

Efficiency in Large Markets with Firm Heterogeneity

Swati Dhingra

CEP, London School of Economics & CEPR

John Morrow

Birkbeck University of London, CEP & CEPR

This Draft: Sep 30, 2017

Abstract

Empirical work has drawn attention to the high degree of productivity differences within industries, and its role in resource allocation. In a benchmark monopolistically competitive economy, productivity differences introduce two new margins for allocational inefficiency. When markups vary across firms, laissez faire markets do not select the right distribution of firms and the market-determined quantities are inefficient. We show that these considerations determine when increased competition from market expansion takes the economy closer to the socially efficient allocation of resources. As market size grows large, differences in market power across firms converge and the market allocation approaches the efficient allocation of an economy with constant markups.

JEL Codes: F1, L1, D6.

Keywords: Efficiency, Productivity, Limit theorem, Market expansion, Competition.

Acknowledgments. We thank Steve Redding, Bob Staiger and Jacques Thisse for encouragement and comments. Contact: s.dhingra@lse.ac.uk; j.morrow@bbk.ac.uk.

1 Introduction

Empirical work has drawn attention to the high degree of heterogeneity in firm productivity, and the constant reallocation of resources across different firms.¹ The focus on productivity differences has provided new insights into market outcomes such as industrial productivity, firm pricing and welfare gains from policy changes.² When firms differ in productivity, the distribution of resources across firms also affects the allocational efficiency of markets.

Symmetric firm models explain when resource allocation is efficient by examining the trade-off between quantity and product variety in imperfectly competitive markets.³ When firms differ in productivity, we must also ask which types of firms should produce and which should be shut down. Firm differences in productivity introduce two new margins of potential inefficiency: selection of the right distribution of firms and allocation of the right quantities across firms. Dhingra and Morrow (2012) show that differences in market power across firms lead to new trade-offs between variety, quantity and productivity. These considerations impact optimal policy rules in a fundamental way, distinct from markets with symmetric costs.

Dhingra and Morrow (2012) examine resource allocation in the standard setting of a monopolistically competitive industry with heterogeneous firm productivity and free entry (e.g. Melitz 2003)⁴. They generalize the demand structure to the symmetric directly additive (SDA)⁵ form of Dixit and Stiglitz (1977) which generates a variable elasticity of substitution, and provides a rich setting for a wide range of market outcomes (Vives 2001; Zhelobodko, Kokovin, Parenti and Thisse 2012). For example, Ushchev (2017) shows the impact of market expansion on product ranges of firms can be modelled flexibly with SDA preferences, while Dhingra and Morrow (2012) and Bertolotti et al. (2017) show that increasing markups could result from market expansion when the elasticity of marginal utility falls with quantity. When elasticities vary with quantity, firms differ in market power and market allocations reflect the distortions of imperfect competition. Under SDA, the market maximizes real revenues, which implies that the private benefits to firms are perfectly aligned with social benefits only under CES demand. More generally, the market allocation is inefficient because of lack of full appropriability of consumer surplus and the business stealing externality that firms impose on each other.

Trade creates larger, more competitive markets, which could reduce the distortions associ-

¹For instance, Görg et al. (2017) show that firm-size distribution affects the relationship between an industry's employment and output.

²Example, Pavcnik (2002); Asplund and Nocke (2006); Foster et al. (2001); Melitz and Redding (2012).

³Example, Spence (1976); Venables (1985); Mankiw and Whinston (1986); Stiglitz (1986).

⁴The first instance of analysis of a CES model of monopolistic competition with heterogeneous firms was in Montagna (1995) in a static, partial equilibrium setting

⁵Labelled as such given the direct utility is additive and the sub utility is the same for all goods.

ated with imperfect competition and provide welfare gains (Krugman 1987). In the presence of productivity differences across firms, misallocation varies with firm productivity, which may be difficult to elicit, especially if how policy treats a firm varies with its observed type. One potential policy option that does not require firm-level information is international integration. The idea of introducing foreign competition to improve efficiency goes back at least to Melvin and Warne (1973). Helpman and Krugman (1985) provide sufficient conditions for welfare gains from trade in symmetric firm trade models. They show when productivity and variety do not decline after integration, then there are gains from trade.⁶ Dhingra and Morrow (2012) show that market integration always provides welfare gains when private and social incentives are aligned, which is characterized by the demand elasticity and the elasticity of utility in their setting. This is true for any cost distribution, but requires a regularity condition on preferences when private markups are decreasing. An increase in market size increases competition and reduces per capita demand for each variety. When preferences are aligned, market expansion alters the private and socially relevant demand elasticities for quantity choice in the same direction. The market therefore incentivizes firms towards the right allocation and provides higher welfare. Building on this result, Bykadorov et al. (2015) show that aligned preferences are necessary and sufficient for welfare gains from trade under symmetric firms and variable marginal costs.

While integration can increase welfare, a more ambitious question is: can we ever expect trade to eliminate the distortions of imperfect competition? Following Stiglitz (1986), this paper examines market and optimal outcomes as market size becomes arbitrarily large. Small markets may have insufficient competition, so looking at large markets allows us to understand where market expansion is headed and when international trade or growth enables markets to eventually mitigate distortions.

As a benchmark for understanding efficiency gains, we follow the literature on imperfect competition in large markets and examine whether integration with large global markets leads to allocational efficiency (Vives 2001, Chapter 6). Integration with large markets will push outcomes towards a new concept, the “CES limit”, where firms converge to charging constant markups. Unlike a perfectly competitive limit (Hart 1985), productivity dispersion and market power persist in the CES limit. Yet the market is efficient and integration with large global markets is therefore a first-best policy to eliminate the distortions of imperfect competition. However, as the limit may require a market size which is unattainable even in fully integrated world markets, integration may be an incomplete tool to reduce distortions.

⁶Specifically, let w denote the wage and $C(w, q) = w(c + f/q)$ denote the average unit cost function for producing q units of variety c . When firms are symmetric in c , trade is beneficial as long as variety does not fall ($M_e \geq M_e^{\text{aut}}$) and average unit cost of the autarky bundle is lower ($C(w, q) \cdot q^{\text{aut}} \leq C(w, q^{\text{aut}}) \cdot q^{\text{aut}}$).

Our results for large markets are related to an earlier literature that examines the pricing strategies of symmetric oligopolistic firms as the number of entrants gets large. Previous work shows oligopolistic firms continue to have market power as the number of entrants gets large when the residual demand elasticity is bounded (Perloff and Salop 1985, Vives 2001 Chapter 6). We instead examine the limit as market size gets large (because entry is not a primitive in the monopolistically competitive model), and find that market power and productivity differences persist in large markets. Such persistence of imperfect competition is consistent with the observation of Samuelson (1967) that “the limit may be at an irreducible positive degree of imperfection” (Khan and Sun 2002). It is remarkable that the large market outcome, which exhibits productivity differences and remains imperfectly competitive, is socially optimal. This shows markups and productivity differences are consistent with large markets and efficiency.

While SDA preferences generate a variable elasticity of substitution, they differ from other symmetric preferences that feature variable elasticities of substitution (VES). For instance, the literature on monopolistic competition generates variable elasticities through a number of different preference forms, such as quasilinear and quadratic (Melitz and Ottaviano 2008; Nocco et al. 2013), indirectly additive (Bertoletti and Etro 2016) and homothetic (Benassy 1996; Feenstra 2003). The optimality results generated in other VES, but non-SDA preferences, heterogeneous firm models differ. For example, Bertoletti et al. (forthcoming) demonstrate for the full class of indirectly additive preferences (which includes CES preferences), market expansion is neutral on prices, production and the equilibrium cutoff for active firms, thus replicating the CES results always, and not just at the limit. However except under CES preferences, the equilibrium is inefficient in terms of pricing, production and selection. Similarly, Nocco et al. (2017) find market allocations are inefficient in their setting: the market cutoff productivity is too low and low cost firms under-produce while high cost firm over-produce in the market.⁷ Furthermore, firm heterogeneity makes entry distortions dependent on the cost distribution. Finally, Arkolakis et al. (forthcoming) consider a demand system encompassing additively separable preferences with a bounded choke price. While they don’t consider optimality, they discuss the gains from trade liberalization as in Arkolakis et al. (2012). SDA preferences differ from these preferences, as they do not have a bounded choke price, and large markets cause heterogeneity across firms to collapse in this case.

