CENTRE *for* ECONOMIC
PERFORMANCE

CEP Discussion Paper No 1130

February 2012

The Impact of Integration on Productivity and Welfare Distortions Under Monopolistic Competition

Swati Dhingra and John Morrow

THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Abstract

A fundamental question in monopolistic competition theory is whether the market allocates resources efficiently. This paper generalizes the Spence-Dixit-Stiglitz framework to heterogeneous firms, addressing when the market provides optimal quantities, variety and productivity. Under constant elasticity demand, each firm prices above its average cost, yet we show market allocations are efficient. When demand elasticities vary, market allocations are not efficient and reflect the distortions of imperfect competition. After determining the nature of market distortions, we investigate how integration may serve as a remedy to imperfect competition. Both market distortions and the impact of integration depend on two demand side elasticities, and we suggest richer demand structures to pin down these elasticities. We also show that integration eliminates distortions, provided the post-integration market is sufficiently large.

1 Introduction

Empirical work has drawn attention to the high degree of heterogeneity in firm productivity and the impact of market integration on firm survival and markups (Bernard et al. 2007, Feenstra 2006). The introduction of firm heterogeneity in monopolistic competition models has provided new insights into the reallocation of resources within industries. A fundamental question within this setting is whether the market allocates resources efficiently. Symmetric firm models explain when market allocations are efficient by examining the trade off between quantity and product variety. When firms are heterogeneous in productivity, we must also ask which types of firms should produce and which should be shut down. This paper examines how firm heterogeneity affects market efficiency. We focus on three key questions. First, does the market allocate resources efficiently? Second, what is the nature of distortions, if any? Third, can economic integration reduce distortions through increased competition?

We answer these questions in the standard setting of a monopolistically competitive industry with heterogeneous productivity draws and free entry (e.g. Melitz 2003). Empirical work shows that firms are rarely symmetric and markups are unlikely to be constant.[1] To allow rich interrelationships between productivity, markups and efficiency, we focus on the general class of variable elasticity demand systems, considered by Dixit and Stiglitz (1977). When demand elasticities vary with quantity and firms vary in productivity, markups will vary within a market. Heterogeneous markups give rise to a range of possible market distortions which would not arise if firms were symmetric or charged constant markups.

Differences in productivity across firms change optimal allocation decisions in a fundamental way. In symmetric firm models, marginal cost pricing and average cost pricing serve as heuristics for first-best and second-best resource allocations. In heterogeneous firm models, inducing each firm to price at its marginal cost or average cost will not maximize welfare because neither scheme takes into account sunk entry costs and the effect of heterogeneity on resource costs. Thus, different levels of production maximize welfare than what marginal or average cost pricing imply. For example, it could be welfare-improving to skew resources towards firms with lower costs (to conserve resources) or towards firms with higher costs (to preserve variety). The relative position of a firm in the cost distribution matters, and incorporating differences in firm costs can alter welfare and policy analysis substantially.

As a heterogeneous cost environment presents information problems for policy, one potential tool to improve efficiency is to integrate with international markets. For instance, the distortions of imperfect competition may be mitigated with increased competition from foreign firms, implying that integration provides opportunities to correct market failure. This idea of welfare gains from

---

[1]CES demand provides a useful benchmark by forcing constant markups that ensure market size plays no role in productivity changes. However, recent studies find market size matters for firm size (Campbell and Hopenhayn 2005) and productivity dispersion (Syverson 2004). Foster et al. (2008) show that "profitability" rather than productivity is more important for firm selection, suggesting a role for richer demand specifications.

mitigating distortions goes back to at least Melvin et al. (1973). Building on this insight, we model international integration as access to new markets and examine whether it can be a policy tool to correct distortions.[2]

Starting with constant elasticity of substitution (CES) demand, we show that when firms vary in productivity, market allocations are efficient but firms earn positive profits. This result seems surprising, based on the logic of marginal or average cost pricing which is designed to return producer surplus to consumers. With productivity differences, the market requires prices above average costs to induce firms to enter and potentially take a loss. Free entry ensures the wedge between prices and average costs exactly finances sunk entry costs. Therefore, the market implements the first-best allocation and laissez faire industrial policy is optimal.[3]

How broadly does this efficiency result hold? We generalize the demand structure to the variable elasticity of substitution (VES) form of Dixit and Stiglitz which provides a rich setting for a wide range of market outcomes (Spence 1976; Vives 2001; Zhelobodko et al. 2011). Within this setting, market efficiency is unique to CES demand. While the market does maximize real revenue, private benefits to firms are perfectly aligned with social benefits only under CES demand.

The nature of distortions under VES demand can be determined by two demand side elasticities: the inverse demand elasticity and the elasticity of utility. Misalignment of these elasticities determines the bias in market allocations relative to optimal allocations. This is in sharp contrast to symmetric firm models where the elasticity of utility is enough to determine this bias in market allocations across quantity and variety. When firms are heterogeneous, the market allocates resources across different types of firms and the variable elasticity of demand shapes the distribution of quantities produced. This distinguishes distortions in heterogeneous firm markets in two respects. First, some firms may over-produce while others under-produce within the same market. This shifts the focus from a general level of production to the distribution of production. Second, the distribution of variable markups affects firm entry, and aggregate quantity and variety now depend on both the elasticity of utility and the elasticity of demand. This alters how one thinks about the trade-off between aggregate quantity versus variety.

Integration with international markets can potentially mitigate distortions across the distribution of firms. To capture the role integration as a policy tool, we first examine the effects of integration with large markets. Such integration will push outcomes towards what we define as the monopolistically competitive limit, which eliminates distortions. In this limit firm heterogeneity and market power persist, but market size is so large that the quantity sold from a firm to each

---

[2]International integration is equivalent to an expansion in market size (e.g., Krugman 1979). As our focus is on efficiency, we abstract from trade frictions which introduce cross-country distributional issues.

[3]Melitz (2003) considers both variable and fixed costs of exporting. We show that the open Melitz economy is efficient, even in the presence of trade frictions. In the presence of fixed export costs, the firms a policymaker would close down in the open economy are exactly those that would not survive in the market. However, a policymaker would not close down firms in the absence of export costs. Thus, the rise in productivity following trade provides welfare gains by optimally internalizing trade frictions. Market allocations are efficient even with asymmetric countries, but the presence of trade frictions introduces distributional concerns which we do not address.

worker is negligible. Most importantly, VES demand operates much like CES demand, and market allocations are efficient. This shows that competition can eliminate distortions while retaining the characteristics of markets with heterogeneous firms who possess market power. Markups become constant across firms, so CES demand might approximate richer demand structures in large economies. However, the monopolistically competitive limit may require a market size which is unattainable even in fully integrated world markets.

Integration with small markets has different effects than integration with large markets. In small markets, heterogeneous markups lead to distortions and small increases in market size may increase welfare, but fail to reduce distortions. We illustrate this with the impact of integration on productivity distortions in small markets. Under VES demand, changes in market productivity depend on the inverse demand elasticity, while changes in optimal productivity depend on the elasticity of utility. A comparison of these elasticities determines how integration affects the gap between market productivity and optimal productivity. When these elasticities are aligned, distortions dissipate. However, when the elasticities are misaligned, integration can exacerbate distortions. This implies that integration may increase the scope for policies that anticipate how integration will alter production in imperfect markets. In this regard, estimation of richer demand structures becomes imperative in understanding how integration impacts market distortions. Our last results provide suggestions about how to assess distortions empirically.

The paper is organized as follows. Section 2 relates this paper to previous work and Section 3 recaps trade models with firm heterogeneity. Section 4 presents efficiency results in a closed economy. Section 5 introduces international trade and contrasts the efficiency of CES demand with inefficiency of VES demand, also deriving a monopolistically competitive limit which shows how integration can eliminate distortions. Section 6 characterizes the nature of static distortions and further analyzes the impact of integration on distortions. Section 7 gathers together some theoretical implications useful for designing empirical strategies and Section 8 concludes.

2   Related Work

Our paper is related to work on welfare gains in industrial organization and international economics. The trade off between variety and quantity occupies a prominent place in the industrial organization literature (e.g., Mankiw and Whinston 1986). We contribute to this literature by studying the effects of firm heterogeneity and international trade. The analysis is motivated by efficiency properties which have been studied at length in symmetric firm models of monopolistic competition.[4] Recently, Bilbiie et al. (2006) show the market equilibrium with symmetric firms is socially optimal if and only if preferences are CES. We generalize the result to heterogeneous firms and show that efficiency is unrelated to the productivity distribution of firms. To the best of

---

[4]Spence (1976); Dixit and Stiglitz (1977); Venables (1985); Bilbiie et al. (2006); Behrens and Murata (2009).

our knowledge, this is the first paper to show market outcomes in Melitz are first best.[5]

To highlight the potential scope of market imperfections, we generalize the well known CES demand structure to VES demand. In contemporaneous work, Zhelobodko et al. (2011) develop complementary results for market outcomes under VES demand and demonstrate its richness and tractability under various assumptions such as multiple sectors and vertical differentiation. Unlike Zhelobodko et al., our focus is on market efficiency.

We also study the limiting behavior of a VES economy. A large literature examines whether monopolistic competition arises as a limit to oligopolistic pricing and when monopolistic competition converges to perfect competition in symmetric firm models (Vives 2001, Chapter 6). This literature considers the limiting behavior as the number of firms tends to infinity. Instead, we examine a monopolistically competitive model with a continuum of firms so there are infinitely many firms even in an economy with finite market size. After establishing the equivalence of increased international trade and increased market size, we study the limiting behavior in terms of the model primitive of market size becoming large.

The findings of our paper are related to an emerging literature on welfare gains in new trade models. Generalizing Krugman (1980) to heterogeneous firms, Melitz shows that opening to trade raises welfare through reallocation of resources towards high productivity firms. Considering 48 countries exporting to the US in 1980-2000, Feenstra and Kee (2008) estimate that rise in export variety accounts for an average 3.3 per cent rise in productivity and GDP for the exporting country. In recent influential work, Arkolakis et al. (forthcoming) show that the mapping between trade data and welfare is the same across several old and new trade models with different production structures. This equivalence holds for models which permit welfare to be summarized by import shares and trade elasticities (that can be derived from gravity equations). Once the Spence-Dixit-Stiglitz demand framework is considered, we find welfare inferences from import shares require additional information about demand and become more structural in nature. Unlike Arkolakis et al., we focus on market efficiency and show that the potential gains from integration are larger because of the opportunity to reduce distortions.

Our results speak directly to the mixed findings about trade liberalization and productivity in the empirical literature. Following trade liberalization, some countries show a reallocation towards high productivity firms while others show a reallocation towards low productivity firms.[6] Tybout

---

[5]We consider this to be the proof of a folk theorem. The idea of efficiency in Melitz has been "in the air." Within the heterogeneous firm literature, Baldwin and Robert-Nicoud (2008) and Feenstra and Kee (2008) discuss certain efficiency properties of the Melitz economy. In their working paper, Atkeson and Burstein (2010) consider a first order approximation and numerical exercises to show that productivity increases are offset by reductions in variety. We provide an analytical treatment to show the market equilibrium implements the unconstrained social optimum. Helpman et al. (2011) consider the constrained social optimum in the presence of a homogeneous good. Their approach differs because the homogeneous good fixes the marginal utility of income.

[6]Interpreting firm size as productivity, Tybout (2003) notes that it was the high productivity firms that lost market share in Chile and Colombia while it was the low productivity firms that suffered a decline in Morocco. While productivity estimation is fraught with difficulties in measuring technical efficiency, we focus on the relationship between productivity and welfare as in the heterogeneous firm literature.

(2003) proposes that these mixed findings could mean that the selection effects emphasized by Melitz are not robust, or that firm size is a poor proxy for productivity. We address the first issue by examining the robustness of selection effects to general demand specifications. Differences in inverse demand elasticities induce different patterns of firm selection, reconciling the mixed evidence for productivity changes across heterogeneous firms. The second issue of productivity measurement has been addressed in several studies. Instead of measurement, we focus on how VES demand can better explain observed patterns. As both observed markups and physical productivity vary with market size under general demand specifications, our findings reiterate the importance of disentangling changes in markups and productivity to understand the sources of welfare gains from trade. We characterize when observed productivity gains reflect a narrowing of the distortionary gap between market and optimal productivity. Therefore, our work is in line with Tybout (2003) and Katayama et al. (2009) who point to the limitations of the empirical literature in mapping observed productivity gains to welfare and optimal policies.

3   Trade Models with Heterogeneous Firms

Trade models with heterogeneous firms differ from earlier trade 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. In this section we briefly recap the implications of asymmetric costs for consumers, firms and equilibrium outcomes.

3.1   Consumers

A mass L of identical consumers in an economy are each endowed with one unit of labor and face a wage rate w normalized to one. Preferences are identical across all consumers in home and foreign countries. Let M_e denote the mass of entering varieties and $q(c)$ denote quantity consumed of variety c by each consumer. A consumer has preferences over differentiated goods $U(M_e, q)$ which take the general VES form:

$$U(M_e, q) \equiv M_e \int u(q(c))dG. \text{ (VES)} \tag{1}$$

Here u denotes utility from an individual variety and $\int u(q)dG$ denotes utility from a unit bundle of differentiated varieties. In a Melitz economy, preferences take the special CES form with $u(q) = q^\rho$.[7] More generally, we assume preferences satisfy usual regularity conditions which guarantee well defined consumer and firm problems.

---

[7]The specific CES form in Melitz is $U(M_e, q) \equiv M_e^{1/\rho} \left(\int (q(c))^\rho dG\right)^{1/\rho}$ but the normalization of the exponent $1/\rho$ in Equation (1) will not play a role in allocation decisions.

Definition 1. (Regular Preferences) u satisfies the following conditions:

1. $u(0)$ is normalized to zero.

2. u is twice continuously differentiable, increasing and concave.

3. $(u'(q) \cdot q)'$ is strictly decreasing in quantity.

4. The elasticity of marginal utility $\mu(q) \equiv |qu''(q)/u'(q)|$ is less than one.

For each good indexed by c, VES preferences induce an inverse demand $p(q(c)) = u'(q(c))/\delta$ where δ is a 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 u.[8] Multiplying both sides of the inverse demand by $q(c)$ and aggregating over all c, the budget multiplier is $\delta = M_e \int_0^{c_a} u'(q(c)) \cdot q(c) dG$.

3.2 Firms

There is a continuum of firms which may enter the market for differentiated goods, by paying a sunk entry cost of f_e. Each firm produces a single variety so the mass of entering firms is the mass of entering varieties M_e. Upon entry, each firm receives a unit cost c drawn from a distribution G with continuously differentiable pdf g.[9]

After entry, should a firm produce for the domestic market it faces a cost function $\text{TC}(q(c)) \equiv cq(c) + f$ where f denotes the fixed cost of production. Each firm faces an inverse demand of $p(q(c)) = u'(q(c))/\delta$ and acts as a monopolist of variety c. Post entry profit of the firm from domestic sales is $\pi(c)$ where $\pi(c) \equiv \max_{q(c)} [p(q(c)) - c]q(c)L - f$. The regularity conditions guarantee the monopolist's FOC is optimal and the quantity choice is given by

$$p + q \cdot u''(q)/\delta = c. \quad \text{(MR=MC)}$$

$MR = MC$ ensures that the markup rate is $(p(c) - c)/p(c) = -qu''(q)/u'(q) = \mu(q(c))$. Therefore, the elasticity of marginal utility summarizes the inverse demand elasticity as $\mu(q) \equiv |qu''(q)/u'(q)| = |d \ln p(q)/d \ln q|$.

When the economy opens to trade, firms incur an iceberg transport cost $\tau \geq 1$ and a fixed cost $f_x \geq 0$ in order to export to other countries. As a result, firms face a cost function $\text{TC}(q_x(c)) \equiv \tau c q_x(c) + f_x$ and a demand function $p(q_x(c))$ for sales to the export market. Profit from foreign sales is $\pi_x(c) \equiv \max_{q_x(c)} [p(q_x(c)) - \tau c] q_x(c) L - f_x$ and the optimal q_x choice is given by a similar $MR = MC$ condition.

---

[8]Utility functions not satisfying inada conditions are permissible but may require parametric restrictions to ensure existence. We will assume inada conditions on utility and revenue, though they are not necessary for all results.

[9]Some additional regularity conditions on G are required for existence of a market equilibrium in Melitz.

3.3 Market equilibrium

Profit maximization implies that firms produce for the domestic and/or export markets if they can earn non-negative profits from sales in the domestic and/or export markets, respectively. We denote the cutoff cost level of firms that are indifferent between producing and exiting from the domestic market as c_a in autarky and c_d in the open economy. The cutoff cost level for firms indifferent between exporting and not producing for the export market is denoted by c_x. Formally, let $\iota = a, d, x$ denote autarky and the domestic and export markets of the open home economy respectively. Each c_ι is fixed by the Zero Profit Condition (ZPC),

$$\pi_\iota(c_\iota) = 0 \qquad \text{for } \iota = a, d, x. \qquad \text{(ZPC)}$$

Since firms with cost draws higher than the cutoff level do not produce, the mass of domestic producers (M_ι) supplying to market ι is $M_\iota = M_e G(c_\iota)$.

