22) shows that $A^* > A^e$ is equivalent to $\frac{v(p(\widehat{c}^e))}{\eta(p(\widehat{c}^e))} < v(\hat{c}^*)$, which is certainly satisfied if $\hat{c}^* \leq \widehat{c}^e$ and v is log-convex. Notice that $A^* > A^e$ is equivalent to $\Phi < 1$ for the CRESS family of surpluses, and thus it is always satisfied in those cases.

[45] It can be verified that the relevant condition:

$$\frac{v'(p(c))}{v'(c)} = \Phi(\widehat{c}^e)\frac{v(\widehat{c}^e)}{v(\hat{c}^*)}$$

also holds.