The paper is organized as follows. Section 2 recaps the standard monopolistic competition framework with firm heterogeneity. Section 3 examines welfare gains from integration, deriving a limit result for large markets. Section 4 concludes.

⁷Nocco et al. (2013) show that the mass of firms cannot be unambiguously ranked even when the model is parameterized with a linear demand and Pareto cost distribution.

2 Model

We adopt the SDA demand structure of Dixit and Stiglitz within the heterogeneous firm framework of Melitz. Monopolistic competition models with heterogeneous firms differ from earlier models with product differentiation in two significant ways. First, costs of production are unknown to firms before sunk costs of entry are incurred. Second, firms are asymmetric in their costs of production, leading to firm selection based on productivity. This Section lays out the model and recaps the implications of asymmetric costs for consumers, firms and equilibrium outcomes.

2.1 Consumers

We explain the SDA demand structure and then discuss consumer demand. The exposition for consumer demand closely follows Zhelobodko et al. (2012) which works with a similar setting and builds on work by Vives (2001).

An economy consists of a mass L of identical workers, each endowed with one unit of labor and facing a wage rate w normalized to one. Workers have identical preferences for a differentiated good. The differentiated good is made available as a continuum N of horizontally differentiated varieties indexed by $i \in [0, N]$. Given prices p_i for the varieties, every worker chooses quantity q_i for each of the varieties to maximize her utility subject to her budget constraint. Preferences over differentiated goods take the general SDA form:

$$U(\mathbf{q}) \equiv \int_0^N u(q_i) di \quad (1)$$

where $u(\cdot)$ is thrice continuously differentiable, strictly increasing and strictly concave on $(0, \infty)$, and $u(0)$ is normalized to zero. The concavity of u ensures consumers love variety and prefer to spread their consumption over all available varieties. Here $u(q_i)$ denotes utility from an individual variety i . Under CES preferences, $u(q_i) = q_i^\rho$ as specified in Dixit-Stiglitz and Krugman (1980).⁸

For each variety i , SDA preferences induce an inverse demand $p(q_i) = u'(q_i)/\delta$ where δ is the consumer's budget multiplier. As u is strictly increasing and concave, for any fixed price vector the consumer's maximization problem is concave. The necessary condition which determines the inverse demand is sufficient, and has a solution provided inada conditions on

⁸The specific CES form in Melitz is $U(\mathbf{q}) \equiv (\int (q_i^\rho di))^{1/\rho}$ but the normalization of the exponent $1/\rho$ in Equation (1) will not play a role in allocation decisions.

u .⁹ Multiplying both sides of the inverse demand by q_i and aggregating over all i , the budget multiplier is $\delta = \int_0^N u'(q_i) \cdot q_i di$. The consumer budget multiplier δ will act as a demand shifter and the inverse demand will inherit the properties of the marginal utility $u'(q_i)$. In particular, the inverse demand elasticity $|d \ln p_i / d \ln q_i|$ equals the elasticity of marginal utility $\mu(q_i) \equiv |q_i u''(q_i) / u'(q_i)|$, which enables us to characterize market allocations in terms of demand primitives. Under CES preferences, the elasticity of marginal utility is constant and the inverse demand elasticity does not respond to consumption ($|d \ln p_i / d \ln q_i| = \mu(q_i) = 1 - \rho$). When $\mu'(q_i) > 0$, the inverse demand of a variety becomes more elastic as its consumption increases. The opposite holds for $\mu'(q_i) < 0$, where the demand for a variety becomes less elastic as its price rises. Zhelobodko et al. (2012) show that the elasticity of marginal utility $\mu(q_i)$ can also be interpreted in terms of substitution across varieties. For symmetric consumption levels ($q_i = q$), this elasticity equals the inverse of the elasticity of substitution between any two varieties. For $\mu'(q) > 0$, higher consumption per variety or fewer varieties for a given total quantity, induces a lower elasticity of substitution between varieties. Consumers perceive varieties as being less differentiated when they consume more, but this relationship does not carry over to heterogeneous consumption levels.

The inverse demand elasticity summarizes market demand, and will enable a characterization of market outcomes. A policymaker maximizes utility, and is not concerned with market prices. Therefore, we define the elasticity of utility $\varepsilon(q_i) \equiv u'(q_i) q_i / u(q_i)$, which will enable a characterization of optimal allocations. For symmetric consumption levels, Vives (2001) points out that $1 - \varepsilon(q_i)$ is the degree of preference for variety as it measures the utility gain from adding a variety, holding quantity per firm fixed. To get an analogue of the discrete good case, consider a consumer who ceases to purchase variety 0, or more formally, a basket of varieties $[0, \alpha]$. The consumer loses average utility of $\int_0^\alpha u(q_i) di / \alpha$ per variety and saves an average income of $\int_0^\alpha p_i q_i di / \alpha$ per variety. The savings can be used to increase consumption of all other varieties proportionally by $\int_0^\alpha p_i q_i di / \int_\alpha^N p_i q_i di$, leading to a rise in average utility per variety of

$$\int_\alpha^N u'(q_i) \left[\int_0^\alpha p_i q_i di / \alpha \int_\alpha^N p_i q_i di \right] q_i di = \delta \int_0^\alpha p_i q_i di / \alpha.$$

Letting α approach zero gives us an expression for how $1 - \varepsilon$ measures the net welfare gain of purchasing additional variety: a welfare gain of $u(q_i)$ at a welfare cost of $\delta p_i q_i$ by proportionally consuming less of other varieties. As $p_i = u'(q_i) / \delta$, $1 - \varepsilon(q_i) = (u(q_i) - u'(q_i) q_i) / u(q_i)$

⁹Additional assumptions to guarantee existence and uniqueness of the market equilibrium are in a separate note available online. Utility functions not satisfying inada conditions are permissible but may require parametric restrictions to ensure existence.

denotes the welfare contribution of variety relative to quantity. With these demand-side elasticities in hand, we turn to firms' production and entry decisions.

2.2 Firms

There is a continuum of firms which may enter the market for differentiated goods, by paying a sunk entry cost of $f_e > 0$. The mass of entering firms is denoted by M_e . Firms are monopolistically competitive and each firm produces a single unique variety. A firm faces an inverse demand of $p(q_i) = u'(q_i)/\delta$ for variety i . It acts as a monopolist of its unique variety but takes aggregate demand conditions δ as given. Upon entry, each firm receives a unit cost $c \geq 0$ drawn from a distribution G with continuously differentiable pdf g . Each variety can therefore be indexed by the unit cost c of its producer.

After entry, should a firm produce, it incurs a fixed cost of production $f > 0$. Profit maximization implies firms produce if they can earn non-negative profits net of the fixed costs of production. A firm with cost draw c chooses its quantity $q(c)$ to $\max_{q(c)} [p(q(c)) - c]q(c)L$ and $q(c) > 0$ if $\pi(c) = \max_{q(c)} [p(q(c)) - c]q(c)L - f > 0$. To ensure the firm's quantity FOC is optimal, we assume marginal revenue is strictly decreasing in quantity and the elasticity of marginal utility $\mu(q) = |qu''(q)/u'(q)|$ is less than one. A firm chooses its quantity to equate marginal revenue and marginal cost ($p + q \cdot u''(q)/\delta = c$), and concavity of the firm problem ensures low cost firms supply higher quantities and charge lower prices.

The markup charged by a firm with cost draw c is $(p(c) - c)/p(c) = -q(c)u''(q(c))/u'(q(c))$. This shows that the elasticity of marginal utility $\mu(q)$ summarizes the markup:

$$\mu(q(c)) = |q(c)u''(q(c))/u'(q(c))| = (p(c) - c)/p(c).$$

When $\mu'(q) > 0$, low cost firms supply higher quantities at higher markups.