In summary, each firm faces a two stage problem: in the second stage it maximizes profits from domestic and export sales given a known cost draw, and in the first stage it decides whether to enter given the expected profits in the second stage. We maintain the standard free entry condition imposed in monopolistic competition models. Specifically, let $\Pi(c)$ denote the total expected profit from sales in all markets for a firm with cost draw c, then ex ante average Π net of sunk entry costs must be zero,

$$\int \Pi(c) dG = f_e. \quad \text{(FE)}$$

The next two Sections examine the efficiency properties of this framework for closed and open economies.

4 Efficiency in the Closed Economy

Having described an economy consisting of heterogeneous imperfectly competitive firms, we now examine efficiency of market allocations in the closed economy. Outside of cases in which imperfect competition leads to competitive outcomes with zero profits, one would generally expect the coexistence of positive markups and positive profits to indicate inefficiency through loss of consumer surplus. Nonetheless, this Section shows that CES demand combined with the Melitz production framework exhibits positive markups and profits for surviving firms, yet it is allocationally efficient. However, we also show that the usual relationship between imperfect competition and welfare, that private incentives are not aligned with optimal production patterns, is true for all VES demand structures except CES.

4.1 Welfare under isoelastic demand

In a closed economy, a policymaker maximizes individual welfare U as given in Equation (1).[10] The policymaker is unconstrained and chooses the mass of entrants, quantities and which firms of various productivities produce. At the optimum, zero quantities will be chosen for varieties above a cost threshold c_a. Therefore, all optimal allocative decisions can be summarized by quantity $q(c)$, potential variety M_e and productivity c_a. Our approach for arriving at the optimal allocation is to think of optimal quantities $q^{\text{opt}}(c)$ as being determined implicitly by c_a and M_e so that per capita welfare can be written as

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

After solving for each q^{opt} conditional on c_a and M_e, Equation (2) can be maximized in c_a and M_e. Proposition 1 shows the market provides the first-best quantity, variety and productivity.

Proposition 1. *Every market equilibrium of a closed Melitz economy is socially optimal.*

Proof. See Appendix. □

The proof of Proposition 1 differs from standard symmetric firm monopolistic competition results because optimal quantity is a nontrivial function of unit cost, variety and cutoff productivity. As the proof is involved, we relegate details to the Appendix and discuss the rationale for optimality below.

In symmetric firm models, we know that firms charge positive markups which result in lower quantities than those implied by marginal cost pricing. However, the markup is constant so the market price (and hence marginal utility) is proportional to unit cost, ensuring proportionate reduction in quantity from the level that would be observed under marginal cost pricing (Baumol and Bradford 1970). Moreover, homogeneous firms choose price equal to average cost so the profit exactly finances the fixed cost of production. Each firm therefore internalizes the effect of higher variety on consumer surplus, resulting in an efficient market equilibrium (Grossman and Helpman 1993, Bilbiie et al. 2006).

With heterogeneous firms, markups continue to be constant, ensuring that market prices across firms are proportionate to unit costs. But, average cost pricing is too low to compensate firms for an efficient allocation, because it will not cover ex ante entry costs. The market ensures that surviving firms internalize the losses faced by exiting firms, losses which are determined by aggregate economic demand that depends on $q(c)$, c_a and M_e. Post entry, surviving firms charge prices higher than average costs ($p(c) \geq [cq(c) + f/L]/q(c)$) which compensates them for the possibility of paying f_e to enter and then being too unproductive to survive. CES demand ensures that c_a and M_e are at optimal levels that fix $p(c_a)$, thereby fixing absolute prices to optimal levels.

---

[10]Free entry implies zero expected profits so the focus is on consumer surplus.

The way in which CES preferences cause firms to optimally internalize aggregate economic conditions can be made clear by defining the elasticity of utility $\varepsilon(q) \equiv qu'(q)/u(q)$ and the social markup $1 - \varepsilon(q)$. We term $1 - \varepsilon(q)$ the social markup because at the optimal allocation, it denotes the utility from consumption of a variety net of its resource cost. At the optimal allocation, there is a multiplier λ which encapsulates the shadow cost of labor and ensures $u'(q(c)) = \lambda c$. Therefore, the social markup is

$$1 - \varepsilon(q) = 1 - u'(q)/u(q) = (u(q) - \lambda cq)/u(q). \quad \text{(Social Markup)}$$

For any optimal allocation, a quantity that maximizes social benefit from variety c solves

$$\max_{q(c)} L(1 - \varepsilon(q(c)))u(q)/\lambda - f = \max_q L\frac{1 - \varepsilon(q)}{\varepsilon(q)}cq - f.$$

In contrast, the incentives that firms face in the market are

$$\max_{q(c)} L\mu(q(c))pq - f = \max \frac{\mu(q)}{1 - \mu(q)}cq - f.$$

Since ε and μ depend only on the primitive $u(q)$, we can examine which preferences would make firms choose optimal quantities. Clearly, if $\mu(q)/(1 - \mu(q))$ is proportional to $(1 - \varepsilon(q))/\varepsilon(q)$, firms will choose optimal quantities q when they produce, but the set of producers might be smaller or larger than optimal, depending on which firms can make enough profits to clear the fixed cost f. For the market to also select the optimal range of productivity, $\mu(q)/(1 - \mu(q))$ must not only be proportional to $(1 - \varepsilon(q))/\varepsilon(q)$, but in fact be the same. Examining CES demand, we see precisely that $\mu(q)/(1 - \mu(q)) = (1 - \varepsilon(q))/\varepsilon(q)$ for all q. Thus, CES demand incentivizes exactly the right firms to produce, in addition to producing optimal quantities.

A direct implication of Proposition 1 is that laissez faire industrial policy is optimal under CES demand. This efficiency result may seem surprising in the context of Dixit and Stiglitz (1977) who find that market allocations are second-best but not first-best under CES demand for differentiated goods. Dixit and Stiglitz consider two sectors (a homogeneous goods sector and a differentiated goods sector) and assume a general utility function to aggregate across these goods. With a general utility function, the elasticity of substitution between the homogeneous and differentiated goods is not constant, leading to inefficient market allocations. In the next subsection, we examine the role of elasticities in greater detail. In keeping with Melitz, we consider a single sector to develop results for market efficiency in terms of elasticities.

4.2 Welfare beyond isoelastic demand

Efficiency of the market equilibrium in a Melitz economy is tied to CES demand. To highlight the role of CES demand, we consider the general class of variable elasticity of substitution (VES) demand studied by Dixit and Stiglitz (1977) as specified in Equation (1). With regard to efficiency, comparison of FOCs for the market and optimal allocation shows constant markups are necessary for efficiency. Therefore, within the VES class, optimality of market allocations is unique to CES preferences.[11]

Proposition 2. *Under VES demand, a necessary condition for the market equilibrium to be socially optimal is that u is CES.*[12]

Proof. Proof available upon request. □

Under general VES demand, market allocations are not efficient and do not maximize individual welfare. Proposition 3 shows that the market instead maximizes aggregate real revenue $(M_e \int u'(q(c)) \cdot q(c) \cdot L dG)$ generated in the economy.

Proposition 3. *Under VES demand, the market maximizes aggregate real revenue in the closed economy.*

Proof. See Appendix. □

Proposition 3 shows that market resource allocation is generally not aligned with the social optimum under VES demand. The market and efficient allocations are solutions to:

$$\max M_e \int_0^{c_a} u'(q(c)) \cdot q(c) dG \quad \text{where } L \geq M_e \left\{ \int_0^{c_a} [cq(c)L + f]dG + f_e \right\} \qquad \text{Market}$$

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

For CES demand, $u(q) = q^\rho$ while $u'(q)q = \rho q^\rho$ implying revenue maximization is perfectly aligned with welfare maximization. Outside of CES, quantities produced by firms are too low or too high and in general equilibrium, this implies the average productivity of operating firms is also too low or too high. Market quantity, variety and productivity reflect distortions of imperfect

---

[11]VES utility is additively separable and therefore does not include the quadratic utility of Melitz and Ottaviano (2008) and the translog utility of Feenstra (2003). However, Zhelobodko et al. (2011) show VES demand captures the qualitative features of market outcomes obtained under these forms of non-additive utility.

[12]For completeness, we note that constant elasticities of demand are necessary but not sufficient for optimality of market allocations. We extend the CES demand of Melitz to CES-Benassy preferences $U(M_e, c_a, q) \equiv v(M_e) \int_0^{c_d} q(c)^\rho g(c) dc$. In this example, u is CES but varieties and the unit bundle are valued differently through $v(M_e)$. Market allocations under CES-Benassy preferences are the same as with CES preferences of Melitz. However, firms do not fully internalize consumers' taste for variety, leading to suboptimal levels of quantity, variety and productivity. Following Benassy (1996), Bilbiie et al. (2006) and Alessandria and Choi (2007), when $v(M_e) = M_e^{\rho(v_B+1)}$, these preferences disentangle "taste for variety" v_B from the markup to cost ratio $(1-\rho)/\rho$. Market allocations are optimal only if taste for variety exactly equals the markup to cost ratio $(v_B = (1-\rho)/\rho)$.
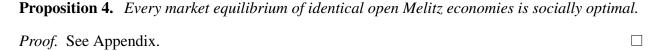
competition, and therefore, increased competition through opening markets to trade might improve efficiency. This leads us to an examination of the impact of trade on market distortions.

5   Efficiency in an Open Economy

Motivated by empirical studies of firm heterogeneity, Melitz (2003) shows that reallocation of resources towards high productivity firms provides a new source of gains from trade. In this Section, we examine how international trade affects market and optimal allocations in a Melitz economy. We start by showing that CES demand continues to induce efficient allocations in an open economy. Under VES demand, market allocations are suboptimal so we examine when market expansion from trade eventually mitigates the distortions of imperfect competition while preserving firm heterogeneity.

5.1   Welfare under isoelastic demand

Trade provides productivity gains by reallocating resources towards low cost firms, and increased productivity is gained at the expense of variety. One might therefore expect artificially selecting low cost firms to produce would improve welfare in autarky. In fact, this is not the case. Proposition 1 shows that the autarkic market equilibrium is efficient. This implies that the open economy productivity level is undesirable in autarky as it generates too little variety. However, as Proposition 4 below shows, the productivity level selected in an open economy is efficient. Thus trade itself makes a new mix of productivity and variety efficient.

Proposition 4. *Every market equilibrium of identical open Melitz economies is socially optimal.*

Proof. See Appendix.   □

Why is the higher productivity level of the open economy inefficient in autarky? Proposition 4 implies that market selection of firms is optimal if an increase in size can only be attained at a cost of exogenous frictions (τ, f_x). Compared to a frictionless world, trade frictions reduce the potential welfare gains from trade. The market minimizes the losses from frictions by weeding out high cost firms. Conditional on trade costs, market selection of firms is optimal and provides a net welfare gain from trade.

Proposition 4 is striking in that the differences in firm costs do not generate inefficiencies despite heterogeneity of profits and the different effects that trade frictions will have on firm behavior. Furthermore, selection of firms performs the function of allocating additional resources optimally without any informational requirements. Under CES demand, laissez faire industrial policy is optimal for the world economy.[13]

---

[13]However, terms of trade externalities may exist and lead to a breakdown of laissez faire policies. Demidova and

Modeling trade between equally sized countries makes the role of trade frictions extremely clear cut. When countries differ in size, trade frictions introduce cross-country distributional issues which obscure the pure efficiency question. Specifically, consider two countries of different sizes with cost distribution $G(c) = (c/c_{\max})^k$ and CES demand. Market allocations are efficient when these countries trade with each other and face no trade frictions. These market allocations maximize social welfare with equal Pareto weights assigned to every individual in the two countries. Introducing trade frictions will continue to induce efficient market allocations, but with unequal Pareto weights. This shows the market is implicitly favoring certain consumers, so that firm selection patterns reflect distributional outcomes in addition to cost competitiveness. The cross-country distribution of welfare gains is important but beyond the focus of this study. In what follows, we wish to study efficiency rather than distribution so we model the stylized case of frictionless trade and consider more general demand structures which can explain a greater range of trade effects.

5.2 Welfare beyond isoelastic demand

This subsection examines market distortions in an open economy with variable elasticities. We abstract from trade costs to focus on efficiency rather than distributional issues. The market equilibrium between freely trading countries of sizes $L_1, ..., L_n$ is identical to the market equilibrium of a single autarkic country of size $L = L_1 + ... + L_n$. Thus, opening to trade is equivalent to an increase in market size, echoing Krugman (1979). This result is summarized as Proposition 5.

Proposition 5. *In the absence of trade costs, trade between countries of sizes $L_1, ..., L_n$ has the same market outcome as a unified market of size $L = L_1 + ... + L_n$.*

Proof. Available upon request, see also Krugman (1979). □

Proposition 5 allows us to think about increased trade as an increase in market size L of a closed economy. An increase in market size has the identical effect of increased competition which will impact efficiency by altering market distortions. We turn to efficiency properties of the open VES economy, and investigate how far increased competition from trade can go towards improving market outcomes.

5.2.1 Market Efficiency under VES Demand

Having established that opening to trade is equivalent to an increase in the size of a VES economy, we can follow the same reasoning as in the closed VES economy to infer that market allocations in an open economy are suboptimal. Marginal revenues do not correspond to marginal utilities so market allocations are not aligned with efficient allocations. This is particularly important when

---

Rodriguez-Clare (2009) incorporate terms of trade considerations and provide domestic policies to obtain the first-best allocation in an open Melitz economy with Pareto cost draws. Chor (2009) also considers when policy intervention is appropriate in a heterogeneous firm model with multinationals and a homogeneous goods sector.

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.

So when does integration mitigate or exacerbate distortions? As acknowledged by Spence, "perfectly general propositions are hard to come by" and the nature of distortions can be highly dependent on parameter magnitudes. To make progress, we follow Stiglitz (1986) and first study market and optimal outcomes as market size becomes arbitrarily large. This allows us to examine when international trade enables markets to eventually mitigate distortions. Later, we also examine distortions in small markets.

5.2.2 Market Efficiency in Large Markets

Looking at efficiency in large markets explains whether integrating with world markets enables a small economy to overcome its market distortions. From a theoretical perspective we will 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.[14]

The large economy concept is similar in spirit to the idea of a competitive limit, in that the number of entrants grows unboundedly large while the quantity supplied from each firm to each worker becomes small. However, when firms are heterogeneous, simply knowing there are a large number of entrants 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 (constant absolute risk aversion).[15] 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 with heterogeneous firms, 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,* $\lim_{q \to 0} \mu(q)$ *and* $\lim_{q \to 0} \varepsilon(q) \in (0, 1)$.

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

---

[14]How large markets need to be to justify this approximation is an open quantitative question.
[15]See Behrens and Murata (2009).

per labor unit at the optimum converges to a non-zero constant (e.g., Solow 1998).[16] An example of a class of utility functions satisfying interior markups is the expo-power utility where $u(q) = [1 - \exp(-\alpha q^{1-\rho})]/\alpha$ for $\rho \in (0,1)$. It nests the CES for $\alpha = 0$.[17] When markups are interior, there is a sharp taxonomy of what may happen to the distribution of costs, prices and total quantities ($Lq(c)$) produced by a firm as follows:

Proposition 6. *Assume markups are interior. Then under the market allocation:*

1. $\lim_{L\to\infty} c_a^{\mathrm{mkt}} = \infty$ *iff* $\lim_{L\to\infty} p\left(c_a^{\mathrm{mkt}}\right) = \infty$ *iff* $\lim_{L\to\infty} Lq\left(c_a^{\mathrm{mkt}}\right) = 0$.

2. $\lim_{L\to\infty} c_a^{\mathrm{mkt}} = 0$ *iff* $\lim_{L\to\infty} p\left(c_a^{\mathrm{mkt}}\right) = 0$ *iff* $\lim_{L\to\infty} Lq\left(c_a^{\mathrm{mkt}}\right) = \infty$.

3. $\lim_{L\to\infty} c_a^{\mathrm{mkt}} \in (0,\infty)$ *iff* $\lim_{L\to\infty} p\left(c_a^{\mathrm{mkt}}\right) \in (0,\infty)$ *iff* $\lim_{L\to\infty} Lq\left(c_a^{\mathrm{mkt}}\right) \in (0,\infty)$.

Similarly, under the optimal allocation:

1. $\lim_{L\to\infty} c_a^{\mathrm{opt}} = \infty$ *iff* $\lim_{L\to\infty} u \circ q\left(c_a^{\mathrm{opt}}\right)/\lambda q\left(c_a^{\mathrm{opt}}\right) = \infty$ *iff* $\lim_{L\to\infty} Lq\left(c_a^{\mathrm{opt}}\right) = 0$.

2. $\lim_{L\to\infty} c_a^{\mathrm{opt}} = 0$ *iff* $\lim_{L\to\infty} u \circ q\left(c_a^{\mathrm{opt}}\right)/\lambda q\left(c_a^{\mathrm{opt}}\right) = 0$ *iff* $\lim_{L\to\infty} Lq\left(c_a^{\mathrm{opt}}\right) = \infty$.

3. $\lim_{L\to\infty} c_a^{\mathrm{opt}} \in (0,\infty)$ *iff* $\lim_{L\to\infty} u \circ q\left(c_a^{\mathrm{opt}}\right)/\lambda q\left(c_a^{\mathrm{opt}}\right) \in (0,\infty)$ *iff* $\lim_{L\to\infty} Lq\left(c_a^{\mathrm{opt}}\right) \in (0,\infty)$.

Proof. See Appendix. □

Proposition 6 shows that when markups are interior and the cost cutoff converges, one of three things must happen. 1) Only the lowest cost firms remain ($\lim_{L\to\infty} c_a^{\mathrm{mkt}} = 0$) and prices go to zero (akin to perfect competition), while the lowest cost firms produce infinite total quantities ($\lim_{L\to\infty} Lq\left(c_a^{\mathrm{mkt}}\right) = \infty$). 2) Post-entry, all firms produce independent of cost ($\lim_{L\to\infty} c_a^{\mathrm{mkt}} = \infty$) while prices become unbounded and the total quantities produced become negligible ($\lim_{L\to\infty} Lq\left(c_a^{\mathrm{mkt}}\right) = 0$), akin to a "rentier" case where firms produce little after fixed costs are incurred. 3) The cost cutoff converges to a positive finite level ($\lim_{L\to\infty} c_a^{\mathrm{mkt}} \in (0,\infty)$), 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 ($\lim_{L\to\infty} c_a^{\mathrm{mkt}} > 0$) but fixed costs still play a role in determining which firms produce ($\lim_{L\to\infty} c_a^{\mathrm{mkt}} < \infty$). We therefore make the following assumption, which by Proposition 6 will guarantee non-degenerate prices and total quantities:

---

[16]Kuhn and Vives (1999) find that $\lim_{q\to0} 1 - \varepsilon(q) = \lim_{q\to0} \mu(q)$ under relatively mild assumptions. Let $v = |V'(q)q/V(q)|$ where $V(q) = u(q)/q$ in our notation. Then their assumption is that there exist positive constants ρ and κ such that $\lim_{q\to0}[v(q) - v(0)]/q^\rho = \kappa$ so $v(q) - v(0)$ has an asymptotic expansion at $q = 0$ with a leading term of the constant elasticity type. Examples of such utility functions include $u(q) = q^\rho e^{\kappa q}$ for $\rho \in (0,1)$ and $\kappa < (1-\rho)/(1+\rho)\rho$, $u(q) = q^\kappa(1-q)$ for $\kappa \in (0,1)$ and $u(q) = q(1-q^\kappa)$ for $\kappa > 0$.

[17]The expo-power utility form was proposed by Saha (1993) and recently used by Holt and Laury (2002) and Post et al. (2008) to model risk aversion empirically.

Assumption (Interior Convergence)**.** *In the large economy, the market and optimum 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 a "perfectly competitive" limit or at the other extreme 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 and each variety contributes at a positive rate to utility even at low levels of quantity.

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). To obtain this result, we introduce one last regularity condition.

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

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

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 monopolistically competitive limit, summarized as Proposition 7.

Proposition 7. *Under the above assumptions, as market size L approaches infinity the market approaches the monopolistically competitive 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 $\alpha = \lim_{q \longrightarrow 0} \mu(q)$.*

4. *The entrant per worker ratio M_e/L converges.*

5. *The market and socially optimal allocations coincide.*

Proof. See Appendix. □

Proposition 7 shows that integration with large markets can push economies based on VES demand to the monopolistically competitive 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. Intuitively, we can explain Proposition 7 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.[18] This convergence to constant markups aligns perfectly with those generated by CES preferences with an exponent equal to $1 - \lim_{q \to 0} \mu(q)$. Thus, large markets reduce market distortions until they are aligned with socially optimal objectives.

It is somewhat remarkable that the large market outcome, which remains imperfectly competitive, is socially optimal. Firms charge positive markups but they exactly recover both average costs and *ex ante* entry costs. Therefore, market allocations are efficient despite positive markups. Such persistence of imperfection in 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).[19] While the monopolistically competitive limit is optimal despite imperfect competition, it is an open empirical question whether markets are sufficiently large for this to be a reasonable approximation to use in lieu of richer VES demand. When integrated markets are small, variable markups are crucial in understanding distortions. We discuss a small VES economy in the next Section.

6 Distortions and the Impact of Integration

The previous section showed how integration with sufficiently large markets can improve welfare and eliminate distortions. However, this outcome is an idealistic statement about the effects of intense competition in an otherwise imperfectly competitive market. When competitive forces are weaker, distortions remain despite integration. Although we have identified the source of distortions as a conflict between private markups $\mu(q)$ captured by firms and social markups $1 - \varepsilon(q)$ that would maximize welfare, we have not detailed the nature of these distortions. In this section we first characterize market distortions in productivity, quantity and entry. We then show that small increases in market size that fall short of large market integration may magnify distortions. Although it is reasonable to expect small increases in market size to improve welfare, additional gains can be captured using policies which mitigate distortions.

6.1 Market Distortions under VES Demand

Here we compare the market and optimal quantity, productivity and variety to understand the nature of distortions in a VES economy. We show that distortions depend on markups $\mu(q)$ and $1 - \varepsilon(q)$. Specifically, the bias in market quantity, productivity and variety is determined by how the markups vary with quantity ($\mu'(q)$ and $(1 - \varepsilon(q))'$). We start with a discussion of the relation

---

[18]The rate at which markups converge of course depends on c and is in any case highly endogenous (see Appendix).

[19]Stiglitz (1986) notes that the CES model violates the assumptions of the competitive limit of the monopolistically competitive economy derived by Hart (1985) who assumes markups are completely wiped out in the limit.

between markups and quantity, and then characterize distortions by these demand characteristics.

6.1.1 Relation between Markups and Quantity

The pattern of markups across firms in a VES economy is determined by μ' and $(1-\varepsilon)'$. When $\mu'(q) > 0$, markups are positively correlated with quantity. This is the case studied by Krugman (1979): firms are able to charge higher markups when they sell higher quantities. Our regularity conditions guarantee low cost firms produce higher quantities (Section 3.1). This means high cost firms have both high q and high markups. When $\mu'(q) < 0$, small "boutique" firms charge higher markups. For CES demand, markups are constant ($\mu' = 0$). The richer VES demand brings out the distinction between $\mu' > 0$ and $\mu' < 0$, which turns out to be important in characterizing distortions.

The sign of $(1-\varepsilon(q))'$ determines how social markups vary with quantity. When it is positive $(1-\varepsilon(q))' > 0$, social markups are higher at higher levels of quantity. As above, this implies a negative correlation between social markups $1-\varepsilon$ and unit costs c. Conversely, when $(1-\varepsilon(q))' < 0$, the "boutique" varieties which are consumed in small quantities provide relatively higher social markups. Under CES preferences, the elasticity of utility is $1-\varepsilon(q) = 1-\rho$ implying $(1-\varepsilon(q))' = 0$.

We use the relationship between markups and quantity to characterize market distortions in an open VES economy. To fix ideas, Table 1 summarizes μ' and $(1-\varepsilon)'$ for commonly used utility functions. Among the forms of $u(q)$ considered are expo-power, HARA and generalized CES (proposed by Dixit and Stiglitz).[20]

Table 1: Private and Social Markups for Common Utility Forms

	$(1-\varepsilon)' < 0$	$(1-\varepsilon)' > 0$
$\mu' > 0$	Generalized CES ($\alpha > 0$): $(q+\alpha)^\rho$	CARA, Quadratic / HARA ($\alpha > 0$): $(1-\rho)\left[(q/(1-\rho)+\alpha)^\rho - \alpha^\rho\right]/\rho$ / Expo-power ($\alpha > 0$): $\left[1-\exp\left(-\alpha q^{1-\rho}\right)\right]/\alpha$
$\mu' < 0$	HARA ($\alpha < 0$): $(1-\rho)\left[(q/(1-\rho)+\alpha)^\rho - \alpha^\rho\right]/\rho$ / Expo-power ($\alpha < 0$): $\left[1-\exp\left(-\alpha q^{1-\rho}\right)\right]/\alpha$	Generalized CES ($\alpha < 0$): $(q+\alpha)^\rho$

6.2 Quantity, Productivity and Entry Distortions

We characterize the bias in market allocations compared to the optimal allocation by demand characteristics. For ease of reference, Table 2 summarizes these biases and discussion of results

---

[20]The relevant parameter restrictions are $\rho \in (0,1)$ for each form, $q/(1-\rho)+\alpha > 0$ for HARA and $q+\alpha > 0$ for Generalized CES.

follows.

Table 2: Distortions by Demand Characteristics

	$(1-\varepsilon)' < 0$	$(1-\varepsilon)' > 0$
$\mu' > 0$	Quantities Too High: $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$ Productivity Too High: $c_d^{\text{mkt}} < c_d^{\text{opt}}$ Entry Ambiguous	Quantities Low-Cost Skewed: $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$ for $c < c^*$ $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$ for $c > c^*$ Productivity Too Low: $c_d^{\text{mkt}} > c_d^{\text{opt}}$ Entry Too Low: $M_e^{\text{mkt}} < M_e^{\text{opt}}$
$\mu' < 0$	Quantities High-Cost Skewed: $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$ for $c < c^*$ $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$ for $c > c^*$ Productivity Too High: $c_d^{\text{mkt}} < c_d^{\text{opt}}$ Entry Too High: $M_e^{\text{mkt}} > M_e^{\text{opt}}$	Quantities Too Low: $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$ Productivity Too Low: $c_d^{\text{mkt}} > c_d^{\text{opt}}$ Entry Ambiguous

We discuss quantity, productivity and entry distortions in turn. Quantity distortions across firms depend on whether private and social markups have the same relationship with quantity. We will say that private and social incentives are *partially aligned* when μ' and $(1-\varepsilon)'$ have the same sign. Conversely, incentives are *misaligned* when μ' and $(1-\varepsilon)'$ have different signs. Proposition 8 shows that when private and social markups are misaligned, market quantities $q^{\text{mkt}}(c)$ are uniformly too high or low relative to optimal quantities $q^{\text{opt}}(c)$. In contrast, Proposition 8 also shows that when private and social markups are partially aligned, the market under or over produces quantity, depending on a firm's costs.

Proposition 8. *When $(1-\varepsilon)'$ and μ' have different signs, $q^{\text{mkt}}(c)$ and $q^{\text{opt}}(c)$ never cross:*

1. *If $\mu' > 0 > (1-\varepsilon)'$, market quantities are too high: $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$.*

2. *If $\mu' < 0 < (1-\varepsilon)'$, market quantities are too low: $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$.*

In contrast, when $(1-\varepsilon)'$ and μ' have the same sign and $\inf_q \varepsilon(q) > 0$, $q^{\text{mkt}}(c)$ and $q^{\text{opt}}(c)$ have a unique crossing c^* (perhaps beyond market and optimal cost cutoffs).

1. *If $\mu' > 0$ and $(1-\varepsilon)' > 0$, $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$ for $c < c^*$ and $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$ for $c > c^*$.*

2. *If $\mu' < 0$ and $(1-\varepsilon)' < 0$, $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$ for $c < c^*$ and $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$ for $c > c^*$.*

Proof. See Appendix. □

The relationship between market and optimal quantities is fixed by FOCs for revenue maximization and welfare maximization. Specifically, the market chooses $[1 - \mu(q^{\text{mkt}})]u'(q^{\text{mkt}}) = \delta c$

and the optimal quantity is given by $u'(q^{\mathrm{opt}}) = \lambda c$. Therefore, the relationship of market and optimal quantities is:

$$\text{Private } \frac{\text{MB}}{\text{MC}} = \frac{\left[1 - \mu\left(q^{\mathrm{mkt}}\right)\right] \cdot u'\left(q^{\mathrm{mkt}}\right)/\delta}{c} = \frac{u'\left(q^{\mathrm{opt}}\right)/\lambda}{c} = \text{Social } \frac{\text{MB}}{\text{MC}}$$

When incentives are misaligned, market and optimal quantities are too high or too low across all varieties. In particular, when $\mu' > 0 > (1 - \varepsilon)'$, all firms over-produce $q^{\mathrm{mkt}}(c) > q^{\mathrm{opt}}(c)$. When $\mu' < 0 < (1 - \varepsilon)'$, market production is too low ($q^{\mathrm{mkt}}(c) < q^{\mathrm{opt}}(c)$). Firms are either over-rewarded ($\mu' > 0$) for producing q or under-rewarded ($\mu' < 0$). When incentives are aligned, the gap between δ and λ is small enough that quantities are not uniformly biased across all firms. Quantities are equal for some c^* when $1 - \mu\left(q^{\mathrm{mkt}}(c^*)\right) = \delta/\lambda$. For all other varieties, quantities are still distorted. When $\mu', (1 - \varepsilon)' > 0$, market production is biased towards low cost firms ($q^{\mathrm{mkt}} > q^{\mathrm{opt}}$ for low c and $q^{\mathrm{mkt}} < q^{\mathrm{opt}}$ for high c). The market over-rewards low cost firms who impose an externality on high cost firms. When $\mu', (1 - \varepsilon)' < 0$, the bias is reversed and quantities are biased towards high cost firms.

Proposition 8 highlights the importance of demand elasticities and firm heterogeneity for policy scope. Dixit and Stiglitz find that only the elasticity of utility matters for quantity bias and elasticity of demand is not relevant for determining efficiency of market production levels. In contrast, variable markups and firm heterogeneity fundamentally change this policy rule. Both the elasticity of utility and the inverse demand elasticity determine the bias in market quantities. Further, the bias varies across firms when markups are partially aligned.

The bias in productivity level is determined by the relation between social markups and quantity. Proposition 9 shows that productivity in the market is either too low or high, depending on whether social markups are increasing or decreasing. Revenue of the cutoff productivity firm is proportional to $u'(q)q$ while its contribution to utility is $u(q)$. Therefore, the gap in productivity cutoffs is determined by $\varepsilon(q)$ and market bias depends on $\varepsilon'(q)$. Increasing social markups $(1 - \varepsilon)' > 0$ encourage higher optimal quantity at lower costs. In general equilibrium, this translates into a lower cost cutoff at the optimum so market costs are too high.

Proposition 9. *Market productivity is too low or high, as follows:*

1. *If $(1 - \varepsilon)' > 0$, market productivity is too low: $c_a^{\mathrm{mkt}} > c_a^{\mathrm{opt}}$.*

2. *If $(1 - \varepsilon)' < 0$, market productivity is too high: $c_a^{\mathrm{mkt}} < c_a^{\mathrm{opt}}$.*

Proof. See Appendix. □

Although a comparison of market entry to optimal entry is generally hard to make, Proposition 10 establishes their relative levels for the case when private and social markups are partially aligned: market entry is too low when private markups are increasing and market entry is too high when private markups are decreasing. When incentives are misaligned, quantity and productivity

distortions have opposing effects on entry so the market entry bias depends on the magnitudes of exogenous parameters.

Proposition 10. *The market over or under produces varieties, as follows:*

1. *If $(1 - \varepsilon)', \mu' < 0$, the market has too much entry: $M_e^{\mathrm{mkt}} > M_e^{\mathrm{opt}}$.*

2. *If $(1 - \varepsilon)', \mu' > 0$ and $\mu'(q)q/\mu \leq 1$, the market has too little entry: $M_e^{\mathrm{mkt}} < M_e^{\mathrm{opt}}$.*

Proof. See Appendix. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

We have detailed the nature of distortions in an open VES economy. Next we show that small increases in market size need not lower distortions. We illustrate this with the impact of a small increase in market size on the productivity gap.

6.3 Productivity changes in the market

With variable markups, opening to trade can have positive or negative effects on productivity. Trade expands market size and has different effects on profitability across firms. We discuss these effects and show that a small increase in market size may exacerbate the productivity gap.

In a VES economy, inverse demand is $p(q(c)) = u'(q(c))/\delta$ where δ is a consumer's budget multiplier. The multiplier δ is an aggregate demand shifter that increases with market size. Differentiating the free entry condition (FE), $\int_0^{c_a} [(p(c) - c)q(c)L - f]\, dG = f_e$ with respect to L and applying the envelope theorem (and noting $\pi(c_a) = f$), we have

$$\int_0^{c_a} [(p(c) - c)q(c) + L\partial p(c)/\partial L \cdot q(c)]\, dG = 0.$$

The first term on the LHS above is the rise in profits from higher sales in a bigger market while the second term reflects the shift in the residual demand curve due to market expansion. Solving for the change in aggregate demand conditions shows

$$-\partial \ln p(c)/\partial \ln L = \int_0^{c_a} (p(c) - c)q(c)dG / \int_0^{c_a} p(c)q(c)dG.$$

Using the fact that $p(c) - c = \mu(c) \cdot p(c)$, we have

$$\partial \ln p(c)/\partial \ln L = -\int_0^{c_a} \mu(c)p(c)q(c)dG / \int_0^{c_a} p(c)q(c)dG < 0.$$

As market size expands, more firms enter so residual demand of each firm falls. The percentage fall is the average markup in the economy, and we now examine how this rise affects the ability of the cutoff firm to survive.