2.3 Market Equilibrium

Profits fall with unit cost c , and the cutoff cost level of firms that are indifferent between producing and exiting from the market is denoted by c_d . The cutoff cost c_d is fixed by the zero profit condition, $\pi(c_d) = 0$. Firms with cost draws higher than the cutoff level earn negative profits and do not produce. The mass of producing firms in equilibrium is therefore $M = M_e G(c_d)$.

In summary, each firm faces a two stage problem: in the second stage it maximizes profits given a known cost draw, and in the first stage it decides whether to enter given the expected profits in the second stage. To study the Chamberlinian tradeoff between quantity and variety,

we maintain the standard free entry condition imposed in monopolistic competition models. Specifically, ex ante average profit net of sunk entry costs must be zero, $\int_0^{c_d} \pi(c) dG = f_e$. This free entry condition along with the consumer's budget constraint ensures that the resources used by firms equal the total resources in the economy, $L = M_e \left[\int_0^{c_d} (cq(c)L + f) dG + f_e \right]$.

2.4 Social Optimum

To assess the efficiency of resource allocation in the market equilibrium, we now describe the policymaker's optimal allocation. A policymaker maximizes individual welfare U as given in Equation (1) by choosing the mass of entrants, quantities and types of firms that produce.¹⁰ The policymaker can choose any allocation of resources that does not exceed the total resources in the economy. However, she faces the same entry process as for the market: a sunk entry cost f_e must be paid to get a unit cost draw from $G(c)$. Fixed costs of production imply that the policymaker chooses zero quantities for varieties above a cost threshold. Therefore, all optimal allocation decisions can be summarized by quantity $q(c)$, potential variety M_e and a productivity cutoff c_d . The policymaker chooses $q(c)$, c_d and M_e to

$$\max M_e \int_0^{c_d} u(q(c)) dG \text{ where } L \geq M_e \left\{ \int_0^{c_d} [cq(c)L + f] dG + f_e \right\}.$$

Our approach for arriving at the optimal allocation is to think of optimal quantities $q^{\text{opt}}(c)$ as being determined implicitly by c_d and M_e so that per capita welfare can be written as

$$U = M_e \int_0^{c_d} u(q^{\text{opt}}(c)) dG. \quad (2)$$

Optimal quantities ensure marginal utility equals the social marginal cost of a variety, $u'(q^{\text{opt}}(c)) = \lambda c$ where λ is the resource multiplier for fixed c_d and M_e . This is equivalent to the observation that efficiency requires the Marginal Rate of Substitution to equal the Marginal Rate of Transformation across varieties, along with an optimal resource multiplier that embodies the endpoint conditions c_d and M_e which depend in turn on the cost distribution G . After solving for each q^{opt} conditional on c_d and M_e , Equation (2) can be maximized in c_d and M_e . Of course, substantial work is involved in showing sufficiency, but we relegate this to the Appendix. The next two Sections compare the market and optimal allocations in this framework.

¹⁰Free entry implies zero expected profits, so the focus is on consumer welfare.

3 Market Efficiency

Having described an economy consisting of heterogeneous, imperfectly competitive firms, we now examine welfare and efficiency from integration with world markets. The existence of gains from international trade is one of the “most fundamental results” in economics (Costinot and Rodriguez-Clare (2013)). Increases in market size encourage competition, so we might expect that integration would reduce market power and improve welfare. However, in a second-best world, there is no guarantee that expanding the economy’s opportunities, through trade or anything else, necessarily leads to a gain (see Helpman and Krugman (1985), pp. 179). Building on this insight, we examine efficiency in large markets to understand the potential of market expansion in eliminating distortions. We show large integrated markets can eliminate distortions, while preserving firm heterogeneity.

3.1 Integration, Market Size and Efficiency

We begin with the equivalence between market expansion and trade. Proposition 1 shows an economy can increase its market size by opening to trade with foreign markets. The market equilibrium between freely trading countries of sizes L_1, \dots, L_n is identical to the market equilibrium of a single autarkic country of size $L = L_1 + \dots + L_n$, echoing Krugman (1979). This result is summarized as Proposition 1.

Proposition 1. *Free trade between countries of sizes L_1, \dots, L_n has the same market outcome as a unified market of size $L = L_1 + \dots + L_n$.*

Proof. Consider a home country of size L opening to trade with a foreign country of size L^* . Suppose the consumer’s budget multipliers are equal in each country so $\delta = \delta^*$ and that the terms of trade are unity. We will show that the implied allocation can be supported by a set of prices and therefore constitutes a market equilibrium. The implied quantity allocation, productivity level and per capita entry are the same across home and foreign consumers, so opening to trade is equivalent to an increase in market size from L to $L + L^*$.

Let e denote the home terms of trade, so

$$e \equiv M_e^* \int_0^{c_d^*} p_x^* q_x^* L dG / M_e \int_0^{c_d} p_x q_x L^* dG$$

and by assumption $e = e^* = 1$. Then the $MR = MC$ condition implies a home firm chooses $p(c)[1 - \mu(q(c))] = c$ in the home market and $e \cdot p_x(c)[1 - \mu(q_x(c))] = c$ in the foreign market. A foreign firm chooses $e^* \cdot p^*(c)[1 - \mu(q^*(c))] = c$ in the foreign market and $p_x^*(c)[1 -$

$\mu(q_x^*(c)) = c$ in the home market. When $\delta = \delta^*$ and $e = e^* = 1$, quantity allocations and prices are identical, i.e. $q(c) = q_x^*(c) = q^*(c) = q_x(c)$ and $p(c) = p_x^*(c) = p^*(c) = p_x(c)$.

This implies cost cutoffs are also the same across countries. The cost cutoff condition for home firms is $\pi + e\pi_x = (p(c_d) - c_d)q(c_d)L + e(p_x(c_d) - c_d)q_x(c_d)L^* = f$. Substituting for optimal q^* and q_x^* in the analogous foreign cost cutoff condition implies $c_d = c_d^*$. From the resource constraint, this fixes the relationship between entry across countries as $L/M_e = \int_0^{c_d} [cq(c) + cq_x(c) + f]dG + f_e = L^*/M_e^*$. Thus, $\delta = \delta^*$ and $e = e^* = 1$ completely determines the behavior of firms. What remains is to check that $\delta = \delta^*$ and $e = e^* = 1$ is consistent with the consumer's problem and the balance of trade at these prices and quantities consistent with firm behavior.

For the consumer's problem, we require at home that $1 = M_e \int_0^{c_d} pqdG + M_e^* \int_0^{c_d^*} p_x^* q_x^* dG$, which from $L/M_e = L^*/M_e^*$ is equivalent to

$$L/M_e = L \int_0^{c_d} pqdG + L^* \int_0^{c_d^*} p_x^* q_x^* dG = L \int_0^{c_d} pqdG + L/M_e - L \int_0^{c_d} p_x q_x dG.$$

Therefore to show the consumer's problem is consistent, it is sufficient to show expenditure on home goods is equal to expenditure on exported goods ($\int_0^{c_d} pqdG = \int_0^{c_d} p_x q_x dG$), which indeed holds by the above equalities of prices and quantities. To show the balance of trade is consistent, we use the consumer budget constraint which gives

$$e = M_e^* \int_0^{c_d^*} p_x^* q_x^* L dG / M_e \int_0^{c_d} p_x q_x L^* dG = M_e^* L / M_e L^* = 1.$$

Similarly, the implied foreign terms of trade is $e^* = 1$. Thus $\delta = \delta^*$ and $e = e^* = 1$ generate an allocation consistent with monopolistic competition and price system consistent with consumer maximization and free trade. \square

Proposition 1 implies that market distortions persist in integrated markets as they are equivalent to larger domestic markets and resource allocation in an integrated market is suboptimal, except under CES demand (see Dhingra and Morrow (2012)). When markups vary, marginal revenues do not correspond to marginal utilities so market allocations are not aligned with efficient allocations. This is particularly important when considering trade as a policy option, as it implies that opening to trade may take the economy further from the social optimum. For example, market expansion from trade may induce exit of low productivity firms from the market when it is optimal to keep more low productivity firms with the purpose of preserving variety.

3.2 Efficiency in Large Markets

We examine when integrating with large global markets enables a small economy to overcome its market distortions. From a theoretical perspective, we term a large market the limit of the economy as the mass of workers L approaches infinity, and in practice we might expect that sufficiently large markets approximate this limiting case.¹¹

Large markets enable us to understand whether competition can eliminate distortions. For instance, when firms are symmetric, large markets eliminate distortions as *per capita* fixed costs fall to zero. This is because free entry leads to average cost pricing ($p = c + f/qL$), so the per capita fixed costs summarize market power. As market size grows arbitrarily large and per capita fixed costs fall to zero, markups disappear leading to perfect competition and efficient allocations in large markets.

Building on this reasoning, we develop the large market concept in two directions to understand the sources of inefficiency. First, we tie the conditions for efficiency to demand primitives, taking into account endogeneity of allocations. In the homogenous firm example above, this amounts to determining how f/qL changes with market size under different model primitives. Second, we examine whether productivity differences are compatible with large markets. When firms are heterogeneous, simply knowing per capita fixed costs does not explain the distribution of productivity, prices and quantity. At least three salient outcomes can occur. One outcome is that competitive pressures might weed out all firms but the most productive. This occurs for instance when marginal revenue is bounded, as when u is quadratic or CARA (e.g. Behrens and Murata 2012). It may also happen that access to large markets allows even the least productive firms to amortize fixed costs and produce. To retain the fundamental properties of monopolistic competition under productivity differences, we chart out a third possibility between these two extremes: some, but not all, firms produce. To do so, we maintain the previous regularity conditions for a market equilibrium. In order to aid the analysis, we make three assumptions on demand at small quantities. The first assumption enables a clear distinction between the three salient outcomes in large markets.

Assumption (Interior Markups). *The inverse demand elasticity and elasticity of utility are bounded away from 0 and 1 for small quantities. Formally, $\lim_{q \rightarrow 0} \mu(q)$ and $\lim_{q \rightarrow 0} \varepsilon(q) \in (0, 1)$.*

The assumption of interior markups guarantees that as the quantity sold from a firm to a consumer becomes small (as happens for all positive unit cost firms), markups remain positive ($\mu > 0$) and prices remain bounded ($\mu < 1$). It also guarantees that the added utility provided per labor unit at the optimum converges to a non-zero constant (e.g., Solow 1998, Kuhn and

¹¹How large markets need to be to justify this approximation is an open quantitative question.

Vives 1999). An example of a class of utility functions satisfying interior markups is the expopower utility where $u(q) = [1 - \exp(-\alpha q^{1-\rho})]/\alpha$ for $\rho \in (0, 1)$. It nests CES preferences for $\alpha = 0$.

When markups are interior, there is a sharp taxonomy of what may happen to the distribution of costs, prices and total quantities ($Lq(c)$), as shown in Proposition 3 in the Appendix. In words, Proposition 3 shows that when markups are interior and the cost cutoff converges, one of three things must happen. 1) Only the lowest cost firms remain and prices go to zero (akin to perfect competition), while the lowest cost firms produce infinite total quantities. 2) Post-entry, all firms produce independent of cost while prices become unbounded and the total quantities produced become negligible, akin to a “rentier” case where firms produce little after fixed costs are incurred. 3) The cost cutoff converges to a positive finite level, and a non-degenerate distribution of prices and total quantities persists. Although each of these possibilities might be of interest, we focus on the case when the limiting cost draw distribution exhibits heterogeneity ($\lim_{L \rightarrow \infty} c_d^{\text{mkt}} > 0$) but fixed costs still play a role in determining which firms produce ($\lim_{L \rightarrow \infty} c_d^{\text{mkt}} < \infty$). We therefore make the following assumption, which by Proposition 3 will guarantee non-degenerate prices and total quantities:

Assumption (Interior Convergence). *In the large economy, the market and optimal allocations have a non-degenerate cost distribution in which some but not all entrants produce.*

Under interior markups and convergence, the economy converges to a monopolistically competitive limit distinct from the extremes of a perfectly competitive limit or a rentier limit. As the economy grows, each worker consumes a negligible quantity of each variety. At these low levels of quantity, the inverse demand elasticity does not vanish and firms can still extract a positive markup μ . This is in sharp contrast to a competitive limit, in which firms are left with no market power and μ drops to zero. Similarly, the social markup $(1 - \varepsilon)$ does not drop to zero in the monopolistically competitive limit, so each variety contributes at a positive rate to utility even at low levels of quantity. The monopolistically competitive limit is therefore consistent with positive markups which become more uniform with increased market size.

In fact, this monopolistically competitive limit has a sharper characterization very close to the conditions which characterize a finite size market under CES demand (including efficiency). We therefore refer to it as a “CES limit” and introduce one last regularity condition to obtain this result.

Assumption (Market Identification). *Quantity ratios distinguish price ratios for small q :*

$$\text{If } \kappa \neq \tilde{\kappa} \text{ then } \lim_{q \rightarrow 0} p(\kappa q)/p(q) \neq \lim_{q \rightarrow 0} p(\tilde{\kappa} q)/p(q).$$

Market identification guarantees production levels across firms can be distinguished if the firms charge distinct prices as quantities sold become negligible. Combining these three assumptions of interior markups, convergence and identification ensures the large economy goes to the CES limit, summarized as Proposition 2. The intuition for the role of these assumptions follows. As market size grows large, $q \rightarrow 0$ so under Interior Markups, $(p - c)/p = \mu(q) \rightarrow \mu(0)$ and, finite but non-zero markups can persist in the large economy. Since profits are $\mu(q)/(1 - \mu(q)) \cdot Lcq$, whether a particular firm survives in the large economy depends on how variable costs Lcq evolve with market size. Clearly, if variable costs diverge to zero for a firm with cost c , that firm must eventually exit, while if variable costs diverge to infinity, the firm must eventually enter. To arrive at the CES limit, necessarily variable costs must converge to a positive level, which requires convergence of the total quantity sold, Lq . However, since firms are embedded in a heterogeneous environment where aggregate conditions impact firm behavior, the pointwise convergence of markups $\{\mu(q(c))\}$ is not sufficient to guarantee that total quantities $\{Lq(c)\}$ are well behaved in aggregate. What is sufficient is that prices $\{p(c)\}$ can distinguish firms as market size grows large, thus the Market Identification condition.¹²

Proposition 2. *Under the above assumptions, as market size approaches infinity, outcomes approach the CES limit. This limit has the following characteristics:*

1. *Prices, markups and expected profits converge to positive constants.*
2. *Per capita quantities $q(c)$ go to zero, while aggregate quantities $Lq(c)$ converge.*
3. *Relative quantities $Lq(c)/Lq(c_d)$ converge to $(c/c_d)^{-1/\alpha}$ with $\alpha = \lim_{q \rightarrow 0} \mu(q)$.*
4. *The entrant per worker ratio M_e/L converges.*
5. *The market and socially optimal allocations coincide.*

Proposition 2 shows that integration with large markets can push economies based on variable elasticity demand to the CES limit. In this limit, the inverse demand elasticity and the elasticity of utility become constant, ensuring the market outcome is socially optimal. Firms charge constant markups which exactly cross-subsidize entry of low productivity firms to preserve variety. This wipes out the distortions of imperfect competition as the economy becomes large. While dealing with the assumptions of the market equilibrium is somewhat delicate (see Appendix), we can explain Proposition 2 intuitively in terms of our previous result that CES preferences induce efficiency. In large markets, the quantity $q(c)$ sold to any individual consumer goes to zero, so markups $\mu(q(c))$ converge to the same constant independent of c .¹³

¹²From a technical standpoint, this guarantees entry is well behaved, avoiding pathological sequences of potential equilibria as market size grows large.