From the cutoff cost condition (ZPC), $(p(c_a) - c_a)q(c_a)L = f$. Differentiating with respect to L and applying the envelope theorem, we have

$$(p(c_a) - c_a)q(c_a) + L(\partial p(c_a)/\partial L - dc_a/dL)q(c_a) = 0.$$

The first terms on the LHS is the rise in profit from higher sales in a bigger market and the second term is the drop in residual demand from market expansion. Rearrangement shows

$$\partial \ln p(c_a)/\partial \ln L + (p(c_a) - c_a)/p(c_a) = (d \ln c_a/d \ln L) \cdot c_a/p(c_a). \tag{3}$$

Substituting for $\partial \ln p(c)/\partial \ln L$ and noting $\mu = (p - c)/p$ and $1 - \mu = c/p$, we see that

$$(d \ln c_a/d \ln L)(1 - \mu(c_a)) = \mu(c_a) - \int_0^{c_a} \mu(c) \cdot p(c)q(c)dG \bigg/ \int_0^{c_a} p(c)q(c)dG. \tag{4}$$

Since $\mu(c_a) \in (0,1)$, $d \ln c_a/d \ln L$ is the same sign as the RHS of Equation (4) which depends on the markup of the cutoff firm relative to the average markup in the economy. High cost firms produce lower quantities so $q(c_a)$ is the lowest quantity produced. Therefore, if $\mu'(q) > 0$, then $\mu(q(c_a)) \le \mu(q(c))$ for all c. Reading off from Equation (4) shows that $d \ln c_a/d \ln L < 0$. Conversely, when $\mu'(q) < 0$, the same argument shows $d \ln c_a/d \ln L > 0$.

The intuitive explanation is as follows. When $\mu'(q) > 0$, high productivity firms sell more q and charge higher markups. With market expansion, the rise in competition (δ) squeezes prices and the less productive firms are the least able to cushion this price drop through markups. They cannot survive so the cutoff cost level drops. Conversely, when $\mu'(q) < 0$, low productivity "boutique" firms have higher markups. With increased competition, they lower per capita quantities (q) but charge higher markups and sell to a bigger market. They have a cushion to survive import competition and the cutoff cost level rises. This is consistent with Holmes and Stevens (2010) who find small US plants were less impacted than large plants during the import surge from China.

Under CES preferences, the inverse demand elasticity is $\mu(q) = 1 - \rho$ implying $\mu'(q) = 0$. Therefore, the cutoff cost level is not affected by market size. The drop in residual demand from higher competition exactly counterbalances the higher sales to a bigger market so firm decisions are unaffected. Under VES demand, this is no longer true: the profitability of the least productive firm is affected by market size. Following market expansion from trade, productivity changes depend on whether the markup of low productivity firms provides enough cushion to absorb the downward shift in demand. We summarize these productivity changes in Proposition 11.

Proposition 11. *Increases in market size (L) change the market cost cutoff (c_a) as follows:*

1. *When private markups increase in quantity, the cutoff decreases with size.*

2. *When private markups decrease in quantity, the cutoff increases with size.*

6.3.1 Optimal productivity changes under VES demand

In a parallel fashion to the market productivity changes, we show that social markups determine the change in optimal productivity after trade. The argument is similar to the one for the impact of trade on market productivity above. Leaving calculations to the Appendix, we arrive at the following comparative static:

$$d \ln c_a / d \ln L = \varepsilon(c_a)^{-1} \left[(1 - \varepsilon(c_a)) - \int_0^{c_a} (1 - \varepsilon(c)) \cdot u(q(c)) dG / \int_0^{c_a} u(q(c)) dG \right]. \quad (5)$$

The change in the cutoff cost level depends on the social markup at the cutoff $(1 - \varepsilon(q))$ relative to the average social markup for all varieties. The sign of the RHS of Equation (5) can be determined and interpreted in a manner similar to the change in market productivity. If the social markup rises with quantity $(1 - \varepsilon(q))' > 0$, social markups are higher for higher productivity firms. Low cost firms make relatively larger contributions to welfare and the cost cutoff falls after trade. Conversely, when $(1 - \varepsilon(q))' < 0$, the cutoff cost level rises with market expansion because the "boutique" varieties which are consumed in small quantities provide relatively higher utility.

Under CES preferences, the elasticity of utility is $1 - \varepsilon(q) = 1 - \rho$ implying $(1 - \varepsilon(q))' = 0$. Constant elasticity of utility implies the optimal cost cutoff does not change with market size. This result explains why it is not optimal to select high productivity firms in the absence of trade frictions in an open Melitz economy.

In summary, the elasticity of utility characterizes optimal productivity changes after trade, as in Proposition 12. When social markups increase in quantity, low productivity varieties have no cushion against the rise in social costs because they provide lower social markups. These varieties are closed down and the optimal cost cutoff falls. When social markups decrease in quantity, low productivity varieties have a cushion against the rise in social costs because they provide higher social markups. Consequently, low productivity varieties are retained and the optimal cost cutoff rises.

Proposition 12. *Increases in market size (L) change the optimal cost cutoff (c_a) as follows:*

1. *When social markups increase in quantity, the cutoff decreases with size.*

2. *When social markups decrease in quantity, the cutoff increases with size.*

Proof. See Supplemental Appendix. □

To summarize how integration affects the gap between market and optimal allocations, Table 3 details the impact of integration on distortions by demand characteristics. Depending on the inverse demand elasticity and the elasticity of utility, productivity distortions may be mitigated or exacerbated for small increases in market size. Specifically, trade reduces the productivity gap when private and social markups move in the same direction but exacerbates it when incentives are misaligned.

Table 3: Impact of Integration on Distortions

	$(1-\varepsilon)' < 0$	$(1-\varepsilon)' > 0$
$\mu' > 0$	Productivity Diverges: $c_d^{\text{mkt}} \downarrow$ & $c_d^{\text{opt}} \uparrow$ $c_d^{\text{mkt}} < c_d^{\text{opt}}$ so Productivity Distortions Integration + Private-Social Misalignment: Productivity Distortions Magnified Scope for complementary policy	Productivity Co-moves: $c_d^{\text{mkt}} \downarrow$ & $c_d^{\text{opt}} \downarrow$ $c_d^{\text{mkt}} > c_d^{\text{opt}}$ so Possible Correction Integration + Private-Social Alignment: Productivity Distortions Partially Corrected
$\mu' < 0$	Productivity Co-moves: $c_d^{\text{mkt}} \uparrow$ & $c_d^{\text{opt}} \uparrow$ $c_d^{\text{mkt}} < c_d^{\text{opt}}$ so Possible Correction Integration + Private-Social Alignment: Productivity Distortions Partially Corrected	Productivity Diverges: $c_d^{\text{mkt}} \uparrow$ & $c_d^{\text{opt}} \downarrow$ $c_d^{\text{mkt}} > c_d^{\text{opt}}$ so Productivity Distortions Integration + Private-Social Misalignment: Productivity Distortions Magnified Scope for complementary policy

Since the theoretical implications of this paper depend heavily on the nature of demand, in particular through $\mu(q)$ and $\varepsilon(q)$, we now turn to how the theory can inform empirical work.

7 Theoretical Insights for Empirical Strategies

This paper has so far illustrated that the underlying demand structure can have very different implications for welfare and productivity distortions. While demand estimation is beyond the scope of this paper, we discuss observable differences that can distinguish different demand characteristics to enable welfare analysis. This Section details these differences for empirical work and discusses some theoretical considerations useful in designing estimation strategies.

The nature of distortions depends on how private and social markups vary with quantity. Empirical work has shown that cross-sectional variation in firm markups and their responses to market expansion can explain how private markups vary with quantity.[21] Social markups are rarely observable, and there is lack of consensus on how they respond to quantity (Vives 2001). Spence suggests social markups decrease with quantity while Dixit and Stiglitz propose increasing social markups. Therefore, the answer to this question must rely on further investigation and we discuss theoretical predictions that can guide this analysis. We begin with how observable implications of VES preferences can distinguish between different types of market demand. We then move on to distinguishing welfare properties which enables policy inference. The section concludes by contrasting productivity and welfare changes in small versus large markets.

---

[21]The bulk of empirical work on pass-through rates and firm selection suggests private markups increase with quantities. However, some studies also suggest markups decrease with quantities as they find a rise in markups after entry (see Zhelobodko et al. 2011). With direct information on prices and costs, Cunningham (2011) finds evidence for decreasing markups among pharmaceutical products.

7.1 Directly observable features of VES demand

How markups change with quantity can be distinguished in at least three ways. First, the cost cutoff for market survival can be directly observed from firm production data. Since increasing private markups imply selection of low cost firms after a rise in market size (Proposition 11), a decrease in the cost cutoff is consistent with increasing markups. Conversely, increases in the cost cutoff following market expansion is consistent with decreasing private markups.[22] We state this relationship as Remark 1.

Remark 1. Following a rise in market size, productivity increases are consistent with increasing private markups, while productivity decreases are consistent with decreasing private markups.

Another direct approach to distinguishing increasing and decreasing markups is estimation using firm pricing and production data. Obviously, increasing markups imply that markups and quantities are positively correlated (Remark 2). Since the theory implies that in a cross-section, quantity falls as unit cost increases, increasing markups also imply markups and unit costs are negatively correlated. The opposite correlations hold for decreasing markups.

Remark 2. Increasing private markups imply $\text{Cov}(\mu, q) > 0$ and $\text{Cov}(\mu, c) < 0$. Decreasing private markups imply $\text{Cov}(\mu, q) < 0$ and $\text{Cov}(\mu, c) > 0$.

Third, if sufficient data is available, the markup function $\mu(q)$ can be estimated semi-parametrically to allow $\mu'(q)$ to vary in sign. Having obtained an estimate $\hat{\mu}(q)$, one can use the VES demand structure directly. One strength of this approach is that recovering $\hat{\mu}(q)$ would allow recovery of the indirectly observable elasticity of utility $\varepsilon(q)$. In fact, $\varepsilon(q)$ and $\mu(q)$ are interrelated through the expression[23]

$$\ln \varepsilon(q)/q = \int_0^q -(\mu(t)/t)\,dt - \ln\left[\int_0^q \exp\left\{\int_0^s -(\mu(t)/t)\,dt\right\}ds\right]. \tag{6}$$

Equation (6) shows that using data on observed markups to semi-parametrically recover $\mu(q)/q$ will allow recovery of $\varepsilon(q)/q$. This fixes the demand system up to the consumer's budget multiplier δ and identifies rich productivity and welfare interrelationships as detailed in the above theory.

7.2 Indirectly observable features of VES demand

Distinguishing increasing and decreasing social markups is more challenging. For instance, we know from the theory that increasing social markups imply it is optimal to select higher productivity firms after a rise in market size, but it is hard to say when such selection would be directly

---

[22]Recently, Mrazova and Neary (2011) show $\mu' > 0$ also influences how firms select into different ways of serving a market.

[23]This equation follows from the observation that $\ln u'(q) = \int_0^q -(\mu(t)/t)\,dt + \kappa$ for some constant κ and by definition $u(0) = 0$. The change in ε is $\varepsilon' = \varepsilon[1 - \varepsilon - \mu]/q$ which can be recovered from μ and ε.

observable. Furthermore, the welfare implications of a change in trade costs no longer take the simple form provided for CES demand in Arkolakis et al. (forthcoming).[24] Consequently, for standard firm level data sets, policy inferences from productivity gains require more structure on demand.

To determine $\varepsilon(q)$ empirically, we propose modeling VES preferences in a way which nests all combinations of increasing and decreasing, private and social markups. This suggests using flexible demand systems that leave determination of these four possibilities up to the data. However, a preliminary question is whether any single demand system can generate all four possibilities. The answer is affirmative for the parametric specification of Equation (7):

$$u(q) = aq^\rho + bq^\gamma. \tag{7}$$

The VES form of Equation (7) allows all sign combinations of $\varepsilon'(q)$ and $\mu'(q)$ (shown in the Appendix).[25] This parametric approach has lower data requirements than using firm pricing and production data semi-parametrically as suggested above.

Another approach to recovering $\varepsilon(q)$ is to directly use price and quantity data. As $\varepsilon(q) = u'(q)q/u(q)$, we can use $u(q) = \int u'(q)dq$ with the initial condition $u(0) = 0$ to infer $u(q)$. Multiplying and dividing $\varepsilon(q)$ by δ, we have a function of firm prices and quantities:

$$\varepsilon(q) = (u'(q)/\delta)q / \int (u'(q)/\delta)dq = p(q)q / \int p(q)dq.$$

To recover the area under the demand curve, $\int p(q)dq$, we need to account for the fact that the observed price-quantity distribution reflects the cost distribution $G(c)$ and not the uniform quantity distribution over which the demand curve should be integrated. For instance, at the mode of the cost distribution, we will observe more price-quantity pairs but these observations over-represent the demand curve at the mode, so these observations need to be appropriately weighted when constructing a sample analog of the integral $\int p(q)dq$.

One approach to recovering $\int p(q)dq$, based on an ordered sample $\{(p_i, q_i)\}$, is to approximate

$$\int_0^{q_k} p(q)dq \approx \sum_{q_i \leq q_k} p_i \cdot (q_{i+1} - q_{i-1})/2$$

which weights each observed price by the length of the "quantity interval" $[(q_i + q_{i-1})/2, (q_{i+1} + q_i)/2]$ over which the demand curve is being integrated. Then the sample analog of the elasticity of utility

---

[24]This is shown graphically in the Appendix for the VES system of Equation (7).

[25]When $\gamma = 1$, the implied demand corresponds to an adjustable pass-through demand system (Bulow and Pfleiderer 1983; Weyl and Fabinger 2009).

is

$$\hat{\varepsilon}(q_k) = p_k q_k / \left[\sum_{q_i \le q_k} p_i \cdot (q_{i+1} - q_{i-1}) / 2 \right]. \tag{8}$$

This equation provides a first pass at recovering the elasticity of utility from firm level data without recourse to semi-parametric or non-linear methods.

Once the elasticity of utility has been recovered, we can determine whether social markups are increasing or decreasing, and thereby compare actual productivity changes with optimal changes. We summarize two distinguishing characteristics, parallel to how private markups can be distinguished, in Remark 3.

Remark 3. Increasing social markups imply $\text{Cov}(1 - \varepsilon, q) > 0$ and $\text{Cov}(1 - \varepsilon, c) < 0$. Decreasing social markups preferences imply $\text{Cov}(1 - \varepsilon, q) < 0$ and $\text{Cov}(1 - \varepsilon, c) > 0$.

A careful empirical approach can address the magnitude of distortions and identify the impact of integration. We now proceed to empirical suggestions for large markets.

7.3 The Role of Market Size

In small markets, differences from a CES approximation are likely to be more pronounced and the relationship between market and optimal outcomes can be addressed by the more detailed VES demand.

Proposition 7 shows that market allocations are efficient in large markets. A theoretical insight that may prove useful in determining how large markets need to be is the idea that markups should tend to align across firms in large markets. Although firms continue to charge positive markups, these markups become more uniform as the per capita quantity becomes negligible for each variety. Therefore decreased dispersion of markups following integration is consistent with positive steps towards the monopolistically competitive limit. We summarize this as Remark 4.

Remark 4. The monopolistically competitive limit is consistent with positive markups which become more uniform with increased market size.

Another consequence of Proposition 7 is that the distribution of firm productivity is stationary in the limit. Thus, to explain substantial productivity changes, the most promising modeling choice is VES demand in small markets far from the monopolistically competitive limit. Since market allocations are not optimal in the presence of variable elasticities, this also highlights the importance of estimating welfare and evaluating potential policies, conditional on both demand and the productivity distribution of the economy. We leave these avenues to further research and conclude in the next Section.

8   Conclusion

This paper examines the efficiency of market allocations when firms vary in productivity and markups. Generalizing the Spence-Dixit-Stiglitz framework to heterogeneous firms, the efficiency of CES demand is valid even with heterogeneous firms. Firms earn positive profits and charge prices higher than their average costs. Yet market allocations are efficient in both closed and open economies, even when trade is costly.

These findings crucially depend on CES preferences which are necessary for market efficiency. Generalizing to variable elasticities of substitution, firms charge heterogeneous markups which affect the trade off between quantity, variety and productivity. The nature of market distortions depends on the elasticity of inverse demand and the elasticity of utility. Under CES demand, these two elasticities are constant and provide strong efficiency properties, but miss out on meaningful trade offs.

Considering variable markups highlights the special role of integration as a policy tool to reduce distortions. Integration with large markets holds out the possibility of approaching the monopolistically competitive 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 empirical work might quantify these relationships and thereby exhibit the scope of integration as a tool to discipline imperfectly competitive markets.

In small markets, distortions persist even after integration. We characterize the nature of market distortions by demand characteristics, which reveals likely targets for policy. As demand conditions vary across industries, empirical work can help identify industry-specific distortions. Careful work using flexible VES demand offers a method to quantify the distortions present in imperfectly competitive markets and gauge the potential for trade to mitigate them. Future work can provide guidance on the design of implementable policies to realize further welfare gains from trade.

References

ALESSANDRIA, G. AND H. CHOI (2007): "Do Sunk Costs of Exporting Matter for Net Export Dynamics?" *The Quarterly Journal of Economics*, 122, 289–336.

ARKOLAKIS, C., A. COSTINOT, AND A. RODRIGUEZ-CLARE (forthcoming): "New trade models, same old gains?" *American Economic Review*.

ATKESON, A. AND BURSTEIN (2010): "Innovation, Firm Dynamics, and international Trade," *Journal of political economy*, 118, 433–484.

BAGWELL, K. AND R. W. STAIGER (2009): "Delocation and trade agreements in imperfectly competitive markets," *NBER Working Paper*.

BALDWIN, R. E. AND F. ROBERT-NICOUD (2008): "Trade and growth with heterogeneous firms," *Journal of International Economics*, 74, 21–34.