¹³The rate at which markups converge depends on c and is in any case endogenous (see Appendix).

This convergence to constant markups aligns perfectly with those generated by CES preferences with an exponent equal to $1 - \lim_{q \rightarrow 0} \mu(q)$. Thus, large markets reduce distortions until market allocations are perfectly aligned with socially optimal objectives.

The CES limit is optimal despite the presence of imperfect competition in the limit, but it is an open empirical question whether markets are sufficiently large for this to be a reasonable approximation to use in lieu of richer variable elasticity demand. When integrated markets are small, variable markups are crucial in understanding distortions and additional gains can be reaped by using domestic policy in conjunction with trade policy.

4 Conclusion

This paper examines the efficiency of market allocations when firms vary in productivity and markups. Considering the Spence-Dixit-Stiglitz framework, the efficiency of CES demand is valid even with productivity differences across firms. This is because market outcomes maximize revenue, and under CES demand, private and social incentives are perfectly aligned.

When firms differ in market power, the market outcomes of quantity, variety and productivity are not socially optimal due to imperfect competition among firms. We examine whether international integration can be a policy tool to mitigate these distortions through increased competition. Market expansion does not guarantee welfare gains under imperfect competition. But we find that integrating with large markets holds out the possibility of approaching the CES limit, which induces constant markups and therefore an efficient outcome.

Even though integration can cause market and social objectives to perfectly align, “How Large is Large?” is an open question. Further work quantifying these relationships could explain the scope for integration as a tool to improve the performance of imperfectly competitive markets.

References

- Arkolakis, C., A. Costinot, and A. Rodriguez-Clare**, “New trade models, same old gains?,” *American Economic Review*, 2012, 102 (1), 94–130.
- Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodriguez-Clare**, “The elusive pro-competitive effects of trade,” *Review of Economic Studies*, forthcoming.
- Asplund, M. and V. Nocke**, “Firm turnover in imperfectly competitive markets,” *The Review of Economic Studies*, 2006, 73 (2).

- Behrens, Kristian and Yasusada Murata**, “Trade, competition, and efficiency,” *Journal of International Economics*, 2012, 87 (1), 1–17.
- Benassy, J. P.**, “Taste for variety and optimum production patterns in monopolistic competition,” *Economics Letters*, 1996, 52 (1), 41–47.
- Bertoletti, Paolo and Federico Etro**, “Preferences, entry, and market structure,” *The RAND Journal of Economics*, 2016, 47 (4), 792–821.
- , **Eileen Fumagalli, and Clara Poletti**, “Price–cost margins and firm size under monopolistic competition: The case of IES preferences,” *Research in Economics*, 2017.
- , **Federico Etro, and Ina Simonovska**, “International Trade with Indirect Additivity,” *American Economic Journal: Microeconomics*, forthcoming.
- Bykadorov, Igor, Alexey Gorn, Sergey Kokovin, and Evgeny Zhelobodko**, “Why are losses from trade unlikely?,” *Economics Letters*, 2015, (129), 35–38.
- Costinot, Arnaud and Andrés Rodríguez-Clare**, “Trade Theory with Numbers: Quantifying the Consequences of Globalization,” in “In: Helpman, E.(Ed.), Handbook of international economics” Citeseer 2013.
- Dhingra, Swati and John Morrow**, “Monopolistic Competition and Optimum Product Diversity Under Firm Heterogeneity,” *London School of Economics, mimeograph*, 2012.
- Dixit, A. K. and J. E. Stiglitz**, “Monopolistic Competition and Optimum Product Diversity,” *The American Economic Review*, 1977, 67 (3), 297–308.
- Feenstra, R. C.**, “A homothetic utility function for monopolistic competition models, without constant price elasticity,” *Economics Letters*, 2003, 78 (1), 79–86.
- Foster, L., J. C. Haltiwanger, and C. J. Krizan**, “Aggregate productivity growth. Lessons from microeconomic evidence,” in “New developments in productivity analysis,” University of Chicago Press, 2001.
- Görg, Holger, Philipp Henze, Viroj Jienwatcharamongkhol, Daniel Kopasker, Hassan Molana, Catia Montagna, and Fredrik Sjöholm**, “Firm size distribution and employment fluctuations: Theory and evidence,” *Research in Economics*, 2017.
- Hart, O. D.**, “Monopolistic competition in the spirit of Chamberlin: A general model,” *The Review of Economic Studies*, 1985, 52 (4), 529.
- Helpman, E. and P. R. Krugman**, *Market Structure and Foreign Trade: increasing returns, imperfect competition, and the international economy*, MIT Press, 1985.
- Khan, M. A. and Y. Sun**, “Non-cooperative games with many players,” *Handbook of Game Theory with Economic Applications*, 2002, 3, 1761–1808.

- Krugman, P.**, “Increasing Returns, Monopolistic Competition, and International Trade,” *Journal of International Economics*, 1979, 9 (4), 469–479.
- Krugman, P. R.**, “Is free trade passé?,” *The Journal of Economic Perspectives*, 1987, 1 (2).
- Krugman, Paul R.**, “Scale Economies, Product Differentiation, and the Pattern of Trade,” *American Economic Review*, 1980, 70 (5), 950–959.
- Kuhn, K. U. and X. Vives**, “Excess entry, vertical integration, and welfare,” *The Rand Journal of Economics*, 1999, 30 (4), 575–603.
- Mankiw, N. G. and M. D. Whinston**, “Free entry and social inefficiency,” *The RAND Journal of Economics*, 1986, pp. 48–58.
- Melitz, M. J. and G. I. P. Ottaviano**, “Market Size, Trade, and Productivity,” *Review of Economic Studies*, 2008, 75 (1), 295–316.
- and **S. J. Redding**, “Heterogeneous Firms and Trade,” *Handbook of International Trade (commissioned)*, August 2012.
- Melitz, Marc J.**, “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 2003, 71 (6), 1695–1725.
- Melvin and R. D. Warne**, “Monopoly and the theory of international trade,” *Journal of International Economics*, 1973, 3 (2), 117–134.
- Montagna, Catia**, “Monopolistic competition with firm-specific costs,” *Oxford Economic Papers*, 1995, pp. 318–328.
- Nocco, Antonella, Gianmarco I. P. Ottaviano, and Matteo Salto**, “Monopolistic Competition and Optimum Product Selection: Why and How Heterogeneity Matters,” *CEP Discussion Paper*, April 2013, 1206.
- , —, and —, “Monopolistic competition and optimum product selection: why and how heterogeneity matters,” *Research in Economics*, 2017.
- Pavcnik, N.**, “Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants,” *The Review of Economic Studies*, 2002, 69 (1), 245–276.
- Perloff, Jeffrey M. and Steven C. Salop**, “Equilibrium with Product Differentiation,” *The Review of Economic Studies*, 1985, 52 (1).
- Rudin, W.**, *Principles of mathematical analysis*, McGraw-Hill New York, 1964.
- Samuelson, P. A.**, “The monopolistic competition revolution,” *Monopolistic competition theory: studies in impact*, 1967, pp. 105–38.
- Solow, R. M.**, *Monopolistic competition and macroeconomic theory*, Cambridge University Press, 1998.

Spence, M., “Product Selection, Fixed Costs, and Monopolistic Competition,” *The Review of Economic Studies*, 1976, 43 (2), 217–235.

Stiglitz, J. E., “Towards a more general theory of monopolistic competition,” *Prices, competition and equilibrium*, 1986, p. 22.

Ushchev, Philip, “Multi-product firms in monopolistic competition: the role of scale-scope spillovers,” *Research in Economics*, 2017.

Venables, A. J., “Trade and trade policy with imperfect competition: The case of identical products and free entry,” *Journal of International Economics*, 1985, 19 (1-2), 1–19.

Vives, X., *Oligopoly pricing: old ideas and new tools*, The MIT press, 2001.

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti, and Jacques-François Thisse, “Monopolistic competition: Beyond the constant elasticity of substitution,” *Econometrica*, 2012, 80 (6), 2765–2784.

A Appendix: The Impact of Large Markets

To arrive at the large market result, we first state Lemmas characterizing convergence in the large market and then show market allocations coincide with optimal allocations. Detailed proofs of the Lemmas are in the Online Appendix.

Lemma 1. *As market size becomes large:*

1. *Market revenue per capita is increasing in market size and goes to infinity.*
2. *At the optimum, utility per capita is increasing in market size and goes to infinity.*
3. *Market entry goes to infinity.*

Proof. From Dhingra and Morrow (2012), the market allocation solves

$$\max_{M_e, c_d, q(c)} LM_e \int_0^{c_d} u'(q(c)) q(c) dG \text{ subject to } L \geq M_e \left(\int_0^{c_d} Lc q(c) + fdG + F_e \right).$$

Let $R(L) \equiv M_e \int_0^{c_d} u'(q(c)) q(c) dG$ be the revenue per capita under the market allocation. Fix L and let $\{q(c), c_d, M_e\}$ denote the market allocation with L resources. Consider an increased resource level $\tilde{L} > L$ with allocation $\{\tilde{q}(c), \tilde{c}_d, \tilde{M}_e\} \equiv \{(L/\tilde{L}) \cdot q(c), c_d, (\tilde{L}/L) \cdot M_e\}$ which direct inspection shows is feasible. This allocation generates revenue per capita of

$$\tilde{M}_e \int_0^{\tilde{c}_d} u'(\tilde{q}(c)) q(c) dG = M_e \int_0^{c_d} u'((L/\tilde{L}) \cdot q(c)) q(c) dG \leq R(\tilde{L}).$$

Since u is concave, it follows that $R(\tilde{L}) > R(L)$. Since $\tilde{q}(c) = (L/\tilde{L}) \cdot q(c) \rightarrow 0$ for all $c > 0$ and $\lim_{q \rightarrow 0} u'(q) = \infty$, revenue per capita goes to infinity as $\tilde{L} \rightarrow \infty$. A similar argument holds for the social optimum.

First note that $q(c)$ is fixed by $u'(q(c)) [1 - \mu(q(c))] = \delta c$, and $\delta \rightarrow \infty$ and $\mu(q(c))$ is bounded, it must be that $u'(q(c)) \rightarrow \infty$ for $c > 0$. This requires $q(c) \rightarrow 0$ for $c > 0$. Since revenue $u'(q(c))q(c)$ is equal to $\varepsilon(q(c))u(q(c))$ and ε is bounded, revenue also goes to zero for each $c > 0$. Revenue is also decreasing in δ for every c , so we can bound revenue with a function $B(c)$. In particular, for any fixed market size \tilde{L} and implied allocation $\{\tilde{q}(c), \tilde{c}_d, \tilde{M}_e\}$, for $L \geq \tilde{L}$:

$$u'(q(c))q(c)\mathbf{1}_{[0, c_d]}(c) \leq u'(\tilde{q}(c))\tilde{q}(c)\mathbf{1}_{[0, \tilde{c}_d]}(c) + u'(\tilde{q}(\tilde{c}_d))\tilde{q}(\tilde{c}_d)\mathbf{1}_{[\tilde{c}_d, \infty]}(c) \equiv B(c) \quad (3)$$

where we appeal to the fact that $q(c)$ is decreasing in c for any market size. Since for any L , $\int_0^{c_d} u'(q(c))q(c)dG = \delta/M_e$, it is clear that $\int_0^\infty B(c)dG = \int_0^{\tilde{c}_d} u'(\tilde{q}(c))\tilde{q}(c)dG + u'(\tilde{q}(\tilde{c}_d))\tilde{q}(\tilde{c}_d) < \infty$. Since $u'(q(c))q(c)$ converges pointwise to zero for $c > 0$, we conclude

$$\lim_{L \rightarrow \infty} \int_0^{c_d} u'(q(c))q(c)dG = \int_0^{c_d} \lim_{L \rightarrow \infty} u'(q(c))q(c)dG = 0$$

by dominated convergence. Therefore $\lim_{L \rightarrow \infty} \delta/M_e = 0$ which with $\delta \rightarrow \infty$ shows $M_e \rightarrow \infty$. The optimal allocation case is similar. \square

Lemma 2. *For all market sizes and all positive marginal cost ($c > 0$) firms:*

1. *Profits ($\pi(c)$) and social profits ($\varpi(c) \equiv (1 - \varepsilon(c)) / \varepsilon(c) \cdot cq(c)L - f$) are bounded.*
2. *Total quantities ($Lq(c)$) in the market and optimal allocation are bounded.*

Proof. For any costs $c_L < c_H$, $q(c_H)$ is in the choice set of a firm with costs c_L and therefore

$$\pi(c_L) \geq (p(c_H) - c_L)q(c_H)L - f = \pi(c_H) + (c_H - c_L)q(c_H)L. \quad (4)$$

Furthermore, for every $\tilde{c} > 0$, we argue $\pi(\tilde{c})$ is bounded. For $\underline{c} \equiv \tilde{c}/2$, $\pi(\tilde{c}) \leq \pi(\underline{c})$ while $\pi(\underline{c})$ is bounded since $\lim_{L \rightarrow \infty} \int_0^{c_d} \pi(c)dG = F_e$ and $\limsup_{L \rightarrow \infty} \pi(\underline{c}) = \infty$ would imply $\limsup_{L \rightarrow \infty} \int_0^{c_d} \pi(c)dG = \infty$. It follows from Equation (4) that $Lq(c)$ is bounded. Substituting ϖ for π leads to similar arguments for the social optimum. \square

Proposition 3. *Assume markups are interior. Then under the market allocation:*

1. $\lim_{L \rightarrow \infty} c_d^{\text{mkt}} = \infty$ iff $\lim_{L \rightarrow \infty} p(c_d^{\text{mkt}}) = \infty$ iff $\lim_{L \rightarrow \infty} Lq(c_d^{\text{mkt}}) = 0$.
2. $\lim_{L \rightarrow \infty} c_d^{\text{mkt}} = 0$ iff $\lim_{L \rightarrow \infty} p(c_d^{\text{mkt}}) = 0$ iff $\lim_{L \rightarrow \infty} Lq(c_d^{\text{mkt}}) = \infty$.

3. $\lim_{L \rightarrow \infty} c_d^{\text{mkt}} \in (0, \infty)$ iff $\lim_{L \rightarrow \infty} p(c_d^{\text{mkt}}) \in (0, \infty)$ iff $\lim_{L \rightarrow \infty} Lq(c_d^{\text{mkt}}) \in (0, \infty)$.

Similarly, under the optimal allocation:

1. $\lim_{L \rightarrow \infty} c_d^{\text{opt}} = \infty$ iff $\lim_{L \rightarrow \infty} u \circ q(c_d^{\text{opt}}) / \lambda q(c_d^{\text{opt}}) = \infty$ iff $\lim_{L \rightarrow \infty} Lq(c_d^{\text{opt}}) = 0$.

2. $\lim_{L \rightarrow \infty} c_d^{\text{opt}} = 0$ iff $\lim_{L \rightarrow \infty} u \circ q(c_d^{\text{opt}}) / \lambda q(c_d^{\text{opt}}) = 0$ iff $\lim_{L \rightarrow \infty} Lq(c_d^{\text{opt}}) = \infty$.

3. $\lim_{L \rightarrow \infty} c_d^{\text{opt}} \in (0, \infty)$ iff $\lim_{L \rightarrow \infty} u \circ q(c_d^{\text{opt}}) / \lambda q(c_d^{\text{opt}}) \in (0, \infty)$ iff $\lim_{L \rightarrow \infty} Lq(c_d^{\text{opt}}) \in (0, \infty)$.