BAUMOL, W. J. AND D. F. BRADFORD (1970): "Optimal Departures From Marginal Cost Pricing," *The American Economic Review*, 60, 265–283.

BEHRENS, K. AND Y. MURATA (2009): "Trade, Competition, and Efficiency," *Cahier de recherche/Working Paper*, 9, 40.

BENASSY, J. P. (1996): "Taste for variety and optimum production patterns in monopolistic competition," *Economics Letters*, 52, 41–47.

BERGE, C. AND KARREMAN (1963): *Topological spaces, including a treatment of multi-valued functions, vector spaces and convexity*, New York: Macmillan.

BERNARD, A. B., J. B. JENSEN, S. J. REDDING, AND P. K. SCHOTT (2007): "Firms in International Trade," *The Journal of Economic Perspectives*, 21, 105–130.

BILBIIE, F. O., F. GHIRONI, AND M. J. MELITZ (2006): "Monopoly power and endogenous variety in dynamic stochastic general equilibrium: distortions and remedies," *manuscript, University of Oxford, Boston College, and Princeton University*.

BULOW, J. I. AND P. PFLEIDERER (1983): "A note on the effect of cost changes on prices," *The Journal of Political Economy*, 91, 182–185.

CAMPBELL, J. R. AND H. A. HOPENHAYN (2005): "Market Size Matters," *Journal of Industrial Economics*, 53, 1–25.

CHOR, D. (2009): "Subsidies for FDI: Implications from a model with heterogeneous firms," *Journal of International Economics*, 78, 113–125.

CUNNINGHAM, T. (2011): "Relative Thinking and Markups," Tech. rep., mimeo Harvard University.

DEMIDOVA, S. AND A. RODRIGUEZ-CLARE (2009): "Trade policy under firm-level heterogeneity in a small economy," *Journal of International Economics*, 78, 100–112.

DIXIT, A. K. AND J. E. STIGLITZ (1977): "Monopolistic Competition and Optimum Product Diversity," *The American Economic Review*, 67, 297–308.

FEENSTRA, R. AND H. L. KEE (2008): "Export variety and country productivity: Estimating the monopolistic competition model with endogenous productivity," *Journal of International Economics*, 74, 500–518.

FEENSTRA, R. C. (2003): "A homothetic utility function for monopolistic competition models, without constant price elasticity," *Economics Letters*, 78, 79–86.

——— (2006): "New Evidence on the Gains from Trade," *Review of World Economics*, 142, 617–641.

FOSTER, L., J. HALTIWANGER, AND C. SYVERSON (2008): "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?" *American Economic Review*, 98, 394–425.

GROSSMAN, G. M. AND E. HELPMAN (1993): *Innovation and Growth in the Global Economy*, MIT Press.

HART, O. D. (1985): "Monopolistic competition in the spirit of Chamberlin: A general model," *The Review of Economic Studies*, 52, 529.

HELPMAN, E., O. ITSKHOKI, AND S. J. REDDING (2011): "Trade and Labor Market Outcomes," *NBER Working Paper*.

HOLMES, T. J. AND J. J. STEVENS (2010): "An alternative theory of the plant size distribution with an application to trade," *NBER Working Paper*.

HOLT, C. A. AND S. K. LAURY (2002): "Risk aversion and incentive effects," *American economic review*, 92, 1644–1655.

KATAYAMA, H., S. LU, AND J. R. TYBOUT (2009): "Firm-level productivity studies: illusions and a solution," *International Journal of Industrial Organization*, 27, 403–413.

KHAN, M. A. AND Y. SUN (2002): "Non-cooperative games with many players," *Handbook of Game Theory with Economic Applications*, 3, 1761–1808.

KRUGMAN, P. (1979): "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, 9, 469–479.

——— (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70, 950–959.

KUHN, K. U. AND X. VIVES (1999): "Excess entry, vertical integration, and welfare," *The Rand Journal of Economics*, 30, 575–603.

MANKIW, N. G. AND M. D. WHINSTON (1986): "Free entry and social inefficiency," *The RAND Journal of Economics*, 48–58.

MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

MELITZ, M. J. AND G. I. P. OTTAVIANO (2008): "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75, 295–316.

MELVIN, R. D. WARNE, AND OTHERS (1973): "Monopoly and the theory of international trade," *Journal of International Economics*, 3, 117–134.

MRAZOVA, M. AND J. P. NEARY (2011): "Selection effects with heterogeneous firms," Tech. rep., Discussion paper.

POST, T., M. J. VAN DEN ASSEM, G. BALTUSSEN, AND R. H. THALER (2008): "Deal or no deal? Decision making under risk in a large-payoff game show," *The American economic review*, 98, 38–71.

SAHA, A. (1993): "Expo-power utility: A'flexible'form for absolute and relative risk aversion," *American Journal of Agricultural Economics*, 905–913.

SAMUELSON, P. A. (1967): "The monopolistic competition revolution," *Monopolistic competition theory: studies in impact*, 105–38.

SOLOW, R. M. (1998): *Monopolistic competition and macroeconomic theory*, Cambridge University Press.

SPENCE, M. (1976): "Product Selection, Fixed Costs, and Monopolistic Competition," *The Review of Economic Studies*, 43, 217–235.

STIGLITZ, J. E. (1986): "Towards a more general theory of monopolistic competition," *Prices, competition and equilibrium*, 22.

SYVERSON, C. (2004): "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 112, 1181–1222.

TROUTMAN, J. L. (1996): *Variational calculus and optimal control: Optimization with elementary convexity*, New York: Springer-Verlag.

TYBOUT, J. R. (2003): "Plant-and firm-level evidence on "new" trade theories," *Handbook of International Trade*, 1, 388–415.

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

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

WEYL, E. G. AND M. FABINGER (2009): "Pass-through as an Economic Tool," *Harvard University, mimeo*.

ZHELOBODKO, E., S. KOKOVIN, M. PARENTI, AND J. F. THISSE (2011): "Monopolistic competition in general equilibrium: Beyond the CES," *PSE Working Papers*.

A   Appendix: Proofs

A.1   Social welfare

To assess the optimality of market allocations resulting from international trade, we need to clarify the planner's objective function over different international pairings between producers and consumers. This is because every linkage between a producer in country j and a consumer in country i may encounter trade frictions distinct from one another, and a planner will factor the costs of each linkage in their decisions. We define social welfare W over allocations of goods $\{Q_{ji}\}$ produced in j and sold in country i to a worker k as

$$W\left(\{Q_{ji}\}\right) \equiv \int_{k \text{ is a worker}} \min_{i,j} \{U(Q_{ji})/\omega_{ji}\} \, dk \qquad (9)$$

where U is each worker's utility and $\omega_{ji} > 0$ is the Pareto weight for country i's consumption of goods from j.

In our setting, workers are treated identically by producers within each country. Accordingly, we constrain the social planner to provide the same allocation to all workers within a country. We identify each worker i with her country I and a country-wide Pareto weight ω_{JI} which weights utility from goods produced in J. Each country has a mass L_I of workers, which allows us to aggregate within each country and write social welfare as

$$W = \sum_{I \text{ is a country}} L_I \min_{I,J} \{U(Q_{JI})/\omega_{JI}\} = \min_{I,J} \{U(Q_{JI})/\omega_{JI}\} \cdot \sum_I L_I. \tag{10}$$

From Equation (10), dividing both sides by the world population shows any socially optimal allocation maximizes per capita welfare, using appropriate Pareto weights for each country pairing (J,I).[26] For any Pareto efficient allocation $\{Q_{JI}^*\}$, defining weights so that $\omega_{JI}/\omega_{J'I'} = U(Q_{JI}^*)/U(Q_{J'I'}^*)$ shows $\{Q_{JI}^*\}$ must maximize W (otherwise a Pareto improvement is possible). Since every Pareto efficient allocation corresponds to some set of weights $\{\omega_{ji}\}$, ranging over all admissible weights $\{\omega_{JI}\}$ sweeps out the Pareto frontier of allocations in which there is a representative worker for each country. Thus, any market allocation can be evaluated for Pareto efficiency in the usual way using Equation (10).

A.2 A Folk Theorem

In this context we need to define the Social Planner's policy space. Provided M_e and $q(c)$, and assuming without loss of generality that all of $q(c)$ is consumed, all allocations are determined. The only question remaining is what class of $q(c)$ the SP is allowed to choose from. A sufficiently rich class for our purposes are $q(c)$ which are positive and continuously differentiable on some closed interval and zero otherwise. This follows from the basic principle that a SP will utilize low cost firms before higher cost firms. Formally, we restrict q to be in sets of the form

$$\mathscr{Q}_{[0,c_d]} \equiv \{q \in \mathscr{C}^1, > 0 \text{ on } [0,c_d] \text{ and } 0 \text{ otherwise}\}.$$

We maintain Melitz's assumptions which imply a unique market equilibrium, and use the following shorthand throughout the proofs: $G(x) \equiv \int_0^x g(c)dc$, $R(x) \equiv \int_0^x c^{\rho/(\rho-1)}g(c)dc$.

Proposition. Every market equilibrium of a closed Melitz economy is socially optimal.

Proof. Assume a market equilibrium exists, which guarantees that $R(c)$ is finite for admissible c. First note that in both the market equilibrium and social planner's problem, $L/M_e = f_e + fG(c_d)$

---

implies utility of zero so in both cases $L/M_e > f_e + fG(c_d)$. The planner problem is

$$\max M_e L \int_0^{c_d} q(c)^\rho g(c)dc \text{ subject to } f_e + fG(c_d) + L\int_0^{c_d} cq(c)g(c)dc = L/M_e \qquad \text{(SP)}$$

where the maximum is taken over choices of M_e, c_d, $q \in \mathcal{Q}_{[0,c_d]}$. We will exhibit a globally optimal $q^*(c)$ for each fixed (M_e, c_d) pair, reducing the SP problem to a choice of M_e and c_d. We then solve for M_e as a function of c_d and finally solve for c_d.

Finding $q^*(c)$ for M_e, c_d fixed. For convenience, define the functionals $V(q), H(q)$ by

$$V(q) \equiv L \int_0^{c_d} v(c,q(c))dc, \qquad H(q) \equiv L \int_0^{c_d} h(c,q(c))dc$$

where $h(c,x) \equiv xcg(c)$ and $v(c,x) \equiv x^\rho g(c)$. One may show that $V(q) - \lambda H(q)$ is strictly concave $\forall \lambda$.[27] Now for fixed (M_e, c_d), consider the problem of finding q^* given by

$$\max_{q \in \mathcal{Q}_{[0,c_d]}} V(q) \text{ subject to } H(q) = L/M_e - f_e - fG(c_d). \qquad (11)$$

Following Troutman (1996), if some q^* maximizes $V(q) - \lambda H(q)$ on $\mathcal{Q}_{[0,c_d]}$ for some λ and satisfies the constraint then it is a solution to Equation (11). For any λ, a sufficient condition for some q^* to be a global maximum on $\mathcal{Q}_{[0,c_d]}$ is

$$D_2 v(c,q^*(c)) = \lambda D_2 h(c,q^*(c)). \qquad (12)$$

This follows because (12) implies for any such q^*, $\forall \xi$ s.t. $q^* + \xi \in \mathcal{Q}_{[0,c_d]}$ we have $\delta V(q^*; \xi) = \lambda \delta H(q^*; \xi)$ (where δ denotes the Gateaux derivative in the direction of ξ) and q^* is a global max since $V(q) - \lambda H(q)$ is strictly concave. The condition (12) is nothing but $\rho q^*(c)^{\rho-1} g(c) = \lambda cg(c)$ which implies $q^*(c) = (\lambda c/\rho)^{1/(\rho-1)}$.[28] From above, this q^* serves as a solution to $\max V(q)$ provided that $H(q^*) = L/M_e - f_e - fG(c_d)$. This will be satisfied by appropriate choice of λ since for fixed λ we have

$$H(q^*) = L \int_0^{c_d} (\lambda c/\rho)^{1/(\rho-1)} cg(c)dc = L(\lambda/\rho)^{1/(\rho-1)} R(c_d)$$

so choosing λ as $\lambda^* \equiv \rho \left(L/M_e - f_e - fG(c_d)\right)^{\rho-1} / L^{\rho-1} R(c_d)^{\rho-1}$ will make q^* a solution. In summary, for each (M_e, c_d) a globally optimal q^* satisfying the resource constraint is

$$q^*(c) = c^{1/(\rho-1)} \left(L/M_e - f_e - fG(c_d)\right)/LR(c_d) \qquad (13)$$

---

[27]Since h is linear in x, H is linear and since v is strictly concave in x (using $\rho < 1$) so is V.

[28]By abuse of notation we allow q^* to be ∞ at $c = 0$ since reformulation of the problem omitting this single point makes no difference to allocations or utility which are all eventually integrated.

which must be > 0 since $L/M_e - f_e - fG(c_d)$ must be > 0 as discussed at the beginning.

Finding M_e for c_d fixed. We may therefore consider maximizing $W(M_e, c_d)$ where

$$W(M_e, c_d) \equiv M_e L \int_0^{c_d} q^*(c)^\rho g(c) dc = M_e L^{1-\rho} [L/M_e - f_e - fG(c_d)]^\rho R(c_d)^{1-\rho}. \quad (14)$$

Direct investigation yields a unique solution to the FOC of $M_e^*(c_d) = (1-\rho)L/(f_e + fG(c_d))$ and $d^2 W/d^2 M_e < 0$ so this solution maximizes W.

Finding c_d. Finally, we have maximal welfare for each fixed c_d from Equation (14), explicitly $\tilde{W}(c_d) \equiv W(M_e^*(c_d), c_d)$. We may rule out $c_d = 0$ as an optimum since this yields zero utility. Solving this expression and taking logs shows that

$$\ln \tilde{W}(c_d) \;=\; \ln \rho^\rho (1-\rho)^{1-\rho} L^{2-\rho} + (1-\rho) [\ln R(c_d) - \ln (f_e + fG(c_d))].$$

Defining $B(c_d) \equiv \ln R(c_d) - \ln (f_e + fG(c_d))$ we see that to maximize $\ln \tilde{W}(c_d)$ we need maximize only $B(c_d)$. In order to evaluate critical points of B, note that differentiating B and rearranging using $R'(c_d) = c_d^{\rho/(\rho-1)} g(c_d)$ yields

$$B'(c_d) = \left\{ c_d^{\rho/(\rho-1)} - R(c_d) f / [f_e + fG(c_d)] \right\} / g(c_d) R(c_d). \quad (15)$$

Since $\lim_{c_d \to 0} c_d^{\rho/(\rho-1)} = \infty$ and $\lim_{c_d \to \infty} c_d^{\rho/(\rho-1)} = 0$ while $R(c_d)$ and $G(c_d)$ are bounded, there is a positive interval $[a,b]$ outside of which $B'(x) > 0$ for $x \le a$ and $B'(x) < 0$ for $x \ge b$. Clearly then we have $\sup_{x \in (0,a]} B(x), \sup_{x \in [b,\infty)} B(x) < \sup_{x \in [a,b]} B(x)$ and therefore any global maximum of B must occur in (a,b). Since B is continuously differentiable, at least one maximum exists in $[a,b]$ and all maxima must occur at critical points of B. From Equation (15), $B'(c_d) = 0$ iff $R(c_d)/c_d^{\rho/(\rho-1)} - G(c_d) = f_e/f$. Now for c_d that satisfy $B'(c_d) = 0$, M_e^* and q^* are determined and inspection shows the entire system corresponds to the conditions for market allocation. Therefore B has a unique critical point, which therefore is a global maximum of B, and therefore maximizes welfare. $\qquad \square$

A.3 Melitz Open Economy

Proposition. Every market equilibrium of identical open Melitz economies is socially optimal.

Proof. Following the discussion of social welfare in the text, we will show that the market allocation is Pareto efficient. Concretely, the products that j produces and are consumed by i are a triple $Q_{ji} = \left(M_e^{ji}, c_d^{ji}, q_{ji} \right)$ which provides welfare of $U(Q_{ji}) \equiv M_e^{ji} L_i \int_0^{c_d^{ji}} (q_{ji}(c))^\rho g(c) dc$. As laid out in the definition of social welfare, these j and i are representative, and the optimal allocation is one that maximizes $W \equiv \min_{i,j} \{ U(Q_{ji})/\omega_{ji} \}$ for some Pareto weights $\{ \omega_{ji} \}$. Since labor is not mobile and resources are symmetric ($L_j = L$ for all j), one can maximize W by considering the

goods produced by each country j separately. Accordingly, fix $j = 1$ so maximizing W amounts to maximizing

$$W^1 \equiv \min_i \{U(Q_{1i})/\omega_{1i}\}. \tag{16}$$

Since U is increasing (if every element of a product vector Q' is strictly greater than a product vector Q then $U(Q') > U(Q)$) it is easy to see that any $\{Q^*_{1i}\}$ that maximizes W^1 is characterized exactly by simultaneously being on the Pareto frontier while $U(Q_{1i})/U(Q_{1j}) = \omega_{1i}/\omega_{1j}$. Since Equation (16) is difficult to deal with directly, we will now maximize an additive social welfare function $\mathscr{W}^1 \equiv U(Q_{11}) + \sum_{j>1} U(Q_{1j})$. This is because any allocation which maximizes $\mathscr{W}^1$ must be Pareto efficient, as any Pareto improvement increases $\mathscr{W}^1$. Since the Pareto weights are free, at any maximum $\{Q^*_{1i}\}$ we may set $\omega_{1i} \equiv U(Q^*_{1i})$ so that $\{Q^*_{1i}\}$ maximizes Equation (16).