Proof. Note the following zero profit relationships that hold at the cost cutoff c_d , suppressing the market superscripts throughout we have:

$$u'(q(c_d)) / \delta - f / [Lq(c_d) \cdot \mu \circ q(c_d) / (1 - \mu \circ q(c_d))] = c_d, \quad (5)$$

$$Lc_d q(c_d) \cdot \mu \circ q(c_d) / (1 - \mu \circ q(c_d)) = f. \quad (6)$$

First, if $\lim_{L \rightarrow \infty} Lq(c_d) = 0$, Equation (6) implies $c_d \cdot \mu \circ q(c_d) / (1 - \mu \circ q(c_d)) \rightarrow \infty$. Clearly $q(c_d) \rightarrow 0$ and since $\lim_{q \rightarrow 0} \mu(q) \in (0, 1)$, $\mu \circ q(c_d) / (1 - \mu \circ q(c_d))$ is bounded, and therefore $c_d \rightarrow \infty$. Now suppose $c_d \rightarrow \infty$ and since $c_d \leq u'(q(c_d)) / \delta$, $u'(q(c_d)) / \delta \rightarrow \infty$. Finally, if $u'(q(c_d)) / \delta \rightarrow \infty$, since $\delta \rightarrow \infty$, necessarily $q(c_d) \rightarrow 0$ so we find $\mu \circ q(c_d) / (1 - \mu \circ q(c_d))$ is bounded. It follows from Equation (6) that $Lc_d q(c_d)$ is bounded, so from Equation (5), $Lq(c_d) \cdot u'(q(c_d)) / \delta$ is bounded so $Lq(c_d) \rightarrow 0$.

If $\lim_{L \rightarrow \infty} Lq(c_d) = \infty$, $q(c_d) \rightarrow 0$ so from $\lim_{q \rightarrow 0} \mu(q) \in (0, 1)$, $\mu \circ q(c_d) / (1 - \mu \circ q(c_d))$ is bounded. Therefore from Equation (6), $c_d \rightarrow 0$. Now assume $c_d \rightarrow 0$ so from (6), $Lq(c_d) \cdot \mu \circ q(c_d) / (1 - \mu \circ q(c_d)) \rightarrow \infty$ which implies with Equation (5) that $u'(q(c_d)) / \delta \rightarrow 0$. Finally, if $u'(q(c_d)) / \delta \rightarrow 0$, (5) shows $c_d \rightarrow 0$.

The second set of equivalences follows from examining the conditions for a firm at the limiting cost cutoff $c_d^\infty \in (0, \infty)$. The argument for the optimal allocation is similar. \square

Lemma 3. *Assume interior convergence. Then as market size grows large:*

1. *In the market, $p(c)$ converges in $(0, \infty)$ for $c > 0$ and $Lq(c_d)$ converges in $(0, \infty)$.*
2. *In the optimum, $u \circ q(c) / \lambda q(c)$ and $Lq(c_d)$ converge in $(0, \infty)$ for $c > 0$.*

Proof. Since $q(c) \rightarrow 0$ for all $c > 0$, $\lim_{q \rightarrow 0} \mu(q) \in (0, 1)$ shows $\lim_{L \rightarrow \infty} p(c)$ aligns with constant markups and thus converges for all $c > 0$. In particular, $p(c_d)$ converges and $L(p(c_d) - c_d)q(c_d) = f$ so it follows $Lq(c_d)$ converges. Similar arguments hold for the social optimum. \square

Lemma 4. *Assume interior convergence and large market identification. Then for the market and social optimum, $Lq(c)$ converges for $c > 0$.*

Proof. Fix any $c > 0$ and first note that for both the market and social planner, $q(c)/q(c_d) = Lq(c)/Lq(c_d)$ and both $Lq(c)$ and $Lq(c_d)$ are bounded, so $q(c)/q(c_d)$ is bounded.

Now consider the market. $q(c)/q(c_d) \geq 1$ has at least one limit point and if it has two limit points, say a and b with $a < b$, there exist subsequences $(q(c)/q(c_d))_{a_n} \rightarrow a$ and $(q(c)/q(c_d))_{b_n} \rightarrow b$. There also exist distinct κ and $\tilde{\kappa}$ in (a, b) so that eventually

$$(q(c))_{a_n} < \kappa q(c_d)_{a_n} < \tilde{\kappa} q(c_d)_{b_n} < (q(c))_{b_n}.$$

With $u'' < 0$ this implies

$$\begin{aligned} (u'(q(c))/u'(q(c_d)))_{a_n} &> (u'(\kappa q(c_d))/u'(q(c_d)))_{a_n} > (u'(\tilde{\kappa} q(c_d))/u'(q(c_d)))_{b_n} \\ &> (u'(q(c))/u'(q(c_d)))_{b_n}. \end{aligned}$$

By assumption, $\lim_{q \rightarrow 0} u'(\kappa q)/u'(q) > \lim_{q \rightarrow 0} u'(\tilde{\kappa} q)/u'(q)$ but since $q(c) \rightarrow 0$,

$$\begin{aligned} \lim_{n \rightarrow \infty} (u' \circ q(c)/u' \circ q(c_d))_{a_n} &= \lim_{n \rightarrow \infty} ([1 - \mu \circ q(c)]c / [1 - \mu \circ q(c_d)]c_d)_{a_n} = c/c_d \\ &= \lim_{n \rightarrow \infty} (u' \circ q(c)/u' \circ q(c_d))_{b_n} \end{aligned}$$

where we have used the fact that $\lim_{q \rightarrow 0} \mu(q) \in (0, 1)$, however by assumption this contradicts $a < b$.

For the social optimum, this argument holds (substituting $\varepsilon \neq 0$ for $u'' < 0$) so long as

$$\kappa \neq \tilde{\kappa} \text{ implies } \lim_{q \rightarrow 0} (u(\kappa q)/\kappa q) / (u(q)/q) \neq \lim_{q \rightarrow 0} (u(\tilde{\kappa} q)/\kappa q) / (u(q)/q). \quad (7)$$

Since $\lim_{q \rightarrow 0} u'(q) = \infty$ and $\lim_{q \rightarrow 0} \varepsilon \in (0, \infty)$ it follows that $\lim_{q \rightarrow 0} u(q)/q = \infty$. By L'Hospital's rule, $\lim_{q \rightarrow 0} (u(\kappa q)/\kappa q) / (u(q)/q) = \lim_{q \rightarrow 0} u'(\kappa q)/u'(q)$ for all κ so the condition (7) in holds because $\kappa \neq \tilde{\kappa}$ implies $\lim_{q \rightarrow 0} u'(\kappa q)/u'(q) \neq \lim_{q \rightarrow 0} u'(\tilde{\kappa} q)/u'(q)$. \square

Lemma 5. *At extreme quantities, social and private markups align as follows:*

1. If $\lim_{q \rightarrow 0} 1 - \varepsilon(q) < 1$ then $\lim_{q \rightarrow 0} 1 - \varepsilon(q) = \lim_{q \rightarrow 0} \mu(q)$.
2. If $\lim_{q \rightarrow \infty} 1 - \varepsilon(q) < 1$ then $\lim_{q \rightarrow \infty} 1 - \varepsilon(q) = \lim_{q \rightarrow \infty} \mu(q)$.

Proof. By assumption, $\lim_{q \rightarrow 0} \varepsilon(q) > 0$. Expanding this limit via L'Hospital's rule shows

$$\begin{aligned} \lim_{q \rightarrow 0} \varepsilon(q) &= \lim_{q \rightarrow 0} q / (u(q)/u'(q)) = \lim_{q \rightarrow 0} 1 / \lim_{q \rightarrow 0} (1 - u(q)u''(q)/(u'(q))^2) \\ &= 1 / \lim_{q \rightarrow 0} (1 + \mu(q) / \varepsilon(q)) = \lim_{q \rightarrow 0} \varepsilon(q) / \lim_{q \rightarrow 0} (\varepsilon(q) + \mu(q)) \end{aligned}$$

which gives the first part of the result. Identical steps for $q \rightarrow \infty$ give the second part. \square

Lemma 6. *Assume interior convergence and large market identification. As market size grows large:*

1. $q(c)/q(c_d) \rightarrow (c/c_d)^{-1/\alpha}$ with $\alpha = \lim_{q \rightarrow 0} \mu(q)$.
2. The cost cutoffs for the social optimum and market converge to the same value.
3. The entrant per worker ratios M_e/L converge to the same value.