$\mathscr{W}^1$ must be maximized subject to a joint cost function $C(\{Q_{1i}\})$ we now detail. For brevity define the two "max" terms $\overline{M} \equiv \max_j\{M_e^{1j}\}$ and $\overline{c} \equiv \max_j\{c_d^{1j}\}$ and the "fixed" cost function $C_f(\overline{M}, \overline{c}) \equiv \overline{M}(f_e + G(\overline{c})f)$ which is incurred from fixed costs at home. Next define "variable" costs at home $C_1(Q_{11})$ and abroad $C_j(Q_{1j})$ by

$$C_1 \equiv M_e^{11} L \int_0^{c_d^{11}} cq_{11}(c)g(c)dc \quad \text{and} \quad C_j \equiv M_e^{1j} \int_0^{c_d^{1j}} \left(L\tau c q_{1j}(c) + f_x\right)g(c)dc$$

where $\tau = \tau_{ji}$ denotes the symmetric transport cost. Then total costs are given by $C(\{Q_{1i}\}) = C_f(\overline{M}, \overline{c}) + C_1(Q_{11}) + \sum_{j>1} C_j(Q_{1j})$.

Now fix $\{M_e^{1j}\}$ and $\{c_d^{1j}\}$ which fixes C_f. Also fix some allocation of labor across variable costs, say $\{\mathscr{L}_j\}$, with $C_f + \sum \mathscr{L}_j = L$, that constrain $C_j \le \mathscr{L}_j$. We may then maximize each $U(Q_{1j})$ subject to the constraint $C_j \le \mathscr{L}_j$ separately and we may assume WLOG that each $\mathscr{L}_j > 0$.[29] As in the argument for the closed economy, sufficient conditions for maximization with $\{M_e^{1j}\}$ and $\{c_d^{1j}\}$ fixed are

$$q^*_{11}(c) = c^{1/(\rho-1)}\mathscr{L}_1/M_e^{11}LR(c_d^{11}), \tag{17}$$

$$q^*_{1j}(c) = c^{1/(\rho-1)}[\mathscr{L}_j/M_e^{1j} - f_xG(c_d^{1j})]/LR(c_d^{1j})\tau. \tag{18}$$

Having found the optimal quantities of Equations (17-18) in terms of finite dimensional variables, we now prove existence of an optimal allocation. Note that for any fixed pair $(\overline{M}, \overline{c})$, the remaining choice variables are restricted to a compact set $K(\overline{M}, \overline{c})$ so that continuity of the objective function (by defining $U(Q_{1j}) = 0$ when $\mathscr{L}_j = 0$) guarantees existence of a solution and we

---

[29]If $\mathscr{L}_j = 0$ for all j then autarkic allocations are optimal, and as shown above the optimal autarkic allocation coincides with the market. Any set of exogenous parameters which result in trade imply welfare beyond autarky, so if countries trade in the market equilibrium, $\mathscr{L}_j = 0$ for all j cannot be optimal. Inada type conditions on $U(Q_{1j})$ imply that if it is optimal to have at least one $\mathscr{L}_j > 0$ then all $\mathscr{L}_j$ are > 0.

denote the value of $\mathscr{W}^1$ at the maximum by $S(\overline{M},\overline{c})$. In fact, $K(\overline{M},\overline{c})$ can be shown to be a continuous correspondence, so by the Theorem of the Maximum $S(\overline{M},\overline{c})$ is continuous on $C_f^{-1}([0,L])$ (Berge and Karreman, 1963). Since C_f is continuous, $C_f^{-1}([0,L])$ is compact and therefore a global max of $S(\overline{M},\overline{c})$ exists. Therefore there is an allocation that maximizes $\mathscr{W}^1$ which we now proceed to characterize.

Now evaluating welfare at the quantities of Equations (17-18) yield respectively

$$U(Q_{11}) = R(c_d^{11})^{1-\rho}L^{1-\rho}M_e^{11}\left(\mathscr{L}_1/M_e^{11}\right)^\rho, \tag{19}$$

$$U(Q_{1j}) = R(c_d^{1j})^{1-\rho}L^{1-\rho}M_e^{1j}\left(\mathscr{L}_j/M_e^{1j} - f_x G(c_d^{1j})\right)^\rho \tau^{-\rho}. \tag{20}$$

Equation (19) is increasing in both M_e^{11} and c_d^{11} so it follows that at any optimum, $M_e^{11*} = \overline{M}$ and $c_d^{11*} = \overline{c}$. Equation (20) is first increasing in M_e^{1j}, attains a critical point at $(1-\rho)\mathscr{L}_j/f_x G(c_d^{1j})$ and is then decreasing, so at any optimum $M_e^{1j*} = \min\left\{(1-\rho)\mathscr{L}_j/f_x G(c_d^{1j}),\overline{M}\right\}$. If $c_d^{1j*} < \overline{c}$ then the first order necessary condition implies

$$M_e^{1j} = (1-\rho)\mathscr{L}_j/f_x\left(\rho R(c_d^{1j})/\left(c_d^{1j}\right)^{\rho/(\rho-1)} + (1-\rho)G(c_d^{1j})\right) < (1-\rho)\mathscr{L}_j/f_x G(c_d^{1j})$$

so $c_d^{1j*} < \overline{c}$ implies $M_e^{1j*} = \overline{M}$ and $M_e^{1j*} < \overline{M}$ implies $c_d^{1j*} = \overline{c}$. Ruling out the latter case, $M_e^{1j*} < \overline{M}$ implies $U(Q_{1j}) = \tau^{-\rho}L^{1-\rho}(1-\rho)^{1-\rho}\rho^\rho \mathscr{L}_j f_x^{\rho-1}\left(R(c_d^{1j})/G(c_d^{1j})\right)^{1-\rho}$ which is decreasing in c_d^{1j} so $c_d^{1j*} = \overline{c}$ cannot be optimal. Therefore we conclude that $M_e^{1j*} = \overline{M}$ and $c_d^{1j*} < \overline{c}$. In particular, c_d^{1j*} must solve the implicit equation

$$\rho R(c_d^{1j*})/\left(c_d^{1j*}\right)^{\rho/(\rho-1)} + (1-\rho)G(c_d^{1j*}) = (1-\rho)\mathscr{L}_j/\overline{M}f_x \tag{21}$$

derived from the first order necessary condition.

With these results in hand, $\mathscr{W}^1$ reduces to

$$\mathscr{W}^1 = \left(\overline{M}L\right)^{1-\rho}\left\{R(\overline{c})^{1-\rho}\mathscr{L}_1^\rho + \tau^{-\rho}\sum_{j>1}R(c_d^{1j})^{1-\rho}\left(\mathscr{L}_j - \overline{M}f_x G(c_d^{1j})\right)^\rho\right\}. \tag{22}$$

Now consider maximizing $\mathscr{W}^1$ as given in Equation (22) over $\overline{M},\overline{c},\mathscr{L}_j,c_d^{1j}$ with c_d^{1j} unconstrained by $\overline{c}$ for $j > 1$. Using a standard Lagrangian approach, the candidate solution from the necessary conditions implies $c_d^{1j*} = (f_x/f)^{(\rho-1)/\rho}\overline{c}/\tau$ and since it is assumed $(f/f_x)^{(1-\rho)/\rho} < \tau$ for trade in a market equilibrium in the Melitz framework, $c_d^{1j*} < \overline{c}$. The candidate solution with c_d^{1j} unconstrained also yields Equation (21) so the unconstrained candidate solution coincides with the solution including the omitted constraints $c_d^{1j*} < \overline{c}$. We conclude the necessary conditions embodied in the candidate solution are also necessary to maximize $\mathscr{W}^1$ with constraints. Since these necessary conditions are exactly those which fix the unique market allocation, the market allocation

maximizes $\mathcal{W}^1$. □

A.4 Results Regarding the Impact of Large Markets

Lemma. *As market size becomes large:*

1. *Under the market, revenue is increasing in market size and goes to infinity.*

2. *Under the optimum, utility per capita is increasing in market size and goes to infinity.*

3. *Market entry goes to infinity.*

Proof. From above, the market allocation solves

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

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 $\tilde{q}(c) = (L/\tilde{L}) \cdot q(c) \longrightarrow 0$ for all $c > 0$ and $\lim_{q \to 0} u'(q) = \infty$, revenue per capita goes to infinity as $\tilde{L} \longrightarrow \infty$. 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 $\delta \longrightarrow \infty$ and $\mu(q(c))$ is bounded, it must be that $u'(q(c)) \longrightarrow \infty$ for $c > 0$. This requires $q(c) \longrightarrow 0$ 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 $L \geq \tilde{L}$:

$$u'(q(c)) q(c) \mathbf{1}_{[0,c_d]}(c) \leq u'(\tilde{q}(c)) \tilde{q}(c) \mathbf{1}_{[0,\tilde{c}_d]}(c) + u'(\tilde{q}(\tilde{c}_d)) \tilde{q}(\tilde{c}_d) \mathbf{1}_{[\tilde{c}_d,\infty]}(c) \equiv B(c) \quad (23)$$

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 $\int_0^\infty B(c) dG = \int_0^{\tilde{c}_d} u'(\tilde{q}(c)) \tilde{q}(c) dG + u'(\tilde{q}(c_d)) \tilde{q}(c_d) < \infty$. Since $u'(q(c)) q(c)$ converges pointwise to zero for $c > 0$, we conclude $\lim_{L \to \infty} \int_0^{c_d} u'(q(c)) q(c) dG = \int_0^{c_d} \lim_{L \to \infty} u'(q(c)) q(c) dG = 0$ by dominated convergence. Therefore $\lim_{L \to \infty} \delta/M_e = 0$ which with $\delta \longrightarrow \infty$ shows $M_e \longrightarrow \infty$. The optimal allocation case is similar. □

Lemma. *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

$$\pi(c_L) \geq (p(c_H) - c_L)q(c_H)L - f = \pi(c_H) + (c_H - c_L)q(c_H)L. \tag{24}$$

Furthermore, for every $\tilde{c} > 0$, we argue that $\pi(\tilde{c})$ is bounded. For $\underline{c} \equiv \tilde{c}/2$, $\pi(\tilde{c}) \leq \pi(\underline{c})$ while $\pi(\underline{c})$ is bounded since $\lim_{L\to\infty} \int_0^{c_d} \pi(c)dG = F_e$ and $\limsup_{L\to\infty} \pi(\underline{c}) = \infty$ would imply $\limsup_{L\to\infty} \int_0^{c_d} \pi(c)dG = \infty$. It follows from Equation (24) that $Lq(c)$ is bounded. Substituting ϖ for π leads to similar arguments for the social optimum. $\qquad\square$

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

1. $\lim_{L\to\infty} c_a^{\text{mkt}} = \infty$ iff $\lim_{L\to\infty} p\left(c_a^{\text{mkt}}\right) = \infty$ iff $\lim_{L\to\infty} Lq\left(c_a^{\text{mkt}}\right) = 0$.

2. $\lim_{L\to\infty} c_a^{\text{mkt}} = 0$ iff $\lim_{L\to\infty} p\left(c_a^{\text{mkt}}\right) = 0$ iff $\lim_{L\to\infty} Lq\left(c_a^{\text{mkt}}\right) = \infty$.

3. $\lim_{L\to\infty} c_a^{\text{mkt}} \in (0,\infty)$ iff $\lim_{L\to\infty} p\left(c_a^{\text{mkt}}\right) \in (0,\infty)$ iff $\lim_{L\to\infty} Lq\left(c_a^{\text{mkt}}\right) \in (0,\infty)$.

Similarly, under the optimal allocation:

1. $\lim_{L\to\infty} c_a^{\text{opt}} = \infty$ iff $\lim_{L\to\infty} u \circ q\left(c_a^{\text{opt}}\right)/\lambda q\left(c_a^{\text{opt}}\right) = \infty$ iff $\lim_{L\to\infty} Lq\left(c_a^{\text{opt}}\right) = 0$.

2. $\lim_{L\to\infty} c_a^{\text{opt}} = 0$ iff $\lim_{L\to\infty} u \circ q\left(c_a^{\text{opt}}\right)/\lambda q\left(c_a^{\text{opt}}\right) = 0$ iff $\lim_{L\to\infty} Lq\left(c_a^{\text{opt}}\right) = \infty$.

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

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

$$u'\left(q\left(c_a\right)\right)/\delta - f/\left[Lq\left(c_a\right) \cdot \mu \circ q\left(c_a\right)/\left(1 - \mu \circ q\left(c_a\right)\right)\right] = c_a, \tag{25}$$

$$Lc_a q\left(c_a\right) \cdot \mu \circ q\left(c_a\right)/\left(1 - \mu \circ q\left(c_a\right)\right) = f. \tag{26}$$

First, if $\lim_{L\to\infty} Lq\left(c_a\right) = 0$, Equation (26) implies $c_a \cdot \mu \circ q\left(c_a\right)/\left(1 - \mu \circ q\left(c_a\right)\right) \longrightarrow \infty$. Clearly $q\left(c_a\right) \longrightarrow 0$ and since $\lim_{q\to 0} \mu\left(q\right) \in (0,1)$, $\mu \circ q\left(c_a\right)/\left(1 - \mu \circ q\left(c_a\right)\right)$ is bounded, and therefore $c_a \longrightarrow \infty$. Now suppose $c_a \longrightarrow \infty$ and since $c_a \leq u'\left(q\left(c_a\right)\right)/\delta$, $u'\left(q\left(c_a\right)\right)/\delta \longrightarrow \infty$. Finally, if $u'\left(q\left(c_a\right)\right)/\delta \longrightarrow \infty$, since $\delta \longrightarrow \infty$, necessarily $q\left(c_a\right) \longrightarrow 0$ so $\mu \circ q\left(c_a\right)/\left(1 - \mu \circ q\left(c_a\right)\right)$ is bounded. It follows from Equation (26) that $Lc_a q\left(c_a\right)$ is bounded, so from Equation (25), $Lq\left(c_a\right) \cdot u'\left(q\left(c_a\right)\right)/\delta$ is bounded so $Lq\left(c_a\right) \longrightarrow 0$.

If $\lim_{L\to\infty} Lq(c_a) = \infty$, $q(c_a) \longrightarrow 0$ so from $\lim_{q\to 0} \mu(q) \in (0,1)$, $\mu \circ q(c_a)/(1 - \mu \circ q(c_a))$ is bounded. Therefore from Equation (26), $c_a \longrightarrow 0$. Now assume $c_a \longrightarrow 0$ so from Equation (26), $Lq(c_a) \cdot \mu \circ q(c_a)/(1 - \mu \circ q(c_a)) \longrightarrow \infty$ which implies with Equation (25) that $u'(q(c_a))/\delta \longrightarrow 0$. Finally, if $u'(q(c_a))/\delta \longrightarrow 0$, Equation (25) shows $c_a \longrightarrow 0$.

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

Lemma. *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, $u \circ q(c)/\lambda q(c)$ converges in $(0,\infty)$ for $c > 0$, $Lq(c_d)$ converges in $(0,\infty)$.*

Proof. Since $q(c) \longrightarrow 0$ for all $c > 0$, $\lim_{q\to 0} \mu(q) \in (0,1)$ shows $\lim_{L\to\infty} p(c)$ 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. *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 $(q(c)/q(c_d))_{a_n} \to a$ and $(q(c)/q(c_d))_{b_n} \to b$. 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

$$\left(u'(q(c))/u'(q(c_d))\right)_{a_n} > \left(u'(\kappa q(c_d))/u'(q(c_d))\right)_{a_n} > \left(u'(\tilde{\kappa}q(c_d))/u'(q(c_d))\right)_{b_n}$$
$$> \left(u'(q(c))/u'(q(c_d))\right)_{b_n}.$$

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

$$\lim_{n\to\infty} \left(u' \circ q(c)/u' \circ q(c_d)\right)_{a_n} = \lim_{n\to\infty} ([1 - \mu \circ q(c)]c/[1 - \mu \circ q(c_d)]c_d)_{a_n} = c/c_d$$
$$= \lim_{n\to\infty} \left(u' \circ q(c)/u' \circ q(c_d)\right)_{b_n}$$

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

For the social optimum, we could repeat this argument (substituting $\varepsilon \neq 0$ for $u'' < 0$ where appropriate) so long as

$$\kappa \neq \tilde{\kappa} \text{ implies } \lim_{q \longrightarrow 0} \left(u(\kappa q)/\kappa q \right) / \left(u(q)/q \right) \neq \lim_{q \longrightarrow 0} \left(u(\tilde{\kappa}q)/\kappa q \right) / \left(u(q)/q \right). \tag{27}$$

Since $\lim_{q \to 0} u'(q) = \infty$ and $\lim_{q \to 0} \varepsilon \in (0, \infty)$ it follows that $\lim_{q \to 0} u(q)/q = \infty$. By L'Hospital's rule, $\lim_{q \longrightarrow 0} \left(u(\kappa q)/\kappa q \right) / \left(u(q)/q \right) = \lim_{q \to 0} u'(\kappa q)/u'(q)$ for all κ so the condition (27) in holds because $\kappa \neq \tilde{\kappa}$ implies $\lim_{q \to 0} u'(\kappa q)/u'(q) \neq \lim_{q \to 0} u'(\tilde{\kappa}q)/u'(q)$. $\qquad\square$

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

1. *If $\lim_{q \to 0} 1 - \varepsilon(q) < 1$ then $\lim_{q \to 0} 1 - \varepsilon(q) = \lim_{q \to 0} \mu(q)$.*

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

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

$$\lim_{q \longrightarrow 0} \varepsilon(q) = \lim_{q \longrightarrow 0} q / \left(u(q)/u'(q) \right) \qquad = \lim_{q \longrightarrow 0} 1 / \lim_{q \longrightarrow 0} \left(1 - u(q)u''(q)/(u'(q))^2 \right)$$

$$= 1/ \lim_{q \longrightarrow 0} \left(1 + \mu(q)/\varepsilon(q) \right) = \lim_{q \longrightarrow 0} \varepsilon(q) / \lim_{q \longrightarrow 0} \left(\varepsilon(q) + \mu(q) \right)$$

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

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

1. *$q(c)/q(c_d) \longrightarrow (c/c_d)^{-1/\alpha}$ with $\alpha = \lim_{q \to 0} \mu(q)$.*

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)

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

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 $\lim_{q \to 0} f_q(z) = \Upsilon^{-1}(z)$ for all $\Upsilon^{-1}(z) \in (0, 1)$. Note

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

40

so since $\lim_{q\to 0}\mu(zq) = \mu^\infty \in (0,1)$ and $\lim_{q\to 0}u'(zq)/zu'(q) = \Upsilon^{-1}(z)/z$, we know $\lim_{q\to 0}f_q'(z) = -\mu^\infty \Upsilon^{-1}(z)/z$.