Proof. Define $\Upsilon(c/c_d)$ by (the above results show this limit is well defined)

$$\Upsilon(c/c_d) \equiv \lim_{q \rightarrow 0} u'(\Upsilon(c/c_d)q) / u'(q) = c/c_d.$$

We will show in fact that $\Upsilon(c/c_d) = (c/c_d)^{-\alpha}$. It follows from the definition that Υ is weakly decreasing, and the results above show Υ is one to one, so it is strictly decreasing. Define $f_q(z) \equiv u'(zq)/u'(q)$ so $\lim_{q \rightarrow 0} f_q(z) = \Upsilon^{-1}(z)$ for all $\Upsilon^{-1}(z) \in (0, 1)$. Note

$$f'_q(z) = u''(zq)q/u'(q) = -\mu(zq) \cdot u'(zq)/zu'(q)$$

so since $\lim_{q \rightarrow 0} \mu(zq) = \mu^\infty \in (0, 1)$ and $\lim_{q \rightarrow 0} u'(zq)/zu'(q) = \Upsilon^{-1}(z)/z$, we know that $\lim_{q \rightarrow 0} f'_q(z) = -\mu^\infty \Upsilon^{-1}(z)/z$. On any strictly positive closed interval I , μ and $u'(zq)/zu'(q)$ are monotone in z so $f'_q(z)$ converges uniformly on I as $q \rightarrow 0$. Rudin (1964) (Thm 7.17) shows

$$\lim_{q \rightarrow 0} f'_q(z) = d \lim_{q \rightarrow 0} f_q(z) / dz = -\mu^\infty \Upsilon^{-1}(z) / z = d\Upsilon^{-1}(z) / dz. \quad (8)$$

We conclude that $\Upsilon^{-1}(z)$ is differentiable and thus continuous. Given the form deduced in (8), $\Upsilon^{-1}(z)$ is continuously differentiable. Since $d\Upsilon^{-1}(z)/dz = 1/\Upsilon' \circ \Upsilon^{-1}(z)$, composing both sides with $\Upsilon(z)$ and using (8) we have $\Upsilon'(z) = -\Upsilon(z)/\mu^\infty z$. Therefore Υ is CES, in particular $\Upsilon(z) = z^{-1/\mu^\infty}$.

Finally, let c_∞^{opt} and c_∞^{mkt} be the limiting cost cutoffs as $L \rightarrow \infty$ for at the social optimum and market, respectively. Letting $q^{\text{opt}}(c)$, $q^{\text{mkt}}(c)$ denote the socially optimal and market quantities,

we know from above that for all $c > 0$:

$$q^{\text{opt}}(c)/q^{\text{opt}}(c_d^{\text{opt}}) \longrightarrow (c_d^{\text{opt}}/c)^{1/\alpha}, \quad q^{\text{mkt}}(c)/q^{\text{mkt}}(c_d^{\text{mkt}}) \longrightarrow (c_d^{\text{mkt}}/c)^{1/\alpha}. \quad (9)$$

Now consider the conditions involving f_e , $\int_0^{c_d^{\text{mkt}}} \pi(c) dG = f_e = \int_0^{c_d^{\text{opt}}} \varpi(c) dG$. Expanding,

$$L \int_0^{c_d^{\text{mkt}}} \frac{\mu \circ q^{\text{mkt}}(c)}{1 - \mu \circ q^{\text{mkt}}(c)} c q^{\text{mkt}}(c) dG - fG(c_d^{\text{mkt}}) = L \int_0^{c_d^{\text{opt}}} \frac{1 - \varepsilon \circ q^{\text{opt}}(c)}{\varepsilon \circ q^{\text{opt}}(c)} c q^{\text{opt}}(c) dG - fG(c_d^{\text{opt}}).$$

It necessarily follows that

$$\begin{aligned} & \lim_{L \rightarrow \infty} L \int_0^{c_d^{\text{mkt}}} \mu \circ q^{\text{mkt}}(c) / (1 - \mu \circ q^{\text{mkt}}(c)) \cdot c q^{\text{mkt}}(c) dG - fG(c_d^{\text{mkt}}) = \\ & \lim_{L \rightarrow \infty} L \int_0^{c_d^{\text{opt}}} (1 - \varepsilon \circ q^{\text{opt}}(c)) / \varepsilon \circ q^{\text{opt}}(c) \cdot c q^{\text{opt}}(c) dG - fG(c_d^{\text{opt}}). \end{aligned} \quad (10)$$

Using Equation (9), we see that $Lq^{\text{opt}}(c)$ and $Lq^{\text{mkt}}(c)$ converge uniformly on any strictly positive closed interval. Combined with the fact that $\lim_{q \rightarrow 0} \mu(q) = \lim_{q \rightarrow 0} 1 - \varepsilon(q)$, we see from Equation (10) the limits of the $\mu/(1 - \mu)$ and $(1 - \varepsilon)/\varepsilon$ terms are equal and factor out of Equation (10), leaving

$$\begin{aligned} & \lim_{L \rightarrow \infty} L c_d^{\text{mkt}} q^{\text{mkt}}(c_d^{\text{mkt}}) \int_0^{c_d^{\text{mkt}}} (c/c_d^{\text{mkt}})(c/c_d^{\text{mkt}})^{-1/\alpha} dG - fG(c_d^{\text{mkt}}) = \\ & \lim_{L \rightarrow \infty} L c_d^{\text{opt}} q^{\text{opt}}(c_d^{\text{opt}}) \int_0^{c_d^{\text{opt}}} (c/c_d^{\text{opt}})(c/c_d^{\text{opt}})^{-1/\alpha} dG - fG(c_d^{\text{opt}}). \end{aligned}$$

Noting $f(1 - \mu^\infty)/\mu^\infty = L c_d^{\text{mkt}} q^{\text{mkt}}(c_d^{\text{mkt}}) = L c_d^{\text{opt}} q^{\text{opt}}(c_d^{\text{opt}})$, we therefore have

$$\begin{aligned} & \lim_{L \rightarrow \infty} \int_0^{c_d^{\text{mkt}}} (c/c_d^{\text{mkt}})^{1-1/\alpha} (c_d^{\text{mkt}}/c_d^{\text{mkt}})^{-1/\alpha} dG - G(c_d^{\text{mkt}}) = \\ & \lim_{L \rightarrow \infty} \int_0^{c_d^{\text{opt}}} (c/c_d^{\text{opt}})^{1-1/\alpha} (c_d^{\text{opt}}/c_d^{\text{opt}})^{-1/\alpha} dG - G(c_d^{\text{opt}}) \end{aligned}$$

so that finally evaluating the limits, we have

$$\int_0^{c_d^{\text{mkt}}} \left[(c/c_d^{\text{mkt}})^{1-1/\alpha} - 1 \right] dG = \int_0^{c_d^{\text{opt}}} \left[(c/c_d^{\text{opt}})^{1-1/\alpha} - 1 \right] dG. \quad (11)$$

Letting $h(w) \equiv \int_0^w \left[(c/w)^{1-1/\alpha} - 1 \right] dG$, we see that $h'(w) = \int_0^w (1/\alpha - 1) c^{1-1/\alpha} w^{1/\alpha-2} dG$

and since $\alpha = \mu^\infty \in (0, 1)$, $h' > 0$. Since h is strictly increasing, there is a unique c_∞^{opt} , namely $c_\infty^{\text{opt}} = c_\infty^{\text{mkt}}$ such that Equation (11) holds. Checking the conditions for L/M_e show they coincide between the market and social optimum as well. \square