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 $q \longrightarrow 0$. It follows (Rudin's Principles, Thm 7.17) that

$$\lim_{q\to 0}f_q'(z) = d\lim_{q\to 0}f_q(z)/dz = -\mu^\infty \Upsilon^{-1}(z)/z = d\Upsilon^{-1}(z)/dz. \tag{28}$$

We conclude that $\Upsilon^{-1}(z)$ is differentiable and thus continuous, and given the form deduced in (28), $\Upsilon^{-1}(z)$ is continuously differentiable. Since $d\Upsilon^{-1}(z)/dz = 1/\Upsilon' \circ \Upsilon^{-1}(z)$, composing both sides with $\Upsilon(z)$ and using Equation (28) 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 $L \longrightarrow \infty$ 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 $c > 0$:

$$q^{\text{opt}}(c)/q^{\text{opt}}\left(c_d^{\text{opt}}\right) \longrightarrow (c/c_\infty^{\text{opt}})^{-1/\alpha} \text{ and } q^{\text{mkt}}(c)/q^{\text{mkt}}\left(c_d^{\text{mkt}}\right) \longrightarrow \left(c/c_\infty^{\text{mkt}}\right)^{-1/\alpha}. \tag{29}$$

Now consider the parallel conditions involving F_e for the market and social optimum, $\int_0^{c_d^{\text{mkt}}}\pi(c)dG = F_e = \int_0^{c_d^{\text{opt}}}\varpi(c)dG$. Expanding these we see that

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

It necessarily follows that

$$\lim_{L\to\infty}L\int_0^{c_d^{\text{mkt}}}\mu \circ q^{\text{mkt}}(c)/\left(1-\mu \circ q^{\text{mkt}}(c)\right)\cdot cq^{\text{mkt}}(c)dG - fG(c_d^{\text{mkt}}) =$$
$$\lim_{L\to\infty}L\int_0^{c_d^{\text{opt}}}\left(1-\varepsilon \circ q^{\text{opt}}(c)\right)/\varepsilon \circ q^{\text{opt}}(c)\cdot cq^{\text{opt}}(c)dG - fG(c_d^{\text{opt}}). \tag{30}$$

Using Equation (29), 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 $\lim_{q\to 0}\mu(q) = \lim_{q\to 0}1-\varepsilon(q)$, we see from Equation (30) the limits of the $\mu/(1-\mu)$ and $(1-\varepsilon)/\varepsilon$ terms are equal and factor out of Equation (30), leaving

$$\lim_{L\to\infty}Lc_\infty^{\text{mkt}}q^{\text{mkt}}(c_\infty^{\text{mkt}})\int_0^{c_d^{\text{mkt}}}(c/c_\infty^{\text{mkt}})(c/c_d^{\text{mkt}})^{-1/\alpha}dG - fG(c_d^{\text{mkt}}) =$$
$$\lim_{L\to\infty}Lc_\infty^{\text{opt}}q^{\text{opt}}(c_\infty^{\text{opt}})\int_0^{c_d^{\text{opt}}}(c/c_\infty^{\text{opt}})(c/c_d^{\text{opt}})^{-1/\alpha}dG - fG(c_d^{\text{opt}}).$$

Noting $f(1-\mu^\infty)/\mu^\infty = Lc_\infty^{\mathrm{mkt}}q^{\mathrm{mkt}}(c_\infty^{\mathrm{mkt}}) = Lc_\infty^{\mathrm{opt}}q^{\mathrm{opt}}(c_\infty^{\mathrm{opt}})$, we therefore have

$$\lim_{L \longrightarrow \infty} \int_0^{c_d^{\mathrm{mkt}}} (c/c_\infty^{\mathrm{mkt}})^{1-1/\alpha}(c_\infty^{\mathrm{mkt}}/c_d^{\mathrm{mkt}})^{-1/\alpha}dG - G(c_d^{\mathrm{mkt}}) =$$

$$\lim_{L \longrightarrow \infty} \int_0^{c_d^{\mathrm{opt}}} (c/c_\infty^{\mathrm{opt}})^{1-1/\alpha}(c_\infty^{\mathrm{opt}}/c_d^{\mathrm{opt}})^{-1/\alpha}dG - G(c_d^{\mathrm{opt}})$$

so that finally evaluating the limits, we have

$$\int_0^{c_\infty^{\mathrm{mkt}}} \left[(c/c_\infty^{\mathrm{mkt}})^{1-1/\alpha} - 1\right]dG = \int_0^{c_\infty^{\mathrm{opt}}} \left[(c/c_\infty^{\mathrm{opt}})^{1-1/\alpha} - 1\right]dG. \tag{31}$$

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 $c_\infty^{\mathrm{opt}} = c_\infty^{\mathrm{mkt}}$ such that Equation (31) holds. Checking the conditions for L/M_e show they coincide between the market and social optimum as well. $\qquad\square$

A.5   Static Distortion Results

Lemma. *For sufficiently high fixed costs, the quantities produced by all firms are close to the maximum quantity produced ($q(0)$).*

Proof. To clarify, we wish to show for sufficiently high f, for any producing firm with cost $c \le c_a$, either $|q(0) - q(c)|$ is arbitrarily small in the case that $q(0)$ is finite, otherwise when $q(0) = \infty$, $q(c)$ grows large. For both of these cases, we need only consider the impact of f on $q(c_a)$ since our assumptions imply it is the lowest quantity produced and the quantity $q(0)$ is unaffected by f in the market or social optimum. Furthermore, both cases hold iff as $f \longrightarrow \infty$, $\delta c_a \longrightarrow 0$ because (considering the market case, similar to the optimum case) we have

$$u'(q(c_a))[1 - \mu(q(c_a))] = \delta c_a$$

and the LHS is marginal revenue which is decreasing in quantity. Since δ is also equal to marginal revenue per capita which by above is maximized by the market, δ is decreasing in f. Direct comparative statics also show that c_a is decreasing in f. Therefore if either δ or $c_a \longrightarrow 0$ we are done and WLOG both δ and c_a are bounded away from 0 (at least on a subsequence, which monotonicity forces to be true on the whole sequence). In particular, $d\delta/df = -\delta G(c_a)M_e/L$ and since $\delta \ge 0$, necessarily $d\delta/df \longrightarrow 0$ which implies $M_e \longrightarrow 0$. Finally, $\delta = M_e \int_0^{c_a} u'(q(c))q(c)dG$ and as δ is bounded away from zero and $M_e \longrightarrow 0$, we conclude $\int_0^{c_a} u'(q(c))q(c)dG \longrightarrow \infty$. Noting that $u'(q(c))q(c) = \varepsilon(q(c)) \cdot u(q(c))$, since c_a is bounded away from zero and G is a probability distribution, $\int_0^{c_a} u'(q(c))q(c)dG \longrightarrow \infty$ implies $q(c) \longrightarrow q(0)$ for $c \in [0, \kappa]$ for some $\kappa > 0$. However this contradicts δc bounded away from zero as $u'(q(c))[1 - \mu(q(c))] = \delta c$. We conclude at

least one of δ or $c_d \longrightarrow 0$, giving the result. $\qquad\square$

Proposition. *Market productivity is too low or high, as follows:*

1. *If $(1-\varepsilon)' > 0$, market productivity is too low: $c_d^{\mathrm{mkt}} > c_d^{\mathrm{opt}}$.*

2. *If $(1-\varepsilon)' < 0$, market productivity is too high: $c_d^{\mathrm{mkt}} < c_d^{\mathrm{opt}}$.*

Proof. For $\alpha \in [0,1]$, define $v_\alpha(q) \equiv \alpha u'(q)q + (1-\alpha)u(q)$ and also define $w(q) \equiv u'(q)q - u(q)$ so $v_\alpha(q) = u(q) + \alpha w(q)$. Consider the continuum of maximization problems (indexed by α) defined as:

$$\max_{M_e, c_d, q(c)} LM_e \int_0^{c_d} v_\alpha(q(c))\, dG \text{ subject to } L \geq M_e \left(\int_0^{c_d} Lcq(c) + f dG + F_e \right). \tag{32}$$

Let the Lagrange multiplier associated with each α in Equation (32) be written as $\beta(\alpha)$. By appealing to the envelope theorem and differentiating Equation (32) in M_e we have $\beta(\alpha) = M_e \int_0^{c_d} v_\alpha(q(c))\, dG$ and that $d\beta/d\alpha = M_e \int_0^{c_d} w(q(c))\, dG = M_e \int_0^{c_d} u(q(c))[\varepsilon(q)-1]\, dG < 0$. The conditions characterizing the solution to every optimum also imply

$$\beta(\alpha) = v_\alpha(q(c_d)) / (c_d q(c_d) + f/L)$$

whereby we arrive at

$$
\begin{aligned}
dv_\alpha(q(c_d))/d\alpha &= (d\beta/d\alpha)(v_\alpha(q(c_d))/\beta) + \beta((dc_d/d\alpha)q(c_d) + c_d(dq(c_d)/d\alpha)) \\
&= w(q(c_d)) + v_\alpha'(q(c_d))(dq(c_d)/d\alpha) \\
&= w(q(c_d)) + \beta c_d(dq(c_d)/d\alpha)
\end{aligned}
$$

so cancellation and rearrangement, using the expressions for β, $d\beta/d\alpha$ above shows

$$
\begin{aligned}
\beta q(c_d)(dc_d/d\alpha) &= w(q(c_d)) - (v_\alpha(q(c_d))/\beta)(d\beta/d\alpha) \\
&= w(q(c_d)) - \left(v_\alpha(q(c_d))/M_e \int_0^{c_d} v_\alpha(q(c))\, dG \right) \cdot M_e \int_0^{c_d} w(q(c))\, dG.
\end{aligned}
$$

We conclude that $dc_d/d\alpha \gtrless 0$ when $w(q(c_d)) \int_0^{c_d} v_\alpha(q(c))\, dG \gtrless v_\alpha(q(c_d)) \int_0^{c_d} w(q(c))\, dG$. Expanding this inequality we have (suppressing $q(c)$ terms in integrands):

$$w(q(c_d)) \int_0^{c_d} u\, dG + \alpha w(q(c_d)) \int_0^{c_d} w\, dG \gtrless u(q(c_d)) \int_0^{c_d} w\, dG + \alpha w(q(c_d)) \int_0^{c_d} w\, dG.$$

Cancellation and expansion again then show this is equivalent to

$$u'(q(c_d))q(c_d) \int_0^{c_d} u\, dG \gtrless u(q(c_d)) \int_0^{c_d} u'q(c)\, dG.$$

43

Finally, this expression can be rewritten $\varepsilon(q(c_d)) \gtrless \int_0^{c_d} \varepsilon(q(c)) u(q(c)) dG / \int_0^{c_d} u(q(c)) dG$ and since $q(c)$ is strictly decreasing in c, we see $dc_d/d\alpha \gtrless 0$ when $\varepsilon' \lessgtr 0$. Note that Equation (32) shows $\alpha = 0$ corresponds to the social optimum while $\alpha = 1$ corresponds to the market equilibrium. It follows that when $\varepsilon' < 0$ that $dc_d/d\alpha > 0$ so we have $c_d^{\text{mkt}} > c_d^{\text{opt}}$ and vice versa for $\varepsilon' > 0$. $\qquad\square$

Proposition. *When $(1-\varepsilon)'$ and μ' have different signs, $q^{\text{mkt}}(c)$ and $q^{\text{opt}}(c)$ never cross:*

1. *If $\mu' > 0 > (1-\varepsilon)'$, market quantities are too high: $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$.*

2. *If $\mu' < 0 < (1-\varepsilon)'$, market quantities are too low: $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$.*

In contrast, when $(1-\varepsilon)'$ and μ' have the same sign and $\inf_q \varepsilon(q) > 0$, $q^{\text{mkt}}(c)$ and $q^{\text{opt}}(c)$ have a unique crossing c^* (perhaps beyond market and optimal cost cutoffs).

1. *If $\mu' > 0$ and $(1-\varepsilon)' > 0$, $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$ for $c < c^*$ and $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$ for $c > c^*$.*

2. *If $\mu' < 0$ and $(1-\varepsilon)' < 0$, $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$ for $c < c^*$ and $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$ for $c > c^*$.*

Proof. This result relies heavily on the following relationship which we first prove:

$$\overline{\sigma} \equiv \sup_{c \leq c_a^{\text{mkt}}} \varepsilon\left(q^{\text{mkt}}(c)\right) > \delta/\lambda > \inf_{c \leq c_a^{\text{opt}}} \varepsilon\left(q^{\text{opt}}(c)\right) \equiv \underline{\sigma}. \tag{33}$$

To see this recall $\delta = M_e^{\text{mkt}} \int_0^{c_d^{\text{mkt}}} u'\left(q^{\text{mkt}}(c)\right) q^{\text{mkt}}(c) dG$ so $\overline{\sigma} > \delta/\lambda$ because

$$\delta/\overline{\sigma} = M_e^{\text{mkt}} \int_0^{c_d^{\text{mkt}}} \left(\varepsilon\left(q^{\text{mkt}}(c)\right)/\overline{\sigma}\right) u\left(q^{\text{mkt}}(c)\right) dG < M_e^{\text{mkt}} \int_0^{c_d^{\text{mkt}}} u\left(q^{\text{mkt}}(c)\right) dG \tag{34}$$

and λ is the maximum welfare per capita so $\lambda > M_e^{\text{mkt}} \int_0^{c_d^{\text{mkt}}} u\left(q^{\text{mkt}}(c)\right) dG > \delta/\overline{\sigma}$. A similar argument shows $\lambda\underline{\sigma} < \delta$, giving Equation (33).

Now note that

$$\left[u''\left(q^{\text{mkt}}(c)\right) q^{\text{mkt}}(c) + u'\left(q^{\text{mkt}}(c)\right)\right]/\delta = c, \qquad u'\left(q^{\text{opt}}(c)\right)/\lambda = c. \tag{35}$$

And it follows from Equations (35) we have

$$\left[1 - \mu\left(q^{\text{mkt}}(c)\right)\right] \cdot u'\left(q^{\text{mkt}}(c)\right) / u'\left(q^{\text{opt}}(c)\right) = \delta/\lambda. \tag{36}$$

Suppose $\mu' > 0 > (1-\varepsilon)'$, and it is sufficient to show $\inf_{c \leq c_a^{\text{mkt}}} 1 - \mu\left(q^{\text{mkt}}(c)\right) \geq \overline{\sigma}$, since then Equations (33) and (36) show that $u'\left(q^{\text{mkt}}(c)\right) < u'\left(q^{\text{opt}}(c)\right)$ which implies $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$. Since $\mu' > 0 > (1-\varepsilon)'$ and by assumption $\lim_{c\to 0} q^{\text{mkt}}(c) = \infty = \lim_{c\to 0} q^{\text{opt}}(c)$,

$$\inf_{c \leq c_a^{\text{mkt}}} 1 - \mu\left(q^{\text{mkt}}(c)\right) = \lim_{q\to\infty} 1 - \mu(q) = \lim_{q\to\infty} \varepsilon(q) + \varepsilon'(q) q/\varepsilon(q) \geq \lim_{q\to\infty} \varepsilon(q) = \overline{\sigma}.$$

Similarly, if $\mu' < 0 < (1 - \varepsilon)'$ one may show that $\sup_{c \leq c_a^{\text{mkt}}} 1 - \mu\left(q^{\text{mkt}}(c)\right) \leq \underline{\sigma}$, implying from Equations (33) and (36) that $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$.

Now consider the cases when μ' and ε' have different signs, and since $\inf_q \varepsilon(q) > 0$, from above in both cases it holds that $\inf_{q>0} 1 - \mu(q) = \inf_{q>0} \varepsilon(q)$ and $\sup_{q>0} 1 - \mu(q) = \sup_{q>0} \varepsilon(q)$. The arguments above have shown that $\sup_{q>0} \varepsilon(q) > \delta/\lambda > \inf_{q>0} \varepsilon(q)$ and therefore

$$\sup_{q>0} 1 - \mu(q) > \delta/\lambda > \inf_{q>0} 1 - \mu(q).$$

It follows from Equation (36) that for some c^*, $1 - \mu\left(q^{\text{mkt}}(c^*)\right) = \delta/\lambda$ and therefore $u'\left(q^{\text{mkt}}(c^*)\right) = u'\left(q^{\text{opt}}(c^*)\right)$ so $q^{\text{mkt}}(c^*) = q^{\text{opt}}(c^*)$. Furthermore, $q^{\text{mkt}}(c)$ is strictly decreasing in c so with $\mu' \neq 0$, c^* is unique. Returning to Equation (36), using the fact that $q^{\text{mkt}}(c)$ is strictly decreasing in c also shows the relative magnitudes of $q^{\text{mkt}}(c)$ and $q^{\text{opt}}(c)$ for $c \neq c^*$. $\square$

Proposition. *The market over or under produces varieties, as follows:*

1. *If $(1 - \varepsilon)', \mu' < 0$, the market has too much entry: $M_e^{\text{mkt}} > M_e^{\text{opt}}$.*

2. *If $(1 - \varepsilon)', \mu' > 0$ and $\mu'(q)q/\mu \leq 1$, the market has too little entry: $M_e^{\text{mkt}} < M_e^{\text{opt}}$.*

Proof. For any preferences v, defining $\varepsilon_v(q) \equiv v'(q)q/v(q)$ and $\mu_v(q) \equiv -v''(q)q/v'(q)$ it holds that at any social optimum that

$$1/M_e = \int_0^{c_d} cq(c)/\varepsilon_v(q(c)) \, dG(c)$$

Defining $B_v(c) \equiv cq(c)/\varepsilon_v(q(c))$ which is the integrand of the equation above, we have

$$B_v'(c) = q(c)/\varepsilon_v(q(c)) + c\left(dq(c)/dc\right)\left[1 - \varepsilon_v'(q(c))q(c)/\varepsilon_v(q(c))\right]/\varepsilon_v(q(c)). \tag{37}$$

Equation (37) can be considerably simplified using two relationships. The first is

$$1 - \varepsilon_v'(q(c))q(c)/\varepsilon_v(q(c)) = \varepsilon_v(q(c)) + \mu_v(q(c)).$$

The second is that manipulating the necessary conditions shows that $dq(c)/dc = -\left(q(c)/c\right) \cdot \left(1/\mu_v(q(c))\right)$. Substituting these relationships into Equation (37) yields

$$B_v'(c) = q(c)/\varepsilon_v(q(c)) \cdot \left[1 - \left[\varepsilon_v(q(c)) + \mu_v(q(c))\right]/\mu_v(q(c))\right] = -q(c)/\mu_v(q(c)).$$

Now consider that the social planner problem corresponds to $v(q) = u(q)$ while the market allocation is generated by maximizing $v(q) = u'(q)q$ so that (suppressing the c argument to q in

integrands)

$$1/M_e^{\text{opt}} - 1/M_e^{\text{mkt}} = \int_0^{c_d^{\text{opt}}} cq^{\text{opt}}/\varepsilon\left(q^{\text{opt}}\right) dG(c) - \int_0^{c_d^{\text{mkt}}} cq^{\text{mkt}}/\left[1 - \mu\left(q^{\text{mkt}}\right)\right] dG \qquad (38)$$

and similarly (suppressing the c arguments):

$$B_u = cq^{\text{opt}}/\varepsilon\left(q^{\text{opt}}\right), \qquad\qquad B_u' = -q^{\text{opt}}/\mu\left(q^{\text{opt}}\right),$$
$$B_{u'q} = cq^{\text{mkt}}/\left[1 - \mu\left(q^{\text{mkt}}\right)\right], \ B_{u'q}' = -q^{\text{mkt}}/\left[\mu\left(q^{\text{mkt}}\right) + \mu'\left(q^{\text{mkt}}\right) q^{\text{mkt}}/\left(1 - \mu\left(q^{\text{mkt}}\right)\right)\right].$$

Now assume $\varepsilon' < 0 < \mu'$, so by above $c_d^{\text{mkt}} > c_d^{\text{opt}}$ and for the result, from Equation (38) it is sufficient to show that $\int_0^{c_d^{\text{opt}}} B_u(c) - B_{u'q}(c) dG(c) \leq 0$. From above, there is also a c^* such that $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$ for $c < c^*$ and $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$ for $c > c^*$. For $c < c^*$, $B_u(c) - B_{u'q}(c) < 0$ as $q^{\text{mkt}}(c) > q^{\text{opt}}(c)$ and $\varepsilon' < 0$ implies

$$cq^{\text{mkt}}/\left[1 - \mu\left(q^{\text{mkt}}\right)\right] > cq^{\text{opt}}/\left[1 - \mu\left(q^{\text{opt}}\right)\right] > cq^{\text{opt}}/\varepsilon\left(q^{\text{opt}}\right).$$

For $c \geq c^*$, $B_u(c) \leq B_{u'q}(c)$ as from continuity $B_u(c^*) \leq B_{u'q}(c^*)$, while $\mu' > 0$ implies

$$\left(B_u(c) - B_{u'q}(c)\right)' = -q^{\text{opt}}/\mu\left(q^{\text{opt}}\right) + q^{\text{mkt}}/\left[\mu\left(q^{\text{mkt}}\right) + \mu'\left(q^{\text{mkt}}\right) q^{\text{mkt}}/\left(1 - \mu\left(q^{\text{mkt}}\right)\right)\right]$$
$$< -q^{\text{opt}}/\mu\left(q^{\text{opt}}\right) + q^{\text{mkt}}/\mu\left(q^{\text{mkt}}\right).$$

Finally, $\mu'(q)q/\mu \leq 1$ implies $q/\mu(q)$ is increasing in q. With $q^{\text{mkt}}(c) < q^{\text{opt}}(c)$ for $c > c^*$, this implies $\left(B_u(c) - B_{u'q}(c)\right)' \leq 0$ so $B_u(c) \leq B_{u'q}(c)$ for $c > c^*$. Put together with above, $\int_0^{c_d^{\text{opt}}} B_u(c) - B_{u'q}(c) dG(c) \leq 0$ giving the result. For the case $\varepsilon' > 0 > \mu'$, the same argument goes through since clearly $\mu'(q)q/\mu(q) \leq 1$. $\qquad\square$

A.6 VES Specific Utility

The VES demand system implied by $u(q) = aq^\rho + bq^\gamma$ can generate all four combinations of increasing and decreasing, private and social markups as we now briefly discuss. First, note that
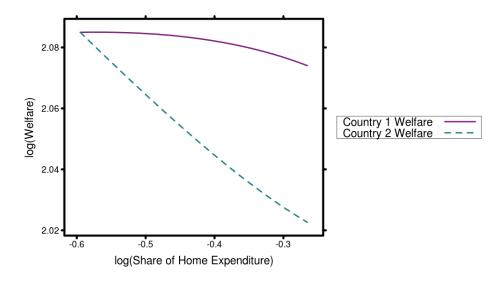
$$\varepsilon'(q) = ab(\rho - \gamma)^2 q^{\rho - \gamma - 1}/\left(aq^{\rho - \gamma} + b\right)^2,$$
$$\mu'(q) = -ab\rho\gamma(\rho - \gamma)^2 q^{\rho - \gamma - 1}/\left(a\rho q^{\rho - \gamma} + b\gamma\right).$$

For $\rho = \gamma$, $\varepsilon'(q) = \mu'(q) = 0$ and we are in a CES economy. For $\rho \neq \gamma$, sign $\varepsilon'(q) = $ sign ab and sign $\mu'(q) = $ sign $-ab \cdot \rho\gamma$, exhibiting all four combinations for appropriate parameter values. In addition, this demand system does not exhibit the log-linear relationship between welfare and share of expenditure on home goods discussed in Arkolakis et al. (forthcoming), as shown in Figure 1

for $u(q) = q^{1/2} + q^{1/4}$.

Figure 1: Welfare and Share of Home Expenditure as Home Tariff Increases

B   Online Appendix

B.1   VES Market Allocation

Proposition. *The market equilibrium, when unique, maximizes aggregate real revenue in the economy. Formally, the market allocation solves*

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

Proof. Consider a social planner who faces a utility function $v(q) \equiv u'(q)q$. Provided $v(q)$ satisfies the regularity conditions used in the proof of optimality, it follows that the following conditions characterize the unique constrained maximum of $LM_e \int_0^{c_d} u'(q(c)) q(c) dG$, where δ denotes the Lagrange multiplier:

$$u''(q(c)) q(c) + u'(q(c)) = \delta c,$$
$$u'(q(c_d)) q(c_d) / (c_d q(c_d) + f/L) = \delta,$$
$$\int_0^{c_d} u'(q(c)) q(c) dG / \left(\int_0^{c_d} [cq(c) + f/L] dG + F_e/L \right) = \delta,$$
$$M_e \left(\int_0^{c_d} Lcq(c) + fdG + F_e \right) = L.$$

Comparing these conditions, we see that if δ is the same as under the market allocation, the first three equations respectively determine each firm's optimal quantity choice, the ex post cost cutoff,

47

and the zero profit condition while the fourth is the resource constraint and must hold under the market allocation. Therefore if this system has a unique solution, the market allocation maximizes $LM_e \int_0^{c_d} u'(q(c))q(c)dG$. Since these conditions completely characterize every market equilibrium, the assumed uniqueness of the market equilibrium guarantees such a unique solution. $\quad\square$

B.2 VES Utility

Proposition. Increases in market size (L) change the optimal cost cutoff (c_a) as follows: When $(1 - \varepsilon(q))' > 0$, the cost cutoff decreases with size. When $(1 - \varepsilon(q))' < 0$, the cost cutoff increases with size.

Proof. Let the normalized resource constraint R be defined as

$$R \equiv 1 - M_e \left(\int_0^{c_a} cq(c)dG + G(c_a)f/L + f_e/L \right).$$

The social planner maximizes $M_e \int_0^{c_a} u(q(c))dG + \lambda R$ where λ is the shadow value of an extra unit of resources. The optimality conditions for the three outcomes of quantity, mass of varieties and cost cutoff determine the optimal allocations along with the resource constraint $R = 0$.

Optimal quantity equates the marginal social benefit to the marginal social cost implying $u'(q(c)) = \lambda c$. The FOC for optimal M_e with the binding resource constraint implies $\int_0^{c_a} u(q(c))dG = \lambda(1-R)/M_e = \lambda/M_e$. The FOC for the optimal cost cutoff shows that the welfare contribution of the marginal variety is equal to its per capita shadow cost, $u(q(c_a)) = \lambda(c_aq(c_a) + f/L)$.

Differentiating the cost cutoff equation wrt to L shows

$$\left(u'(q(c_a)) - \lambda c_a\right)(dq(c_a)/dL) - (c_aq(c_a) + f/L)(d\lambda/dL) - \lambda q(c_a)(dc_a/dL) + \lambda f/L^2 = 0.$$

Substituting for $u'(q(c)) = \lambda c$ and multiplying through by L/λ, we have

$$(c_aq(c_a) + f/L)(d\ln\lambda/d\ln L) + c_aq(c_a)(d\ln c_a/d\ln L) = f/L. \tag{39}$$

Equation (39) shows dc_a/dL is tied to $d\lambda/dL$. Changes in the cost cutoff depend on how the shadow value of labor changes with market size, namely $d\lambda/dL$. Differentiating the M_e FOC wrt to L and rearranging shows

$$d\ln\lambda/d\ln L = d\ln M_e/d\ln L$$
$$+ LM_e \int_0^{c_a} c(dq(c)/dL)dG + LM_e(c_aq(c_a) + f/L)g(c_a)(dc_a/dL). \tag{40}$$

The binding resource constraint shows $0 = d(1-R)/dL$ and substituting for Equation (40) implies $d\ln\lambda/d\ln L - M_e(G(c_a)f/L + f_e/L) = 0$. The shadow value of labor rises with market size and

the percentage rise in λ reflects the better amortization of fixed and sunk costs in bigger markets. Using the expression for $R = 0$, we have $d \ln \lambda / d \ln L = 1 - M_e \int_0^{c_a} cq(c) dG$. Substituting this into Equation (39) and rearranging gives

$$d \ln c_a / d \ln L = [c_a q(c_a) / (c_a q(c_a) + f/L)]^{-1} \left[M_e \int_0^{c_a} cq(c) dG - c_a q(c_a) / (c_a q(c_a) + f/L) \right]$$

$$= [c_a q(c_a) / (c_a q(c_a) + f/L)]^{-1} \int_0^{c_a} u'(q(c)) \cdot q(c) dG / \int_0^{c_a} u(q(c)) dG - 1$$

where the second line follows from $u'(q(c)) = \lambda c$ and the M_e FOC. Substituting for the elasticity of utility $\varepsilon \equiv qu'(q)/u(q)$, we have Equation (5). $\qquad \square$

B.3 Large Market Results

Lemma. *Integration will select a few highly productive firms for the market and social optimum under the following conditions:*

1. *If $\lim_{q \to 0} \mu(q) = 0$ then $c_d^{\text{mkt}} \longrightarrow 0$ as $L \longrightarrow \infty$.*

2. *If $\lim_{q \to 0} 1 - \varepsilon(q) = 0$ then $c_d^{\text{opt}} \longrightarrow 0$ and $c_d^{\text{mkt}} \longrightarrow 0$ as $L \longrightarrow \infty$.*

Proof. Considering the market case, by above for each fixed $c > 0$, $Lq(c)$ and $\pi(c)$ are bounded in L. Since $\pi(c) = L\mu(q(c))/(1 - \mu(q(c))) \cdot cq(c)$, and $\mu(q(c))/(1 - \mu(q(c))) \longrightarrow 0$ (because necessarily $q(c) \longrightarrow 0$), $\pi(c) \longrightarrow 0$ as $L \longrightarrow \infty$ so $c_d^{\infty} < c$ for each $c > 0$, implying $c_d^{\text{mkt}} \longrightarrow 0$. The case for the social optimum is similar, with the addition that $\lim_{q \to 0} 1 - \varepsilon(q) < 1$ implies $\lim_{q \to 0} \mu(q) = 0$, giving the result. $\qquad \square$

B.4 Trade and Market Size

Proposition. In the absence of trade costs, trade between countries with identical VES demand and with sizes $L_1, ..., L_n$ has the same market outcome as a unified market of size $L = L_1 + ... + L_n$.

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 economy 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 and productivity level is symmetric across home and foreign consumers while entry 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

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

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 $\pi + e\pi_x = (p(c_a) - c_a) q(c_a)L + e(p_x(c_a) - c_a) q_x(c_a)L^* = f$. Substituting for optimal q^* and q_x^* in the analogous foreign cost cutoff condition implies $c_a = c_a^*$. From the resource constraint, this fixes the relationship between entry across countries as $L/M_e = \int_0^{c_a}[cq(c) + cq_x(c) + f]dG + f_e = L^*/M_e^*$. 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 $1 = M_e \int pq dG + M_e^* \int_0^{c_a} p_x^* q_x^* dG$, which from $L/M_e = L^*/M_e^*$ is equivalent to

$$L/M_e = L \int pq dG + L^* \int_0^{c_a} p_x^* q_x^* dG = L \int pq dG + L/M_e - L \int_0^{c_a} p_x q_x dG.$$

Therefore to show the consumer's problem is consistent, it is sufficient to show the expenditure on home goods is equal to expenditure on exported goods ($\int pq dG = \int_0^{c_a} p_x q_x dG$), 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

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

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. □

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1116	Ghazala Azmat Nagore Iriberri	The Provision of Relative Performance Feedback Information: An Experimental Analysis of Performance and Happiness
1115	Pascal Michaillat	Fiscal Multipliers over the Business Cycle
1114	Dennis Novy	Gravity Redux: Measuring International Trade Costs with Panel Data
1113	Chiara Criscuolo Ralf Martin Henry G. Overman John Van Reenen	The Causal Effects of an Industrial Policy
1112	Alex Bryson Richard Freeman Claudio Lucifora Michele Pellizzari Virginie Perotin	Paying for Performance: Incentive Pay Schemes and Employees' Financial Participation
1111	John Morrow Michael Carter	Left, Right, Left: Income and Political Dynamics in Transition Economies
1110	Javier Ortega Gregory Verdugo	Assimilation in Multilingual Cities
1109	Nicholas Bloom Christos Genakos Rafaella Sadun John Van Reenen	Management Practices Across Firms and Countries
1108	Khristian Behrens Giordano Mion Yasusada Murata Jens Südekum	Spatial Frictions
1107	Andrea Ariu Giordano Mion	Service Trade and Occupational Tasks: An Empirical Investigation
1106	Verónica Amarante Marco Manacorda Edward Miguel Andrea Vigorito	Do Cash Transfers Improve Birth Outcomes? Evidence from Matched Vital Statistics, Social Security and Program Data
1105	Thomas Sampson	Assignment Reversals: Trade, Skill Allocation and Wage Inequality
1104	Brian Bell Stephen Machin	Immigrant Enclaves and Crime